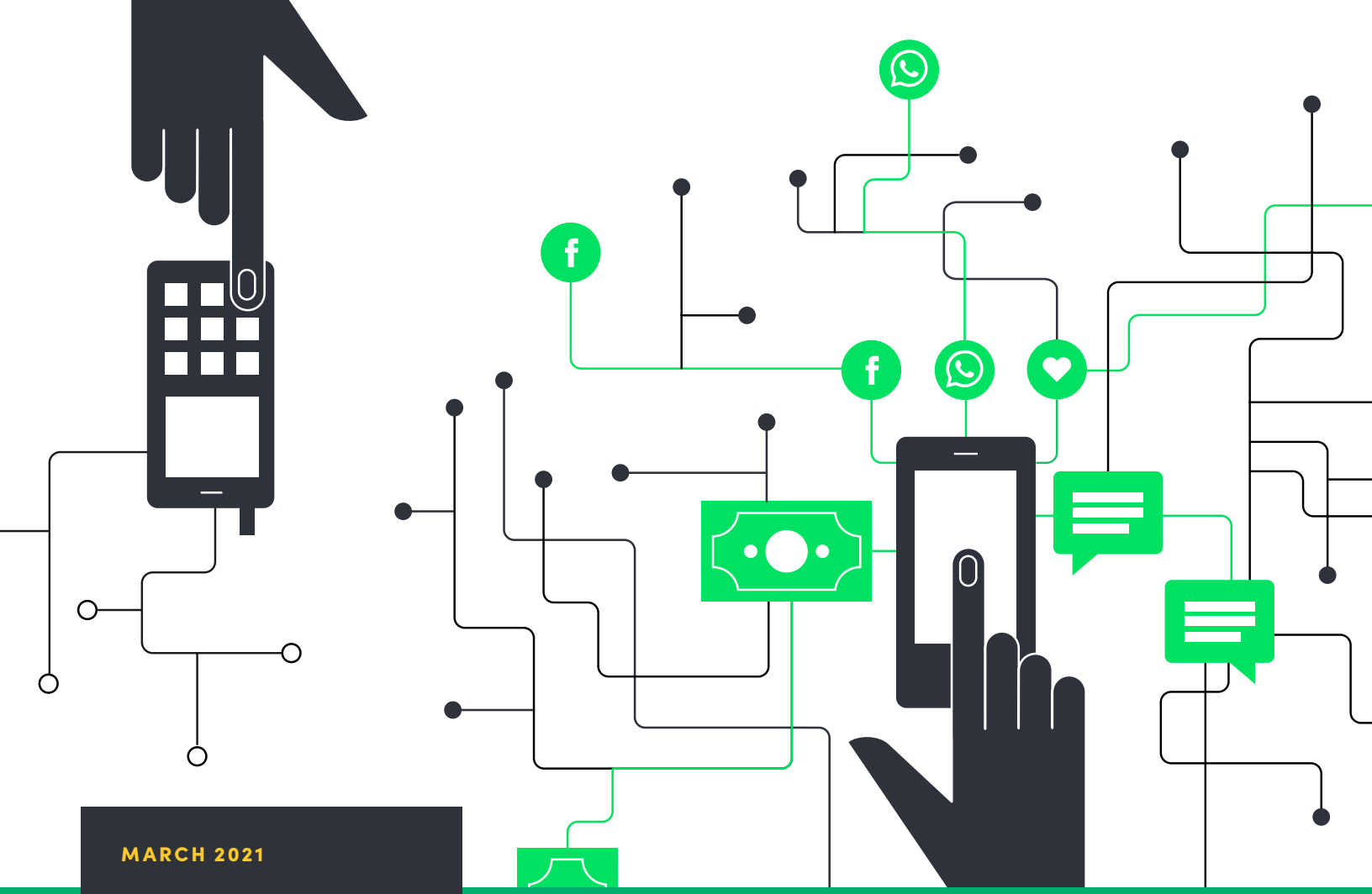


MARCH 2021

AUTHORS

Alexandra Rizzi, Alexandra Kessler, and Jacobo Menajovsky

The Stories Algorithms Tell: Bias and Financial Inclusion at the Data Margins

CENTER *for* FINANCIAL INCLUSION | ACCION

Acknowledgements

CFI completed this work as part of our partnership with FMO to advance responsible digital finance.

We would like to acknowledge the many people who took the time to give us helpful insights and feedback during the course of this work, including David Hernández-Velázquez, Thelma Brenes Munoz, and Mitzi Padilla at FMO; Salmana Ahmed at Luminate; Amy Paul and Paul Nelson at USAID; Paul Randall at CreditInfo; Eric Duflos at CGAP; Stephan Dreyer and Florian Wittner at the Hans-Bredow-Institut for Media Research; Andrew Selbst at UCLA; Reema Patel and Carly Kind at the Ada Lovelace Institute; Sarayu Natarajan at the Aapti Institute; Rafael Zanatta and Bruno Bioni at Data Privacy Brazil; Tarunima Prabhakar at Carnegie India; independent consultant Jayshree Venkatesan; and Aaron Riecke at Upturn. We'd also like to thank the regulators and fintechs whom we interviewed—thank you for your candor and openness on a sensitive topic.

Introduction: New Visibilities, New Stories

Algorithms—mathematical recipes ranging from the simple to the complex—have a long history in the field of banking.[i] But in recent years, several trends have converged to supercharge their application, especially in emerging markets. The growth in mobile phone ownership and internet use continues to march ahead; by the time you finish reading this paper, more than 3,500 new users from emerging markets will be on the internet, largely through their mobile devices.[ii] Average internet use, as measured from any type of device, is staggering: 9 hours and 45 minutes per day in the Philippines, 9 hours and 17 minutes in Brazil, and 6 hours and 30 minutes in India, with more than a third of that time on social media.[1] Digitalization in the wake of the COVID-19 pandemic, in part encouraged by governments through temporary reductions in mobile money fees, has further pushed consumers into using their mobile devices as financial tools. In Rwanda, for instance, this resulted in a doubling, *within two weeks*, of unique mobile money subscribers sending a P2P transfer, from 600,000 to 1.2 million.[2]

The "data fumes" generated from the seismic increases in digital activity have found a home in ever-increasing computational power as well as advanced algorithms and machine learning techniques. These practical superpowers are being applied by financial service providers and regulators alike with the intention of lowering costs, expanding economic opportunity, and improving how markets function.[3] The applications are seemingly boundless, from customer segmentation, product design, marketing, and portfolio monitoring to underwriting, ID verification, fraud detection, and collection.[4] For example, the Mexican National Banking and Securities Commission recently built machine-learning models to enhance its anti-money laundering supervision over financial technology companies (fintechs). Their model flags suspicious transactions, clients or reports—flags that feed into individual and on-site supervisory reports for follow-up.[5] Natural Language Processing (NLP) and other AI-powered techniques allow providers to leverage chatbots to address customer problems 24/7. The opportunities have ushered in highly skilled technologists, data scientists, and engineers who build internal data infrastructure as well as test, prototype, monitor, and tweak models.

Across all industries, predictive, data-driven algorithms are being used to tell stories about individuals and, depending on how they are wielded, can drive high-stakes decisions: who receives a loan, what sentencing a judge will recommend, what therapeutics a doctor will provide. The exploding data ecosystem has created billions of new stories for financial service providers; at the Center for Financial Inclusion (CFI) we are most interested in the ones they try to tell (or don't tell) about low-income consumers.

In this paper, we explore the stories algorithms can tell about who is creditworthy in emerging markets, the risks of that narrative for those it leaves out, and what it all might mean for inclusive finance. As data ethicist Professor David Robinson writes, "There's often a gap between how much of a person's story an algorithm can tell, and how much we want it to tell."[6] We have two main objectives: a) to ground some of the universal challenges on the use of algorithms, automated decisions, alternative data, and bias

i For example, international credit cards have long used scores to immediately recommend what type of credit card to offer customers. ("From Catalogs to Clicks: The Fair Lending Implications of Targeted, Internet Marketing")

ii Using 2018-2019 data from https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx on individuals using the internet in developing markets, we calculated approximately 48.33 new users per minute and use a reading rate of 200 words per minute.

in the context of inclusive financial services; and b) to present the current state of play among inclusive finance actors from desk research and interviews with a sample of fintechs, regulators, and other experts. It is aimed at the stakeholders that can influence the trajectory of the inclusive finance industry, with specific recommendations for regulators, investors, and donors. Our broader goal is to break down silos between data science teams and those that view themselves in non-technical positions while playing a crucial role in shaping investments, business processes, partnerships, staff composition, project scope, and legal frameworks.

Exploring Algorithms and Bias in Inclusive Finance

IMPROVEMENTS ON THE STATUS QUO

When designed and used to maximize benefits, algorithm-driven decisions can counter human biases and increase the speed and accuracy of disbursing appropriate loans to people who need them but were previously denied access to credit. Algorithms have the potential to overcome some of the entrenched implicit and explicit biases of face-to-face interactions. In India, mystery shopping audits showed that individual bank staff can strongly influence financial access, even when regulation and eligibility rules should not give such discretion.[7] A U.S.-based study conducted by the Haas School of Business found that fintech algorithms discriminated 40 percent less on average than loan officers in loan prices, and the algorithms did not discriminate at all in accepting and rejecting loans.[8] At CFI, we share in the inclusive finance community's optimism for the power of increased digitalization, data processing capabilities, and troves of data trails to increase financial inclusion.

BIAS IS A UNIVERSAL CONCERN

However, the pace of change and the opacity of the technology has outstripped the ability of most in the sector to understand potential risks and issues. Underwriting, and many other operational functions within financial services, are being digitized and increasingly automated. Whether it's a decision-supporting algorithm or a decision-making algorithm, humans are less in control than ever before.

Issues have cropped up with real-world consequences and harms, across all sectors. The now-infamous AppleCard (a partnership between Goldman Sachs and Apple) came under investigation by financial regulators for discrimination against women when complaints surfaced that for couples with comparable credit scores, husbands had received 10 to 20 times the credit limit of their wives.[9] The U.S. Department of Housing and Urban Development (HUD) filed a lawsuit against Facebook in 2019 for violations of the Fair Housing Act by limiting a person's housing choices based on protected characteristics. The suit alleged that Facebook allowed its advertising algorithms to exclude housing ads for people classified as parents, non-Christian, or interested in Hispanic culture; it also alleged that through its massive collection of online and offline data and machine learning techniques, Facebook recreated groups defined by their protected class.[10] An algorithm used by commercial healthcare providers to identify individuals for "high-risk care management" programs recommended that white patients receive more comprehensive care than equally sick black patients.[11] Carnegie Mellon researchers uncovered that, despite treating gender as a sensitive attribute, Google's ad listings for high-earning positions were shared with men at almost six times the rate they were presented to women.[12]

The scale of harm or exclusion that could be wrought by a discriminatory algorithm dwarfs that of a biased individual; in economics literature this distinction is known as statistical vs. taste-based discrimination, respectively.[13] For instance, in the healthcare example, the flawed algorithm was applied commercially to over 200 million people annually.[14] How do these misfires happen? We categorize the issues into three buckets: inputs, code, and context.[iii]

INPUTS, CODE, AND CONTEXT

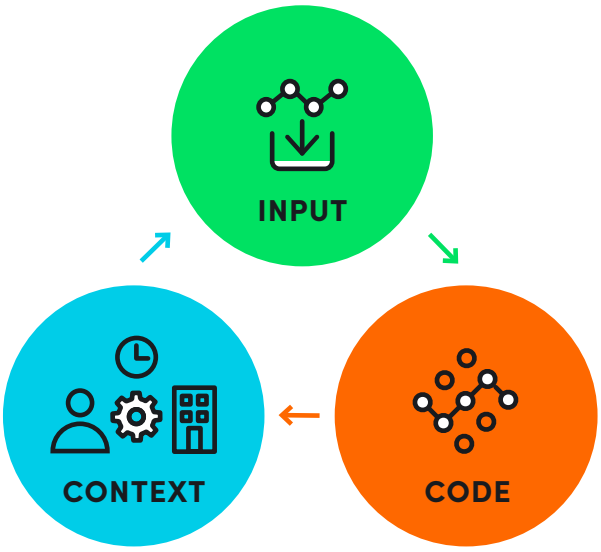
Evidence has demonstrated how, despite good intentions, bias can seep into algorithms from a variety of entry points. Most foundationally, data leveraged for a predictive algorithm can unintentionally reflect existing societal biases and historical discrimination. A country's legacy of inequality, such as mandatory migration, entrenched gender norms, racial segregation,

[iii] We borrow the input, code, and context Framework from Hunt and McKelvey who study the use of algorithms in media.

or other types of discrimination in education and employment, for example, will inevitably reflect itself in the data trails crunched by algorithms. In the healthcare example cited above, the algorithm relied on past healthcare expenditures to predict what care a patient would require going forward. But Black Americans have had to deal with decades of institutional and cultural barriers in healthcare access, resulting in lower past expenditures. The story the algorithm was telling, then, was not the patients' actual medical need but rather the history of disparate access to healthcare between white and Black America.[15] Beyond challenges of representativeness, data inputs face issues in stability, quality, and control, which is particularly relevant in a fast-moving world of digital finance where small tweaks in mobile money platforms or apps lead to big changes in consumer behavior and the stability of data trails.

