

Estratexia Galega
de Intelixencia Artificial 2030

Guía práctica para a xestión da intelixencia artificial nas Administracións públicas



XUNTA
DE GALICIA



AXENCIA PARA A
MODERNIZACIÓN
TECNOLÓXICA DE GALICIA

Edita. Xunta de Galicia. Axencia para a Modernización Tecnolóxica de Galicia (AMTEGA)

Santiago de Compostela. Ano 2023

Os textos deste documento distribúense baixo licenza Creative Commons Atribución-Compartir Igual 4.0 Internacional (CC BY-SA 4.0). Dispoñible en: <https://creativecommons.org/licenses/by-sa/4.0/deed.es>. As imaxes foron obtidas de «Deloitte Brand Space»; non está permitida a súa extracción ou redistribución como imaxes independentes, separadas do documento orixinal.



AXENCIA PARA A
MODERNIZACIÓN
TECNOLÓXICA DE GALICIA



Julián Cerviño Iglesia

Director da Axencia para a Modernización Tecnolóxica de Galicia

O desenvolvemento da intelixencia artificial está chamado a marcar un antes e un despois na traxectoria da Humanidade. Por primeira vez, estamos a crear máquinas coa capacidade de procesar con criterios intelixentes volumes inxentes de datos e con capacidade para facer moitas cousas. E iso debe ser obxecto dun saber, onde habemos de ter claro o que estamos a facer e onde queremos chegar con iso. Porque o reto das sociedades modernas e humanistas é non perder a referencia; as máquinas fanas persoas.

Agora estamos perante un horizonte totalmente novo e con enormes incertezas. Cómpre que pensemos que o desenvolvemento da intelixencia artificial supón unha enorme oportunidade para procesar rapidamente e resolver dilemas que se están cuestionando sobre o noso dereito á intimidade. O noso continente foi pioneiro ao propoñer a primeira lexislación de IA do mundo, que se publicou en 2021, baseada na importancia dunha xestión adecuada dos riscos asociados a esta tecnoloxía.







No marco do sector público, a intelixencia artificial permitirá transformar a relación entre a Administración pública e a cidadanía, e impulsará a evolución cara a un goberno máis intelixente e centrado nas persoas. Ao permitir análises de datos máis rápidas e precisas, esta tecnoloxía creará unha Administración máis áxil e con capacidade para tomar mellores decisións, á vez que impulsará unha maior transparencia e adaptabilidade ás necesidades cambiantes da cidadanía.

Consciente deste reto, a Xunta de Galicia asumiu xa en 2021 un compromiso estratéxico por garantir o uso eficaz, eficiente, ético e seguro da intelixencia artificial na Administración autonómica. Así se plasma na Estratexia Galega de Intelixencia Artificial 2030, folla de ruta coa que a Xunta de Galicia aspira a liderar a creación das condicións propicias para o desenvolvemento e a implantación dunha intelixencia artificial ética e fiable, que respecte os valores e normas da rexión, reforce a súa presenza nas cadeas de valor da industria e, ante todo, sirva á cidadanía. Para iso, a Xunta de Galicia propuxo un texto no que será a primeira lexislación de intelixencia artificial dunha rexión europea co anteproxecto de lei para o desenvolvemento e impulso da intelixencia artificial en Galicia, que busca establecer un marco legal uniforme para o deseño, a adquisición, a implantación e o uso de sistemas de intelixencia artificial na Administración Xeral e do Sector Público Autonómico de Galicia e fixar os principios éticos reitores do deseño, a adquisición e a implantación dos sistemas de intelixencia artificial en Galicia. En definitiva, o obxectivo desta Lei é garantir que os sistemas de intelixencia artificial que se empregan na comunidade son fiables e respectuosos cos dereitos fundamentais das persoas.

Deste xeito, esta guía constitúe un compoñente esencial dentro deste marco estratéxico e normativo co propósito de proporcionar as claves necesarias para que calquera Administración pública identifique as distintas fases e axentes intervinientes na xestión de proxectos que integren sistemas de intelixencia artificial.

Acerca da Axencia para a Modernización Tecnolóxica de Galicia

A Axencia para a Modernización Tecnolóxica de Galicia (AMTEGA) é o instrumento estratéxico e executivo da Xunta de Galicia para impulsar a modernización da Administración pública, así como o desenvolvemento económico e social da rexión. Constituída a través do Decreto 252/2011, do 15 de decembro, a súa actuación responde aos seguintes obxectivos:

-  Poñer a disposición da cidadanía e das empresas de Galicia contidos e servizos públicos dixitais avanzados
-  Dotar á cidadanía das competencias e recursos para participar activamente no desenvolvemento da sociedade da información
-  Transformar, mediante un uso acomodado das TIC, a Administración pública autonómica, coa finalidade de mellorar a calidade de prestación dos servizos públicos
-  Impulsar o hipersector TIC de forma que se converta nun soporte para incrementar a competitividade e o emprego
-  Impulsar o uso dos servizos TIC por parte das empresas galegas
-  Implantar unha rede de infraestruturas moderna e sostible que garanta a integración de Galicia na sociedade da información



Por que esta guía?

A intelixencia artificial é unha tecnoloxía disruptiva que ten a capacidade de xerar un impacto significativo en diversos aspectos da sociedade e a economía. Entre as razóns que fan da intelixencia artificial unha tecnoloxía revolucionaria, cabe destacar a súa contribución a unha mellor toma de decisións por parte das persoas, en tanto que é capaz de analizar grandes conxuntos de datos e extraer información valiosa deles de forma áxil. A intelixencia artificial tamén permite a automatización de procesos en diversos ámbitos. En definitiva, no caso do sector público, estas e outras aplicacións da intelixencia artificial permiten que as persoas centren os seus esforzos en tarefas de valor engadido e tomen decisións apoiadas en datos para dar lugar a mellores servizos ou a unha mellor atención, entre outros beneficios.

Con todo, é importante sinalar que o uso da intelixencia artificial tamén expón desafíos éticos e sociais que deben abordarse de maneira responsable. A importancia da intelixencia artificial radica non só nos seus beneficios potenciais, senón tamén na necesidade de abordar as súas implicacións éticas e sociais a medida que se avanza no seu desenvolvemento e adopción. Así o recoñece a Unión Europea, que asumiu o compromiso de liderar a nivel mundial o establecemento de normas harmonizadas en materia de xestión de sistemas de IA e dos riscos asociados aos devanditos sistemas.

Por iso, esta guía, de forma coherente coa normativa europea, pretende servir como soporte na xestión de proxectos que integren sistemas de intelixencia artificial na Administración pública e achegar os pasos que se han de dar e os controis de riscos que se han de realizar desde que se concibe unha oportunidade ata que se converte nun proxecto, de forma que se garanta a óptima adopción desta tecnoloxía.

Para quen é esta guía?

Esta guía diríxese ao conxunto das Administracións públicas, organismos e entidades que conforman o sector público. A guía achega unha serie de pautas e indicacións que poden ser adaptadas por cada Administración pública, de acordo coa súa idiosincrasia.

Índice de contidos

Capítulo 1

A intelixencia artificial nas políticas públicas

7

1.1 Que é a intelixencia artificial

9

1.2 A intelixencia artificial no ciclo de vida das políticas públicas

12

Capítulo 2

O ciclo de vida dun sistema de intelixencia artificial

21

2.1 O sistema de intelixencia artificial

23

2.2 O ciclo de vida dun sistema de IA

25

2.3 A xestión do ciclo de vida dun sistema de IA

27

Capítulo 3

Fases do ciclo de vida dun sistema de intelixencia artificial

45

3.1 Fase 1. Identificación e avaliación de oportunidades vinculadas á adopción da IA

48

3.2 Fase 2. Deseño, desenvolvemento e implantación do sistema de IA

62

Capítulo 4

Xestión de riscos dun sistema de IA

69

4.1 Introducción á xestión de riscos

71

4.2 Modelo de xestión de riscos dun sistema de IA

72

4.3 Exemplo de xestión de riscos no deseño, desenvolvemento e implantación dun sistema de IA

84

Conclusións

96

Referencias

98

Capítulo 1

A intelixencia artificial nas políticas públicas

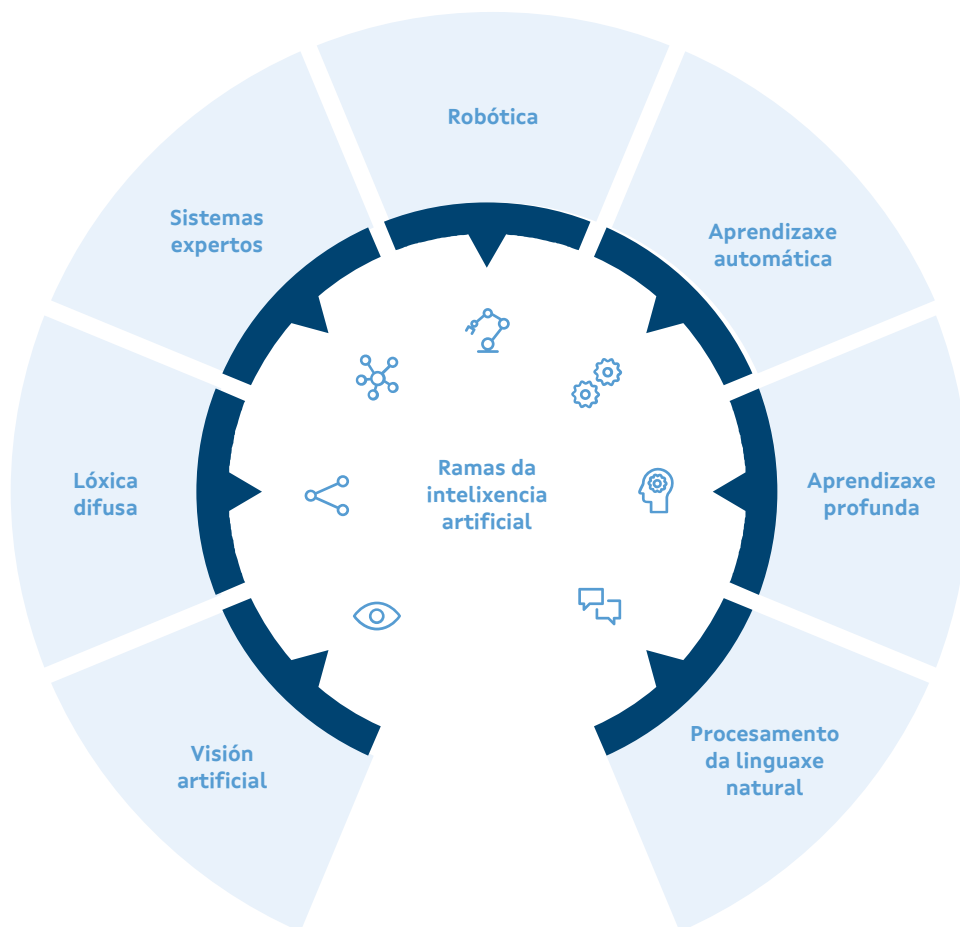
Nun mundo tecnolóxico en constante evolución, a intelixencia artificial (IA) emerxe como unha forza transformadora no ciclo de vida das políticas públicas. Este capítulo explora como a IA, ao automatizar procesos administrativos e mellorar a toma de decisións, convértese nun catalizador clave para a eficiencia e a adaptabilidade na Administración pública. Desde a realización de actos administrativos ata a anticipación de necesidades mediante análises predictivos, a IA redefine a interacción entre a Administración e a cidadanía, e proporciona eficiencia, personalización e transparencia. A implantación responsable da IA non só aborda ineficiencias, senón que tamén abre oportunidades para unha relación máis eficiente e satisfactoria coa sociedade.



1.1. Que é a intelixencia artificial

Neste contexto, a intelixencia artificial (IA) terá un impacto relevante en todos os ámbitos da sociedade. Esta tecnoloxía busca emular a capacidade cognitiva dos seres humanos a través de algoritmos e modelos matemáticos. Trátase dun campo da informática que se enfoca en desenvolver sistemas e programas capaces de realizar tarefas que normalmente requirirían intelixencia humana. Estas tarefas inclúen o razoamento, a aprendizaxe, a percepción, a toma de decisións ou a comprensión da linguaxe natural, entre outras.

