INVESTING IN EQUITABLE AI FOR INCLUSIVE FINANCE

A RISK MANAGEMENT GUIDE FOR IMPACT INVESTORS

August 2023

Lead Author: Lucciana Alvarez, Research Manager, Center for Financial Inclusion
Co-Author: Alexandra Rizzi, Senior Research Director, Center for Financial Inclusion

CENTER *for* FINANCIAL INCLUSION | ACCION

USAID FROM THE AMERICAN PEOPLE

ACKNOWLEDGMENTS

This work is the result of USAID's Equitable AI Challenge. Implemented by DAI's Digital Frontiers, the challenge aimed to identify innovative approaches to address artificial intelligence inequitable outcomes. The Equitable AI Challenge asked for proposals that critically consider holistic and creative approaches to identify and address gender biases in AI systems within global development contexts. This work is the result of CFI's winning proposal and provides a guide that considers gender-inequity issues from the outset when designing algorithms for inclusive finance.

This guide was developed in partnership with:

Introduction

Artificial intelligence (AI) enables innovations in digital finance by increasing efficiency, reducing costs, and serving consumers at scale. Whether automating back-end processes or customer-facing decisions, AI has led to breakthroughs in digital payments, digital credit, insurtech, and various support functions such as chatbots, personalized marketing, and robo-advice.[i] Investors have noticed this catalytic potential; venture capital in fintechs and start-ups using AI in the financial and insurance sectors jumped from USD $227 million in 2012 to $15.4 billion in 2022, becoming the third largest amount by sector, after IT and healthcare (see Figure 1). [ii] In 2023, despite a difficult environment for venture capital, investments in generative AI have defied the trend and soared.[iii]

For investors focused on financial inclusion, AI-driven innovations — coupled with the growth of digital data trails from mobile phones, satellites, and other sources — have the potential to create viable business models for historically underserved market segments. AI provides scalable ways to determine the identity or creditworthiness of individuals and businesses that historically lacked identification, collateral, and credit history.[iv] By automating processes, AI can enable higher volumes of low-value transactions that make harder-to-reach segments, like women microentrepreneurs, more viable customers.[v]

Figure 1: VC Investment in AI in Financial and Insurance Sectors
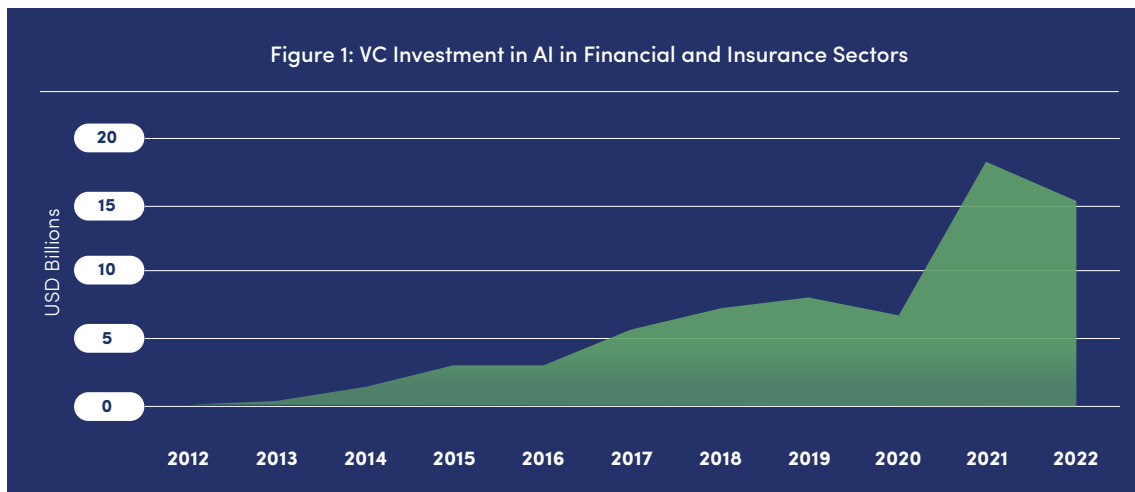
While there are certainly arguments for the use of AI, the deployment of AI also comes with a bevy of risks. One prominent risk is inequitable outcomes for marginalized consumers.[vi] For women, harmful outcomes of AI include lower quality of service, unfair allocation of opportunities, and reinforcement of existing stereotypes.[vii] In financial services, gender-biased AI has resulted in credit discrimination, differential pricing of goods and services, and reduced choice. When applied at scale, the harms caused by AI counter

the financial inclusion goals on which many impact investors and their portfolio companies focus. AI gender bias also has business, regulatory, and reputational repercussions.

Impact investors are well positioned to take a proactive role in selecting and supporting companies to build and deploy equitable AI systems in inclusive finance services. However, evaluating the potential hazards of an AI system requires technical expertise, which many investors do not have.

To fill this gap, the Center for Financial Inclusion (CFI) developed a short guide for impact investors to examine how their investee companies use AI and to help understand the potential for harmful bias, all with a gender lens. There is considerable anecdotal and academic evidence on the gaps and barriers women face in digital financial services (DFS), and this guide exposes many of these examples in the context of gender bias and fairness. However, while this guide is focused on women and gender bias, many of the underlying themes and questions within this document can be broadly applied to other traditionally or historically marginalized groups.

> "Impact investors are well positioned to take a proactive role in selecting and supporting companies to build and deploy equitable AI systems in inclusive finance services."
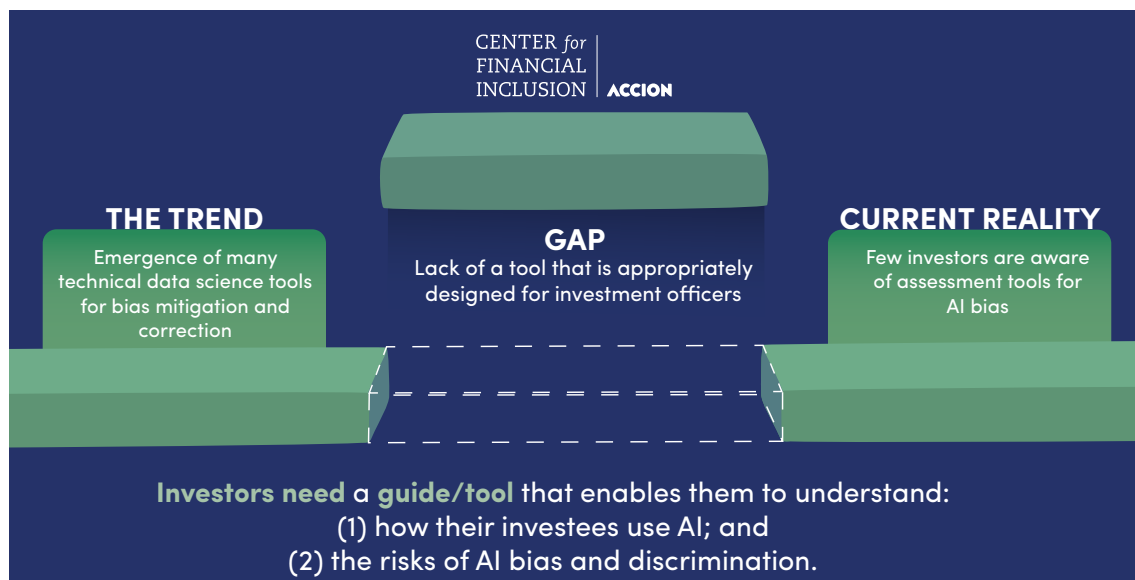
Objectives and Organization of This Guide

This guide offers conversational prompts for investment officers and investees, with the goal of fostering a stronger mutual understanding of risk and emerging risk management practices. When these conversations do not take place, investors and investees risk missing issues and failing to put in place equitable practices.