Even if developers take pains to avoid using data on protected categories, particular variables could easily proxy for such sensitive data in the code—for instance, using geolocation in a country that has clear geographic divisions by race or religion, or the educational level of the applicant in a country that has traditionally limited access to education for certain groups, or mobility data as a sign of stability in a country where internal migration is common. Additionally, the opacity of many models can make it even harder to detect, with machine learning techniques undecipherable sometimes even for the developers themselves, creating challenges to auditing.

FRAMEWORK TO UNDERSTAND ALGORITHMIC BIAS



Organizational diversity and grounding in local context are important dynamics that, when absent, can lead to oversights, incorrect assumptions, and exclusion. Additionally, increasing reliance on automated algorithms to make decisions, such as credit approval, may distance organizational leaders from decisions that could harm consumers. Numerous financial service providers interviewed for this paper report that data science solutions are created by short-term consultants, purchased through off-the-shelf packages, or developed by teams that are relatively siloed off from senior management. In one case in East Asia, an investor seconded an entire data science team to a financial service provider, but the team had little interaction with the rest of the organization and did not know the context or client base well. Senior management had only a superficial idea how the data science solutions were being designed or deployed, which is problematic both for monitoring for harms and for accountability, should things go amiss.

While the framework of inputs, code, and context help explain algorithm development and facilitate the categorization of risks and tools, in practice they overlap and addressing one area without the others is limiting. Long-term solutions for organizations should aim to be holistic and address all three areas through an iterative process. For instance, context will determine what kind of data is available and the methods necessary to evaluate your model. Data science skills will come into play, but fear of the "black box" should not stop sector and country specialists from getting involved, as they have critical knowledge that will help guide choices about algorithm development and deployment.

Taken to scale in emerging markets, bias could run counter to the goals for inclusive financial services and result in the denial of economic opportunities to consumers at the data margins.

RESPONSES

A multitude of approaches across the finance, technology, engineering, and medical sectors, largely in developed markets, have worked towards "fairness-aware" algorithm development and testing. These approaches are often part of bigger discussions around building responsible technology and equitable data economies given historic marginalization as well as the power imbalances between big tech and consumers.

There has been a focus on building technical tools, such as experiments to quantify disparate impact,[16] black box testing methods, and code reviews.[17,18] Other approaches have endeavored to make algorithms more transparent, through "white box" testing or logging processes and disclosure of source code and data sources.[19] Initiatives have sprung up to build awareness and tools, whether from the data science community itself like the Fairness, Accountability and Transparency in Machine Learning (FATML), or multilaterals like the OECD's Principles on Artificial Intelligence.

Governments are just beginning to create regulatory strategies to address, let alone enforce, algorithmic accountability.[20] The World Bank tallied in 2017 that only 44 percent of low-income markets had laws prohibiting discrimination in financial services, though the purview of these was often for regulated institutions, leaving out large swaths of the market.[21] Beyond what already exists in the financial sector, the newest contributions have come from the slew of recently passed omnibus data protection laws, the gold standard being the General Data Protection Regulation (GDPR). Much like policymakers and regulators, consumers are in a constant state of catch-up as to what data is collected about them, who collects it, and how it is processed and even monetized.

WHY IT MATTERS FOR INCLUSIVE FINANCE

While responsible algorithms and ethical AI debates have received attention in sectors such as criminal justice, access to healthcare, and mainstream finance, there has been little exploration in the inclusive finance space, particularly around bias, discrimination, and exclusion.

In credit scoring, the application that this paper focuses on, inaccurate and incomplete data presents risks of incorrectly categorizing individuals' creditworthiness. This risk is heightened for vulnerable groups since the data trails of vulnerable individuals can encode realities of their environment and the types of experimental or predatory products they've been exposed to, making their individual profile appear riskier due to the conditions under which they are accessing credit.[22] This has been documented in traditional credit scoring mechanisms in the U.S., where communities of color are exposed to more payday and "fringe" lenders, a parallel of which in the inclusive finance space has existed in Kenya, where a digital lending laboratory exposed low-income consumers to credit bureau blacklisting which may have barred them from loans or negatively marked their digital footprints.[23]

Taken to scale in emerging markets, this could run counter to the goals for inclusive financial services and result in the denial of economic opportunities to consumers at the data margins. Recent research conducted by MSC shows that digital credit customers tend to be younger, male, and living in urban areas, generally fitting into categories of those who tend to be more financially included and digitally savvy.[24] A 2018 study of digital credit transaction data in Tanzania also revealed striking gender and rural/urban gaps in digital credit users.[25] This challenges the story that alternative, mobile phone data will inevitably solve the thin file problem of many rural or female consumers.

What We Want to Know

For CFI, the proliferation of these tools raises a host of fundamental questions that deserve further inquiry. Our questions range from the empirical (e.g., What are providers and other stakeholders doing today to identify and mitigate against bias?) to the ethical (e.g., How to define fairness in inclusive finance?). Few of these can be definitively answered, but the sectoral conversations around them must start today:

- Are algorithm-driven tools helping providers and markets achieve inclusive finance goals or further cementing the digital divide? How can inclusive finance algorithms become biased and exclusionary?
- What are providers and other stakeholders doing today to identify and mitigate against bias? What are the incentives and challenges for providers to do anything about it? Can advances in other fields be applied in inclusive financial services?
- How can marketplaces be effectively supervised as these complex tools are being deployed? Does increased use of algorithms change market competition or influence competitive dynamics?
- How do the new universal approaches to data protection intersect with algorithms, bias, and inclusive financial services?
- How do consumers think about the decisions made about them using algorithms, the data they share, and their nascent data rights?

APPROACH & LIMITATIONS

This paper represents the results of a multi-pronged exploratory effort. The CFI team conducted key informant interviews with more than 30 stakeholders across 12 countries. Among them, the team spoke with financial service providers, largely fintech companies, as well as several third-party companies that conduct analytics and partner with lenders. These discussions centered on levels of awareness and concern over the issues and what tools for accountability currently exist. We also interviewed market actors including a mix of regulators and consumer organizations in Uganda, Rwanda, Brazil, the Philippines, and India. Given the sensitive nature of the discussions, we have kept the names of the interviewees confidential and will only be referring to them by their country or region, and for fintechs, their business model. Finally, we identified and spoke with a handful of data protection scholars, all based in the United States or Europe, with expertise in emerging data protection frameworks, as well as several cutting edge data rights and ethics organizations, like the Ada Lovelace Institute.

We hover around the use of algorithms for underwriting, admittedly a tiny slice of the of use cases, for several reasons. When it comes to questions of how fintechs are advancing inclusive financial services, credit decisions are often made by the automated system that determines who becomes a customer and begins to build a credit history, and who is denied access and continues to be excluded from credit and other follow-on financial products.[iv] Despite the focus on underwriting, our observations have implications for other use cases of algorithms in inclusive finance, and more broadly, in development interventions as well.

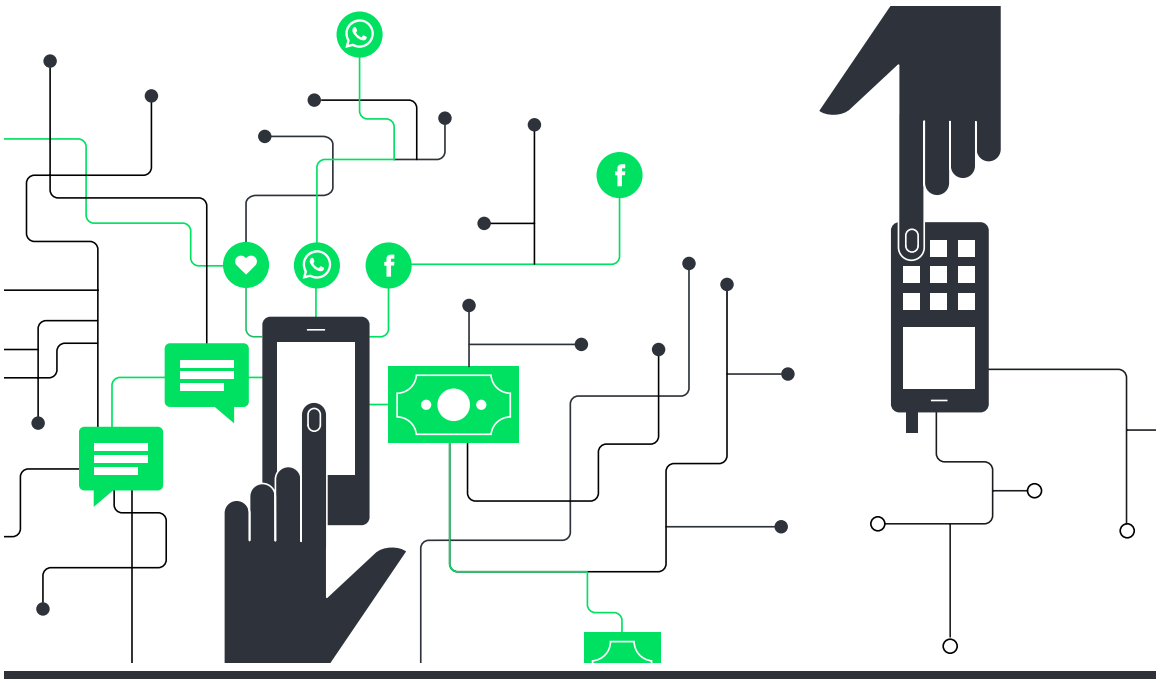
The rest of the paper is organized as follows: a) Section 1 touches on data trails in the digital economy; b) Section 2 digs into the risks of bias and emerging tools through the three aforementioned categories of Inputs, Code, and Context; c) Section 3 lays out suggestions for various inclusive finance stakeholders to advance evidence, solutions, and incentives for responsible algorithms.

This is not meant as a definitive treatise on the topic, but a first step in a wider portfolio of research. We are limited in the sample of providers, business models, and their varied adoption of algorithmic systems and machine learning techniques. Additionally, our interviews with providers did not include a review of their proprietary algorithms, codes, or data sources. Much more work will be needed in this area, which will be addressed in Section 3.

iv N.B. While not the focus of this paper, we also believe that the inclusive fintech sector should focus on opening other pathways beyond credit, especially given building evidence of debt stress in countries with advanced digital lending markets.

1 Section 1: Data Trails in the Digital Economy

DATA TRAILS IN A DIGITAL ECONOMY



Old to New Data Environments for Underwriting

The application of algorithms and alternative data to credit scoring is meant to solve for the limited availability of traditional financial metrics, especially for unbanked customers, as well as introduce efficiencies in operations like customer acquisition and decision-making. When forecasting creditworthiness, the ideal data has always been the past credit history of individuals coupled with a cash-flow analysis. This approach weights their credit exposure, credit line usage, and repayment behaviors. In more developed markets, credit bureaus and/or credit registries collect information (both positive and negative records) from across the market and develop generic and ad hoc credit scoring models that are widely used for underwriting. Most credit agencies develop scorecards for thin file customers too, although the algorithm for thin file scorecards traditionally has been less predictive than for those with more credit history.

Credit agencies and bureaus in many markets function like a club; financial institutions share their data (and in many markets they must do so based on regulatory requirements) to access data-driven products and services. After decades of work, the subjectivity of manual underwriting is long gone, the scoring models are monitored, and on balance, the market is better understood because there is more transparency about debt levels, non-performing loans, and repayment behaviors. Of course this "club" is not the same everywhere—there

are stark differences in coverage (percent of institutions participating), type of data collected, technology to maintain and process data, ability to develop high-quality reports, scoring models, and quality of data-driven tools. While the top 20 markets globally cover an average of 83 percent of the adult population through either a credit bureau or registry, the bottom 50 markets cover on average only 10 percent of adults.[26] Although there have been recent gains, millions of individuals remain "invisible" and have a limited financial footprint to leverage for greater access to financial services.

In emerging economies, mobile phones have contributed to an explosion of additional data, often labeled "alternative," and raised the visibility of millions of consumers. Researchers have tested mobile phone metadata to derive behavioral indicators that can be used to accurately predict repayment. These approaches have looked at phone usage measures such as transactions (derived from SMS, calls, or data usage), location data, payments activity across one's social network, and phone characteristics (model). Björkegren and Grissen posit that this subset of phone data can be linked to repayment capacity: "Phone usage captures many behaviors that have some intuitive link to repayment. A phone account is a financial account and captures a slice of a person's expenditure. Most of our indicators measure patterns in how expenses are managed, such as variation (is usage erratic?), slope (is usage growing or shrinking over time?), and periodicity (what are the temporal patterns of usage?)."[27]

Fainter Digital Footprints

While this has leapfrogged certain barriers, phone access and usage are driven by and tangled up in digital capability, social norms, and other powerful forces. And since digital credit models leverage phone-specific variables, the interaction with these forces is a critical but little-understood component of determining how inclusive these new tools are.

