Trustworthy AI

7 pasos hacia una IA fiable

Ejercicio

Co-funded by the Erasmus+ Programme of the European Union

Nota sobre los derechos de autor:

7 PASOS HACIA UNA AI FIABLE

Ejercicio

En este ejercicio aprenderás a evaluar las implicaciones éticas de la IA, basándote en casos de IA de la vida real de una manera práctica, basada en problemas y sistemática. Aumentarás tus habilidades analíticas y de resolución de problemas analizando casos de uso siguiendo 7 pasos que integran las Directrices Éticas de la UE para una IA fiable como parte del proceso. Esto te ayudará a abordar las cuestiones éticas que rodean el uso de la IA de una manera holística. De este modo, aprenderas a:

- Desarrollar una mentalidad ética para su aplicación práctica: Identificar, aplicar y equilibrar elementos y dilemas éticos, morales y sociales.
- Aplicar la mentalidad ética al proceso de desarrollo y uso de la IA: Plantear soluciones conscientes para el problema y someterlo a un análisis ético.
- Justifica tu posición: Poner a prueba tus soluciones para comunicar y defender tu posición de forma convincente.

¿CÓMO FUNCIONA?

Tu profesor/instructor te ha dado a ti o a tu grupo un problema o caso para analizar. Este ejercicio te permite evaluar y valorar el caso desde la perspectiva de las Directrices éticas para la IA fiable, siguiendo 7 pasos:

1. **Exponer el problema**
2. **Identificar los factores y las partes interesadas pertinentes**
3. **Listar soluciones**
4. **Evaluar las soluciones con respecto a los 7 requisitos para una IA fiable**
5. **Someter las soluciones a “pruebas de daño”**
6. **Elejir la solución**
7. **Revisar y reflexionar**

PASO 1

Exponga el problema

- Describa el caso
- Identifique y defina claramente el problema
- Compruebe los hechos

Resuma brevemente el caso o el problema y pregúntese "¿Cuál es exactamente el problema que hay que resolver? "Puedes hacerlo con preguntas como "¿Veo un conflicto de intereses? "o "¿Hay partes de este caso que me incomodan? ". Algunos problemas desaparecen cuando se examina la situación más de cerca. Investiga (en línea) los detalles relacionados con el problema que has identificado. Por último, trata de ver el problema desde una perspectiva diferente, por ejemplo, desde la perspectiva del consumidor en lugar de la de la empresa, o desde la perspectiva del trabajador.

Para ayudarle a entender mejor los pasos, cada paso tendrá explicaciones adicionales basadas en un **breve ejemplo** de un caso de uso:

Rastrear y prevenir el fraude bancario – Paso 1

Empresas como MasterCard o los bancos tienen que enfrentarse a personas que intentan cometer un fraude. El fraude bancario implica la obtención ilegal de dinero, activos u otros bienes en poder de una institución financiera. También incluye hacerse pasar por un banco u otra institución para estafar a personas inocentes.

Problema: los bancos sufren las consecuencias de las personas y organizaciones que cometen fraudes bancarios. El fraude bancario puede detectarse a veces observando los patrones de las transacciones. Sin embargo, es mucho trabajo comprobar todas las transacciones y no siempre estamos seguros de lo que buscamos. ¿Puede la IA ayudarnos a luchar contra el fraude bancario reconociendo los patrones de las transacciones sospechosas?

¿Sigue siendo el mismo problema? ¿Ha empeorado el problema?

PASO 2

Identifique factores y las partes interesadas pertinentes

Cuando identifiques a las partes interesadas, intenta pensar en todas las personas implicadas, incluidas las posibles organizaciones u organismos institucionales. Otros factores relevantes que pueden ser importantes para el problema y sus soluciones son los códigos profesionales, las leyes y reglamentos y otras limitaciones prácticas.