PROJECT APPROACH

To build this guide, CFI conducted a needs assessment with four impact investors to map their existing due diligence and post-investment processes. The needs assessment suggested that despite concerns around inequitable AI, investors were not equipped with the knowledge or tools to have meaningful discussions with prospective and existing portfolio companies. CFI then reviewed academic literature on AI and harmful bias in financial services and analyzed over 120 existing guides and checklists on ethical AI. Despite a recent proliferation of resources, CFI's review revealed a dearth of practical guidelines for non-technical stakeholders, including impact investors, to assist with risk identification and management.[viii]

CENTER for FINANCIAL INCLUSION | ACCION

THE TREND
Emergence of many technical data science tools for bias mitigation and correction

GAP
Lack of a tool that is appropriately designed for investment officers

CURRENT REALITY
Few investors are aware of assessment tools for AI bias

Investors need a **guide/tool** that enables them to understand:
(1) how their investees use AI; and
(2) the risks of AI bias and discrimination.

STRUCTURE OF THE GUIDE

This guide is intended for use by impact investors and highlights the value of prioritizing equitable AI, while also providing practical suggestions by which to assess how AI is used by fintech investees and identify any areas of bias.

Section 1 shares use cases of AI in inclusive finance, the drivers of harmful AI bias towards women, and features a snapshot of the state of practice in bias identification and mitigation.

Section 2 provides investment officers with an actionable set of questions to help them understand the use of AI among their investees and identify potential risks of harmful gender-based bias and discrimination. Investment officers can use these questions during due diligence or post-investment, depending on the stage of the fintech development and/or the investors' internal processes.

Why Prioritize Equitable AI?

CONTEXT

Why should impact investors prioritize AI and assess whether it is equitable? Because doing so aligns with their business interests and serves the best interests of the end consumers. Despite the central role that AI can play in digital finance, assessing artificial intelligence usually does not make up a substantial part of impact investors' due diligence or post-investment engagement. The time allocated to due diligence processes tends to extend proportionally with the ticket size of the investment — with the progression from prescreen to due diligence to closing often occurring in less than six months. Given the technical barriers in assessing AI systems and the current lack of standards, it is not surprising that assessing AI is not prioritized.

Additionally, investors may do their due diligence processes before a fintech is fully developed or launched. In these instances, impact investors scrutinize the founder's vision and promise to hire capable, responsible data scientists, rather than an actual product with a track record of results (e.g., lending outcomes).

RISKS TO WOMEN CONSUMERS

Unfortunately, biased algorithms often impact already marginalized groups and perpetuate historical inequities.[ix] For example, in 2018, Amazon recalled its AI-based hiring tool after it proved to be biased against women applicants.[x] The flawed AI system had been trained to vet applicants by observing patterns in resumes submitted to Amazon over the preceding 10 years, and with most applicants being men, the system became biased toward preferring men candidates.[xi]

Women are already underserved and underbanked; 7 percent fewer have bank accounts than men, they are 23 percent less likely than men to borrow from a financial institution, and

Algorithms are any series of mathematical rules that define a sequence of operations. They can be used in a computer program, in analog human-driven decisions, or a hybrid form.

Artificial intelligence (AI) systems automate decisions and behave in ways that mimic and go beyond human capabilities.

Machine learning (ML) is a subcategory of AI that uses algorithms to automatically learn insights and recognize patterns from data, applying that learning to make predictions and inform decisions.

Source: Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment

they are more than 27 percent less likely to access fintech products and services.[xii] Algorithmic decisions are increasingly influencing women's financial access and economic opportunities. These high-stakes decisions include which products or services companies market to her and at what price, whether a company approves her for a loan, whether a company approves her

insurance claim, and whether others perceive her as a fraudster. When AI systems' decisions perpetuate harmful gender bias, they can set back a woman's financial trajectory and, when applied more broadly, women's economic empowerment.

BUSINESS IMPLICATIONS OF EQUITABLE AI

Companies increasingly recognize biased AI systems as a material business risk. For example, Microsoft outlined this possibility in its internal risk analysis and filings to the U.S. Securities and Exchange Commission (SEC).[xiii] In a survey of 350 U.S.- and U.K.-based technologists, more than one-third reported that their business had suffered due to a biased algorithm. The survey also found that 62 percent reported lost revenue, 61 percent lost customers, and 43 percent lost employees.[xiv]

For a fintech, excluding or mistreating potential customers, including women, could mean the loss of future revenue for the business as well as legal, reputational, and correctional costs. For investors in these companies, the impacts could trickle upward in lower investment returns, unsustainable business growth, and reputational costs.[xv] In an Economist Intelligence Unit survey of senior non-tech executives, 94 percent of respondents believed responsible AI would produce a greater long-term return on investment (ROI) for investors or shareholders.[xvi]

OPERATIONAL AND REPUTATIONAL RISKS

According to an Accenture global survey of 1,500 C-suite executives, firms that implemented responsible AI practices were 1.7 times more likely to scale up their AI successfully.[xvii] Correcting or canceling a biased AI system can entail high costs from lost employee time and resources associated with developing and correcting the system.[xviii] Employee retention and talent recruitment may also suffer; Google faced protests and walkouts from employees showing opposition to its AI-driven drone, and Meta struggled with recruitment after the Cambridge Analytica scandal, with job acceptance rates decreasing from 85 to approximately 45 percent.[xix]

