

INVERTIR EN INTELIGENCIA ARTIFICIAL EQUITATIVA PARA LAS FINANZAS INCLUSIVAS



UNA GUÍA DE GESTIÓN DE RIESGOS PARA INVERSORES DE IMPACTO

Agosto 2023

Autora principal: Lucciana Alvarez, Gerente de Investigación, Centro para la
Inclusión Financiera (CFI)

Coautora: Alexandra Rizzi, Directora Sénior de Investigación,
Centro para la Inclusión Financiera (CFI)

CENTER for
FINANCIAL
INCLUSION

ACCION



AGRADECIMIENTOS

Las autoras desean agradecer a Dr. Jonathan Fu y Dr. Mrinal Mishra por realizar la revisión de literatura académica que fue fundamental para este trabajo y por comentar los borradores de la guía. Además, agradecen a Stella Dawson, Mayada El-Zoghbi, Aerial Emig, Elizabeth Miller, Michael Schlein, Edoardo Totolo y al Consejo Asesor del Centro para la Inclusión Financiera (CFI) por su orientación editorial. Asimismo, agradecemos a los miembros del equipo de Equitable AI (Inteligencia Artificial Equitativa) en USAID y DAI, quienes brindaron excelente orientación y sugerencias durante todo el proyecto, con mención especial a Shachee Doshi, Stefanie Falconi, Paul Nelson, Meredith (Beth) Perry, Priya Sethi y Genevieve Smith. Finalmente, agradecemos a los equipos de inversión de Quona, DFC, FMO, Accion Venture Lab y su investee Kuunda, quienes generosamente brindaron tiempo para la evaluación de las necesidades de los inversionistas y brindaron comentarios útiles a los borradores de la guía.

Este trabajo es el resultado del Equitable AI Challenge de USAID y DAI, una competencia organizada en el 2022 para identificar soluciones innovadoras para abordar los resultados desiguales de género como producto del uso de la inteligencia artificial. El Equitable AI Challenge solicitó propuestas que consideraran enfoques holísticos y creativos para identificar y abordar los sesgos de género en los sistemas de inteligencia artificial dentro de contextos de desarrollo global. La propuesta ganadora de CFI fue desarrollar un módulo de debida diligencia que permita identificar los riesgos de inequidad de género generados por los algoritmos aplicados a las finanzas inclusivas.

Esta guía fue desarrollada en colaboración con:



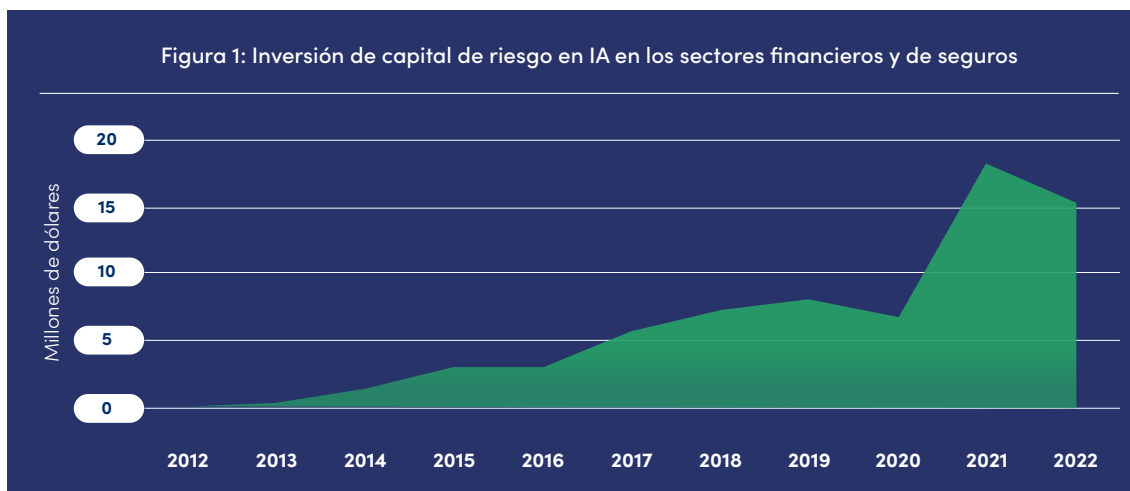
Introducción	3
Objetivos y organización de esta guía	5
¿Por qué priorizar la IA equitativa?	6
Sección 1: Casos de uso de la IA, fuentes de sesgo y estado de la mitigación, identificación y corrección de sesgos	9
Sección 2: Guía de conversaciones para promover una IA equitativa	15
Referencias	22



Introducción

La inteligencia artificial (IA) permite innovaciones en las finanzas digitales al aumentar la eficiencia, reducir los costos y atender a consumidores en gran escala. Ya sea automatizando los procesos de back-end o las decisiones de cara al cliente, la IA ha generado avances en pagos digitales, crédito digital, seguros y en diversas funciones de soporte como chatbots, marketing personalizado y el asesoramiento automatizado.ⁱ Los inversionistas han notado este potencial catalizador de la IA. En efecto, el capital de riesgo (venture capital) invertido en empresas emergentes que utilizan IA en el sector financiero y de seguros aumentó de USD \$227 millones en 2012 a \$15.4 mil millones en 2022, convirtiéndose en el tercer sector con mayor monto de inversiones, después de los sectores de tecnología de la información y atención médica.ⁱⁱ En el año 2023, a pesar del entorno difícil para el capital de riesgo, las inversiones en IA generativa desafiaron la tendencia y se han disparado.ⁱⁱⁱ

Para los inversores centrados en la inclusión financiera, las innovaciones impulsadas por la IA, junto con el crecimiento de los datos digitales de teléfonos móviles, satélites y otras fuentes, tienen el potencial de crear modelos de negocio viables para segmentos de mercado históricamente desatendidos. La IA proporciona soluciones escalables para determinar la solvencia de personas y empresas que históricamente carecían de identificación, garantía e historial crediticio.^{iv} Estas tecnologías automatizan procesos de negocio permitiendo mayores volúmenes de transacciones de bajo valor que permitan incluir a los segmentos más difíciles de alcanzar, como las mujeres microempresarias, de forma más rentable y sostenible en el largo plazo.^v



Si bien existen argumentos a favor del uso de la IA, su implementación también conlleva una serie de riesgos. Uno de los más importantes es la inequidad en los resultados para ciertos grupos de consumidores por motivos vinculados al género, raza o etnia.^{vi} Para las poblaciones vulnerables, como las mujeres y personas afrodescendientes, el uso de IA puede generar exclusión en el acceso a servicios, discriminación en el precio o en la calidad del producto ofrecido a ciertas poblaciones comparado con grupos privilegiados, y el refuerzo de los estereotipos existentes.^{vii}

En los servicios financieros, el sesgo hacia las mujeres en la IA ha dado lugar, por ejemplo, a discriminación crediticia reflejada en líneas de crédito diferenciadas por género y, en consecuencia, a un desigual acceso y uso de productos financieros. Cuando se aplica en escala, los daños causados por la IA contrarrestan las metas de inclusión financiera en las que se centran muchos inversores de impacto y las empresas en su portafolio. Más allá de los efectos en los consumidores, el sesgo de género de la IA también tiene repercusiones comerciales, regulatorias y de reputación en las empresas que utilizan IA y en sus inversionistas.

Los inversores de impacto están bien posicionados para asumir un papel proactivo e incentivar que las empresas en sus portafolios desarrollen e implementen sistemas de IA equitativos para ofrecer servicios financieros inclusivos. Los inversores pueden influir en la adopción de prácticas de IA responsables y equitativas en el proceso de debida diligencia y en el seguimiento de su portafolio. Sin embargo, evaluar los riesgos potenciales de un sistema de IA requiere una experiencia técnica que muchos inversores no tienen.

Para llenar este vacío, el Centro para la Inclusión Financiera (CFI) desarrolló una breve guía para

que los inversores de impacto examinen de qué manera las empresas invertidas utilizan IA y, con ello, ayudar a comprender la potencial presencia de sesgos con una perspectiva de género. Existe considerable evidencia anecdótica y académica sobre las brechas y barreras que enfrentan las mujeres en el uso de los servicios financieros digitales (DFS), y este documento expone muchos de estos ejemplos. Vale aclarar que, si bien esta guía se centra en los sesgos hacia las mujeres, muchos de los temas y las preguntas incluidas pueden aplicarse ampliamente a otros grupos históricamente marginados por motivos étnicos, de raza o por pertenecer a la población LGBTQ+.

“

Los inversores de impacto están bien posicionados para asumir un papel proactivo e incentivar que las empresas en sus portafolios desarrollen e implementen sistemas de IA equitativos para ofrecer servicios financieros inclusivos.”





Objetivos y organización de esta guía

Esta publicación ofrece una guía de conversación para inversionistas y empresas que utilizan IA, con la meta de fomentar una comprensión mutua más sólida del riesgo de sesgos algorítmicos y las prácticas emergentes de gestión de estos. Cuando estas conversaciones no se llevan a cabo, los inversores y las empresas en las que invierten corren el riesgo de pasar por alto estos problemas y no implementar prácticas equitativas a tiempo.

