



A3

Ejemplos de aplicación de la metodología desarrollada en este manual

Dos casos completos: matriz de relaciones funcionales, cadenas causales y estrategias de mejora.

01 Uso de IA para apoyar decisiones judiciales sobre medidas cautelares

02 Uso de IA en un sistema de calificación crediticia con enfoque de género

Ejemplos de aplicación de la metodología desarrollada en este manual

Este anexo contiene dos ejemplos de aplicación de la metodología propuesta.

Los mapas sistémicos (la representación gráfica en red de las matrices de relaciones funcionales de cada ejemplo) son piezas visuales que este anexo aún no incluye como diagrama. Las matrices de relaciones y todo el análisis que se deriva de ellas sí se presentan completos a continuación.

EJEMPLO 1 Medidas cautelares

- 1.1 La matriz de relaciones funcionales
- 1.2 Explicación de las cadenas funcionales
- 1.3 Fundamentos del análisis sistémico: ámbitos y rutas
- 1.3.5 Estrategias de mejora

EJEMPLO 2 Calificación crediticia con enfoque de género

- 2.1 Relaciones funcionales por ámbito
- 2.2 Fundamentos del análisis sistémico: ámbitos y rutas
- 2.2.5 Estrategias de mejora
- 2.2.6 Cómo combinar las estrategias



EJEMPLO 1

Uso de IA para apoyar decisiones judiciales sobre medidas cautelares

Este ejemplo aplica la metodología a un caso en el que se usa un sistema de IA como apoyo a decisiones judiciales sobre medidas cautelares. Fiel al pensamiento sistémico, el método no se aplica al sistema judicial, o una de sus partes, como un todo indiferenciado, sino a los ámbitos y las cadenas del mapa sistémico elaborado expreso, y en función de los roles de sus elementos y de los resultados de las evaluaciones de riesgos e impactos a la población.

1.1

La matriz de relaciones funcionales

El primer paso es la construcción de la matriz de relaciones funcionales del sistema objetivo, la cual es una matriz simétrica formada por los elementos que conforman el sistema y que sirve de base para el despliegue del mapa sistémico. Cada flecha del mapa (aún no ilustrado) representa una relación funcional dirigida (causa → efecto): el signo (+) indica refuerzo, el signo (–) compensación.

Significado de las siglas que representan los nodos en el mapa sistémico

IOP Intensidad de lo oscuro de la piel	NS Nivel socioeconómico de la población objetivo
PA Probabilidad de arresto	DRSJ Disponibilidad de recursos en sistema judicial
GCL Gradiente visión conservadora-liberal en sistema judicial	CT Carga de trabajo en juzgados
IAJ Influencia algorítmica en jueces	DAOC Decisiones algorítmicas (basadas en recomendación de IA) en otros circuitos
NR Nivel de riesgo asignado al imputado	PP Probabilidad prisión preventiva
EPSI Efectos psicoemocionales negativos sobre imputado	PD Probabilidad desempleo futuro del imputado
ERN Efectos reputacionales negativos sobre imputado	DEP Daño por efectos procesales
DF Daño a familiares del imputado	DI Daños intergeneracionales

TABLA A3.1

Matriz de relaciones funcionales entre elementos del subsistema judicial de medidas cautelares

	IOP	NS	PA	DRSJ	GCL	CT	IAJ	DAOC	NR	PP	EPSI	PD	ERN	DEP	DF	DI
IOP	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
NS	—	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
PA	+	—	—	NO	+	NO	+	+	NO	NO	NO	NO	NO	NO	NO	NO
DRSJ	NO	NO	NO	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
GCL /	NO	NO	NO	NO	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
CT	NO	NO	NO	—	NO	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
IAJ	NO	NO	NO	NO	NO	NO	—	+	NO	NO	NO	NO	NO	NO	NO	NO
DAOC	NO	NO	NO	NO	NO	NO	NO	—	NO	NO	NO	NO	NO	NO	NO	NO
NR	NO	NO	NO	—	+	+	+	+	—	NO	NO	NO	NO	NO	NO	NO
PP	+	—	+	—	+	+	+	+	+	—	NO	NO	NO	NO	NO	NO
EPSI	NO	NO	+	NO	NO	NO	NO	NO	+	+	—	+	+	+	+	+
PD	NO	NO	NO	NO	NO	NO	NO	NO	+	+	+	—	+	+	+	+
ERN	NO	NO	NO	NO	NO	NO	NO	NO	+	+	+	+	—	+	NO	NO
DEP	NO	—	+	NO	NO	+	+	+	+	+	+	+	+	—	+	+
DF	NO	—	+	NO	NO	NO	NO	NO	NO	+	+	+	+	+	—	+
DI	NO	—	+	NO	NO	NO	NO	NO	NO	+	+	+	+	+	+	—

Fuente: Elaboración propia, metodología interna de Centro-i.

1.2

Explicación de las cadenas funcionales

A continuación se describen las relaciones funcionales registradas en la matriz, organizadas por ámbito. La cadena algebraica indica el tipo de relación: el signo (+) señala que un incremento del emisor incrementa al receptor; el signo (–) señala que lo reduce.

Ámbito 1 — Causas estructurales

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
IOP / Intensidad de lo oscuro de la piel	NS / Nivel socioeconómico	(IOP) - (NS)	En muchos países se hacen patentes desigualdades histórico-estructurales de origen étnico, asociadas por ejemplo al color de la piel. En este ejemplo se asume que a mayor intensidad de la piel oscura, menor nivel socioeconómico en promedio.
IOP / Intensidad de lo oscuro de la piel	PA / Probabilidad de arresto / PP / Probabilidad de prisión preventiva	(IOP) + (PA) + (PP)	A mayor intensidad de piel oscura, mayor probabilidad de arresto —la sobrevigilancia policial sobre ciertos grupos genera más detenciones independientemente de la conducta real— y mayor probabilidad de prisión preventiva, arrastrando el sesgo estructural hasta la decisión cautelar.
NS / Nivel socioeconómico	PA / Probabilidad de arresto / PP / Probabilidad de prisión preventiva	(NS) - (PA) - (PP)	A mayor nivel socioeconómico, menor probabilidad de arresto —los recursos y el entorno reducen la exposición a la vigilancia y la detención— y menor prisión preventiva, por mayor acceso a defensa de calidad y a alternativas (fianza, arraigo).
NS / Nivel socioeconómico	DEP / Daño por efectos procesales / DF / Daño a familiares del imputado / DI / Daños intergeneracionales	(NS) - (DEP) - (DF) - (DI)	Un mayor nivel socioeconómico amortigua el daño procesal, familiar e intergeneracional: hay más recursos y capacidades para absorber el golpe económico y social.

Ámbito 2 — Factores del sistema judicial

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
DAOC / Decisiones algorítmicas en otros circuitos	IAJ / Influencia algorítmica en jueces / PA / Probabilidad de arresto / NR / Nivel de riesgo del imputado / PP / Probabilidad de prisión preventiva	(DAOC) + (IAJ) + (PA) + (NR) + (PP)	Cuanto más decisiones algorítmicas existan en otros circuitos, mayor es la influencia algorítmica sobre los jueces (normalización del uso de la herramienta) y mayor el efecto de imitación institucional, lo que genera decisiones inerciales a favor de niveles altos de riesgo asignado y decretos más frecuentes de prisión preventiva.
GCL / Gradiente de visión conservadora-liberal en el sistema judicial	PA / Probabilidad de arresto / NR / Nivel de riesgo del imputado / PP / Probabilidad de prisión preventiva	(GCL) + (PA) + (NR) + (PP)	Un gradiente cargado hacia lo conservador en el sistema judicial aumenta la probabilidad de arresto, el nivel de riesgo asignado y la prisión preventiva.
CT / Carga de trabajo en juzgados	NR / Nivel de riesgo del imputado / PP / Probabilidad de prisión preventiva	(CT) + (NR) + (PP)	A mayor carga de trabajo en los juzgados, mayor riesgo asignado y prisión preventiva: bajo presión de tiempo, los jueces descansan más en el puntaje algorítmico.
IAJ / Influencia algorítmica en jueces	NR / Nivel de riesgo del imputado / PP / Probabilidad de prisión preventiva	(IAJ) + (NR) + (PP)	Como resultado del arrastre de sesgos de las bases de datos de entrenamiento, a mayor influencia algorítmica los jueces acaban aceptando con mayor frecuencia niveles altos de riesgo de los imputados y recomendando más veces prisión preventiva.
DRSJ Disponibilidad de recursos en el sistema judicial	NR / Nivel de riesgo del imputado / PP / Probabilidad de prisión preventiva / CT / Carga de trabajo en juzgados	(DRSJ) - (NR) - (PP) - (CT)	A mayor disponibilidad de recursos judiciales, menor nivel de riesgo asignado — hay más capacidad de valorar cada caso individualmente — y menor prisión preventiva, al existir más alternativas y capacidad de supervisión en libertad. Más recursos judiciales también reducen la carga de trabajo por juzgado.