COMPLIANCE

Although dozens of countries and multilaterals like the Organization for Economic Cooperation and Development (OECD) have published high-level principles and strategies on ethical artificial intelligence, the actual regulation of AI in financial services is still nascent and untested. While the European Union is marching ahead with a draft Artificial Intelligence Act (AIA), most other jurisdictions, particularly those in emerging markets, have not advanced with a comprehensive legislative model.[xx] At the same time, fintechs and their investors must be aware of how existing policy and regulatory provisions within their countries of operation address data privacy, equality, and anti-discrimination — and how financial regulators apply those approaches to algorithmic decision-making. While financial regulation related explicitly to AI is still new, fintech companies that do not address bias in AI may be liable and incur penalty fees.[xxi] In 2022, the Consumer Financial Protection Bureau fined a fintech called Hello Digit more than $2.7 million for a faulty algorithm that depleted consumers' savings. [xxii]

TABLE 1: How Harmful Gender Bias Trickles Up: Example of Credit Underwriting	
WHAT DOES INEQUITABLE AI LOOK LIKE?	Sitha, a woman from an ethnic minority in a peri-urban market in Southeast Asia, applies for a digital loan at a fintech company to help her business grow. The company rejects her application. The AI model that calculated Sitha's creditworthiness leveraged data on previous borrowers — including income, education, gender, age, and credit access — reflecting historical exclusion for women and Sitha's ethnic group. The model does not accurately predict Sitha's ability or willingness to repay. Sitha is not alone. The biased AI model renders women applicants significantly more likely to be a "false negative," meaning that they are rejected despite having a higher propensity to repay than men.[xxiii]
IMPACT ON THE CONSUMER	Sitha does not gain access to the digital loan that she needs to grow her business. Instead, she turns to the local moneylender whose exorbitant interest rates make it nearly impossible to repay, eliminate debt, and gain enough profit margin to grow her business.
IMPACT ON THE FINTECH	By rejecting women "false negatives," who likely would have had no repayment issues, the digital lender loses "good" customers and their ability to increase market share. Should the company identify the problem, it would likely incur expenses to correct the AI system. Additionally, given that the digital lender provides better credit scoring for male customers with the same socio-economic characteristics but different gender or ethnicity, the company could become legally liable for discrimination.
IMPACT ON THE INVESTOR	Investors might not receive the same return on investment if the digital lender's AI system was not generating "false negatives." Additionally, it creates the potential for reputational damage through association with a digital lender that may be held liable for discrimination.

AI USE CASES IN INCLUSIVE FINANCE

While credit underwriting is the most well-known case of artificial intelligence in inclusive finance, fintechs leverage AI across many business functions. Investors must be able to ask and understand how a fintech integrates AI into its business model, particularly if the fintech is making or supporting high-stakes decisions that positively, or negatively, impact consumers' financial lives. This section adheres to the classifications developed by the Cambridge Centre for Alternative Finance and the World Economic Forum in their global study on the adoption of AI in fintech:[xxiv]

❶ REVENUE GENERATION

Fintechs use AI primarily for data analytics and to inform decision-making when designing new products or processes.[xxv] Two relevant applications are:

↗ Credit Underwriting: Companies use machine learning to accelerate lending decisions and reduce credit default risk.[xxvi] Fintechs can apply these models using more traditional, structured data — including payment, transaction, or credit bureau data. They can also use unstructured and semi-structured data sources — including social media activity, satellite data, mobile phone, and text message activity — which can provide a more complex view of creditworthiness, especially for those without formal credit records. [xxvii]

↗ Insurtech Underwriting: Like credit underwriting, insurtech companies use AI and machine learning to automate processes and to analyze non-traditional data and information. Existing insurtech solutions in Sub-Saharan Africa and

South Asia use AI and alternative data — including telematics, satellite data, and wearables — for risk classification and pricing.[xxviii]

❷ CLIENT ACQUISITION

Natural language processing engines provide increasingly realistic and useful interactions. They learn from previous interactions and adapt to different types of consumers and behaviors. Top AI use cases in this domain include AI-enabled customer communication channels and real-time service adjustments to clients' needs.[xxix] For example, Allstate developed an internal chatbot to provide accurate quotes and advice to its commercial clients.[xxx]

❸ RISK MANAGEMENT

Companies regularly deploy AI to classify and detect transactions as potentially fraudulent or anomalous.[xxxi] Fintechs also commonly use AI for preventive pattern analysis of new datasets and AI-enabled risk management.[xxxii] For example, FICO's Falcon platform learns from merchant-owned data to identify behavioral profiles of customers and detect fraud.[xxxiii] Additional use cases include:

↗ Know-Your-Customer (KYC): Companies can also use AI and ML to transform or upgrade identity checks within a financial institution; while they historically relied on usernames and passwords, they

can now use voice recognition, facial recognition, or other similar biometric data, which can reduce the risk of compromised or shared usernames and passwords.[xxxiv] For example, Alipay's Smile Pay uses facial recognition to authenticate and consent to a transaction, allowing retail customers a frictionless checkout process.[xxxv]

↗ Cybersecurity: Companies can use AI and ML to automate cyber threat detection, identify compromised data and information, steer investigations after incidents, and extract information to share with other financial institutions and relevant authorities.

4 CUSTOMER SERVICE

Chatbots and virtual assistants offer more options for customers to get help when they need it and provide more customized profiling and advice. Generative AI has recently garnered intense interest for its potential to revolutionize customer-facing interactions and chatbots. Top use cases in this domain have included AI-enabled add-on services, digital account opening solutions, and AI-enabled marketing.[xxxvi]

5 AUTOMATION AND PROCESS RE-ENGINEERING

Often, fintechs deploy AI to consolidate and automate administrative tasks, like reporting and compliance.[xxxvii] For example, in the insurance sector, companies use AI automation to assess and process claims.[xxxviii]

Rethinking the Term "Bias":

The term "*algorithmic bias*" almost always carries a negative connotation. However, at a technical level, it is a neutral term and conveys that an algorithm's output deviates in a systematic way relative to a norm or standard. These norms or standards could be statistical, institutional, moral, or even legal — and change from fintech to fintech. And while the term "*algorithmic bias*" is typically applied to a decision that generates unfair outcomes, not all biases are necessarily bad, and some may even be desirable. For example, a fintech may create a model that deliberately builds in statistical bias toward women, such that the algorithm's output deviates from the predictions of the training data.

Sources: Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment ; Fazelpour and Danks, 2021; Antony, 2016; Johnson, 2021

Not all uses of AI in fintech impact consumers in the same way. There are use cases where AI drives operations such as a back-end transactional database or a spam filter, applications which do not have a direct consequence for a consumer. But there are instances where AI makes or assists decisions, such as credit and insurance underwriting, that directly impact consumers and their economic opportunities. This guide casts these latter use cases as inherently riskier from a financial inclusion perspective and gives them more scrutiny.

This distinction in use cases pulls in part from the AI risk classifications that the European Union proposed in their draft Artificial Intelligence Act (AIA). The AIA has much higher obligations for high-risk AI systems, including credit-scoring, than it does for minimal or limited risk AI systems, like a chatbot.[xxxix]

SOURCES OF HARMFUL BIAS

Harmful AI bias is not solely a function of the underlying input data or code and can emerge from various sources. Biases also depend on the domains of application, goals for use, and other contextual factors. For example, an insurance claims processing algorithm may be fair at one insurtech but could be potentially

harmful if used by a different company or ported from one country to another within the same company.