This can be most easily demonstrated through a gender lens, due to available research, but it *must* be explored for other marginalized groups. For example, 44 percent of women in low-income countries lack access to ID, an integral part of financial onboarding, compared to 28 percent of men.[28] Women in low-and middle-income countries are 8 percent less likely than men to own a phone and 20 percent less likely to use mobile internet, which means that in low- and middle-income countries, there are 300 million fewer women accessing mobile internet than men.

The primary barrier to mobile ownership and internet access in Africa and Latin America is affordability, and prices are still relatively high, with a 500MB data plan in sub-Saharan Africa costing around 15 percent gross national income per capita, as compared to the global average of 10 percent GNI per capita.[29] In Asia, literacy and skills are the main barrier, followed by affordability.[30] Even when women own phones, their access is strongly correlated with income and education. They also face challenges in using different add-on services and features and are less inclined to use their phone to promote their business or gather market intelligence.[31,32]

Qualitative research in Ghana and India found that women's digital access was often moderated or monitored by their social network. In Ghana, all male study participants had created their own Facebook accounts, while female participants' accounts were created and regularly monitored by a male family member or friend.[33] In India, women's access to phones was mediated by a male relative who often owned the phone.[34] Women self-censor their presence online due to fear of harassment or damage to their reputation. They use apps that are "closed-circuit" (e.g., WhatsApp) as opposed to "open" (like Facebook).[35] They might create several distinct online identities—one for their close friends, one for their family, etc.

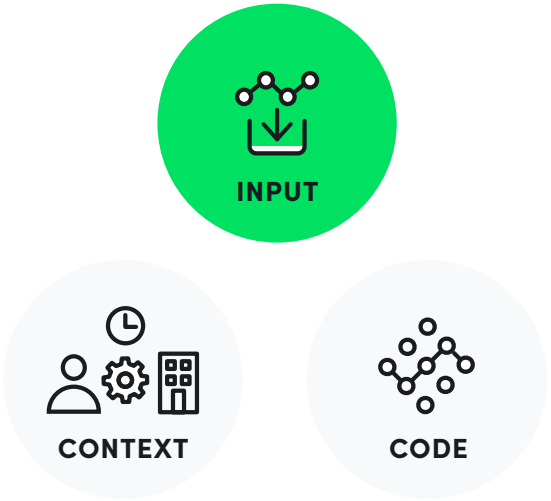
Data availability has increased, and while this is bringing many new customers into the fold, there is the potential that it increases exclusion for those with limited data trails. And as data trails are gobbled up by an algorithm, they likely tell an incomplete story about the consumer, including their creditworthiness. It's understandable that providers must make choices about who is likely to repay a loan based on available data, but certain assumptions and structural data limitations could mask these more complicated stories about creditworthiness.

DIGITAL DATA TRAILS TELL A STORY, BUT IS IT THE RIGHT STORY?

YOUR DATA TRAIL SAYS...	BUT THE REAL STORY IS...
↗ Your relatively small number of monthly mobile money transactions show a low level of business activity and you are a poor risk for an MSME loan.	↗ The majority of your transactions are conducted in cash, offline and, despite COVID, your business is thriving.
↗ The small contact list on your phone and/or your small number of social media connections shows you don't have a robust social network that could support you if you are behind on your payments.	↗ Women like you in your peri-urban town are strongly discouraged from adding contacts to their phone for fear of harassment and/or reputational damage.
↗ You have infrequent mobile money receipts on your SMS log on your phone, and therefore insufficient evidence of cash flows to qualify for a loan.	↗ You share your phone with your multigenerational household, and you or another family member frequently needs to delete old SMS messages because your phone storage is limited.
↗ Your identity is associated with too many SIM cards, so the telecom operator has flagged your account for fraud.	↗ You are a refugee with no national ID card, which is required to register a SIM in the country where you live. As a workaround, you've paid a local to register a SIM for you using their ID. That person has been flagged for registering too many SIMs and all accounts associated with that name have been frozen, barring you from using your mobile money account or building a transaction history.

2 Section 2: Understanding Bias and Potential Solutions

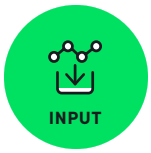
INPUT



The first, and arguably most important, stage of algorithm development for applications like underwriting is identifying and curating data inputs. Data quality refers to the utility of a given dataset for easy processing and analysis. For instance, the utility of a dataset increases when developers are confident the data contains no errors, which is difficult to verify for many datasets used in the development context. Fintechs strongly prefer data inputs that they control rather than source externally or gain from a third party; data they collect directly from or about customers enables them to know why certain data points are included or excluded, how up-to-date the information is, and any other limitations of that dataset.

When fintechs regularly source data from external entities like credit bureaus, telecom companies, or data brokers, it introduces unknowns. There is limited understanding of how the third-party data was collected, why certain decisions were made about what to collect, and how up-to-date it is. Not knowing why and how a third-party dataset has been collected, regardless of source, hampers the ability of companies to understand what might be missing. Without this information, the dataset may be suitable for one use but not another, especially if the stakes of the decisions or model outcome are very different. For example, if a data collection process prioritized number of responses over accuracy, this dataset might be useful to understand broad trends but not predictions about individuals. It is also more difficult to judge the accuracy of third-party datasets; analysis of U.S. data brokers, for instance, showed reports that were "riddled with inaccuracies."[36] While standardization of credit bureau data has advanced, a number of fintechs expressed frustration with credit bureaus and registries, noting that the lack of information about the data and the inability to access it in real time can lower its value in their credit models.

The opacity of external data also makes it more difficult to interrogate trends and can introduce unknowns into credit models. For instance, a fintech accessing credit bureau data may observe an increase in default rates, but won't know if that increase is a result of true numbers of defaults or increased reporting. In many markets, fintechs sit in a regulatory gray area and are not required by law to report defaults to credit bureaus, or there is a reluctance on the part of regulators to impose penalties on lenders that are failing to report data or are providing error-riddled data. While some fintechs report data voluntarily, they may not do so consistently, and users of credit bureau data would not know when other companies started or stopped reporting, or what mistakes are present in that data. Using this data to inform changes to credit models might base future decisions on trends that reflect company behavior rather than consumer behavior.



Another issue relates to the underlying stability of data used as inputs for credit models; stability refers to the consistency of inputs and conditions over time. Applied in our context, this focuses attention on the stability of the types of mobile phone data used in credit models, such as device information, storage, and app usage. The introduction of a new app or service could cause people to behave differently on their phones and drastically change battery usage, storage metrics, or other features of a credit model. Similarly, the cost of devices or airtime could change as more powerful phones become cheaper over time, destabilizing the underlying logic of the credit model. According to one data scientist, this instability can "be heavily correlated with things that can lead to discriminatory decisions." As models depend on past performance to predict future default rates, external shocks or changes to the environment that shift behaviors can destabilize the model over time.[v]

REPRESENTATIVENESS AND MISSING LABELS

Representativeness refers to whether the data sufficiently reflects the context for that model's deployment. For instance, does the training data, which is the source material used to train the scoring algorithm, reflect demographic, socioeconomic, and/or behavioral characteristics of borrowers in the market where the credit product will be made available?

Bias can arise when complete data is available for certain segments of borrowers in a market but is incomplete for other segments. Training data that is not representative can result in misclassifications of those customers with limited digital footprints, which may be excluded from data systems due to legacies of discrimination, as described earlier.[37] Depending on the options available and the level of data invisibility, it may be inappropriate or unethical to apply algorithmic decisions to certain situations.[38] For instance, developing and training a model in one market does not mean that model should be used in another, even if there are similar demographics or characteristics between the countries. Representative data must reflect the context where the model will be used, and fintechs that redeploy lending algorithms with minor tweaks in new markets are making assumptions that may not hold and could exclude consumers.

Representativeness of data takes on another dynamic after a model is deployed and learns from the market. Ongoing model training depends on who has been onboarded onto the fintech platform as a customer. Fintechs highlight that their data is limited to people who apply for loans, and there is no way of understanding the behavior of people who were rejected, whether by a loan officer or an algorithm. This creates a "missing labels" problem in training data, where data for people who never received credit is unavailable as a counterfactual.[39] Without a counterfactual for people denied credit by the scoring algorithm, the model could be perpetuating exclusion of some market segments. "Any kind of application-style data processing or machine learning is particularly challenging because you don't get the results of the people you say no to," one interviewee explained.

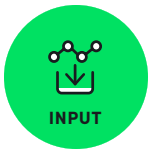
EMERGING TOOLS

Audits and Data Documentation Systems

While much attention has been dedicated to technological challenges of testing or explaining an algorithm's logic, AI accountability researchers have identified more basic gaps of documentation and reporting around "a system's purpose, policies, inputs, and outputs," which does not necessarily require technical methods or breaking open the black box.[40] Machine learning practitioners, both from our interviews and other domains, are aligned that anti-bias measures should first focus on the data available for training models, prior to examining the models themselves. Most fintechs interviewed described ad hoc discussions around bias, often at an inflection point of integrating a new data source, but had few internal processes or documentation approaches.

Audits, conducted internally or by an independent party, as well as documentation tools from other industries managing high stakes outcomes and risks, such as aerospace, medical devices, and traditional financial services, might be useful. One example of process documentation comes from the aerospace

v Destabilization of the Google Flu Trends model provides an example of how media coverage of a bad flu season and the H1N1 pandemic changed search behavior and predictive accuracy of Google's flu-tracking methodology: https://www.nature.com/news/polopoly_fs/1.12413!/menu/main/topColumns/topLeftColumn/pdf/494155a.pdf



industry, which has long used design checklists, simple tools that assist designers in "having a more informed view of important questions, edge cases, and failures."[41] The checklists are not structured to be yes/no box-checking, but instead require designers or engineers to describe a process they undertook. This could be adapted to the machine learning product development cycle to help inform others of decisions made or processes used at different inflection points in the pipeline. Another standard engineering procedure called Failure Modes and Effects Analysis (FMEA) could be applicable. FMEA is a methodology that ex ante examines a proposed technology or design for potential failures, by conducting research and literature reviews on similar technology deployments and known risks associated with the use of that technology.[42]

Tools such as the Datasheets for Datasets, developed by practitioners at Microsoft Research, have also emerged to fill this need, recognizing that the selection of data is the fundamental determinant of a model's behavior.[43] Datasheets for Datasets are modeled off best practices in the electronics industry, which require that components of electronics products each be documented with a datasheet that describes requirements, recommended uses, and other information and characteristics of the component in order to prevent inappropriate applications. These researchers have suggested that all datasets should be accompanied by similar documentation about the sources, recommended uses, known characteristics, and known concerns of a particular dataset. While not a technical tool, the datasheet would facilitate information sharing about the use of data for certain purposes, an especially important step considering fintechs' use of third-party datasets—especially from credit bureaus—and their limitations.[44]

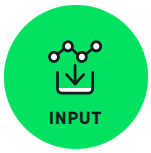
The Datasheets for Datasets tool contains sets of questions about dataset creation, composition, and other characteristics. Tailoring these tools to the inclusive finance sector would require a set of questions probing whether the dataset is sufficiently representative of low-income consumers to achieve inclusion goals, as well as the ethical considerations that are unique to these model deployment contexts.[45] As most documentation efforts are currently ad hoc, standardization would support continuity for the provider as well as facilitate communication about inclusive finance goals and model development decisions with management, investors, third parties, and between technical and non-technical stakeholders.

There is a cost involved with making input data more representative, and fintechs are not aligned on whether they should bear the burden to ensure that the inputs they use account for the broader population of excluded customers. Efforts to access more representative data through market research or other investments are expensive. As one fintech founder articulated, "At the end of the day, whoever you're marketing to is what your algorithm is being exposed to and there's a constant battle between cost of acquisition and the bias in the algorithm." A cofounder of a different Asia-based fintech questioned whether it was their role to make decisions based on data that isn't generated by the market, noting it's "too hard to make decisions based on data on who is not applying for loans."

Regulatory Levers on Data Inputs

Rights to Data Access and Rectification

Article 16 of GDPR and similar provisions elsewhere give individuals the right to have errors and inaccurate or incomplete personal data corrected or rectified by a data processor. The 2019 Kenyan Data Protection Act, for instance, states that individuals have the right to request the rectification of personal data that is "inaccurate, out-of-date, incomplete or misleading."[46] In a digital credit scenario, this would give potential borrowers the right to (theoretically) request that a lender rectifies data about them that feeds into their risk profile or creditworthiness. For instance, perhaps credit bureau records were inaccurate, or the lender has analyzed the social media activity of the wrong individual, or income levels were incorrectly recorded.