ENFOQUE DEL PROYECTO

Para elaborar esta guía, CFI inició aplicando una evaluación de necesidades a cuatro inversores de impacto en la que se mapeó los procesos de debida diligencia y post-inversión existentes. Los resultados sugieren que, a pesar de las preocupaciones sobre la inequidad generada por la IA, los inversores no poseían el conocimiento necesario ni las herramientas para mantener conversaciones significativas cuando evaluaban invertir en empresas que utilizan algoritmos. En una siguiente etapa, se revisó la literatura académica sobre la IA, los sesgos en los servicios financieros, y las prácticas de mitigación, identificación y corrección de estos. Adicionalmente, CFI analizó más de 120 guías y herramientas existentes sobre IA responsable y ética. A pesar de que existe una multitud de recursos, la revisión de CFI reveló una escasez de guías prácticas para ayudar con la identificación y gestión de riesgos enfocada en los actores no técnicos, incluidos los inversores de impacto.^{viii}



ESTRUCTURA DE LA GUÍA

Esta guía se enfoca en los inversores de impacto y destaca el valor de priorizar la IA equitativa. Al mismo tiempo, ofrece sugerencias prácticas para entender cómo las fintech utilizan la IA e identificar sesgos algorítmicos. En la primera sección, se explica el uso de la IA en las finanzas inclusivas, los factores que explican el sesgo algorítmico dañino hacia las mujeres, y presenta un panorama actual de las prácticas de mitigación, identificación, y corrección de sesgos. Esta sección es útil para cualquier actor en el sector de la inclusión financiera.

La sección 2 brinda una guía conversacional dirigida a inversionistas, compuesta por un conjunto de preguntas prácticas para ayudarlos a comprender el uso de la IA por parte de las empresas en las que evalúan invertir. La guía también permite identificar riesgos potenciales de discriminación y los efectos negativos diferenciados por motivos de género. Los inversionistas pueden utilizar estas preguntas durante la debida diligencia o después de la inversión, según la etapa de desarrollo de la tecnología y/o los procesos internos de los inversores.



¿Por qué priorizar la IA equitativa?

CONTEXTO

¿Por qué los inversores de impacto deberían priorizar el uso responsable de la IA y evaluar si la aplicación de esta tecnología es equitativa? Porque hacerlo se alinea con sus intereses comerciales y garantiza que se busque el mayor impacto social y beneficio para los consumidores finales. A pesar del papel central que la IA puede desempeñar en las finanzas digitales, la evaluación de riesgos de sesgo algorítmico generalmente no es una parte sustancial de la debida diligencia de los inversores o del compromiso post-inversión. El tiempo asignado a los procesos de debida diligencia tiende a extenderse proporcionalmente al tamaño de la operación de inversión. De hecho, la progresión desde la preselección hasta la debida diligencia y el cierre de la inversión suele ocurrir en menos de seis meses.

Dadas las barreras técnicas para evaluar los sistemas de IA y la actual falta de estándares de IA responsable, no sorprende que no se dé prioridad a la evaluación de los algoritmos en la evaluación de proyectos de inversión. Sin embargo, los inversionistas pueden realizar los procesos de debida diligencia antes de que una fintech esté completamente conformada o que la tecnología haya sido desarrollada. En estos casos, los inversionistas de impacto pueden utilizar esta guía para examinar la visión de la empresa y sugerir contratar científicos de datos responsables y capaces.

RIESGOS PARA LAS CONSUMIDORAS

Desafortunadamente, los algoritmos sesgados a menudo impactan a grupos ya marginados y perpetúan las desigualdades históricas.^{ix} Por ejemplo, en el 2018, Amazon retiró su herramienta de contratación basada en IA después de que se demostró que estaba sesgada contra las mujeres solicitantes.^x El defectuoso sistema de IA había sido entrenado para examinar



Los algoritmos son una serie de reglas matemáticas que definen una secuencia de operaciones. Se pueden utilizar en un programa de computadora, en decisiones analógicas impulsadas por humanos o en una forma híbrida.



Los sistemas de inteligencia artificial (IA) automatizan decisiones en forma similar y más allá de las capacidades humanas.



El machine learning (ML) o aprendizaje automático es una subcategoría de la IA que utiliza algoritmos para aprender automáticamente información y reconocer patrones a partir de datos, aplicando ese aprendizaje para hacer predicciones e informar decisiones.

Fuente: [Lista de Evaluación de Inteligencia Artificial Confiable \(ALTAI\) para autoevaluación](#)

a los solicitantes en base a la observación de patrones en los currículums enviados a Amazon durante los 10 años anteriores. Sin embargo, como la mayoría de los solicitantes históricamente eran hombres, el sistema dio resultados sesgados al preferir candidatos hombres.^{xi}

En el sector financiero, las mujeres ya están desatendidas y no cuentan con servicios bancarios suficientes. La diferencia en la posesión de cuentas bancarias es 7 por ciento menor con relación a los hombres.^{xii} Además, las mujeres

tienen un 23 por ciento menos de probabilidad que los hombres de pedir prestado a una institución financiera y tienen 27 por ciento menos de probabilidad de acceder a productos y servicios de una fintech.^{xiii}

Las decisiones algorítmicas influyen cada vez más en el acceso a los servicios financieros y en las oportunidades económicas de las mujeres. Estas decisiones de alto riesgo incluyen la selección de productos o servicios que se ofrecen a las mujeres y el precio de estos, la aprobación de una solicitud de préstamo o el reclamo de seguro, y si los sistemas de seguridad perciben ciertas transacciones como fraudulentas o no. Cuando las decisiones de los sistemas de IA perpetúan sesgos de género dañinos, estos pueden afectar el acceso y uso de servicios financieros de una mujer y sus oportunidades económicas.

IMPLICACIONES DE NEGOCIO DE UNA IA EQUITATIVA

Las empresas reconocen cada vez más que los sistemas de IA sesgados presentan un riesgo comercial importante. Por ejemplo, Microsoft reconoció este problema en su análisis de riesgo interno y en sus presentaciones ante la Comisión de Bolsa y Valores de EE. UU.^{xiv} En una encuesta realizada a 350 fintech de los EE. UU. y el Reino Unido, más de un tercio informó que su negocio se había visto afectado debido a un algoritmo sesgado. La encuesta también encontró que el 62 por ciento de fintech informó pérdida de ingresos, el 61 por ciento perdió clientes y el 43 por ciento perdió empleados debido a este problema.^{xv}

Para una fintech, excluir o discriminar a potenciales clientes, incluidas las mujeres, podría causar pérdidas de ingresos futuros, así como costos legales, reputacionales y correccionales. Para los inversores en estas empresas, los impactos podrían extenderse en línea ascendente, en forma de menores retornos de la inversión, crecimiento empresarial insostenible y costos reputacionales.^{xvi} En una encuesta de la Unidad de Inteligencia de The Economist realizada a altos ejecutivos, el 94 por ciento de los encuestados opina que la IA responsable produciría un mayor retorno de la inversión (ROI) a largo plazo para los inversores o accionistas.^{xvii}

RIESGOS OPERATIVOS Y REPUTACIONALES

Según una encuesta global de Accenture realizada a 1.500 ejecutivos de alto nivel, las empresas que implementaron prácticas responsables de IA tenían 1.7 veces más probabilidad de ampliar esta tecnología con éxito.^{xviii} Corregir o cancelar un sistema de IA sesgado puede generar altos costos debido a la pérdida de tiempo y de recursos de los empleados asociados con el desarrollo del sistema sesgado y la corrección de este.^{xix} Asimismo, la retención de empleados y la contratación de talento también pueden verse afectados. Tal es el caso de Google, que enfrentó protestas y huelgas de empleados que mostraban oposición a su dron impulsado por inteligencia artificial, y de Meta, que tuvo problemas con el reclutamiento de empleados después del escándalo de Cambridge Analytica, con tasas de aceptación laboral que disminuyeron aproximadamente del 85 al 45 por ciento.^{xx}

CUMPLIMIENTO

Aunque decenas de países y organismos multilaterales como la Organización para la Cooperación y el Desarrollo Económicos (OCDE) han publicado principios y estrategias de alto nivel sobre inteligencia artificial ética o responsable, la regulación real de la IA en los servicios financieros aún es incipiente y no ha sido probada. Si bien la Unión Europea avanza con un proyecto de Ley de Inteligencia Artificial (AIA), la mayoría de las demás jurisdicciones, en particular las de los mercados emergentes, no han avanzado con un modelo legislativo integral.^{xxi}

Las fintech y sus inversores deben ser conscientes de las políticas y las disposiciones regulatorias existentes en sus países de operación que abordan la privacidad de los datos, la igualdad y la discriminación, además de la manera en que los reguladores financieros aplican esos enfoques en la toma de decisiones algorítmicas. Si bien la regulación financiera relacionada explícitamente con la IA aún es nueva, las fintech que no cumplan con estas normas en el futuro cercano podrían ser halladas infractoras e incurrir en multas.^{xxii} En 2022, la Oficina de Protección Financiera del Consumidor impuso una multa de más de \$2.7 millones a una fintech llamada Hello Digit por un algoritmo defectuoso que agotó los ahorros de los consumidores.^{xxiii}

TABLA 1:

¿Cómo se propaga el sesgo algorítmico de género? Un ejemplo de decisión de crédito

¿CÓMO SE VE LA IA NO EQUITATIVA?



Sitha, una mujer de una minoría étnica en un mercado periurbano del sudeste asiático, solicita un préstamo digital en una fintech para ayudar a que su negocio crezca. Sin embargo, la empresa rechazó su solicitud. El modelo de IA que evaluó la solvencia crediticia de Sitha se apoyó en datos de clientes anteriores (incluidos ingresos, educación, género, edad y acceso al crédito), lo que refleja la exclusión histórica de las mujeres y el grupo étnico de Sitha. Por ello, el modelo no predice con precisión la capacidad o la voluntad de pago de Sitha.

