

Regulación de la Inteligencia Artificial y Gestión de Riesgos

Un nuevo marco para la
innovación en el Sistema Financiero

III Jornada Anual de Gestión de Riesgos
Club de Gestión de Riesgos de la
República Dominicana (CGRRD)

Omar Bairan CFA, FRM, CQF

Santo Domingo, D.N.

18 de septiembre, 2025

Notificación

Esta presentación se realiza exclusivamente para fines informativos.

El contenido de esta presentación no debe ser interpretado como asesoría o recomendación.

Las opiniones presentadas son únicamente las del expositor.

Agenda

01

Diseño regulatorio: protección vs. innovación

02

Casos de implementación fallida

03

Riesgos de la IA en los servicios financieros

04

Gobernanza de la IA

05

IA explicable

06

Conclusiones

Diseño regulatorio: protección vs. innovación

Argumentos a favor de regular la IA

Derechos Fundamentales	Prevenir discriminación algorítmica y proteger privacidad, igualdad y dignidad humana.
Seguridad y Robustez	Reducir riesgos de errores, ciberataques y uso malicioso en sectores críticos.
Responsabilidad	Obligar a desarrolladores de sistemas de IA a rendir cuentas por decisiones automatizadas.
Confianza Social	Aumentar confianza de consumidores e inversionistas, favorecer adopción segura.
Prevención de Abusos	Limitar el uso fraudulento, manipulación política y vigilancia masiva.

Argumentos en contra de regular la IA

Innovación Tecnológica	Frena desarrollo ágil, encarece investigación, puede reducir la inversión.
-------------------------------	--

Costos Operativos	Eleva costos operativos, dificulta entrada al mercado para startups.
--------------------------	--

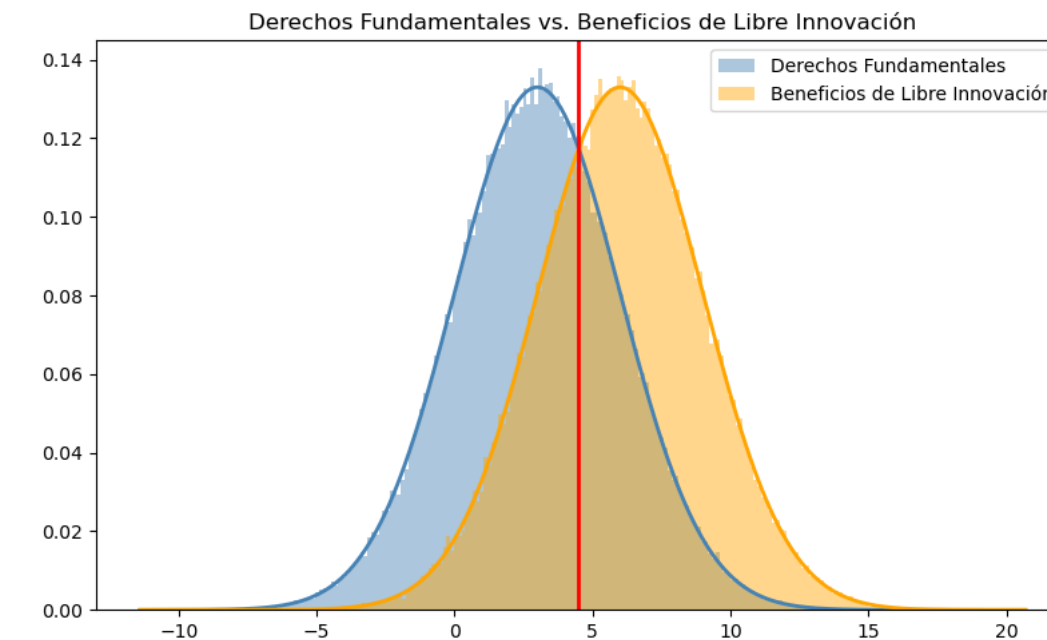
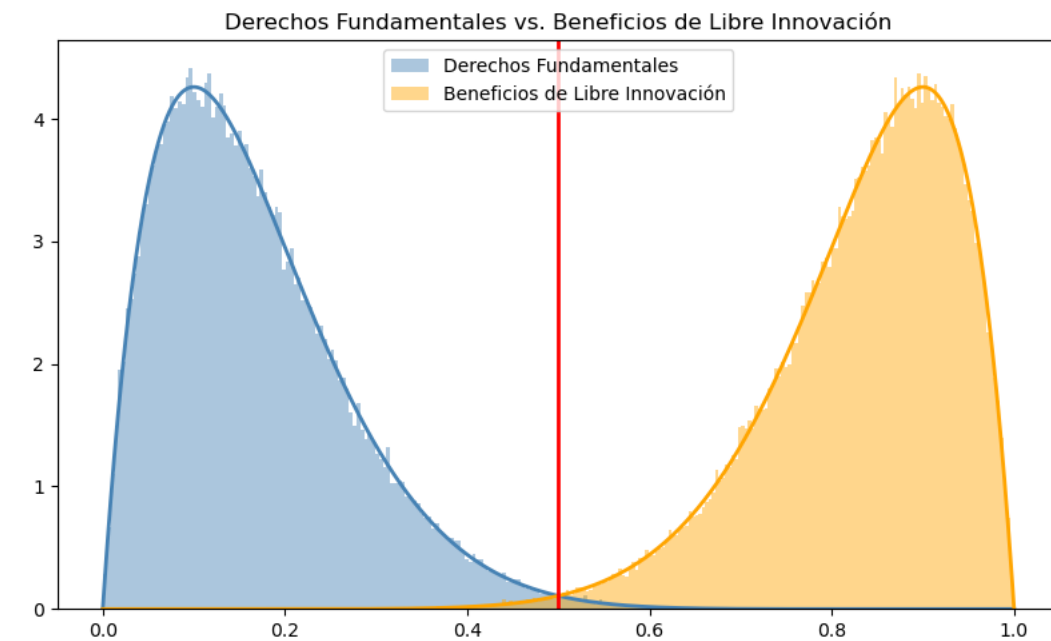
Adaptabilidad Regulatoria	Regulación puede volverse obsoleta rápidamente, posible incertidumbre jurídica.
----------------------------------	---