Most frameworks apply this right to the rectification of data inputs or observed data, but not to inferences or predictions made by the provider. Inferences in a digital lending algorithm might include a prediction of an individual's cash flow over time and/or income volatility, and strength of their social network—key in determining whether they should receive a loan, and not impervious to bias.[47] Applying data rectification at the inference and prediction stage veers into murky terrain as it involves proprietary code or businesses' intellectual property. Thus far, the California Consumer Privacy Act (CCPA) is the only well-known framework that gives protections for data inferences and predictions.[48]

Regardless of whether rectification is applied to input data or data inferences, in order for such a right to be exercised, it assumes that individuals are aware that digital lenders, for instance, have their personal data. Additionally, it implies that data processors like digital lenders have control over the data they are collecting and using, which, in the era of data brokers and alternative data, could be a daunting and near-impossible task.

Fair Treatment of Sensitive and Protected Data

Emerging data protection frameworks require that data processors, such as digital lenders, treat data "fairly" or "lawfully" to avoid mistreatment or harm. The draft Indian Personal Data Protection Bill (PDPB), for instance, specifically mentions "discriminatory treatment" as a harm resulting from improper handling of personal data.

Although there are differences across jurisdictions, in most cases, new data privacy bills call out certain types of personal data for special provisions and treatment. Medical and criminal records, political affiliation, religion, race, sexual preference, and in some cases financial records are all part of what certain jurisdictions call sensitive or special data. Automated decision-making, such as loan approval, using such sensitive or special data is not allowed without explicit consent of the consumer. An American law professor called this a prophylactic or "sledgehammer" approach, as it aims to exert control at the data collection phase. While blunt, it may be necessary as once data is collected, trying to control the methodology or testing end results will leave you "lost in most scenarios."[49]

Since GDPR-style regulation is universal and not sector-specific, it does not include provisions on what type of data could be used (or not used) in digital lending algorithms, for instance. Rather, it is meant to piggyback off of and reinforce other existing laws, like financial sector regulation, that prevent discrimination.[50] Without taking into account the specificities of the financial sector, the context where they operate, and the data that could be used to score individuals, the risk of bias and exclusion increases.

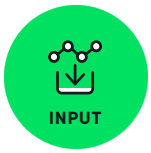
GENERAL DATA PROTECTION REGULATION (GDPR)

In 2016 the European Union (EU) announced the General Data Protection Regulation (GDPR), replacing previous directives to reflect the ways the world had changed and to bring data protection regulation into the 21st century.* GDPR governs how companies can collect, process, handle, store, and use personal data, and also enumerates novel data rights for individuals, some of which this section will explain. The influence and impact of GDPR is difficult to overstate, as it applies not only within the 28 member states but also to data that is exported out of the EU to anywhere else, as well as any individuals "in the Union," even if they are not physically located in the EU. The focus of GDPR has been on individual rights and protections, and the implications on businesses like MSMEs or group/societal harms, is less clear.

Quickly becoming the global standard, GDPR has inspired other countries—and U.S. states, in the case of the California Consumer Privacy Act (CCPA)—to develop (or update) their own data protection frameworks. As of writing, 128 out of 194 countries surveyed had put in place data protection legislation, many inspired by GDPR, with many more in draft form.†

* https://www.financialdirector.co.uk/2018/06/21/gdpr-how-is-it-affecting-banks/

† https://unctad.org/page/data-protection-and-privacy-legislation-worldwide Accessed January 15th 2021



Consent for Data Processing

The majority of regulatory frameworks justify data sharing, usage, and processing under the notion of informed consent, whereby a customer agrees to the privacy notice of an app or provider. That is to say if a consumer reads (or more likely scrolls on their phone) through a privacy notice and assents, with a click, they have agreed to whatever a provider will do with their data. In GDPR, consent is one of the six legal justifications for data processing outlined in Article 6, and compared to the other justifications appears to be the easiest way for businesses to avoid fines down the road.

Contrary to popular belief, asking for consent is not always mandatory for data processors like digital lenders. GDPR and similar data privacy frameworks allow the processing of personal data if it is necessary for the performance of a contract. In the case of the draft PDPB in India, it explicitly mentions credit scoring and recovery of debt as two of the seven exceptions for processing personal data without explicit consent from individuals. Article 7 of Brazil's General Data Protection Law (Lei Geral de Protecao de Dados or LDPG), mentions a similar provision in which data can be processed legally and without consent for the protection of credit markets.

The explicit inclusion of credit scoring and collection as exceptions to individuals' consent is a nod to policy goals of fostering inclusive finance. These provisions allow digital lenders not only to run algorithms to rescore individuals without consent but also for collection purposes. It is common practice for financial institutions to re-score individuals to evaluate and price portfolios that are at different stages of debt collection. This helps financial organizations decide when they might sell that debt to collection agencies. If that's the case, then individuals' consent is no longer attached to the original financial provider but to a different entity that now owns that debt, and at least some of their personal data.

From the more advanced to emerging markets, a fundamental question remains as to whether individuals are aware enough or have the autonomy to give their free, specific, informed, and unambiguous consent to data processing.

SENSITIVE DATA: FAIRNESS THROUGH UNAWARENESS

A counterintuitive trade-off concerning protected attribute data is that the (often legal) prohibition against collecting such data can undercut efforts to evaluate algorithms for fairness. For instance, without collecting data on race, it is much more difficult to verify if an algorithm is delivering biased decisions based on race. While the limits on collecting protected data are understandable, research across sectors and in the fintech industry show that eliminating sensitive attribute data from machine learning processes does not eliminate decisions based on proxies for these attributes and can actually make discrimination harder to detect, known as "fairness through unawareness." More stakeholders are coming to support the idea of collecting sensitive data to help identify and prevent unfair outcomes.

Privacy scholars have stated that consent is linked to autonomy, and can be easily undermined by issues stemming from poverty, such as low financial capability.[51] At the same time, the narrative that all consumers will happily consent to trade away their data in exchange for a service or product has gained traction. The Ipsos Global Trends Survey found that in the last few years there has been a 7 percent rise, to nearly half, in respondents who would trade away their data for personalized services or products; this rises to about two-thirds in China and India.[52]

But is this the full story? Researchers at CGAP have skewered the utility of informed consent along several dimensions:

- Choice: Due to its binary nature, consumers are given the choice between consent and getting the product or not consenting and losing access.
- Comprehension: People do not take the time to fully read or understand what they are consenting to, especially third-party data sharing.
- Complexity/bias: Consent is not valid when consumers do not understand what is done or could be done with their data—including the possibility of excluding them.[53]

CGAP has also conducted small-scale research in Kenya and India showing that low-income consumers value privacy enough to pay extra for financial products with extra protection—though this kind of comparison shopping is not available on the market.[54] Therefore, we must not assume that because low-income consumers currently consent, that it is their preference, nor that they wouldn't prefer a more privacy-intensive financial product.

Some advocates have pushed for regulatory approaches that go beyond consent and move the burden from the consumer to the provider. The PDPB in India contains some innovative approaches, including "consent managers," "data trust scores," and "legitimate purposes test."[55] A legitimate purpose test is not rooted in consent but requires data processors like lenders to use personal information in a way that aligns with the original intent behind collection and is beneficial to consumers. It is a provision that could help avoid data collection beyond what is necessary, but in a world where "all data is credit data" there are open questions on how to differentiate which data points are relevant and which ones are not. While the Indian bill is not yet law, the privacy community eagerly awaits how these beyond-consent approaches will be implemented.

INPUT SUMMARY

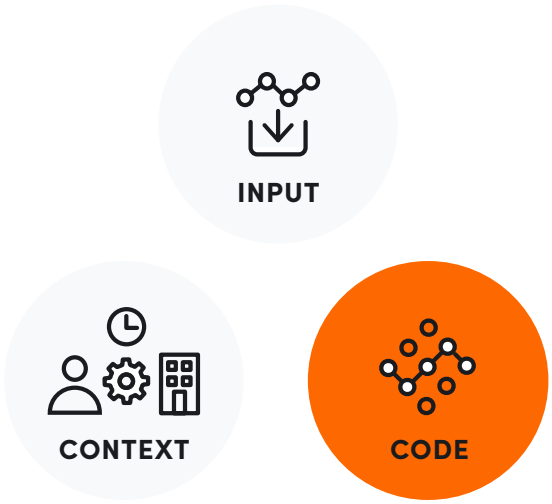
Biases and Limitations

The data inputs to an algorithm are the first opening to introduce unintended biases. Whether using their own data or third-party data, providers must understand the quality, representativeness, and limitations of data used for decision-making.

Potential Solutions

While regulations are emerging in some markets to improve protections for sensitive data and introduce rules governing use of personal data, providers can also take steps to improve the quality of their data. Leveraging lessons from other industries, **providers should consider introducing tools and processing to better document their use of data** to diagnose the potential for bias.

CODE



While predictive analytics and algorithms are not new in finance, the introduction of more sophisticated machine learning methods has given rise to opaque or "black box" models. Such models go far beyond linear regressions and statistical models in which a person can identify features of the model that drive certain outcomes or classifications. Models that rely on complex processing and transformation of thousands of data points into a single decision—yes or no to a credit application—can result in a situation where humans are not capable of interpreting which specific behaviors or data points drove the ultimate credit decision.[56] Whether using supervised or unsupervised machine learning techniques, black box models obscure decision-making because the algorithm has developed complex, non-linear rules and processes that, even if analyzed by humans, would be difficult to understand and explain.[57]

One CEO of a third-party technology provider that has partnered on model development observed that incumbent MFIs or banks are interested in machine learning models but frustrated by the lack of clarity in the decision-making process. She described "an algorithm that doesn't even use variables and was designed through machine learning and has 200,000 different combinations of characteristics. It is more accurate, but it's not explainable, so you can use it, but you can never

ask us why a decision was made." In this situation, bank leadership is unable to understand a specific loan decision or interrogate a pattern of bias in loan decisions without a team of data scientists on staff with expendable time. If the company is relying entirely on a third-party provider for data science expertise, a proprietary model could end up a black box even for the company using it in their credit decisions. With additional data also comes the possibility of introducing spurious or unintended correlations with sensitive or protected attributes into decision-making.[58] For example, studies from developed markets have documented how mobile phone data such as device type and installed applications can be predictive of, or proxies for, gender or age.[59] Machine learning fairness researchers have pointed to the risk of encoding such correlations into decision engines and unintentionally making decisions that are actually discriminatory.[60]

Mobile phone and geolocation data are two types of data that are already commonly associated with proxy risk. Geographic data has long been understood as a proxy for race in many markets, including the U.S., where redlining has been well documented.[61] Some digital lenders create default rate heat maps to avoid certain geographic areas that show higher than average default rates. But the result is that any individual that lives in that area will be negatively impacted by the algorithm. Location data is directly collected by network operators so having aggregated financial geodata is not only a possibility but a must in order to manage their portfolio. The risks of "networked privacy" harms, where individuals are assessed not by their own behaviors but by those in their networks, comes squarely into play with proxy data.[62]

POTENTIAL SOLUTIONS

Technical Tools to Detect Bias

AI fairness researchers have designed open source tools to detect bias in models; interactive tools from IBM, Google, and Pymetrics, among others, provide algorithmic methods to audit models for statistically significant differences in treatment of different demographic groups, or probe different model results by tweaking features or inputs.[63] A complement to the "Datasheets for Datasets" approach in the previous section, "Model Cards for Model Reporting" is a proposed technique that focuses on the performance characteristics and intended context for the model's application. Model cards take the next step in logging the performance benchmarks of a trained machine learning model across demographic and other groups, as well as the methods used to evaluate performance, intended uses and contexts for deployment, and any other ethical considerations.[64] While these tools are publicly available and tailored to meet a range of technical skill sets, they do require some level of data expertise to operate and have not been widely validated, let alone in the inclusive finance sector.

Model Reviews at Fintechs

Fintechs interviewed acknowledged the importance of anti-bias testing for models, both prior to and after deployment in the field, but also acknowledged that technical approaches to detecting discrimination are limited, as is the capacity of many smaller fintechs. Model reviews mostly looked at performance and accuracy, though not often focused on exclusion or bias.

Reviews of models prior to deployment take the form of "monitoring at the model training level, what are the driving features we see of predictive power, and what's the subsegment performance of the models, and do those raise any red flags."[65] The management of these reviews can be led by credit teams, product teams, or data science teams, but often involve reading out findings to leadership from other parts of the company that need to have a working, if not technical, understanding of the models and their drivers. A few organizations keep a running record of all decisions associated with how credit decisions are made and model changes, but there is no standard for this type of documentation, nor has the practice been widely adopted.

Without any regular monitoring process in place, ad hoc discussions about bias are sometimes triggered by observed performance irregularities against established business rules: different approval thresholds for different segments, levels of variance that trigger a review, or some basic requirements around data sufficiency. In Asia, one fintech noticed a gendered pattern in loan decisions where women clients were being rejected at a significantly higher rate compared to male clients. The company recognized that the gender imbalance of phone ownership was affecting their ability to onboard women and for their algorithm to understand female borrowers.

