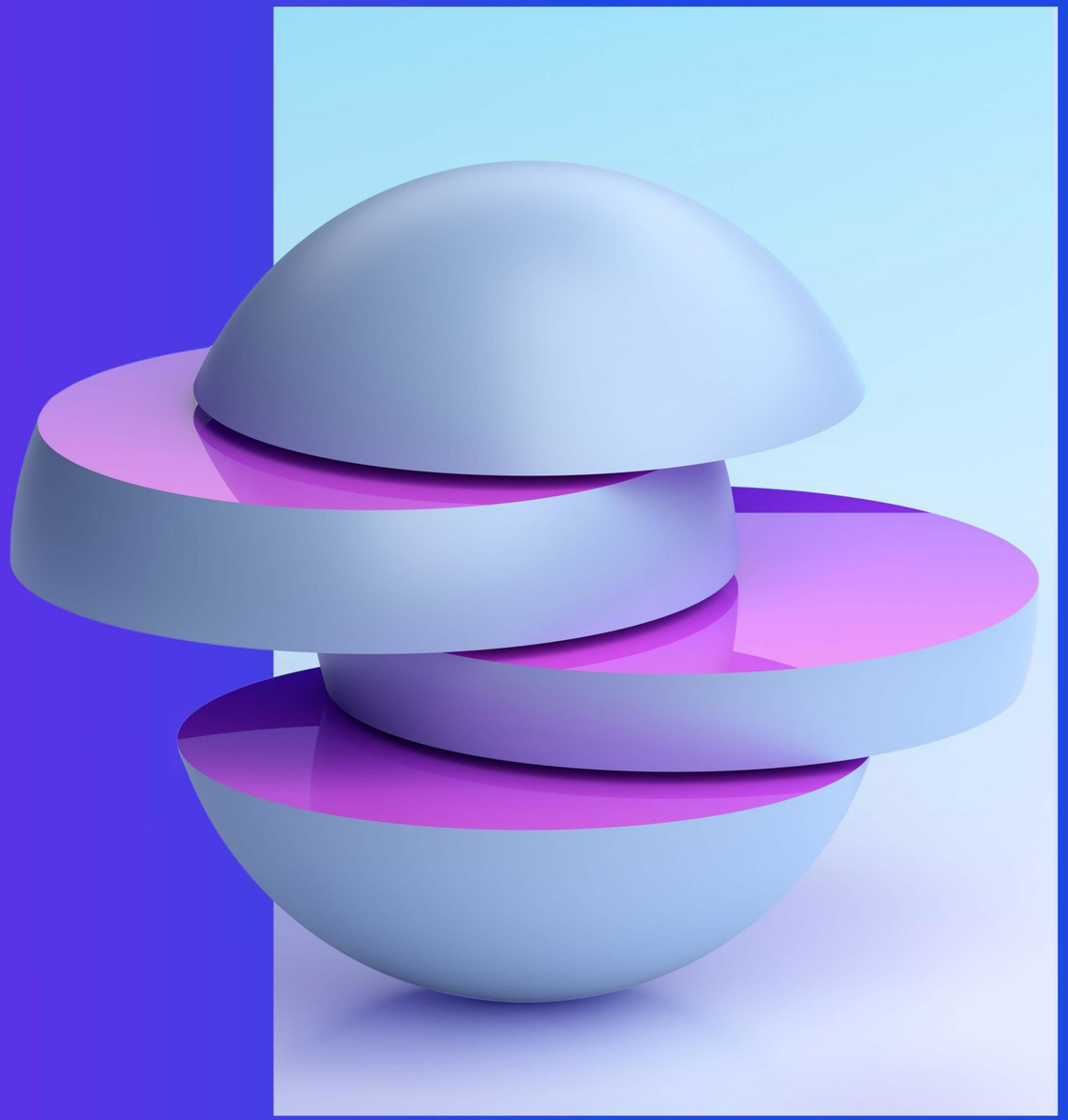
Privacidad en el nuevo mundo de la IA

Cómo generar confianza en la IA a través de la privacidad.



Contenido

Privacidad de los datos y confianza en la IA: una promesa

La Inteligencia Artificial (IA) promete transformar nuestras vidas, ayudándonos a ser más eficientes, productivos, sanos e innovadores.