Aínda que o seu ritmo de evolución supón a aparición incesante de novas ramas de intelixencia artificial, poden identificarse as seguintes:





Visión artificial

Ten como obxectivo desenvolver técnicas que axuden as computadoras a ver e comprender imaxes e vídeos dixitais. Aplicar modelos de aprendizaxe automática a imaxes permite ás computadoras identificar obxectos, rostros, persoas, animais e máis. Os modelos algorítmicos axudan as computadoras a aprender por si mesmas sobre os contextos dos datos visuais e, con suficientes datos proporcionados a través dun modelo, as computadoras poden aprender a distinguir unha imaxe doutra.

Unha rede neuronal convolucional traballa xunto a un modelo para descompoñer as imaxes en píxeles, asignándolles etiquetas. A rede neuronal logo utiliza estas etiquetas para realizar convolucións, que é unha operación matemática en dúas funcións para producir unha terceira función, e facer predicións sobre o que ve.



Lóxica difusa

É unha técnica que axuda a resolver problemas ou afirmacións que poden ser verdadeiros ou falsos. Este método imita as decisións humanas ao considerar todas as posibilidades existentes entre os valores dixitais de «si» e «non». En poucas palabras, mide o grao en que unha hipótese é correcta. É unha forma conveniente e flexible de implantar técnicas de aprendizaxe automática e imitar o pensamento humano de maneira lóxica.



Sistemas expertos

Un sistema experto é un programa especializado nunha tarefa singular. Estes sistemas están deseñados principalmente para resolver problemas enleados con capacidades de toma de decisións similares ás humanas. Utilizan un conxunto de regras, chamadas regras de inferencia (por exemplo, nocións lóxicas de tipo «se-entón»), que unha base de coñecementos alimentada por datos define para eles.



Robótica

A robótica baseada en IA refírese á integración de sistemas e algoritmos de intelixencia artificial en robots para permitirlles realizar tarefas de maneira máis autónoma, adaptativa e eficiente. Neste contexto, a IA permite que os robots procesen información, aprendan de experiencias anteriores e tomen decisións baseadas en datos e patróns. En lugar de depender de programación ríxida e algoritmos predefinidos, os robots baseados en IA poden aprender e mellorar o seu rendemento co tempo. Isto implica a capacidade de adaptarse a ambientes cambiantes, tomar decisións en tempo real e enfrontarse a situacións novas sen intervención humana constante.



Machine learning (aprendizaxe automática)

A aprendizaxe automática é a capacidade das máquinas para aprender automaticamente a partir de datos e algoritmos. A aprendizaxe automática mellora o rendemento utilizando experiencias pasadas e pode tomar decisións sen estar programada especificamente para facelo. O proceso comeza coa recompilación de datos históricos, como instrucións e experiencia directa, para que se poidan construír modelos lóxicos para inferencias futuras. A precisión da saída depende do tamaño dos datos: unha maior cantidade de datos construírá un mellor modelo, o que á súa vez aumenta a súa precisión. Os algoritmos de aprendizaxe automática clasifícanse en tres tipos:

- Aprendizaxe supervisada: as máquinas adéstranse con datos etiquetados para predicir o resultado.
- Aprendizaxe non supervisada: as máquinas adéstranse con datos non etiquetados, e o modelo extrae información da entrada para identificar características e patróns, de modo que poida xerar un resultado.
- Aprendizaxe por reforzo: as máquinas aprenden a través de ensaio e erro, utilizando retroalimentación para formar accións.



Redes neuronais/aprendizaxe profunda (deep learning)

As redes neuronais tamén son coñecidas como redes neuronais artificiais ou redes neuronais simuladas. No corazón dos algoritmos de aprendizaxe profunda, as redes neuronais inspíranse no cerebro humano e copian como as neuronas biolóxicas envían sinais entre si.

As redes neuronais artificiais teñen capas de nodos, que constan dunha capa de entrada, unha ou máis capas ocultas e unha capa de saída. Cada nodo, tamén chamado unha neurona artificial, conéctase con outras neuronas e ten un limiar e un peso asociados. Cando a saída dun nodo individual supera un valor de limiar especificado, o nodo actívase para enviar datos á seguinte capa da rede. As redes neuronais necesitan datos de adestramento tanto para aprender como para mellorar a precisión



Procesamento da linguaxe natural

O procesamento da linguaxe natural permite ás computadoras entender tanto o texto como as palabras faladas, de maneira similar aos humanos. Combinando aprendizaxe automática, lingüística e modelos de aprendizaxe profunda, as computadoras poden procesar a linguaxe humana en datos de voz ou texto para comprender o significado completo, a intención e o sentimento.



1.2. A intelixencia artificial no ciclo de vida das políticas públicas

O ciclo de vida da política pública representa a forma na que se desenvolven as políticas públicas e serve, ademais, para planificar e analizar as diferentes fases deste proceso. O ciclo iníciase coa identificación e definición dun problema ou asunto a resolver. Neste punto, as Administracións han de avaliar as alternativas propostas polos distintos órganos xestores responsables desa política. Durante este proceso, a tecnoloxía actúa como un instrumento que pode servir para desenvolver as alternativas de política pública propostas a partir do contexto e a viabilidade da solución. Nese caso, a intelixencia artificial álzase como unha

ferramenta esencial para impulsar a calidade dos servizos públicos. Na Administración pública implántanse solucións de IA para analizar grandes volumes de datos, identificar patróns no comportamento cidadán e optimizar a asignación de recursos, entre outros múltiples usos. O uso da IA contribúe polo tanto ao impulso dunha Administración máis áxil e adaptativa, o que permite atinxir respostas máis efectivas aos retos e adaptadas aos cambios nas necesidades da cidadanía.

1.2.1. A relación da cidadanía coa Administración pública

A evolución da relación entre a Administración pública e a cidadanía experimentou un cambio drástico coa aparición da vía dixital. Antes, as interaccións limitábanse a procesos presenciais e trámites pouco transparentes, marcados por longas esperas e limitada accesibilidade. A participación cidadá no ciclo de vida das políticas públicas estaba condicionada por barreiras burocráticas e unha experiencia desigual.

Coa transición cara á vía dixital, este escenario sufriu unha transformación fundamental. A cidadanía agora pode interactuar coa Administración pública de maneira remota e aproveitar plataformas en liña, portais e aplicacións móbiles. Este cambio xerou un novo paradigma, onde a conveniencia e a axilidade definen a relación entre a cidadanía e a Administración.

A vantaxe primordial da vía dixital é a comodidade que ofrece. A posibilidade de realizar unha variedade de trámites de forma telemática e acceder a servizos públicos en calquera momento do día representa un cambio paradigmático na dinámica de interacción. A axilidade converteuse nunha característica distintiva desta vía, facilitada pola presentación electrónica de documentos, a realización de pagos en liña e a capacidade de dar seguimento en tempo real ao estado das solicitudes. A dixitalización de trámites non só reduciu os tempos de resposta, senón que tamén mellorou significativamente a experiencia do usuario.

A pesar destes avances, a vía dixital non está exenta de desafíos. A complexidade dalgúns procesos, a falta de uniformidade na atención á cidadanía e a fenda dixital son aspectos críticos que cómpre que se aborden para garantir unha relación efectiva entre a Administración e os cidadáns. A superación destes desafíos será crucial para consolidar os beneficios da transformación dixital e garantir unha interacción cidadanía-Administración inclusiva e equitativa. Este novo paradigma dixital, aínda que prometedor, demanda unha atención constante a fin de optimizar a relación e fortalecer a confianza nos servizos gobernamentais.

1.2.2. A IA como forza transformadora da relación Administración-cidadanía

O poder transformador da IA na relación entre a Administración e a cidadanía é innegable. A implantación de solucións baseadas en IA redefine a dinámica tradicional ao ofrecer eficiencia, personalización e anticipación na prestación de servizos públicos. Neste contexto, pódense diferenciar dúas grandes aplicacións da IA na Administración pública:

- Sistemas para realizar actos administrativos: refírese ao ámbito da IA no que a tecnoloxía é capaz de «tomar» decisións e «realizar» as accións resultantes sen a intervención directa de persoas. Isto implica que o sistema asume a execución de tarefas que anteriormente eran realizadas por persoas.

A IA actúa como un catalizador clave para automatizar procesos administrativos. A capacidade de sistemas avanzados para automatizar tarefas e procesos permite que as persoas se centren na resolución de cuestións complexas e estratéxicas. Esta automatización non só acelera os tempos de resposta na resolución de procedementos, senón que tamén contribúe a unha toma de decisións máis informada por parte das persoas, o que garante unha maior precisión e consistencia nos servizos públicos.

- Sistemas para realizar actuacións administrativas: no contexto desta tecnoloxía, estes sistemas abarcan aquelas instancias nas que a información producida polos modelos de IA se emprega como elemento clave no proceso de toma de decisións humanas.

A IA brinda a capacidade de realizar análises predictivas, o que permite á administración anticipar necesidades e axustar recursos e servizos de maneira proactiva. Esta capacidade de prever posibles obstáculos nos procesos administrativos mellora a eficiencia e a capacidade de resposta, e así garante que a Administración estea preparada para afrontar desafíos emerxentes e cambios nas demandas da cidadanía.

A personalización da interacción, facilitada pola IA, convértese nun elemento esencial para fortalecer a conexión coa cidadanía. Os sistemas de asistentes virtuais poden adaptarse ás necesidades específicas de cada persoa e proporcionar información e asistencia de maneira contextualizada. Este enfoque personalizado non só mellora a experiencia do usuario, senón que tamén contribúe a construír unha relación máis próxima e satisfactoria entre a Administración e a cidadanía.

Así mesmo, a IA facilita unha Administración máis transparente e eficiente. Os algoritmos de IA poden analizar grandes conxuntos de datos para identificar patróns, tendencias e áreas de mellora. Esta capacidade de análise profunda proporciona información valiosa para a toma de decisións informadas e permite que a Administración axuste as súas políticas públicas de maneira máis eficaz.

En conclusión, a implantación responsable da IA non só se erixe como a solución para as ineficiencias na Administración pública, senón que tamén representa unha fonte de oportunidades a futuro. Esta integración, levada a cabo de maneira consciente, xera unha transformación positiva na prestación de servizos públicos e establece unha relación máis eficiente e satisfactoria coa cidadanía. É imperativo destacar que a correcta adopción da IA debe ir da man das persoas da Administración, quen son en última instancia partícipes da aplicación desta tecnoloxía no ciclo de definición e execución de políticas públicas, o que asegura un enfoque equitativo e adaptativo ás necesidades cambiantes da sociedade.

Posibles aplicacións da IA na Administración pública

Resolución de consultas da cidadanía

Os *chatbots* e asistentes virtuais baseados en IA poden proporcionar respostas rápidas e precisas ás preguntas da cidadanía, o que mellora a calidade da atención e reduce a carga de traballo das persoas empregadas.

Reforzo da ciberseguridade

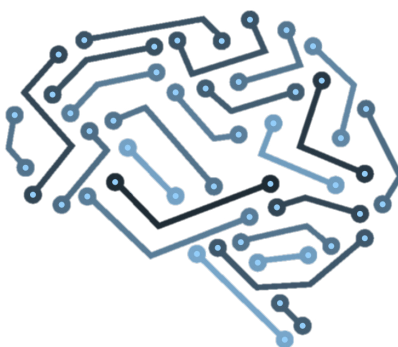
A IA utilízase na detección de ameazas cibernéticas e na protección de infraestruturas críticas, o que é fundamental nun mundo cada vez máis dixitalizado.

Detección e resposta a desastres naturais

A IA pode axudar na detección temperá de desastres naturais, como terremotos, inundacións e incendios forestais, e na coordinación da resposta de emerxencia, o que pode salvar vidas e reducir o impacto.