- Identificar a las partes interesadas, tanto internas como externas
- Identificar otros factores que influyen en el problema

Rastrear y prevenir el fraude bancario – Paso 2

Las partes interesadas incluyen: las personas u organizaciones titulares de cuentas bancarias, las personas de los bancos responsables de la seguridad y la prevención de la delincuencia, los expertos en IA, los analistas de datos, los responsables jurídicos, los gestores de clientes, los directivos, los miembros del consejo de administración, etc. También personas ajenas al banco, como supervisores bancarios/financieros u organizaciones que se ocupan de la prevención de la delincuencia como las fuerzas del orden.

Los factores relevantes podrían incluir: Regulación (por ejemplo, leyes y regulaciones financieras, ley de protección de dator (GDPR), regulación relacionada con la IA, derechos fundamentales, derecho penal), presión política o social, la competencia, los resultados financieros, etc.

PASO 3

Liste soluciones

Piensa entre 1-3 posibles soluciones o enfoques del problema. Al menos una solución debe estar basada en la IA (también puede ser una solución de IA predeterminada sobre la que quieras reflexionar). Si enumeras más de una solución, al menos una debe ser no basada en IA. Describe cómo las soluciones resolverían el problema e incluye cómo pueden realizarse tus soluciones. En este paso puedes seguir siendo imaginativo.

Rastrear y prevenir el fraude bancario – Paso 3

Para este ejemplo hemos limitado el problema al reconocimiento de patrones de transacciones dudosas y planteamos una solución de IA para discutir. Esta solución consiste en un modelo de deep learning que aprende a reconocer posibles patrones de transacciones fraudulentas, basándose en datos históricos. Los datos tienen que ser una mezcla realista de transacciones normales y transacciones fraudulentas tanto de ciudadanos normales como de organizaciones más grandes. El modelo aprenderá a detectar las transacciones fraudulentas y podrá aplicarse en tiempo real y bloquear automáticamente las cuentas cuando se detecten supuestas transacciones fraudulentas. Necesitaríamos datos de alta calidad de los bancos para entrenar este

PASO 4

Evalúe su(s) solución(es) de con respecto a las Directrices éticas para una IA fiable

En este paso evaluará **cada una de sus soluciones de IA** (hágalo una por una) con respecto a las Directrices éticas para una IA fiable[1] y **configure, adapte o sustituya** su solución en consecuencia. Las Directrices describen 7 requisitos que es importante tener en cuenta al contemplar el desarrollo, la implantación o el uso de la IA.

Los 7 requisitos para una IA de confianza[2]

1. Agencia humana y supervisión
2. Robustez técnica y seguridad
3. Privacidad y gobernanza de los datos
4. Transparencia
5. Diversidad, no discriminacion y equidad
6. Bienestar social y medioambiental
7. Rendición de cuentas

No todos los requisitos serán igual de relevantes para tu caso y/o solución de IA. Sin embargo, tómate un momento para pensar cómo un requisito específico puede seguir siendo relevante, ya que algunas conexiones pueden no ser tan obvias como otras. Te animamos a volver a pensar y reformular tu(s) solución(es) de IA según los problemas que puedas encontrar o las conclusiones que puedas sacar al ejecutar este paso.

Las siguientes secciones contienen una explicación rápida de cada requisito y algunas preguntas que podría plantearse en relación con las soluciones propuestas. No es necesario que analice cada pregunta en profundidad. Elija las preguntas que considere más relevantes para su caso y su(s) solución(es) de IA. Le animamos a que lea las que contienen una explicación exhaustiva de todos los requisitos.

Requisito nº 1 - Agencia humana y supervisión

Los sistemas de IA deben apoyar la autonomía humana y permitirles tomar decisiones con conocimiento de cla ausa. Para lograrlo, los sistemas de IA deben actuar como facilitadores de una sociedad democrática y equitativa, apoyando la acción del usuario, fomentando los derechos fundamentales y permitiendo la supervisión humana. La supervisión puede lograrse a través de mecanismos de gobernanza como el enfoque humano en el ciclo (human-in-the-loop), el humano sobre el ciclo (human-on-the-loop) o el humano al mando (human-in-command). HITL se refiere a la capacidad de intervención humana en cada ciclo de decisión del sistema, que en muchos casos no es posible ni deseable. HOTL se refiere a la capacidad de intervención humana durante el ciclo de diseño del sistema y la supervisión del funcionamiento del mismo. HIC se refiere a la capacidad de supervisar la actividad global del sistema de IA (incluyendo su impacto económico, social, legal y ético más amplio) y la capacidad de decidir cuándo y cómo utilizar el sistema en cualquier situación. Esto puede incluir la decisión de no utilizar un sistema de IA en una situación concreta.