Esta tecnología ya se utiliza en los sectores privado y público, aprovechando el poder de los datos para mejorar las previsiones, mejorar productos y servicios, reducir costes y liberar a los trabajadores de tareas rutinarias.

En el sector sanitario, los médicos pueden predecir con mayor precisión y rapidez los riesgos para la salud y llevar a cabo tratamientos complejos con mayor eficacia. En la minería, los robots dotados de IA están realizando tareas peligrosas como extracción de carbón, exploración marina y también ayudando en operaciones de rescate.

En la banca comercial, la IA y el ML ayudan a los equipos de ventas y marketing a identificar clientes potenciales, predecir sus necesidades y su propensión a comprar. También permiten fijar precios dinámicos para micro-segmentos y automatizar los procesos de toma de decisiones, los conjuntos de reglas de crédito y las excepciones. En consumo y venta al por menor, la IA está ayudando a predecir y analizar tendencias, crear modelos virtuales que pueden mostrar conjuntos, anticipar las necesidades de los clientes y ayudarles a disfrutar de una experiencia de compra más personalizada.

Según el Global Tech Report 2023 de KPMG, los líderes tecnológicos identifican la IA y el aprendizaje automático como la tecnología más importante para alcanzar las ambiciones a corto plazo. En la encuesta mundial sobre la confianza en la inteligencia artificial a más de 17.000 personas de todo el mundo, el 85% cree que la IA puede aportar una serie de beneficios. Sin embargo, como ocurre con cualquier tecnología emergente, existen riesgos. La misma encuesta revela que el 61% de las personas desconfían de los sistemas de IA, y sólo la mitad cree que los beneficios de la IA superan a los riesgos. El uso generalizado y no regulado de esta tecnología suscita inquietud por su impacto en los derechos humanos y la privacidad de las personas.

Esto es especialmente cierto en el caso de la IA generativa (GenAI), que utiliza potentes modelos básicos que se entrenan con cantidades masivas de datos no etiquetados. Los líderes de la IA han publicado cartas abiertas en las que solicitan una pausa en el desarrollo de la GenAI, instando a los legisladores a que garanticen su uso en el futuro mediante barreras de protección. Algunos de los riesgos citados incluyen un diseño defectuoso, una lógica sesgada, errores de codificación, vulnerabilidades de seguridad y, lo que es más importante, juicios que discriminen a individuos o grupos (después de todo, los datos utilizados fueron creados originalmente por humanos y pueden reflejar prejuicios existentes).

Además, los modelos de IA pueden generar resultados inexactos que den lugar a noticias falsas o desinformación. Los algoritmos también son impredecibles, complejos y difíciles de explicar. Dada su naturaleza patentada, carecen de transparencia y generan resultados basados en el procesamiento de datos a gran escala procedentes de Internet, lo que aumenta el riesgo de filtración de datos confidenciales y de violación de información personal protegida legalmente. Las leyes internacionales de protección de la intimidad se aplican a la recopilación de datos en todas las fases del ciclo de vida de una IA, por lo que no es de extrañar que la extracción y recolección de datos de IA haya atraído el escrutinio de las Autoridades Europeas de Protección de Datos (AEPD) y los organismos de control de la privacidad de todo el mundo inician investigaciones sobre la legalidad de las actividades de tratamiento de datos relacionadas con GenAI. Este documento investiga las implicaciones para la privacidad de la adopción generalizada de la IA. Pretende desvelar lo que esto significa para las empresas y esboza los pasos clave que las organizaciones pueden dar para utilizar la IA de forma responsable. Al mantenerse informadas sobre las implicaciones para la privacidad de la adopción de la IA y tomar medidas proactivas para mitigar los riesgos, las empresas pueden aprovechar el poder de esta tecnología y, al mismo tiempo, salvaguardar la privacidad de las personas.

[1] KPMG Global Tech Report 2023, KPMG International, 2023.
[2] Trust in artificial intelligence, The University of Queensland and KPMG Australia, 2023.
[3] Ibid.
[4] KPMG Global Tech Report 2023, KPMG International, 2023.
[5] Pause Giant AI Experiments: An Open Letter, Future of Life Institute, March 22, 2023.

