Ministry of Infrastructure
and Water Management

Evaluación del impacto de la IA

La herramienta para un proyecto de IA responsable

Versión 2.0, diciembre de 2024. En colaboración con colegas del Ministerio de Infraestructura y Gestión del Agua (I&W) en la Oficina del CIO (CDIB), la Inspección de Medio Ambiente Humano y Transporte (ILT IDlab y departamento de análisis) y la Dirección General de Obras Públicas y Gestión del Agua (RWS Datalab).

Si tiene alguna pregunta o comentario, envíela a teamai@minienw.nl

Tabla de contenido

La evaluación de impacto de la IA ayuda a garantizar una gestión responsable IA por diseño

La inteligencia artificial (IA) ofrece oportunidades, pero también conlleva riesgos. Es importante tener claridad con respecto al impacto de un sistema de IA antes de implementarlo para evitar consecuencias negativas no deseadas. Esto ayuda a garantizar que se puedan aprovechar al máximo las oportunidades que ofrece la IA. Para permitir que la IA se use de manera responsable, el laboratorio de identificación y el departamento de análisis de ILT, el laboratorio de datos de RWS y la Oficina del CIO del Ministerio de Industria y Comercio han desarrollado la Evaluación de impacto de la IA (AIIA). La AIIA está destinada a usarse como una herramienta para respaldar el proceso de pensamiento con el fin de aumentar la responsabilidad, la calidad y la reproducibilidad de la implementación de la IA. La AIIA analiza los obstáculos en la recopilación de datos, el sistema de IA y los algoritmos y tiene en cuenta la legislación y las regulaciones pertinentes. Una vez completada, una AIIA brinda transparencia con respecto a las evaluaciones realizadas al decidir si se utiliza o no un sistema de IA.

En este caso, asumimos que la IA se está aplicando en un contexto específico . Si el sistema de IA (o parte de él) se utiliza para un propósito diferente, será necesario llevar a cabo otra IAIA. Algunos ejemplos de esto podrían incluir un modelo de reconocimiento de imágenes para barcos que también se esté implementando para otros vehículos.

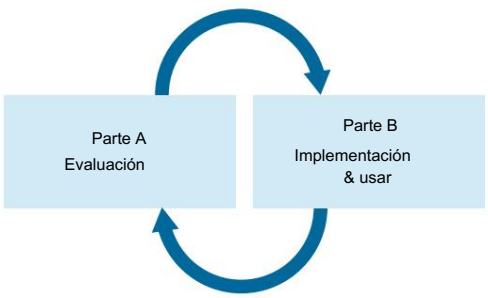
La Ley de IA describe una serie de áreas de riesgo para los sistemas de IA, desde el riesgo inaceptable hasta el riesgo mínimo. Las medidas que se deben tomar dependerán en última instancia del nivel de riesgo. Por este motivo, siempre se deben aplicar los niveles de riesgo descritos en la Ley de IA (consulte el Apéndice 1). Por ejemplo, si el sistema de IA pasa de una aplicación de bajo riesgo a una aplicación de alto riesgo, se necesitarán requisitos más estrictos para cumplir con la Ley de IA.

Completar el AIIA con un grupo multidisciplinario

El AIIA debe ser completado por un grupo multidisciplinario de la organización, ya que se necesita un tipo diferente de conocimiento para completar las diferentes secciones. El encabezado "¿Quién hace qué?" ofrece una descripción general de los roles que pueden ser relevantes para cada capítulo. Algunas preguntas deberán ser respondidas por un científico de datos y otras por un experto legal.

Las diferentes partes de la AIIA

El AIIA se divide en dos partes. La Parte A analiza los factores que se deben tener en cuenta para utilizar un sistema de IA: ¿cuál es el propósito y qué efectos se esperan? Esta información se utiliza luego para evaluar la posible aplicación del sistema de IA y las medidas necesarias. Esto garantiza que el debate ético en torno a la conveniencia de su aplicación sea demostrable. La Parte B analiza el diseño, la implementación y el uso del sistema de IA.



Aunque ambas partes se deben completar por separado, hay muchas interrelaciones entre ellas. Al completar la Parte A, a veces será necesario abordar algunos temas de la Parte B para lograr una evaluación efectiva. Por ejemplo, el tipo de algoritmo tiene un impacto en la sostenibilidad y, por lo tanto, afectará la evaluación. Las decisiones que se toman durante la implementación pueden influir en la decisión de utilizar o no el sistema de IA. Esto puede ocurrir, por ejemplo, cuando una suposición sobre la implementación en la Parte A resulta no ser correcta en la Parte B (por ejemplo, debido a los datos disponibles o al uso de una infraestructura de TI diferente), como resultado de lo cual la evaluación realizada en la Parte A ya no es correcta. La siguiente tabla muestra las diferentes aplicaciones potenciales del AIIA.

Utilice la AIIA en todas las fases de su proyecto de IA

El AIIA se puede utilizar en todas las fases de desarrollo y adquisición de un sistema de IA. El AIIA es más beneficioso si se utiliza desde el principio de un proyecto de IA. Es obligatorio haber completado el AIIA cuando el sistema entra en producción (incluso en el caso de un piloto). La siguiente tabla describe las distintas posibilidades.
No existe una pre-AIIA porque la AIIA es obligatoria para todos los sistemas de IA, independientemente del riesgo basado en la Ley de IA.

Escaneo rápido AIA	Utilice el Análisis rápido de AIIA para investigar si una idea de IA es factible y deseable. Los resultados dejarán en claro si es una buena idea adquirir o desarrollar un sistema de IA. Al investigar las opciones, utilice las preguntas azules de la Parte A y utilice las preguntas azules de la Parte B para un análisis más profundo si es necesario. Este paso se puede utilizar para tomar una decisión de seguir adelante o no.
Compilando Un proyecto plan	Si está elaborando un plan de proyecto para el uso de un sistema de IA, es obligatorio completar la AIIA. Utilice la AIIA para brindar transparencia en lo que respecta a la evaluación del sistema de IA y dar cuenta de las decisiones tomadas. También es una herramienta eficaz para debatir y verificar que se hayan tenido en cuenta todos los aspectos relevantes. Las directrices tituladas AI voor opdrachtgevers (IA para clientes de encargos) brindan consejos prácticos para el desarrollo, la implementación y la gestión de un sistema de IA.
Durante desarrollo	Las decisiones que se toman en torno a la implementación influyen en el impacto de un sistema de IA (por ejemplo, los datos utilizados o el tipo de modelo). Utilice la Parte A para determinar el impacto y la decisión de utilizar un sistema de IA. Utilice la Parte B para comprobar si se han tenido en cuenta los aspectos pertinentes en la implementación.
Sistema de IA en producción	La AIIA es obligatoria en el momento en que el sistema de IA entra en producción. Una vez que se encuentra en producción, debe evaluar si el proyecto de IA aún cumple con los requisitos. Compruebe si se ha modificado el área de aplicación.

AIIA para algoritmos de impacto

El AIIA ha sido diseñado para su uso con sistemas de IA, sistemas basados en máquinas con un componente de aprendizaje. También existen algoritmos más simples, que son básicamente una "receta" con reglas predefinidas. También se puede utilizar el AIIA para algoritmos de este tipo, pero no es obligatorio. Las preguntas ayudarán a evaluar el impacto del algoritmo y las opciones de implementación que son relevantes para garantizar una aplicación responsable.

¿Necesitar ayuda?

¿Considera que la AIIA no es la herramienta adecuada para un proyecto o sistema de IA que está considerando? ¿Quizás tenga otros comentarios o preguntas sobre la AIIA? ¿La finalización de la AIIA genera dudas? Si es así, comuníquese con el equipo de IA en I&W.

Versión 2.0 rendición de cuentas

La primera versión de la AIIA fue desarrollada por el departamento de análisis e IDlab de ILT, el Datalab de RWS y la Oficina de I&W del CIO (CDIB). La AIIA fue aprobada por el Consejo Administrativo de I&W el 4 de julio de 2022. Esta versión 2.0 (2024) ha sido adaptada para añadir información sobre la IA generativa, y también tiene en cuenta las experiencias de los usuarios e información sobre la versión definitiva de la Ley de IA. El vínculo con el

Derechos fundamentales y evaluación del impacto de los algoritmos (IAMA) También se ha mejorado1 . Como resultado, la AIIA ahora no solo aborda la cuestión de cómo se puede aplicar la IA, sino también si es deseable implementarla.

Cómo utilizar este documento

Los siguientes factores son importantes para completar la AIIA:

- La AIIA es obligatoria en el momento en que el sistema de IA entra en producción.
- El grado de cumplimiento de la AIIA dependerá de la experiencia del líder del proyecto y también reflejará el impacto del sistema de IA. A menos que se indique lo contrario, un simple "sí" o "no" no será suficiente como respuesta a las preguntas.
- Las preguntas de color azul son obligatorias y deben ser respondidas siempre.
- Las preguntas verdes tienen como objetivo proporcionar información adicional. Deben completarse si son necesarias. importante.
- Las palabras en negrita son términos clicables que se definen en el Glosario de términos incluido en el apéndice.
- Está disponible una plantilla para rellenar.

Tenga en cuenta que el Servicio de Auditoría del Gobierno Central (Auditdienst Rijk) y el Tribunal de Cuentas de los Países Bajos (Algemene Rekenkamer) pueden comprobar la corrección y la seguridad del sistema de IA. Una auditoría de la IA completada no significa necesariamente que el sistema de IA sea seguro. Para garantizar el cumplimiento de la Ley de IA (Reglamento (UE) 2024/1689), El Apéndice 1 debe completarse si su sistema de IA es de alto riesgo.

1 Esto significa que ya no es necesario completar tanto una AIIA como una IAMA, excepto en los casos en que se necesite asistencia adicional para evaluar los derechos fundamentales. Véase el Capítulo 2.1.

¿Quién hace qué?

Es importante que el AIIA sea completado por un equipo multidisciplinario porque se necesitan conocimientos o experiencia específicos para las diferentes secciones. La siguiente tabla incluye una sugerencia para la participación de diferentes roles para un capítulo específico del AIIA. Por supuesto, dependerá del alcance y tamaño del proyecto si estos roles son realmente necesarios o si se debe involucrar a más personas de las que se sugieren en la siguiente tabla.

	Propósito	Impacto	Evaluación	Técnico Robustez	Gestión de la confianza	Gestión	Responsabilidad
Partes interesadas:		insignia					
Director de Informática	insignia	insignia	insignia	insignia	insignia	insignia	insignia
Oficina de Seguridad de la Información				insignia	insignia	insignia	insignia
Consultor de comunicación:		insignia					insignia
Científicos de datos:	insignia	insignia	insignia	insignia	insignia	insignia	insignia
Responsable del tratamiento de datos o titular de los datos de origen:					insignia		
Experto en el dominio:	insignia	insignia	insignia	insignia	insignia	insignia	
Profesional de la privacidad:	insignia	insignia	insignia	insignia	insignia		
Experto legal:		insignia	insignia		insignia		insignia
Cliente encargado de la puesta en servicio:	insignia	insignia	insignia	insignia	insignia	insignia	insignia
Otros miembros del equipo del proyecto:							
Líder del proyecto:	insignia	insignia	insignia	insignia	insignia	insignia	insignia
Asesor estratégico en ética:	insignia	insignia	insignia				

Parte A: Evaluación

¿El impacto del sistema de IA es proporcional a los objetivos previstos? Esta es una pregunta clave en la primera parte de la AIIA. Repase todas las preguntas y concéntrese en los propósitos, la solución prevista y los efectos esperados. La proporcionalidad de la implementación de la IA es de particular importancia. ¿Quizás existan otras formas menos radicales de lograr el objetivo?