Competitividad Global	Motiva fuga de talentos/empresas a jurisdicciones menos reguladas.
------------------------------	--

Riesgo de Regulación Inadecuada	Normas mal diseñadas pueden crear efectos adversos no previstos.
--	--

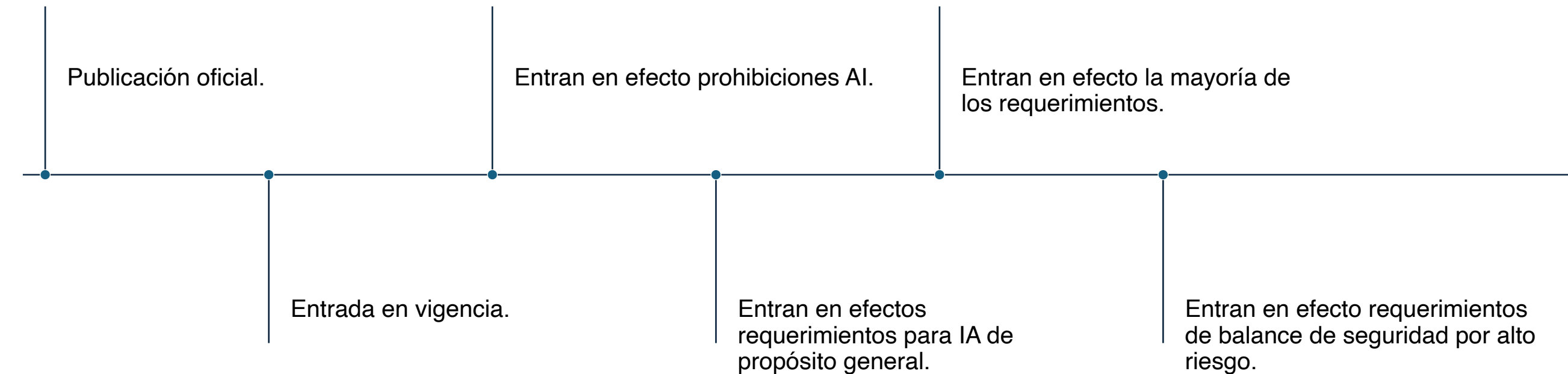
Protección de derechos vs. libre innovación

- Definir dónde trazar la línea divisoria para establecer el nivel de exigencias regulatorias es un proceso complejo.
- Establecer el nivel de regulación considerando potenciales daños por violaciones de derechos fundamentales y la promoción de la libre innovación es finalmente un ejercicio basado en juicio experto.



Ley de IA en la Unión Europea

- Es un referente global, orientado a fomentar la innovación responsable, la protección de derechos fundamentales, la privacidad y el desarrollo ético de la IA.
- Entró en vigor desde 2024 y la conformidad completa con la mayoría de las obligaciones debe alcanzarse en **agosto de 2026**.



Ley de IA en la Unión Europea

Registro y evaluación de conformidad antes de comercialización.

Supervisión humana y mitigación de riesgos a lo largo de todo el ciclo de vida.

Transparencia: Documentación clara, gobernanza de datos e informes sobre capacidades y limitaciones del sistema.

Monitoreo robusto y evaluación post-comercialización, incluyendo auditorías independientes y evaluación de impacto.

Oficina Europea de Inteligencia Artificial supervisa la implementación junto con las autoridades nacionales, coordinadas por una nueva Junta Europea de la IA.

Ley de IA en la Unión Europea

Requisitos conforme Nivel de Riesgo		
Nivel de riesgo	Ejemplos	Requisitos regulatorios
Riesgo inaceptable	Puntuación social, IA para manipulación	Prohibición total; restricción inmediata.
Alto riesgo	Empleo, migración, seguridad pública	Registro, evaluación de conformidad, supervisión humana, gestión de ciclo de vida.
Riesgo limitado	Chatbots	Transparencia y obligación de divulgación.
Riesgo mínimo	Juegos, filtros de spam	No hay obligaciones obligatorias; se recomiendan buenas prácticas.

Regulación de la IA en Estados Unidos

El enfoque estadounidense para la regulación de la IA es fragmentado y en gran medida orientado a incentivos más que a imposiciones prescriptivas.

Las órdenes y marcos federales fijan la dirección estratégica y guías de riesgo, mientras que los estados promulgan leyes vinculantes en áreas de riesgo clave.

La política ejecutiva actual favorece la innovación y la competitividad, con pocos requisitos técnicos directos para desarrolladores privados; no obstante, los marcos voluntarios y la regulación sectorial configuran buenas prácticas y obligaciones en ámbitos concretos

La regulación de la IA en Estados Unidos se caracteriza actualmente por una mezcla de **órdenes ejecutivos federales, leyes sectoriales, orientaciones de agencias federales y una extensa legislación a nivel estatal**, en lugar de un único estatuto integral.

Regulación de la IA en Estados Unidos

Orden Ejecutiva 14179 (enero de 2025): Revoca restricciones regulatorias previas (notablemente la EO 14110 de 2023), con el objetivo de **impulsar la innovación, la competitividad económica y la seguridad nacional**, mediante la **reducción de barreras regulatorias y la suspensión de requisitos que puedan obstaculizar el desarrollo, uso o adquisición de IA por parte del gobierno federal**. Encarga a las agencias federales revisar y, en su caso, revertir o revisar políticas de IA previas y presentar un **nuevo plan de acción nacional sobre IA en un plazo de 180 días**.