Ámbito 3 — Decisión

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
NR / Nivel de riesgo del imputado	PP / Probabilidad de prisión preventiva	(NR) + (PP)	A mayor nivel de riesgo asignado por el sistema, mayor probabilidad de prisión preventiva: es la conexión directa entre el cálculo algorítmico y la decisión.

Ámbito 3 — Decisión (continuación)

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
PA / Probabilidad de arresto	PP / Probabilidad de prisión preventiva	(PA) + (PP)	La probabilidad de arresto eleva la prisión preventiva: quien es detenido entra al circuito donde se decide la medida cautelar.

Ámbito 4 — Efectos sobre el imputado

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
PA / Probabilidad de arresto	EPSI / Efectos psicoemocionales sobre el imputado / DEP / Daño por efectos procesales / DF / Daño a familiares del imputado / DI / Daños intergeneracionales	(PA) + (EPSI) + (DEP) + (DF) + (DI)	Un incremento en la probabilidad de arresto desencadena directamente daños psicoemocionales, procesales, familiares e intergeneracionales, aun antes de la sentencia.
NR / Nivel de riesgo del imputado	EPSI / Efectos psicoemocionales sobre el imputado / PD / Probabilidad de desempleo futuro del imputado / ERN / Efectos reputacionales sobre el imputado / DEP / Daño por efectos procesales	(NR) + (EPSI) + (PD) + (ERN) + (DEP)	El nivel de riesgo asignado incrementa el daño psicoemocional, el desempleo futuro, el daño reputacional y el procesal: la etiqueta de 'riesgo alto' estigmatiza y condiciona.
PP / Probabilidad de prisión preventiva	EPSI / Efectos psicoemocionales sobre el imputado / PD / Probabilidad de desempleo futuro del imputado / ERN / Efectos reputacionales sobre el imputado / DEP / Daño por efectos procesales	(PP) + (EPSI) + (PD) + (ERN) + (DEP)	La prisión preventiva amplifica todos los daños sobre el imputado: el encierro previo al juicio deteriora la salud mental, el empleo, la reputación y la capacidad de defensa.
EPSI / Efectos psicoemocionales sobre el imputado	PD / Probabilidad de desempleo futuro del imputado / ERN / Efectos reputacionales sobre el imputado / DEP / Daño por efectos procesales	(EPSI) + (PD) + (ERN) + (DEP)	Los daños sobre el imputado se retroalimentan mutuamente: el deterioro psicoemocional, el desempleo, el daño reputacional y el procesal se refuerzan entre sí formando bucles de agravamiento.

Ámbito 5 — Efectos sobre familia y comunidad

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
PP / Probabilidad de prisión preventiva	DF / Daño a familiares del imputado / DI / Daños intergeneracionales	(PP) + (DF) + (DI)	La prisión preventiva daña a la familia (pérdida de ingreso, separación) y produce efectos intergeneracionales (impactos psicoemocionales sobre hijas e hijos).
ERN / Efectos reputacionales sobre el imputado	DF / Daño a familiares del imputado / DI / Daños intergeneracionales	(ERN) + (DF) + (DI)	El daño reputacional del imputado se transmite a la familia por estigma de asociación y alcanza a la siguiente generación.
EPSI / Efectos psicoemocionales sobre el imputado	DF / Daño a familiares del imputado / DI / Daños intergeneracionales	(EPSI) + (DF) + (DI)	El daño psicoemocional del imputado se propaga a la familia y a los hijos, amplificando la desventaja.
PD / Probabilidad de desempleo futuro del imputado	DF / Daño a familiares del imputado / DI / Daños intergeneracionales	(PD) + (DF) + (DI)	El desempleo futuro del imputado se propaga a la familia y a los hijos, amplificando la desventaja económica.
DEP / Daño por efectos procesales	DF / Daño a familiares del imputado / DI / Daños intergeneracionales	(DEP) + (DF) + (DI)	El daño por efectos procesales del imputado se propaga a la familia y a los hijos, amplificando la desventaja.
DF / Daño a familiares del imputado	DI / Daños intergeneracionales	(DF) + (DI)	El daño a la familia y el daño intergeneracional se refuerzan mutuamente: el deterioro del núcleo familiar agrava las perspectivas de la siguiente generación.

En conjunto, las cadenas funcionales describen un sistema en el que un sesgo estructural de origen (IOP, NS) se propaga a través de los factores judiciales y la decisión cautelar, y termina amplificándose en una red de daños sobre el imputado, su familia y las generaciones siguientes.

1.3

Fundamentos del análisis sistémico: ámbitos y rutas

El mapa sistémico admite dos formas de análisis complementarias. La primera divide el sistema en ámbitos según la función causal de sus nodos; la segunda sigue rutas concretas que atraviesan esos ámbitos para diseñar estrategias de intervención. Los dos criterios son complementarios y se usan en secuencia: primero el análisis por ámbitos organiza el sistema y asigna las familias de instrumentos; después el análisis por rutas selecciona los eslabones concretos donde actuar.

TABLA 1.1

Complementariedad del análisis por ámbitos y del análisis por rutas

CRITERIO DE COMPARACIÓN	ANÁLISIS POR ÁMBITOS	ANÁLISIS POR RUTAS
Pregunta que responde	¿Qué instrumentos gobiernan cada zona del sistema?	¿En qué eslabón o nodo concreto conviene actuar para cortar la propagación del daño?
Unidad de análisis	Grupos de nodos con función causal homogénea.	Secuencias de nodos conectados que atraviesan varios grupos.
Qué revela	La estructura del sistema: origen, mediación, decisión y destino del daño.	Los efectos indirectos /de segundo orden) y los puntos de palanca a lo largo de un recorrido.
Conexión con la metodología	Asigna a cada ámbito su cuadrante en la matriz impacto-certidumbre y sus instrumentos.	Selecciona, dentro de ese marco, los eslabones donde aplicar la acción concreta.
Tipo de resultado	Un mapa de gobernabilidad: qué se regula y con qué familia de instrumentos.	Una estrategia de intervención: dónde actuar y en qué orden para reducir riesgo e impacto.
Principal ventaja	Evita aplicar el instrumento equivocado en el lugar equivocado.	Permite cortar cadenas completas de daño interviniendo en un solo eslabón clave.
Cuándo se usa	Primero: para organizar el sistema y enmarcar las opciones de regulación.	Después: para diseñar y priorizar las acciones concretas dentro de ese marco.

Fuente: Elaboración propia.

El siguiente análisis identifica, para cada ámbito, la vulnerabilidad y los instrumentos que le corresponden; donde la certidumbre es suficiente, se ofrece una estimación numérica de riesgo e impacto conforme al método V-R-I (ver Anexo 2).

Ámbito de causas estructurales

- Las cadenas IOP → PA (+) y NS → PP (–) muestran que la vulnerabilidad estructural condiciona la exposición antes de que el algoritmo intervenga.
- V se estima alta: la población afectada combina marcadores de susceptibilidad socioeconómica y de discriminación.
- Esta sección alimenta la estimación de V para las demás secciones, pero su tratamiento es de política pública, no de regulación de IA.

Ámbito de factores judiciales

- Corregulación con el órgano de administración judicial para fijar estándares de uso de la herramienta.
- Auditoría obligatoria de los efectos de la carga de trabajo y la influencia algorítmica sobre las decisiones.
- Protocolos adaptativos para ajustar el uso conforme se acumula evidencia.
- La cadena DRSJ → PP (–) y DRSJ → CT (–) señala una palanca positiva: invertir en recursos judiciales reduce la prisión preventiva y la saturación.

Ámbito de decisión y bucle de sesgo

- Vulnerabilidad V = 5 (muy alta): la decisión recae sobre personas en situación de vulnerabilidad estructural heredada del ámbito de causas estructurales.
- Probabilidad P = 4 (alta): las flechas que entran a PP (NR, GCL, CT, IAJ, DAOC) elevan la probabilidad de prisión preventiva; solo DRSJ la reduce.
- Riesgo R = $P \times V = 4 \times 5 = 20$ (crítico en la matriz de riesgo).
- Personas expuestas N = 4 (amplio): decenas de miles de personas imputadas en el ámbito de despliegue.
- Impacto I = $R \times N = 20 \times 4 = 80$ → severidad crítica. Confirma el cuadrante D y la necesidad de salvaguardas máximas.
- Instrumentos: sandbox estricto con participación del Poder Judicial, criterios de parada explícitos, SupTech con acceso del regulador en tiempo real, evaluación de impacto previa obligatoria.
- Condición indispensable: la decisión judicial es siempre humana; el sistema solo debe usarse como apoyo.