Machine-based algorithmic processes typically involve several steps, including problem specification, data collection, data preprocessing, modeling, validation, and deployment. Investors must understand that bias can creep in at any stage and because these processes are highly iterative, it can be quite dependent on the staff involved.[xl] To help categorize potential sources of bias in inclusive finance AI, CFI has used a framework of inputs, code, and context.[xli] While the framework helps explain algorithm development and facilitates the categorization of risks, in practice, the categories overlap and an issue might be relevant for both the input and code categories, for instance. While not exhaustive, the following examples are a useful place to start.

INPUTS

↗ Historical Bias: Historical bias can enter algorithmic systems due to preexisting and longstanding cultural, social, or institutional dynamics. For example, if an AI system that evaluates creditworthiness uses historical income data to predict whether an individual can repay a loan, the systemic gender gap in salaries could lead to unfair outcomes for women applicants.

↗ Sampling Bias: Sampling bias can occur when gathering data or sampling a population to support the development of a model. Sampling methods may be limited in that they only reach a narrow population, which can lead to an algorithm that produces a result that disadvantages under-sampled groups. For example, the well-documented gender gap in access to finance and digital financial services may mean that women are under-sampled in data sets.[xlii] Additionally, given the gender digital divide, women may be more hesitant to share sensitive data.

↗ Selection Bias: Selection bias happens when outcomes for specific individuals, such as women or other marginalized groups, are unavailable. Credit models, for instance, are trained on individuals whose applications for loans were accepted rather than rejected. Given that the financial system has historically excluded women, the data used to train algorithms might reflect the characteristics of a subset of the loan applicants: men whose loan applications were historically accepted.

CODE

↗ Biases in Problem Specification: Bias can appear when the data science team chooses how to optimize the algorithm to align with a specific strategy. Outcomes largely depend on what goals are set — such as maximizing profits, ensuring client privacy, or minimizing disparate treatment or outcomes for women — each of which can have different results. If a fintech cares about loan applicants who are most likely to repay, but the prediction algorithm is optimized to identify those who will return the highest profits to the institution, then the algorithm outputs will not provide the appropriate information.

↗ Biases in Preprocessing: Algorithms can't consume incomplete and inconsistent data, and the "noise" disrupts the true pattern of the sample. Data preprocessing aims to solve these challenges. However, during the preprocessing stage, an algorithm must make choices that could potentially introduce bias. These choices include how to handle missing data for underrepresented groups, including women, or how to identify and address outliers. Other issues could occur due to manual errors, unexpected events, and technical issues.

↗ Bias by Proxy: Some well-intentioned fintechs defend their AI systems by saying that gender bias is impossible because they intentionally exclude data on sensitive attributes like gender from their models. Unfortunately, this argument does not stand up in practice. Inclusive finance algorithms are crunching ever-

larger amounts of data as well as data coming from new sources. As a result, AI models can inadvertently use proxies for sensitive attribute data in decisions, even if it was unintentional. For instance, phone or app types can be proxies for gender, age, or other characteristics, which can sometimes be misleading.

CONTEXT

↗ **Transfer Context Bias:** Fintechs deploy algorithms for specific uses or purposes and in a particular context. If employed outside of those contexts, they may not perform according to appropriate standards. For example, risk assessment algorithms trained on historical data in the United States may not translate well to an emerging or developing economy setting.

↗ **Interpretation Bias:** Even if employed in the proper context, algorithmic bias can arise from a staff member's misinterpretation of the algorithm's outputs or functioning. These failures sometimes can happen without staff being aware; the algorithms might be described to staff simply as "predicting loan applicant success" without defining what constitutes success.

↗ **Representation Bias:** When the personnel who design, operate, and govern models lack diversity, it can heighten the risk of bias and discrimination issues in machine learning models. A lack of representation weakens a fintech's ability to recognize and respond to problems while developing and using models. This issue becomes particularly problematic considering the documented gender gap in representation within leadership positions and STEM fields.

THE STATE OF PRACTICE IN MITIGATING, DETECTING, AND CORRECTING HARMFUL AI BIAS

To better understand the current landscape of practice with mitigating, detecting, and correcting harmful biases, CFI sorted approaches into three

categories: 1) approaches for general awareness-raising on ethical AI and developing strong codes of conduct; 2) tools oriented to business processes and sound data documentation practices; and 3) techniques centered on specific data science practices to detect bias and implement corrective measures.

Through this analysis, CFI identified a gap in available resources for concerned, non-technical stakeholders, such as investment officers, to help discern whether a provider has taken appropriate steps to mitigate, detect, and correct harmful bias. This guide aims to help close this gap.

ETHICS PRINCIPLES AND CODES OF CONDUCT

These principles can foster a broader understanding of ethical algorithms, including fairness, transparency, explainability, and auditability principles within a company. While they provide broader guidance at an organizational level, they are not the detailed, technical approaches required for bias detection and mitigation.

In general, a fintech's code of conduct should advocate that its algorithms abide by the following principles:

i. Be based on motives of non-maleficence (i.e., they should do no harm);

ii. Be oriented toward providing benefits for humanity;

iii. Assure privacy and data security for individuals whose data is being collected and used;

iv. Build transparency and explainability into their processes; and

v. Ensure that products perform their intended function consistently and correctly. [xliii]

However, while these principles are useful for creating awareness and alignment on broader goals, they lack enforcement mechanisms and concrete tools that can be applied toward bias mitigation.

In 2019, Google Cloud began to receive requests from enterprise customers asking for solutions for AI-based lending using non-traditional data. In 2020, Google Cloud conducted a sprint using Google's AI Principles and ultimately decided against building an AI lending tool. The team believed that a "product built — with today's technologies and data — could create disparate impact related to gender, race and other marginalized groups, and conflict with Google AI's principle to 'avoid creating or reinforcing unfair bias.'"

Sources: McElhaney et al, 2022; Google AI Progress Report

BUSINESS PROCESS AND DATA DOCUMENTATION TOOLS

AUDITING FRAMEWORKS AND BIAS TOOLKITS

Frameworks and toolkits in this category can include governance, impact assessment, auditing tools, and standards that aim to ensure AI systems are responsible. "Auditing frameworks" go beyond codes of conduct and principles by adding relevant tools for improved monitoring and enforcement.[xliv] Most frameworks provide a combination of worksheets, questionnaires, and checklists that fintechs can use to promote alignment with the stated guiding principles or codes of conduct. Some frameworks involve comprehensive end-to-end auditing and governance processes that companies can complete with internal or external support.