Principios clave para privacidad de la IA

La integración de la privacidad desde el diseño en los sistemas de IA debería ayudar a generar confianza y a sortear los posibles problemas de privacidad.

La confianza es un factor clave para los ingresos y el crecimiento. Las organizaciones que utilizan la IA deben integrar la privacidad en los procesos de desarrollo de la IA y en los sistemas de IA para garantizar que sean seguros, eficaces e imparciales, con el apoyo de una gobernanza sólida, una rendición de cuentas clara y una supervisión robusta. A la espera de que la legislación se ponga al día con la tecnología, las organizaciones que deseen la tecnología o las organizaciones que deseen poner en marcha la IA deben integrar la privacidad en todas las fases del ciclo de vida de la IA como mejor práctica. Adoptar un enfoque de privacidad desde el diseño puede ayudar a garantizar a los clientes, reguladores y otras partes interesadas la fiabilidad de la IA y minimizar cualquier impacto negativo. Para ayudar a las organizaciones a adoptar un enfoque proactivo de ingeniería de privacidad por diseño, a continuación se indican los principios clave de privacidad que deben tenerse en cuenta a lo largo del ciclo de vida de la IA.

01

Legalidad y equidad

La IA tiene un propósito legítimo, legal y claramente definido con un impacto mínimo en la privacidad.

Riesgo para la privacidad

Fallos en el diseño y la implementación del modelo o ausencia de elementos de seguridad. Estos riesgos pueden producirse cuando las especificaciones de diseño son inadecuadas para las tareas previstas, debido a una mala elección de las variables, a suposiciones erróneas o a dar prioridad a la eficiencia frente a la eficacia.

Qué puede salir mal

Un organismo del sector público creó un sistema basado en IA para detectar el fraude en las prestaciones que arrojaba resultados imprecisos, aleatorios e injustos. Entre los fallos fundamentales del diseño y la aplicación del modelo estaba la exigencia de datos invasivos, irrelevantes, banales y subjetivos. Por ejemplo, utilizaba el estado civil, la duración y la frecuencia de las relaciones para asignar puntuaciones de alto riesgo a determinados solicitantes, pero no a otros. El modelo también establecía correlaciones injustificables entre distintas entradas que socavaban la coherencia de las decisiones. Por último, el proceso de toma de decisiones carecía de un proceso de apelación funcional (una característica de seguridad clave). Como resultado, el algoritmo asignaba de forma desproporcionada puntuaciones de alto riesgo de fraude a las personas que más necesitaban las prestaciones sin ninguna prueba o justificación convincente. Los solicitantes que reunían los requisitos y eran marcados se veían sometidos a investigaciones invasivas y estigmatizantes, sufriendo retrasos innecesarios en la entrega de prestaciones muy necesarias.

Mitigar el riesgo

Seleccione cuidadosamente los insumos que sean pertinentes, legales y no discriminatorios. Antes de tomar decisiones basadas en correlaciones, debe demostrarse que tienen una relación causa-efecto.

02

Transparencia y explicabilidad

La transparencia es fundamental tanto para la rendición de cuentas como para la optimización del producto. La explicabilidad y la interpretabilidad ayudan a comprender cómo se toman las decisiones y ofrecen garantías cuando la IA funciona bien, o recursos cuando no lo hace.

Riesgo para la privacidad

Acciones de cumplimiento de la normativa; riesgos de propiedad intelectual; ausencia de características de seguridad; defectos de diseño del modelo.

Qué puede salir mal

La incapacidad de explicar y justificar las decisiones de los sistemas de IA protegidos por algoritmos de "caja negra" impide a las personas cuestionar el proceso y el resultado. La falta de transparencia tanto en los datos de entrada como en los de salida dificulta evitar resultados discriminatorios o perjudiciales.

Mitigar el riesgo

La información personal para formar conjuntos de datos debe ser transparente, fiable, exacta, completa y correlacionada con los datos de salida, que pueden impugnarse si son sesgados o inexactos. Todos los datos recopilados deben contener avisos de privacidad (derecho a ser informado).

03

Gobernanza y accountability

La legislación sobre privacidad exige estructuras de gobierno sólidas y programas de privacidad que definan y comuniquen claramente las funciones y responsabilidades.