Ámbito efectos sobre el imputado

- Vulnerabilidad $V = 5$ (muy alta): los efectos recaen directamente sobre la persona imputada y se amplifican por los bucles de refuerzo entre los daños.
- Probabilidad $P = 4$ (alta): una vez decidida la prisión preventiva, los daños psicoemocionales, laborales, reputacionales y procesales son altamente probables.
- Riesgo $R = 4 \times 5 = 20$ (crítico).
- Personas expuestas $N = 4$ (amplio): el mismo orden de magnitud que la población imputada afectada.
- Impacto $I = 20 \times 4 = 80 \rightarrow$ severidad crítica. Requiere mecanismos de remedio y monitoreo continuo de equidad.

Ámbito de efectos sobre la familia

- Vulnerabilidad $V = 5$ (muy alta): incluye a hijas e hijos menores y dependientes, grupos altamente vulnerables.
- Probabilidad $P = 3$ (media): no toda prisión preventiva produce daño intergeneracional grave, pero ocurre con frecuencia apreciable.
- Riesgo $R = 3 \times 5 = 15$ (muy alto).
- Personas expuestas $N = 5$ (masivo): el número se multiplica al pasar del imputado a su familia y a la siguiente generación.
- Impacto $I = 15 \times 5 = 75 \rightarrow$ severidad alta, en el límite con crítica. La N elevada compensa la menor probabilidad.

Tabla 1.3 · Resumen de instrumentos sugeridos por el análisis de ámbitos funcionales

ÁMBITO	IMPACTO	CERTIDUMBRE	INSTRUMENTOS SUGERIDOS
Causas estructurales (IOP, NS)	Alto	Alta	Fuera del alcance de la gobernanza de IA. Corresponde a política pública estructural (antidiscriminación, reducción de desigualdad). Aquí se identifica la vulnerabilidad (V) que alimenta las demás secciones.
Factores judiciales (DRSJ, GCL, CT, IAJ, DAOC)	Alto	Alta	Corregulación con el órgano de administración judicial · Auditoría obligatoria · Protocolos adaptativos · Inversión en recursos judiciales (palanca DRSJ).
Decisión y bucle de sesgo (NR, PP)	Alto	Baja	Sandbox estricto con el Poder Judicial · Criterios de parada · SupTech en tiempo real · EIDF previa obligatoria · Decisión humana irrenunciable.
Efectos sobre el imputado (EPSI, PD, ERN, DEP)...	Alto	Media	Mecanismos de remedio accesibles · Monitoreo continuo de equidad · Transparencia hacia las personas afectadas · Vigilancia de los bucles de refuerzo.
y su familia (DF, DI)	Alto	Baja	Gobernanza anticipatoria · Horizon scanning · Evaluación longitudinal de efectos intergeneracionales.

1.3.5

Estrategias de mejora: acciones a lo largo de cada cadena

Cada cadena del mapa se convierte en una estrategia de mejora cuando se define un menú de acciones aplicables en cada uno de sus nodos.

Estrategia 1 · Cadena socioeconómica del daño

Recorre la propagación desde el bajo nivel socioeconómico hasta el daño intergeneracional. Como su origen (NS) está fuera del alcance de la gobernanza de IA, la estrategia concentra sus acciones en los eslabones intermedios y finales.

	NODO 1	NODO 2	NODO 3	NODO 4	NODO 5	NODO 6
Acciones o medidas	Esta es una variable estructural que pertenece al ámbito de la política económica y social, por lo que escapa a las acciones del modelo.	Monitoreo automatizado continuo del sesgo de muestreo policial con umbrales de disparo predefinidos (SupTech). Al superar el umbral, se activa automáticamente el ciclo de corrección: revisión de variables proxy → reentrenamiento parcial → validación de equidad → reimplementación. Habría que establecer casuísticamente un periodo máximo de cumplimiento del ciclo.	Criterios de parada dinámicos ajustables por umbral sin reforma legislativa (vía protocolo del órgano de administración judicial). Revisión humana obligatoria con motivación explícita cuando el puntaje supera el umbral o cuando el imputado pertenece a un grupo con desigualdad documentada.	Indicador de seguimiento de la duración del proceso desagregado por tipo de decisión algorítmica, para medir si el uso del puntaje acorta o alarga el DEP y ajustar el uso en consecuencia. Este indicador es insumo del ciclo de evaluación del sandbox.	Mecanismo de remedio accesible vinculado a la decisión algorítmica: cuando se revoca una medida cautelar fundada en el puntaje, se activa automáticamente la elegibilidad para apoyo económico y psicosocial familiar. El vínculo entre revocación y activación del apoyo no requiere ley; basta un protocolo interinstitucional.	Evaluación longitudinal del impacto intergeneracional con cohortes anuales (indicador de seguimiento de largo plazo). Los resultados retroalimentan la revisión del modelo: si el daño intergeneracional es medible y atribuible al puntaje, activa el ciclo de recalibración aguas arriba.
Recursos necesarios	Altos	Medios	Altos	Medios	Medios	Medios-altos
Dificultad	Alta	Media	Alta	Media	Baja	Media

Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.

Estrategia 2 · Cadena de automatización algorítmica

Recorre la propagación desde la normalización del uso de algoritmos hasta el daño psicoemocional. Sus acciones se concentran en limitar la deferencia del juez al puntaje, donde la intervención es más barata y eficaz.

	NODO 1	NODO 2	NODO 3	NODO 4	NODO 5
Acciones o medidas	Sandbox de uso restringido del algoritmo en circuitos piloto seleccionados, con evaluación comparativa (con/sin puntaje) antes de expandir. Revisión de precedentes algorítmicos como mecanismo de aprendizaje judicial iterativo.	Diseñar el procedimiento de manera que la/el juez no vea el puntaje algorítmico antes de revisar el expediente, para evitar fenómeno de anclaje cognitivo. Establecer ciclo de aprendizaje institucional mediante experimentos controlados de presentación (pruebas A/B), para identificar el formato que más reduce el anclaje y que genera mejores resultados.	Validación técnica periódica por un organismo permanente con capacidad de actualización continua. Explicabilidad obligatoria del puntaje en lenguaje comprensible para la/el juez y para la persona imputada. Frecuencia mínima de revalidación: cada vez que se actualice el modelo o anualmente, lo que ocurra primero.	Criterios de paro ajustables vía protocolo (sin reforma legal). Decisión humana irrenunciable como principio de responsabilidad clara y no delegable. Segunda opinión obligatoria en casos límite ($\pm X$ puntos del umbral de decisión), con el umbral X revisable mediante evidencia acumulada.	Recurso administrativo expedito ante el órgano de administración judicial, activado específicamente cuando la decisión se basó en el puntaje algorítmico — distinto del proceso judicial ordinario. Plazo de resolución máximo fijado en el protocolo. Los resultados del recurso alimentan el ciclo de mejora del modelo.
Recursos necesarios	Medios	Medios	Medios-altos	Altos	Medios
Dificultad	Media	Baja-media	Media-alta	Alta	Baja

Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.

Estrategia 3 · Cadena de sesgo estructural

Recorre el sesgo desde su origen estructural hasta el estigma reputacional y familiar. Como IOP no es modificable, la estrategia ataca el punto donde el sesgo entra al modelo y se traslada a la decisión.

	NODO 1	NODO 2	NODO 3	NODO 4	NODO 5
Acciones o medidas	Fuera del alcance de la gobernanza de IA. Acción ágil alcanzable: documentar y publicar periódicamente la correlación empírica entre la variable IOP (u otras correlacionadas) y el puntaje asignado por el modelo. Esta transparencia activa la presión institucional y social sobre el modelo sin requerir reforma legislativa.	Ciclo automatizado de monitoreo-disparo-corrección para incidir sobre probabilidad de arresto. Por ejemplo, crear indicador de disparidad por grupo definido de antemano (ej. diferencial de puntaje medio por grupo ≥ X%) con umbral que avisa para revisar aumento desproporcionado de arrestos por grupo.	Criterios de paro sensibles a disparidad por grupo, ajustables conforme cambia la distribución de los datos (proporcionalidad dinámica). Revisión humana reforzada en casos de alto riesgo: la revisión es automática, no discrecional, cuando el puntaje del imputado pertenece a un grupo con disparidad documentada.	Protocolo de acceso restringido al puntaje implementable por vía administrativa del poder judicial (sin reforma legal): define quiénes pueden consultar el puntaje y con qué fines. Prohibición de uso del puntaje en procesos distintos al penal. Tiempo de retención del puntaje en el expediente: máximo definido en el protocolo, renovable con justificación.	Prohibición administrativa de que el puntaje del imputado sea consultado en trámites de bienestar social, escuelas o empleo.
Recursos necesarios	Altos	Medios-altos	Altos	Medios	Medios
Dificultad	Alta	Media-alta	Alta	Media	Media