El caso de Sitha no es aislado. El modelo de IA sesgado hace que las mujeres solicitantes tengan muchas más probabilidades de volverse un “falso negativo”, lo que significa que son rechazadas a pesar de tener una mayor propensión a pagar que los hombres.^{xxiv}

EL IMPACTO EN EL CONSUMIDOR



Sitha no tiene acceso al préstamo digital que necesita para hacer crecer su negocio. En consecuencia, recurre a un prestamista informal local, cuyas tasas de interés exorbitantes hacen que le sea casi imposible a Sitha saldar la deuda y obtener suficiente margen de beneficio para hacer crecer su negocio.

IMPACTO EN LAS FINTECH



Al rechazar a las mujeres que probablemente no habrían tenido problemas de pago de deudas con “falsos negativos”, el prestamista digital pierde “buenos” clientes y su capacidad para aumentar su participación en el mercado. Si la empresa identifica el problema de sesgo algorítmico en el futuro, probablemente incurrirá en gastos para corregir el sistema de IA. Además, dado que el prestamista digital ofrece una mejor calificación crediticia para clientes con las mismas características socioeconómicas pero diferente género o etnia, la empresa podría ser legalmente culpable de discriminación.

EL IMPACTO EN EL INVERSOR



Es posible que los inversores reciban un mayor retorno a su inversión si el sistema de inteligencia artificial del prestamista digital no generara “falsos negativos” o si no estuviera sesgado. Además, invertir en una tecnología sesgada o asociarse con un prestamista digital que puede ser considerado culpable de discriminación crea la posibilidad de que se produzcan costos legales o daños a su reputación.





Sección 1: Casos de uso de la IA, fuentes de sesgo y estado de la mitigación, identificación y corrección de sesgos

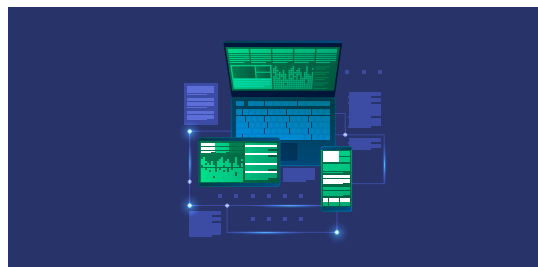
USOS DE LA IA EN LAS FINANZAS INCLUSIVAS

Si bien la evaluación crediticia es el uso más conocido de inteligencia artificial en las finanzas inclusivas, las fintech aplican la IA en diferentes procesos comerciales. Los inversores deben preguntar y comprender de qué manera una fintech integra la IA en su modelo de negocio, especialmente si está tomando o apoyando decisiones de alto riesgo que impactan positiva o negativamente la vida financiera de los consumidores. Este apartado sigue la clasificación desarrollada por el Cambridge Centre for Alternative Finance y el Foro Económico Mundial en su estudio global sobre la adopción de la IA en las fintech:^{xxv}

1 GENERACIÓN DE INGRESOS

Las fintech utilizan la IA principalmente para el análisis de datos y para informar la toma de decisiones al diseñar nuevos productos o procesos.^{xxvi} Dos aplicaciones relevantes son:

- **Créditos:** Las empresas utilizan machine learning para automatizar las evaluaciones crediticias y reducir el riesgo de incumplimiento de pagos de deuda.^{xxvii} Las fintech pueden aplicar estos modelos utilizando datos tradicionales, incluida información sobre transacciones o historial crediticio. Pero principalmente innovan al utilizar fuentes de datos no estructurados y semiestructurados (incluyendo la actividad en las redes sociales, los datos satelitales, la actividad de teléfonos móviles y mensajes de texto) que pueden dar una visión más compleja de la solvencia crediticia, especialmente para aquellos sin registros o historial bancario formal.^{xxviii}
- **Seguros:** Al igual que la suscripción de créditos, las empresas de seguros (Insurtech) utilizan la IA y el machine



learning para automatizar procesos y analizar datos e información no tradicionales. Las soluciones de Insurtech existentes en África subsahariana y el sur de Asia utilizan inteligencia artificial y datos alternativos (incluidos datos satelitales y dispositivos portátiles) para clasificar riesgos y fijar precios.^{xxix}

2 GESTIÓN DE RIESGOS

Las empresas implementan periódicamente la IA para clasificar y detectar transacciones como potencialmente fraudulentas o anómalas.^{xxx} Las fintech también suelen utilizar la IA para el análisis preventivo de patrones de nuevos conjuntos de datos y la gestión de riesgos.^{xxxi} Por ejemplo, la plataforma Falcon de FICO aprende de los perfiles de comportamiento de los clientes y detecta fraudes.^{xxxii} Algunos casos de uso adicionales incluyen:

- **Conozca a su cliente (Know Your Client, KYC):** Las empresas también pueden utilizar la IA para transformar o mejorar los controles de identidad dentro de una institución financiera. Si bien estos procesos históricamente dependían de nombres de usuario y contraseñas, ahora pueden usar reconocimiento de voz, reconocimiento facial u otros datos biométricos similares, lo que puede reducir el riesgo de que los nombres de usuario y las contraseñas se vean vulnerados o se compartan.^{xxxiii} Por ejemplo, Smile Pay de Alipay utiliza el reconocimiento facial para autenticar y

dar consentimiento a una transacción, lo que permite a los clientes minoristas un proceso de pago sin fricciones.^{xxxiv}

- **Ciberseguridad:** Las empresas pueden utilizar la IA y el machine learning para automatizar la detección de amenazas cibernéticas. Por ejemplo, esta tecnología puede identificar datos e información comprometidos, dirigir investigaciones después de incidentes y extraer información para compartir con otras instituciones financieras y autoridades pertinentes.

3 SERVICIO AL CLIENTE

Los chatbots y los asistentes virtuales ofrecen más opciones para que los clientes obtengan ayuda cuando la necesitan y brinden perfiles y consejos más personalizados. La IA generativa (generative AI) ha despertado recientemente un gran interés por su potencial para revolucionar las interacciones y los chatbots de trato directo con los clientes. Los principales casos de uso en este dominio incluyen servicios complementarios habilitados para IA, soluciones de apertura de cuentas digitales y marketing habilitado para IA.^{xxxv}

4 AUTOMATIZACIÓN Y REINGENIERÍA DE PROCESOS

A menudo, las fintech implementan la IA para consolidar y automatizar tareas administrativas, como informes y auditoría.^{xxxvi} Por ejemplo, en el sector de los seguros, las empresas utilizan esta tecnología para evaluar y procesar los reclamos.^{xxxvii}

No todos los usos de la IA en las fintech tienen el mismo impacto en los consumidores. Hay casos de uso en los que la IA impulsa operaciones como una base de datos transaccional de back-end o un filtro de spam, aplicaciones que no tienen una consecuencia directa para el consumidor. Pero hay casos en los que la IA toma o ayuda a tomar decisiones, como el otorgamiento de créditos y seguros, que tiene un impacto directo en los consumidores y sus oportunidades económicas. Esta guía presenta estos últimos casos de uso como intrínsecamente más riesgosos desde una perspectiva de inclusión financiera y les da un mayor escrutinio.

Esta distinción en los casos de uso se basa en parte en las clasificaciones de riesgo de la IA que la Unión Europea propuso en su proyecto de Ley de Inteligencia Artificial (Artificial Intelligence Act, AIA). Esta regulación contiene responsabilidades mucho mayores para los sistemas de IA de alto riesgo, incluida la calificación crediticia, que para los sistemas de riesgo mínimo o limitado, como un chatbot.^{xxxviii}

FUENTES DE SESGOS DAÑINOS

El sesgo perjudicial de la IA no es únicamente una función de los datos o códigos de entrada subyacentes y puede surgir de varias fuentes. Los sesgos también dependen de la aplicación de la tecnología, las metas de uso y otros factores contextuales. Por ejemplo, un algoritmo de procesamiento de reclamaciones de seguros puede ser justo en una Insurtech, pero podría ser potencialmente dañino si lo utiliza una empresa diferente o es replicado de un país a otro dentro de la misma empresa.



Repensar el término “sesgo”:

El término “sesgo algorítmico” casi siempre tiene una connotación negativa. Sin embargo, a nivel técnico, es un término neutral ya que existe cuando los resultados de un algoritmo se desvían de manera sistemática en relación con una norma o estándar. Estas normas o estándares podrían ser estadísticos, institucionales, morales o incluso legales, y cambian de una fintech a otra. Si bien el término “sesgo algorítmico” suele aplicarse a una decisión que genera resultados injustos, no todos los sesgos son necesariamente malos y algunos pueden incluso ser deseables. Por ejemplo, una fintech puede crear un modelo que deliberadamente incorpore un sesgo estadístico en beneficio de las mujeres, de modo que el resultado del algoritmo se desvíe de las predicciones de los datos de entrenamiento.

Fuentes: [Lista de evaluación de inteligencia artificial confiable \(ALTA\) para autoevaluación](#); [Fazelpour y Danks, 2021](#); [Anthony, 2016](#); [Johnson, 2021](#)

Los sistemas algorítmicos computacionales suelen implicar varios pasos que incluyen la especificación del problema, la recopilación de datos, el preprocesamiento de datos, el modelado, la validación y la implementación. Los inversores deben comprender que el sesgo puede aparecer en cualquier etapa y, dado que estos procesos son altamente iterativos, pueden depender bastante del personal involucrado.^{xxxix} Para ayudar a categorizar posibles fuentes de sesgo en la IA de las finanzas inclusivas, CFI ha utilizado su marco conceptual de insumos, código y contexto.^{xi} Si bien este marco ayuda a explicar el desarrollo de algoritmos y facilita la categorización de riesgos, en la práctica, las categorías se superponen y un problema puede ser relevante tanto para las categorías de insumos y código a la vez, por ejemplo. Si bien no son exhaustivos, los siguientes ejemplos son un punto de partida útil.