DATA DOCUMENTATION FOR ALGORITHMIC DEVELOPMENT: DATASHEETS AND MODEL OR METHOD CARDS

To build transparency and allow for stronger engagement between model developers and other stakeholders, a fintech can use robust data and model documentation practices to develop an algorithmic system. This transparency can help a fintech to identify risks up front, allow for clearer investigations should issues crop up after deployment, and facilitate future data science team members to make informed adaptations.[xlv]

Popular examples of data documentation include datasheets for datasets and model or method cards. Datasheets document the motivation for developing new datasets, data products, collection objectives, collection processes, data formatting and labeling procedures, uses, distribution, and maintenance of the datasets. Model and method cards document key attributes of an algorithmic model and their performance characteristics. An additional tool is the Dataset Nutrition Label.[xlvi]

DATA SCIENCE BIAS DETECTION AND CORRECTIVE MEASURES

These approaches are grounded in responsible AI frameworks and involve specific bias mitigation and detection tactics. They typically follow a two-stage strategy for addressing algorithmic bias: 1) define and use one or more mathematical fairness measures to quantify the amount of bias in the algorithmic output; and 2) develop mitigation responses that reduce any problematic bias. For the latter step, fintechs can integrate a growing range of technical approaches into the algorithmic design at preprocessing, in-processing, and/or post-processing stages.[xlvii] Women's World Banking, for example, created a Python-based tool to audit fairness among accepted and rejected credit applications.[xlviii] FinRegLab published research that analyzed machine learning underwriting models with advanced "explainability" tools (that themselves used machine learning) from seven technology companies and found some techniques reliably identified underwriting model features that were key for consumer disclosure and fair lending analysis.[xlix]

A Note on Defining Fairness:

Many of the recent advances in testing and auditing for bias are framed around the concept of fairness. While the proliferation of fairness measures has provided a useful starting point for the development of bias mitigation techniques, there is no consensus regarding the optimal choice of fairness metric(s) to apply, particularly when it comes to financial services provision. The meaning of fairness varies widely across cultures. This makes defining fairness more challenging, and necessitates the consideration of the cultural context in which AI system will be deployed.

Defining an approach to fairness means making choices and trade-offs. These decisions should be well documented for internal communication and to allow for healthy discussion and debate. Trade-offs may lead algorithms to achieve certain fairness metrics — parity between men and women borrowers, for instance — but may come at the cost of less accurate models.

Fintechs that that are building algorithms and AI systems must ensure that they give adequate attention to defining what their AI system is trying to achieve and what harmful biases it is trying to avoid (e.g., ethical, contextual, institutional, or legal objectives). The team should then carefully select fairness metrics that are aligned with these goals. While management and C-suite executives might not delve into the technical details of how fairness metrics are operationalized, there should be a documented understanding between management and data science teams regarding the definition, measurement, and accountability for performance against the relevant fairness metrics.

Impact investors should also understand how fintechs define, operationalize, and monitor for fairness, and decide whether those indicators align with their own investment goals.

Sources: Demirguc-Kunt et al., 2022; Chen et al., 2021; Mulligan et al., 2019; Berkeley Haas; Mitchell et al., 2021; Mehrabi et al., 2021; Green and Hu, 2018; Srivastava et al., 2019; Green, 2021

Section 2: Conversational Prompts to Promote Equitable AI

This section provides investment officers with a practical set of questions to understand the use of AI among their investees and identify potential risks of harmful gender-based bias. These questions are also relevant for other marginalized groups. The conversational guide is designed to be used during the due diligence process and can help to inform investment decisions or covenants or post-investment decisions as part of board governance and portfolio engagement. It is important to note that some early-stage fintechs may not have built out their models by the time due diligence occurs, and these questions may be more appropriate in a post-investment stage.

CFI curated the prompts and questions included below from a review of more than 120 existing ethical AI guides, checklists, and tools. The selection of questions to ultimately include was informed by the investor needs assessment and the context of inclusive finance, and prompts were tailored for investment officers in the inclusive fintech space who are at least one step removed from the data science team.

NOTES FOR USE

The prompts highlight key areas where investors and investees who are aiming to promote the equitable use of AI should discuss and arrive at mutual expectations. CFI does not recommend "scoring" a fintech's practices based on the included questions, as it would give a false sense of precision. Given the dynamic nature of the data science that fuels AI, the lack of consensus around targeted bias mitigation techniques, and uncertainty around how responsible AI regulation emerges, the prompts are more directional than exhaustive or immutable.

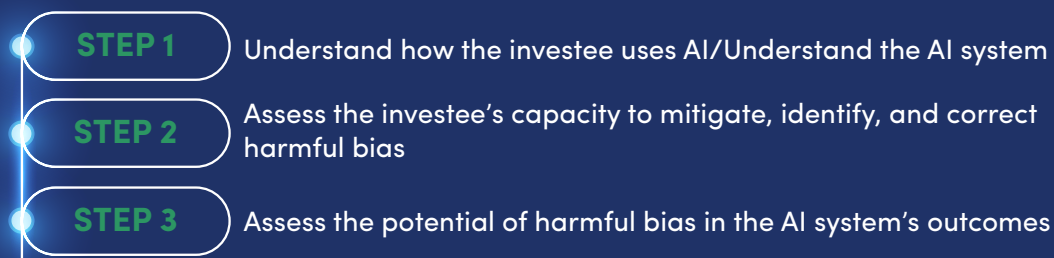Some prompts pertain to areas where a self-reported answer from the investee is sufficient. In contrast, other prompts might reveal potential areas for improvement or indicate the need for more in-depth discussions to establish expectations. This guide elevates the business processes, such as data management and governance, that support algorithmic development and sees them to be key priorities along with the code and outcomes. Because many models are dynamic, looking at outcomes alone provides only a snapshot of a constantly evolving system.

The questions and accompanying strategies are structured around three steps. Each step has a brief explanation at the beginning, followed by questions, and concludes with an "Explore Further" section. The "Explore Further" sections provide references used in developing each section and offer opportunities to dive deeper and review pertinent examples.

Even though the future of AI within inclusive finance is constantly evolving, it is important for investors to avoid a passive "wait-and-see" approach. While the industry lacks consensus on effective targeted bias mitigation techniques and there is uncertainty surrounding the development of responsible AI regulation, the malleability of this moment represents an opportunity for impact investors to effect meaningful change. By actively incorporating the principles of equitable AI into the development of fintech products, investors can ensure a more inclusive future.

SNAPSHOT VIEW

STEP 1 Understand how the investee uses AI/Understand the AI system

STEP 2 Assess the investee's capacity to mitigate, identify, and correct harmful bias

STEP 3 Assess the potential of harmful bias in the AI system's outcomes

SUMMARY VIEW

STEP 1
Understand how the investee uses AI/Understand the AI system

Does the investee use AI in consumer–facing decisions (e.g., KYC, credit underwriting, insurance underwriting, automation, customer service, or client acquisition)?