Riesgo para la privacidad

Deriva de conceptos; desviación de funciones; diseño defectuoso de algoritmos.

Los desarrolladores de IA o los promotores de proyectos pueden exagerar o incluso tergiversar las afirmaciones sobre sus modelos. Cuando los compradores aceptan sin rechistar estas afirmaciones y confían en estos sistemas o les dan luz verde, las deficiencias sólo pueden hacerse patentes cuando se producen daños en el mundo real u otros fallos de la IA. Por ejemplo, los "falsos positivos" del reconocimiento facial pueden dar lugar a falsas detenciones, arrestos de personas inocentes e intrusiones injustificadas en sus vidas..

Qué puede salir mal

Un hipotético barco autoconducido diseñado para encontrar la ruta más rápida a través del puerto podría causar daños a los delicados sistemas acuáticos o a los bañistas desprevenidos si no se programa para hacerlo de forma segura. Por eso, las entradas a los conjuntos de datos de entrenamiento deben incluir imágenes de bañistas.

Mitigar el riesgo

Los sistemas de IA requieren líneas claras de responsabilidad para garantizar que Los riesgos de la IA se gestionan eficazmente; se comunica a todos los implicados un propósito, una estrategia y un conjunto de expectativas claros; se supervisa y se informa adecuadamente, especialmente sobre las deficiencias; se responsabiliza a terceros (por ejemplo, proveedores de datos o desarrolladores de IA), que cooperarán para resolver cualquier problema que surja.

04

Minimización de datos

Los datos personales deben ser adecuados, pertinentes y limitados a su finalidad.

Riesgo para la privacidad

Fallos posteriores al despliegue (solidez; ataques de adversarios; interacciones imprevistas).

En el mundo de la inteligencia artificial, se suele suponer que "cuanto más, mejor", pero no todos los datos son de la misma calidad. Los sistemas entrenados en conjuntos de datos sin validación externa pueden no funcionar correctamente en el mundo real. Esto puede dar lugar a un uso indebido o a una reutilización injusta de la información personal de alguien. Además, es probable que no se consulte a los hablantes de idiomas poco representados antes de utilizar sus idiomas para entrenar y desarrollar modelos de lenguaje natural sin comprensión cultural.

Qué puede salir mal

Los conjuntos de datos no validados externamente podrían dar lugar a recomendaciones o conversaciones inapropiadas con menores, difusión de información errónea y comentarios racistas, respuestas lógicamente incoherentes o a mentiras descaradas. Además, el uso de fotos de baja calidad para entrenar sistemas de reconocimiento facial puede afectar negativamente a la precisión en aplicaciones del mundo real y crear sesgos de selección.

Mitigar el riesgo

No todos los datos son útiles, relevantes o fiables. La minimización de datos impulsa una mejor selección de datos para el entrenamiento, filtrando los datos inapropiados, lo que da como resultado menos conjuntos de datos pero de mayor calidad. Utilizar principalmente el inglés y otras lenguas dominantes para entrenar los LLM puede dar lugar a modelos sesgados y excluir otras lenguas en los datos de entrenamiento, especialmente si solo se utilizan unas pocas fuentes de medios (web, redes sociales).

05

Limitación de la finalidad

El tratamiento de datos personales debe tener una finalidad claramente definida y comunicada para proteger los derechos, respetar la autonomía y prevenir cualquier daño potencial.

Riesgo para la privacidad

Fallos en el diseño y la aplicación del modelo (desviación de conceptos y desviación de datos por fallos de alineación); tareas imposibles; medidas de cumplimiento de la normativa; modelo de negocio fuera de la ley.

La falta de propósito abre la posibilidad de la desinformación. Por ejemplo, tomar el trabajo escrito de un autor y utilizarlo para crear nuevos artículos falsamente atribuidos al autor.

Qué puede salir mal

Un LLM entrenado con artículos de noticias disponibles públicamente y otros datos extraídos de la web puede proporcionar respuestas incorrectas o promover o la desinformación. Si se escriben artículos controversiales, difamatorios o que no se ajustan a la realidad, con el estilo de un autor cuyo contenido se utilizó para entrenar el modelo de IA, podría causar daños si el público cree que el autor lo escribió.