Áreas de enfoque: La orden enfatiza la **seguridad nacional**, el **crecimiento económico** y la **eliminación de “sesgos ideológicos o agendas sociales diseñadas”**. No impone regulaciones directas a desarrolladores privados de IA, **favoreciendo un entorno permisivo de innovación** para actores privados y gubernamentales, con especial atención a **aplicaciones de defensa e infraestructuras críticas**.

Regulación de la IA en Estados Unidos

Área de supervisión	Enfoque pre-2025	Cambios por decretos 2025
Enfoque regulatorio	Gestión de riesgo, DEI, confiabilidad.	Desregulación, innovación, supremacía del mercado.
Contratación federal	Énfasis en seguridad/ética y acceso amplio.	Neutralidad ideológica, requisitos de elegibilidad más estrictos.
Orientación de agencias (NIST)	Apoyo a controles de riesgo.	Revisión/suspensión de controles estrictos.
Coordinación intergubernamental	Patchwork, autonomía estatal.	Armonización federal mediante incentivos financieros.
Estrategia internacional	Multilateral, controles cautelosos.	Proactividad industrial y controles de exportación.

Casos de implementación fallida

Alta tasa de proyectos fallidos

1. Fallas en pilotos de IA generativa

- **Informe del MIT (2025):** el 95 % de los proyectos piloto de IA generativa en instituciones financieras no consiguieron producir ahorros financieros medibles ni aumentos de beneficios.
- **Problemas comunes:** pilotos mal acotados, ausencia de ROI y incapacidad para pasar de la “prueba de concepto” al despliegue en el mundo real.

2. Altas tasas de abandono

- **Datos de S&P Global:** el 42 % de las empresas canceló la mayoría de sus iniciativas de IA en 2025, un aumento fuerte frente al 17 % del año anterior.
- **En finanzas:** solo el 38 % de los proyectos de IA alcanzaron las expectativas de ROI, y más del 60 % sufrieron retrasos significativos.

Alta tasa de proyectos fallidos

3. Causas subyacentes

- **Problemas de datos (35% de los proyectos fallidos):** proyectos lanzados con datos incompletos, sesgados o incompatibles, lo que socava la fiabilidad del modelo desde el inicio.
- **Resistencia humana y organizacional (28%):** los gestores de riesgo desconfían de algoritmos en cajas negras, los equipos de cumplimiento frenan la adopción por temores regulatorios y los usuarios de negocio se aferran a flujos de trabajo heredados.
- **Deuda técnica e integración (22%):** los sistemas legados no están diseñados para IA, por lo que los proyectos quedan atrapados en la fase piloto y no pueden integrarse o escalar.
- **Complejidad regulatoria (15%):** leyes en evolución (GDPR, AI Act, regulaciones financieras) crean minas de cumplimiento que paralizan la toma de decisiones y frenan los despliegues.

Alta tasa de proyectos fallidos

4. Decepciones con chatbots

- **Hallazgos de reguladores (CFPB):** muchos chatbots de bancos decepcionaron a los usuarios por no responder consultas complejas ni escalar adecuadamente a agentes humanos. La mayoría eran sistemas basados en reglas que ofrecían respuestas inconsistentes o incompletas, generando insatisfacción y posibles riesgos legales.

5. Promesa exagerada vs. realidad

- **Informes de RAND, BCG, Deloitte:** eran comunes los “puntos ciegos” en la dirección - los decisores malinterpretaron la naturaleza probabilística y los límites de la IA. Expectativas poco realistas provocaron pérdida de confianza y abandono de proyectos cuando los modelos no entregaron certezas ni ahorros sustanciales.

Alta tasa de proyectos fallidos

Proyectos	Tipo de institución	Resultado
Pilotos de GenAI	Varios grandes bancos de EE. UU. / UE (informe MIT 2025)	95% sin impacto financiero esperado
Despliegues de chatbots	Bancos en EE. UU. y Reino Unido	Decepción de clientes, revisión regulatoria
Trading algorítmico	Múltiples hedge funds	Fallos de modelo bajo estrés de mercado
Scoring crediticio	Multinacionales / hallazgos regulatorios	Sesgos / incumplimiento regulatorio
Monitoreo transacciones	Multinacionales / hallazgos regulatorios	Incorrecto diseño del algoritmo, fallas en integración de datos y una gobernanza insuficiente / incumplimiento regulatorio.

Causa raíz de proyectos fallidos

Causa raíz	Ejemplo típico
Mala calidad de datos	Datos heredados incompletos, sesgados y no integrados
Pobre ajuste al negocio	Sin ROI estratégico, falta de claridad ejecutiva
Resistencia organizacional	Equipos compartimentados, falta de apoyo, aversión al cambio
Deuda técnica	Sistemas legados que bloquean integración/escalamiento
Complejidad regulatoria	Entorno regulatorio cambiante, posibles sanciones
Desalineación de proveedores	Contrataciones impulsadas por expectativas exageradas, mal encaje interno
Escasez de talento	Falta de perfiles con capacidades en IA + finanzas + regulación

Riesgos de la IA en los servicios financieros

Tendencias actuales

IA Generativa	Nuevas amenazas, incluidas las alucinaciones y el potencial de uso indebido para la creación de información fraudulenta.
Agentes de IA	Pueden tomar acciones complejas y ejecutar ciertas acciones de forma autónoma. Replantan las preguntas sobre responsabilidad y control en la IA.
Aprendizaje por refuerzo en asignación dinámica de riesgos	Pueden aprenden políticas óptimas a partir de prueba y error en sistemas dinámicos y pueden obtener mejores resultados que modelos estáticos.
IA Causal	Es un cambio de paradigma que va más allá del reconocimiento de patrones y predicciones en base a correlaciones. Busca identificar relaciones de causa y efecto y realizar inferencias.
Redes neuronales de grafos	Arquitecturas que permiten modelar sistemas complejos para análisis de riesgo sistémico. Pueden ser aplicadas en la generación de escenarios para pruebas de esfuerzo considerando eventos sistémicos.