INSUMOS

- **Sesgo histórico:** Puede entrar en los sistemas algorítmicos debido a dinámicas culturales, sociales o institucionales preexistentes y de larga data. Por ejemplo, si un sistema de inteligencia artificial que evalúa la solvencia crediticia utiliza datos históricos de ingresos para predecir si una persona puede pagar un préstamo, la brecha sistémica de género en los salarios, ingresos, y activos podría generar resultados injustos para las mujeres solicitantes.
- **Sesgo muestral:** Puede ocurrir al recopilar datos o tomar muestras de una población para respaldar el desarrollo de un modelo. Los métodos de muestreo pueden estar limitados en el sentido de que solo llegan a una población reducida, lo que puede llevar a un algoritmo que produzca un resultado que ponga en desventaja a los grupos no representados igualitariamente en el muestreo. Por ejemplo, la brecha de género bien documentada en el acceso a las finanzas y a los servicios financieros digitales puede significar que las mujeres no estén representadas suficientemente en las muestras de los conjuntos de datos.^{xii}
- **Sesgo de selección:** El sesgo de selección ocurre cuando los datos para individuos específicos, como mujeres u otros grupos

marginados, no están disponibles. Los modelos de crédito, por ejemplo, generalmente se basan en personas cuyas solicitudes de préstamo fueron aceptadas en lugar de rechazadas. Dado que el sistema financiero históricamente ha excluido a las mujeres, los datos utilizados para entrenar algoritmos podrían reflejar las características de un subconjunto de solicitantes de préstamos: Hombres cuyas solicitudes de préstamo fueron históricamente aceptadas.

CÓDIGO

- **Sesgos en la especificación de problemas:** El sesgo puede aparecer cuando el equipo de ciencia de datos elige cómo optimizar el algoritmo para alinearlo con una estrategia específica. Los resultados dependen en gran medida de las metas que se establecen (como maximizar las ganancias, garantizar la privacidad del cliente o minimizar el trato o los resultados desiguales para las mujeres), cada uno de los cuales puede tener resultados diferentes. Si una fintech se enfoca en los solicitantes de préstamos que tienen más probabilidades de pagar, pero el algoritmo de predicción está optimizado para identificar aquellos que devolverán las mayores ganancias a la institución, entonces el algoritmo no dará los resultados adecuados.
- **Sesgos en el preprocesamiento:** Los algoritmos no pueden consumir datos incompletos e inconsistentes, y el “ruido” altera el verdadero patrón de la muestra. El procesamiento previo de los datos tiene como objetivo resolver estos desafíos. Sin embargo, durante esta etapa, un algoritmo debe tomar decisiones que podrían introducir sesgos. Este proceso implica importantes decisiones sobre cómo manejar los datos faltantes de grupos subrepresentados, incluidas las mujeres, o cómo identificar y abordar los valores atípicos. Podrían ocurrir otros problemas debido a errores manuales, eventos inesperados y problemas técnicos.
- **Sesgo por proximidad:** Algunas fintech bien intencionadas defienden sus

sistemas de IA bien intencionados diciendo que el sesgo de género es imposible porque intencionalmente excluyen datos sobre atributos sensibles como el género de sus modelos. Lamentablemente, este argumento no se sostiene en la práctica. Los algoritmos de finanzas inclusivas están procesando cantidades cada vez mayores de datos, así como datos provenientes de nuevas fuentes. Como resultado, los modelos de IA pueden utilizar inadvertidamente sustitutos de datos de atributos sensibles en las decisiones, incluso si no fue algo intencional. Por ejemplo, los tipos de teléfonos o apps pueden reflejar patrones de género, edad u otras características, lo que a veces puede resultar engañoso.

CONTEXTO

- **Sesgo por transferencia de contexto:** Las fintech implementan algoritmos para usos o propósitos específicos y en un contexto particular. Si se emplean fuera de esos contextos, es posible que no se desempeñen de acuerdo con los estándares apropiados. Por ejemplo, los algoritmos de evaluación de riesgos entrenados con datos históricos en Estados Unidos pueden no traducirse bien a un entorno de economía emergente.
- **Sesgo de interpretación:** Incluso si se emplea en el contexto adecuado, el sesgo algorítmico puede surgir de una mala interpretación que un miembro del personal haga de los resultados o del funcionamiento del algoritmo. Estas fallas a veces pueden ocurrir sin que el personal se dé cuenta. Los algoritmos podrían describirse al personal simplemente como “predecir el éxito del solicitante de préstamo” sin definir qué constituye o cómo se mide el éxito.
- **Sesgo de representación:** Cuando el personal que diseña, opera y gobierna los modelos carece de diversidad, puede aumentar el riesgo de problemas de sesgo y discriminación en los modelos de machine learning. La falta de representación debilita la capacidad de las fintech de reconocer y responder a los problemas mientras desarrollan y

utilizan modelos. Esta cuestión se vuelve particularmente problemática cuando se tiene en cuenta la brecha de género documentada dentro de los puestos de liderazgo y los campos de STEM.

EL ESTADO DE LA PRÁCTICA EN LA MITIGACIÓN, DETECCIÓN Y CORRECCIÓN DE SESGOS DAÑINOS EN LA IA

Para comprender mejor el panorama actual de la práctica de mitigar, detectar y corregir sesgos dañinos, CFI clasificó los enfoques en tres categorías: 1) enfoques para la sensibilización general sobre la IA ética y el desarrollo de códigos de conducta sólidos; 2) herramientas orientadas a procesos de negocio y prácticas de documentación de datos; y 3) técnicas centradas en prácticas específicas de ciencia de datos para detectar sesgos e implementar medidas correctivas.

A través de este análisis, CFI identificó una brecha en los recursos disponibles para los actores interesados no técnicos, como los funcionarios de inversiones, para ayudar a discernir si una empresa en la que evalúan invertir ha tomado las medidas adecuadas para mitigar, detectar y corregir sesgos dañinos. Esta guía tiene como objetivo ayudar a cerrar esta brecha.

PRINCIPIOS ÉTICOS Y CÓDIGOS DE CONDUCTA

Estos principios pueden fomentar una comprensión más amplia de los algoritmos éticos, incluidos los principios de equidad, transparencia, explicabilidad y auditabilidad dentro de una empresa. Si bien brindan una orientación más amplia a nivel organizacional, no proveen enfoques técnicos detallados necesarios para la detección y mitigación de sesgos.

En general, el código de conducta de una fintech debe abogar por que sus algoritmos respeten los siguientes principios:

- i. Estar basados en motivos de no maleficencia (es decir, no deberían causar daño);
- ii. Estar orientados a brindar beneficios a la humanidad;

- iii. Garantizar la privacidad y la seguridad de los datos de las personas cuyos datos se recopilan y utilizan;
- iv. Incorporar transparencia, explicabilidad y auditabilidad en sus procesos; y
- v. Asegurarse de que los productos. ^{xiii}

Sin embargo, si bien estos principios son útiles para crear conciencia y alineación con objetivos más amplios, carecen de mecanismos de aplicación y herramientas concretas que puedan aplicarse para mitigar los sesgos en las finanzas inclusivas.

En 2019, Google Cloud comenzó a recibir solicitudes de clientes empresariales que pedían soluciones para préstamos basados en inteligencia artificial utilizando datos no tradicionales. En 2020, a pesar de realizar un sprint para desarrollar prototipos para cubrir estas necesidades, finalmente decidió no crear una aplicación algorítmica para préstamos. El equipo creía que un “producto creado (con las tecnologías y los datos actuales) podría crear un impacto dispar relacionado con el género, la raza y otros grupos marginados, y entrar en conflicto con el principio de Google AI de ‘evitar crear o reforzar prejuicios injustos’”.

Fuentes: [McElhanev et al, 2022](#); [Informe de progreso de la IA de Google](#)

HERRAMIENTAS DE DOCUMENTACIÓN DE DATOS Y PROCESOS DE NEGOCIO

MARCOS DE AUDITORÍA Y KITS DE HERRAMIENTAS SOBRE SESGOS

Los marcos y herramientas de esta categoría pueden incluir la gobernanza, la evaluación

de potenciales impactos, las herramientas de auditoría y los estándares que tienen como objetivo garantizar que los sistemas de IA sean responsables. Los marcos de auditoría van más allá de los códigos de conducta y los principios, al agregar herramientas relevantes para mejorar el seguimiento y la diligencia. ^{xiii} La mayoría de estos marcos proporcionan una combinación de hojas de trabajo, cuestionarios y listas de verificación que las fintech pueden utilizar para promover la alineación con los principios rectores o códigos de conducta establecidos. Algunos implican procesos integrales de auditoría y gobernanza de principio a fin de que las empresas pueden completar la identificación de sesgos algorítmicos con apoyo interno o externo.