Si esto resulta difícil de evaluar, por ejemplo, debido a un conflicto de intereses o porque el sistema de IA viola derechos fundamentales, existen diversas herramientas disponibles para llevar a cabo una conversación estructurada sobre aspectos éticos. Entre ellas, se incluyen el diálogo sobre datos del IBDS, DEDA, Moreel beraad (Consultas Morales) y muchas otras Métodos. El equipo de IA de I&W ofrece talleres internos para apoyar esta conversación.

1 Propósito y necesidad del sistema

Estas preguntas tratan sobre el propósito del sistema de IA y su función prevista dentro de la organización.

1.1 Propósito del sistema

La "aplicación de la IA" no es un objetivo en sí misma. Un sistema de IA se implementa para lograr un objetivo determinado dentro de la organización, como permitir que el trabajo se realice de manera más eficiente o eficaz. ¿Qué problema debe resolverse? En este caso, observe todo el proceso en el que la IA desempeña un papel. Puede utilizar las respuestas a estas preguntas al completar el resto de la AIIA.

1. Proporcione una breve descripción del propósito previsto y el resultado previsto del sistema de IA (título, descripción general, definición del problema, plazo previsto, ubicación, grupos objetivo, dominio y proceso operativo).

2 ¿En qué nivel de riesgo del Reglamento de IA se encuentra su sistema de IA: inaceptable, alto o ¿riesgo mínimo?

3 ¿En qué parte de la organización (en qué procesos) se pretende utilizar el sistema de IA?

Para identificar el nivel de riesgo de la Ley de IA que se aplica a su sistema, le recomendamos realizar la evaluación del Apéndice 1. Esto le brindará mayor claridad con respecto a las obligaciones legales subyacentes a su sistema.

Si su sistema se encuentra en la categoría "inaceptable" , no se permitirá el sistema de IA. En ese caso, no es necesario completar el resto de la AIIA.

Si su sistema de IA está en la categoría de "alto riesgo" , también deberá completar las preguntas restantes en el Apéndice 1. Esto le proporcionará una descripción general de las obligaciones adicionales que deben cumplir los sistemas de alto riesgo.
Para encontrarse.

Para todos los demás sistemas, categorizados como de riesgo bajo o mínimo, las preguntas del AIIA serán suficientes.

1.2 Solución prevista

En esta parte, analizamos la solución prevista para el problema descrito anteriormente, como las tecnologías de IA que se aplicarán y los datos que se utilizarán. Cuando sea necesario, utilice también las preguntas de la Parte B para ayudarle a responder completamente las preguntas siguientes.

1. Proporcione una breve descripción del sistema de IA previsto (tecnología, datos y tipo de algoritmo).

2 ¿Por qué se eligió esta forma de IA (por ejemplo, IA generativa, regresión lineal o red neuronal)?

3 ¿Qué alternativas se consideraron (por ejemplo, sin IA, IA menos compleja, tipo diferente de algoritmo)?

1.3 Rol dentro de la organización

Como cualquier otro sistema de TI, un sistema de IA tiene un cliente encargado y una parte con responsabilidad final.
La propiedad es esencial. En estas preguntas, se determina la división de tareas en el desarrollo y uso del sistema. Estas funciones se definen en el glosario de términos. Base sus respuestas en estas definiciones.

1 Describe la división de tareas en la configuración del sistema de IA (como el desarrollador,
cliente encargado, responsable del proyecto, organizaciones de gestión de TI y persona con responsabilidad final).
Si un tercero es responsable del desarrollo: ¿qué acuerdos contractuales existen?

2 ¿Quién será el usuario del sistema de IA, quiénes son los usuarios finales que trabajarán con el sistema?
¿Y qué partes involucradas se verán afectadas por el sistema de IA?

3 ¿Qué partes interesadas, personas y/o grupos han sido consultados en el desarrollo de la IA?
¿sistema?

4 ¿Qué comentarios se han recopilado de equipos o grupos que representan diferentes orígenes y experiencias? ¿Y cómo se ha hecho el seguimiento de estos comentarios?

1.4 Mantenimiento y administración

Al igual que cualquier otro sistema informático, un sistema de IA requiere mantenimiento y administración. En el caso de una prueba de concepto o piloto, también es una buena idea investigar dónde y cómo se pondrá en producción el sistema de IA.
Esto permite tomar las decisiones correctas con antelación y evita, por ejemplo, una infraestructura de TI o tecnologías de IA incompatibles que darían como resultado un sistema de IA que no se puede gestionar.

1 Describa la división de tareas para la administración y el mantenimiento del sistema de IA (por ejemplo, el desarrollador, el cliente encargado, el líder del proyecto, las organizaciones de gestión y la parte con la responsabilidad final). Si una parte externa es responsable del desarrollo del sistema: ¿qué acuerdos contractuales existen?

2 ¿Cómo se aplican las nuevas leyes y reglamentos que se puedan introducir o actualizar durante la vigencia de una ley? ¿Se tiene en cuenta el sistema de IA?

3 ¿Se ha documentado la experiencia necesaria para gestionar el sistema de IA?

4 ¿Cómo se tienen en cuenta los cambios en el contexto del sistema de IA?

2 Impact

Las preguntas de esta sección tienen como objetivo identificar el impacto de la aplicación del sistema de IA en un contexto específico. En este caso se mencionan explícitamente los derechos fundamentales y la sostenibilidad, pero también hay otros efectos que pueden tener un impacto positivo o negativo sobre la prosperidad en general. Se trata de impacto en el sentido más amplio de la palabra, como el impacto sobre grupos destinatarios específicos o sobre la prosperidad en general.

Por favor, lea el capítulo paso a paso porque está organizado en una secuencia determinada. Por ejemplo, los valores públicos también se relacionan con los derechos humanos y estos ya se tratan en el capítulo "Derechos fundamentales".
La Parte B también cubre gran parte del impacto.

2.1 Derechos fundamentales

Todas las personas que tengan un interés en el funcionamiento del sistema de IA deben recibir un trato adecuado. Esto significa que deben protegerse los derechos (fundamentales) de todas las partes involucradas . A la hora de responder a las preguntas, se aplican los derechos humanos fundamentales establecidos en la Constitución holandesa. y el Convenio Europeo de Derechos Humanos. Apéndice de la Evaluación de Impacto sobre los Derechos Fundamentales y los Algoritmos (IAMA) incluye una lista de grupos de derechos fundamentales y esto podría ser potencialmente útil para responder las preguntas siguientes.

El sistema de IA puede dar lugar a una infracción leve, media o grave de los derechos fundamentales. Será necesario prestar especial atención a las infracciones medias o graves y también pueden ser necesarias medidas adicionales.

La Parte B de la AIIA incluye medidas para proteger una serie de derechos fundamentales, como la protección de los datos personales, el derecho de acceso a la información y un proceso justo. También puede haber otros derechos fundamentales que podrían verse afectados por el uso del sistema de IA.

¿Necesita ayuda? La Parte 4 de la IAMA ofrece un plan paso a paso sobre los derechos fundamentales que incluye información adicional sobre el tema2 .

1 ¿Cuál será el impacto potencial sobre los derechos fundamentales de los ciudadanos al utilizar el sistema de IA?

2 ¿Qué base jurídica sustenta el uso del sistema de IA y las decisiones que se pretenden adoptar?
¿Basado en el sistema de IA?

3 ¿Qué disposiciones constitucionales pueden ser aplicables?

4 ¿Cuáles de estas disposiciones constitucionales pueden ser vulneradas en caso de una incorrecta implementación del sistema de IA?

2 Las preguntas 1, 2 y 3 del Capítulo 4 del IAMA corresponden con las preguntas del Capítulo 2.1, la pregunta 4 del IAMA corresponde con el Capítulo 1.1 y las preguntas 5, 6 y 7 pueden utilizarse para realizar la evaluación del Capítulo 3.

2.2 Sostenibilidad

A lo largo de su vida útil, un sistema de IA genera una huella ecológica a través del consumo de energía, agua y materias primas. Por ejemplo, para que funcione, se necesita un suministro continuo de agua . Además, un sistema de IA también consume energía. Más concretamente, es importante preguntarse si el consumo de energía es proporcional a la definición del problema y también a cómo se generó esa energía.

Actualmente es difícil medir el impacto ambiental de un sistema de IA: las empresas, por ejemplo, no son transparentes en lo que respecta al consumo de energía y la investigación en este ámbito todavía está en sus primeras etapas. Mientras no haya cifras fiables disponibles sobre el impacto, se debe utilizar esta pregunta principalmente como una forma de considerar opciones sostenibles. A veces es posible reducir el impacto ambiental mediante el uso de una tecnología, una infraestructura o un modelo diferentes, y esto a menudo tiene la ventaja añadida de hacer que el modelo sea más rápido.

En términos generales, es posible afirmar que un modelo "más simple" consumirá menos energía que un modelo más extenso, como un modelo de lenguaje grande. Además, el tamaño y el tipo de entrada y salida también tienen una influencia importante en el consumo de energía (por ejemplo, el texto requiere menos potencia de procesamiento que las imágenes).

Por otra parte, también es posible utilizar un sistema de IA para lograr beneficios medioambientales. Este impacto debería compensarse, por ejemplo, con los costes medioambientales de funcionamiento del sistema.

1 ¿Cuál será el impacto ambiental de la introducción del sistema de IA (desarrollo, instalación y uso) y cómo se medirá?

2 ¿Qué medidas se han tomado para minimizar el impacto ambiental (negativo) de la IA?
¿sistema?

2.3 Otros efectos

En esta sección se analizan los efectos o consecuencias que no se han mencionado anteriormente. La parte B del AIIA incluye numerosos efectos, por lo que es importante analizarlos para asegurarse de que se ha incluido todo. En este capítulo también es posible mencionar cuestiones que son de particular importancia para este proyecto a fin de garantizar que se les preste la atención suficiente.

A continuación se presenta una lista de posibles efectos. La medida en que estos puntos sean relevantes dependerá del alcance del proyecto.

Otros efectos a considerar:

- Valores públicos: ideas cambiantes y relacionadas con el contexto sobre lo que percibimos como valioso como sociedad. Se puede encontrar más información y conocimientos sobre los valores públicos en la caja de herramientas del Ministerio del Interior y Relaciones del Reino (BZK) para la innovación éticamente responsable3
- La misión y visión de la organización: para I&W, por ejemplo, esto es trabajar para lograr un mundo mejor, más limpio, Holanda segura, sostenible y accesible
- Corto y largo plazo
- Efectos sobre el individuo, la organización y la sociedad.
- Aspectos positivos y negativos del uso del sistema de IA

En esto no sólo se deben considerar los riesgos, sino también las oportunidades y los efectos positivos de aplicar el sistema de IA.

1 ¿Cómo contribuye el sistema de IA a la misión de la organización?

2 Además de las preguntas anteriores, ¿existen otros efectos relevantes (positivos, negativos, ¿Qué riesgos (para grupos objetivo específicos, en diferentes niveles y en términos de prosperidad general) del sistema de IA deben tenerse en cuenta?

3 https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/publieke-waarden/toolbox-voor-ethisch-verantwoorde-innovatie/publieke-waarden-centraal/

3 Evaluar si se debe o no utilizar el sistema de IA

En este paso, se determina si tiene sentido implementar el sistema de IA o no, analizando el propósito, la solución y el impacto del sistema de IA. ¿Es proporcionado aplicar el sistema de IA? ¿La aplicación prevista realmente resuelve el problema? ¿O también son adecuadas otras tecnologías? Describa claramente cómo se ha realizado esta evaluación.