Predición de epidemias

A IA pode analizar grandes volumes de datos para identificar tendencias e patróns que poderían levar á detección temperá de problemas ou a anticipación de necesidades. Por exemplo, no ámbito da saúde pública, a IA pode predicir gromos de enfermidades antes de que se propague.



Tradución e servizos multilingües

A IA utilízase para a tradución automática de documentos e servizos multilingües, o que é fundamental nunha Unión Europea con múltiples idiomas oficiais. Isto facilita a comunicación e a colaboración nun contorno diverso.

Detección de fraudes e corrupción

A IA pode axudar na identificación de actividades fraudulentas e corruptas mediante a análise de datos financeiros e transaccionais. Isto é esencial para garantir a integridade e a transparencia no sector público.

Optimización do tráfico e transporte público

En áreas urbanas, a IA utilízase para optimizar o tráfico e o transporte público. Os algoritmos de IA poden mellorar a xestión de semáforos, predicir conxestións e proporcionar rutas máis eficientes para o transporte público, o que reduce a conxestión e mellora a mobilidade.

Planificación urbana e xestión de residuos

A IA pódese utilizar na planificación urbana para optimizar o uso de recursos, prever necesidades de infraestrutura e xestionar eficientemente os residuos e a enerxía nas cidades.

1.2.3. Aplicacións da IA na Administración pública a escala mundial

O sector público global non é alleo ao impacto das aplicacións de IA e comezou xa a aplicar esta tecnoloxía nos seus procesos. Estes exemplos ilustran como a IA se integrou en diversas áreas das Administracións públicas en todo o mundo e mellorou a eficiencia, a calidade dos servizos e a toma de decisións, entre outros aspectos.







Transformando a atención médica: o NHS do Reino Unido adopta chatbots e IA para mellorar o acceso e a investigación médica

No Reino Unido, a implantación de intelixencia artificial no Servizo Nacional de Saúde (NHS) revolucionou a atención médica. A introdución de *chatbots* e asistentes virtuais mellorou significativamente a accesibilidade á información de saúde para os pacientes, o que ofreceu respostas rápidas a consultas comúns e facilitou a xestión de citas médicas. Ademais, exploráronse aplicacións máis avanzadas de IA na investigación médica, o que mellorou a precisión no diagnóstico e impulsou o desenvolvemento de tratamentos personalizados.



Transformando a atención médica: o NHS do Reino Unido adopta chatbots e IA para mellorar o acceso e a investigación médica

Nos Estados Unidos, a Administración de Alimentos e Medicamentos (FDA) adoptou a intelixencia artificial para axilizar a revisión e aprobación de medicamentos e dispositivos médicos. Os algoritmos de IA analizan datos clínicos e científicos de maneira eficiente, acelerando así o proceso de aprobación e mellorando a capacidade de tomar decisións informadas sobre a seguridade e eficacia de novos produtos médicos.



Smart Nation: Singapur utiliza a IA para unha xestión urbana eficiente e sostible

Singapur, a través da súa iniciativa «Smart Nation», impulsou o uso da intelixencia artificial na xestión do tráfico, a seguridade pública e a mellora de servizos públicos. A cidade utiliza tecnoloxías avanzadas, como sensores e análises de datos, para tomar decisións informadas en tempo real e crear un contorno urbano máis eficiente e sostible.



Infraestruturas intelixentes: Países Baixos moderniza a xestión de estradas e pontes con IA

A implantación da intelixencia artificial nos Países Baixos centrouse na xestión de infraestruturas críticas. A Axencia Holandesa de Estradas adoptou tecnoloxías avanzadas para o mantemento e a inspección de estradas e pontes, o que mellorou a eficiencia operativa e reduciu os custos asociados coa infraestrutura vital.



**Avanzando na saúde e a mobilidade:
Francia implanta a IA en diagnósticos
médicos e transporte urbano**

Francia aplicou a intelixencia artificial na saúde e a mobilidade urbana. No sector da saúde, a IA utilizouse para analizar imaxes médicas e xestionar electronicamente rexistros médicos, o que mellorou a precisión do diagnóstico e a eficiencia na atención médica. No ámbito da mobilidade, proxectos de IA contribuíron a unha xestión máis eficiente do transporte público e a redución de conxestións nas cidades francesas.



**Educación personalizada e robótica:
Xapón lidera con sistemas de titoría
baseados en IA e asistencia a persoas
maiores**

Xapón destaca pola aplicación de intelixencia artificial na educación e a robótica. Os sistemas de titoría personalizados baseados en IA melloraron a experiencia educativa dos estudantes xaponeses ao adaptarse ás súas necesidades individuais. Ademais, a implantación da IA en proxectos de robótica e asistencia a persoas maiores contribuíu á mellora da calidade de vida e a autonomía da poboación maior.



**Energía do futuro: Australia impulsa
a sostibilidade con sistemas de IA na
xestión de redes eléctricas**

Australia implantou a intelixencia artificial cun enfoque especial na xestión da enerxía. Os sistemas de IA utilízanse na distribución de recursos en redes eléctricas, o que mellora a eficiencia e a resiliencia do sistema enerxético. Este enfoque contribuíu á transición cara a fontes de enerxía máis sostibles e fortaleceu a infraestrutura enerxética do país.



**Fronteiras eficientes: a Axencia
Canadense de Servizos Fronteirizos de
Canadá implanta a IA para facilitar o
comercio transfronteirizo e reforzar a
seguridade**

A Axencia Canadense de Servizos Fronteirizos (CBSA) adoptou a intelixencia artificial para axilizar os procesos nos cruzamentos fronteirizos. Os sistemas de IA axudan na detección de riscos e facilitan o comercio transfronteirizo ao mellorar a eficiencia no procesamento de viaxeiros e mercadorías. Esta implantación fortaleceu a seguridade fronteiriza e impulsou a fluidez no intercambio comercial.

Mobilidade

Educación

Asistencia

Energía

Comercio

Seguridade

1.2.2. Aplicacións da IA na Administración pública a escala nacional, autonómica e local

A Administración pública en España tamén é coñecedora do potencial da IA e está a demostrar un crecente interese e compromiso na súa adopción. Tanto a escala nacional como autonómica e local, a Administración pública conta con numerosos exemplos de aplicación práctica da IA na súa actividade.

A Secretaría de Estado de Dixitalización e Intelixencia Artificial (SEDIA) do Ministerio de Asuntos Económicos e Transformación Dixital impulsou a elaboración dun selo nacional que certifique a seguridade, calidade e responsabilidade dos sistemas de intelixencia artificial desenvolvidos en España.





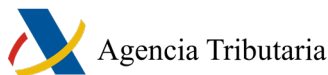
Detección temprá de enfermidades

O Ministerio de Sanidade utiliza a IA na detección temprá de gromos de enfermidades, como a gripe, a través da análise de datos e o seguimento de síntomas. Así mesmo, durante a pandemia de COVID-19, desenvolvéronse modelos de IA para predicir a propagación do virus e xestionar recursos de atención médica.



Chatbots de atención á cidadanía

O SEPE implantou *chatbots* de IA no seu sitio web para axudar a cidadanía a obter información sobre o desemprego, a procura de traballo e outros servizos relacionados co emprego.



Identificación de fraude

A Axencia Tributaria emprega a IA para identificar evasión fiscal e fraude a través da análise de grandes conxuntos de datos. A IA axuda a detectar patróns de evasión e a tomar medidas para previlo.



Xestión do tráfico

O Goberno Vasco emprega a IA na optimización do tráfico e o transporte público. Desenvolvéronse sistemas de IA para mellorar a xestión de semáforos e a mobilidade urbana.



Asistente virtual de atención turística

O Goberno de Navarra creou un asistente virtual impulsado por IA xenerativa co obxectivo de mellorar a atención turística a visitantes e cidadanía e promover a adopción de novas tecnoloxías no sector turístico.



Xestión da mobilidade urbana

Barcelona utiliza a IA para mellorar a xestión da mobilidade urbana. Implantaron sensores e algoritmos de IA para controlar o tráfico, predicir atascos e mellorar a mobilidade na cidade.



Xestión da calidade do aire

O Concello de Valencia desenvolveu sistemas de IA para controlar a calidade do aire e a xestión da contaminación. Estes sistemas proporcionan información en tempo real sobre a calidade do aire na cidade e recomendacións de actuación para mellorar os indicadores.

Capítulo 2

O ciclo de vida dun sistema de intelixencia artificial

Neste capítulo exploramos o rápido avance da intelixencia artificial (IA) e a definición precisa dun sistema de IA segundo a normativa europea. Detállase o ciclo de vida destes sistemas, dividido en dúas fases: identificación e avaliación de oportunidades, e deseño, desenvolvemento e implantación. A xestión integral, destacada na futura Lei de intelixencia artificial, enfócase na transparencia, a obxectividade e fiabilidade. Resáltase a importancia da xestión de riscos en todas as etapas, coa clasificación dos sistemas segundo o seu nivel de risco e a imposición de obrigacións proporcionais.



2.1. O sistema de intelixencia artificial

A intelixencia artificial evoluciona a un ritmo imparabile e xera innovacións constantes en tempos cada vez máis curtos que impulsan novos debates sobre o seu uso ético e sobre a xestión dos riscos asociados a esta tecnoloxía. Este vertixinoso progreso expón desafíos e oportunidades, e require unha adaptabilidade continua na comprensión desta tecnoloxía en constante cambio.

A Unión Europea, pioneira na regulación da IA a escala global, é consciente de que é necesario que a definición dun sistema de IA sexa tan concreta que permita acoutar a súa natureza, pero que ao mesmo tempo sexa o suficientemente flexible para abarcar calquera innovación emerxente arredor desta tecnoloxía. Por iso, a normativa comunitaria achega unha definición que combina claridade con adaptabilidade, o que garante unha comprensión completa do seu alcance e permite á súa vez entender mellor os elementos clave para a súa correcta xestión.

Un sistema de IA defínese como «o *software* que se desenvolve empregando unha ou varias das técnicas e estratexias de IA e que pode, para un conxunto determinado de obxectivos definidos por seres humanos, xerar información de saída como contidos, predicións, recomendacións ou decisións que inflúan nos contornos cos que interactúa».