[1] Ethics Guidelines for Trustworthy AI, Grupo de Expertos de Alto Nivel en IA, 2019: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[2] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

- Derechos fundamentales:
  - ¿Puede identificar cualquier impacto negativo sobre los derechos fundamentales que su solución pueda tener?
  - ¿Puede identificar y documentar las posibles compensaciones entre los diferentes principios y derechos?

- Agencia humana:
  - ¿ Mejora o aumenta el sistema de IA las capacidades humanas?
  - ¿Está el sistema de IA centrado en el ser humano?: ¿deja una oportunidad significativa para la elección humana?
  - ¿Permite a los individuos tener más control sobre sus vidas o limita su libertad y autonomía?

- Supervisión humana:
  - ¿Puede describir el nivel de control o participación humana en su solución?
  - ¿Ha pensado en un "humano en el ciclo" o un "humano en el ciclo" o un "humano al mando"?
  - ¿Tiene la persona que realiza la supervisión las competencias, los conocimientos y la autoridad pertinentes?

Rastrear y prevenir el fraude bancario

Esta solución de IA podría complementar las capacidades humanas. Aprenderá patrones que un humano podría pasar por alto y podrá aplicar sus conocimientos automáticamente a gran escala.

Sin embargo, la solución de la IA también podría limitar la autonomía humana, cuando, por ejemplo, puede conducir al "sesgo de confirmación" del individuo que trabaja con el sistema, o al efecto "el ordenador dice no".

Los derechos fundamentales afectados podrían ser: el derecho a la privacidad (si hay un uso no autorizado de los datos personales o afecta a nuestra libertad de gastar nuestro dinero como queramos); el derecho a la sospecha razonable (si la detección del fraude se basa en características que alguien comparte con otros, en lugar de en una sospecha real); el derecho a la defensa (si el modelo es una caja negra y resulta imposible explicar el fundamento de la decisión); el derecho a la no discriminacion (si el modelo resulta producir resultados sesgados).

Nuestra solución carece por ahora de supervisión o participación humana. Esto podría solucionarse añadiendo una capa de control que compruebe manualmente las decisiones de los modelos en busca de errores e impactos en los derechos humanos. También habría que hacer posible que determinadas personas estuvieran capacitadas para deshacer decisiones o para impedir que el modelo funcione/compruebe una determinada cuenta bancaria.

❄ Reflexione: Si tu solución no cumple este requisito, ¿hay alguna forma de adaptarla en consecuencia?

Requisito nº 2 – Robustez técnica y seguridad

La robustez técnica requiere que los sistemas de IA se desarrollen con un enfoque preventivo de los riesgos y de manera que se comporten de forma fiable según lo previsto, minimizando los daños no intencionados e inesperados, y evitando los daños inaceptables. Esto también debe aplicarse a los posibles cambios en el entorno operativo del sistema o a la presencia de otros agentes (humanos y artificiales) que puedan interactuar con ellos de forma adversa. Además, debe garantizarse la integridad física y mental de los seres humanos. En definitiva, este principio hace hincapié en el establecimiento de un equilibrio entre la robustez técnica y las limitaciones éticas, pero también en la capacidad de evaluar las posibles compensaciones entre ambas.

- Resistencia a los ataques y seguridad:
  - ¿Puede identificar alguna forma potencial de ataque a la que el sistema de IA pueda ser vulnerable?
- Plan de emergencia y seguridad general:
  - ¿Existe una posibilidad probable de que el sistema de IA pueda causar daños o perjuicios a los usuarios o a terceros?
- Precisión:
  - ¿Cómo se puede medir y garantizar la precisión del sistema?