Justifique claramente sus decisiones. Intente tenerlo todo en cuenta. Considere a los diferentes grupos de partes implicadas: no debe permitirse que ningún grupo de personas se vea injustamente desfavorecido. Por supuesto, en algunos casos el impacto será tan significativo que no será necesario realizar una evaluación adicional, por ejemplo, si se violan los derechos fundamentales o si la ley y las regulaciones no lo permiten.

La parte B incluye medidas para garantizar la aplicación responsable del sistema de IA. Es posible que sean necesarias medidas adicionales o específicas, por ejemplo, dentro del proceso. Debes mencionarlas aquí.

1 ¿El impacto es proporcional a los objetivos previstos y existen otras formas menos radicales de alcanzarlos? En otras palabras: ¿es proporcional y se ajusta a la subsidiariedad desplegar el sistema para alcanzar los objetivos planteados?

2 ¿Existen medidas adicionales (por ejemplo, como parte de los procesos) que podría tomar para utilizar el sistema? ¿Responsablemente?

Parte B: Implementación y uso del sistema de IA

La Parte B de la AIIA analiza el diseño, la implementación y el uso del sistema de IA. Incluye más detalles y, en muchos casos, requiere un mayor conocimiento sobre la tecnología de IA. Asegúrese de involucrar a esta persona experta, por ejemplo, un analista de datos, al completar y utilizar este capítulo.

4 Robustez técnica

Que un sistema de IA funcione o no de la forma para la que fue previsto depende de su solidez técnica.

4.1 Sesgo

El sesgo consiste en hacer suposiciones sobre cosas, personas o grupos. Esto tiene dos caras. Por un lado, es necesario sacar conclusiones de antemano sobre el uso de los datos en una nueva situación. Al generalizar, siempre hacemos suposiciones. Por otro lado, es importante evitar distorsiones injustificadas causadas por formas de injusticia y sesgos indeseables que pueden estar en contradicción con los derechos humanos o las regulaciones.

El sesgo puede ocurrir en cualquier parte del sistema. Por ejemplo, incluye: sesgo en la entrada, sesgo en el modelo y sesgo en el resultado. Al diseñar y utilizar IA, es importante tener en cuenta el sesgo de datos y el sesgo de diseño. Estos tipos de sesgo suelen estar causados por suposiciones socioeconómicas en los datos o hechas por un desarrollador. El modelo puede exacerbar estas suposiciones. Esto puede significar que un sistema de IA no funcione de manera efectiva para todas las partes involucradas si no se es consciente del sesgo y se corrige.

En esencia, se trata de una cuestión de concienciación e integridad. Es imposible trabajar completamente sin prejuicios. Los prejuicios pueden perdurar a menudo durante décadas y seguir sin ser reconocidos, incluso cuando deberían serlo. Tomemos, por ejemplo, los modelos de lenguaje (ChatGPT, Google Bing, Microsoft Co-pilot): probablemente se hayan entrenado utilizando información de Internet y todos los supuestos que esta información ha incluido durante años.
Por lo tanto, en lugar de centrarnos en una IA libre de sesgos, deberíamos intentar ser lo más conscientes posible de cualquier posible discriminación.

El sesgo está estrechamente relacionado con la diversidad, la igualdad y la justicia humanas, pero es importante tener en cuenta que las suposiciones también pueden estar relacionadas con aspectos no humanos, como la naturaleza o el entorno vital. Al considerar cómo pretende mitigar el sesgo, también puede ser relevante establecer una distinción entre cualquier impacto negativo, ningún impacto positivo y el impacto positivo que podría tener el sesgo.

1 ¿Cómo se controlarán los sesgos potencialmente indeseables , como el sesgo en la entrada, el sesgo en el modelo y el sesgo en la ¿Se tendrá en cuenta la salida del sistema de IA ?

Sesgo en la entrada (datos)

2 ¿Los datos de entrada son pertinentes y representativos, teniendo en cuenta el propósito previsto? (pregunta 1 de 1.1) del sistema de IA?

3 En el muestreo aleatorio, ¿se han protegido subpoblaciones en caso necesario?

4 ¿Se ha fundamentado y coordinado con las partes la elección de las variables de entrada? ¿involucrado?

Sesgo en el modelo

5 ¿Qué medidas se han adoptado para evitar que se creen o sesgos injustificados o
¿exacerbado en un sistema de IA?

6 ¿Puede el sistema de IA ser utilizado por los usuarios finales previstos (en otras palabras, independientemente de su
características, como la edad, el género o la capacidad)?

Sesgo en la salida (datos)

7 ¿Existen mecanismos de control, supervisión o seguimiento para evitar que los grupos de la sociedad se vean afectados
de manera desproporcionada por las consecuencias negativas del sistema de IA? En el caso concreto de la
tecnología de la información, es necesario distinguir entre los sujetos supervisados (OTS) y el resto de la sociedad.

4.2 Precisión

Un sistema de IA preciso funciona de manera eficaz y es eficaz en sus evaluaciones. Es importante medir continuamente el rendimiento de un sistema de IA (tanto durante las fases de desarrollo como de producción). La calidad de los datos de entrenamiento utilizados también es importante. Cualquier sistema de IA es un trabajo en progreso: sigue siendo necesario probarlo y volver a entrenarlo periódicamente. Es deseable tener alguna forma de cuantificar la probabilidad de que el sistema realice una evaluación incorrecta.

Puede determinar la precisión del sistema de antemano estableciendo criterios de aceptación tanto para los datos (de entrenamiento) como para el sistema. Algunos ejemplos de criterios de aceptación pueden ser una cantidad mínima de datos o valores umbral específicos para el sistema de medición. Existen numerosos tipos diferentes de sistemas de medición (a los que los científicos de datos suelen denominar "métricas de rendimiento") disponibles para cuantificar la calidad de los modelos , como una puntuación de precisión, una puntuación de exactitud y una puntuación de recuperación o F1.

Es importante asegurarse de que el sistema de medición y los criterios de aceptación estén correctamente adaptados a los datos y al propósito previsto del sistema de IA4 . Entre otras cosas, esto debe coordinarse con los resultados del análisis de riesgos (véase "Gestión de riesgos"), ya que con el paso del tiempo pueden surgir nuevos riesgos al utilizar un sistema de IA. También es importante asegurarse de que la calidad del sistema se controle de forma continua. Durante el reentrenamiento o el desarrollo posterior, debe volver a evaluar los criterios de aceptación y la elección de los sistemas de medición.

4 El sistema de medición elegido debe ser adecuado al modelo y a los datos utilizados para medir la calidad. Tomemos, por ejemplo,
Por ejemplo, un sistema que tiene que etiquetar cinco palabras de cada cien palabras de un texto específico. Si el sistema etiqueta
0 palabras en este texto, el modelo tiene una precisión del 95 %. Si se determina la calidad del sistema en función de la precisión,
el modelo parece funcionar muy bien cuando, de hecho, la recuperación es 0 y, en realidad, no funciona bien en absoluto. Por este
motivo, la precisión no es adecuada para determinar el rendimiento de este modelo.

1 ¿Cómo se medirá y garantizará la precisión continua del sistema?

2 ¿Qué criterios de aceptación se han establecido para medir la calidad de la entrada (datos) y ¿Salida (datos) del modelo?

3 ¿Son los criterios de aceptación apropiados para los datos y el propósito del sistema de IA?

4 ¿Cómo se comprobarán regularmente los resultados (datos) de forma aleatoria y se supervisarán continuamente? ¿exactitud?

5 ¿Cómo se analizarán y evaluarán las desviaciones en los resultados (datos) con respecto a los criterios de aceptación? ¿corregirse de manera oportuna?

6 ¿Cuáles serían los resultados si se utilizaran modelos alternativos?

4.3 Confiabilidad

Un sistema de IA fiable produce resultados consistentes en casos similares. En lo que respecta a la fiabilidad, la cuestión clave es si los resultados individuales (datos) se pueden reproducir utilizando el mismo modelo y la misma entrada (datos), la misma configuración y los mismos parámetros. También es importante que el sistema proporcione una indicación fiable de qué tan bien funcionará el modelo en nuevas situaciones.

1. ¿Es confiable el sistema de IA?

2. ¿Cuáles son los factores más importantes que influyen en el rendimiento del sistema de IA?

3. ¿Se excluye una parte del (sub)conjunto de datos del proceso de aprendizaje del modelo y solo se utiliza para determinar la confiabilidad, o se calcula la confiabilidad del modelo mediante validación cruzada?

4. ¿Cómo se ha fundamentado y evaluado el ajuste de (hiper)parámetros?

4.4 Implementación técnica

La implementación técnica describe cómo se integra técnicamente el sistema de IA en el entorno de TIC de la organización. Los requisitos específicos de hardware y software del sistema de IA se documentan para que se puedan tener en cuenta en la implementación y la gestión del sistema.

Además, la arquitectura del sistema muestra claramente cómo se relacionan entre sí los distintos componentes del software. Una arquitectura bien pensada reduce los riesgos operativos que conlleva la creación de una solución tecnológica y crea un puente entre los requisitos operativos y técnicos. En muchos casos, ya existe documentación y conviene consultarla. También conviene echar un vistazo a la arquitectura empresarial de gestión de infraestructuras y agua (IWEA).

1. ¿Cómo se ha implementado técnicamente el sistema de IA?

2. ¿Se ha considerado cómo encaja el sistema de IA en la infraestructura técnica y del sistema existente y se han tomado las medidas adecuadas para su implementación (si corresponde)?

3. Describe la arquitectura del sistema (¿cómo se interrelacionan los componentes del software)?

4. ¿Se han documentado requisitos específicos de hardware y software?

5. Si la aplicación está alojada externamente, ¿bajo qué condiciones sucede esto?

6. ¿Cómo se configura el acceso al sistema de IA y sus componentes (pensemos en las medidas genéricas de gestión de TI)?

7. ¿Cómo puede el sistema de IA interactuar con otro hardware o software (si corresponde)?

8. ¿Cómo se configura el registro y la monitorización?

4.5 Reproducibilidad

La reproducibilidad tiene que ver con el entrenamiento, la validación y la prueba. La reproducibilidad se refiere a cuestiones como el registro de los datos utilizados, el desarrollo del modelo, el registro de los cambios en los datos, si la misma entrada (datos) produce resultados consistentes y si existen determinadas situaciones o condiciones que pueden afectar el resultado (datos).

La reproducibilidad está estrechamente relacionada con la trazabilidad. El objetivo principal de la trazabilidad es garantizar que los conjuntos de datos y los procesos estén debidamente documentados. La gestión de versiones de datos, el modelo y el entrenamiento del modelo son aspectos importantes de esto.

1. ¿Es reproducible el sistema de IA ? ¿Se ha establecido un proceso para medirlo?

2. ¿Es posible reconstruir ahora o en el futuro los resultados (datos) obtenidos (es decir, se han guardado versiones anteriores del modelo, conjuntos de datos y condiciones mediante la gestión de versiones)?

3. ¿Es posible reconstruir el modelo basándose en los parámetros dados y una semilla fija ?

4. ¿Es posible reproducir las líneas generales del sistema de IA utilizando la documentación?

5. ¿Cómo se documentarán los cambios durante la vida útil del sistema?

4.6 Explicabilidad

La explicabilidad técnica se relaciona con la capacidad de comprender tanto los procesos técnicos como las decisiones humanas relacionadas con ellos. También debe quedar claro qué opciones de diseño se han elegido y cuál es la razón para utilizar el sistema de IA. Véase también: "Rendición de cuentas" para obtener más información sobre la explicabilidad, la transparencia y la comunicación con los usuarios y otras partes implicadas.