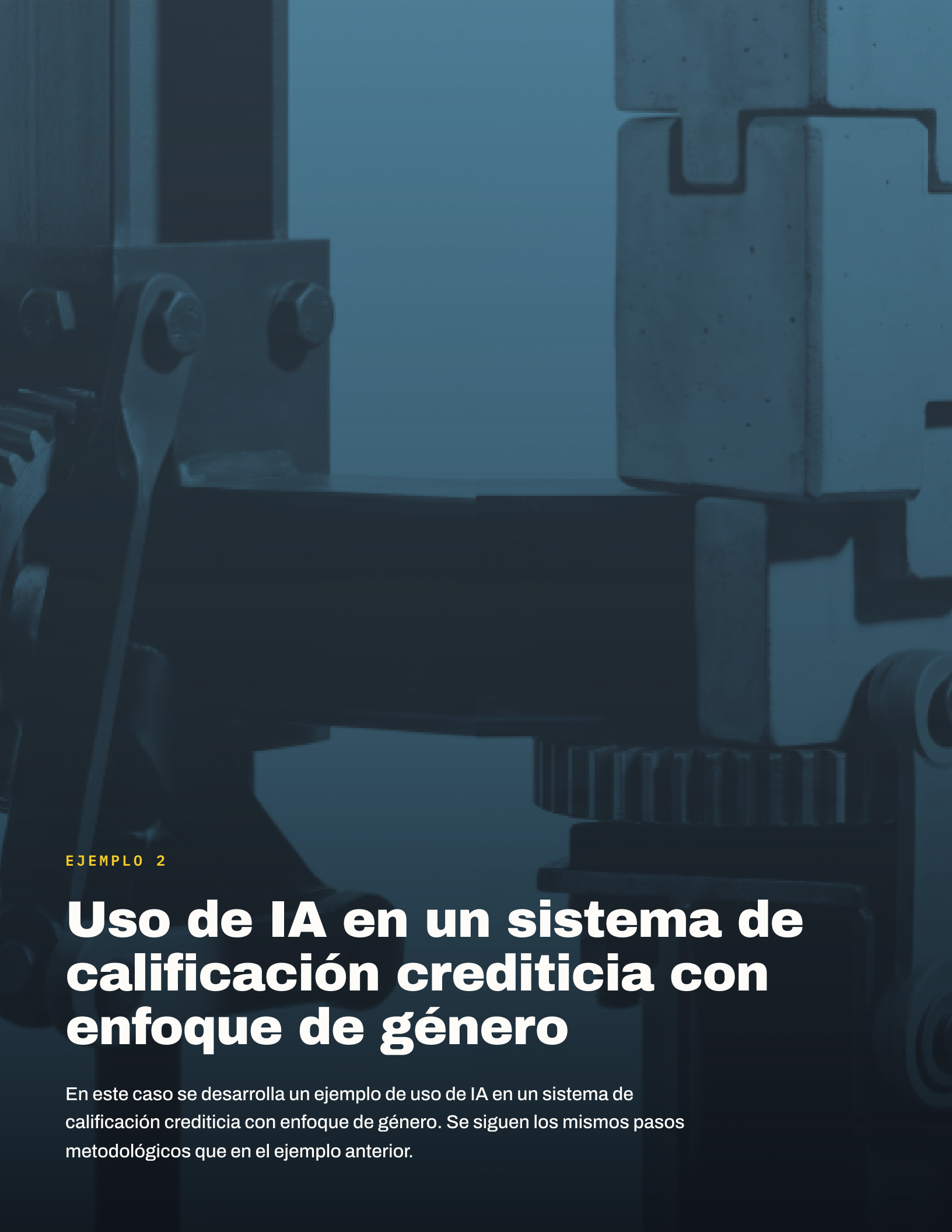
Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.

Estrategia 4 · Cadena de saturación judicial

Recorre cómo la falta de recursos del sistema judicial, vía la saturación de los juzgados, infla la prisión preventiva y el daño procesal. La acción de mayor palanca está en el origen de la cadena.

	NODO 1	NODO 2	NODO 3	NODO 4
Acciones o medidas	Usar los datos del propio sistema de IA para identificar los cuellos de botella específicos: ¿en qué etapa se acumula el rezago?, ¿en qué circuitos?, ¿para qué tipos de casos? Intervenir con evidencia, no con inversión general. Para hacer que el algoritmo que contribuye al problema se convierta en un insumo que ayude a la solución, hay que hacerlo parte del diagnóstico.	La administración judicial debe estar abierta a la reingeniería de procesos y al monitoreo de la carga de trabajo de los juzgados en tiempo real, por ejemplo, para aprovechar las ventajas de la regulación ágil. El monitoreo en tiempo real visibiliza cuándo una/un juez tiene sobrecarga para activar protocolos de redistribución x como instrumento de regulación adaptativa. Digitalización de trámites para descongestionar. Piloto de redistribución automática de casos en circuitos con datos suficientes antes de expansión generalizada.	Criterios de paro sensibles a disparidad por grupo, ajustables conforme cambia la distribución de los datos (proporcionalidad dinámica). Revisión humana reforzada en casos de alto riesgo: la revisión es automática, no discrecional, cuando el puntaje del imputado pertenece a un grupo con disparidad documentada.	Indicador de duración del proceso desagregado por tipo de decisión algorítmica: medir si el uso del puntaje acorta o alarga el DEP en promedio y por circuito. Este indicador es insumo del ciclo de evaluación del sandbox y del ajuste del protocolo de uso del puntaje.
Recursos necesarios	Muy altos	Altos	Altos	Medios
Dificultad	Alta	Media-alta	Alta	Media

Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.



EJEMPLO 2

Uso de IA en un sistema de calificación crediticia con enfoque de género

En este caso se desarrolla un ejemplo de uso de IA en un sistema de calificación crediticia con enfoque de género. Se siguen los mismos pasos metodológicos que en el ejemplo anterior.

TABLA A3.2

Matriz de relaciones funcionales del sistema de IA de calificación de crédito con enfoque de género

	PSDH	UVPG	VEA	FCM	CCM	CPA	PCE	PCR	PAC	NCA	PC	DHCM	DSR	IF	SERC	SSG	CAC	PIC	BEFI	EGO	TRYE	MEPGA	PSE	RAP
PSDH	—	NO	–	NO	NO	NO	NO	–	NO	NO	NO	+	NO	NO	NO	+	NO	NO	NO	NO	NO	NO	NO	NO
UVPG	NO	—	–	NO	NO	NO	NO	NO	NO	NO	NO	NO	+	NO	+	+	–	+	NO	NO	NO	NO	NO	NO
VEA	NO	NO	—	NO	NO	NO	NO	+	NO	NO	NO	NO	NO	NO	–	NO	NO	NO	NO	NO	NO	NO	NO	NO
FCM	NO	NO	+	—	NO	NO	NO	NO	NO	NO	NO	NO	–	NO	–	–	+	NO	NO	NO	NO	NO	NO	NO
CCM	NO	NO	+	NO	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	–	–	NO	NO	NO	NO	NO	NO	NO	NO
CPA	NO	NO	NO	+	+	—	+	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	+	NO	NO	+
PCE	NO	NO	+	NO	NO	+	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
PCR	NO	NO	NO	NO	NO	NO	NO	—	+	NO	–	NO	NO	NO	NO	NO	NO	NO	+	NO	NO	NO	NO	NO
PAC	NO	NO	NO	NO	NO	NO	NO	NO	—	+	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
NCA	–	NO	NO	NO	NO	NO	NO	NO	NO	—	NO	NO	NO	NO	NO	NO	+	NO	+	NO	NO	NO	NO	NO
PC	NO	NO	NO	NO	NO	NO	NO	NO	NO	–	—	NO	NO	NO	NO	NO	NO	NO	–	NO	NO	NO	NO	NO
DHCM	NO	NO	NO	NO	NO	NO	NO	–	NO	–	NO	—	NO	–	NO	NO	NO	NO	–	NO	NO	NO	NO	NO
DSR	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	—	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
IF	NO	NO	NO	NO	NO	NO	NO	NO	NO	+	NO	–	NO	—	NO	–	+	NO	+	NO	NO	NO	NO	NO
SERC	NO	NO	NO	NO	NO	NO	NO	–	NO	NO	NO	NO	NO	NO	—	+	NO	NO	NO	NO	NO	+	NO	NO
SSG	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	+	NO	–	NO	—	NO	NO	NO	NO	NO	+	NO	NO
CAC	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	—	NO	NO	NO	NO	NO	NO	NO
PIC	NO	NO	NO	NO	NO	NO	NO	NO	–	NO	NO	NO	NO	NO	NO	NO	–	—	NO	NO	NO	NO	NO	NO
BEFI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	—	NO	NO	NO	NO	NO
EGO	NO	NO	NO	+	+	+	+	NO	NO	NO	NO	NO	NO	NO	–	–	+	NO	NO	—	+	NO	+	+
TRYE	NO	NO	NO	NO	NO	NO	NO	+	NO	NO	NO	NO	NO	NO	–	NO	+	NO	NO	NO	—	NO	NO	NO
MEPGA	NO	NO	NO	+	NO	+	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	+	NO	—	+	NO
PSE	NO	NO	+	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	–	–	+	NO	NO	NO	NO	–	—	NO
RAP	NO	NO	NO	NO	NO	NO	+	NO	NO	NO	NO	NO	NO	NO	NO	–	NO	NO	NO	+	+	NO	NO	—

Fuente: Elaboración propia, metodología interna de Centro-i.