DOCUMENTACIÓN DE DATOS PARA EL DESARROLLO ALGORÍTMICO: HOJAS DE DATOS Y TARJETAS DE MODELO O MÉTODO

Para generar transparencia y permitir un compromiso más sólido entre los desarrolladores de tecnología y otras partes interesadas, una fintech puede utilizar datos robustos y prácticas de documentación de modelos para desarrollar un sistema algorítmico. Esta transparencia puede ayudar a la empresa a identificar riesgos desde el principio, facilitar auditorías en investigaciones en caso de que surjan problemas después de la implementación y facilitar que los futuros miembros del equipo de ciencia de datos realicen adaptaciones informadas. ^{xiv}

Entre los ejemplos populares se incluyen las hojas de datos (datasheets) o tarjetas para documentar las bases de datos y modelos. Las hojas de datos documentan el objetivo para recolectar información, la metodología y procesos de recopilación de datos, así como los procedimientos de limpieza, etiquetado y manipulación de datos. Asimismo, se documentan los detalles de la distribución de datos y del mantenimiento de los conjuntos de datos. Las tarjetas de modelo y método documentan los atributos clave de un modelo algorítmico y sus características de rendimiento. Un ejemplo de la aplicación de estas herramientas para evitar sesgos algorítmicos es la etiqueta nutricional de bases de datos (Dataset Nutrition Label), que permite el uso informado y estandarizado de bases de datos. ^{xv}

DETECCIÓN DE SEGOS Y MEDIDAS CORRECTIVAS EN CIENCIA DE DATOS

Estos enfoques se basan en marcos de IA responsable e implican técnicas específicas de mitigación y detección de sesgos. Por lo general, siguen una estrategia de dos etapas para abordar el sesgo algorítmico. La primera implica definir y utilizar una o más medidas de equidad matemática para cuantificar la cantidad de sesgo en el resultado algorítmico; y la segunda etapa incluye desarrollar respuestas de mitigación que reduzcan cualquier sesgo problemático. Para el último paso, las fintech pueden integrar una gama cada vez mayor de enfoques técnicos en el diseño algorítmico, preprocesamiento, procesamiento, y post-procesamiento.^{xvi} Women's World Banking, por ejemplo, creó una herramienta basada en Python para auditar la equidad entre las solicitudes de crédito aceptadas y rechazadas.^{xvii} FinRegLab publicó una investigación que aplicó herramientas avanzadas de "explicabilidad" para analizar modelos de evaluación de crédito basados en machine learning. Tras analizar los modelos de siete empresas, FinRegLab encontró que herramientas identificaron de manera confiable características de modelos claves para entender la solvencia crediticia de los consumidores y otorgar préstamos de manera justa.^{xviii}



Nota sobre la definición de equidad:

Muchos de los avances recientes en las pruebas y auditorías para detectar sesgos se enmarcan en torno al concepto de equidad. Si bien la proliferación de medidas de equidad ha proporcionado un punto de partida útil para el desarrollo de técnicas de mitigación de sesgos, no existe un consenso sobre la elección óptima de métricas a aplicar, particularmente cuando se trata de la prestación de servicios financieros.

Definir un enfoque de equidad implica tomar decisiones y hacer concesiones. Estas decisiones deben estar bien documentadas para la comunicación interna y para permitir una discusión y debates saludables. Las definiciones de equidad pueden llevar a los algoritmos a aplicar ciertas métricas de equidad (paridad entre prestatarios hombres y mujeres, por ejemplo), pero pueden tener como costo modelos menos precisos.

Las fintech que están creando algoritmos y sistemas de IA deben asegurarse de prestar la atención adecuada a la definición de lo

que esta tecnología intenta lograr y los sesgos dañinos que intenta evitar (por ejemplo, objetivos éticos, contextuales, institucionales o legales). Luego, el equipo debe seleccionar cuidadosamente las métricas de equidad que estén alineadas con estas metas. Si bien es posible que la gerencia y los ejecutivos de alto nivel no profundicen en los detalles técnicos de cómo se operacionalizan las métricas de equidad, debe haber un entendimiento documentado entre la gerencia y los equipos de ciencia de datos con respecto a la definición, medición y responsabilidad del desempeño frente a las métricas de equidad relevantes. Los inversores de impacto también deben comprender de qué manera la fintech define, operacionaliza y monitorea la equidad, y decidir si esos indicadores se alinean con sus propias metas de inversión.

Fuentes: [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](#)



Sección 2: Guía de conversaciones para promover una IA equitativa

Esta sección proporciona a los oficiales de inversión un conjunto práctico de preguntas para comprender el uso de la IA por parte de las empresas en portafolio e identificar riesgos potenciales de sesgos nocivos de género. Resaltamos que estas preguntas también son aplicables a otros grupos minoritarios o marginados. La guía conversacional está diseñada para usarse durante el proceso de debida diligencia y puede ayudar a informar decisiones de inversión o decisiones que son parte del seguimiento al portafolio de inversión. Es importante tener en cuenta que algunas fintech en fase inicial pueden tener pendiente la construcción de los modelos en el momento en que se produce la diligencia debida, y estas preguntas pueden ser más apropiadas en una etapa posterior a la inversión.

CFI adaptó las preguntas que se incluyen en esta guía a partir de una revisión de más de 120 herramientas éticas de IA existentes. La selección de las preguntas que se incluyeron en última instancia se basó en la evaluación de las necesidades de los inversores, el contexto de las finanzas inclusivas, y consultas con expertos y actores implicados. La guía se adaptó a los inversionistas en el espacio fintech que no tienen contacto directo con el equipo de ciencia de datos.



NOTAS DE USO

La guía destaca áreas clave sobre las que los inversores y las empresas que reciben capital deben conversar para promover el uso equitativo de la IA y satisfacer expectativas mutuas al respecto. CFI no recomienda calificar cuantitativamente las prácticas de una fintech basándose en las preguntas incluidas en esta guía, ya que daría una falsa sensación de precisión. Dada la naturaleza dinámica de la ciencia de datos que impulsa la IA, la falta de consenso sobre las

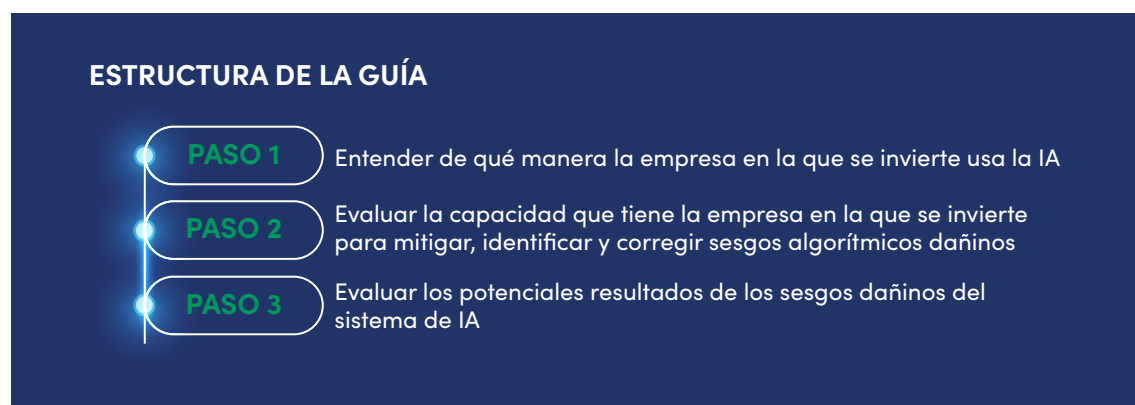


técnicas específicas de mitigación de los sesgos y la incertidumbre sobre la regulación sobre el uso responsable de esta tecnología, la guía tiene un fin más orientativo que exhaustivo o inmutable.

Algunas preguntas se refieren a áreas en las que es suficiente una respuesta autoreportada por parte de la empresa que busca inversión. Por el contrario, ciertas respuestas podrían revelar áreas potenciales de mejora o indicar la necesidad de discusiones más profundas para establecer expectativas. Esta guía eleva los procesos de negocio, como la gestión y la gobernanza de datos, que respaldan el desarrollo algorítmico y los considera prioridades clave junto con el código y los resultados. Dado que muchos modelos son dinámicos, observar únicamente los resultados ofrece sólo una fotografía de un sistema en constante evolución.

En esta guía, las preguntas y las estrategias que las acompañan se estructuran en torno a tres pasos. Cada paso tiene una breve explicación al principio, está seguido de preguntas y concluye con una sección de “Explorar más”. Esta última sección brinda referencias utilizadas en el desarrollo de cada sección y ofrece oportunidades para profundizar y revisar ejemplos pertinentes.

Si bien el futuro de la IA dentro de las finanzas inclusivas está en constante evolución, es importante que los inversores eviten un enfoque pasivo de “esperar a ver qué pasa”. Aunque la industria carece de consenso sobre técnicas efectivas de mitigación de sesgos específicos y existe incertidumbre en torno al desarrollo de una regulación responsable de la IA, la maleabilidad de este momento representa una oportunidad para que los inversores de impacto promuevan cambios significativos. Al incorporar activamente los principios de la IA equitativa en el desarrollo de productos de las fintech, los inversores pueden garantizar un futuro más inclusivo para los servicios financieros.



ESTRUCTURA DE LA GUÍA

PASO 1

Entender de qué manera la empresa en la que se invierte usa la IA

- ¿La empresa utiliza la IA en decisiones que pueden afectar las oportunidades económicas de los consumidores, por ejemplo: Evaluación de créditos, suscripción de seguros, adquisición de clientes, KYC, etc.?
- ¿La estrategia y el objetivo de negocio de AI de la empresa están alineadas con los del inversionista?
- ¿La empresa en la que se invierte ha identificado y documentado a los individuos y grupos en riesgo que podrían ser afectados o resultar sistemáticamente en desventaja debido al uso del algoritmo?