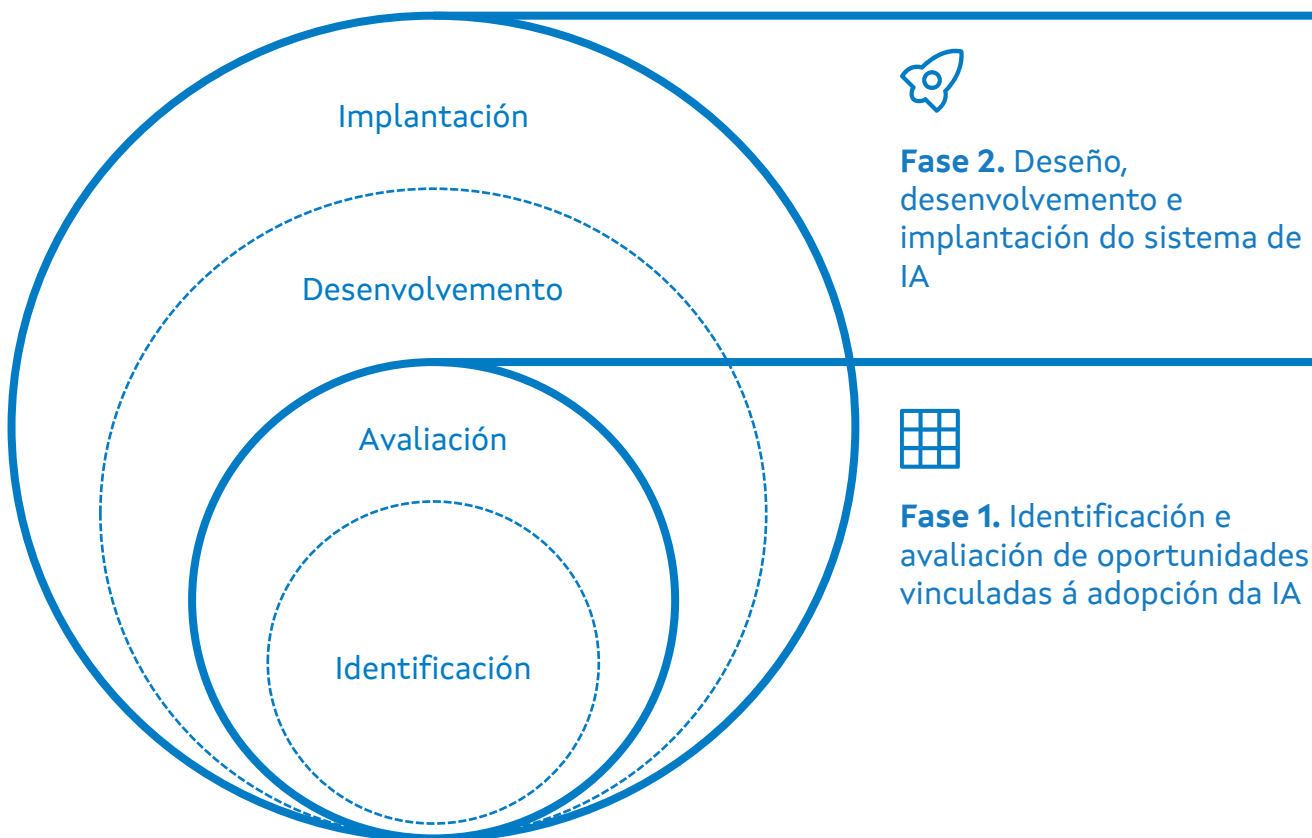
Constátase que esta definición reflicte unha comprensión integral e coherente da IA no ámbito da Administración pública e recoñécese a súa capacidade de autonomía e a súa función asistida na toma de decisións humanas. No primeiro caso, a capacidade de «xerar información de saída» e «tomar decisións» sen intervención directa de persoas concorda coa noción de sistemas para levar a cabo actos administrativos. No segundo caso, a referencia a «obxectivos definidos por seres humanos» e a xeración de información que inflúe nos contornos alíñase coa idea de sistemas para levar a cabo actuacións administrativas, onde a información producida polos modelos de IA se emprega no proceso de toma de decisións humanas.



2.2. Xestión do ciclo de vida dun sistema de IA

2.2.1.O ciclo de vida dun sistema de IA

O ciclo de vida dun sistema de IA é o proceso polo cal unha Administración pública identifica unha barreira ou área de mellora que pode ser cuberta mediante a aplicación de IA, analiza a idoneidade de adoptar un sistema de IA e conceptualiza e executa un proxecto para a súa implantación. Este proceso implica a análise continua do risco da implantación da IA, así como a avaliación das capacidades das que dispón a Administración para adoptar o sistema. En termos globais, poden diferenciarse dúas grandes fases no ciclo de vida dun sistema de IA:





Fase 1. Identificación e avaliación de oportunidades vinculadas á adopción da IA

O ciclo de vida dun sistema de IA na Administración pública comeza coa identificación da barreira ou área de mellora e a análise da idoneidade de aplicar IA para a súa cobertura. Neste punto, vólvese esencial que aquelas persoas encargadas de identificar oportunidades poidan definir de maneira nítida o desafío en termos de políticas públicas ou xestión interna ao cal buscan dar resposta mediante a incorporación dun sistema de IA. É necesario que a Administración pública teña no seu ámbito de actuación a competencia para facer fronte ao problema ou circunstancia da que deriva a oportunidade.

Unha vez identificada a barreira ou área de mellora, resulta imprescindible que a Administración pública conte con elementos obxectivos e transparentes para a súa avaliación e potencial transformación nun proxecto para a implantación dun sistema de IA. Estes elementos permiten, por unha banda, asegurar a viabilidade do futuro proxecto e, doutra banda, asegurar a súa aliñación cos requirimentos normativos, éticos ou doutra índole que deban observarse na adopción da IA.



Fase 2. Deseño, desenvolvemento e implantación do sistema de IA

A partir da avaliación das oportunidades e da verificación da súa viabilidade, a Administración pública ha de concretar os requisitos de arquitectura e funcionalidade que deberá incorporar o sistema de IA que materializará esa oportunidade. Unha vez desenvolvido o sistema de IA con base nos devanditos requisitos, adquire especial relevancia o adestramento do sistema, en tanto que permite que o sistema aprenda e mellore a súa capacidade para facer análises máis precisas.

O ciclo de vida do sistema de IA conclúe coa súa validación, despregadura e mantemento. Estas etapas son procesos continuos nos que se require unha colaboración constante para asegurar que o sistema de IA opera de forma fiable, segundo os obxectivos que se definiron nun inicio e mantendo a observancia dos requirimentos normativos asociados á adopción do sistema.

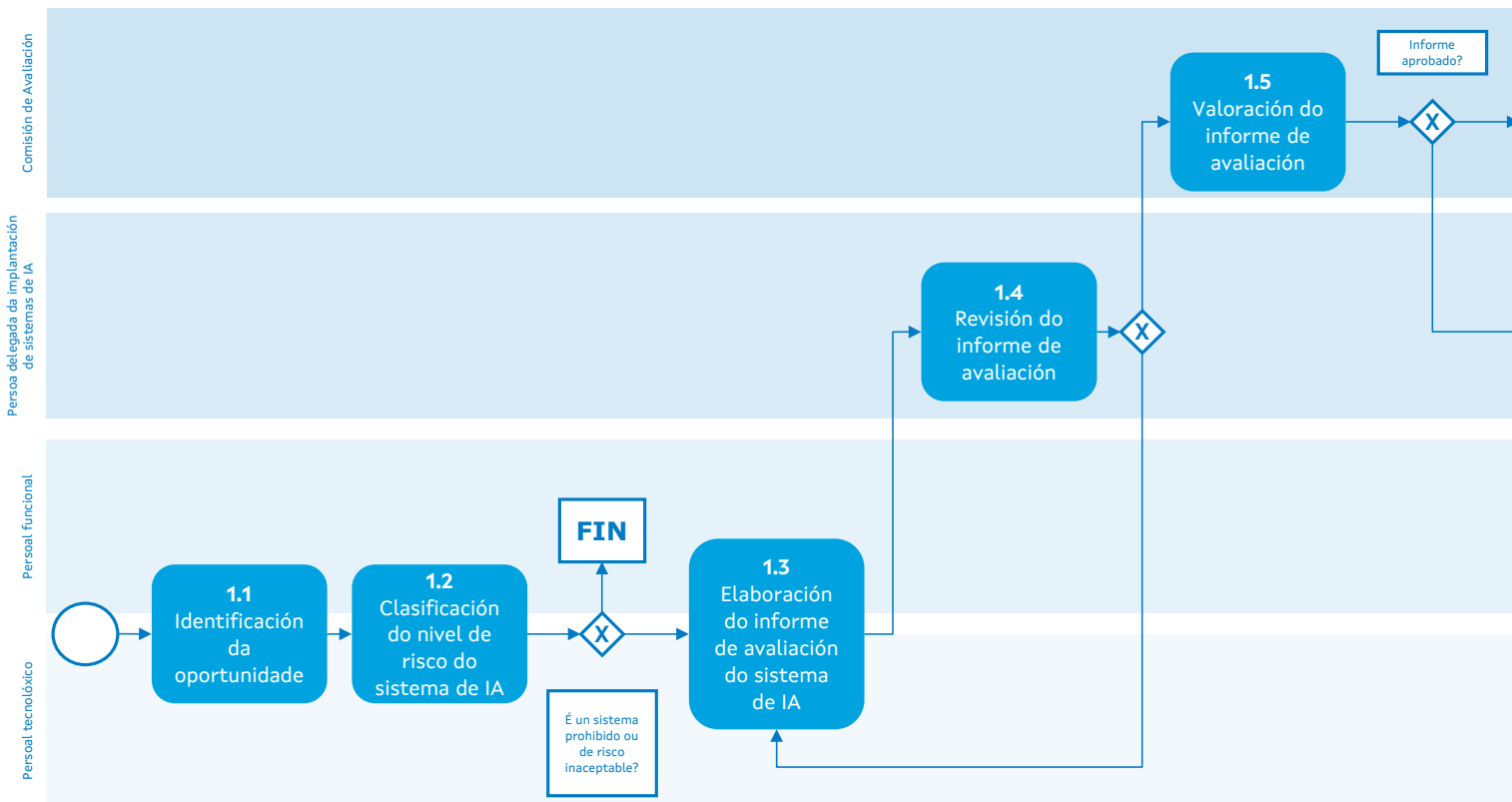
2.3. Diagrama de fluxo de xestión do ciclo de vida dun sistema de IA

A xestión dun sistema de IA abarca non só as actividades necesarias para a xestión do seu ciclo de vida, senón especialmente as actividades vinculadas coa xestión dos riscos inherentes ao deseño, desenvolvemento e implantación das solucións de IA. A utilización de

ferramentas estándar de soporte a estas actividades achega ao proceso global a transparencia, obxectividade e fiabilidade necesarias para a óptima xestión do sistema de intelixencia artificial.

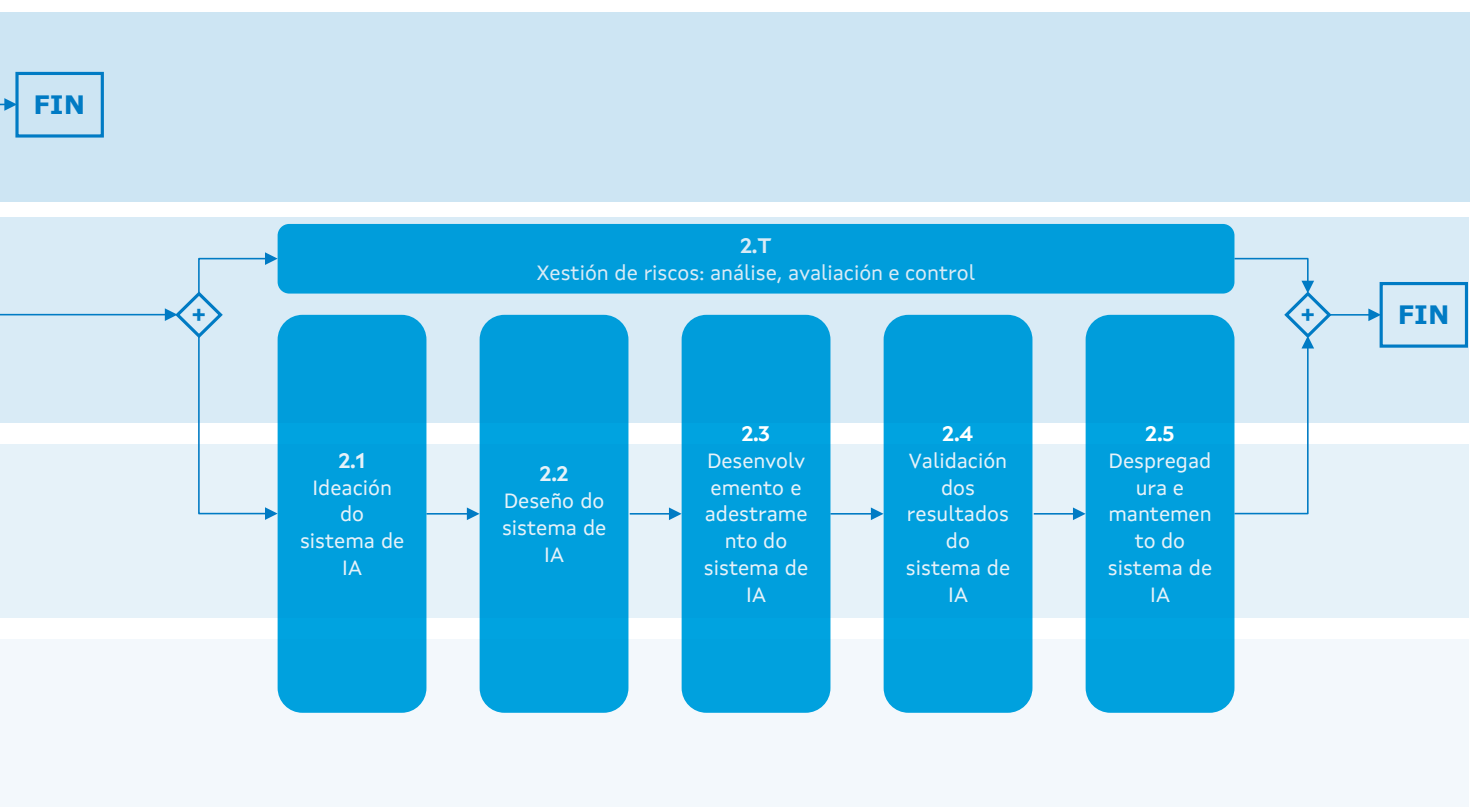


Fase 1. Identificación e avaliación de oportunidades vinculadas á adopción da IA





Fase 2. Diseño, desenvolvemento e implantación do sistema de IA



2.3.1. Axentes intervinientes na xestión dun sistema de intelixencia artificial

A relación dos axentes que interveñen na xestión dun sistema de IA pode variar dunha Administración pública a outra, tanto na súa nomenclatura como no seu ámbito de actuación específico. No entanto, as actividades que se desenvolven na xestión do ciclo de vida dun sistema de IA han de ser cubertas en calquera Administración pública e deben ser executadas por axentes cuxo rol e funcións deben asemellarse aos que aquí se expoñen.