As a result, they instituted a lower threshold to approve women for credit. They have observed similar issues around age and hold a monthly review of these rules to decide whether and how to tweak thresholds for different segments. Of course, the decision to lower thresholds for certain segments increases risk as it could allow for more customers to default. In this case, the company decided to accept that risk in favor of a more inclusive outcome, knowing that their predictions will be less accurate until they have sufficient data from onboarding more female clients.[66]

Internal limitations and a lack of incentive structure hamper efforts by fintechs to monitor for bias or disparate impact. Detecting and understanding proxies is one aspect of preventing discriminatory decisions, but it can be prohibitively difficult to reveal these relationships in complex models, especially when it is not well understood how data points such as phone behavior, social media interactions, or online browsing connect to sensitive attributes. Among fintech lenders that use geolocation data, developers were aware that geography can be highly correlated with demographic data, such as tribe or ethnicity, upon which they actively want to avoid basing credit decisions. Fintechs using geographic data need to understand the local cultural context and power relationships, such as why different tribes or ethnic groups live in certain areas or other spatially correlated variables. While some of these variables can be identified, correlations with other types of data are not as well understood and can introduce bias.

Regulatory Levers

One approach taken in the United States is testing models' disparate impacts. This outcomes-based tool, wielded under the Equal Credit Opportunity Act (ECOA), predates the use of alternative data and fintech and requires that lenders change their model if they notice that it leads to disparate outputs based on race, religion, sex, or other protected attributes.

Even in developed markets with a history of regulation and supervision of credit markets, the capacity to deploy actual monitoring practices and algorithm testing is limited. In the EU, GDPR created the Data Protection Impact Assessment (DPIA) tool for data processors to ex ante assess their practices if they are carrying out high risk data processing. While a DPIA might touch on issues of algorithmic accountability, it does not amount to a government model audit.

A government-led audit of algorithms would require a high level of technical expertise to allow the effective inspection of the internal designs of automated decision-making tools, and full disclosure of the source code and data. Many legal experts believe it is essential to have regulators with the technical expertise to understand the complexities of these technologies so they do not feel intimidated by the task, in addition to a strong commitment to consumer protection.[67] These criteria are likely difficult to meet due to lack of government resources, especially in emerging markets, and a reluctance by companies to reveal trade secrets.

Right to an Explanation

GDPR and GDPR-inspired legislation imparts consumers with the right to an explanation of automated decision-making ; essentially, individuals are empowered to obtain meaningful information about the logic involved in an algorithm's decision-making, as well as the decision's significance and the consequences for the individual. This condition applies only to fully automated decisions. In the context of digital credit, those denied a loan have standing to inquire as to why they were refused. Beyond informing individuals about the nature of such decisions, this provision might help individuals assess their willingness to share their data going forward.

The draft Rwandan Data Privacy framework takes explainability a step further and mandates in article 39 that data controllers should inform individuals about the logic involved in their automated decision at the time of personal data collection. Brazil's Data Protection Act affords the consumer the ability to ask for a review of a decision made with their personal data taken solely through automated processing; included in the review should be the criteria and procedures used for the decision.[68] The 2019 Ugandan Data Protection and Privacy Act gives consumers the ability to ask a data controller that decisions about them are not based solely by automated decision means; in practice, the process is arduous, as the request must come in writing and is likely at odds with the consumer's often immediate need for digital credit.

What explainability means in practice and how it will be enforced is still a very open debate. GDPR and other frameworks do not go beyond a general description, which leaves room for arguing about what level of detail an explanation should have. Would a robust explanation of the rationale behind a decision require developers and data processors to share how the code was developed? Explainability runs contrary to the opaque methodologies used when developing algorithms and their protection by intellectual property laws.

Enforcement of explainability would also require a supervisory body that has the capacity to sift through algorithms and monitor and test the efficacy of explanations used by providers. It is hard to imagine how this will play out in emerging markets given capacity constraints of supervisors and where most of the algorithms and credit scoring decision tools are developed leveraging alternative data and multiple, third-party sources of information.

It is also unclear what constitutes a fully automated decision that would trigger additional protections for consumers. In a digital lending context, if an algorithm automatically collects, collates, analyzes, and recommends who is rejected for a loan but a human, unaware of any reasoning behind the decision, pushes "approve," does this constitute an automated decision for which consumers deserve an explanation?

Dr. Sandra Wachter, research fellow at the Oxford Internet Institute, predicts that GDPR is likely to grant individuals information about the existence of an automated decision algorithm, but no actual explanation about the rationale of the decision.[69] Her work also suggests explanations that use counterfactuals, for instance: "Your loan application stated that your monthly income is ₹4000. If your monthly income was ₹5500 you would have been offered a loan."[70]

There is some concern that should consumers become too proficient in how algorithms make decisions, it creates incentives for them to "reverse engineer" or "game" the system. Twitter users taught Microsoft's AI chatbot, Tay, to produce racist, antisemitic, and misogynistic content in less than a day and hackers can trick self-driving cars into crashing by plastering fake stickers on the road.[71] In the digital lending space, consumers might game an underwriting algorithm by seizing on what behaviors increase their credit limits, such as number of outgoing SMS messages per day, and changing their behavior accordingly, even if it is not aligned with their repayment capacity.

Researchers from the Digital Credit Observatory demonstrated through a lab experiment in Kenya that even for new adopters, smartphone users are savvy enough to change their behavior in response to information about decisions. However, the same research team also devised decision rules for the algorithm that anticipated consumer manipulation, and found the cost of transparency (or loss in predictive performance) to be 8 percent, versus a 23 percent loss in predictive performance when such manipulation is not anticipated in the decision rules.[72] Additionally, it stands to reason that transparency about underwriting algorithms, when decision rules are in part driven by potentially beneficial financial behaviors, like saving on your mobile wallet or not borrowing at 3 a.m., might nudge consumers towards positive decisions. Much more work is needed in this area.

CODE SUMMARY

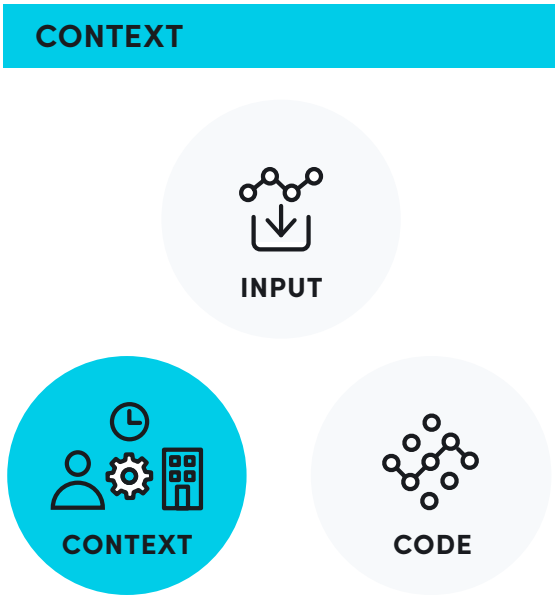
Biases and Limitations

As the data inputs become more diverse and complex, so too do the codes processing data to make decisions. While not currently the predominant models in play, there has been an emergence of opaque models with complex, non-linear rules which make it nearly impossible to understand how they make decisions. And with additional data also comes the possibility that the code builds off of unintended correlations with sensitive or protected attributes.

Potential Solutions

Technical capacity constraints among both providers and governments to understand algorithms and automated decision-making are a challenge but steps can be taken which do not require deep technical knowledge.

Providers can employ tools designed to test for biases and conduct model reviews before and after they are deployed. Policymakers are also exploring rules on explainability which require providers to be able to explain to a customer how a decision was made.



Issues of bias are often only detected after deployment, usually when cases arise in the market with consumers, making bias mitigation efforts generally reactive rather than proactive.[73] Anticipating or addressing bias is often not built into the product development cycle which, depending on the company, can look very different. The process can be further fragmented when providers work with third-party analytics companies to build models. As one CEO described, third-party companies can develop complex machine learning models that are subsequently challenging to integrate into the internal decision-making processes as well as manage over time. In these third-party arrangements, providers often return with questions about outcomes of the machine learning model but, as the same CEO stated, the companies "cannot afford to pay data scientists to go answer that question on a customer by customer basis."

Beyond product development challenges, defining bias and fairness is thorny and context-specific. Evaluating the fairness of an outcome is not solely a technical question, but a social and ethical question. A risk for algorithm-driven systems can link back to insufficient clarity or poorly defined thinking about the purpose and outcomes of the model, and how those link to technical decisions, such as defining target variables.[74] There can be a disconnect between statistical and ethical or societal concepts of fairness.[75]

Ethical and value systems are context-dependent and determinations of fairness also depend on local laws, customs, and concepts of rights and responsibilities.[76] It may be inappropriate to ask a machine learning practitioner to decide questions where a society lacks consensus. This was captured by the sentiment of a data scientist we interviewed: "I feel quite strongly that just because I'm the person implementing these things doesn't mean that I should be the person deciding them."

Cross-cultural dynamics create additional complexity for lenders and machine learning practitioners in the inclusive finance sector. Analysis of the 2020 Inclusive Fintech 50 cohort shows the bulk of investments in early-stage fintechs operating in emerging markets are directed at companies largely led by men and headquartered in developed markets.[77] The significance of these gaps for reducing algorithmic bias lies in the ability of leadership and staff at fintechs to assess the local context and understand the structural vulnerabilities that could affect data or lead to the destabilization of model performance. Surveys of machine learning practitioners across sectors have identified biases and lack of diverse experiences among the teams developing algorithms as a constraint to anticipating and surfacing fairness issues.[78]

POTENTIAL SOLUTIONS

Tools to Integrate into Product Development

Academics, think tanks, and auditing companies have put forward numerous methodologies based on ethics and other impact frameworks to help organizations think through intended and unintended consequences of their products on people in the real world. One such tool is the Ethical Matrix, created by Cathy O'Neil and offered through her consultancy, which goes beyond model inputs to assess the impact on the people affected by the system's decisions, mapping out different stakeholders and the consequences they may experience due to an algorithm's intended use or failures.[79]

The AI Blindspots project posits questions through a set of cards designed to be used during the model planning, delivery, and deployment stages to help teams avoid unconscious bias and replicating structural inequality.[80] AI Now, a think tank, developed a framework to assess public sector or government applications of AI

through its algorithmic impact assessments.[81] One of the few tools grounded in the international development context, but not specifically focused on digital finance, is the Net Hope Toolkit, which serves as an introduction for nonprofit and development practitioners to AI ethics and fairness principles, as well as workshops and materials to guide conversations on the suitability of AI-based solutions for development.[82]

Another group of techniques aims to foster discussion of impact and ethical considerations at the outset of model development to anticipate potential risks, document intentions and concerns, design testing, and determine procedures at different points in the model development and deployment stages to monitor and revisit assumptions. Given the highly iterative product development lifecycle and different needs across industries, these tools are structured to be flexible mechanisms for accountability. While all of these tools have adaptable elements, they share a common purpose of defining goals, risks to stakeholder groups, or some form of bias impact statement at the outset.[83]

For inclusive fintech, understanding how to select and apply a framework to use for fairness evaluations is still unsettled, with little pilot testing of these different available approaches. Fintechs report talking about structural vulnerabilities at a strategic level but often do not have the resources or staff to customize models for deployment in different contexts. While some fintechs have documentation practices in place and ad hoc forums for discussion, none of those interviewed had piloted specific frameworks in the product development process.

Others expressed skepticism of business rules as a wholesale solution to mitigating bias without other measures in place. As one CEO said, "If there is an opaque model, whatever it's driven by, that is excluding a customer, whatever business rules are on top of it don't matter... Only in retrospect and only if they have an incentive to, is any lender going to go back and retroactively look at the cases where customers were denied and try to change it." Some level of understanding around the drivers of a model are required for business rules to have an impact.

Staff Diversity

A fintech operating in Asia with predominantly local leadership and staff highlighted an example of how their contextual knowledge helped them shift their lending models from one context to another. An important driver of their decisions in one city rested on a customer's ownership of their shop or home. The credit team recognized that these criteria would eliminate migrants as customers since they did not own property, especially those seeking opportunities in expensive cities, yet could still be creditworthy customers.

Unfortunately, the inclusive finance industry is a story of concentration rather than diversity of leadership. There is limited data on industry-wide trends, but recent analysis has shown that the bulk of investments in inclusive fintech are concentrated among companies with headquarters in a small number of wealthier cities. Mostly, these companies are led by men and there is a large investment gap between expat and local founders.[84] This matters because humans building technology have biases of their own. A lack of diverse experiences among teams, especially experiences in the local context where models are set to deploy, will hinder the company's ability to anticipate and surface fairness issues.

CONTEXT SUMMARY

Biases and Limitations

The operating context of a provider creates additional openings for biases. For example, lack of diversity among teams can make it difficult to detect issues. And when third-parties are hired to develop algorithms, they may produce a solution that no one at the company fully understands and which is difficult to manage over time.

Potential Solutions

Providers should integrate tools into the product development process to define goals, identify risks, and explicitly aim to mitigate against bias before model development begins. Diversifying teams is also an important step to managing potential openings for bias.