Implicaciones sistémicas de la IA para la estabilidad financiera

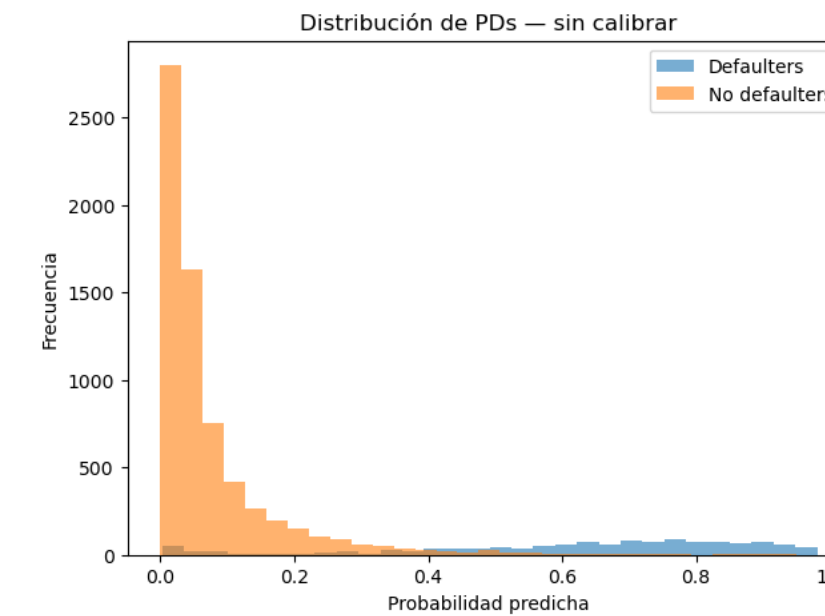
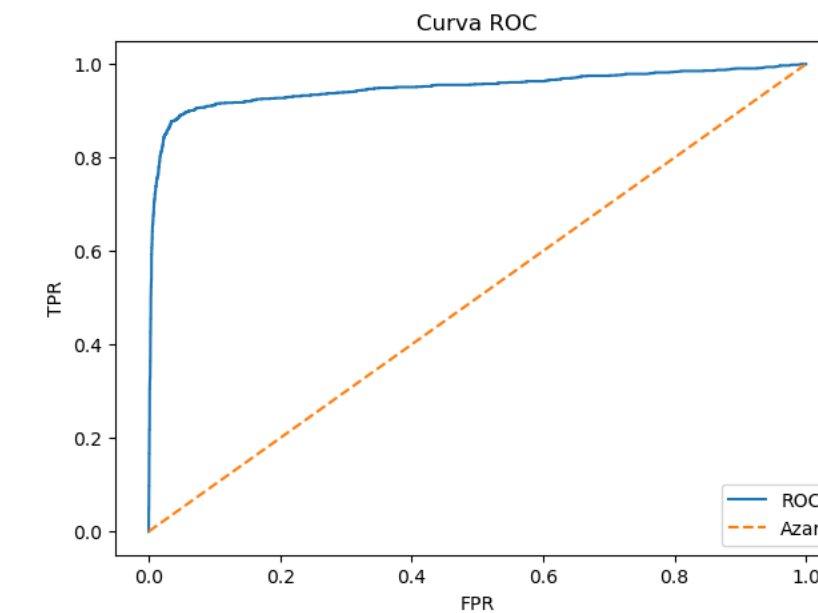
- La integración de la inteligencia artificial en el sistema financiero global presenta **riesgos sistémicos sin precedentes** que van mucho más allá de las vulnerabilidades financieras tradicionales.
- A medida que las instituciones financieras adoptan cada vez más la IA en funciones críticas —desde algoritmos de negociación hasta decisiones de crédito—, la posibilidad de **disrupciones a escala del sistema y choques amplificados** se ha convertido en una preocupación principal para la Consejo de Estabilidad Financiera (Financial Stability Board – FSB).
- En la publicación titulada “The Financial Stability Implications of Artificial Intelligence”, el FSB identifica cómo los desarrollos de la IA afectan vulnerabilidades del sector financiero.

Implicaciones sistémicas de la IA para la estabilidad financiera

Vulnerabilidad	Efectos del desarrollo de la IA
Concentración y dependencia de proveedores de servicios	<ul style="list-style-type: none">• Mayor uso de IA, incremento en la dependencia de hardware especializado, y servicios en la nube incrementan dependencia en terceros.• Complejidad de cadenas de abastecimiento para IA con mercados concentrados.• La dependencia en proveedores y la concentración de mercado puede incrementar la dependencia sistémica.
Correlaciones de mercado	<ul style="list-style-type: none">• Uso generalizado de la IA puede llevar a metodologías de modelado y uso de data similares.• Mayor uso de modelo pre-entrenados puede incrementar correlaciones de mercado.
Ciberseguridad	<ul style="list-style-type: none">• GenAI, LLM, LRM pueden incrementar las capacidades de actores en ciberataques, incrementando la frecuencia y severidad.• Uso intenso de datos y nuevas formas de interactuar con los datos pueden incrementar las oportunidades de ataques.• Uso de proveedores especializados puede incrementar el riesgo de ciberataques a los proveedores.
Riesgo de modelo, calidad de datos y gobernanza.	<ul style="list-style-type: none">• Mayor cantidad de metodologías de IA puede incrementar riesgo de modelos para entidades que no pueden validar y monitorearlos.• Mayor importancia de datos no estructurados para desarrollar modelos de IA, con baja transparencia y calidad.• Accesibilidad a herramientas de IA puede incentivar adopción sin controles y gobernanza.