Is the investee's AI strategy and vision and business strategy aligned with the investor's business strategy and mission?

Has the investee identified and documented the individuals and groups at risk of being systematically disadvantaged by the algorithm?

STEP 2
Assess the investee's capacity to mitigate, identify, and correct harmful bias

Input: Data, model, and testing	Has the investee defined what a "fair" algorithmic system means, associated measures, and thresholds?
	Has the investee implemented measures to reduce unfairness in the data, modeling, and testing?
Context: Governance	Is the investee committed to a proactive and systematic AI fairness framework or a responsible AI framework?
Context: Diversity and training	Does the data science team have gender diversity?
	Has the data science team been trained on how bias can enter AI tools and ways to mitigate it?
Context: Mechanisms for complaint resolution	Does the company have a response plan, redressal, or recourse mechanism if the results harm women?

STEP 3

Assess the potential of harmful bias in the AI system's outcomes

Customer acquisition and outreach	Is there a gender balance in the demand/incoming leads for the fintech product?
	Are men and women equally prioritized in customer outreach or marketing campaigns?
KYC and automatization	Do men and women applicants have the same probability of successfully passing KYC steps?
	Do men and women applicants have the same success meeting other documentation requirements and automatic filters?
For credit and insurance products	Do men and women applicants have the same credit/risk score after accounting for other relevant factors?
	Do men and women applicants have the same likelihood of receiving a credit/insurance offer after accounting for other relevant factors?
For cybersecurity/ fraud detection	Do transactions from men and women have a similar likelihood of being correctly flagged as fraudulent activity?

FULL VERSION

STEP 1

Understand how the investee uses AI/Understand the AI system

Guidance to complete this step

When evaluating a new investment or sitting on the board, the investor should be familiar with the use of AI by the investee, particularly if it makes high-stakes decisions that impact consumers. An investor should also become familiar with whether the fintech's AI operations fall under anti-discrimination laws, data privacy regulation, financial regulation on algorithms, and broader regulations that govern ethical AI.

Questions:	YES	NO	DK	Sources of information to check
Does the investee use AI in decisions that can impact consumers' economic opportunities (e.g., credit underwriting, insurance underwriting, client acquisition)?				↗ Strategy ↗ Business plan ↗ Annual report ↗ Client application ↗ Interviews with investees
Does the investee have a documented AI vision and strategy?				
Is the investee's AI vision and strategy aligned with the investor's business strategy and mission?				
Has the investee identified and documented the individuals and groups at risk of being systematically disadvantaged by the algorithm?				

Explore Further

- ↗ Transforming Paradigms: A Global AI in Financial Services Survey
- ↗ Responsible Investing in AI: A Responsible AI Due Diligence for VCs
- ↗ AI for good: Research insights from financial services
- ↗ The Stories Algorithms Tell: Bias and Financial Inclusion at the Data Margins
- ↗ Algorithmic Bias, Financial Inclusion, and Gender
- ↗ Artificial Intelligence: Practical Superpowers: The Case for AI in Financial Services in Africa
- ↗ Machine Learning Explainability and Fairness: Insights from Consumer Lending
- ↗ Reflecting the Past, Shaping the Future: Making AI Work for International Development
    - ↗ Addendum of Use Cases at USAID

STEP 2

Assess the investee's capacity to mitigate, identify, and correct harmful bias

Guidance to complete this step

When an investee uses an AI system in high-stakes consumer-facing decisions, investors should assess the investee's capacity to mitigate, identify, and correct these problems. Since unfair outcomes can come from multiple operational areas, this section probes for organizational practices, including governance, training, diversity, modeling, testing, and redressal and complaint mechanisms. This step provides a general diagnosis of the areas and practices that need support. If a company purchased an algorithm "off-the-shelf," many of the questions in Input and Code would be directed at that vendor.

Questions:	YES	NO	DK	Sources of information to check
Input: Data				
Has the investee defined what a "fair" algorithmic system means, along with associated measures and thresholds?				
Is the definition of fairness grounded in the social context of the operating market(s)?				
Has the investee defined a process for identifying at-risk groups?				
Does the investee check if the data they are using to train models is representative of the population of men and women they are trying to serve?				↗ Strategy ↗ Business plan ↗ Client application
Has the data science team checked if data availability or quality differs based on gender?				↗ Interviews with investees and/or third-party vendor
If the data is not equally representative, or if the availability and quality differ based on gender, have they implemented corrective measures (e.g., oversampling and synthetic augmentation) to address gaps in training data?				↗ Board reports ↗ Client data ↗ Interviews with AI team
Does the investee's data science team have a transparent process that documents the rationale for the algorithm's development, data collection objectives, data collection processes, data labeling procedures, distribution, and management of the datasets (e.g., datasheets for datasets, Data Nutrition Labels, or comparable)?				↗ Data used for modeling ↗ Algorithm code ↗ Data reports or statistics
Code: Modeling and testing				
Has the investee's data science team documented AI system design choices and key attributes of the algorithmic model, the context in which the team intended to use the model, and performance metrics (e.g., model or method cards)?				
Has the investee's data science team documented the human judgment in the design of the algorithm?				
Is the investee's definition of fairness measured and prioritized when modeling and testing?				

Questions:	YES	NO	DK	Sources of information to check
Has the investee's data science team assessed and documented quantitative estimates of the system's performance against its fairness objectives (e.g., an equal number of women and men with loan applications accepted, controlling for other relevant characteristics)?				
Has the investee assessed any trade-offs between the system's fairness objectives and its commercial objectives? Is there a process for flagging issues related to harmful bias, discrimination, or poor performance of the AI system?				
Is there a process to test for model drift to ensure accountability over time?				
Context: Governance				
Does management understand, at a high level, how the team optimized the algorithm, what the fairness goals are, and any risk assessments for disproportionate impact against women?				↗ Strategy ↗ Business plan ↗ Client application
Is the investee committed to a proactive and systematic AI fairness framework or a responsible AI framework?				↗ Interviews with investees and/or third-party vendor
Does the investee have an active responsible AI ethics council (or comparable body) aligned with a governance structure?				↗ Board reports ↗ Client data
Does the investee have the ability to design AI solutions that follow regulations on fairness and responsible AI?				↗ Interviews with AI team
Does the investee have updated and readily available documents for auditors (e.g., risk assessment, data assessment, model assessment, model monitoring logs), if required?				↗ Data used for modeling ↗ Algorithm code ↗ Data reports or statistics
Does the board of directors receive regular management reports on algorithmic performance against fairness metrics?				
Can management or the board of directors promptly shut down the AI system if necessary?				
Context: Diversity and training				
Does the data science team have gender diversity?				
Has the company trained the data science team on how bias can enter AI tools and ways to mitigate it?				
Has the investee designed a road map prioritizing organization-wide AI bias awareness and training programs?				
Context: Mechanisms for complaint resolution				
Are the reasonings for decisions made by the algorithm transparent to the consumers they are targeting?				
Does the company have a response plan, redressal, or recourse mechanism if AI decisions harm women?				