State of Practice in Inclusive Finance: Early Days

While fintechs show an awareness of the importance of bias and exclusion, most are only at an early stage of mitigating against these risks. This reality is also reflected in larger cross-sectoral surveys of AI developers who have called for domain-specific tools as well as voiced concern around internal capacity, such as time or staff dedicated to understanding fairness.[85] Many fintechs are also operating amid regulatory uncertainty, as new data frameworks are being passed but the capacity for enforcement is limited and unclear.

Additionally, the tradeoffs for regulators between risk and opportunity currently seem tilted towards the latter. One data protection law professor described the attitude as an approach that sees that "the benefits [of algorithms] are immediate and real; the potential harm is gradual and distributed."[86] Another regulator from East Africa noted that until an algorithm has been proven to be risky, "Let's have an algorithm before we think about risks related to algorithms. The risks are something that come afterward...to be frank, it's something we haven't started to think so much about."[87]

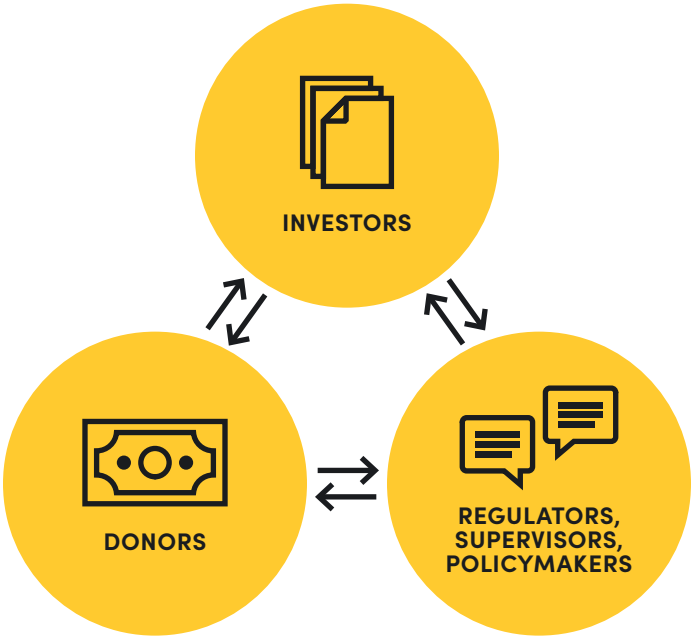
Technological developments will always keep regulators searching for the best ways to approach the challenge of protecting consumers while fostering growth and business opportunities. And while new data protection regulation ostensibly gives consumers new rights, they place the onus of action on the individual. Realistically, how likely are low-income consumers to take advantage of these rights and understand their responsibilities? Additionally, as regulatory-based rights and recourse are currently framed around harm at the individual level and the exclusionary impacts of algorithms might be occurring at the group level, this deserves more attention and concern.[88] CFI is conducting small-scale open-ended qualitative work in Rwanda with digital borrowers to shed some light on these dynamics and plans to do more in the future.

3 Section 3: A Learning Agenda for the Path Forward

Given the swirl of unknowns around the deployment of algorithms in inclusive finance, we recommend a learning agenda to support responsible and inclusive lending; many of the topics can also be applied to a wider set of products and business models. These are a broad set of questions and topics, and their breadth signals the large unmet need for useful, focused, feasible, and inclusive evidence to guide the field forward.[89] It's critical that we begin the search for answers now given the rapidly transforming global data ecosystem as well as the gaps in agency between those designing the algorithms and those impacted by them.

While inclusive finance is transforming from unprecedented data science and technical capabilities, advancing this learning agenda requires all stakeholders. Only a subset of inclusive finance employees has the technical prowess to code and create the mathematical recipes that determine who becomes a customer, but these same data scientists and developers would be among the first to say that, given the stakes of the decisions that their algorithms make, other types of actors must be part of the conversation. As one data scientist told us, "We know that [AI bias] reflects the reality, and if the reality is unjust or inequitable you can't really program it out of your algorithms." We would emphasize that the perspective of other actors, both provider staff and outside stakeholders, bring crucial insights on the contexts in which, and the consumers for which, algorithms are being deployed. Most of us will not become data scientists, but that should not dissuade from engaging in the data-driven reshaping of inclusive finance. The work must of course involve providers, but our focus here is on recommendations for the broader ecosystem including donors, investors, and governments—who can support the vast learning agenda.

ACTION REQUIRED BY MULTIPLE STAKEHOLDERS



Donors

Donors have a vital role to play in supporting the generation of evidence and testing scalable solutions that advance fair and inclusive financial services. Donor support should be directed at a robust learning agenda rather than direct subsidies to providers. An advocacy agenda may also be necessary as well, particularly in putting these issues on the radar of governments. Engagement with these issues, which we found to be relatively low in emerging markets, is the first step towards building knowledge and skills to address them. Following awareness building, donors should also support supervisors to engage with providers and develop accountability measures.

SUPPORT INCLUSIVITY FRAMEWORKS FOR FINTECHS

The questions around algorithms and exclusion feed into a broader chorus of voices asking whether fintechs are reaching new, underserved populations or merely reaching consumers who already have digital options. While the potential to reach underserved populations is widely touted, actual fintech outreach to those segments is anecdotal and not captured systematically. Donors should support the development of inclusivity frameworks that leverage relevant, reportable fintech-level data as well as customer data to understand outreach and ideally impact; this would likely involve demand-side research to validate the selection of reported data.

Complementary to this framework is support for providers to invest in systems to identify and onboard previously unbanked customers. As an example, one early stage fintech in South America set up an assessment framework that included consumer surveys and other portfolio metrics to monitor alongside model performance. The company keeps this data and demographic data collected at the know-your-customer (KYC) stage completely separate from the data used in the credit model, and only uses it for disparate impact assessments.

TEST AND ADAPT TOOLS TO IMPROVE FAIRNESS OF ALGORITHMS

This paper has presented many new tools to mitigate the risk of bias at the levels of inputs, code, and context. They include: a) methods for understanding and anticipating bias from data inputs; b) technical tools to test code for disparate impact; and c) documentation techniques that enable auditing and improved communication among stakeholders, among others.

These methods need to be tested with inclusive finance companies to understand gaps and limitations that may be sector-specific, as well as to identify the cost of compliance. While these methodologies come from researchers in many industries that use machine learning, their feasibility has not been tested widely in terms of effectiveness in mitigating bias against the time and resources required for implementation.

Fintech associations have emerged in many markets, and while initially focused on incubating companies, in several markets like Indonesia and Kenya, they have taken on responsible finance agendas and could play a role in developing industry guidance on responsible use of algorithms. They should be supported to introduce issues around bias to their members and encourage collaboration on pilot testing bias mitigation strategies and sharing learnings with local authorities and other actors.

PRIORITIZE CONSUMER RESEARCH

The (already limited) demand-side work with consumers on data protection, perceptions of algorithms, and emerging data rights tends to be clustered in advanced economies. While useful, this work is not easily extrapolated to markets that are grappling with large-scale onboarding to digital finance, gaps in digital capability, and large swaths outside the formal financial system. Opportunities abound to learn more about consumer attitudes, perceptions, and trust of alternative data and algorithms, as well as to test implementation approaches for data rights.

SUPPORT THE EVOLUTION OF FINANCIAL INFRASTRUCTURE

While credit bureaus are not new, functioning credit bureaus take on greater urgency in the context of real-time credit decisions. For digital lenders to use credit bureau data in real time, credit bureaus must address integration challenges, like APIs, and accuracy concerns in their data. Similarly, information sharing between fintechs and credit bureaus that allows for flagging of reports during experimental deployments of an underwriting algorithm could reduce blacklisting and consumer harm.[90] Capacity building for data reporting, integration of new sources, and trend analysis to meet the scale and speed of fintech should be integrated into inclusive finance strategies so that donors and others can offer support.

DEVELOP FRAMEWORKS TO CONDUCT MARKET MAPPINGS

Datasets used for inclusive finance objectives must be employed with an awareness of the systemic ways in which existing data collection mechanisms may fail to capture the full picture of a marginalized segment's behavior and experiences. Market-level research should be carried out to map and interrogate the existing data sources leveraged for inclusive finance and how they intersect with marginalized groups, access to technology, historic deprivations, social norms, and other data idiosyncrasies particular to the context. They should also surface other potentially overlooked sources of data that could be digitized, such as supply chains, that might help to create data trails to lessen the reliance on phone-generated data trails. These exercises should be interdisciplinary, and ideally also involve impacted communities as well as social justice and consumer protection organizations.

As part of this, donors should also look for ways to strengthen local capacity for both data science as well as privacy law, so that developers and programmers come to the table with stronger contextual knowledge and background. Projects like Datahack4FI that were a mix of capacity building with competitions or hackathons could provide lessons for other likeminded donors who are looking for the most effective ways to champion those efforts.[91]

IDEA: Donors might also think about how to advance their own internal approach through monitoring, evaluation, and learning (MEL) plans with grantees. While they may not give grants solely for the deployment of algorithms, automated decision-making and algorithms are increasingly components of larger intervention. For instance, an underwriting algorithm might help decide which farmers receive a multi-pronged "credit +" capacity building intervention which includes a loan, additional technical assistance, and training for new agricultural technology. While a larger MEL framework focuses on outcomes and impacts on the farmers, we suggest having several indicators that push grantees to think critically about the fairness and inclusivity of the algorithm. This may require greater adaptability and flexibility of monitoring and evaluation to better align with machine learning projects.

SUPPORT EFFORTS TO IMPROVE CONSUMER DIGITAL CAPABILITY AND CONSUMER RIGHTS AGENDAS

To avoid existing digital footprints driving future financial access, donors should support efforts to build digital and financial capability for low-income customers. This is not as simple as handing out a smartphone to everyone, but a deeper process of transforming norms and capabilities. It might also entail targeted support for the digitalization of existing financial service providers, such as Village Loan and Savings Associations (VSLA) and microfinance institutions, in addition to supporting the digital capability of individuals.

International consumer advocacy organizations, such as Consumers International, are increasingly recognizing the importance of digital rights, but few local consumer organizations have adopted robust advocacy agendas around data. As technology adoption increases in developing markets, consumer organizations must build awareness of how consumer harm is linked to data privacy infringement and can play a crucial role in documenting the impact of different policy actions on consumers.

Investors

Without incentives for providers, adoption of any systematic strategies to look for and mitigate bias will likely continue to be a low priority for capacity-constrained fintechs. Investors have an important role to play in establishing incentives to draw attention to the importance of responsible algorithms and digital lending. Especially in environments where regulatory and supervisory oversight is limited, investors can play a role in shaping provider practice, for instance through past support and implementation of the Client Protection Principles.

Investors should leverage key moments such as due diligence and the drafting of covenants or grant agreements to incorporate monitoring for responsible practices. These efforts would help nudge providers toward reflection and action as well as help investors determine their own internal vocabulary for how to define and incorporate responsible lending practices into their processes. Investors might even consider introducing "fairness KPIs" for their investees to measure outputs like disparate impact. Investors should also work with their portfolio companies to ensure that training data is representative. These efforts may require a greater acceptance of defaults not only by the provider, but also by the investor. Investors and their portfolio companies should also acknowledge and work to better understand the potential tradeoffs these efforts may have on rapid growth, profit, and risk tolerance. As investors are engaging in algorithm-driven initiatives across development sectors, including health, finance, agriculture, and the environment, there are likely to be cross-portfolio learnings as well.

DOCUMENT AND SCREEN FOR RESPONSIBLE ALGORITHM PRACTICES

That algorithms are proprietary and walled off from scrutiny should not put off robust discussions around the connective tissue that governs the decision-making process itself. On the following page are questions we recommend investors bring forth with prospective investees.[vi]

REQUIRE AUDITS

Given the early stage of anti-bias tools, investors should not yet require strict auditing standards among their portfolios, but use their relationships to incentivize and test different approaches. While there are various auditing frameworks and a long set of tools to test for discrimination, there is no consensus on which tools will work best in the context of inclusive finance.[92,93,94] We suggest drawing from some of the tools mentioned in this paper as a starting point to engage with portfolio companies.

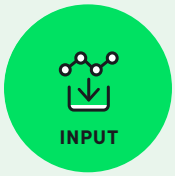
ALIGN KPIS TO INCENTIVIZE INCLUSIVE LENDING

Investors can leverage existing reporting mechanisms such as KPIs to incorporate metrics and learning agendas around algorithms and exclusion, as it aligns with their own missions. One CEO characterized KPIs as investor directives rather than incentives: "Usually [startups] have one or two or three KPIs that they're trying to optimize for and nothing else really matters…that tends to be…associate[d] with growth—not even quality of portfolio just customer acquisition." Goals around impact can be at odds with investor-set KPIs, especially those coming from investors outside the social impact sector. Multiple fintechs spoke about the pressure to cut the cost of customer acquisition and focus chiefly on growth: "When it comes to day-to-day working, growth takes over everything else," one CEO remarked.