Leyenda de siglas y variables

PSDH Permanencia de sesgos en datos históricos — A — Datos y Modelo	UVPG Uso de variables proxy de género — A — Datos y Modelo
VEA Validez del entrenamiento algorítmico y del modelo — A — Datos y Modelo	FCM Frecuencia de calibración del modelo — A — Datos y Modelo
CCM Calidad de la calibración del modelo — A — Datos y Modelo	CPA Cumplimiento del proveedor con auditorías éticas y tecnológicas. — A — Datos y Modelo
PCE Probabilidad de certificación del proveedor — A — Datos y Modelo	PCR Puntaje del análisis crediticio — B — Decisión
PAC Probabilidad de aprobación del crédito — B — Decisión	NCA Número de créditos aprobados — B — Decisión
PC Precio del crédito (tasa de interés) — B — Decisión	DHCM Debilidad del historial crediticio de las mujeres — C — Solicitantes
DSR Desvío de solvencia real de la solicitante — C — Solicitantes	IF Inclusión financiera de las mujeres — C — Solicitantes
SERC Sesgo estructural del resultado crediticio — D — Impactos	SSG Sesgo sistémico de género — D — Impactos
CAC Calidad del análisis crediticio — D — Impactos	PIC Probabilidad de impago en cartera — D — Impactos
BEFI Bienestar económico familiar e intergeneracional — D — Impactos	EGO Eficacia de la gobernanza interna de IA — E — Gobernanza
TRYE Transparencia y explicabilidad del modelo — E — Gobernanza	MEPGA Mecanismo de paro de gobernanza activado — E — Gobernanza
PSE Prueba sistémica de equidad (prueba controlada de generación sintética de sesgos) — E — Gobernanza	RAP Reporte de auditoría públicamente accesible — E — Gobernanza

2.1

Relaciones funcionales por ámbito

A continuación se describen las relaciones funcionales registradas en la matriz y representadas en el mapa, organizadas por ámbito. La cadena algebraica indica el tipo de relación entre variables: el signo (+) señala que un incremento del emisor incrementa al receptor; el signo (–) señala que lo reduce.

Ámbito A — Datos y Modelo

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
PSDH Permanencia de sesgos en datos históricos	VEA / Validez del entrenamiento algorítmico y del modelo	(PSDH) – (VEA)	La no eliminación de los sesgos históricos de género en las bases de datos reduce la validez del modelo de IA. A mayor permanencia de sesgos, menos válido y confiable el modelo.
PSDH Permanencia de sesgos en datos históricos	DHCM / Debilidad del historial crediticio de las mujeres / SSG / Sesgo sistémico de género	(PSDH) + (DHCM) + (SSG)	La permanencia de sesgos en bases de datos perpetúa el historial crediticio escaso de las mujeres. A mayor permanencia de sesgos, menor utilidad de los historiales crediticios. El historial de menor acceso femenino al crédito formal se codifica como “mayor riesgo” y refuerza el sesgo sistémico de género.
PSDH Permanencia de sesgos en datos históricos	PCR / Puntaje del análisis crediticio / DHCM / Debilidad del historial crediticio de las mujeres / SSG / Sesgo sistémico de género	(PSDH) – (PCR) + (DHCM) + (SSG)	La permanencia de sesgos en bases de datos reduce la calidad de los resultados de la evaluación crediticia, lo que perpetúa la debilidad del historial crediticio e incrementa los sesgos sistémicos de género.
UVPG Uso de variables proxy de género	VEA / Validez del entrenamiento algorítmico y del modelo	(UVPG) – (VEA)	Las variables proxy (tipo de empleo, sector, antigüedad) correlacionan con género. A mayor uso de variables proxy, menor validez del modelo.
UVPG Uso de variables proxy de género	SERC / Sesgo estructural del resultado crediticio / CAC / Calidad del análisis crediticio	(UVPG) + (SERC) – (CAC)	El uso de variables proxy penaliza el resultado de la solicitante sin que exista evidencia directa de mayor riesgo. A mayor uso de variables proxy, mayor el sesgo estructural del resultado y menor la calidad del análisis crediticio.

Ámbito A — Datos y Modelo (continuación 1 de 2)

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
UVPG / Uso de variables proxy de género	SSG / Sesgo sistémico de género	(UVPG) + (SSG)	Las variables proxy amplifican el sesgo de género al asociar patrones de empleo femenino con mayor probabilidad de incumplimiento. A mayor uso de variables proxy, mayor sesgo sistémico.
UVPG / Uso de variables proxy de género	DSR / Desvío de solvencia real de la solicitante / CAC / Calidad del análisis crediticio	(UVPG) + (DSR) – (CAC)	El uso de variables proxy subestima la solvencia real de la solicitante. A mayor uso de dichas variables, mayor el desvío del resultado respecto de la solvencia real y menor la calidad del análisis crediticio.
UVPG / Uso de variables proxy de género	PIC / Probabilidad de impago en cartera / CAC / Calidad del análisis crediticio	(UVPG) + (PIC) – (CAC)	A mayor uso de variables proxy, mayor la probabilidad estimada de impago (por sobreestimación del riesgo) y menor la calidad del análisis crediticio.
FCM / Frecuencia de calibración del modelo	SSG Sesgo sistémico de género / DSR / Desvío de solvencia real de la solicitante / CAC / Calidad del análisis crediticio	(FCM) – (SSG) – (DSR) + (CAC)	A mayor frecuencia de calibración del modelo, menos sesgo sistémico de género, menor desvío del resultado respecto de la solvencia real de la solicitante y mayor calidad del análisis crediticio.
FCM / Frecuencia de calibración del modelo	VEA / Validez del entrenamiento algorítmico y del modelo / SSG / Sesgo sistémico de género	(FCM) + (VEA) – (SSG)	A mayor frecuencia de calibración del modelo, mayor validez de los resultados y menor sesgo sistémico de género.
FCM / Frecuencia de calibración del modelo	SERC / Sesgo estructural del resultado crediticio / SSG / Sesgo sistémico de género / CAC / Calidad del análisis crediticio	(FCM) – (SERC) – (SSG) + (CAC)	La recalibración del modelo —mediante reentrenamiento, ajuste de umbrales o corrección de proxies— reduce el sesgo estructural y el sesgo sistémico de género, lo que mejora la calidad del análisis crediticio.
CPA / Compliance del proveedor con auditorías éticas y tecnológicas	FCM / Frecuencia de calibración del modelo / CCM / Calidad de la calibración del modelo / PCE / Probabilidad de certificación del proveedor	(CPA) + (FCM) + (CCM) + (PCE)	El compliance con las auditorías del proveedor obliga a calibrar frecuentemente el modelo, lo que ayuda a la identificación de fallas en el entrenamiento y mejora la calibración, contribuyendo a que el proveedor se mantenga certificado.

Ámbito A — Datos y Modelo (continuación 2 de 2)

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
CPA / Compliance del proveedor con auditorías éticas y tecnológicas	TRYE Transparencia y explicabilidad del modelo	(CPA) + (TRYE)	El compliance con las auditorías exige al proveedor documentar los atributos del modelo y sus pesos, aumentando la transparencia y explicabilidad.

Ámbito B — Decisión

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
PCR Puntaje del análisis crediticio	PAC / Probabilidad de aprobación del crédito	(PCR) + (PAC)	Un mejor puntaje en el análisis crediticio incrementa la probabilidad de aprobación del crédito.
PCR Puntaje del análisis crediticio	PC / Precio del crédito (tasa de interés) / BEFI Bienestar económico familiar e intergeneracional	(PCR) – (PC) – (BEFI)	Un puntaje más alto se traduce en una tasa de interés menor, lo que reduce el costo del crédito y por lo tanto incrementa el bienestar económico familiar e intergeneracional.
NCA Número de créditos aprobados	PSDH Permanencia de sesgos en datos históricos / CAC Calidad del análisis crediticio / BEFI Bienestar económico familiar e intergeneracional	(NCA) – (PSDH) + (CAC) + (BEFI)	A mayor número de créditos aprobados, menor permanencia de sesgos históricos (se enriquece la base con datos más diversos), mejor calidad del análisis crediticio, y mayor bienestar económico de los hogares.