Mitigar el riesgo

Todas las partes interesadas en el sistema de IA, desde los desarrolladores hasta los vendedores y usuarios finales, deben comprender y respetar el objetivo previsto de la IA. Esto guiará la selección de los elementos de datos utilizados para entrenar el modelo, los casos de uso durante el despliegue y el funcionamiento, los valores y supuestos integrados en él, su configuración, las salvaguardias y mucho más. De este modo, se respetan el consentimiento original o la excepción de consentimiento y las expectativas de las personas, lo que aumenta la confianza y reduce el riesgo de actividades de aplicación de la ley o reacciones públicas.

06

Precisión

Bajo la ley de privacidad, los datos personales deben estar actualizados, ser completos y exactos antes de ser utilizados. Además, las personas tienen derecho a corregir sus datos.

Riesgo para la privacidad

La calidad de los datos puede repercutir en la eficacia de la IA y causar diversos perjuicios. Por ejemplo, los datos inexactos pueden ser perjudiciales al tomar decisiones sobre políticas gubernamentales o planificación comunitaria. Incluso cuando los datos introducidos son correctos, el modelo puede ofrecer un perfil incorrecto de una persona debido a suposiciones erróneas, una puntuación deficiente o la incapacidad de procesar datos desconocidos. Unos datos personales inadecuados o irrelevantes pueden dar lugar a una desviación del modelo, degradando su rendimiento (por ejemplo, una menor precisión de la predicción). Si falta una revisión humana significativa, el sistema de IA puede tomar decisiones inexactas, aunque funcione según lo programado y no detecte anomalías.

Qué puede salir mal

La policía detiene falsamente a un hombre porque su sistema de reconocimiento facial lo identifica como sospechoso sin otras pruebas. A pesar de que la policía no está segura del parecido entre el hombre y la foto, cree en las afirmaciones de precisión del proveedor del sistema de IA. El hombre sufre perjuicios económicos debido a la pérdida de salario y honorarios de abogados, humillación, ansiedad, molestias, pérdida de libertad y posible estigma debido a la detención ilegal.

Mitigar el riesgo

Los principios de minimización de datos y precisión pueden ayudar a mejorar la calidad de los datos para evitar muchos de estos perjuicios. Por ejemplo, la Oficina del Comisionado de Información del Reino Unido (ICO) aconseja a los desarrolladores de IA que consideren la compensación entre la minimización de datos y la precisión estadística en la fase de prueba para garantizar que el modelo sigue siendo preciso. Es esencial supervisar el rendimiento del modelo para la toma de decisiones o los sistemas de IA predictiva para garantizar que los datos siguen siendo relevantes, actualizados, adecuados y se vuelven a entrenar cuando sea necesario.

07

Limitación de almacenamiento

Las leyes de privacidad prohíben a las empresas conservar datos personales cuando ya no son necesarios. Aun así, algunas leyes permiten conservar los datos si se anonimizan adecuadamente.

Riesgo para la privacidad

Una vez entrenado un modelo, los datos de entrenamiento subyacentes sólo deben conservarse si es necesario volver a entrenarlo. Incluso entonces, el riesgo de desviación del modelo exige reevaluar la calidad de los datos para eliminar cualquier dato irrelevante u obsoleto. Conservar los conjuntos de datos de entrenamiento más allá de su uso legítimo, incluso si están "anonimizados", también supone una importante carga de cumplimiento, ya que el riesgo de re-identificación debe gestionarse continuamente.

A los reguladores les preocupan especialmente los riesgos de violación y de reutilización ilícita de los conjuntos de datos o el enriquecimiento ilícito y el riesgo de re-identificación asociado a los datos conservados más tiempo del necesario. Además, incluso si se anonimizan, los datos de formación filtrados pueden enriquecerse o someterse a ingeniería inversa para volver a identificar a las personas.

Qué puede salir mal

Un hipotético modelo de IA diseñado para evaluar la idoneidad para el empleo se basa en un antiguo conjunto de datos que prioriza la capacidad de un empleado para asistir al trabajo en persona. El modelo no ha sido reentrenado para abordar las nuevas realidades del trabajo a distancia. En consecuencia, los candidatos que no pueden asistir en persona se ven desfavorecidos y excluidos o reciben una calificación inferior en el proceso de contratación.