PASO 2

Evaluar la capacidad que tiene la empresa en la que se invierte para mitigar, identificar y corregir sesgos algorítmicos dañinos

Insumos: datos, modelo y pruebas	<ul style="list-style-type: none"> ¿La empresa ha definido lo que significa un sistema algorítmico “justo”, las medidas o indicadores asociados y los umbrales de medición? ¿La empresa ha implementado medidas para reducir la injusticia en los datos, el modelamiento y el testeo?
Contexto: Gobernanza	<ul style="list-style-type: none"> ¿La empresa está comprometida con un marco de IA equitativo o con un marco de IA responsable de forma proactiva y sistemática?
Contexto: Diversidad y formación	<ul style="list-style-type: none"> ¿El equipo de ciencia de datos tiene diversidad de género? ¿Se ha capacitado al equipo de ciencia de datos sobre la manera en que el sesgo puede introducirse a las herramientas de IA y las formas de mitigarlo?
Contexto: Mecanismos de resolución de quejas	<ul style="list-style-type: none"> ¿La empresa tiene un plan de respuesta, reparación o mecanismo de recurso si los resultados perjudican a las mujeres?

PASO 3

Evaluar los potenciales resultados de los sesgos dañinos del sistema de IA

Customer acquisition and outreach	¿Existe un equilibrio de género en la demanda del producto de la fintech?
	¿Se da la misma prioridad a hombres y mujeres en las campañas de marketing o de servicio al cliente?
KYC y automatización	¿Los solicitantes hombres y mujeres tienen la misma probabilidad de superar con éxito los pasos de KYC?
	¿Los solicitantes hombres y mujeres tienen el mismo éxito al momento de cumplir con otros requisitos de documentación y filtros automáticos?
Para productos de crédito y seguros	¿Los candidatos hombres y mujeres tienen la misma puntuación de crédito/riesgo después de tener en cuenta otros factores relevantes?
	¿Los candidatos hombres y mujeres tienen la misma probabilidad de recibir una oferta de crédito/seguro después de tener en cuenta otros factores relevantes?
Para ciberseguridad/detección de fraudes	¿Las transacciones de hombres y mujeres tienen una probabilidad similar de ser marcadas correctamente como actividad fraudulenta?

VERSIÓN EXTENDIDA

PASO 1

Entender de qué manera la empresa en la que se planea invertir usa la inteligencia artificial (IA)

Orientación para completar este paso

Al evaluar una nueva inversión o formar parte de la junta directiva, el inversor debe estar familiarizado con el uso de la IA por parte de la empresa participada, especialmente si toma decisiones de alto riesgo que impactan a los consumidores. Un inversor también debería familiarizarse con el saber si las operaciones de IA de la fintech están sujetas a leyes antidiscriminatorias, regulaciones de privacidad de datos, regulaciones financieras sobre algoritmos y regulaciones más amplias que rijan la IA ética.

Preguntas:	SÍ	NO	NS	Fuentes de información a verificar
¿La empresa utiliza la IA en decisiones que pueden afectar las oportunidades económicas de los consumidores, por ejemplo: Evaluación de créditos, suscripción de seguros, adquisición de clientes, KYC, etc.?				<ul style="list-style-type: none"> ➤ Estrategia ➤ Plan de negocios ➤ Reporte anual ➤ Solicitud de cliente ➤ Entrevistas con funcionarios de la empresa en la que se planea invertir
¿La empresa tiene una estrategia y objetivos de negocio de IA documentados?				
¿La estrategia y el objetivo de negocio de AI de la empresa están alineadas con los del inversionista?				
¿La empresa en la que se invierte ha identificado y documentado a los individuos y grupos en riesgo que podrían ser afectados o resultar sistemáticamente en desventaja debido al uso del algoritmo?				

Explorar más

- [Transformando paradigmas: una encuesta global sobre IA en servicios financieros](#)
- [Inversión responsable en la IA: una debida diligencia sobre IA responsable para los capitalistas de riesgo](#)
- [IA para siempre: conocimientos de investigación de los servicios financieros](#)
- [Las historias que cuentan los algoritmos: sesgo e inclusión financiera en los márgenes de los datos](#)
- [Sesgo algorítmico, inclusión financiera y género](#)
- [Inteligencia artificial: superpoderes prácticos: el caso de la IA en los servicios financieros en África](#)
- [Explicabilidad y equidad del machine learning: perspectivas de los préstamos al consumo](#)
- [Reflejar el pasado, dar forma al futuro: hacer que la IA funcione para el desarrollo internacional](#)
- [Anexo de casos de uso en USAID](#)

PASO 2

Evaluar la capacidad de la empresa para mitigar, identificar y corregir sesgos dañinos

Orientación para completar este paso

Cuando una empresa utiliza un sistema de inteligencia artificial en decisiones de alto riesgo que impactan al consumidor, los inversionistas interesados en financiarlas deben evaluar la capacidad de estas empresas para mitigar, identificar y corregir sesgos algorítmicos. Dado que este problema puede ser originado en los procesos operativos de la empresa, esta sección analiza las prácticas organizacionales, incluida la dirección, capacitación, diversidad, modelamiento, testeo, así como los mecanismos de reporte y reparación. Este paso proporciona un diagnóstico general de las áreas y prácticas que necesitan apoyo. Si una empresa comprara un algoritmo diseñado por otro proveedor, muchas de las preguntas relacionadas a los insumos y código estarían dirigidas a ese proveedor.

Preguntas:	SÍ	NO	NS	Fuentes de información a verificar
Insumos: Datos				
¿La empresa ha definido lo que significa un sistema algorítmico "justo", incluyendo los indicadores y umbrales asociados para su medición?				
¿La empresa ha implementado medidas para reducir la injusticia en los datos, el modelamiento y el testeo?				
¿La definición de equidad se basa en el contexto social del mercado en el que opera la empresa?				
¿La empresa ha definido un proceso para identificar grupos en riesgo?				
¿La empresa ha comprobado si los datos que utiliza para entrenar a los modelos son representativos de la población de hombres y mujeres a la que intenta atender?				<ul style="list-style-type: none"> ➤ Estrategia ➤ Plan de negocios ➤ Solicitud de cliente ➤ Entrevistas con empresas participadas y/o proveedor externo
¿El equipo de ciencia de datos ha verificado si la disponibilidad o calidad de los datos difiere según el género?				<ul style="list-style-type: none"> ➤ Informes de la junta
Si los datos no son igualmente representativos, o si la disponibilidad y la calidad difieren según el género, ¿se han implementado medidas correctivas (por ejemplo, sobre muestreo y aumento sintético) para reducir las brechas en los datos de entrenamiento?				<ul style="list-style-type: none"> ➤ Datos del cliente ➤ Entrevistas con el equipo de IA
¿El equipo de ciencia de datos de la empresa tiene un proceso transparente que documente el propósito del desarrollo del algoritmo, los objetivos de recopilación de datos, la metodología de recopilación de datos, los procedimientos de etiquetado de datos, la distribución y la gestión de los datos (por ejemplo, hojas de datos para bases de datos, etiquetas nutricionales de datos, etc.)?				<ul style="list-style-type: none"> ➤ Datos utilizados para modelar ➤ Código de algoritmo ➤ Informes de datos o estadísticas
Código: Modelado y pruebas				
¿El equipo de ciencia de datos de la empresa ha documentado las opciones de diseño del sistema de IA y los atributos clave del modelo algorítmico, el contexto en el que el equipo pretendía utilizar el modelo y las métricas de rendimiento (por ejemplo, tarjetas de modelo o método)?				
¿El equipo de ciencia de datos de la empresa ha documentado el juicio humano en el diseño del algoritmo?				
¿Se mide y prioriza la equidad de la empresa al modelar y probar?				
¿El equipo de ciencia de datos de la empresa ha evaluado y documentado estimaciones cuantitativas del desempeño del sistema en comparación con sus objetivos de equidad (por ejemplo, un número igual de mujeres y hombres con solicitudes de préstamo aceptadas, controlando otras características relevantes)?				