1. ¿Es el sistema de IA suficientemente explicable e interpretable para los desarrolladores?

1. Al desarrollar el sistema de IA, ¿cómo se ha tenido en cuenta la explicabilidad del modelo, por ejemplo para los usuarios?

2. ¿Qué tecnologías se han utilizado para garantizar que el sistema de IA sea explicable y por qué?
   ¿Esta tecnología elegida?

5 Gobernanza de datos

La gobernanza de datos se refiere a los procedimientos establecidos en relación con los datos en lo que respecta al acceso, la propiedad, la usabilidad, la integridad y la seguridad. También abarca la calidad de los datos utilizados.

La gobernanza de datos también incluye la privacidad. La privacidad es uno de los derechos humanos fundamentales que podría verse comprometido por la IA. Por lo tanto, es importante garantizar que exista una gobernanza de datos eficaz y la protección de los datos personales, al menos de acuerdo con el Reglamento General de Protección de Datos (RGPD) y la política de privacidad de I&W.

5.1 Calidad e integridad de los datos

La calidad de los datos es esencial para el funcionamiento eficaz de un sistema de IA. Existen buenas razones por las que existe el término "basura que entra = basura que sale". Además, los datos recopilados pueden, por ejemplo, contener sesgos, imprecisiones, errores y equivocaciones construidos socialmente (véase también "Sesgo"). Esto debe abordarse antes de que estos datos se utilicen de nuevo.

Para ello, se deben utilizar como inspiración los principios FAIR (Findable, Accessible, Interoperable y Reutilizable) y los principios FACT (Fair, Accurate, Confidential and Transparent), estrechamente relacionados. Estos principios se utilizan en el RGPD. Los conjuntos de datos y el procedimiento de trabajo deberán probarse y documentarse en cada paso: formación, prueba, fase de implementación y fase operativa. Esto también se aplica a los sistemas de IA que no se han desarrollado internamente, sino que se han adquirido.

La calidad de los datos es especialmente importante si se van a utilizar datos personales. Según el RGPD, es esencial que los datos personales sean correctos y estén actualizados si es necesario.5 Los datos también deben ser necesarios para el propósito del análisis. Por lo tanto, también debe describir el propósito del análisis y cómo está tratando los datos y cualquier error en los conjuntos de datos y los resultados.6

Las 'Directrices para la aplicación de algoritmos' del Ministerio de Justicia y Seguridad ('Richtlijnen voor het toepassen van algoritmen')7 hacen explícita la calidad de los datos a un nivel muy operativo.

1. ¿Qué datos de entrenamiento se utilizarán como entrada para el algoritmo y de qué fuentes provienen los datos? ¿originar?

2. ¿Cómo se protegerá la calidad de los datos?

5 Artículo 5.1.(d) del RGPD. Véase también el apartado 3 Rectificación y supresión de datos del RGPD.

6 Richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses, JNV, p.20.

7 https://open.overheid.nl/documenten/ronl-1411e45f-b822-49fa-9895-2d76e663787b/pdf

General

3. ¿Son los datos utilizados necesarios para el sistema de IA?

4. ¿Cómo se evitan las duplicaciones de datos no deseadas?

5. ¿Es posible actualizar los datos de entrenamiento y prueba cuando la situación lo requiera? ¿Cuándo decidirá volver a entrenar, detener temporalmente o seguir desarrollando el sistema de IA?8

6. ¿Los datos cumplen los supuestos subyacentes al modelo?

7. ¿Cómo se han recopilado y cotejado los datos de entrada utilizados en el sistema de IA?

8. ¿Cómo se etiquetarán los datos?

9. ¿Qué factores (piense en limitaciones en el método de recolección, almacenamiento, etc.) afectan la calidad de ¿Qué se puede hacer con la entrada (datos)?

10. ¿Se han evaluado los datos de entrada para detectar los cambios que se producen durante el entrenamiento, las pruebas y la evaluación? ¿También durante el uso del algoritmo a lo largo del tiempo?

Salida (datos)

11. Si la salida (datos) se utiliza como nueva entrada, ¿cómo se almacenará y comprobará la salida (datos)? ¿Corrección y completitud?

12. ¿Cómo se garantizará que los resultados (datos) estén disponibles en el momento oportuno?

8 Artículo 14.4.

5.2 Privacidad y confidencialidad

Al diseñar un sistema de IA, se debe tener en cuenta la legislación sobre privacidad. En definitiva, es más fácil asegurarse de que esto se haga correctamente con antelación que repararlo en una etapa posterior. Por supuesto, la privacidad debe protegerse durante todo el ciclo de vida del sistema de IA.

Al procesar datos personales, se debe completar un análisis previo de la Evaluación de Impacto de Protección de Datos (EIPD) para determinar si es necesario completar la EIPD completa.

Además de los datos personales, se podrá utilizar otra información confidencial que no debe hacerse pública. Esto se aplica, por ejemplo, al uso de información confidencial, como información clasificada o secretos comerciales.
Estos datos también deben protegerse. La Ley de IA complementa el RGPD al incluir reglas adicionales para el uso de datos (personales) en sistemas de IA. Los datos confidenciales deben protegerse y asegurarse de manera suficiente (véase Gestión de riesgos).

1. ¿Qué enfoques se están adoptando para los datos personales o confidenciales?

Respecto a los datos personales

2. ¿El sistema de IA trabaja con datos personales (es aplicable el RGPD)? En caso afirmativo, responda también a las siguientes preguntas. En caso contrario, continúe con “Respecto a los datos confidenciales”.

3. ¿El tratamiento de datos personales es proporcional y se ajusta a la subsidiariedad (utilice la evaluación del Capítulo 3 como base para esto)?

4. ¿Se puede rastrear la salida del sistema de IA de forma directa o indirecta hasta las personas (es el RGPD)? aplicable)?

5. ¿Han estado involucrados funcionarios, como el Director de Privacidad, el Oficial de Seguridad de la Información, ¿Director de Información, Responsable de Privacidad, etc.?

6. ¿Con qué frecuencia se evalúa la calidad y necesidad del tratamiento de datos personales?

Respecto de los datos confidenciales (es decir, no personales)

7. ¿Se utilizarán o almacenarán datos confidenciales?

8. ¿Cómo se garantizará la seguridad de esta información?

6 Gestión de riesgos

Es importante estar atento a los riesgos. Los riesgos imprevistos pueden hacer que un sistema de IA produzca resultados poco fiables. Esto puede provocar daños, como resultado de un rendimiento deficiente del sistema de IA o debido a ataques de piratas informáticos.

6.1 Prevención de riesgos

El desarrollo y la puesta en servicio de un sistema de IA entrañan peligros que esta AIIA pretende abordar en la medida de lo posible. Sin embargo, pueden surgir problemas imprevistos. Es importante determinar cómo se abordarán estos posibles peligros. Deben existir mecanismos para gestionar los riesgos y estos mecanismos deben haber sido probados. Esto implica áreas como la prevención del envenenamiento de datos, el alcance de las medidas de gestión y la seguridad de los lugares de almacenamiento de los resultados. Los riesgos también pueden surgir después de que se haya implementado el sistema de IA. Verifique las medidas de gestión de riesgos de forma periódica, al menos una vez cada tres años o cuando se realicen cambios importantes.

En el caso de los sistemas de IA de alto riesgo, es obligatorio contar con un sistema de gestión de riesgos. Encontrará preguntas adicionales al respecto en el Anexo 1. En el caso de los demás sistemas, será suficiente con un análisis de riesgos puntual.

1. ¿Cómo se ha probado el sistema para garantizar que las medidas de gestión de riesgos sean adecuadas y específicas?

6.2 Procedimiento alternativo

Es recomendable tener un plan en caso de que surjan problemas con el sistema de IA. Esto significa que debe haber un procedimiento alternativo disponible. Algunos ejemplos pueden incluir la opción de volver de un modelo de aprendizaje automático a un modelo basado en reglas más limitado. Incluso se puede ir un paso más allá y realizar el proceso manualmente. Por supuesto, si es aceptable que el sistema no esté disponible temporalmente, no tener ningún plan en absoluto también es una opción.

El uso de un sistema de IA puede provocar el deterioro de ciertas habilidades humanas. El efecto de la calculadora sobre nuestra capacidad de cálculo mental es un ejemplo de ello. Por este motivo, cualquier procedimiento alternativo no debe depender de repente de esta habilidad.

1. ¿Cuál será el plan en caso de problemas con el funcionamiento del sistema de IA?

2. ¿Cuál sería el impacto si fallara el sistema?

3. Vea el ejemplo de la calculadora anterior. ¿Qué efecto equivalente podría ocurrir si se coloca el sistema de IA? ¿En servicio y es esto deseable?

6.3 Riesgos de seguridad de la información

En la medida de lo posible, un sistema de IA debe construirse de forma que sea seguro desde el principio, y la seguridad debe tenerse en cuenta en la fase de diseño. De esta forma se pueden gestionar los riesgos de seguridad de la información, como la manipulación del modelo, el acceso no autorizado o los ataques de piratería . Elabore una lista de riesgos previsibles utilizando el proceso de gestión de riesgos de la organización. Esto incluye áreas como: trazar el mapa de la tríada CIA; niveles de clasificación de la información; implementación de medidas de línea base de seguridad de la información gubernamental (BIO); pruebas de seguridad; y, si el nivel de seguridad BIO no es suficiente, posiblemente realizar un análisis de riesgos (técnico) adicional.

Para ello, son esenciales unas autorizaciones configuradas correctamente y un proceso de modificación sólido. Además, es importante comprobar si los errores e irregularidades se pueden detectar y solucionar técnicamente. Se pueden encontrar orientaciones más prácticas en el Marco de evaluación de algoritmos (Toetsingskader Algoritmes van de Rekenkamer) 9 o en el Marco de investigación de algoritmos ADR (Onderzoekskader algoritmes ADR) 10 del Tribunal de Cuentas.
La AIVD tiene una guía para el desarrollo seguro de sistemas de IA.11 Además, el Proyecto Abierto de Seguridad de Aplicaciones Mundiales (OWASP) tiene numerosos recursos sobre la aplicación segura de la IA.

1. ¿Cómo se identifican los riesgos de seguridad de la información, se reducen a un nivel aceptable y se prueban (desde un punto de vista ¿perspectiva técnica)?

2. ¿Cómo se evita que terceros no autorizados aprovechen las vulnerabilidades de la IA? ¿sistema?

3. ¿Cuál sería el impacto si terceros tuvieran acceso no autorizado al código fuente, los datos o los resultados del sistema de IA?

4. ¿Es posible que las personas aprovechen el hecho de que se esté utilizando un sistema de IA en lugar de... ¿Una decisión humana?

5. ¿Cuál es el sistema para registrar quién utiliza el sistema de IA y durante cuánto tiempo?

6. Además de las medidas de seguridad estándar de I&W, ¿se han tomado medidas adicionales para ¿Asegurar el sistema de IA?

9 https://www.rekenkamer.nl/onderwerpen/algoritmes/toetsingskader
10 https://www.rijksoverheid.nl/documenten/rapporten/2023/07/11/onderzoekskader-algoritmes-adr-2023
11 https://www.aivd.nl/documenten/publicaties/2023/02/15/ai-systemen-ontwikkel-ze-veilig

7 Responsabilidad

El gobierno central holandés debe rendir cuentas de sus acciones dentro de la organización, ante la Cámara de Representantes y ante la sociedad en general. Aunque la tecnología se utiliza cada vez más, también existen preocupaciones sobre el uso de la IA. Para rendir cuentas con respecto al uso y los resultados de los sistemas de IA, es esencial establecer un proceso para este propósito.