Ámbito C — Solicitantes

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
DHCM Debilidad del historial crediticio de las mujeres	PCR / Puntaje del análisis crediticio / NCA / Número de créditos aprobados / IF / Inclusión financiera de las mujeres / BEFI / Bienestar económico familiar e intergeneracional	(DHCM) – (PCR) – (NCA) – (IF) – (BEFI)	Una mayor debilidad del historial crediticio es penalizada por el modelo, que lo interpreta como señal de riesgo, reduciendo el puntaje en el análisis crediticio, lo cual reduce el número de créditos aprobados y disminuye la inclusión financiera de las mujeres, reduciendo el bienestar económico familiar e intergeneracional.

Ámbito D — Impactos

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
IF Inclusión financiera de las mujeres	SSG / Sesgo sistémico de género / DHCM / Debilidad del historial crediticio de las mujeres / CAC / Calidad del análisis crediticio / NCA / Número de créditos aprobados / BEFI / Bienestar económico familiar e intergeneracional	(IF) – (SSG) – (DHCM) + (CAC) + (NCA) + (BEFI)	A medida que más mujeres son incluidas en el sistema financiero, el sesgo de género pierde sustento estadístico y se reduce, lo cual fortalece el historial crediticio (reduce su debilidad), mejora la calidad del análisis, incrementa el número de créditos aprobados y aumenta el bienestar económico familiar e intergeneracional.

Ámbito E — Gobernanza y Regulación

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
EGO Eficacia de la gobernanza interna de IA	FCM / Frecuencia de calibración del modelo / CCM / Calidad de la calibración del modelo / SSG / Sesgo sistémico de género	(EGO) + (FCM) + (CCM) – (SSG)	La gobernanza interna establece los protocolos de calibración periódica, lo cual eleva la frecuencia y calidad de la calibración del modelo. Cuando esto se cumple, se reduce el sesgo sistémico de género.
EGO Eficacia de la gobernanza interna de IA	CPA / Compliance del proveedor con auditorías éticas y tecnológicas / PCE / Probabilidad de certificación del proveedor / TRYE Transparencia y explicabilidad del modelo	(EGO) + (CPA) + (PCE) + (TRYE)	La gobernanza ordena la auditoría periódica del proveedor como condición de continuidad del contrato. Cuando la gobernanza es eficaz, se incrementan los compliances del proveedor con sus auditorías éticas y tecnológicas, lo que hace que el proveedor se mantenga certificado y mejore la transparencia y explicabilidad del modelo.
EGO Eficacia de la gobernanza interna de IA	CAC / Calidad del análisis crediticio / SERC / Sesgo estructural del resultado crediticio	(EGO) + (CAC) – (SERC)	La gobernanza eficaz mejora la calidad del análisis crediticio al establecer criterios y métricas de equidad para la evaluación, y reduce el sesgo estructural del resultado al exigir auditorías y correcciones sistemáticas.

Ámbito E — Gobernanza y Regulación (continuación 1 de 2)

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
TRYE Transparencia y explicabilidad del modelo	CAC / Calidad del análisis crediticio / PCR / Puntaje del análisis crediticio / SERC / Sesgo estructural del resultado crediticio	(TRYE) + (CAC) + (PCR) – (SERC)	La transparencia y explicabilidad del modelo mejora la calidad del análisis al hacer visibles los atributos y pesos del modelo. Permite que las solicitantes conozcan los factores que afectan su puntaje y los corrijan, mejorando el PCR. Al visibilizar el sesgo estructural, facilita su reducción.
PCE Probabilidad de certificación del proveedor	VEA / Validez del entrenamiento algorítmico y del modelo / CPA / Compliance del proveedor con auditorías éticas y tecnológicas	(PCE) + (VEA) + (CPA)	Un proveedor certificado garantiza estándares de entrenamiento que elevan la validez del modelo. La certificación también refuerza el compliance del proveedor con las auditorías, pues la pérdida de certificación tiene consecuencias contractuales.
MEPGA Mecanismo de paro de gobernanza activado	EGO / Eficacia de la gobernanza interna de IA / FCM / Frecuencia de calibración del modelo / CPA / Compliance del proveedor con auditorías éticas y tecnológicas / PSE / Prueba sistémica de equidad (red-teaming de sesgos)	[(lim SSG) o (lim SERC)] → (MEPGA) + (EGO) + (FCM) + (CPA) → (PSE)	Cuando SSG o SERC superan el nivel de disparo predefinido, el mecanismo de paro se activa y obliga a escalar la gobernanza: exige calibración inmediata, activa la auditoría del proveedor y ordena una prueba sistémica de equidad. Es el mecanismo de “no desplegar si”, el sistema no puede operar normalmente hasta que las métricas de sesgo vuelvan a los niveles aceptables predeterminados.
PSE Prueba sistémica de equidad (red-teaming de sesgos)	SERC / Sesgo estructural del resultado crediticio / SSG / Sesgo sistémico de género / CAC / Calidad del análisis crediticio / VEA / Validez del entrenamiento algorítmico y del modelo / MEPGA Mecanismo de paro de gobernanza activado	(PSE) – (SERC) – (SSG) + (CAC) + (VEA) – (MEPGA)	La prueba sistémica de equidad es una evaluación adversarial periódica: simula distintos perfiles de solicitantes mujeres para detectar sesgos antes de que lleguen a producción. A mayor frecuencia e intensidad de la PSE, menor el sesgo estructural y el sesgo sistémico detectados, mejor la calidad del análisis y mayor la validez del modelo. Cuando la prueba es exitosa, reduce la probabilidad de que el umbral de gobernanza se active, cerrando el ciclo de control preventivo.

Ámbito E — Gobernanza y Regulación (continuación 2 de 2)

EMISOR	RECEPTORES	CADENA	DESCRIPCIÓN
RAP Reporte de auditoría públicamente accesible	TRYE Transparencia y explicabilidad del modelo / PCE / Probabilidad de certificación del proveedor / EGO / Eficacia de la gobernanza interna de IA / SSG / Sesgo sistémico de género	(RAP) + (TRYE) + (PCE) + (EGO) – (SSG)	El reporte de auditoría públicamente accesible —ante CNBV, CONDUSEF, inversionistas o sociedad civil— es la forma más robusta de transparencia: no solo interna sino verificable externamente. A mayor accesibilidad del reporte, mayor la transparencia y explicabilidad del modelo, mayor la probabilidad de que el proveedor mantenga su certificación, más eficaz la gobernanza (accountability externo) y menor el sesgo sistémico de género, pues la exposición pública crea un incentivo poderoso para corregir desviaciones antes de que sean noticias.

2.2

Fundamentos del análisis sistémico: ámbitos y rutas

El mapa sistémico admite dos formas de análisis complementarias: por ámbitos funcionales homogéneos, y por rutas o cadenas causales que atraviesan varios ámbitos. Los dos criterios son complementarios y se usan en secuencia: primero el análisis por ámbitos organiza el sistema y asigna las familias de instrumentos; después el análisis por rutas selecciona los eslabones concretos donde actuar.

TABLA 2.1

Complementariedad de los dos enfoques analíticos

CRITERIO DE COMPARACIÓN	ANÁLISIS POR ÁMBITOS	ANÁLISIS POR RUTAS
Pregunta que responde	¿En qué eslabón o nodo concreto conviene actuar para cortar la propagación del sesgo de género y sus efectos?	¿Qué instrumentos gobiernan cada zona del sistema?
Unidad de análisis	Secuencias de nodos conectados que atraviesan varios grupos.	Grupos de nodos con función causal homogénea.
Qué revela	Los efectos indirectos (de segundo orden) y los puntos de palanca a lo largo de la cadena de sesgo.	La estructura del sistema: origen del sesgo, mediación algorítmica, decisión crediticia y destino del daño.
Conexión con la metodología	Selecciona, dentro de ese marco, los eslabones donde aplicar la acción concreta de corrección del sesgo.	Asigna a cada ámbito su cuadrante en la matriz impacto-certidumbre y sus instrumentos de regulación.
Tipo de resultado	Una estrategia de intervención: dónde actuar y en qué orden para reducir el sesgo y ampliar la inclusión financiera.	Un mapa de gobernabilidad: qué se regula y con qué familia de instrumentos.
Principal ventaja	Permite cortar cadenas completas de sesgo interviniendo en un solo eslabón clave.	Evita aplicar el instrumento equivocado en el lugar equivocado.
Cuándo se usa	Después: para diseñar y priorizar las acciones concretas dentro de ese marco.	Primero: para organizar el sistema y enmarcar las opciones de regulación.

Fuente: Elaboración propia.

El análisis examina cada uno de los cinco ámbitos del sistema, identifica la vulnerabilidad y los instrumentos que le corresponden, y en los ámbitos donde la certidumbre es suficiente, ofrece una estimación numérica de riesgo e impacto conforme al método V-R-I (ver Anexo 2).