Preguntas:	SÍ	NO	NS	Fuentes de información a verificar
¿La empresa ha evaluado alguna compensación entre los objetivos de equidad del sistema y sus objetivos comerciales?				<ul style="list-style-type: none"> ➤ Estrategia ➤ Plan de negocios ➤ Solicitud de cliente ➤ Entrevistas con empresas participadas y/o proveedor externo ➤ Informes de la junta ➤ Datos del cliente ➤ Entrevistas con el equipo de IA ➤ Datos utilizados para modelar ➤ Código de algoritmo ➤ Informes de datos o estadísticas
¿Existe un proceso para señalar problemas relacionados con sesgos perjudiciales, discriminación o mal desempeño del sistema de IA?				
¿Existe un proceso para verificar si hay desviación del modelo, a fin de garantizar la responsabilidad a lo largo del tiempo?				
Contexto: Gobernanza				
¿La gerencia, a alto nivel, comprende de qué manera el equipo optimizó el algoritmo, cuáles son las metas de equidad y cualquier evaluación de riesgo de impacto desproporcionado contra las mujeres?				
¿La empresa está comprometida con un marco de IA equitativo o con un marco de IA responsable de forma proactiva y sistemática?				
¿La empresa tiene un consejo de ética de IA responsable y activo (u organismo comparable) alineado con una estructura de gobernanza?				
¿La empresa tiene la capacidad de diseñar soluciones de IA que sigan las regulaciones sobre equidad e IA responsable?				
¿La empresa tiene documentos actualizados y fácilmente disponibles para los auditores (por ejemplo, evaluación de riesgos, evaluación de datos, evaluación de modelos, registros de seguimiento de modelos), si es necesario?				
¿La junta directiva recibe informes de gestión periódicos sobre el desempeño algorítmico en comparación con métricas de equidad?				
¿La gerencia o la junta directiva puede cerrar rápidamente el sistema de IA si es necesario?				
Contexto: Diversidad y formación				
¿El equipo de ciencia de datos tiene diversidad de género?				
¿La empresa ha capacitado al equipo de ciencia de datos sobre cómo el sesgo puede introducirse en las herramientas de inteligencia artificial y las formas de mitigarlo?				
¿La empresa ha diseñado una hoja de ruta que priorice los programas de capacitación y concientización sobre el sesgo de la IA en toda la organización?				
Contexto: Mecanismos de resolución de quejas				
¿Los razonamientos de las decisiones tomadas por el algoritmo son transparentes para los consumidores a los que se dirige?				
¿La empresa tiene un plan de respuesta, reparación o mecanismo de recurso si las decisiones de IA perjudican a las mujeres?				
Explorar más				
<ul style="list-style-type: none"> ➤ Inversión responsable en IA: una debida diligencia sobre IA responsable para los capitalistas de riesgo ➤ Mitigar el sesgo en la inteligencia artificial ➤ Autoridad Monetaria de Singapur – Documento 4 de Veritas – Estudios de caso de evaluación de los principios FEAT ➤ Cómo diseñar un consejo de ética de IA ➤ Guía NIST AI RMF ➤ Conjunto de herramientas de ética y algoritmos: un marco de gestión de riesgos para gobiernos (y otras personas también) ➤ Estándares para servicios financieros digitales responsables – SPTF junio de 2022 				

PASO 3

Evaluar el potencial de sesgos dañinos en los resultados del sistema de IA

Orientación para completar este paso

Las siguientes preguntas pueden ayudar a detectar riesgos potenciales de sesgo perjudicial al observar los resultados y los impactos del sistema de IA, pero es posible que se necesiten más auditorías para evaluar y diagnosticar estos problemas con precisión. Las respuestas "No" o "No sé" indican que el sistema de inteligencia artificial de la empresa fintech podría estar produciendo resultados inequitativos contra las mujeres, y se requerirían más conversaciones e investigaciones con el equipo de ciencia de datos. Recuerde que las preguntas centradas en los resultados proporcionan una fotografía en un momento en el tiempo y deben revisarse periódicamente.

NB: El paso 3 puede ser más complicado en un mercado donde la regulación prohíbe la recopilación de características sensibles. Si bien esta regulación tiene la intención de proteger la privacidad de las personas y prohibir a los prestamistas tomar decisiones basadas en estas características sensibles, dificulta la prueba de los resultados de la IA en todos los grupos demográficos.

Pregunta:	SÍ	NO	NS	Fuentes de información a verificar
Adquisición y alcance de clientes				
¿Existe un equilibrio de género en los prospectos de demanda y de entrada para el producto fintech?				
¿Se da la misma prioridad a hombres y mujeres en las campañas de marketing o de servicio al cliente?				
KYC y automatización				
¿Los solicitantes hombres y mujeres tienen la misma probabilidad de superar con éxito los pasos de KYC?				
¿Los solicitantes hombres y mujeres tienen el mismo éxito al momento de cumplir con otros requisitos de documentación y filtros automáticos?				
Para productos de crédito y seguros				
¿Los candidatos hombres y mujeres tienen la misma puntuación de crédito/riesgo después de tener en cuenta otros factores relevantes?				
¿Los candidatos hombres y mujeres tienen la misma probabilidad de recibir una oferta de crédito/seguro después de tener en cuenta otros factores relevantes?				
¿Los hombres y mujeres a los que se les ofrecen productos de crédito/seguro reciben los mismos términos del servicio después de tener en cuenta otros factores relevantes?				
¿Los hombres y mujeres cuyas solicitudes han sido rechazadas tienen la misma puntuación crédito/riesgo después de tener en cuenta otros factores relevantes?				

- ↗ Estrategia
- ↗ Plan de negocios
- ↗ Plan de negocios
- ↗ Aplicaciones de cliente
- ↗ Entrevistas con funcionarios de la empresa
- ↗ Datos administrativos y estadísticas.
- ↗ Entrevistas con el equipo de ciencia de datos, funcionarios de préstamos o seguros

Pregunta:	SÍ	NO	NS	Fuentes de información a verificar
¿Los motivos de rechazo de las solicitudes de hombres y mujeres difieren sistemáticamente?				
¿Los solicitantes hombres y mujeres tienen la misma probabilidad de convertirse en clientes habituales de forma incondicional y condicional?				
Para asesoramiento financiero				
¿Los solicitantes hombres y mujeres reciben el mismo tipo de orientación, motivación o sugerencias (por ejemplo, una recomendación para utilizar tipos de productos específicos, o asignación de activos) después de tener en cuenta otros factores relevantes?				<ul style="list-style-type: none"> ↗ Estrategia ↗ Plan de negocios ↗ Plan de negocios ↗ Aplicaciones de cliente ↗ Entrevistas con funcionarios de la empresa
En los casos en los que un asesor de IA hace recomendaciones, pero un cliente puede decidir seguirlas o rechazarlas, ¿tienen los clientes hombres y mujeres una probabilidad similar de aceptar o rechazar las recomendaciones después de tener en cuenta otros factores relevantes?				<ul style="list-style-type: none"> ↗ Datos administrativos y estadísticas. ↗ Entrevistas con el equipo de ciencia de datos, funcionarios de préstamos o seguros
En los casos en que un asesor de IA toma decisiones automatizadas, ¿los solicitantes hombres y mujeres exhiben los mismos resultados o indicadores de desempeño después de tener en cuenta otros factores relevantes?				
Para ciberseguridad/detección de fraudes				
¿Las transacciones de hombres y mujeres tienen una probabilidad similar de ser marcadas correctamente como actividad fraudulenta?				
Explorar más				
<ul style="list-style-type: none"> ↗ Verifique su sesgo: una guía práctica para prestamistas ↗ Eliminación del sesgo de la IA en los seguros ↗ ORCAA: Consultoría de Riesgos y Auditoría Algorítmica 				

Recursos adicionales

- ↗ [Confrontando Sesgos: Marco Conceptual De BSA Para Construir Confianza En La Inteligencia Artificial](#)
- ↗ [Kit de recursos de Inteligencia Artificial Responsable de Nasscom](#)
- ↗ [Sobre votos y listas](#)
- ↗ [Checklist para la Inteligencia Artificial ética](#)
- ↗ [Checklist De Autoevaluación Para Una Inteligencia Artificial De Confianza](#)
- ↗ [Marco Conceptual de Datos Éticos](#)



Referencias

- i** Junta de Estabilidad Financiera, “Inteligencia artificial y machine learning en servicios financieros: evolución del mercado e implicaciones para la estabilidad financiera”, noviembre de 2017. <https://www.fsb.org/wp-content/uploads/P011117.pdf><https://www.fsb.org/wp-content/uploads/P011117.pdf>
- ii** [OECD.AI](https://oecd.ai/en/data?selectedArea=investments-in-ai-and-data&selectedVisualization=vc-investments-in-ai-by-industry), “Inversiones de capital de riesgo en IA por industria”, consultado el 6 de diciembre de 2022. <https://oecd.ai/en/data?selectedArea=investments-in-ai-and-data&selectedVisualization=vc-investments-in-ai-by-industry>
- iii** Sraders. Anne. “¿Cuánto valen realmente las startups de IA generativa? Los capitalistas de riesgo están pagando mientras luchan por participar en el próximo gran avance”. Fortune, abril de 2023. <https://fortune.com/2023/04/06/how-much-are-generative-ai-startups-worth-venture-capital/><https://fortune.com/2023/04/06/how-much-are-generative-ai-startups-worth-venture-capital/>
- iv** Biallas, Margarete y Felicity O’Neill. “Innovación en inteligencia artificial en servicios financieros”. Nota 85 de IFC, junio de 2020.
- v** Ibid.
- vi** Pazarbasioglu, Ceyla, Alfonso García Mora, Mahesh Uttamchandani, Harish Natarajan, Erik Feyen y Mathew Saal. “Servicios financieros digitales”. Grupo del Banco Mundial, abril de 2020. <https://pubdocs.worldbank.org/en/230281588169110691/Digital-Financial-Services.pdf>
- vii** Smith, Genevieve e Ishita Rustagi. “Cuando los buenos algoritmos se vuelven sexistas: por qué y cómo promover la equidad de género en la IA”. Stanford Social Innovation Review, marzo de 2021. <https://doi.org/10.48558/A179-B138>
- viii** NIST, “Marco de gestión de riesgos de inteligencia artificial”, enero de 2023. <https://doi.org/10.6028/NIST.AI.100-1>
- ix** Townson, Sian. “La IA puede hacer que los préstamos bancarios sean más justos”. Harvard Business Review, noviembre de 2020. <https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair>
- x** Smith, Genevieve e Ishita Rustagi. “Mitigación del sesgo en la inteligencia artificial: un manual de estrategias de liderazgo fluido y equitativo”. Centro Berkeley Haas para la Equidad, el Género y el Liderazgo, julio de 2020. <https://haas.berkeley.edu/equity/industry/playbooks/mitigating-bias-in-ai/>
- xi** Dastin, Jeffrey. “Amazon elimina una herramienta secreta de reclutamiento de IA que mostraba prejuicios contra las mujeres”. Reuters, octubre de 2018. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCNIMK08G>
- xii** Alianza Financiera para Mujeres, “Diseño fintech con inteligencia de género: cómo las fintech pueden captar la economía femenina”, 2021. <https://financialallianceforwomen.org/download/how-fintechs-can-capture-the-female-economy/>; Alianza Financiera para Mujeres, “Manual de datos de género para la inclusión financiera de las mujeres”, 2023. <https://financialallianceforwomen.org/download/the-gender-data-playbook-for-womens-financial-inclusion/>
- xiii** Alianza Financiera para Mujeres, “Diseño fintech con inteligencia de género: cómo las fintech pueden captar la economía femenina”, 2021. <https://financialallianceforwomen.org/download/how-fintechs-can-capture-the-female-economy/>; Alianza Financiera para Mujeres, “Manual de datos de género para la inclusión financiera de las mujeres”, 2023. <https://financialallianceforwomen.org/>