7.1 Transparencia hacia los usuarios

Los usuarios finales deben tener una idea clara del funcionamiento de un sistema de IA (además de la facilidad de explicación que se exige a los desarrolladores, véase "Robustez técnica"). Esto se aplica en particular a los empleados que utilizan la IA como parte de su proceso de trabajo. No es necesario que comprendan completamente el sistema de IA. Sin embargo, el funcionamiento de un sistema de IA y sus limitaciones deben estar claros en líneas generales.

Si es necesario tomar una decisión sobre el uso de la IA, además de comprender su funcionamiento y sus limitaciones, es esencial tener un grado significativo de influencia sobre la decisión.

1. ¿De qué manera proporciona a sus usuarios finales una idea del funcionamiento y las limitaciones del sistema de IA? ¿Se les presta la atención suficiente mientras persistan?
2. ¿Qué papel desempeñan las personas en las decisiones basadas en la información del sistema de IA ('humanos en el circuito') y cómo? ¿Les permites desempeñar este papel?
3. ¿Cómo puede el sistema ser monitoreado y comprendido por todos (supervisión humana)?

7.2 Comunicación a las partes implicadas

En esta sección se tratan dos tipos de comunicación con los usuarios finales. En primer lugar, los usuarios finales deben saber que están tratando con los resultados de un sistema de IA (y no de un ser humano, por ejemplo). En segundo lugar, los usuarios finales tienen derecho en todo momento a saber cómo un algoritmo determina los resultados de un sistema de IA. Esto también significa que el propósito y las limitaciones del sistema deben comunicarse de forma clara, justa y transparente. Tanto los procesos tecnológicos como las decisiones humanas relacionadas deben ser comprensibles, recuperables y corregibles en caso necesario. Esto se puede lograr, por ejemplo, designando a una persona de contacto con conocimientos sustanciales sobre el sistema de IA.
Debido a la naturaleza de autoaprendizaje de la IA, no siempre es posible rastrear completamente el funcionamiento del sistema.
Sin embargo, siempre debe ser posible proporcionar una explicación adecuada del proceso a los usuarios finales.

Además, los ciudadanos deben poder acceder a la información sobre el sistema de IA o hacer valer sus derechos de conformidad con el RGPD. Los ciudadanos deben poder impugnar los resultados del sistema de IA. Esto también significa que los datos y las condiciones en las que se ponen a disposición deben almacenarse (véase Archivado).

1. ¿En qué medida es usted transparente respecto de los sistemas de IA frente a los diferentes grupos de partes involucradas y de qué manera?
2. ¿Se están estableciendo mecanismos para permitir que los usuarios finales hagan comentarios sobre el sistema (datos, tecnología, grupo objetivo, etc.)? ¿Y cómo y cuándo se validan (se analizan y se hace seguimiento de ellos)?
3. De conformidad con la Ley de IA, ¿es necesario incluir el sistema en el registro de algoritmos y/o (para aplicaciones de alto riesgo) en la base de datos de la UE?

4 ¿Se informa al usuario final y a las partes involucradas en el sistema de IA de que los resultados son generados por un sistema de IA y qué implica esto para ellos?

5 ¿Se ha elaborado un manual?

6 ¿Cuáles son los posibles efectos secundarios (psicológicos), como el riesgo de confusión, preferencia o ¿Fatiga cognitiva en el usuario final al utilizar el sistema de IA?

7. ¿De qué manera se les da a los diferentes grupos de partes involucradas (ciudadanos, colegas, gerentes, etc.) ¿Una perspectiva de los diferentes aspectos del sistema de IA? Esto incluye áreas como el uso de datos, el modelo o los resultados.

8. ¿Cómo habéis tomado medidas para conseguir explicabilidad específicamente dirigida al usuario final?

9 ¿Es el sistema lo suficientemente transparente para permitir a los implementadores interpretar la salida del sistema (datos) y utilizarla adecuadamente?20

10 ¿Se han tomado medidas para proporcionar capacitación a los usuarios finales si es necesario?

11 ¿Cómo se asegura de que los comentarios realizados por las partes involucradas y los usuarios finales se gestionen adecuadamente a nivel interno?

12 Si una parte involucrada desea presentar una objeción21 o presentar una queja sobre el sistema de IA,22 ¿Está claro qué pasos deben seguir? Lo mismo se aplica a la presentación de un recurso.

7.3 Verificabilidad

La verificabilidad se refiere a la forma en que se evalúan los datos, el modelo y los resultados. Este proceso puede realizarse mediante auditorías internas o externas. Se aplican requisitos más estrictos cuando se utiliza un sistema de IA en situaciones de mayor riesgo. áreas.

Es fundamental que exista conocimiento de las fuentes, el sistema y el resultado. Esto es responsabilidad del propietario del sistema.

1 ¿Cómo se verificará el sistema de IA y quién lo hará?

2 ¿De qué manera se rinde cuentas sobre el sistema de IA?

3 ¿Quién realiza la auditoría independiente del sistema de IA? ¿Y de qué manera?

7.4 Archivado

El archivado es el almacenamiento de información para que pueda reutilizarse en el futuro. Dicha reutilización puede incluir la reconstrucción del modelo (véase "Reproducibilidad"), la explicación del funcionamiento del sistema a un nuevo miembro del personal (véase "Explicabilidad") o la rendición de cuentas a una parte involucrada (véase "Rendición de cuentas"). El archivado también es importante para garantizar el cumplimiento de la legislación y las regulaciones. Por ejemplo, se aplica un período mínimo de retención para las aplicaciones de alto riesgo de los registros en un sistema de IA, que también se describen en las listas de selección de I&W.

Entrada (datos)

1. ¿Cómo se almacena la entrada (datos) ?

2. ¿Cuál es el período de retención de la entrada (datos)?

Modelo

3. ¿Cómo se almacena el modelo ?

4. ¿Cómo se organiza la gestión de versiones?

Salida (datos)

5. ¿Cuál es el período de retención de la salida (datos)?

Glosario de términos

En este documento se utilizan diversos términos que pueden tener definiciones diferentes en la literatura. La siguiente lista define claramente los términos utilizados en este documento.

Criterios de aceptación	Condiciones basadas en el propósito previsto y los datos acordados que el sistema de IA debe cumplir. Estas condiciones pueden referirse a la cantidad de datos, métricas de precisión para el resultado (datos) o un mecanismo de verificación de resultados independiente. Siempre que sea posible, los criterios de aceptación deben ser cuantificables, lo que permite su seguimiento mediante un sistema de medición adecuado. Los buenos criterios de aceptación son SMART y suficientemente diferenciados para permitir que todos los aspectos del sistema de IA se supervisen de manera efectiva.
Exactitud	Muy preciso o meticuloso: se refiere a un sistema que es capaz de tomar decisiones correctas y
Evaluaciones precisas	Expresado como una fórmula: TP+TN/(TP+TN+FP+FN). TP= Verdadero positivo, TN= Verdadero negativo, FP= Falso positivo, FN= Falso negativo. Cuanto mayor sea el número de resultados verdaderos en relación con los resultados falsos, mayor será la precisión.
Ley de IA	Legislación europea que establece reglas para el desarrollo y uso de sistemas de IA.
Sistema de IA	Un sistema basado en máquinas que está diseñado para funcionar con distintos niveles de autonomía y que puede exhibir capacidad de adaptación después de su implementación y que, para objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar resultados tales como predicciones, contenido, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales.
Algoritmo	Un conjunto de reglas e instrucciones que una computadora sigue automáticamente al realizar cálculos para resolver un problema o responder una pregunta.
Área de aplicación	Término utilizado en la legislación sobre IA para referirse al contexto en el que se utilizará un sistema de IA. Un ejemplo podría ser la infraestructura.
Inteligencia artificial	No existe una definición fija de IA. Aplicamos la descripción de IA utilizada por el Tribunal de Cuentas de los Países Bajos: “la capacidad [...] de interpretar correctamente datos externos, aprender de dichos datos y utilizar estos aprendizajes para lograr objetivos y tareas específicos mediante una adaptación flexible”. Aunque no se utiliza en este documento, también nos gustaría llamar la atención sobre la definición de la Comisión Europea, que es la siguiente: La IA se refiere a los sistemas que muestran un comportamiento inteligente al analizar su entorno y tomar medidas, con cierto grado de autonomía, para lograr objetivos específicos.
Inclinación	Prejuicio. Hacer suposiciones sobre cosas, personas o grupos, en muchos casos sin basarse en mediciones reales.
Sesgo en la entrada	La calidad, la consistencia y la integridad de los datos son una condición previa importante para un análisis imparcial.

Sesgo en la salida	La forma en que se utilizan los resultados (datos) puede tener un impacto en la vida de las personas. Es importante garantizar que una correlación injustificada no dé lugar a una causalidad.
Sesgo en el modelo	¿En qué medida son correctos los modelos ? ¿En qué medida corrigen los fallos conocidos en la representatividad de los datos? Esto puede estar relacionado, por ejemplo, con lo que aprende el sistema de IA y lo que se considera efectos de aprendizaje no deseados.
BIO	Línea base de seguridad de la información gubernamental (Baseline Informatiebeveiliging Overheid) 12, el marco de estándares básicos para la seguridad de la información dentro de los niveles de gobierno.
A propósito	Tener en cuenta la legislación pertinente sobre inteligencia artificial, privacidad y seguridad en el proceso de diseño. Algunos ejemplos son la inteligencia artificial, la privacidad y la seguridad desde el diseño.
Director de Información	Director de Información.
Director de Seguridad de la Información	Director de Seguridad de la Información.
Puesta en servicio del cliente	Una persona o división organizacional que contrata a un contratista. Esta persona también es responsable en última instancia de completar una AIIA (junto con el líder del proyecto).
Corrupción	El mal uso o explotación de errores en el sistema o la explotación de características neutrales del sistema. 25 A diferencia de la corrupción no intencionada.
Sesgo de datos	Se refiere a muestras aleatorias que no son representativas de toda la población.
Canalización de datos	Cómo se mueven los datos del campo al modelo; el proceso por el que pasan los datos.
Titular de los datos	Persona física u organización que tiene (o cree tener) un interés en el uso o los resultados del sistema. Se ha tomado la decisión deliberada de no utilizar la palabra "parte interesada", ya que este término es más amplio que la definición de "parte interesada" utilizada en el derecho administrativo holandés. Algunos ejemplos pueden ser los ciudadanos, las personas supervisadas, pero también los propios usuarios finales.
Sesgo de diseño	Problemas en el diseño técnico, incluyendo limitaciones en las herramientas informáticas, como hardware y software.
Revelador	Una organización o persona que diseña, desarrolla y/o entrena un sistema de IA.
Diversidad	Esto se refiere al reconocimiento de diferentes tipos de “sujetos” en nuestros análisis. Con esto tratamos de evitar que grupos de sujetos relevantes sean excluidos injustamente del desarrollo de un sistema de IA, como resultado de lo cual el sistema no satisface sus necesidades.