STEP 3

Assess the potential of harmful bias in the AI system's outcomes

Guidance to complete this step

The following questions can help flag potential risks of harmful bias by looking at the outcomes and impacts of the AI system, but further audits may be needed to assess and diagnose these problems accurately. "No" or "Don't know" answers flag that the investee's AI system might be producing unfair outcomes against women, and further conversations and investigations would be required with the data science team. Remember that outcome-centered questions provide a snapshot in time and should be revisited on a regular basis.

N.B.: Step 3 may be trickier in a market where regulation forbids the collection of sensitive characteristics. This was done with the intent of protecting individuals' privacy as well as prohibiting lenders from making decisions based on these sensitive characteristics, but it makes testing AI outcomes across demographics more difficult.

Questions:	YES	NO	DK	Sources of information to check
Customer acquisition and outreach				
Has the investee defined what a "fair" algorithmic system means, along with associated measures and thresholds?				
Are men and women equally prioritized in customer outreach or marketing campaigns?				
KYC and automatization				↗ Strategy
Do men and women applicants have the same probability of successfully passing KYC steps?				↗ Business plan
Do men and women applicants have the same success meeting other documentation requirements and automatic filters?				↗ Business plan ↗ Client applications ↗ Interviews with investees
For credit and insurance products				↗ Administrative data and statistics
Do men and women applicants have the same credit/risk score after accounting for other relevant factors?				↗ Interviews with the data science team, loan or insurance officers
Do men and women applicants have the same likelihood of receiving a credit/insurance offer after accounting for other relevant factors?				
Do men and women who are extended credit/insurance offers receive the same credit/insurance terms after accounting for other relevant factors?				
Do men and women with rejected applications have the same average credit/risk score after accounting for other relevant factors?				

Questions:	YES	NO	DK	Sources of information to check
Do men and women with rejected applications systematically differ in their rejection reasons?				
Do men and women applicants have the same likelihood of becoming repeat clients unconditionally and conditionally?				↗ Strategy
For financial advice				↗ Business plan
Do men and women applicants receive the same type of advice/nudges (e.g., a recommendation to use specific product types, allocation of assets) after accounting for other relevant factors?				↗ Business plan ↗ Client applications ↗ Interviews with investees
In cases where an AI advisor makes recommendations, but a client can decide to follow or reject them, do men and women applicants have a similar likelihood of accepting recommendations after accounting for other relevant factors?				↗ Administrative data and statistics ↗ Interviews with the data science team, loan or insurance officers
In cases where an AI advisor makes automated decisions, do men and women applicants exhibit the same performance outcomes after accounting for other relevant factors?				
For cybersecurity/fraud detection				
Do transactions from men and women have a similar likelihood of being correctly flagged as fraudulent activity?				

Explore Further

↗ Check Your Bias: A Field Guide for Lenders
↗ Eliminating AI Bias in Insurance
↗ ORCAA: Risk Consulting and Algorithmic Auditing

Additional Resources

↗ Confronting Bias: BSA's Framework to Build Trust in AI
↗ Nasscom Responsible AI Resource Kit
↗ Of Oaths and Checklists
↗ Risk Mitigation Checklist
↗ Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment
↗ Data Ethics Framework

References

i   Financial Stability Board, "Artificial intelligence and machine learning in financial services: Market developments and financial stability implications," Nov. 2017. https://www.fsb.org/wp-content/uploads/P011117.pdf

ii   OECD.AI, "VC investments in AI by industry," accessed on Dec. 6, 2022. https://oecd.ai/en/data?selectedArea=investments-in-ai-and-data&selectedVisualization=vc-investments-in-ai-by-industry

iii   Sraders. Anne. "How much are generative A.I. startups really worth? VCs are paying up as they scramble to get in on the next big thing." Fortune, Apr. 2023. https://fortune.com/2023/04/06/how-much-are-generative-ai-startups-worth-venture-capital/

iv   Biallas, Margarete and Felicity O'Neill. "Artificial Intelligence Innovation in Financial Services." IFC Note 85, Jun. 2020.

v   Ibid.

vi   Pazarbasioglu, Ceyla, Alfonso Garcia Mora, Mahesh Uttamchandani, Harish Natarajan, Erik Feyen, and Mathew Saal. "Digital Financial Services." World Bank Group, Apr. 2020. https://pubdocs.worldbank.org/en/230281588169110691/Digital-Financial-Services.pdf

vii   Smith, Genevieve and Ishita Rustagi. "When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity." Stanford Social Innovation Review, Mar. 2021. https://doi.org/10.48558/A179-B138

viii   NIST, "Artificial Intelligence Risk Management Framework," Jan. 2023. https://doi.org/10.6028/NIST.AI.100-1

ix   Townson, Sian. "AI Can Make Bank Loans More Fair." Harvard Business Review, Nov. 2020. https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair

x   Smith, Genevieve and Ishita Rustagi. "Mitigating Bias in Artificial Intelligence: An Equity Fluent Leadership Playbook." Berkeley Haas Center for Equity, Gender and Leadership, Jul. 2020. https://haas.berkeley.edu/equity/industry/playbooks/mitigating-bias-in-ai/

xi   Dastin, Jeffrey. "Amazon scraps secret AI recruiting tool that showed bias against women." Reuters, Oct. 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

xii   Financial Alliance for Women, "Gender-Intelligent Fintech Design: How Fintechs Can Capture the Female Economy," 2021. https://financialallianceforwomen.org/download/how-fintechs-can-capture-the-female-economy/; Financial Alliance for Women, "The Gender Data Playbook for Women's Financial Inclusion," 2023. https://financialallianceforwomen.org/download/the-gender-data-playbook-for-womens-financial-inclusion/

xiii   Smith and Rustagi, "Mitigating Bias in Artificial Intelligence."

xiv   DataRobot, "State of AI Bias," 2022. https://www.datarobot.com/resources/state-of-ai-bias-report/thank-you/

xv   Rizzi, Alexandra, Alexandra Kessler, and Jacobo Menajovsky. "The Stories Algorithms Tell: Bias and Financial Inclusion at the Data Margins." Center for Financial Inclusion, Mar. 2021. https://content.centerforfinancialinclusion.org/wp-content/uploads/sites/2/2021/03/The-Stories-Algorithms-Tell-CFI-publication-MAR21.pdf

xvi The Economist Intelligence Unit, "Staying ahead of the curve: The business case for responsible AI," Oct. 2020. https://pages.eiu.com/rs/753-RIQ-438/images/EIUStayingAheadOfTheCurve.pdf

xvii Reilly, Athena, Joe Depa, and Greg Douglass. "AI: Built to Scale." Accenture, 2019. https://www.accenture.com/_acnmedia/Thought-Leadership-Assets/PDF-2/Accenture-Built-to-Scale-PDF-Report.pdf#zoom=50

xviii Smith and Rustagi, "Mitigating Bias in Artificial Intelligence."