❇ Reflexione: Si tu solución no cumple este requisito, ¿hay alguna forma de adaptarla en consecuencia?

Requisito nº 3 - Privacidad y gobernanza de los datos

La privacidad es un derecho fundamental especialmente afectado por los sistemas de IA, ya que los datos son el combustible de un sistema de IA. A menudo, estos datos son sobre personas muy reales y no cuidar bien los datos tiene efectos muy reales sobre las personas. Por lo tanto, la prevención del daño a la privacidad debe ser una prioridad. Este requisito requiere, lógicamente, una gobernanza de datos adecuada. Esto abarca la calidad e integridad de los datos utilizados, su relevancia considerando el ámbito en el que se desplegarán los sistemas de IA, sus protocolos de acceso y la capacidad de procesar los datos de manera que se proteja la privacidad.

- Respeto a la privacidad y a la protección de datos:
  - ¿Existen formas de desarrollar el sistema de IA o de entrenar el modelo sin o con un uso mínimo de datos potencialmente sensibles o personales?
- Calidad e integridad de los datos:
  - ¿Se le ocurren mecanismos de supervisión para la recogida, el almacenamiento, el tratamiento y el uso de los datos?
- Acceso a los datos:
  - ¿Qué protocolos, procesos y procedimientos se le ocurren para gestionar y garantizar la correcta gobernanza de los datos?

- ¿Quién debe estar autorizado a acceder a los datos de los usuarios y en qué circunstancias? (Piense en las cualificaciones y los conocimientos/competencias para comprender los detalles de la política de protección de datos).

Rastrear y prevenir el fraude bancario

Para entrenar este modelo tenemos que utilizar datos reales para encontrar patrones que podamos utilizar en nuevas transacciones. Sin embargo, podemos tomar medidas y mecanismos de supervisión que ayuden a respetar la privacidad y mantener los datos seguros. Todos los datos serán anónimos y estarán encriptados. Se almacenarán de forma segura y sólo serán accesibles a petición de un grupo selecto de personas con razones válidas para investigarlos. No se eliminarán inmediatamente después de la formación, por si necesitamos encontrar errores en los datos en caso de que el sistema funcione mal. Sin embargo, debemos ser críticos con el aspecto de los "datos indirectos", en los que datos no personales o sensibles, datos personales anonimizados, pueden, en combinación con otros datos, proporcionar una información indirecta sobre personas.

* Reflexione: Si tu solución no cumple este requisito, ¿hay alguna forma de adaptarla en consecuencia?

Requisito nº 4 - Transparencia

Las cuestiones sobre a la "transparencia" varían. Una primera cuestión consiste en reconocer los sistemas transparentes y en qué se diferencian de los opacos. En segundo lugar, la transparencia consiste en poder explicar la decisión o el razonamiento de un sistema. También abarca la transparencia de los elementos relevantes para los sistemas de IA. Esto incluye los datos recogidos, la formación y el funcionamiento del sistema, y los modelos de negocio pertinentes. Para ello, el uso de los datos y otras decisiones tomadas en el proceso de diseño deben documentarse y comunicarse adecuadamente.

- Trazabilidad:
  - ¿Qué mecanismos podría establecer que faciliten la auditabilidad del sistema, como garantizar la trazabilidad y el registro de sus procesos y resultados?
  - ¿Ha revisado los resultados o las decisiones tomadas por el sistema, así como otras posibles decisiones que se derivarían de diferentes casos (por ejemplo, para otros subgrupos de usuarios)?

- Explicabilidad:
  - ¿Puede explicar por qué el sistema tomará una determinada decisión de forma comprensible para todos los usuarios?

- Comunicación:
  - ¿Qué mecanismos se pueden poner en marcha para informar a los usuarios (finales) sobre las razones y los criterios de los resultados del sistema?
  - ¿Cuál es el objetivo exacto de su sistema de IA y quién o qué puede beneficiarse de él?