12 https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/cybersecurity/bio-en-ensia/baseline-informatiebeveiliging-overheid/

Experto en el dominio	Alguien que tenga mucho conocimiento sobre el área problemática en la que se está construyendo el sistema de IA .
————————	Evaluación de impacto de la protección de datos. Esta herramienta se utiliza para evaluar los riesgos de privacidad que implica el procesamiento de datos.
Usuario final	Los usuarios finales son las personas que aplican el sistema de IA en la práctica dentro de la organización. Esto se refiere a personas físicas. ¿Quién tiene en sus manos los controles? ¿Quién dentro de la organización recopila información del sistema de IA? Algunos ejemplos podrían ser un inspector o un controlador de tráfico vial.
Entidad	Un puesto dentro de un departamento de la organización.
Igualdad	Esto se refiere a la noción de que el mismo tipo de sujeto recibe el mismo trato. tratamiento.
Explicable	Una declaración de cómo las variables de entrada contribuyen a una salida del algoritmo que debe explicarse.
Justicia	Si no se da el mismo trato a todos los sujetos, debe ser posible explicarlo. En este sentido, es importante que exista una imagen lo más completa posible de las características distintivas de los sujetos, tanto para demostrar qué características realmente desempeñan un papel (y atribuyen un riesgo menor a la parte A que a la parte B) como para demostrar cuáles no lo desempeñan (como resultado de lo cual las partes A y B tienen un riesgo demostrado igual).
RGPD	Reglamento General de Protección de Datos. Esta legislación de privacidad regula el cuidado y la protección de los datos personales por parte de empresas y organizaciones.
IA generativa	Un tipo específico de IA, en el que se utilizan algoritmos para generar contenido. Mediante un simple mensaje, los usuarios pueden generar texto, imagen, sonido o código informático en un instante. El ejemplo más conocido de esto es ChatGPT.
Gobernancia	La acción o forma de gobernar, el código de conducta y las organizaciones de supervisión. Se relaciona con las decisiones que determinan las expectativas, otorgan poder o verifican el desempeño. Consiste en un proceso separado o en una parte específica de los procesos de gestión o liderazgo.
Ataque de piratería	Intrusión en el sistema de IA, con consecuencias como contaminación de datos, filtraciones no deseadas sobre un sistema de IA (o su funcionamiento) o corrupción del software o hardware.
IA de alto riesgo	La Ley de IA establece qué es la IA de alto riesgo. A menudo se trata de productos estrechamente relacionados con los derechos fundamentales o la seguridad de los productos. En todos los casos: un sistema de IA que perfila a las personas mediante el tratamiento de datos personales. Además, todos los sistemas de IA en los ámbitos de: datos biométricos no prohibidos; infraestructuras críticas; educación y formación profesional; empleo; acceso y disfrute de servicios públicos y privados esenciales; aplicación de la ley; gestión de la migración, el asilo y el control de fronteras; administración de justicia y procesos democráticos.

Entrada (datos)	Los datos procesados para un fin predeterminado. En el contexto de un sistema de IA, esto puede referirse a datos sin procesar, como observaciones de la realidad. En el contexto del modelo, esto normalmente se refiere a datos preprocesados.
Grupo de interés	Grupo de partes interesadas para medir la diversidad. Puede ser un grupo de usuarios finales o un grupo de personas afectadas por el sistema.
Modelo de lenguaje grande	Un LLM es una IA generativa que puede generar texto. Se entrena con conjuntos de datos muy grandes y contiene numerosos parámetros.
IA de riesgo limitado	La Ley de IA establece qué es la IA de riesgo limitado, es decir, la IA que está diseñada para interactuar con humanos, reconocer emociones o producir imágenes manipuladas. Algunos ejemplos son los filtros de spam, los resúmenes de textos, la clasificación de temas relacionados con incidentes de aviación o los sistemas de IA que regulan la iluminación de las oficinas.
Sistema de IA de bajo riesgo	Sistemas de IA con riesgo de manipulación o engaño. Estos sistemas de IA deben ser transparentes y los usuarios deben estar informados sobre su interacción con la IA.
Organización de gestión	Una organización que configura y optimiza la aplicación del sistema de IA. gestión.
Metadatos	Datos que describen las características de otros datos. Por ejemplo, de quién son, quién los envió o cuándo se modificaron por última vez.
IA de riesgo mínimo	Cualquier sistema de IA que no esté prohibido o categorizado como IA de alto riesgo o IA de riesgo limitado.
Modelo	Representación matemática (simplificada) de la realidad que se utiliza para procesar información. En un sistema de IA, la representación matemática suele aprenderse parcial o totalmente según un algoritmo, por lo que ni siquiera los desarrolladores pueden explicar con exactitud cómo el modelo ha obtenido sus resultados.
Consultas morales	Consejo Consultivo del Medio Ambiente Físico (mayo de 2021), Moreel Beraad.
Impacto negativo	Partes involucradas que experimentaron consecuencias negativas como resultado de la aplicación del sistema de IA, por ejemplo porque fueron discriminadas con base en un sesgo en el sistema de IA.
Sin impacto positivo	Partes implicadas que, por definición, no experimentan ningún impacto negativo por el uso del sistema de IA, pero que, por ejemplo, permanecen en la misma situación que antes. En este sentido, puede existir el peligro de que estas partes implicadas no experimenten el mismo "impacto positivo" del sistema de IA que el que experimentaron otras partes implicadas.
Salida (datos)	Los datos producidos por el sistema de IA. Estos son los resultados del modelo.
Parámetro	Una variable dentro del modelo. Si se modifica esta variable, también cambiará el tamaño resultante del modelo o del cálculo.

Fiesta con lo último responsabilidad	Un rol dentro de la organización que es responsable del sistema de IA. Por ejemplo, esto se refiere a la responsabilidad de garantizar que se obtengan los resultados correctos para el sistema de IA. Normalmente, esta responsabilidad recae en el propietario del proceso.
Impacto positivo	Partes implicadas que experimentan consecuencias favorables como resultado del uso del sistema de IA. Podría tratarse de un grupo minoritario que se ve favorecido. Esto puede conllevar el peligro de que este sesgo positivo sea excesivamente optimista y, por lo tanto, no refleje la realidad. El inconveniente de esto también podría ser un "impacto negativo" para otras partes involucradas.
Líder del proyecto	La persona con la máxima responsabilidad del proyecto del que forma parte el sistema de IA . Junto con el cliente encargado, esta persona también es la responsable final de realizar una evaluación de impacto de IA.
Proporcional	La proporcionalidad consiste en garantizar una relación razonable entre el objetivo y la solución aplicada. ¿El uso de la IA es proporcional al problema que se pretende resolver con el algoritmo? La ventaja esperada debe ser mayor que el riesgo que implica la IA.
Puesta en servicio	El momento en que un sistema de IA se "pone en servicio" es el momento en que se utiliza por primera vez en un proceso. Esto va precedido de una prueba externa o piloto. La AIIA debe haberse completado cuando el sistema se pone en servicio.
Confiable	Que tiene la característica de comportamiento consistente y resultados consistentes.
Reproducible	La capacidad de lograr repetidamente un resultado similar siempre que se ejecute un procedimiento descrito.
Responsable	Las acciones de una entidad pueden atribuirse a esa entidad de una manera única y esta entidad es responsable de esas acciones. ¿Quién es quién? Ingrese qué personas han desempeñado un papel en la respuesta a esta pregunta.
Sistema de IA sin riesgos	Toda IA que no entre en las otras categorías. Estos sistemas no están contemplados en la Ley de IA.
Robustez	Desarrollado con un enfoque preventivo, comportándose como se predijo y describió con anticipación, evitando daños inaceptables.
Semilla	Una "semilla" es el punto de partida de un generador de números aleatorios. A partir de este punto de partida, este generador siempre sigue la misma ruta para crear nuevos números (pseudo) aleatorios. Al documentar la semilla, es posible repetir la ruta de los números (pseudo) aleatorios. Esto significa que la semilla es necesaria para verificar la reconstrucción de un modelo siempre que el modelo haga uso de números aleatorios. La semilla en sí misma también es un número. No hay requisitos específicos para este número y a menudo se elige algo "reconocible" (por ejemplo, "123456", o "0, 42, 1234" o la fecha de nacimiento de un desarrollador).
Lista de selección	Una lista que describe durante cuánto tiempo deben conservarse los registros de archivo, por ejemplo, de conformidad con la Ley de Registros Públicos de los Países Bajos (Archiefwet).

Tenedor de apuestas	Persona u organización que puede influir en una decisión o actividad, verse influida por ella o considerarse influida. Un interesado también puede ser, por ejemplo, el propietario de los datos utilizados.
Trazabilidad	Cuando los procesos y los resultados se pueden verificar.
Transparente	Cuando el funcionamiento y los propósitos del sistema de IA se comunican claramente y los resultados del sistema de IA son explicables.
Tipos de algoritmos	Es posible utilizar distintas tecnologías para crear IA, como redes neuronales, bosques aleatorios u otras formas de aprendizaje automático. Sin embargo, también se pueden utilizar algoritmos menos complejos, como reglas de negocio o árboles de decisión.
Sistema de IA de riesgo inaceptable	Sistemas de IA que: aplican técnicas subliminales, manipuladoras o engañosas; hacer uso inapropiado de vulnerabilidades, categorizar biométricamente; aplicar puntuación social; predecir el potencial criminal de los individuos; crear bases de datos para reconocimiento facial; inferir emociones; reconocer emociones de forma remota en tiempo real.
Corrupción no intencionada	Influir en el funcionamiento del sistema de IA sin intención maliciosa, por ejemplo, introduciendo datos incorrectos o pulsando los botones equivocados. La corrupción no intencionada se incluye en el ámbito de la fiabilidad. Distinguimos entre esta y la corrupción (intencionada).
Usuario	Según la Ley de IA: “una autoridad pública, agencia u otro organismo que utilice un sistema de IA bajo su propia autoridad […]”. El usuario es quien pone en uso el sistema. Nunca se trata de una persona física. Algunos ejemplos podrían ser el ILT o el RWS.

Apéndice 1: Evaluación del nivel de riesgo

La Ley de IA describe una serie de áreas de riesgo de aplicación de los sistemas de IA. Cuanto mayor sea el riesgo, más medidas deben tomarse. El nivel de riesgo se determina en función del área de aplicación en la que se implementa el sistema de IA. Las preguntas que aparecen a continuación ayudan a determinar el nivel de riesgo de su sistema de IA.

La clasificación de los niveles de riesgo en la Ley de IA es absoluta, lo que significa que no se tienen en cuenta otros riesgos, como los derivados de la legislación sobre privacidad. Por ejemplo, un sistema de IA puede tener un impacto significativo en el ámbito de la privacidad, pero puede estar en un ámbito de aplicación con un riesgo mínimo según la Ley de IA. Esto significa que no se trata de un sistema de IA de alto riesgo. Sin embargo, las preguntas estándar de la AIIA pueden ser útiles para mitigar riesgos de este tipo.

Definición de sistema de IA de alto riesgo (Ley de IA)

Un sistema de IA que elabora perfiles de personas mediante el procesamiento de datos personales. Además, todos los sistemas de IA en los campos de: datos biométricos no prohibidos (por la Ley de IA); infraestructura crítica; educación y formación profesional; empleo; acceso y disfrute de servicios públicos y privados esenciales; aplicación de la ley; gestión de la migración, el asilo y el control de fronteras; administración de justicia y procesos democráticos.

¿Aún no está seguro de qué nivel de riesgo se aplica al sistema de IA? Si es así, responda las siguientes preguntas.

A) Evaluación del riesgo inaceptable

1. ¿El sistema de IA utiliza mensajes subliminales? ¿O se trata de técnicas deliberadamente manipuladoras o engañosas? ¿usado?
2. ¿El sistema de IA hace uso de la manipulación o el abuso de un grupo vulnerable, lo que podría resultar en daños físicos o psicológicos?