REVIEW

ASSESS

[Red sticky note]



Persoa delegada da implantación de sistemas de IA

A persoa delegada da implantación de sistemas de IA é unha persoa da Administración pública cuxas funcións se vinculan a supervisar as oportunidades relacionadas coa IA que se identifiquen na organización, ofrecer asesoramento e avalialas para a súa elevación e debate na Comisión de Avaliación; e apoiar a xestión dos seus riscos ao longo do seu deseño, desenvolvemento e implantación.

Cal é o seu rol?

Supervisar e validar a análise das oportunidades de IA para a súa elevación á Comisión de Avaliación de IA.

Cales son as súas funcións?

- Garantir o cumprimento da normativa de IA
- Controlar o estado da arte da IA e apoiar a identificación de oportunidades de IA
- Supervisar a avaliación de oportunidades de IA
- Apoiar a xestión de riscos de proxectos e sistemas de IA



Comisión de Avaliación

A Comisión de Avaliación é un órgano formado por persoal funcional, persoal TI e persoal experto da Administración, encargado de avaliar as oportunidades vinculadas á IA que lle sexan elevadas, así como de supervisar a xestión de riscos ao longo do deseño, desenvolvemento e implantación da solución.

A Comisión de Avaliación reúne de forma periódica para revisar as oportunidades de IA que lle foron elevadas, así como para revisar a xestión de riscos das solucións de IA que se atopen en fase de deseño, desenvolvemento e implantación. O método de decisión deste órgano ha de fomentar a unanimidade dos seus integrantes e ha de garantir que as persoas responsables das oportunidades de IA que foron elevadas sexan escoitadas para a súa valoración.

Cal é o seu rol?

Avaliar as oportunidades vinculadas á IA que lle sexan elevadas, e autorizar e supervisar a implantación de solucións de IA.

Cales son as súas funcións?

- Avaliar as oportunidades de IA que lle sexan elevadas
- Revisar a xestión de riscos das solucións de IA



Persoal funcional e persoal TI

O persoal funcional e o persoal TI están involucrados no conxunto das actividades de carácter executivo que compoñen o proceso de xestión do ciclo de vida dun sistema de IA na Administración pública, desde a identificación e avaliación preliminar da oportunidade de IA ata a súa implantación final. A responsabilidade destas actividades é compartida entre o persoal funcional e o persoal TI.

Cal é o seu rol?

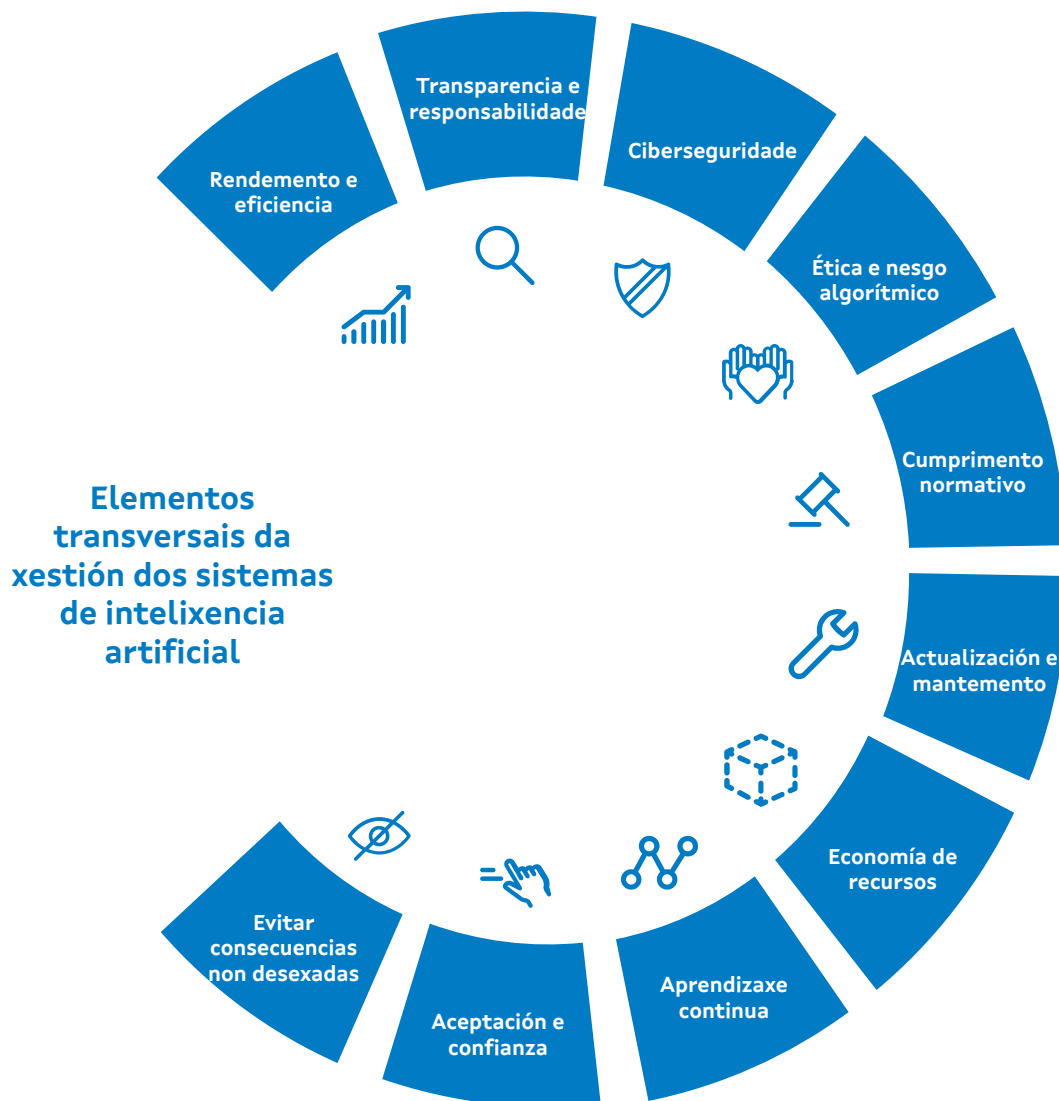
Liderar as actividades de carácter executivo que compoñen o proceso de xestión do ciclo de vida dun sistema de IA na Administración pública.

Cales son as súas funcións?

- Identificar oportunidades de IA
- Clasificar o nivel de risco do sistema de IA
- Avaliar o sistema de IA de forma global a fin de que o valore a Comisión de Avaliación
- Liderar o deseño, desenvolvemento e implantación do sistema de IA

2.3.2. Elementos transversais de xestión

A xestión do ciclo de vida dun sistema de IA é un proceso integral que permite asegurar que o sistema observa en todo momento as condicións que garanten a súa idoneidade, en termos de adecuación á necesidade que cobre e de rendemento; en termos de uso fiable e sostible no tempo; en termos de cumprimento normativo e dos requirimentos éticos asociados á tecnoloxía, e en termos de mellora continua.





Rendemento e eficiencia

Unha xestión adecuada garante que o sistema de IA funcione de maneira eficiente e cumpra cos seus obxectivos de rendemento. Isto implica optimizar os algoritmos, os recursos informáticos e os conxuntos de datos para lograr un funcionamento eficiente e maximizar o rendemento.



Transparencia e responsabilidade

A xestión adecuada dun sistema de IA inclúe a documentación e a rastrexabilidade das decisións que toma. Isto é esencial para garantir a transparencia e a rendición de contas no caso de problemas ou decisións incorrectas.



Ciberseguridade

A ciberseguridade é unha preocupación importante na xestión de sistemas de IA, especialmente cando se trata de aplicacións críticas ou sensibles. Unha xestión adecuada debe abordar a seguridade dos datos e protexer o sistema contra ameazas e ataques maliciosos.



Ética e nesgo algorítmico

A xestión da IA implica abordar cuestións éticas, como o nesgo nos algoritmos. É importante garantir que o sistema de IA non discrimine certos grupos ou individuos e que se axuste a principios éticos sólidos.



Cumprimento normativo

A regulación arredor da IA está en constante evolución. A xestión adecuada debe garantir o cumprimento das leis e regulacións aplicables, o que inclúe o cumprimento de regulacións de privacidade de datos e dereitos dos cidadáns.



Economía de recursos

A xestión eficiente dos recursos, como o poder de procesamento e o almacenamento de datos, é esencial para evitar o desperdicio e garantir un uso rendible da infraestrutura de intelixencia artificial.



Aprendizaxe continua

A IA beneficiase da aprendizaxe continua. A xestión adecuada inclúe a capacidade de adaptar e mellorar o sistema a medida que se recompilan e analizan máis datos e se obteñen máis coñecementos.



Aceptación e confianza

A xestión da IA tamén se relaciona coa percepción e a confianza dos usuarios e as partes interesadas. Un sistema de IA xestionado adecuadamente é máis propenso a gañar a aceptación e a confianza dos usuarios e, en última instancia, a ser máis efectivo.



Evitar consecuencias non desexadas

A IA pode ter impactos significativos na sociedade e na toma de decisións. A xestión adecuada axuda a evitar consecuencias non desexadas, como a discriminación, a falta de privacidade e a perda de emprego, e a mitigar posibles riscos.