Materialización del riesgo de modelo

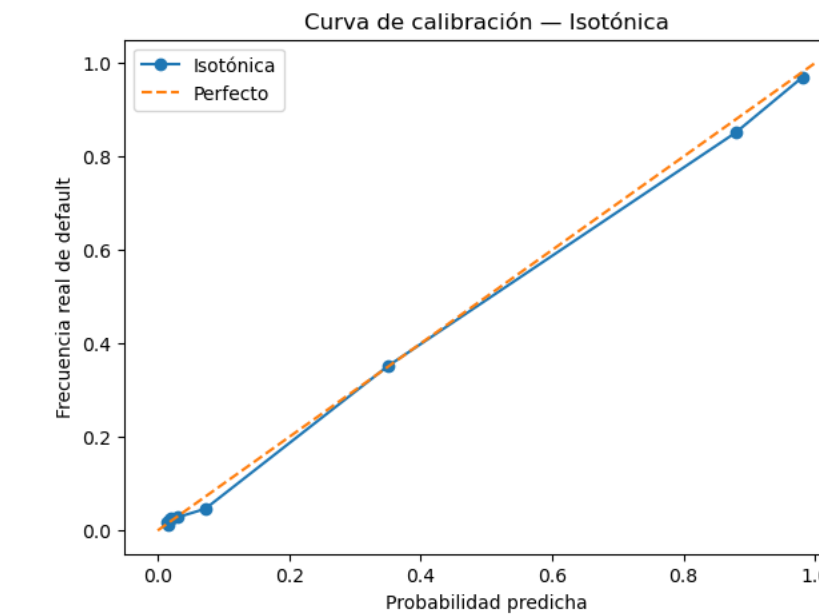
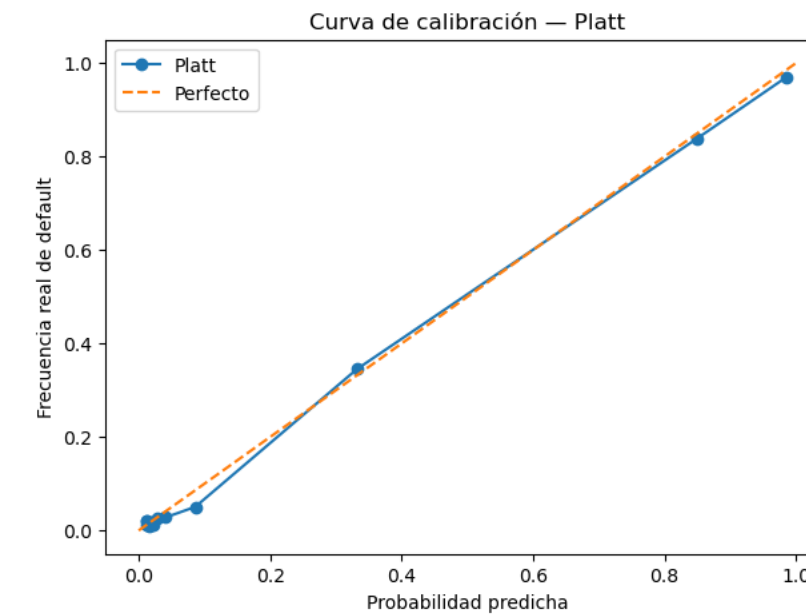
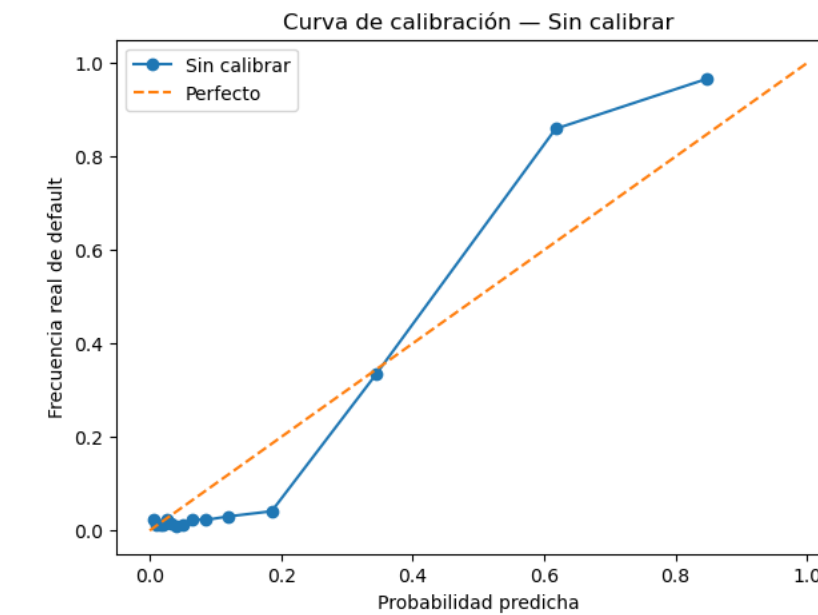
- Ejemplo de cartera de crédito con datos sintéticos con 20,000 observaciones.
- $PD = 0.15$, $LGD=0.45$.
- Modelo de bosques aleatorios para obtener PD.
- Optimizamos el área bajo la curva (ABC).