- ¿Puede especificar los escenarios de uso del sistema y comunicarlos claramente para garantizar que el sistema sea comprensible y adecuado para el público al que va dirigido?

❇ Reflexione: Si tu solución no cumple este requisito, ¿hay alguna forma de adaptarla en consecuencia?

Requisito nº 5 - Diversidad, no discriminación y equidad

Tendemos a pensar que los sistemas de IA son objetivos y sin prejuicios, ya que se basan en datos y lógica. Sin embargo, los datos imparciales no existen. La finalidad de la recogida de datos, la persona que los mide y la decisión de lo que se va a medir es siempre una elección humana, por lo que la IA y los datos siempre llevan nuestra subjetividad. Por lo tanto, para lograr una IA digna de confianza, debemos permitir la inclusión y la diversidad a lo largo de todo el ciclo de vida del sistema de IA. Además de tener en cuenta y hacer partícipes a todas las partes interesadas en todo el proceso, esto también implica garantizar la igualdad de acceso mediante procesos de diseño inclusivos, así como la igualdad de trato.

- Evitar el sesgo injusto:
  - Evaluar y reconocer las posibles limitaciones derivadas de la composición de los conjuntos de datos utilizados.
  - Evaluar si pudiese haber personas o grupos que pudieran verse afectados de forma desproporcionada por las consecuencias negativas.

- Accesibilidad y diseño universal:
  - Evaluar si el sistema de IA es utilizable por personas con necesidades especiales o discapacidades o en riesgo de exclusión. ¿Cómo se puede diseñar esto en el sistema y cómo se puede verificar?

- Participación de las partes interesadas:
  - ¿Se te ocurren ideas para incluir la participación de las diferentes partes interesadas en el desarrollo y uso del sistema de IA?

Rastrear y prevenir el fraude bancario

Debemos asegurarnos de que el sistema no produce resultados sesgados. El sesgo (accidental) presente en el conjunto de datos no debe repetirse ni amplificarse. Para ello, debemos asegurarnos de que determinados grupos de personas no sean tratados de forma injusta o discriminatoria. Debemos considerar cuidadosamente los conjuntos de datos que utilizamos para entrenar nuestro modelo para asegurarnos de que los sesgos del pasado no se repiten ni amplifican. También debemos asegurarnos de no introducir características o indicadores sesgados durante el proceso de diseño del modelo. También debemos comprobar los resultados del sistema a lo largo de todo su ciclo de vida para asegurarnos de que sus decisiones siguen siendo imparciales.

El proceso de desarrollo debe contar con la participación de las partes interesadas, tanto de dentro como de fuera de la organización. Entre ellos se encuentran la dirección, los responsables jurídicos, el departamento comercial, el front office y los asesores, pero también la autoridad bancaria, los expertos en fraude y la autoridad de consumo.

❇ Reflexione: Si tu solución no cumple este requisito, ¿hay alguna forma de adaptarla en consecuencia?

Requisito nº 6 - Bienestar social y medioambiental

El bienestar del medio ambiente y de la sociedad en su conjunto debe considerarse una "parte interesada" durante todo el ciclo de vida del sistema de IA. Este requisito incluye el impacto en la democracia y el discurso público y el fomento de la sostenibilidad y la responsabilidad ecológica. Implica tanto la investigación de soluciones de IA que aborden el cambio climático u otras preocupaciones sociales, como ser conscientes de la huella ecológica de la formación y el despliegue de un sistema de IA. La interdisciplinariedad es un factor importante para cumplir este requisito, así como la realización de evaluaciones de impacto.

- Impacto social:
  - ¿Se ha asegurado de que se conocen bien las repercusiones sociales del sistema de IA? Por ejemplo, ¿se ha evaluado si existe riesgos de pérdida de empleo o de descualificación de la mano de obra? ¿Qué medidas se han tomado para contrarrestar estos riesgos?