A fintech interviewed that focuses on including more women in its customer base has dealt with this tension between incentives and social mission. The fintech recognizes that onboarding more women will require investments in research to understand women's needs and barriers, product design changes, and marketing initiatives, which is in conflict with investor KPIs on lowering customer acquisition costs to meet growth targets. A restructuring of some of those incentives is needed to tackle questions of bias and to fuel long-term sustainable business growth, before companies scale. As one fintech put it, "…You can't make those [anti-bias] changes when you're doing a million loans a day or 500,000 loans a day. You can do those changes when you're doing maybe 15,000–20,000 loans a day…we're aware of where we want to go as a business and a little bit of a loss now and a gain later is better than taking a loss at that scale."

vi If rejection rates are high, the algorithm is most likely very selective. Thus, it has a negative impact on financial inclusion. If the rejection rate is low, the difference between individuals might be in the pricing, term, and amount.

DUE DILIGENCE QUESTIONS FOR INVESTORS

	QUESTIONS	EVIDENCE SHOULD DEMONSTRATE
	↗ What are your data sources? What are the quality control measures for data quality, accuracy, and stability? ↗ How much of the data used in your model is generated in-house or through your own observations vs. acquired through third parties or data brokers? ↗ How are data sources representative of the population you are aiming to serve? What market dynamics, such as historical or current discrimination, are you tracking that may affect data availability and quality?	↗ The company shows awareness of data quality risks and plans to improve representativeness or monitor for potential bias based on data inputs and knowledge of local market context. ↗ Established timetables for reviewing in-house data quality and relationships with third-party vendors.
	↗ What are the rejection rates of your model? Is the rejection justifiable or indicative of a possible bias? ↗ Which features drive credit decisions? ↗ Do you test for and track disparate impacts of your model? What is the methodology? ↗ Have you surfaced or do you suspect that there might be proxies for excluded categories that drive credit decisions?	↗ An established definition of fairness and corresponding tools identified to monitor based on that definition. ↗ Management has a basic understanding of model features and drivers, and can explain whether they are based on criteria that is directly or indirectly related to financial behaviors. ↗ Established thresholds and improvement targets for rejection rates and priority groups.
	↗ Who is responsible for developing and managing the data sources and the code? ↗ Does your data science team interact with staff that is knowledgeable around the market context where the model will be deployed? ↗ What documentation processes are in place for model performance and decisions and are they shared with management/board, etc.? ↗ How do you communicate to consumers who have been denied a loan? What are the key messages?	↗ Diverse staff that is representative of the country where the product is deployed. ↗ Activities to build non-technical staff capacity, especially when third parties are involved in algorithm development. ↗ Policy on customer communications that shows sensitivity to different digital literacy levels.

Regulators, Supervisors, and Policymakers

Governments need to engage stakeholders, build enforcement capacity and independent controls, and better understand the challenges and opportunities around algorithms and AI. The path forward should balance capacity and priorities as well as the need to advance responsible data practices.

The enforceability of data privacy and rights frameworks are still being tested, especially in markets with lower supervisory capacity. With such a large learning agenda ahead, it is too early to provide universal policy prescriptions on how to incorporate oversight of algorithmic decisions into policy frameworks to address financial exclusion. However, we suggest that in drafting and implementing legislation and regulation, policymakers keep consumers top of mind, especially as data protection frameworks are often followed or accompanied by open banking frameworks. Regulatory frameworks introduced into these contexts must contend with the gap between giving people new digital rights and their capacity to manage and benefit from those rights.

In countries that have enacted data privacy or data rights regimes, consumers have newfound rights and responsibilities. While a positive development, even in developed markets consumers do not fully grasp what it all means in practice, showing limited understanding of consent terms or how their data is used.[95] The March 2019 Eurobarometer, a direct survey of residents in 28 EU member-states, asked respondents about their awareness and action vis-à-vis their newfound data rights. Slightly more than half of respondents (57 percent) were aware of their right to have their data deleted, while 41 percent had heard of their right to have a say when decisions are automated. Eighteen percent of respondents had exercised their right to access their data, though the survey does not specify in which sector these rights were exercised.[96] It is unlikely that consumers in emerging markets will have visibility on whether a lender had collected inaccurate or incomplete data, and if third parties like data brokers were the source of the mistake, the sources are further obscured from the consumer view.[97]

MARKET MONITORING

Given capacity constraints, there is an opportunity to create incentives for companies to do the bulk of the monitoring and testing while at the same time build internal supervisory capacity to monitor those institutions and the market. Even in this scenario, government agencies will need to play a central role and develop a robust internal capacity to engage with tech-driven companies. A regulator from California mentioned during our interviews that integrating the "Do not touch our business" mentality of the private sector with the "Help us better understand what you are doing" mindset of the regulators could be in the best interest of both parties and a path forward.[98]

SPACE TO TEST AND LEARN

Innovation facilitators such as fintech associations and incubators, regulatory sandboxes, innovation hubs, or hotlines can act as an iterative tool for governments to collect, test, and acquire evidence on emerging and innovative technology as part of its determination for how and whether to regulate it.[99] The use of alternative data and advanced algorithms presents risks that are difficult to evaluate in the abstract, and several governments have constructed engagements with innovators in order for them to deepen their knowledge and test approaches.[100] Since 2017 in the United States, the Consumer Financial Protection Bureau (CFPB) has worked with Upstart Networks, an online lender that leverages alternative data, through a no-action letter. The no-action letter signals that the CFPB does not intend to take any enforcement action against Upstart under the Equal Credit Opportunity Act, and in exchange, Upstart will share with CFPB information about loan applications, its decision methodology, and whether its model actually expands access to the underserved.[101]

Conclusion

This paper focuses on bias in digital credit underwriting, but there are dozens of other products, services, platforms, and business models that merit their own inquiries and careful study. CFI is committed to doing research, sharing evidence, and testing solutions to advance its learning agenda and this paper is just the first step.

The inclusive finance sector is not alone in addressing the challenge of defining fairness and mitigating against bias in its use of algorithms. Every week there seems to be a headline around bias and algorithms gone awry, whether from criminal justice, healthcare, or even the tech sector itself.[102] The forces shaping how the issues play out are fluid and unpredictable and, as this paper illustrates, the list of open questions is vast and multifaceted. The stories told by data are increasingly important to self-determination and economic opportunity, and we are committed to ensuring that those stories reflect the full potential of all consumers.

Appendix: Referenced Tools

INPUT:

Design Checklists—Used in the aerospace industry, these are tools that ask designers for descriptions of processes they've undertaken to address important questions, prior failures, or edge cases.

Failure Modes and Effects Analysis (FMEA)—A standard engineering procedure, this a methodology for ex ante examination of a proposed technology or design for potential failures through research and literature reviews on similar technology deployments and known risks.

Datasheets for Datasets—Developed for machine learning use by Microsoft Research, this tool uses best practice from the electronics industry to document requirements, recommended uses, and other information and characteristics about a dataset to facilitate information sharing about the use of data for certain purposes. Available at: https://www.microsoft.com/en-us/research/project/datasheets-for-datasets/

CODE:

IBM Toolkit: AI Fairness 360—An open source toolkit in Python or R code to examine and mitigate bias in machine learning models, structured around AI product development and application lifecycle. Available at: https://aif360.mybluemix.net/

What-If Tool—A tool developed by Google AI researchers to allow users with minimal coding skills to visually investigate trained regression and classification models to test performance of hypothetical situations, importance of features, and subsets of input data for different machine learning definitions of fairness. Available at: https://pair-code.github.io/what-if-tool/

Pymetrics/audit-ai—An open source tool to measure the effects of discriminatory patterns in training data and determine what traits fed into an algorithm are driving outcomes that lead to adverse impacts on some groups of people. Available at: https://github.com/pymetrics/audit-ai

Model Cards for Model Reporting—A tool to log the performance benchmarks of a trained machine learning model across demographic and other groups, as well as the methods used to evaluate performance, intended uses and contexts for deployment, and any other ethical considerations.

CONTEXT:

Ethical Matrix—A tool designed by Cathy O'Neil (author of *Weapons of Math Destruction*) and offered through her consultancy ORCAA, which maps the potential impacts of an algorithm on different stakeholder groups, and the consequences they may experience due to an algorithm's intended use or failures.

AI Blindspots—A tool developed by practitioners at the MIT Media Lab and Berkman Klein Center that uses a set of cards to pose questions to AI developers through an "AI Blindspot Discovery Process," beginning with defining the purpose of the system through to providing guidance for how individuals can contest decisions. Available at: https://aiblindspot.media.mit.edu/index.html

Algorithmic Impact Assessment—A practical framework designed by researchers at the think tank AI Now, modeled off of environmental impact assessments, to shed light on automated decision systems used in the public sector and set up a process for greater accountability with affected communities.

Net Hope Toolkit—A set of guides and workshop materials to facilitate discussions at nonprofit organizations about fairness, bias, and the suitability of AI-based solutions in international development. Available at: https://solutionscenter.nethope.org/artificial-intelligence-ethics-for-nonprofits-toolkit

Notes

1 "Digital 2020: 3.8 Billion People Use Social Media," We Are Social, January 30, 2020, https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media.

2 Carboni, Isabelle and Hennie Bester, "When Digital Payment Goes Viral: Lessons from COVID-19's impact on mobile money in Rwanda," Cenfri, May 19, 2020, https://cenfri.org/articles/covid-19s-impact-on-mobile-money-in-rwanda/.

3 Meka, Sushmita et al., "Artificial Intelligence: Practical Superpowers," BFA Global, May 2018, https://bfaglobal.com/wp-content/uploads/2019/03/FIBR-Artificial_Intelligence_FINAL_MAY2018-1.pdf.

4 "The New Physics of Financial Services: How Artificial Intelligence Is Transforming the Data Ecosystem," Deloitte & World Economic Forum, August 2018, http://www3.weforum.org/docs/WEF_New_Physics_of_Financial_Services.pdf.

5 Di Castri, Simone et al., "An AML Suptech Solution for the Mexican National Banking and Securities Commission (CNBV)," August 2018, https://www.r2accelerator.org/aml-data-infrastructure-prototype.

6 Robinson, David G. "Reclaiming the stories that algorithms tell," O'Reilly, May 27th 2020, https://www.oreilly.com/radar/reclaiming-the-stories-that-algorithms-tell/.

7 Mowl, Amy Jenson and Camille Boudout, "Barriers to Basic Banking: Results from an Audit Study in South India," IFMR, 2015.

8 Bartlett, Robert, Adair Morse, Richard Standon and Nancy Wallace, "Consumer Lending Discrimination in the FinTech Era," Working Paper 25943, Working Paper Series, National Bureau of Economic Research, 2019.

9 Vincent, James, "Apple's credit card is being investigated for discrimination against women," The Verge, November 11, 2019, https://www.theverge.com/2019/11/11/20958953/apple-credit-card-gender-discrimination-algorithms-black-box-investigation.

10 "HUD Charges Facebook with Housing Discrimination Over Company's Targeted Advertising Practices," HUD News Release, March 28, 2019, https://archives.hud.gov/news/2019/pr19-035.cfm.

11 Obermeyer, Ziad et al., "Dissecting racial bias in an algorithm used to manage the health of populations," *Science* Vol. 366, Issue 6464 (October 25, 2019): 447-453, https://science.sciencemag.org/content/366/6464/447.

12 Spice, Byron, "Questioning the Fairness of Targeting Ads Online," *Carnegie Mellon News*, July 7, 2015, https://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html.

13 Garz, Seth et al., "Consumer Protection for Inclusive Finance in Low and Middle Income Countries: Bridging Regulator and Academic Perspective," *National Bureau of Economic Research*, Working Paper 28262, December 2020, http://www.nber.org/papers/w28262.

14 Obermeyer et al., "Dissecting Racial Bias."

15 Obermeyer et al., "Dissecting Racial Bias."

16 "Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment," Consumer Financial Protection Bureau, 2014, https://www.consumerfinance.gov/data-research/research-reports/using-publicly-available-information-to-proxy-for-unidentified-race-and-ethnicity/.

17 Reike, Aaron et al., "Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods," Upturn & Omidyar Network, February 2018, https://omidyar.com/wp-content/uploads/2020/09/Public-Scrutiny-of-Automated-Decisions.pdf.

18 Rolf, Esther et al., "Balancing Competing Objectives for Welfare-Aware Machine Learning with Imperfect Data," AI for Social Good workshop at NeuroIPS, (2019).

19 Kroll, Joshua et al., "Accountable Algorithms," UPenn Law Review, (2017).

20 Hunt, Robert and Fenwick McKelvey, "Algorithmic Regulation in Media and Cultural Policy," *Journal of Information Policy* Vol. 9 (2019): 307-335.

21 World Bank, 2017 Global Inclusive finance and Consumer Protection Survey. As cited in Seth Garz et al., "Consumer Protection."

22 Rice, Lisa and Deidre Swesnik, "Discriminatory Effects of Credit Scoring on Communities of Color," *Suffolk University Law Review* (2013).

23 Kessler, Alex, "It's Time to Protect Kenyans from a Digital Lending Laboratory," Center for Financial Inclusion, February 26, 2020, https://www.centerforfinancialinclusion.org/its-time-to-protect-kenyans-from-a-digital-lending-laboratory.