Mitigar el riesgo

Es fundamental conocer a fondo todas las leyes pertinentes sobre conservación de datos , revisar y depurar periódicamente los datos cuando corresponda.

08

Seguridad

Las empresas que tratan datos personales deben garantizar su confidencialidad, integridad y disponibilidad.

Riesgo para la privacidad

Fracaso bajo ataques adversarios; actividad de aplicación de la normativa; pérdida de credibilidad.

Las malas prácticas de seguridad pueden dar lugar a la violación de los datos de entrenamiento, que pueden incluir información sensible como detalles financieros, datos demográficos y códigos postales. Una violación de este tipo podría exponer a las personas del conjunto de datos de formación al riesgo de fraude de identidad, perjuicios financieros, ansiedad y molestias en sus esfuerzos por evitar posibles daños.

Qué puede salir mal

Algunos de los principales riesgos para la seguridad son :

- Re-identificación mediante ataques de "caja negra" y "caja blanca".
- Riesgo de revelación de atributos (riesgo de inferir información adicional a partir de datos anonimizados).
- Vulneración de datos mediante ataques de adversarios, por ejemplo, sistemas de juego que permiten a un impostor obtener acceso no autorizado.

Mitigar el riesgo

Un análisis más amplio de la seguridad requerirá conocimientos especializados en materia de seguridad. Sin embargo, algunos aspectos de la seguridad relacionados con la privacidad (como los ejemplos anteriores) requieren una atención específica, ya que afectan a todo el conjunto de requisitos.

09

Respeto a la privacidad del usuario final

La IA debe respetar los derechos de privacidad, incluidos los derechos de información, rectificación, explicación, supresión y toma de decisiones automatizada.

Riesgo para la privacidad

Medidas reglamentarias; pérdida de credibilidad y quebraderos de cabeza para las relaciones públicas; ausencia de elementos de seguridad.

Qué puede salir mal

Supongamos que un sistema de IA no reconoce los derechos de privacidad y no tiene suficientes controles y equilibrios. En este caso, la empresa se expone a riesgos normativos y de reputación.

Mitigar el riesgo

Estos derechos se aplican a todo el ciclo de vida de la IA, pero pueden variar ligeramente según la fase. Los mecanismos de explicabilidad deben integrarse en todo el ciclo de vida de la IA, ya que el diseño, las decisiones de entrada y el modelado pueden influir en la decisión resultante. Los derechos de revisión e impugnación pueden garantizar que las personas sigan ejerciendo el control sobre sus datos personales. El derecho a corregir los datos de entrada o a cuestionar los supuestos en los que se basa un modelo es fundamental para garantizar una toma de decisiones justa y precisa.

El camino por recorrer: Construir una IA fiable

La privacidad de los datos es la base de una empresa que da prioridad a la IA, con transparencia sobre cómo se utiliza la información personal y plena responsabilidad por cualquier uso indebido, con procesos para mitigarlo rápidamente.

La velocidad y eficacia de la IA están transformando el mundo, y es comprensible que las empresas estén deseando aprovechar su potencial. La ventaja tecnológica solo puede aportar ventajas competitivas si los clientes y otras partes interesadas confían en que los datos se utilizan de forma responsable. Aunque los organismos reguladores siguen esforzándose por seguir el ritmo de los avances de la IA, la Ley de IA de la UE envía un mensaje claro de que la regulación será exhaustiva y tendrá consecuencias significativas en caso de incumplimiento. Las empresas que establezcan controles sólidos sobre el uso de la IA basados en directrices éticas claras, deberían estar bien posicionadas para aprovechar los beneficios de la IA y, al mismo tiempo, proteger a la sociedad de posibles riesgos y cumplir los requisitos normativos.

Five key steps that can help companies build trust in AI.

1

Comprenda su entorno normativo y adopte una estrategia de privacidad de la IA

Legisladores, responsables políticos y reguladores insisten constantemente en alinear los sistemas de IA con normas reconocidas. Por lo tanto, es esencial identificar qué marcos normativos se aplican a su empresa, determinar con cuáles decide cumplir y planificar cómo se desplegará su IA. Cree una línea de base para el uso de la IA que satisfaga los distintos regímenes y racionalice su desarrollo de IA o las actividades empresariales relacionadas con la IA en consecuencia.