Ámbito A — Datos y Modelo

- Qué se identifica aquí — Vulnerabilidad (V)
- Las cadenas (PSDH) + (DHCM) y (UVP) + (SSG) muestran que la vulnerabilidad se origina en datos que reflejan exclusión histórica y en proxies que correlacionan con género. Las mujeres con historial delgado son sistemáticamente subevaluadas, no por menor solvencia real, sino por menor presencia en los datos formales de crédito.
- V se estima alta (4-5): la población afectada combina marcadores de susceptibilidad estructural (historial débil, empleo informal, ausencia del sistema financiero formal) con una opacidad del modelo que dificulta la identificación y corrección del sesgo desde adentro.
- Esta sección alimenta la estimación de V para los ámbitos B y D, pero su tratamiento requiere instrumentos de cuadrante B: correulación con el proveedor, auditoría y certificación.
- Qué se aplica aquí — instrumentos de cuadrante B
- Correulación con el proveedor para fijar estándares de equidad algorítmica como condición contractual.
- Auditoría obligatoria de los efectos del modelo por grupo de género en el puntaje, para lo cual es conveniente hacer un análisis estadístico de razón de proporciones que nos dé un indicador o parámetro de la medida en la que el modelo produce resultados sistemáticamente desfavorables para las mujeres. Debe tomarse en cuenta que este tipo de pruebas no distinguen causas, dependen de qué tan bien están definidos los grupos y no captura sesgos de acceso, sólo mide a las personas que ya solicitaron crédito, por lo que este análisis no concluye la investigación sino que la detona.
- Protocolo de certificación del proveedor (PCE): el proveedor solo puede operar si sus modelos pasan pruebas periódicas de equidad definidas por la institución y por el regulador.
- La cadena (CPA) + (FCM) + (CCM) + (PCE) señala la palanca positiva: el compliance con las auditorías eleva la frecuencia y calidad de la calibración, lo que reduce PSDH y UVP aguas abajo.

Ámbito B — Decisión

- Justificación de la estimación cualitativa
- Vulnerabilidad $V = 5$ (muy alta): la decisión recae sobre mujeres en situación de vulnerabilidad financiera estructural, con historial delgado y solvencia real subestimada.
- Probabilidad $P = 4$ (alta): las relaciones que llegan a PCR ((PSDH)–(PCR), (UVPG)+(SERC), (DHCM)–(PCR), (DSR)–(PCR)) elevan conjuntamente la probabilidad de un puntaje sesgado; solo las variables de calibración y gobernanza la compensan.
- Riesgo $R = P \times V = 4 \times 5 = 20$ (crítico en la matriz de riesgo).
- Personas expuestas $N = 4$ (amplio): el 60% de la cartera potencial son mujeres; decenas de miles de solicitudes semanales en una institución mediana.
- Impacto $I = R \times N = 20 \times 4 = 80 \rightarrow$ severidad crítica. Confirma el cuadrante D para la decisión y la necesidad de sandbox con criterios de parada (MEPGA).
- Condición indispensable: ninguna decisión automatizada de rechazo debe ser irreversible sin revisión humana; el sistema solo puede recomendar, no decidir de forma vinculante sin mecanismo de recurso.

Ámbito C — Solicitantes

- Justificación de la estimación
- Vulnerabilidad $V = 5$ (muy alta): el historial delgado no refleja menor solvencia, sino exclusión histórica del sistema formal; las solicitantes afectadas no tienen mecanismo para demostrar su solvencia real dentro del modelo actual.
- Probabilidad $P = 4$ (alta): la cadena (DHCM)–(PCR)–(NCA)–(IF) es la ruta de exclusión más corta y directa del sistema; opera en cada solicitud sin necesidad de condiciones especiales.
- Riesgo $R = 4 \times 5 = 20$ (crítico).
- Personas expuestas $N = 5$ (masivo): las mujeres con historial delgado representan la mayoría de la cartera potencial no atendida; en México, más del 50% de la población adulta sin acceso formal al crédito son mujeres.
- Impacto $I = 20 \times 5 = 100 \rightarrow$ severidad crítica máxima. Requiere historial alternativo de crédito, datos de solvencia no convencionales y sandbox para productos piloto de inclusión.

Ámbito D — Impactos

- Justificación de la estimación
- Vulnerabilidad $V = 5$ (muy alta): los efectos incluyen la perpetuación del ciclo de exclusión financiera intergeneracional ($BEFI\downarrow$), el refuerzo del sesgo sistémico ($SSG\uparrow$) y el deterioro de la calidad de la cartera ($PIC\uparrow$), daños que se retroalimentan y son difíciles de revertir.
- Probabilidad $P = 3$ (media): no todo sesgo en el modelo produce una crisis sistémica inmediata, pero el riesgo latente crece con el tiempo y el volumen de solicitudes afectadas.
- Riesgo $R = 3 \times 5 = 15$ (muy alto).
- Personas expuestas $N = 5$ (masivo): el sesgo sistémico de género en el crédito tiene efectos que se multiplican al pasar de la solicitante a su hogar y a generaciones futuras.
- Impacto $I = 15 \times 5 = 75 \rightarrow$ severidad alta, en el límite con crítica. La N elevada compensa la probabilidad media. Requiere gobernanza anticipatoria, horizon scanning y monitoreo longitudinal.

Ámbito E — Gobernanza y Regulación

- Justificación de la estimación
- Vulnerabilidad $V = 3$ (moderada): la gobernanza de IA es un campo emergente en México; hay margen para construir capacidades antes de que la regulación se endurezca, pero el riesgo de inacción crece.
- Probabilidad $P = 3$ (media): el marco regulatorio mexicano (CNBV, CONDUSEF, Ley Fintech) avanza hacia exigencias de IA ética; la probabilidad de intervención regulatoria aumentará en los próximos años.
- Riesgo $R = 3 \times 3 = 9$ (moderado, pero con tendencia alcista).
- Personas expuestas $N = 4$ (amplio): las sanciones y la pérdida de licencia afectan a la institución completa y a toda su cartera de clientes.
- Impacto $I = 9 \times 4 = 36 \rightarrow$ severidad moderada-alta. La oportunidad de intervención proactiva es alta: actuar hoy en gobernanza cuesta menos que responder a una sanción regulatoria o a un escándalo reputacional mañana.

Tabla 2.2 · Estimaciones numéricas de riesgo e impacto

ÁMBITO	V	P	R	N	I	SEVERIDAD / INSTRUMENTOS PRIORITARIOS
A — Datos y Modelo (PSDH, UVPG, VEA, FCM, CCM, CPA, PCE)	4-5	4	16-20	4	64-80	Crítico · Corregulación + Auditoría + Certificación
B — Decisión (PCR, PAC, NCA, PC)	5	4	20	4	80	Crítico · Sandbox + Criterios de parada (MEPGA) + Revisión humana
C — Solicitantes (DHCM, DSR, IF)	5	4	20	5	100	Crítico máximo · Historial alternativo + Sandbox piloto inclusión
D — Impactos (SERC, SSG, CAC, PIC, BEFI)	5	3	15	5	75	Muy alto · Monitoreo continuo + SupTech + Gobernanza anticipatoria
E — Gobernanza (EGO, TRYE, MEPGA, PSE, RAP)	3	3	9	4	36	Moderado-alto · Soft law + RegTech + RAP · Oportunidad proactiva

Tabla 2.3 · Resumen de instrumentos sugeridos

ÁMBITO	IMPACTO	CERTIDUMBRE	INSTRUMENTOS SUGERIDOS
A — Datos y Modelo (PSDH, UVPG, VEA, FCM, CCM, CPA, PCE)	Alto	Alta	Corregulación con el proveedor · Auditoría obligatoria · Certificación del proveedor (PCE) · Prueba sistémica de equidad (PSE)
B — Decisión (PCR, PAC, NCA, PC)	Alto	Alta	Corregulación con el proveedor · Auditoría del modelo · Límites algorítmicos al precio diferenciado por género · Revisión humana de rechazos
C — Solicitantes (DHCM, DSR, IF)	Alto	Media	Historial de crédito alternativo · Educación financiera · Datos alternativos de solvencia · Sandbox para productos piloto de inclusión
D — Impactos (SERC, SSG, CAC, PIC, BEFI)	Alto	Baja	Sandbox estricto · Criterios de parada (MEPGA) · Monitoreo continuo de equidad · SupTech en tiempo real
E — Gobernanza y Regulación (EGO, TRYE, MEPGA, PSE, RAP)	Alto	Media	Principios éticos y soft law · RegTech / SupTech · Reporte público de auditoría (RAP) · Regulación por resultados · Gobernanza anticipatoria

2.2.5

Estrategias de mejora: acciones a lo largo de cada cadena

Cada cadena del mapa se convierte en una estrategia de mejora cuando se define un menú de acciones aplicables en cada uno de sus nodos.

Estrategia 1 · Cadena de sesgos en los datos

Recorre la propagación desde la permanencia de sesgos en los datos históricos y el uso de proxies de género hasta el sesgo estructural del resultado crediticio y el sesgo sistémico. Como su origen (PSDH, UVPG) está en el modelo y los datos del proveedor, la estrategia se concentra en la cadena de calibración y auditoría: es la intervención más aguas arriba del sistema y la que mayor efecto multiplicador produce sobre todas las demás cadenas.