2.3.3. A importancia da xestión de riscos

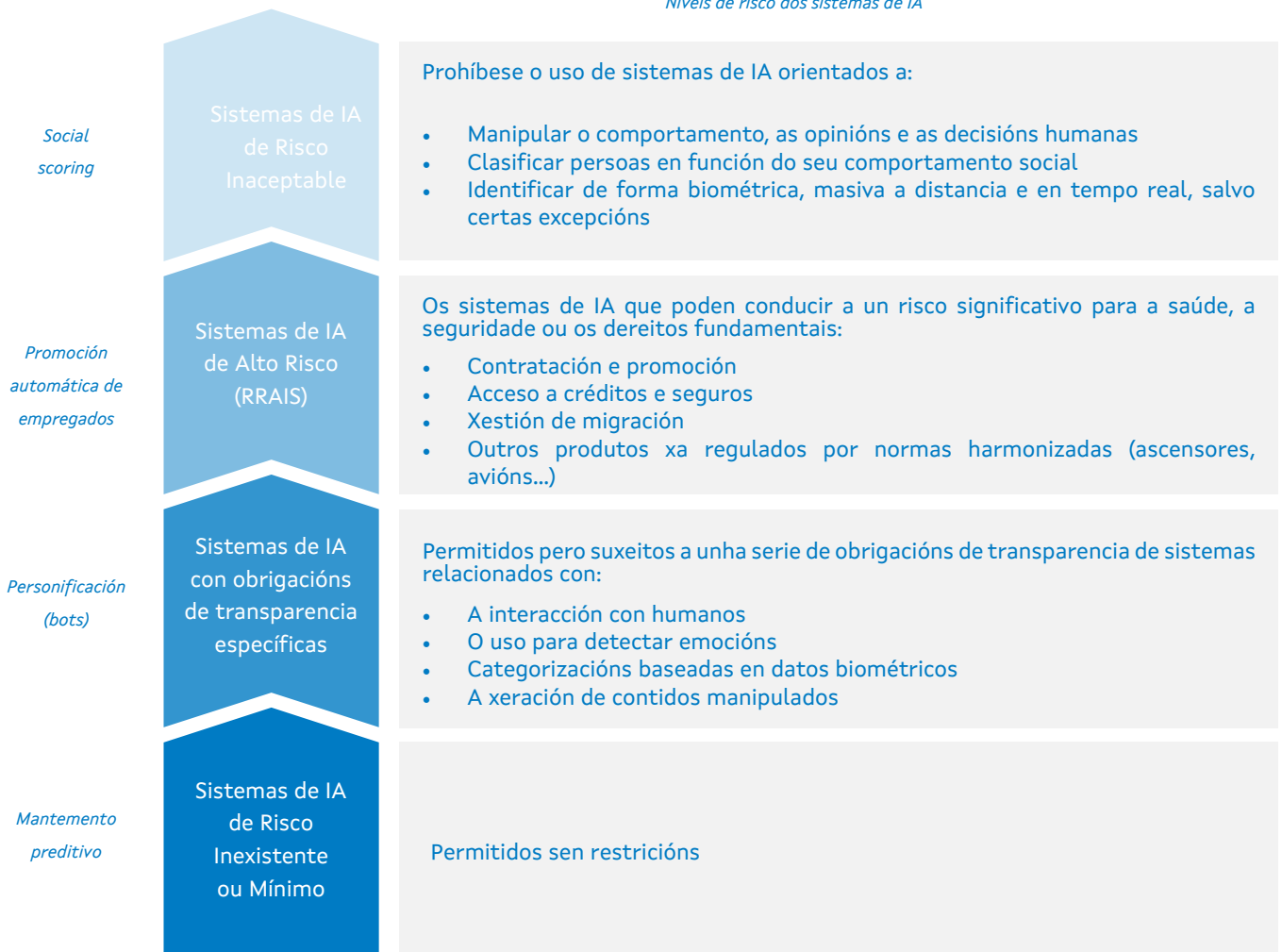
A xestión de riscos desempeña un papel fundamental no éxito de calquera proxecto, especialmente no ámbito tecnolóxico. Os proxectos tecnolóxicos a miúdo involucran avances rápidos e cambios nas condicións do contorno, o que pode xerar riscos imprevistos. A identificación temperá destes riscos permite desenvolver estratexias de mitigación e continxencia, e así mitigar o impacto negativo no desenvolvemento dos proxectos.

No caso específico de proxectos de IA, a xestión de riscos cobra unha relevancia aínda maior. A complexidade inherente á creación de sistemas de IA, como a dispoñibilidade de datos adaptados, a interpretación dos modelos de IA ou o uso ético desta tecnoloxía, pon de manifesto esta importancia. Por iso, implantar medidas de prevención ou mitigación de riscos é esencial para garantir

a equidade e a seguridade na despregadura de solucións de IA.

Consciente desta realidade, a normativa europea sitúa o risco como elemento central da xestión dos sistemas de IA. En concreto, a normativa permite trasladar os elementos transversais de xestión previamente sinalados a unha escala de riscos a través dos cales se materializa a xestión do ciclo de vida do sistema. Así, os sistemas de IA deben clasificarse en función do seu nivel de risco (inaceptable, alto, limitado e mínimo) e as persoas responsables da súa xestión deben, de forma consecuenta, aplicar determinadas obrigacións e restricións en función de cada nivel de risco.

Niveis de risco dos sistemas de IA





A normativa comunitaria orienta unha serie de reflexións para que a Administración pública que está a valorar un potencial sistema de IA coñeza o seu nivel de risco. Como se indicaba anteriormente, en función do resultado das devanditas reflexións, as persoas responsables deberán aplicar unhas ou outras medidas de xestión de riscos segundo o modelo de xestión de riscos propio da Administración pública.

Reflexións para a clasificación do nivel do risco dun sistema de IA:



O sistema de IA emprega técnicas subliminares que poden transcender a conciencia dunha persoa co obxectivo de alterar de maneira substancial o seu comportamento dun modo que provoque ou sexa razoablemente probable que provoque prexuízos físicos ou psicolóxicos a esa persoa ou a outra, ou que teña ese efecto?

O sistema de IA pode aproveitar algunha das vulnerabilidades dun grupo específico de persoas derivadas da súa idade ou discapacidade, ou dunha situación social ou económica específica, co obxectivo de alterar de maneira substancial o comportamento dunha persoa que pertenza ao devandito grupo dun modo que provoque ou sexa razoablemente probable que provoque prexuízos físicos ou psicolóxicos a esa persoa ou a outra, ou que teña ese efecto?

O sistema de IA foi deseñado co fin de avaliar ou clasificar as persoas físicas durante un período determinado de tempo atendendo ao seu comportamento social ou a características persoais ou da súa personalidade coñecidas ou preditas, de forma que a puntuación cidadá resultante provoque unha ou varias das situacións seguintes mencionadas a continuación?

i. un trato prexudicial ou desfavorable cara a determinadas persoas físicas ou grupos de persoas físicas en contextos sociais que non garden relación cos contextos onde se xeraron ou obtiveron os datos orixinalmente.

ii. un trato prexudicial ou desfavorable cara a determinadas persoas físicas ou grupos de persoas físicas que é inxustificable ou desproporcionado con respecto ao seu comportamento social ou a gravidade de este.

Segundo a finalidade do sistema

A finalidade do seu sistema de IA é a identificación biométrica remota (*)?

(Designa un sistema de IA destinado a identificar persoas físicas xeralmente a distancia, sen a súa participación activa, comparando os seus datos biométricos cos que figuran nun repositorio de datos de referencia.*

A finalidade do seu sistema de IA é ser utilizado como compoñentes de seguridade na xestión e funcionamento de infraestruturas dixitais críticas, o tráfico rodado e da subministración de auga, gas, calefacción e electricidade?

A finalidade do seu sistema de IA é determinar o acceso ou a admisión de persoas físicas a programas ou centros educativos e de formación profesional a todos os niveis ou para asignar a persoas físicas aos devanditos programas ou centros?

A finalidade do seu sistema de IA é avaliar os resultados da aprendizaxe, tamén cando eses resultados se utilicen para orientar o proceso de aprendizaxe das persoas físicas en programas ou centros educativos e de formación profesional a todos os niveis?

A finalidade do seu sistema de IA é a contratación ou a selección de persoas físicas, en particular para publicar anuncios de emprego específicos, analizar e filtrar as solicitudes de emprego e avaliar os candidatos?

A finalidade do seu sistema de IA é tomar decisións relativas á promoción e á rescisión de relacións contractuais de índole laboral, para a asignación de tarefas a partir de comportamentos individuais ou trazos ou características persoais e para realizar un seguimento e unha avaliación do rendemento e o comportamento das persoas no marco das devanditas relacións?

A finalidade do seu sistema de IA é ser utilizado polas autoridades públicas ou no seu nome para avaliar a admisibilidade das persoas físicas para acceder a servizos e axudas esenciais de asistencia pública, así como para conceder, reducir, retirar ou recuperar eses servizos e axudas?



A finalidade do seu sistema de IA é avaliar a solvencia de persoas físicas ou establecer a súa cualificación crediticia, salvo os sistemas de IA postos en servizo por provedores que son microempresas e pequenas empresas, tal como se define no anexo da Recomendación 2003/361/CE da Comisión, para o seu uso propio?

A finalidade do seu sistema de IA é o envío ou o establecemento de prioridades no envío de servizos de primeira intervención en situacións de emerxencia, por exemplo bombeiros e servizos de asistencia médica?

A finalidade do seu sistema de IA é a avaliación de riscos e a fixación de prezos en relación coas persoas físicas no caso dos seguros de vida e de saúde, con excepción dos sistemas de IA postos en servizo por provedores que sexan microempresas e pequenas empresas, tal como se definen no anexo da Recomendación 2003/361/CE da Comisión, para o seu propio uso?

A finalidade do seu sistema de IA é ser utilizado polas autoridades encargadas da aplicación da lei ou no seu nome para avaliar o risco de que unha persoa física cometa unha infracción ou reincida ou o risco de que unha persoa física se converta en posible vítima de infraccións penais?

A finalidade do seu sistema de IA é ser utilizado polas autoridades encargadas da aplicación da lei ou no seu nome como polígrafos e ferramentas similares, ou para detectar o estado emocional dunha persoa física?

A finalidade do seu sistema de IA é ser utilizado polas autoridades encargadas da aplicación da lei ou no seu nome para avaliar a fiabilidade das probas durante a investigación ou o axuizamento de infraccións penais?

A finalidade do seu sistema de IA é ser utilizado polas autoridades encargadas da aplicación da lei ou no seu nome para predicir a comisión ou reiteración dunha infracción penal real ou potencial a partir da elaboración de perfís de persoas físicas mencionada no artigo 3, punto 4, da Directiva (UE) 2016/680, ou para avaliar trazos e características da

personalidade ou comportamentos delituosos pasados de persoas físicas ou grupos?

A finalidade do seu sistema de IA é ser utilizado polas autoridades encargadas da aplicación da lei ou no seu nome para elaborar perfís de persoas físicas, como se menciona no artigo 3, punto 4, da Directiva (UE) 2016/680, durante a detección, a investigación ou o axuízamento de infraccións penais?

A finalidade do seu sistema de IA é ser utilizado polas autoridades públicas competentes ou no seu nome como polígrafos e ferramentas similares, ou para detectar o estado emocional dunha persoa física?

A finalidade do seu sistema de IA é ser utilizado polas autoridades públicas competentes ou no seu nome para avaliar un risco, como un risco para a seguridade, a saúde ou relativo á migración irregular, que expoña unha persoa física que teña a intención de entrar ou entrarse no territorio dun Estado membro?

A finalidade do seu sistema de IA é ser utilizado polas autoridades públicas competentes ou no seu nome para examinar as solicitudes de asilo, visado e permiso de residencia, e as reclamacións asociadas con respecto á admisibilidade das persoas físicas solicitantes?

A finalidade do seu sistema de IA é ser utilizado por unha autoridade xudicial ou no seu nome para interpretar feitos ou a lei, así como para aplicar a lei a un conxunto concreto de feitos?





Segundo o sector de aplicación do sistema

O seu sistema constitúese en si mesmo como un produto ou un compoñente de seguridade deste dentro dos sectores ou ámbitos listados a continuación?

- a. Máquinas
- b. Seguridade dos xoguetes
- c. Embarcacións de recreo e motos acuáticas
- d. Ascensores e compoñentes de seguridade para ascensores
- e. Aparellos e sistemas de protección para uso en atmosferas potencialmente explosivas
- f. Equipos radioeléctricos
- g. Equipos a presión
- h. Instalacións de transporte por cable
- i. Equipos de protección individual
- j. Aparellos que queiman combustibles gasosos
- k. Produtos sanitarios
- l. Produtos sanitarios para diagnóstico in vitro
- m. Seguridade da aviación civil
- n. Homologación de vehículos de dúas ou tres rodas e cuadríciclos
- o. Homologación de vehículos agrícolas ou forestais, e a vixilancia do mercado dos devanditos vehículos
- p. Equipos mariños
- q. Interoperabilidade do sistema ferroviario dentro da Unión Europea
- r. Homologación e vixilancia do mercado dos vehículos de motor e os seus remolques e dos sistemas, os compoñentes e as unidades técnicas independentes destinados aos devanditos vehículos
- s. Normas comúns no ámbito da aviación civil e polo que se crea unha Axencia da Unión Europea para a Seguridade Aérea na medida na que afecte o deseño, a produción e a comercialización de aeronaves contemplados no artigo 2 (IA Act), apartado 1, letras a) e b), cando se refira a aeronaves non tripuladas e os seus motores, hélices, compoñentes e equipos para controlalas a distancia.
- t. Requisitos de homologación de tipo dos vehículos de motor e dos seus remolques, así como os sistemas, compoñentes e unidades técnicas independentes destinados a eses vehículos, no que respecta á súa seguridade xeral e á protección dos ocupantes dos vehículos e dos usuarios vulnerables da vía pública.