ABC: 0.949
Brier: 0.04572
LogLoss: 0.18248

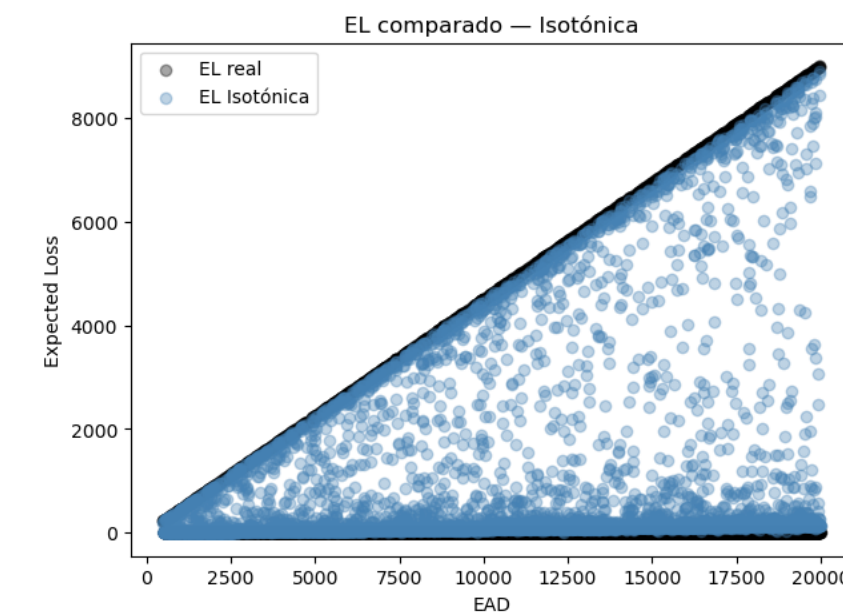
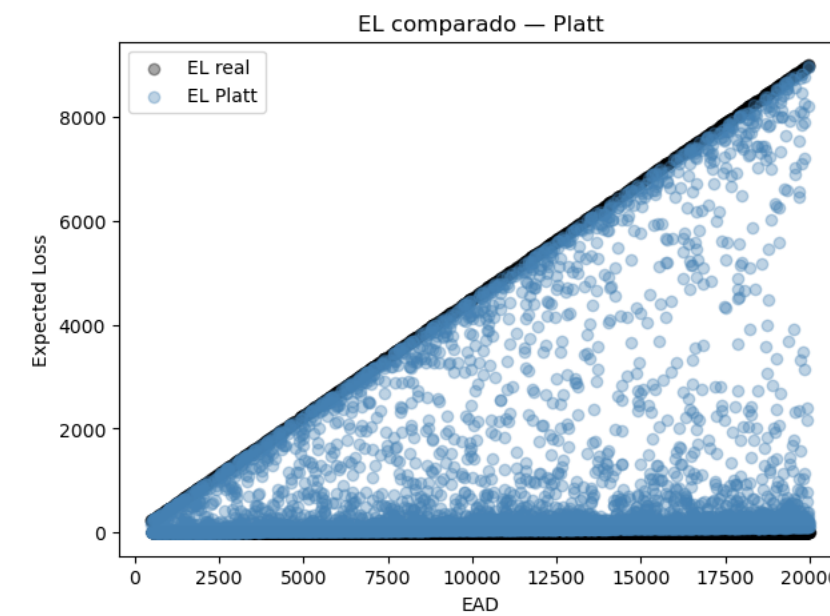
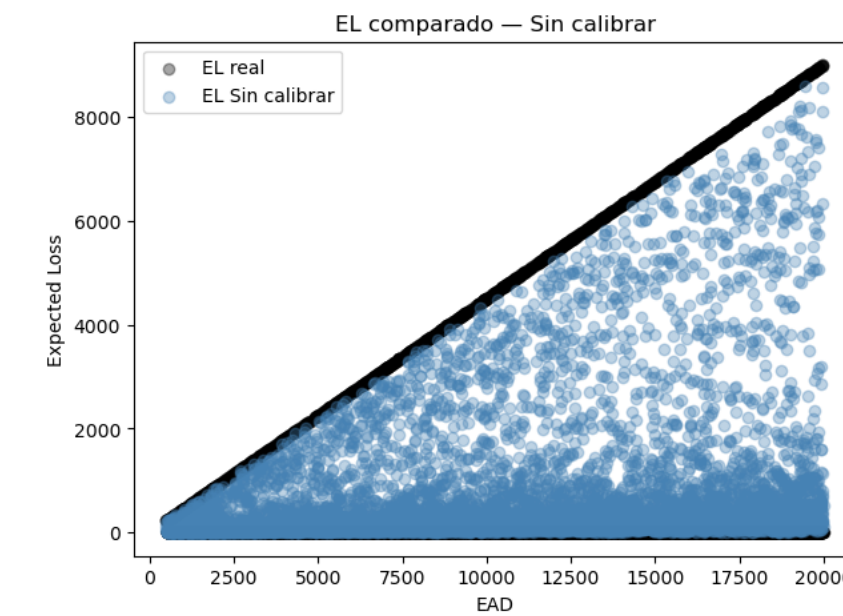
Materialización del riesgo de modelo

- El ABC mide el orden de las observaciones no la probabilidad de ocurrencia.
- Si optimizamos ABC podemos sobreestimar o subestimar la PD.
- El ABC es invariante a transformaciones monótonas de las probabilidades estimadas.
- Por eso podemos aplicar calibraciones con métodos como Platt (sigmoide) o Regresión Isotónica.



Materialización del riesgo de modelo

- En el gráfico de pérdida esperada comparamos EAD con EL, para ver alineación por cliente.
- Si la estimación de la PD está sesgada, la pérdida esperada estimada se desvía de la real.