3. ¿El sistema de IA utiliza puntuación social?13
4. ¿El sistema de IA pretende inferir las emociones de una persona en el trabajo, más allá de consideraciones de seguridad médica?
5. ¿El sistema de IA utiliza la identificación biométrica en un espacio público? ¿O la categorización biométrica de las personas en función de datos personales especiales?

¿Ha respondido no a todas las preguntas? Si es así, es probable que el sistema de IA no sea una aplicación con riesgos inaceptables.

B)Evaluación de alto riesgo

1. ¿El sistema de IA elabora perfiles de individuos?
2. ¿El sistema de IA es un producto o un componente de seguridad de un producto dentro de uno de los siguientes campos?

- Maquinaria (Directiva 2006/42/CE)
- Juguetes (Directiva 2009/48/CE)
- Embarcaciones de recreo y motos acuáticas (Directiva 2013/53/UE)
- Ascensores (Directiva 2014/33/UE)
- Equipos y sistemas de protección destinados a ser utilizados en una atmósfera potencialmente explosiva (Directiva 2014/34/UE)
- Equipos de radio (Directiva 2014/53/UE)
- Equipos a presión (Directiva 2014/68/UE)
- Instalaciones de transporte por cable (Reglamento (UE) 2016/424)
- Equipos de protección individual (Reglamento (UE) 2016/425)
- Aparatos que queman combustibles gaseosos (Reglamento (UE) 2016/425)

13 Puntuación social: sistemas de IA para la evaluación o clasificación de personas físicas o grupos de personas durante un período específico en función de su comportamiento social o de características personales de personalidad conocidas, derivadas o previstas (p. 51 Ley de IA)

• Productos sanitarios (Reglamento (UE) 2017/745)
• Productos sanitarios para diagnóstico in vitro (Reglamento (UE) 2017/746)

El Reglamento también hace referencia a una lista de productos que también se consideran aplicaciones de IA de alto riesgo. Con excepción de los artículos 6.1, 102 a 109 y 122, las obligaciones de alto riesgo de conformidad con la Ley de IA aún no se aplican a estos productos. Sin embargo, los requisitos de la Ley de IA se utilizarán en una etapa posterior para dar forma a la legislación de productos específica que se aplicará a estos productos. Todavía no se sabe cuándo sucederá esto y variará según el producto. Esto se aplica a los siguientes productos y legislación:

• (Protección de la) aviación civil (Reglamento (CE) 300/2008 y Reglamento (UE) 2018/1139)
• Vehículos de dos o tres ruedas y cuatriciclos (Reglamento (UE) 168/2013)
• Vehículos agrícolas y forestales (Reglamento (UE) 167/2013)
• Equipos marinos (Directiva 2014/90/UE)
• Interoperabilidad del sistema ferroviario en la UE (Directiva (UE) 2016/797)
• Vehículos de motor y sus remolques (Directiva (UE) 2018/858 y Directiva (UE) 2019/2144)

3. ¿El sistema de IA utiliza identificación biométrica remota de personas?
4. ¿Se utilizará el sistema de IA como un componente de seguridad en la gestión/operación de: datos digitales críticos? ¿Infraestructura14, transporte por carretera, suministro de agua, gas, calefacción o electricidad?
5. ¿El sistema de IA influye en la contratación de personas y en el acceso al trabajo?
6. ¿Se utilizará el sistema de IA en relación con el acceso y el uso de servicios públicos y privados esenciales y ¿pagos?
7. ¿El sistema de IA estará activo en la aplicación de la ley?
8. ¿Se utilizará el sistema de IA en áreas relacionadas con la gestión de la migración, el asilo y el control de fronteras?

¿Ha respondido no a todas las preguntas? Si es así, es probable que el sistema de IA no sea una aplicación de alto riesgo.

C)Evaluación de la obligación de transparencia

1. ¿Implica un sistema de IA que interactúa con personas?
2. ¿Puede el sistema de IA generar o manipular contenido artificial (IA generativa/GPAI)?
3. ¿Puede el sistema de IA reconocer emociones o categorizar biométricamente?
4. ¿El sistema utiliza tecnologías deep-fake?15

Si ha respondido “sí” a alguna de las preguntas anteriores, debe tomar las medidas de transparencia necesarias.16 En términos generales, esto significa: garantizar que los usuarios sepan que están tratando con un sistema de IA.

14 Infraestructura digital crítica: puntos de intercambio de Internet, proveedores de servicios DSN, registros de nombres de dominio de nivel superior, nube informática, centros de datos.

15 Material de imagen, audio o video o texto creado o manipulado por IA que las personas pueden ver incorrectamente como auténtico y La verdad.

16 Ley de Inteligencia Artificial, artículo 50.

Excepciones

Si el sistema se aplica exclusivamente a una de las siguientes áreas, las preguntas estándar de la AIIA serán suficientes. Aún será necesario que registres el sistema de IA (esto se puede hacer en el registro de algoritmos) si se trata de un sistema de alto riesgo que está cubierto por las excepciones.

Tenga en cuenta: esto solo se aplica si el sistema de IA no crea perfiles de personas. De lo contrario, el sistema siempre ser una aplicación de alto riesgo.17

Excepciones, IA para:

1. Fines militares
2. Aplicación de la ley y justicia
3. Investigación sobre IA
4. IA de código abierto en fase de desarrollo (antes de entrar al mercado)
5. Actividades no profesionales
6. El sistema de IA no influye fundamentalmente en el resultado de la toma de decisiones. Por ejemplo, un sistema que lleve a cabo tareas puramente procedimentales o que sirva para verificar y mejorar una actividad humana previa (por ejemplo, corregir el lenguaje). También se incluyen en este grupo los sistemas utilizados para detectar patrones de toma de decisiones.18

Nota: Muchas definiciones relacionadas con las aplicaciones de la IA aún están en proceso de desarrollo, por lo que es posible que, después de responder a estas preguntas, aún no esté totalmente claro en qué categoría se encuentra el sistema de IA.
Es posible que desee plantear esta pregunta y otras al Sandbox de IA regulatoria del gobierno central holandés (más información).

17 Ley de IA, artículo 6.3.
18 Ley de IA, artículo 6.3.

Apéndice 2: Sistemas de alto riesgo

¿El sistema de IA es una aplicación de alto riesgo (consulte el Apéndice 1)? Utilice las siguientes preguntas adicionales como ayuda. Esto le permitirá aplicar las medidas de seguridad adecuadas para garantizar que utiliza el sistema de IA de manera responsable.

La Ley de IA describe una serie de partes diferentes, cada una de las cuales tiene su propia función o responsabilidad con respecto a un sistema de IA. Este apéndice incluye las preguntas destinadas al usuario (implementador) y al desarrollador (proveedor) de un sistema de IA. Un proveedor, por ejemplo, adquiere un sistema de IA y lo proporciona en su nombre.
Alternativamente, introducen cambios sustanciales en el sistema.19

Preguntas si desea utilizar un sistema de IA de alto riesgo

Según la Ley de Inteligencia Artificial, los sistemas de alto riesgo deben cumplir varios requisitos. Las preguntas que aparecen a continuación le ayudarán a hacerlo.

Efectos

- ¿El sistema de IA constituye un riesgo significativo para la salud, la seguridad o los derechos fundamentales de las personas?20 Por favor Proporcione a continuación las razones de por qué/por qué no.

Mantenimiento y gestión

- ¿Cómo se garantizará el seguimiento y el funcionamiento del sistema?21
- ¿Cómo se conservarán los registros producidos por el sistema de IA durante al menos seis meses?22
- ¿Se han adoptado medidas para garantizar la supervisión humana por parte de personas con la competencia, la formación y la autoridad necesarias?

Robustez técnica

- ¿Son los datos (de entrada) pertinentes y representativos, teniendo en cuenta el propósito previsto (pregunta 1 de 1.1)? del sistema de IA?23
- ¿Se ha proporcionado documentación técnica para el sistema de IA?24
- A los efectos de la supervisión humana, ¿es posible verificar, interpretar o posiblemente ignorar los resultados? (datos)?25

Fiabilidad

- ¿Se han incluido las métricas para los niveles de precisión en las instrucciones de usuario?

Gobernanza de datos

- ¿Los datos de entrenamiento del sistema de IA cumplen los siguientes requisitos de calidad?
- Opciones de diseño relevantes para conjuntos de datos.
- Proporcionar transparencia sobre el origen de los datos.
- Actividades de procesamiento relevantes, como anotación, etiquetado, limpieza, actualización, mejora y agregación.
- El establecimiento de supuestos que los datos deben medir y representar.
- Una evaluación de la disponibilidad, cantidad e idoneidad de los conjuntos de datos necesarios.
- Una evaluación del sesgo potencial en los datos con consecuencias negativas para la salud, la seguridad y la salud fundamental. Derechos y discriminación.

19 Ley de Inteligencia Artificial, artículo 25.
20 Ley de IA, artículo 6.2a.
21 Ley de IA, artículo 26.5.
22 Ley de IA, artículo 26.6.
23 Ley de IA, artículo 26.4.
24 Ley IA, Anexo IV.
25 Ley de IA, artículo 14.4.

• Medidas para detectar, prevenir y limitar los sesgos.
• Identificar y contrarrestar las deficiencias que impiden el cumplimiento normativo.

Gestión de riesgos

• ¿Cómo se probó el sistema de IA para su propósito previsto y para garantizar que cumple con la gestión de riesgos?
¿Requisitos antes de su puesta en servicio?
• ¿Se ha determinado y documentado un sistema de gestión de riesgos? Este incluirá los siguientes pasos:
- Un análisis de riesgos del sistema de IA para la salud, la seguridad o los derechos fundamentales
- Una evaluación y valoración de los riesgos que puedan ocurrir
- Una evaluación de nuevos riesgos tras la entrada al mercado, basada en el sistema de seguimiento del sistema de IA26
- La elaboración de medidas de gestión de riesgos.
• ¿Cómo se salvaguardará la posible colaboración con las autoridades supervisoras y otras autoridades competentes?27 Esto se refiere
a áreas como las personas de contacto, la accesibilidad, etc.
• ¿Es probable que los grupos vulnerables (como los niños) tengan acceso al sistema de IA? En ese caso, el riesgo
Los sistemas de gestión deberán ser especialmente estrictos.

NB: La oficina de IA está desarrollando una plantilla para un cuestionario, utilizando en parte una herramienta automatizada, para proporcionar a los implementadores una forma simplificada de cumplir con estas obligaciones.

Comunicación

• ¿Cómo se comunicará sobre la puesta en servicio del sistema de IA de alto riesgo?28
• ¿Se ha registrado el sistema de IA en la base de datos de la UE para sistemas de alto riesgo?29
• ¿Se han elaborado instrucciones de uso? Éstas deben contener al menos lo siguiente:30
- La identidad y datos de contacto del prestador;
- Características, capacidades y limitaciones (finalidad);
- Posibles cambios futuros;
- Medidas relativas a la supervisión humana;
- Los recursos computacionales y de hardware necesarios, la vida útil esperada y cualquier medida de mantenimiento y cuidado necesaria
(incluida la frecuencia);
- Una descripción de los mecanismos incluidos en el sistema de IA; - Los niveles
de precisión y las métricas de precisión relevantes.

IA generativa

• ¿Existe información y documentación para aclarar las posibilidades y limitaciones del sistema de IA?
• ¿Existe una política respecto a cómo el sistema de IA puede proteger los derechos de autor?
• ¿Se ha proporcionado un resumen detallado del contenido con el que se entrenó la IA?