SISTEMAS DE IA CON OBRIGACIÓNS DE TRANSPARENCIA ESPECÍFICAS



Se o sistema de IA está destinado a interactuar con persoas físicas, está deseñado e desenvolvido de forma que esas persoas estean informadas de que están interactuando cun sistema de IA, excepto nas situacións nas que isto resulte evidente desde o punto de vista dunha persoa xurídica que estea razoablemente informada, observadora e circunspecta, dadas as circunstancias e o contexto de utilización?

(Conteste non a esta pregunta se o sistema de IA será autorizado pola lei para fins de detección, prevención, investigación ou axuízamento de infraccións penais, con suxeición ás correspondentes salvagardas dos dereitos e liberdades de terceiros, salvo que este sistema estea á disposición do público para denunciar unha infracción penal).

Se o sistema de IA é ou utiliza un sistema de categorización biométrica, infórmase do funcionamento do sistema ás persoas físicas expostas a el?

(Conteste non a esta pregunta se o sistema de IA será utilizado para a categorización biométrica, autorizado por lei para detectar, impedir e investigar infraccións penais, con suxeición ás correspondentes salvagardas dos dereitos e liberdades de terceiros).

Se o sistema de IA é ou utiliza un sistema de IA que xere ou manipule contido de imaxe, son ou vídeo que se asemelle notablemente a persoas, obxectos, lugares ou outras entidades ou sucesos existentes, e que poida inducir erroneamente a unha persoa a pensar que son auténticos ou verídicos (ultrafalsificación), fai público que o contido foi xerado de forma artificial ou manipulado?

(Conteste non a esta pregunta se o uso do sistema de IA será autorizado por lei para detectar, impedir, investigar e axuizar infraccións penais ou cando o contido forme parte dunha obra ou programa manifestamente creativo, satírico, artístico ou ficticio, con suxeición ás correspondentes salvagardas dos dereitos e liberdades de terceiros).

Se o sistema de IA é ou utiliza un sistema de IA que xere ou manipule contido de imaxe, son ou vídeo que se asemelle notablemente a persoas, obxectos, lugares ou outras entidades ou sucesos existentes, e que poida inducir erroneamente a unha persoa a pensar que son auténticos ou verídicos (ultrafalsificación), fai público que o contido foi xerado de forma artificial ou manipulado?

(Conteste non a esta pregunta se o uso do sistema de IA será autorizado por lei para detectar, impedir, investigar e axuizar infraccións penais ou cando o contido forme parte dunha obra ou programa manifestamente creativo, satírico, artístico ou ficticio, con suxeición ás correspondentes salvagardas dos dereitos e liberdades de terceiros).



MODELO FUNDACIONAL OU SISTEMA DE IA XENERATIVA



O sistema de IA constitúe un sistema de IA xenerativa ou está baseado no desenvolvemento dun modelo fundacional?

Modelo fundacional: un modelo de IA que se adestra con gran cantidade de datos a escala, deseñado para fornecer resultados xerais e que pode adaptarse a unha ampla gama de tarefas distintas.

Sistema de IA xenerativa: modelos fundacionais utilizados en sistemas de IA destinados especificamente a xerar, con distintos niveis de autonomía, contidos como texto complexo, imaxes, audio ou vídeo.

Capítulo 3

Fases do ciclo de vida dun sistema de intelixencia artificial

Neste capítulo exploraremos en detalle as distintas fases que conforman o ciclo de vida dun sistema de intelixencia artificial (IA), desde a súa concepción ata a súa implantación total. Proporcionaremos unha visión integral do proceso, destacando as actividades esenciais que compoñen o fluxo de xestión deste ciclo. En cada etapa abordaranse as cuestións clave, identificaranse os roles responsables e analizaranse erros comúns que poden xurdir durante a execución.



A continuación preséntase a visión global e desenvólvense as actividades que conforman o diagrama de fluxo de xestión do ciclo de vida dun sistema de IA, reflectindo en cada caso as cuestións clave, os roles responsables e erros comúns na súa execución.



Fase 1. Identificación e avaliación de oportunidades vinculadas á adopción da IA

1.1 Identificación da área de mellora e análise da idoneidade de aplicar IA

1.2 Clasificación do nivel de risco do sistema de IA

1.3 Avaliación estándar do sistema de IA

1.4 Revisión preliminar da avaliación

1.5 Valoración final da avaliación



Fase 2. Deseño, desenvolvemento e implantación do sistema de IA

2.1 Ideación do sistema de IA

2.2 Deseño do sistema de IA

2.3 Desenvolvemento e adestramento do sistema de IA

2.4 Validación dos resultados do sistema de IA

2.5 Despregadura e mantemento do sistema de IA

3.1

Fase 1. Identificación e avaliación de oportunidades vinculadas á adopción da IA

As actividades de identificación e avaliación de oportunidades de IA permiten que todas as partes interesadas na implantación de sistemas de IA poidan coñecer e avaliar de forma obxectiva e compartida posibles sistemas de IA para implantar na Administración pública. En todo caso, estas actividades poden modularse ás capacidades ou modelo de funcionamento de cada Administración pública que as implanta.

3.1.1 Identificación da área de mellora e análise da idoneidade de aplicar IA



O proceso global de incorporación da IA na Administración pública empeza por definir de maneira precisa a barreira, problema ou área de mellora que se busca abordar a través da implantación da IA. É esencial acoutar a área de mellora, identificar as súas implicacións tanto para usuarios externos como internos, e xustificar a implantación da IA como unha solución valiosa, desde a perspectiva dun acto administrativo (toma de decisións autónoma) ou unha actuación administrativa (apoio á toma de decisións ou mellora de servizos).

Para iso, é necesario determinar se a Administración pública ten a competencia legal e normativa para abordar o problema, se se require asociación con outros axentes, e se conta cos recursos necesarios. Ademais, tras definir o problema, é crucial avaliar o estado actual de resposta da Administración, identificar limitacións e oportunidades (viabilidade previa) e entender como podería mellorar a situación a implantación da IA.

Cuestións clave

A oportunidade pódela identificar tanto o persoal funcional como o persoal TI da Administración pública.

A persoa que identifique a oportunidade deberá comunicala á persoa delegada da implantación de sistemas de IA, dando inicio ao proceso de xestión dun sistema de IA.

Algunhas reflexións

Cal é o problema que se busca solucionar ou a necesidade que se busca cubrir?

Cales son os axentes (internos ou externos) que se ven afectados polo problema ou a necesidade? É posible cuantificar os devanditos axentes?

Por que é prioritario resolver este problema ou cubrir esta necesidade?

Existe algún caso parecido de uso da IA que fose implantado anteriormente? Cal?

Cales son os obxectivos da potencial implantación da IA (ou os beneficios que se espera obter) e como impactan nos axentes afectados polo problema ou a necesidade?

Erros comúns

- **Falta de comprensión:** non comprender completamente as capacidades e limitacións da IA pode levar a identificar oportunidades pouco realistas ou inapropiadas.
- **Falta de enfoque nos problemas correctos:** identificar áreas que non se beneficiarían significativamente da IA ou que teñen necesidades máis urxentes pode levar á asignación ineficiente de recursos.
- **Falta de colaboración:** non involucrar a axentes clave ou non buscar colaboracións externas pode limitar a eficacia da implantación.
- **Sobreestimar os beneficios:** ter expectativas pouco realistas sobre os beneficios da IA pode levar a decepcións e desilusión.
- **Non involucrar os colectivos impactados:** non consultar os colectivos e partes interesadas pode levar a solucións que non se axusten ás súas necesidades ou valores.

3.1.2 Clasificación do nivel de risco do sistema de IA



A clasificación do nivel de risco dun sistema de IA é unha actividade esencial no proceso de xestión dun sistema de IA, en tanto que permite determinar se a oportunidade pode dar lugar a un sistema aceptable nos termos de risco que establece a normativa europea. A propia normativa concreta unha serie de requisitos para que a Administración pública que está a valorar un potencial sistema de IA coñeza o seu nivel de risco.

Cuestións clave

Como primeiro elemento de filtrado das oportunidades de IA, encherase un cuestionario para clasificar o potencial sistema de IA segundo os niveis de risco previstos na AI Act.

No caso de que se trate dun sistema prohibido ou un nivel de risco inaceptable, o proceso terminará neste punto.

En función do nivel de risco resultante (salvo no caso previamente indicado), incorporaranse preguntas adicionais no «Informe de avaliación».



A clasificación do nivel de risco dun sistema de IA debe realizarse a través dun «**Cuestionario de clasificación do nivel de risco do sistema de IA**» que recolle todas as preguntas necesarias para que a Administración pública coñeza en que nivel de risco se clasifica o sistema de IA que está a valorar. En función do risco, poderase optar por continuar coa avaliación detallada do sistema de IA.

Erros comúns

- **Subestimar riscos éticos e de privacidade:** centrarse excesivamente nos riscos técnicos e operativos sen considerar os aspectos éticos e de privacidade pode resultar en falta de cumprimento da normativa.
- **Falta de avaliación do impacto social:** non analizar como pode afectar o sistema de IA a sociedade no seu conxunto, incluíndo o emprego, a economía e a equidade, pode levar a consecuencias non desexadas.
- **Falta de transparencia:** non proporcionar información adecuada ás partes interesadas sobre como funciona o sistema de IA e como se toman as decisións pode xerar desconfianza.

3.1.3 Avaliación estándar do sistema de IA



Despois de establecer a área de mellora e confirmar que a IA é a ferramenta adecuada para apoiar a solución, e antes de determinar o interese de acometer un proxecto para a súa implantación, é crucial avaliar unha serie de aspectos fundamentais para garantir a viabilidade do proxecto. Esta avaliación debe conter elementos estándares que permitan que todas as partes interesadas poidan analizar o sistema de forma obxectiva, o que potencia a transparencia e a cohesión das decisións que se adopten respecto da implantación da IA.

Cuestións clave

Unha vez clasificado o nivel de risco do sistema de IA, elaborárase un informe de avaliación do sistema.

A colaboración do persoal funcional e o persoal TI é indispensable para o éxito desta avaliación, así como a comunicación continua coa persoa delegada da implantación de sistemas de IA.



A avaliación dun sistema de IA estrúcturase a través dun **«Informe de avaliación»** que valora a dispoñibilidade de datos adaptados a un modelo de IA, a dispoñibilidade de capacidades e competencias internas para a implantación do sistema de IA, os requisitos de integración do sistema de IA cos sistemas informáticos ou a necesidade de recursos económicos e humanos para a implantación do sistema

Erros comúns

- **Non avaliar o rumbo e a equidade:** non considerar a posibilidade de nesgo nos datos de adestramento e nos algoritmos pode levar a decisións discriminatorias e prexudiciais.
- **Non considerar o tamaño do conxunto de datos:** contar cun conxunto de datos pequeno ou non representativo pode dar como resultado modelos de IA deficientes e pouco fiables.
- **Ignorar a ciberseguridade:** non avaliar adecuadamente as vulnerabilidades de seguridade no sistema de IA pode dar lugar a riscos de ataques cibernéticos e brechas de datos.
- **Non considerar a robustez e a resiliencia:** non avaliar como pode comportarse o sistema de IA en situacións inesperadas ou baixo presión pode ser un erro significativo.
- **Ignorar a infraestrutura necesaria:** non avaliar adecuadamente se a infraestrutura tecnolóxica existente é compatible co sistema de IA pode levar a problemas de integración e atrasos.
- **Subestimar a carga computacional:** non estimar adecuadamente os recursos necesarios para adestrar e executar o sistema pode dar como resultado problemas de rendemento e custos imprevistos.