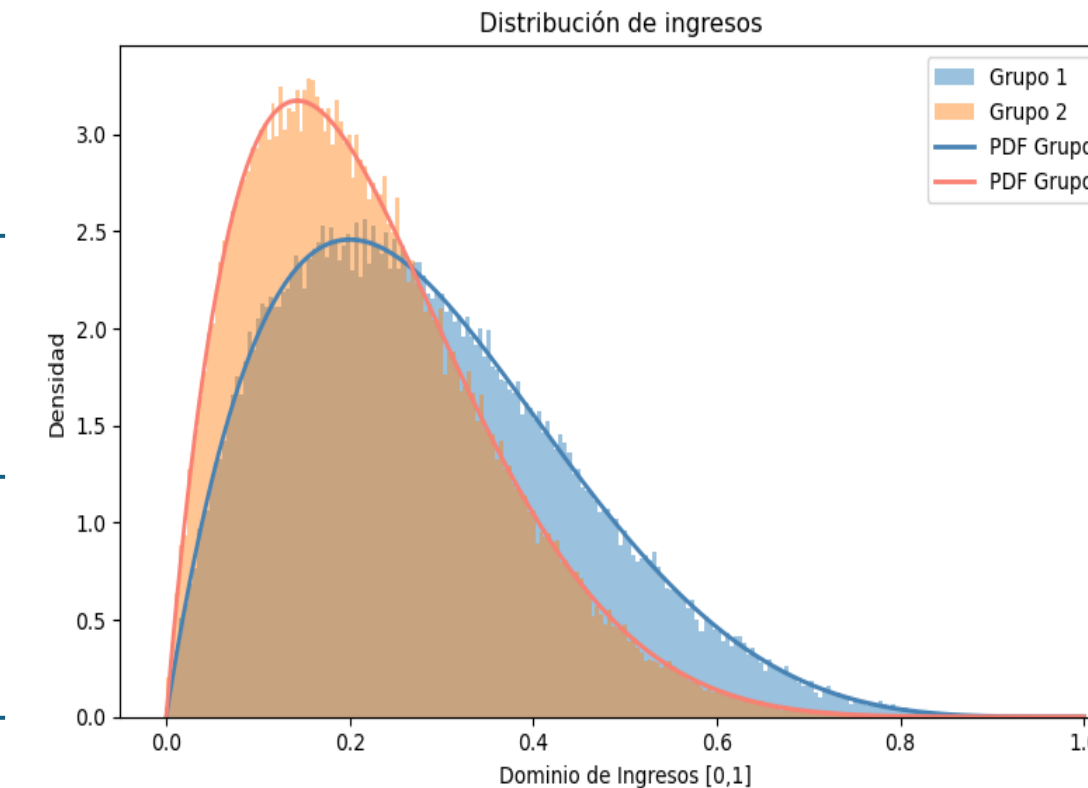
Desafíos de sesgo y equidad

La **discriminación algorítmica** sigue siendo una preocupación crítica en aplicaciones financieras de IA. La investigación muestra que los sistemas de IA pueden perpetuar sesgos históricos presentes en los datos de entrenamiento, produciendo resultados injustos.

El **sesgo en la puntuación crediticia** puede perjudicar sistemáticamente a ciertos grupos demográficos, incluso sin usar características protegidas de forma explícita.

Variables indirectas como códigos postales o historial laboral pueden señalar indirectamente características protegidas, dando lugar a resultados discriminatorios.

Los **sistemas de IA para préstamos** reciben un escrutinio especial, y los reguladores se centran en garantizar prácticas de préstamo equitativas.



Gobernanza de la IA

Gobernanza de la IA en el Sector Financiero



Gobernanza de la Inteligencia Artificial (IA): es el conjunto de **políticas, normas, y prácticas** que regulan el uso y desarrollo de la IA de manera **ética, transparente y responsable**.

Esta gobernanza busca asegurar que la implementación y operación de la IA estén **alineadas con los valores sociales**, los **derechos humanos** y la **normativa**.

La gobernanza de la IA es crucial para **fomentar la confianza pública** en estas tecnologías y para asegurar que su **impacto** en la sociedad sea **positivo y justo**.

La **transparencia** incluye la posibilidad de **explicar** los resultados de los modelos y herramientas utilizadas.

Principios de la OCDE sobre la IA

Crecimiento inclusivo, desarrollo sostenible y bienestar

La IA debe diseñarse para generar beneficios sociales amplios, promover la sostenibilidad y abordar las desigualdades.

Valores centrados en la persona y equidad

La IA debe respetar los derechos humanos, la dignidad y la democracia, y evitar la discriminación y el sesgo; se enfatiza el uso de datos diversos y representativos y auditorías periódicas.

Transparencia y explicabilidad

Los sistemas de IA deben ser transparentes en su lógica, capacidades e impactos, permitiendo que usuarios y partes interesadas entiendan y cuestionen los resultados.

Robustez, seguridad y fiabilidad

La IA debe ser técnicamente fiable, resistente frente a ataques o errores, probada en materia de seguridad y desplegada de forma segura con supervisión y controles adecuados.

Rendición de cuentas

Desarrolladores y responsables de decisión deben ser identificables y responsables, con mecanismos claros de escalado, auditoría y reparación frente a impactos u daños.

Declaración de gobernanza de la IA de BRICS

Principio	BRICS	Occidente
Autoridad central	Liderazgo ONU, multilateralismo inclusivo	UE/G7/OCDE, armonización técnica
Soberanía digital	Máxima, autonomía nacional	Compartida, alineación regulatoria
Ética y valores	Pluralista, sensible al contexto	Universal, prescriptiva
Estructura de mercado	Antimonopolio, acceso equitativo	Mercado abierto, controles por riesgo
Enfoque de desarrollo	Equidad, creación de capacidades	Derechos, protección del consumidor
Tecnología	Infraestructura soberana, independencia	Nube global, gigantes tecnológicos

Comparación de distintos marcos de Gobernanza de IA

Característica	NIST AI RMF	ISO/IEC 42001	Ley IA de la UE	Principios de la OCDE
Estado	Voluntario, basado en orientación	Obligatorio para certificación, estándar internacional	Regulación vinculante, aplicación estricta	Principios, directrices de alto nivel
Alcance	Gestión de riesgos para IA confiable	Sistema de gestión para gobernanza responsable de IA	Regula toda la IA, por niveles de riesgo, exigible	Ética, equidad, rendición de cuentas
Estructura	Cuatro funciones (Gobernar, Mapear, Medir, Gestionar)	PDCA: ciclo Plan-Hacer-Verificar-Actuar	Categorización por riesgo, protocolos de cumplimiento	Valores / recomendaciones
Certificable	No (alineación interna o voluntaria)	Sí (auditoría externa y certificación posible)	Sí (auditorías de cumplimiento, supervisión regulatoria)	No
Sanciones	Ninguna por incumplimiento	Requisitos para certificación; sanciones por declaraciones falsas	Multas administrativas por infracciones	Ninguna
Aplicabilidad global	Liderado por EE. UU., práctica reconocida globalmente	Estándar internacional, alcance global	De alcance UE y con impacto global	Miembros de la OCDE / internacional
Enfoque de riesgo	Autoevaluado, narrativo y con métricas	Gestión de riesgo estructurada, basada en evidencia	Niveles obligatorios, documentación, supervisión humana	Innovación responsable

IA explicable

¿Por qué IA Explicable?