	NODO 1	NODO 2	NODO 3	NODO 4	NODO 5
Acciones o medidas	Como parte del ciclo continuo de mejoramiento de las bases de datos, hay que incorporar todo tipo de ingresos, remesas y créditos informales al historial. Revisar continuamente los resultados de entrenamiento para identificar y eliminar variables que codifican exclusión histórica. Utilizar umbrales que midan, por ejemplo, proporciones de género para detener y corregir los datos antes de desplegarlos. de datos antes de corregirlos.	Análisis de correlación de variables con género; eliminar o ponderar diferencialmente las proxies identificadas. Hacer un análisis estadístico de razón de proporciones antes de cada actualización del modelo.	Establecer protocolos de calibración por umbral, no por calendario. Incluir métricas de equidad de género en cada ciclo de calibración.	Monitoreo automatizado y continuo del diferencial de puntaje por género (SupTech) con controles en tiempo real.	Auditoría de equidad sistémica periódica; publicar métricas de equidad en el RAP; si el sesgo sistémico supera el umbral, activar acciones.
Recursos necesarios	Altos	Medios	Medios	Medios	Medios
Dificultad	Alta	Media	Media	Media	Baja

Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.

Estrategia 2 · Cadena de gobernanza del proveedor

Recorre la cadena de control sobre el proveedor externo del motor de evaluación : desde el compliance con las auditorías hasta la certificación y la transparencia. Sus acciones se concentran en el vínculo contractual con el proveedor, donde la intervención es la más eficiente: una cláusula bien diseñada en el contrato activa toda la cadena de mejora del modelo.

	NODO 1	NODO 2	NODO 3	NODO 4	NODO 5
Acciones o medidas	Incluir cláusulas de auditoría ética y técnica periódica en el contrato con el proveedor; pérdida de certificación como condición de rescisión contractual.	Exigir certificación del proveedor conforme a estándares de equidad algorítmica (ISO/IEC 42001 o equivalente); renovación anual condicionada a resultados de PSE.	Explicación automática de rechazos: la solicitante debe recibir no solo el motivo del rechazo sino las acciones concretas que mejorarían su perfil en el siguiente ciclo. Documentación de atributos y pesos del modelo publicada y verificable por el regulador sin necesidad de solicitud.	Validación técnica por organismo permanente (no perito puntual) con autoridad explícita para bloquear el despliegue si el modelo no pasa las pruebas de rendimiento diferencial por género. “No desplegar si no pasa”; es la condición, no una recomendación. Periodicidad mínima: en cada actualización del modelo.	Pantalla de métricas de auditoría accesible en tiempo real, no solo reporte anual. Actualización mínima trimestral con métricas de equidad. El reporte anual formal ante CNBV/CONDUSEF se construye a partir de la información en pantalla: no es un documento separado sino una síntesis verificable del historial del año.
Recursos necesarios	Medios	Medios-altos	Medios	Medios-altos	Bajos
Dificultad	Media	Media-alta	Media	Media	Baja

Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.

Estrategia 3 • Cadena de inclusión financiera

Recorre la propagación desde la debilidad del historial crediticio hasta la exclusión financiera y el deterioro del bienestar. DHCM activa un sandbox de producto piloto cuyos datos retroalimentan el modelo. IF abandona la educación financiera como acción de fe y la convierte en experimento controlado con criterio de evaluación.

	NODO 1	NODO 2	NODO 3	NODO 4	NODO 5
Acciones o medidas	Sandbox de producto piloto: créditos con límites bajos y escalonamiento para mujeres con historial delgado, con evaluación comparativa de tasa de impago real vs. estimada. Los resultados del piloto alimentan el reentrenamiento del modelo; el sandbox genera los datos que reducen DHCM en el siguiente ciclo.	Datos alternativos de solvencia (comportamiento digital, pagos de renta) sujetos al mismo ciclo de validación de equidad que los datos tradicionales. Hacer un análisis estadístico de razón de proporciones antes de cada actualización del modelo. La validación de los datos alternativos es parte del protocolo PSE.	Revisión humana obligatoria de todos los rechazos de mujeres con historial delgado: no discrecionalmente, sino activada por regla de disparo. Segunda opinión en casos límite (± margen predefinido del umbral). Hacer que el margen sea revisable mediante evidencia acumulada del sandbox.	Experimento continuo y controlado de educación financiera: grupo de tratamiento vs. grupo de control, midiendo el efecto en la tasa de aprobación futura para definir retroalimentación y mejorar resultados en aprobaciones futuras. La educación financiera se valida como instrumento de aplicación continua.	Indicador de bienestar de la cartera femenina con umbral de alerta: si el diferencial de bienestar respecto a la cartera masculina supera X% predeterminado, activa revisión del modelo. Los resultados retroalimentan el ciclo de calibración, no solo se publican en el RAP.
Recursos necesarios	Altos	Medios	Medios	Bajos	Medios
Dificultad	Media	Media	Media	Baja	Baja

Recursos y dificultad: verde = bajo/favorable • ámbar = intermedio • rojo = alto/desfavorable.

Estrategia 4 · Cadena de gobernanza interna y mecanismo de paro

Recorre cómo la eficacia de la gobernanza interna activa o previene el mecanismo de paro (MEPGA) y la prueba sistémica de equidad (PSE). La acción de mayor palanca está en el origen de la cadena: una gobernanza interna bien diseñada (EGO alta) reduce la probabilidad de que el umbral se dispare y, cuando se dispara, garantiza una respuesta institucional rápida y efectiva.

	NODO 1	NODO 2	NODO 3	NODO 4
Acciones o medidas	Mecanismos dinámicos de monitoreo de gobernanza, por ejemplo, uso de métricas de equidad con umbrales predefinidos: cuando se superan, activan protocolo de escalamiento con plazo máximo de respuesta (ej. 48 horas para notificar al consejo, 30 días para corregir). La política de no discriminación es operativa, no declarativa. Responsable nombrado de gobernanza de IA con autoridad ejecutiva para pausar el sistema, independiente de la dirección comercial.	Prueba controlada de generación sintética de sesgos con umbrales predefinidos. Cuando éstos se rebasan se produce un informe con criterio de pase/no-pase que alimenta la decisión para activar el MEPGA.	Si los sesgos reportados por la PSE no se corrigen en un tiempo preestablecido, el sistema activa el MEPGA hasta que las correcciones sean implantadas y validadas.	Reentrenamiento con criterio de aceptación/rechazo: el modelo recalibrado no se despliega hasta pasar la PSE. Esta condición de despliegue asegura la corrección de dinámica de sesgos. Cada ciclo documentado genera nuevos insumos para la auditoría del proveedor (CPA).
Recursos necesarios	Medios	Medios	Bajos	Medios-altos
Dificultad	Media	Media	Baja	Alta

Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.

Estrategia 5 · Cadena de precio y riesgo de cartera

Recorre cómo el sesgo en el puntaje se traduce en precios diferenciados y en subestimación del riesgo real de la cartera. Esta cadena tiene un argumento de negocio directo: una institución que discrimina en precio por sesgo de género pierde a las mejores pagadoras del mercado y acumula riesgo de cartera de manera invisible.

	NODO 1	NODO 2	NODO 3	NODO 4
Acciones o medidas	Diferencial de puntaje por género documentado como disparador contractual, no como argumento de negociación. Si el diferencial supera el umbral pactado en el contrato con el proveedor, activa automáticamente la cláusula de recalibración, sin necesidad de negociación caso a caso.	Bandas de precio con regla de paridad: la tasa ofrecida a una mujer no puede superar la ofrecida a un hombre con igual perfil de riesgo. Implementable por política interna sin reforma legal. Revisión humana de excepciones con motivación explícita registrada.	Comparación continua de tasa de impago real vs. estimada por género. Cuando la tasa real de la cartera femenina es menor que la estimada, esa brecha se convierte en evidencia contractual que activa la cláusula de recalibración del proveedor. Es el argumento de negocio más potente: el modelo subestima la solvencia de las mejores pagadoras.	Indicador de aprobación por género con umbral de alerta: si el diferencial de aprobación mujer/hombre con igual perfil supera X %, activa revisión automática del modelo. El indicador mensual se publica en el tablero de equidad y retroalimenta el ciclo de calibración.
Recursos necesarios	Medios	Bajos	Bajos	Bajos
Dificultad	Media	Baja	Baja	Baja

Recursos y dificultad: verde = bajo/favorable · ámbar = intermedio · rojo = alto/desfavorable.

2.2.6 · Cómo combinar las estrategias

Las cinco estrategias convergen en PCR (puntaje del análisis crediticio). Las acciones en PCR — revisión humana obligatoria de rechazos, segunda opinión en casos límite, disparador contractual de recalibración— benefician simultáneamente a las cinco cadenas y son la inversión de mayor palanca del sistema.