xix Ibid.

xx Feldstein, Steven. "Evaluating Europe's push to enact AI regulations: how will this influence global norms?" Democratization, Apr. 2023. https://doi.org/10.1080/13510347.2023.2196068

xxi Dotan, Ravit. "Responsible Investing in AI: A Guidebook for VCs." VentureESG, Dec. 2022. https://www.ravitdotan.com/_files/ugd/f83391_899994a46dfc4cfdb253cabbd3f841c7.pdf; Smith and Rustagi, "Mitigating Bias in Artificial Intelligence."

xxii Consumer Financial Protection Bureau, "CFPB Takes Action Against Hello Digit for Lying to Consumers About Its Automated Savings Algorithm," Aug. 2022. https://www.consumerfinance.gov/about-us/newsroom/cfpb-takes-action-against-hello-digit-for-lying-to-consumers-about-its-automated-savings-algorithm/ Consumer Financial Protection Bureau, "CFPB Takes Action Against Hello Digit for Lying to Consumers About Its Automated Savings Algorithm," Aug. 2022. https://www.consumerfinance.gov/about-us/newsroom/cfpb-takes-action-against-hello-digit-for-lying-to-consumers-about-its-automated-savings-algorithm/

xxiii Women's World Banking, "Check Your Bias! A Field Guide for Lenders," Dec. 2022. https://www.womensworldbanking.org/insights/report-check-your-bias-a-field-guide-for-lenders/

xxiv Cambridge Centre for Alternative Finance and World Economic Forum, "Transforming Paradigms: A Global AI in Financial Services Survey," Jan. 2020. https://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf

xxv Ibid.

xxvi Koide, Melissa. "AI for good: Research insights from financial services." Center on Regulation and Markets at Brookings, Aug. 2022. https://www.brookings.edu/wp-content/uploads/2022/08/AI-for-good-Research-insights-from-financial-services-4.pdf

xxvii Biallas and O'Neill, "Artificial Intelligence Innovation in Financial Services."; OECD, "OECD Business and Finance Outlook 2021: AI in Business and Finance," Sep. 2021. https://doi.org/10.1787/ba682899-en

xxviii Pazarbasioglu et al., "Digital Financial Services."

xxix Cambridge Centre for Alternative Finance and World Economic Forum, "Transforming Paradigms."

xxx McWaters, R. Jesse. "The New Physics of Financial Services: Understanding how artificial intelligence is transforming the financial ecosystem." World Economic Forum, Aug. 2018. https://www3.weforum.org/docs/WEF_New_Physics_of_Financial_Services.pdf

xxxi World Economic Forum, "Empowering AI Leadership: AI C-Suite Toolkit," Jan. 2022. https://www3.weforum.org/docs/WEF_Empowering_AI_Leadership_2022.pdf; Boukherouaa, El Bachir, Khaled AlAjmi, Jose Deodoro, Aquiles Farias, and Rangachary Ravikumar. "Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance." IMF, Oct. 2021. https://www.elibrary.imf.org/view/journals/087/2021/024/article-A001-en.xml

xxxii Cambridge Centre for Alternative Finance and World Economic Forum, "Transforming Paradigms."

xxxiii McWaters, "The New Physics of Financial Services: Understanding how artificial intelligence is transforming the financial ecosystem."

xxxiv Buolamwini, Joy and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," MIT Conference on Fairness, Accountability,

and Transparency, Feb. 2018. https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/

xxxv McWaters, "The New Physics of Financial Services."

xxxvi Cambridge Centre for Alternative Finance and World Economic Forum, "Transforming Paradigms."

xxxvii Ibid.

xxxviii McWaters, "The New Physics of Financial Services."

xxxix European Parliament, "AI Act: a step closer to the first rules on Artificial Intelligence," May 2023. https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence

xl Leslie, David. "Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector." The Alan Turing Institute, Jun. 2019. https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

xli Rizzi et al., "The Stories Algorithms Tell."

xlii Demirgüç-Kunt, Asli et al. "The Global Findex Database: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19." World Bank, 2022. https://www.worldbank.org/en/publication/globalfindex; Chen, S. et al. "The fintech gender gap." Journal of Financial Intermediation, Vol. 54, Apr. 2023. https://www.sciencedirect.com/science/article/pii/S1042957323000098#bib1

xliii Floridi, Luciano et al. "AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." Minds & Machines, Nov. 2018. https://link.springer.com/article/10.1007/s11023-018-9482-5; Floridi, Luciano and Josh Cowls. "A Unified Framework of Five Principles for AI in Society." Harvard Data Science Review, 1(1), Jul. 2019. https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8

xliv ICO, "Guidance on the AI auditing framework: Draft guidance for consultation," Feb. 2020. https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf; Ethical OS, "A guide to anticipating the future impact of today's technology," Aug. 2018. https://ethicalos.org/

xlv Adkins, David et al. "Method Cards for Prescriptive Machine-Learning Transparency." 2022 IEEE/ACM 1st International Conference on AI Engineering, May 2022. https://ieeexplore.ieee.org/document/9796452

xlvi Data Nutrition Project, accessed Jun. 2023. https://datanutrition.org/

xlvii Danks, David and Alex John London. "Algorithmic Bias in Autonomous Systems." 26th International Joint Conference on Artificial Intelligence, Aug. 2017. https://www.cmu.edu/dietrich/philosophy/docs/london/IJCAI17-AlgorithmicBias-Distrib.pdf; Fazelpour, Sina and David Danks. "Algorithmic Bias: Senses, sources, solutions." Philosophy Compass, Vol. 16, Issue 8, Aug. 2021. https://compass.onlinelibrary.wiley.com/doi/full/10.1111/phc3.12760

xlviii Mirpourian, Mehrdad. "A Python Toolkit for Auditing and Detecting Gender Bias in Credit Assessment Processes." Github repository, Jan. 2023. https://github.com/WomensWorldBanking/A-Python-toolkit-for-auditing-and-detecting-gender-bias-in-credit-assessment-processes

xlix FinRegLab, "Explainability and Fairness in Machine Learning for Credit Underwriting: Policy and Empirical Findings Overview, " July 2023. https://finreglab.org/wp-content/uploads/2023/07/FRL_ML-FindingsOverview_Final.pdf

CENTER *for*
FINANCIAL
INCLUSION | **ACCION**