- [download/the-gender-data-playbook-for-womens-financial-inclusion/](#)
- xiv** Smith y Rustagi, “Mitigación del sesgo en la inteligencia artificial”.
- xv** DataRobot, “Estado del sesgo de la IA”, 2022. <https://www.datarobot.com/resources/state-of-ai-bias-report/thank-you/>
- xvi** Rizzi, Alexandra, Alexandra Kessler y Jacobo Menajovsky. “Las historias que cuentan los algoritmos: sesgo e inclusión financiera en los márgenes de los datos”. Centro para la Inclusión Financiera, marzo de 2021. <https://content.centerforfinancialinclusion.org/wp-content/uploads/sites/2/2021/03/The-Stories-Algorithms-Tell-CFI-publication-MAR21.pdf>
- xvii** The Economist Intelligence Unit, “Mantenerse a la vanguardia: el caso empresarial de una IA responsable”, octubre de 2020. <https://pages.eiu.com/rs/753-RIQ-438/images/EIUStayingAheadOfTheCurve.pdf>
- xviii** Reilly, Athena, Joe Depa y Greg Douglass. “IA: construida a escala”. Accenture, 2019. https://www.accenture.com/_acnmedia/Thought-Leadership-Assets/PDF-2/Accenture-Built-to-Scale-PDF-Report.pdf#zoom=50
- xix** Smith y Rustagi, “Mitigación del sesgo en la inteligencia artificial”.
- xx** Ibid.
- xxi** Feldstein, Steven. “Evaluando el impulso de Europa para promulgar regulaciones sobre IA: ¿cómo influirá esto en las normas globales?”+ Democratization, abril de 2023. <https://doi.org/10.1080/13510347.2023.2196068>
- xxii** Dotan, Ravit. “Inversión responsable en IA: una guía para capitalistas de riesgo”. VentureESG, diciembre de 2022. https://www.ravitdotan.com/_files/ugd/f83391_899994a46dfc4cfdb253cabbd3f841c7.pdf; Smith y Rustagi, “Mitigación del sesgo en la inteligencia artificial”.
- xxiii** Oficina de Protección Financiera del Consumidor, “CFPB toma medidas contra Hello Digit por mentir a los consumidores sobre su algoritmo de ahorro automatizado”, agosto de 2022. www.consumerfinance.gov/about-us/newsroom/cfpb-takes-action-against-hello-digit-for-lying-to-consumers-about-its-automated-savings-algorithm/
- xxiv** Banco Mundial de Mujeres, “¡Revise su sesgo! Una guía práctica para prestamistas”, diciembre de 2022. <https://www.womensworldbanking.org/insights/report-check-your-bias-a-field-guide-for-lenders/>
- xxv** Centro de Cambridge para Finanzas Alternativas y Foro Económico Mundial, “Transformar paradigmas: Una encuesta global sobre la IA en los servicios financieros”, enero de 2020. https://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf
- xxvi** Ibid.
- xxvii** Koide, Melissa. “IA para siempre: conocimientos de investigación de los servicios financieros”. Centro de Regulación y Mercados de Brookings, agosto de 2022. <https://www.brookings.edu/wp-content/uploads/2022/08/AI-for-good-Research-insights-from-financial-services-4.pdf>
- xxviii** Biallas y O’Neill, “Innovación en inteligencia artificial en servicios financieros”; OCDE, “Perspectivas empresariales y financieras de la OCDE 2021: IA en los negocios y las finanzas”, septiembre de 2021. <https://doi.org/10.1787/ba682899-en>
- xxix** Pazarbasioglu et al., “Servicios financieros digitales”.
- xxx** Foro Económico Mundial, “Empoderando el liderazgo en IA: Kit de herramientas de la alta dirección de IA”, enero de 2022. https://www3.weforum.org/docs/WEF_Empowering_AI_Leadership_2022.pdf; Boukherouaa, El Bachir, Khaled AlAjmi, José Deodoro, Aquiles Farias y Rangachary Ravikumar. “Impulsar la economía digital: oportunidades y riesgos de la inteligencia artificial en las finanzas”. FMI, octubre de 2021. <https://www.elibrary.imf.org/view/journals/087/2021/024/article-A001-en.xml>
- xxxi** Centro de Cambridge para Finanzas Alternativas y Foro Económico Mundial, “Transforming Paradigms”.

- xxxii** McWaters, “La nueva física de los servicios financieros: comprender cómo la inteligencia artificial está transformando el ecosistema financiero”.
- xxxiii** Buolamwini, Joy y Timnit Gebru. “Matices de género: Disparidades de exactitud interseccional en la clasificación comercial de sexos”, Conferencia del MIT sobre equidad, responsabilidad y transparencia, febrero de 2018. <https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/>
- xxxiv** McWaters, “La nueva física de los servicios financieros”.
- xxxv** Centro de Cambridge para Finanzas Alternativas y Foro Económico Mundial, “Transforming Paradigms”.
- xxxvi** Ibid.
- xxxvii** McWaters, “La nueva física de los servicios financieros”.
- xxxviii** Parlamento Europeo, “Ley de IA: un paso más hacia las primeras normas sobre inteligencia artificial”, mayo de 2023. <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>
- xxxix** Leslie, David. “Comprensión de la ética y la seguridad de la inteligencia artificial: una guía para el diseño e implementación responsables de sistemas de IA en el sector público”. Instituto Alan Turing, junio de 2019. https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf
- xl** Rizzi et al., “Las historias que cuentan los algoritmos”.
- xli** Demirgüç-Kunt, Asli et al. “La base de datos global Findex: inclusión financiera, pagos digitales y resiliencia en la era de COVID-19”. Banco Mundial, 2022. <https://www.worldbank.org/en/publication/globalindex>; Chen, S. y col. “La brecha de género en las fintech”. Revista de Intermediación Financiera, vol. 54, abril de 2023. <https://www.sciencedirect.com/science/article/pii/S1042957323000098#bibl>
- xlii** Floridi, Luciano et al. “AI4People: un marco ético para una buena sociedad de IA: oportunidades, riesgos, principios y recomendaciones”. Minds & Machines, noviembre de 2018. <https://link.springer.com/article/10.1007/s11023-018-9482-5>; Floridi, Luciano y Josh Cowls. “Un marco unificado de cinco principios para la IA en la sociedad”. Harvard Data Science Review, 1(1), julio de 2019. <https://hdsr.mitpress.mit.edu/pub/10jsh9dl/release/8>
- xliii** ICO, “Orientación sobre el marco de auditoría de IA: Borrador de orientación para consulta”, febrero de 2020. <https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf>; Ethical OS, “Una guía para anticipar el impacto futuro de la tecnología actual”, agosto de 2018. <https://ethicalos.org/>
- xliv** Adkins, David y cols. “Tarjetas de método para la transparencia prescriptiva del machine learning”. 2022 IEEE/ACM 1.^a Conferencia Internacional sobre Ingeniería de IA, mayo de 2022. <https://ieeexplore.ieee.org/document/9796452>
- xlv** Proyecto Data Nutrition, consultado en junio de 2023. <https://datanutrition.org/>
- xlvi** Danks, David y Alex John London. “Sesgo algorítmico en sistemas autónomos”. 26a Conferencia Internacional Conjunta sobre Inteligencia Artificial, agosto de 2017. <https://www.cmu.edu/dietrich/philosophy/docs/london/IJCAI17-AlgorithmicBias-Distrib.pdf>; Fazelpour, Sina y David Danks. “Sesgo algorítmico: sentidos, fuentes, soluciones”. Philosophy Compass, vol. 16, número 8, agosto de 2021. <https://compass.onlinelibrary.wiley.com/doi/full/10.1111/phc3.12760>
- xlvii** Mirpourian, Mehrdad. “Un conjunto de herramientas de Python para auditar y detectar sesgos de género en los procesos de evaluación crediticia”. Repositorio de Github, enero de 2023. <https://github.com/WomensWorldBanking/A-Python-toolkit-for-auditing-and-detecting-gender-bias-in-credit-assessment-processes>
- xlviii** FinRegLab, “Explicabilidad y equidad en el machine learning para la suscripción de créditos: descripción general de políticas y hallazgos empíricos”, julio de 2023. https://finreglab.org/wp-content/uploads/2023/07/FRL_ML-FindingsOverview_Final.pdf

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