Misceláneas

• Si se utiliza identificación biométrica remota, ¿se ha otorgado autorización judicial?
¿autoridad?31

26 El sistema de seguimiento posterior a la comercialización deberá recopilar, documentar y analizar de forma activa y sistemática los datos pertinentes que:
pueden ser proporcionados por los implementadores o pueden ser recopilados a través de otras fuentes sobre el rendimiento de los sistemas de IA de alto riesgo a lo largo de su vida útil, y que permiten al proveedor evaluar el cumplimiento continuo de los sistemas de IA con los requisitos establecidos en el Capítulo III, Sección 2.

27 Ley de IA, artículo 26.12.

28 Ley de IA, artículo 26.7.

29 Ley de Inteligencia Artificial, artículo 26.8 y artículos 49 y 71.

30 Ley de Inteligencia Artificial, artículos 13.2 y 13.3.

31 Ley de IA, artículo 26.10.

Preguntas para desarrolladores (proveedores) de sistemas de IA de alto riesgo

Los desarrolladores (o para usar el término en la Ley de IA: proveedores32) de un sistema de IA de alto riesgo también deben responder las siguientes preguntas.

- ¿Se ha recopilado documentación que demuestre el cumplimiento de los requisitos? Véase el Anexo IV de la Ley de IA para los requisitos.
- ¿Cómo pueden los usuarios encontrar la documentación técnica? Incluso si no hay participación en la implementación de la ¿su sistema?
- ¿Se ha indicado que se trata de un sistema de alto riesgo y, en caso afirmativo, cómo?33
- ¿Se ha establecido un sistema de gestión de calidad?34 También deberá documentar esto. Lo siguiente Se requerirá información mínima:
  - Una estrategia para el cumplimiento normativo
  - Procedimientos de diseño
  - Procedimientos de control de calidad
  - Un procedimiento de inspección
  - Una descripción general de los requisitos técnicos y lo que se necesita para hacerlos cumplir.
  - Procedimientos de gestión de datos
  - Un sistema de gestión de riesgos -
  Un procedimiento de seguimiento
  - Un procedimiento para notificar incidentes graves
  - Una estrategia de comunicación con las autoridades competentes
  - Sistemas y procedimientos para el registro de la documentación pertinente - Gestión de recursos y medidas relacionadas con la seguridad del suministro
  - Un marco de rendición de cuentas
- ¿Cómo se almacenarán los registros del sistema de IA?35
- ¿Existe un procedimiento de evaluación de la conformidad y se ha registrado en una declaración UE de conformidad? ¿ Cómo se demostrará que se han cumplido los requisitos determinados para el sistema de IA? Estos deben presentarse a un organismo designado y conservarse durante al menos diez años.
- ¿La declaración de conformidad ha sido aprobada por la autoridad supervisora pertinente?37 Esto es necesario sólo si el sistema de IA de alto riesgo se relaciona con un producto de alto riesgo relevante (Apéndice I).
- Si el sistema de IA está destinado a cumplir con la Ley de IA, ¿se ha proporcionado el marcado CE, por ejemplo en el ¿documentación?38
- Si el sistema ya no cumple con la Ley de IA, ¿cómo se tomarán medidas correctivas, como ¿Retirar este sistema del mercado, desactivarlo o retirarlo del mercado? ¿Cómo se informará a los usuarios sobre esto?39
- ¿El sistema cumple con los requisitos europeos de accesibilidad?40

32 'Proveedor': persona física o jurídica, autoridad pública, agencia u otro organismo que desarrolla un sistema de IA o un modelo de IA de propósito general o que tiene un sistema de IA o un modelo de IA de propósito general desarrollado y lo comercializa o pone el sistema de IA en servicio bajo su propio nombre o marca registrada, ya sea a cambio de un pago o de forma gratuita.

33 Ley de Inteligencia Artificial, artículo 16.

34 Ley de IA, artículo 17.

35 Ley de Inteligencia Artificial, artículo 19.

36 Ley de Inteligencia Artificial, artículos 43 y 47.

37 Ley de Inteligencia Artificial, artículo 16k.

38 Ley de Inteligencia Artificial, artículo 48.

39 Ley de Inteligencia Artificial, artículo 20.

40 Directivas (UE) 2016/2102 y (UE) 2019/882.

Apéndice 3: Puntos a tener en cuenta en relación con la IA generativa

Para el uso de la IA generativa, como los modelos de lenguaje extensos (LLM), hay una serie de puntos importantes que se deben tener en cuenta al completar el AIIA. Este apéndice se centra en las formas más importantes en las que la IA generativa se diferencia de otros sistemas de IA.

¿Qué es la IA generativa?

La IA generativa es un tipo de IA en el que se utilizan algoritmos para generar contenido. En su visión para todo el gobierno sobre la IA generativa41, el Gabinete holandés ha declarado que la IA generativa debe servir al propósito de mejorar el bienestar y la autonomía humanos, la sostenibilidad, la prosperidad, la justicia y la seguridad. Según la visión, al aplicar aplicaciones responsables de la IA generativa, estamos aprovechando las oportunidades que esta tecnología tiene para ofrecer.

En esencia, la IA generativa utiliza una red neuronal compuesta por miles de millones de parámetros. El resultado que se elige se basa en estadísticas y no hay ninguna lógica subyacente ni conocimiento sobre la realidad. El resultado varía y es difícil de reproducir o explicar. Esto nos lleva inmediatamente al punto más importante que se debe considerar antes de completar el AIIA: solo se debe utilizar la IA generativa si es aceptable que el resultado no sea explicable o verificable.

> Posición provisional para las organizaciones del gobierno central: en principio no se permite
> La posición provisional sobre el uso de la IA generativa en las organizaciones del gobierno central establece actualmente requisitos estrictos para el uso de LLMS en el gobierno central: “Las aplicaciones de IA generativa no contratadas, como ChatGPT, Bard y Midjourney, por lo general no cumplen de manera demostrable con la legislación pertinente en materia de privacidad y derechos de autor. Debido a esto, su uso por parte de (o en nombre de) organizaciones del gobierno central en principio no está permitido en aquellos casos en los que exista riesgo de que se infrinja la ley, a menos que el proveedor y el usuario cumplan de manera demostrable con las leyes y regulaciones pertinentes”.

Puntos a tener en cuenta al completar el AIIA

A continuación se describe las formas más importantes en las que la IA generativa se diferencia de otros sistemas de IA.
Esto será de ayuda a la hora de completar la AIIA. Si alguna de las áreas de la AIIA no se incluye a continuación, esto no significa que no sea relevante, sino que no hay puntos específicos a tener en cuenta en relación con el hecho de que se trata de una IA generativa.

Propósito y necesidad

- Valores públicos: no hay puntos específicos que considerar en el caso de la IA generativa. La decisión de utilizarla puede a menudo socavar la transparencia, la explicabilidad y la sostenibilidad.
- Derechos fundamentales: comprobar si el modelo ha sido entrenado y ajustado de forma ética. Este es el paso final en el que se utiliza el feedback humano para mejorar las respuestas. Como ocurre con muchos otros productos, estas personas ('click workers') no siempre tienen las mejores condiciones de trabajo.42
- Sostenibilidad: averiguar si es posible alcanzar los mismos objetivos utilizando un método menos complejo.
  Sistema o herramienta. La IA generativa consume enormes cantidades de energía tanto para el entrenamiento como para el uso del modelo.
  Investigar si se pueden aplicar tecnologías de ahorro de energía, como un modelo más pequeño, optimización o una arquitectura de software diferente.
- Consideraciones: al decidir si utilizar o no IA generativa, debe tener en cuenta lo siguiente:
  Decisión provisional sobre el uso de IA generativa en organizaciones del gobierno central (véase la introducción).

41 https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2024Z00480&did=2024D01191
42 AI draait op werk van miljoenen onzichtbare, slechtbetaalde mensen. ¿Wie komt er voor ze op? - El corresponsal

Robustez técnica

• Sesgo: compruebe siempre si el resultado del sistema tiene sesgos y evalúe los riesgos de que existan. Los datos utilizados para entrenar un modelo de IA generativa suelen contener sesgos, y la naturaleza del modelo (modelo de lenguaje estadístico) exacerba estos sesgos. Aunque los proveedores de software utilizan la retroalimentación humana y barandillas para eliminar los bordes afilados, el sesgo permanece en el modelo. Además, hay otros métodos disponibles para reducir y/o detectar sesgos.43

• Precisión, fiabilidad y reproducibilidad: solo debe utilizar un modelo de IA generativa para fines que no requieran un alto nivel de precisión, fiabilidad y reproducibilidad y/o debe tomar medidas al configurar el sistema para reducir la probabilidad de alucinaciones. Realice una evaluación de los riesgos potenciales de un resultado poco fiable. Un LLM es un modelo de lenguaje estadístico sin conocimiento de la realidad.
El resultado siempre es diferente y no necesariamente se basa en la realidad. Un modelo de IA generativa también tiene el potencial de "alucinar", generando respuestas que pueden parecer lógicas y convincentes, pero no son ciertas.
Otra característica desafiante de la IA generativa es que es difícil medir la calidad del resultado.
• Implementación técnica: en el caso de un modelo de IA generativa alojado externamente, se deben hacer acuerdos contractuales sobre el uso de los datos de la organización (entrada). A menudo existe una demanda de utilizar estos datos como datos para entrenar más el modelo. Existe la posibilidad de que otras personas (no autorizadas) vean estos datos (durante el uso) si hacen uso de las indicaciones adecuadas. Al adquirir un sistema, el proveedor debe demostrar que cumple con todos los requisitos establecidos. La posición del Gabinete holandés sobre la IA generativa expresa una preferencia por el uso de IA generativa de código abierto.

• Explicabilidad: Sólo se deben utilizar modelos de IA generativos si la explicabilidad de los resultados no es un requisito. Realice una evaluación de los riesgos potenciales de esta falta de explicabilidad. Explique cuáles son las limitaciones para el personal que tenga que trabajar con el sistema. Una IA generativa es un sistema altamente complejo con miles de millones de parámetros, cuyo funcionamiento no es explicable: es una "caja negra". Es prácticamente imposible demostrar cómo el modelo llega a una respuesta, aunque actualmente existen algunas iniciativas para mejorar la transparencia de esto.44

Gobernanza de datos

• Calidad e integridad de los datos: es probable que los datos de entrenamiento de la IA generativa contengan datos personales y material protegido por derechos de autor. En el caso de la mayoría de los modelos de IA, se desconoce qué datos de entrenamiento (y cuál es su calidad) se han utilizado en el modelo. La Ley de IA exige transparencia. El RGPD no permite que los datos personales se procesen sin más.

• Privacidad: llegar a acuerdos con el proveedor respecto a la responsabilidad del tratamiento de los datos
Introducir los datos (por ejemplo, si se utilizarán como datos de formación). Redactar un contrato de tratamiento adecuado. Si se celebran los acuerdos adecuados, se puede reducir la probabilidad de que se produzcan violaciones de datos.
• Seguridad de la información: la OWASP ha compilado una lista de las diez vulnerabilidades más críticas y, por lo tanto, riesgos de seguridad de la información, de los sistemas de IA generativa45.

Rendición de cuentas

• Verificabilidad: véase “explicabilidad”.

43 https://www.datacamp.com/blog/understanding-and-mitigating-bias-in-large-language-models-llms
44 https://arxiv.org/abs/2309.01029
45 OWASP Top 10: LLM y riesgos de seguridad de IA generativa

Esta es una publicación de:

Ministerio de Infraestructura y Gestión del Agua

Apartado de correos 20901

2500 EX La Haya

Diciembre de 2024 | 73263