24 "Making Digital Credit Truly Responsible: Insights from analysis of digital credit in Kenya," MSC, September 2019, https://content.centerforfinancialinclusion.org/wp-content/uploads/sites/2/2019/09/Digital-Credit-Kenya-Final-report.pdf.

25 Izaguirre, Juan Carlos et al., "Digital Credit Market Monitoring in Tanzania," CGAP, September 2018, https://www.cgap.org/research/slide-deck/digital-credit-market-monitoring-tanzania.

26 "Doing Business 2020: Comparing Business Regulation in 190 Economies," World Bank Group, 2020, https://openknowledge.worldbank.org/bitstream/handle/10986/32436/9781464814402.pdf.

27 Björkegren, Daniel and Darell Grissen, "Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment," 2019, https://arxiv.org/ftp/arxiv/papers/1712/1712.05840.pdf.

28 World Bank 2019 as cited in Seth Garz et al., "Consumer Protection."

29 Wang, Andy, "The Digital Desert," *Harvard International Review*, Vol. 41, No. 1 (Winter 2020): 37-40.

30 Rowntree, Oliver et al., "The Mobile Gender Gap Report 2020," GSMA, March 2020, https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/02/GSMA-The-Mobile-Gender-Gap-Report-2020.pdf.

31 Kapinga, Alsen et al., "Mobile marketing applications for entrepreneurship development: Codesign with women entrepreneurs in Iringa, Tanzania," The Electronic Journal of Information Systems in Developing Countries, January 2019, https://onlinelibrary.wiley.com/doi/10.1002/isd2.12073.

32 Wyche, Susan et al., "Understanding women's mobile phone use in rural Kenya: An affordance-based approach," Mobile Media and Communication, July 2018, https://journals.sagepub.com/doi/full/10.1177/2050157918776684.

33 Roggemann, Kristen et al., "User Perceptions of Trust and Privacy on the Internet," DAI, October 2020, https://www.dai.com/fi-cyber-user-trust-summary.pdf.

34 Sonne, Lina, "What do we Know about Women's Mobile Phone Access and Use? A review of evidence," Dvara Research Working Paper Series No. WP-2020-03, August 2020, https://www.dvara.com/research/wp-content/uploads/2020/06/What-Do-We-Know-About-Womens-Mobile-Phone-Access-Use-A-review-of-evidence.pdf.

35 Ibid.

36 Yu, Persis and Jillian McLaughlin, "Big Data: A Big Disappointment for Scoring Consumer Credit Risk," National Consumer Law Center, March 2014, https://www.nclc.org/images/pdf/pr-reports/report-big-data.pdf.

37 "How to Prevent Discriminatory Outcomes in Machine Learning," World Economic Forum, March 2018, http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf.

38 Awwad, Yazeed et al., "Exploring Fairness in Machine Learning for International Development," MIT D-Lab, March 2020, https://d-lab.mit.edu/resources/publications/exploring-fairness-machine-learning-international-development.

39 Björkegren and Grissen, "Behavior Revealed."

40 Reike et al., "Public Scrutiny."

41 Raji, Inioluwa Deborah et al., "Closing the AI Accountability Gap: Defining an internal end-to-end framework for internal algorithmic auditing," In Conference on Fairness, Accountability, and Transparency (FAT* '20), January 27–30, 2020, https://dl.acm.org/doi/abs/10.1145/3351095.3372873.

42 Ibid.

43 Gebru, Timnit et al., "Datasheets for Datasets," Microsoft Research, 2018, https://www.microsoft.com/en-us/research/uploads/prod/2019/01/1803.09010.pdf.

44 Gilman and Green, "The Surveillance Gap: The Harms of Extreme Privacy and Data Marginalization," *NYU Review of Law and Social Change* (2017).

45 Awwad, Yazeed et al., "Exploring Fairness in Machine Learning for International Development," MIT D-Lab, (March 2020).

46 Republic of Kenya, "The Data Protection Act 2019," Kenya Gazette Supplement, November 11, 2019, http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/2019/TheDataProtectionAct__No24of2019.pdf.

47 MacMillan, Rory, "Big Data, Machine Learning, Consumer Protection and Privacy," Security, Infrastructure and Trust Working Group and FIGI, 2019, https://www.itu.int/en/ITU-T/extcoop/figisymposium/Documents/FIGI_SIT_Techinical%20report_Big%20data%2C%20Machine%20learning%2C%20Consumer%20protection%20and%20Privacy_f.pdf.

48 California Consumer Protection Privacy Act 2018. Available at: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.

49 Interview with Andrew Selbst, October 2, 2020.

50 MacMillan,"Big Data."

51 Interview with Stephan Dreyer, October 2, 2021.

52 "Data Dilemmas," Ipsos Global Trends, 2020, https://www.ipsosglobaltrends.com/2020/02/data-dilemmas/.

53 Medine, David and Gayatri Murthy, "Making Data Work for the Poor: New Approaches to Data Protection and Privacy," CGAP Focus Note, January 2020, https://www.cgap.org/sites/default/files/publications/2020_01_Focus_Note_Making_Data_Work_for_Poor_0.pdf.

54 Medine, David and Maria Fernandez Vidal, "Study Shows Kenyan Borrowers Value Data Privacy, Even During Pandemic," CGAP Blog Series: Data Privacy and Protection, July 30, 2020, https://www.cgap.org/blog/study-shows-kenyan-borrowers-value-data-privacy-even-during-pandemic.

55 Medine, David and Gayatri Murthy, "India's Proposed Data Protection Bill Breaks from Notice and Consent," CGAP, March 9, 2020, https://www.cgap.org/blog/indias-proposed-data-protection-bill-breaks-notice-and-consent.

56 Hurley, Mikella and Julius Adebayo, "Credit Scoring in the Era of Big Data," *Yale Journal of Law and Technology* (2017), https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1122&context=yjolt.

57 Johnson, Kristin, Frank Pasquale, and Jennifer Chapman. "Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation." *Fordham Law Review* (2019), https://ir.lawnet.fordham.edu/flr/vol88/iss2/5/.

58 Robinson, David and Harlan Yu, "Knowing the Score: New Data, Underwriting, and Marketing in the Consumer Credit Marketplace," 2014, https://www.upturn.org/static/files/Knowing_the_Score_Oct_2014_v1_1.pdf.

59 Al-Zuabi, Ibrahim et al., "Predicting Customer's Gender and Age Depending on Mobile Phone Data," *Journal of Big Data*, February 19, 2019, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0180-9; and Seneviratne, Suranga et al., "Your Installed Apps Reveal Your Gender and More," 2015, https://dl.acm.org/doi/10.1145/2721896.2721908.

60 Barocas, Solon and Andrew Selbst, "Big Data's Disparate Impact," *California Law Review* 104, issue 3 (2016), https://lawcat.berkeley.edu/record/1127463.

61 Rice, Lisa and Deidre Swesnik, "Discriminatory Effects of Credit Scoring on Communities of Color," *Suffolk University Law Review* 46, no 3 (2013).

62 Madden, Mary, "Privacy, Security, and Digital Inequality," Data and Society, September 27, 2017, https://datasociety.net/library/privacy-security-and-digital-inequality/.

63 Tools include: IBM Fairness 360 toolkit (https://aif360.mybluemix.net/), Pymetrics audit-AI (https://github.com/pymetrics/audit-ai), and Google's What-If Tool (https://pair-code.github.io/what-if-tool/).

64 Mitchell, Margaret et al., "Model Cards for Model Reporting," FAT* '19: Conference on Fairness, Accountability, and Transparency, January 29–31, 2019, https://arxiv.org/abs/1810.03993.

65 Interview with Fintech.

66 Hurley and Adebayo, "Credit Scoring."

67 Interview with Andrew Selbst, October 2, 2020.

68 Article 22 of GDPR, as cited in Macmillan, "Big Data."

69 Wachter, Sandra, "Towards accountable AI in Europe?" The Alan Turing Institute, July 2017, https://www.turing.ac.uk/news/towards-accountable-ai-europe.

70 Wachter, Sandra, Brent Middlestadt, and Chris Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," Harvard Journal of Law & Technology, 2018, https://arxiv.org/abs/1711.00399. As cited in Macmillan, "Big Data."

71 Samuel, Sigal, "It's disturbingly easy to trick AI into doing something deadly," Vox, April 8, 2019, https://www.vox.com/future-perfect/2019/4/8/18297410/ai-tesla-self-driving-cars-adversarial-machine-learning.

72 Bjorkegren, Daniel et al., "Manipulation-Proof Machine Learning," Working Paper, May 28, 2020, https://arxiv.org/abs/2004.03865.

73 Holstein, Kenneth et al., "Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?" CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, https://dl.acm.org/doi/10.1145/3290605.3300830.

74 Hurley and Adebayo, "Credit Scoring."

75 Holstein et al., "Improving Fairness."

76 Awwad et al., "Exploring Fairness."

77 "Inclusive Fintech 50: Driving Inclusive Finance Amid Crisis." Inclusive Fintech 50, December 2020, https://www.inclusivefintech50.com/white-paper-2020.

78 Holstein et al., "Improving Fairness."

79 Sekinah, Toni, "Assess biased algorithms with an ethical matrix," dataIQ, December 2019, https://www.dataiq.co.uk/articles/assess-biased-algorithms-with-an-ethical-matrix.

80 Calderon, Ania et al., "The AI Blindspots cards," Berkman Klein Center and MIT Media Lab, 2019, https://aiblindspot.media.mit.edu/.

81 Reisman, Dillon et al., "Algorithmic Impact Assessments: A practical framework for public agency accountability," AINOW, April 2018, https://ainowinstitute.org/aiareport2018.pdf.

82 Toplic, Leila and Nora Lindstrom, "AI Ethics for Nonprofits Toolkit," Net Hope Solutions Center, December 2020, https://solutionscenter.nethope.org/artificial-intelligence-ethics-for-nonprofits-toolkit.

83 Lee, Nicol Turner, Paul Resnick, and Genie Barton, "Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms," Brookings, 2019, https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.

84 Inclusive Fintech 50, "Driving Inclusive Finance."

85 Raji et al., "Closing the AI Accountability Gap."

86 Interview with regulator from sub-Saharan Africa, November 2020.

87 Interview with regulator from BNR, November 2020.

88 Taylor, Linnet, "What is data justice? The case for connecting digital rights and freedoms globally," Big Data and Society, July-December 2017.

89 "USAID Learning Questions Checklist," USAID Learning Lab, December 11, 2018, https://usaidlearninglab.org/library/learning-questions-checklist.

90 Mazer, Rafe, "Emerging Data Sharing Models to Promote Financial Service Innovation: Global trends and their implications for emerging markets," FSD Kenya, 2018, https://www.fsdkenya.org/publications/emerging-data-sharing-models-to-promote-financial-service-innovation-global-trends-and-their-implications-for-emerging-markets/.

91 See website for now closed project DataHack4FI https://datahack4fi.org/.

92 Raji et al., "Closing the AI Accountability Gap."

93 Barocas, Solon et al. "Fairness and Machine Learning," FAIRMLBook, 2019, https://fairmlbook.org/.

94 Holstein et al., "Improving Fairness."

95 Whitley, Edgar and Roser Pujadas, "Report on a study of how consumers currently consent to share their financial data with a third party," London School of Economics and Political Science, March 2018, fscp_report_on_how_consumers_currently_consent_to_share_their_data.pdf (fs-cp.org.uk).

96 European Commission, Directorate-General for Justice and Consumers, "Special Eurobarometer 487a," June 2019.

97 Grauer, Yael, "What Are Data Brokers and Why are they Scooping Up Information about You?" Vice, March 27, 2018, https://www.vice.com/en/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection.

98 Interview with US-based regulator.

99 Jeník, Ivo, and Schan Duff, "How to Build a Regulatory Sandbox: A Practical Guide for Policy Makers," CGAP, 2020, https://www.cgap.org/research/publication/how-build-regulatory-sandbox-practical-guide-policy-makers.

100 Ibid.

101 "CFPB Announces First No-Action Letter to Upstart Network," CFPB, September 14, 2017, https://www.consumerfinance.gov/about-us/newsroom/cfpb-announces-first-no-action-letter-upstart-network/.

102 Allyn, Bobby, "Ousted Black Google Researcher: They Wanted to Have My Presence, but Not Me Exactly," NPR Technology, December 17, 2020, https://www.npr.org/2020/12/17/947719354/ousted-black-google-researcher-they-wanted-to-have-my-presence-but-not-me-exactl.

The Center for Financial Inclusion (CFI) works to advance inclusive financial services for the billions of people who currently lack the financial tools needed to improve their lives and prosper. We leverage partnerships to conduct rigorous research and test promising solutions, and then advocate for evidence-based change. CFI was founded by Accion in 2008 to serve as an independent think tank on inclusive finance.

www.centerforfinancialinclusion.org

@CFI_Accion

CENTER *for* FINANCIAL INCLUSION | ACCION