2

Incorpore la privacidad desde el diseño a sus proyectos de IA

Evalúe el impacto sobre la privacidad y aborde los problemas de cumplimiento en la fase de ideación -y durante todo el ciclo de vida de la IA- mediante una evaluación sistemática del impacto sobre la privacidad (PIA) o una evaluación del impacto sobre la protección de datos (DPIA). La privacidad desde el diseño, tal como se describe en la norma ISO31700 y el Marco de Evaluación de la Privacidad por Diseño de KPMG, puede ayudar a las organizaciones a incorporar la privacidad a los sistemas de IA. Incluso si cree que su sistema sólo utiliza datos anónimos o no personales, puede surgir Riesgo para la privacidad, incluida la re-identificación a partir de conjuntos de datos de entrenamiento e incluso modelos de IA e impactos descendentes de datos no personales utilizados para entrenar modelos que afectan a individuos y comunidades. Una evaluación sólida también incluirá la modelización de las amenazas a la seguridad y la privacidad a lo largo del ciclo de vida de la IA y la consulta a las partes interesadas cuando proceda. Considerar cuestiones de privacidad más amplias, como la justicia de datos y la soberanía de los datos indígenas (los derechos de los pueblos indígenas a gobernar los datos sobre sus comunidades, pueblos, tierras y recursos).

3

Evaluar el riesgo de la IA para la privacidad

Evaluar el riesgo para la privacidad asociado al desarrollo de soluciones de IA internas o al uso de modelos públicos que se entrenan con datos públicos. Asegúrese de que estos modelos se adhieren a las nuevas normas éticas, reglamentos, mejores prácticas y códigos de conducta para hacer operativos los requisitos.Esto se aplica tanto si usted es el desarrollador como si es un cliente que desarrolla o adquiere e integra un sistema de IA. Si usted es un cliente, pida al desarrollador la documentación que respalde su PIA y las evaluaciones relacionadas con el Riesgo para la privacidad de la IA y lleve a cabo sus propios modelos. Si no pueden proporcionar esta documentación, considere otro proveedor. En muchas jurisdicciones, incluidas las del Reino Unido y la UE, la PIA/DPIA ya es un requisito legal y un punto de referencia que debería incluir consideraciones de IA. La PIA/ DPIA debe abordar el uso inicial de la IA y las consideraciones de diseño. Hay que centrarse en la articulación de la necesidad y proporcionalidad de la recogida de datos, así como en el consentimiento.

4

Audite su sistema de IA

Si es usted un desarrollador de sistemas de IA o un tercero/proveedor de IA, debe garantizar a clientes y reguladores que ha tomado las medidas necesarias para crear una IA digna de confianza. Una forma de hacerlo es mediante una auditoría con respecto a normas reconocidas, marcos reguladores y mejores prácticas, incluida una evaluación del impacto algorítmico. Por ejemplo, probando el sistema de IA mediante guiones de prueba que puedan abordar escenarios del mundo real para obtener comentarios de los usuarios y ayudar a garantizar su eficacia, fiabilidad, imparcialidad y aceptación general antes de su despliegue. Esto incluye explicar qué datos se utilizaron y cómo se aplicaron al usuario final, así como la forma en que el usuario final puede impugnar o cuestionar el uso de la IA con fines de toma de decisiones automatizada para evitar resultados sesgados.

5

Respetar los derechos y las opciones explicando y haciendo transparentes las entradas y salidas de datos.

Esté preparado para responder a las preguntas y gestionar las preferencias de las personas afectadas por el desarrollo o el uso de sistemas de IA. Las organizaciones que deseen utilizar la IA para la toma de decisiones automatizada deben ser capaces de explicar en un lenguaje sencillo cómo la IA puede afectar a sus usuarios finales. La explicabilidad es la capacidad de articular por qué un sistema de IA ha llegado a una determinada decisión, recomendación o predicción. Considere la posibilidad de desarrollar flujos de trabajo documentados para identificar y explicar qué datos se utilizaron, cómo se aplicaron al usuario final y cómo éste puede impugnar o cuestionar el uso de la IA para la toma de decisiones.