- Sociedad y democracia:
  - Evaluar si la lógica de la IA puede simplificar y polarizar el discurso público.
  - Evaluar si el sistema de IA podría utilizarse para manipular o confundir a las personas.

- IA sostenible y respetuosa con el medio ambiente:
  - ¿Qué mecanismos podría establecer para medir el impacto medioambiental del desarrollo, la implantación y el uso del sistema de IA? (Por ejemplo, piensa en el tipo de energía que utilizan los centros de datos).
  - ¿Qué medidas se le ocurren para reducir el impacto medioambiental del ciclo de vida de su sistema de IA?

❄ Reflexione: Si tu solución no cumple este requisito, ¿hay alguna forma de adaptarla en consecuencia?

Requisito nº 7 - Responsabilidad

Este requisito exige que se establezcan mecanismos que garanticen la responsabilidad y la rendición de cuentas sobre los sistemas de IA y sus resultados, antes, durante y después de su desarrollo, despliegue y uso. Esto requiere una auditoría y un registro adecuados, así como marcos legales de responsabilidad. Los desarrolladores e implantadores de sistemas de IA deben ser capaces de demostrar la minimización de los efectos negativos.

- La auditabilidad:
  - ¿Qué mecanismos podría establecer que faciliten la auditabilidad del sistema, como garantizar la trazabilidad y el registro de sus procesos y resultados?

- Minimizar e informar del impacto negativo:
  - Realice una evaluación del riesgo o del impacto de su sistema de IA, teniendo en cuenta las diferentes partes interesadas que se ven afectadas (in)directamente.
  - Piense en los procesos que puede establecer para que terceras partes (por ejemplo, proveedores, consumidores, distribuidores) o trabajadores informen de posibles vulnerabilidades, riesgos o sesgos en el sistema de IA.

- Documentar las compensaciones:
  - ¿Cuáles son los intereses y valores relevantes a los que afecta el sistema de IA?
  - ¿Cuáles son las posibles compensaciones entre ellos?
  - ¿Cómo se decide este tipo de intercambios? Documente el comercio de la decisión.

- Posibilidad de recurrir a la justicia:
  - ¿Qué mecanismos pueden establecerse para permitir la reparación en caso de que se produzca algún daño o impacto adverso?

Rastrear y prevenir el fraude bancario

Tenemos que pensar en un mecanismo que nos muestre por qué un determinado patrón de transacción se considera dudoso. El sistema debe tener alguna forma de explicabilidad, esto ayudará a su auditabilidad. Tenemos que establecer un proceso que integre un mecanismo de evaluación de riesgos y de información en caso de efectos adversos. Tenemos que establecer una estructura de responsabilidad que documente todo el proceso de diseño, desarrollo y uso del sistema. Tenemos que garantizar que los ciudadanos puedan pedir explicaciones y compensaciones cuando se enfrenten a un resultado del sistema con el que no estén de acuerdo.

PASO 5

Someta su(s) solución(es) de IA a pruebas

En este paso pondrás a prueba tu(s) solución(es) de IA utilizando diferentes pruebas. Puedes responder a las preguntas (relevantes) que te proporciona cada prueba para ayudarte a decidir si tu solución pasa la prueba.

- Prueba de daño - ¿Hace su solución menos daño que las otras soluciones?
  - ¿Tiene una solución algunos criterios en cuenta mejor que otros?
  - ¿Es posible combinar lo mejor de distintas soluciones en una sola?
  - ¿Es necesaria su solución para resolver el problema y, se limita a resolverlo?
  - ¿Presta su solución atención a los grupos vulnerables y garantiza que no sean tratados con parcialidad?

- Prueba de publicidad: ¿Querría que mi solución se publicara en el periódico?
  - ¿Qué tipo de preguntas e inquietudes del público plantearía su solución?
  - ¿Tiene la solución en cuenta a todas las partes interesadas? ¿De qué manera favorece o perjudica a algunas partes interesadas en detrimento de otras?
  - ¿Tiene la solución algún impacto social más amplio?
  - ¿Hasta qué punto es democrática la solución? Considere el efecto sobre la agencia y el poder de los ciudadanos.