- **Tendencia regulatoria** hacia una mayor transparencia de los resultados de las herramientas de IA.
- En muchas jurisdicciones los usuarios de servicios financieros tienen **vías legales** para **obtener** los resultados de estas herramientas y **cuestionarlos**.
- La **tendencia tecnológica** se dirige hacia la implementación de modelos y metodologías cada vez más complejos.
- Estos modelos son llamados **cajas negras** porque no permiten observar su funcionamiento de forma directa y sus resultados no son fácilmente explicables, aunque fueren certeros.
- Incluso si estos modelos estuvieren abiertos para inspección, su **tamaño y complejidad** hacen casi imposible su entendimiento.

IA Explicable

La **Inteligencia Artificial Explicable** (IAE o XAI).

- Métodos y técnicas que buscan hacer que los resultados de los sistemas de IA sean más **comprensibles** para las personas.
- Esto facilita que los usuarios comprendan y **confíen** en las decisiones o predicciones realizadas por estos sistemas.
- La IAE ayuda a cumplir con las regulaciones al proporcionar explicaciones claras de las decisiones tomadas por los sistemas de IA.

IA Explicable

- La IAE también aborda los requerimientos **éticos** relacionadas con la IA.
- La IAE ayuda a **identificar y mitigar sesgos**, para asegurar un uso justo y responsable de las tecnologías de IA.
- **Técnicas y Herramientas:** Se utilizan diversas técnicas para lograr la explicabilidad en la IA, que incluyen **puntuaciones de importancia** de características, **visualización de componentes internos** del modelo (como capas de redes neuronales) y **modelos sustitutos** que aproximan modelos complejos con otros más simples y interpretables.
- **Desafío:** Hacer que modelos complejos (e.g., redes neuronales profundas) sean interpretables a menudo requiere técnicas sofisticadas y no siempre se logran resultados idóneos.
- Usualmente existe un desbalance entre el **rendimiento** de un modelo y su **explicabilidad**.

IA Explicable - Metodologías

Herramientas Agnósticas a Modelos

- Aplicables a una variedad de modelos de IA (ML).
- Útiles para modelos simples y complejos.
- Ejemplos: LIME, CA-LIME, SHAP, Alibi.
- Ofrecen flexibilidad en el enfoque de “explicabilidad.”

Herramientas Específicas a Modelos

- Diseñadas para tipos específicos de modelos.
- Ejemplo: la herramienta What-If de Google en Tensorboard.
- Ofrecen información detallada para arquitecturas de modelo particulares.
- Pueden proporcionar explicaciones más matizadas para modelos específicos.

IA Explicable - Metodologías

Herramientas de Explicación Global

- **Alcance Amplio:** Proporcionan información sobre el comportamiento general y el proceso de toma de decisiones con el modelo.
- **Comprensión General:** Ofrecen una comprensión de cómo funciona el modelo en un nivel general.
- **Aplicaciones:** Útiles para obtener una visión general de la lógica del modelo y para el cumplimiento regulatorio donde sea necesario entender el modelo en su totalidad.

Herramientas de Explicación Local

- **Alcance Específico:** Se centran en explicar decisiones o predicciones específicas hechas por el modelo.
- **Perspectiva Detallada:** Ofrecen explicaciones para situaciones o rangos de datos específicos.
- **Aplicaciones:** Beneficiosas para entender o justificar predicciones individuales cuando los modelos son muy complejos para aplicar herramientas de explicación global.

IA Explicable - Metodologías

Herramientas de Explicación Intrínsecas

- **Transparencia Integrada:** La arquitectura del modelo es inherentemente interpretable y transparente.
- **Sencillez y Comprensibilidad:** A menudo involucran modelos más simples (como la regresión lineal o los árboles de decisión) que son fácilmente interpretables.
- **Limitación:** Pueden sacrificar complejidad y precisión en aras de la transparencia.

Herramientas de Explicación Post-Hoc

- **Aplicadas a Modelos Complejos:** Diseñadas para explicar modelos que no son inherentemente interpretables (como los modelos de aprendizaje profundo).
- **Análisis Después de los Hechos:** Proporcionan información sobre las decisiones del modelo después de que este ha sido entrenado y ha realizado predicciones.
- **Flexibilidad:** Pueden aplicarse a una variedad de modelos complejos, pero pueden ofrecer explicaciones menos intuitivas o fiables en comparación con los métodos intrínsecos.

Conclusiones

- La **adopción acelerada de IA en finanzas** (fraude, crédito, mercados) trae beneficios, pero también **riesgos sistémicos inéditos** que deben ser gestionados de manera proactiva.
- Los principales riesgos identificados son: **riesgo de modelo y sesgo algorítmico, vulnerabilidades operativas y cibernéticas, y dependencias de terceros altamente concentradas.**
- Los **casos de implementación fallida** evidencian que la falta de **gobernanza, transparencia.**
- Las **iniciativas regulatorias de la UE y EE. UU.** muestran enfoques divergentes (prescriptivo y legalmente vinculante vs. flexible y orientado a riesgos), lo que incrementa la complejidad para actores globales.
- La **gobernanza internacional** (ONU, OCDE, UNESCO, G7, ASEAN, BRICS) refleja un escenario multipolar: desde principios éticos universales hasta modelos soberanistas de gobernanza de IA.
- Los **estándares técnicos (ISO/IEC 42001, NIST AI RMF)** se perfilan como puntos de convergencia para la gestión responsable de la IA, facilitando la interoperabilidad regulatoria.

Muchas gracias