- Prueba de defendibilidad - ¿Podría defender la elección de esta solución ante un comité gubernamental, un comité de mis compañeros o mis padres?
  - ¿Es la solución lícita (permitida legalmente)?
  - ¿Socava los derechos humanos (por ejemplo, la vida, la seguridad, la privacidad, la no discriminación, la libertad de información, la libertad de manifestación, un lugar de trabajo sano y seguro, un juicio justo, etc.)?
  - ¿Cuál es el razonamiento para elegir esta solución en lugar de otras? ¿Puedo defender ese razonamiento?

- Prueba de virtudes: ¿Cómo te refleja tu solución?
  - ¿Qué tipo de creencias, suposiciones, actitudes y valores refleja su solución?
  - ¿Qué tipo de creencias, suposiciones, actitudes y valores refleja el proceso de selección de su solución?
  - ¿Qué tipo de valores e ideales quiere promover con su solución?
  - ¿Fue la solución elegida de forma independiente? ¿Sirve a los intereses de alguien?

- Prueba profesional - ¿Qué podría decir un comité de ética sobre su solución?
  - ¿Promueve la ética del sector?
  - ¿Qué dirían sus colegas, compañeros de clase y colegas sobre la alineación ética de su solución?
  - ¿Qué diría su superior cuando describa el problema y sugiera esta solución?

Rastrear y prevenir el fraude bancario – Paso 5

Para este ejemplo realizamos la prueba de publicidad: Nos sentiremos cómodos de inmediato si nuestra solución se publica en el periódico. Por un lado, podría ayudar a prevenir la delincuencia y hacer más seguras las operaciones bancarias, pero por otro, es probable que el público se preocupe por su intimidad, por si entrarán en algún tipo de "lista negra" y por los elementos que informarían de los resultados de mi solución. Por ejemplo, podrían preguntarse si se les juzgará en función del tipo de compras que realicen, o dónde compren cosas, o a qué horas, o si los detalles de sus compras siguen siendo privados y cómo pueden evitar o combatir los resultados erróneos.

* Repita los pasos 4 y 5 para cada una de sus soluciones de IA.

PASO 6

Elija

En este paso se hace una elección tentativa basada en los pasos 4 y 5 y de sus 3 (o más) soluciones (tanto de IA como de no IA) se elige la solución que es -o tiene el potencial de ser- la más confiable. En este paso, considere lo siguiente:

- ¿Cuál de las soluciones de IA que cumple más requisitos en el paso 4?
- ¿Qué solución de IA supera el mayor número de pruebas del paso 5?
- ¿Cuál de las soluciones no relacionadas con la IA resuelve el problema?
- ¿Tienen las soluciones sin IA menos impacto ético que las soluciones con IA?

PASO 7

Revisar y reflexionar

Rastrear y prevenir el fraude bancario – Paso 7

El seguimiento de los patrones de transacción es bastante invasivo y podría no ser necesario si tuviéramos una mejor seguridad. Podría ser más razonable encontrar una solución de seguridad. También podríamos pensar en una forma de probar si esta solución sería realmente eficaz para limitar el fraude bancario, antes de considerarla. Podríamos comparar estas pruebas con un enfoque actuarial en el que aceptamos un cierto riesgo de fraude como parte de nuestros riesgos empresariales.

Revise suproceso de los pasos 1 a 6. ¿Cómo ve los pasos anteriores ahora que ha pensado en sus soluciones? ¿Ve el problema de otra manera? ¿Hay soluciones diferentes que podría haber pensado? ¿Hay soluciones mejores que no impliquen sistemas de IA u otras formas de tecnología?

Piense también en qué podría hacer menos probable que tuviera que tomar una decisión así en el futuro. ¿Hay precauciones que pueda tomar? ¿Hay formas de conseguir más apoyo a la hora de gestionar este problema? ¿Hay aspectos organizativos que deban cambiar?

Trustworthy AI

Co-funded by the
Erasmus+ Programme
of the European Union