Estrutura e contido do informe de avaliación dun sistema de IA na Administración pública

A partir destas premisas, a continuación preséntanse a estrutura e o contido do informe de avaliación dun sistema de IA na Administración pública:



a) Descrición da oportunidade de IA


Descrición do sistema de IA obxecto de avaliación, indicando a necesidade á que dá resposta e os principais beneficios esperados mediante a súa implantación (ver apartado «3.1.1. Identificación da área de mellora e análise da idoneidade de aplicar IA»).

b) Tarefas que realizar e ramas da IA empregadas polo sistema

Descrición de en que medida é idóneo que a necesidade sexa cuberta mediante unha das seguintes tecnoloxías de IA (ou por unha tecnoloxía de IA alternativa), e en que medida esa necesidade non pode ser cuberta por unha tecnoloxía que non aplique IA.

Tarefas do sistema	Descrición	Rama da IA que empregar	Xustificación
Recoñecemento	Categorización de imaxes, textos e vídeos por medio da identificación das súas características clave	<i>*texto libre*</i>	<i>*texto libre*</i>
Detección de eventos	Detección de patróns e anomalías nun conxunto de datos	<i>*texto libre*</i>	<i>*texto libre*</i>
Predición	Predición dun estado futuro baseado en datos históricos	<i>*texto libre*</i>	<i>*texto libre*</i>
Personalización	Desenvolvemento de perfís de usuario que, con base en datos xerados polas súas accións, aumentan o nivel de personalización co tempo	<i>*texto libre*</i>	<i>*texto libre*</i>



Tarefas do sistema	Descrición	Rama da IA que empregar	Xustificación
Apoio á interacción	Desenvolvemento de ferramentas que potencian a interacción entre persoas e máquinas	<i>*texto libre*</i>	<i>*texto libre*</i>
Optimización enfocada en obxectivos	Optimización dos procesos dentro dunha organización a través da simulación de escenarios	<i>*texto libre*</i>	<i>*texto libre*</i>
Razoamento con estruturas de coñecemento	Establecemento de relacións causais entre os datos dispoñibles e un evento futuro non existente, focalizándose na causalidade entre variables	<i>*texto libre*</i>	<i>*texto libre*</i>
	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>

c) Nivel de risco do sistema de IA

Identificación do nivel de risco do sistema de IA e concreción das medidas para a xestión dos riscos do sistema (ver apartado «2.3.3. A importancia da xestión de riscos» e apartado «3.1.2. Clasificación do nivel de risco do sistema de IA»).



d) Disponibilidade de datos adaptados a un modelo de IA

Descrición do sistema de IA obxecto de avaliación, indicando a necesidade á que dá resposta e os principais beneficios esperados mediante a súa implantación (ver apartado «3.1.1. Identificación da área de mellora e análise da idoneidade de aplicar IA»).

Preguntas	Si	Non	Comentarios
Ten suficientes datos para que o modelo aprenda?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Entende os efectos posteriores do uso de datos desta maneira?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Os datos son precisos e completos, e con que frecuencia se actualizan?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Os datos foron recompilados utilizando fontes de medición adecuadas, fiables e imparciais?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Os datos están seguros e ten permiso para usalos?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Onde se atopan os datos?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Ten acceso aos datos e con que rapidez pode acceder a eles?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
En que formato están os datos? Requiren unha preparación significativa para estar listos para o modelado?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Os datos están estruturados? Por exemplo, pódense almacenar nunha táboa ou non están estruturados, como correos electrónicos ou páxinas web?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Existen restricións nos datos? Por exemplo, conteñen información confidencial, como domicilios particulares?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>



Preguntas	Si	Non	Comentarios
Podes vincular variables clave dentro e entre conxuntos de datos?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
O conxunto de datos contén patróns claros para que o modelo explore ao facer unha predición?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
O conxunto de datos contén patróns claros pero accidentais, o que resulta en nesgos de aprendizaxe do modelo?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>
Ten suficientes datos para que o modelo aprenda?	<input type="checkbox"/>	<input type="checkbox"/>	<i>*texto libre*</i>

e) Capacidades e competencias internas

Descrición de en que medida se dispón actualmente no equipo interno dos perfís necesarios para a implantación da IA, concretando o seu rol na devandita implantación; e, en caso de non dispoñer deles, identificación das accións de mitigación que se expoñen para dar cobertura ás habilidades que se asocian ao devandito rol.

Perfís	Carácter	Dispoñibilidade	Rol	Xustificación	Mitigación
Arquitecto de datos Establecer a visión para o uso de datos da organización, a través do deseño de datos, para satisfacer as necesidades comerciais	Recomendable	Si/Non	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>
Científico de datos Identificar problemas comerciais complexos mentres se aproveita o valor dos datos	Recomendable	Si/Non	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>
Enxeñeiro de datos Desenvolver a entrega de produtos e servizos de datos en sistemas e procesos comerciais	Obrigatorio	Si/Non	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>



Perfís	Carácter	Disponibilidade	Rol	Xustificación	Mitigación
Responsable ético Proporcionar xuízos éticos e avaliacións sobre as entradas (datos) do modelo de IA	Obrigatorio	Si/Non	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>
Técnico funcional Identificar problemas comerciais complexos mentres se aproveita o valor dos datos	Obrigatorio	Si/Non	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>

f) Integración cos sistemas da Administración pública

Identificación das principais necesidades de integración da solución de IA cos sistemas actuais da Administración pública, segundo as diferentes capas arredor das que resulta necesario realizar esa integración. Así mesmo, identificación das accións necesarias para levar a cabo esa integración.

Capa	Necesidades de integración	Accións de integración
Capa de adquisición/interconexión Integra a información desde as fontes de datos e fornece información á capa de coñecemento	<i>*texto libre*</i>	<i>*texto libre*</i>
Capa de coñecemento Modela semanticamente os datos da cidade e procesa analiticamente os datos das capas de adquisición e interoperabilidade	<i>*texto libre*</i>	<i>*texto libre*</i>
Capa de interoperabilidade Facilita a prestación de servizos. Para a interoperabilidade semántica, a información debe ser compatible cun modelo de datos estándar	<i>*texto libre*</i>	<i>*texto libre*</i>
Capa de servizos Constituída por servizos conectados a través da capa de interoperabilidade	<i>*texto libre*</i>	<i>*texto libre*</i>



Capa	Necesidades de integración	Accións de integración
Capa de soporte Capa transversal encargada de ofrecer servicios comúns como auditoría, control ou seguridade	<i>*texto libre*</i>	<i>*texto libre*</i>

g) Interese de adquirir IAaaS, desenvolver ou comprar solucións

Descrición de en que medida é idónea a selección dunha alternativa (adquirir IAaaS, desenvolver ou comprar solucións) para a implantación da solución de IA.

Alternativa	Descrición	Xustificación
IAaaS (IA as a Service) É unha alternativa de interese para aquelas situacións nas que se desexe implantar de forma áxil e flexible servizos baseados en IA, sen a necesidade de integrar solucións nos sistemas da Administración pública	<ul style="list-style-type: none">• Necesidade exclusiva ou xenérica: achega flexibilidade en ambos os escenarios.• Madurez dos produtos dispoñibles: máis adecuada para a implantación de produtos maduros no mercado, aínda que achega flexibilidade en ambos os escenarios.• Necesidade de integración cos sistemas de Administración pública: máis adecuada para solucións que non requiran unha integración cos sistemas de Administración pública.	<ul style="list-style-type: none">• Necesidade exclusiva ou xenérica: <i>*texto libre*</i>• Madurez dos produtos dispoñibles: <i>*texto libre*</i>• Necesidade de integración cos sistemas de Administración pública: <i>*texto libre*</i>



Alternativa	Descrición	Xustificación
<p>Compra dunha solución</p> <p>Esta opción é máis adecuada se se está a buscar unha aplicación común de IA (por exemplo, recoñecemento óptico de caracteres), aínda que presenta maiores barreiras en termos de integración cos sistemas da Administración pública</p>	<ul style="list-style-type: none">• Necesidade exclusiva ou xenérica: resulta adecuada para responder a necesidades xenéricas.• Madurez dos produtos dispoñibles: máis adecuada para a implantación de produtos moi maduros no mercado.• Necesidade de integración cos sistemas de Administración pública: presenta maiores barreiras en termos de integración cos sistemas da Administración pública.	<ul style="list-style-type: none">• Necesidade exclusiva ou xenérica: <i>*texto libre*</i>• Madurez dos produtos dispoñibles: <i>*texto libre*</i>• Necesidade de integración cos sistemas de Administración pública: <i>*texto libre*</i>
<p>Desenvolvemento dunha solución</p> <p>Pódese optar por crear ou adaptar modelos de IA listos para usar ou algoritmos de código aberto, a condición de que o equipo interno conte coas habilidades para construír un proxecto de IA internamente e o equipo de operacións poida executar e manter unha solución de IA interna</p>	<ul style="list-style-type: none">• Necesidade exclusiva ou xenérica: resulta adecuada para responder a necesidades específicas (non xenéricas no mercado).• Madurez dos produtos dispoñibles: non require a existencia de produtos maduros no mercado.• Necesidade de integración cos sistemas de Administración pública: a integración cos sistemas da Administración pública non supón unha alta barreira, no sentido de que o desenvolvemento se realizará desde un primeiro momento considerando esa integración.	<ul style="list-style-type: none">• Necesidade exclusiva ou xenérica: <i>*texto libre*</i>• Madurez dos produtos dispoñibles: <i>*texto libre*</i>• Necesidade de integración cos sistemas de Administración pública: <i>*texto libre*</i>



h) Estimación de custo total de propiedade (TCO) vinculado ao sistema

Indicación da estimación de recursos económicos necesarios para a implantación e o mantemento do sistema de intelixencia artificial.

Fase	Tipoloxía	Tipoloxía	Xustificación
Implantación da solución	Custos directos de adquisición	<i>*texto libre*</i>	<i>*texto libre*</i>
	Custos indirectos de adquisición	<i>*texto libre*</i>	<i>*texto libre*</i>
	Outros eventuais custos	<i>*texto libre*</i>	<i>*texto libre*</i>
	Gastos xerais de estrutura	<i>*texto libre*</i>	<i>*texto libre*</i>
Mantemento da solución	Custos de propiedade	<i>*texto libre*</i>	<i>*texto libre*</i>
	Custos de mantemento	<i>*texto libre*</i>	<i>*texto libre*</i>
	Custos de uso	<i>*texto libre*</i>	<i>*texto libre*</i>
	Custos de eliminación	<i>*texto libre*</i>	<i>*texto libre*</i>

i) Estimación de recursos humanos internos vinculados ao sistema

Indicación da estimación de recursos humanos internos (FTE) necesarios para a implantación e mantemento do sistema, así como a área do que deberán proceder (da consellería correspondente e/ou a Administración pública).

Fase	Tipoloxía	Tipoloxía	Xustificación
Implantación da solución	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>
Mantemento da solución	<i>*texto libre*</i>	<i>*texto libre*</i>	<i>*texto libre*</i>

3.1.4 Revisión preliminar da avaliación



É esencial que a avaliación preliminar dun sistema de IA conte coa supervisión dunha figura experta en IA, de forma que se garanta que, con relación á decisión final sobre a implantación dun sistema de IA, se valoren unicamente aqueles sistemas con garantías de viabilidade.

Cuestións clave

A persoa delegada da implantación de sistemas de IA deberá revisar o informe de avaliación para que este poida ser elevado á Comisión de Avaliación.

No caso de identificar aspectos que modificar, a persoa delegada da implantación de sistemas de IA informará o persoal funcional e o persoal TI, que procederá á súa incorporación no informe ata a validación definitiva do mesmo.

3.1.5 Valoración final da avaliación



A Comisión de Avaliación ha de tomar a decisión final respecto da adopción dun sistema de IA con base en criterios coñecidos vinculados aos elementos da avaliación preliminar. Esta garantía de transparencia e obxectividade é clave para asegurar a confianza nas decisións que tome a Administración pública en relación coa IA e para fomentar o consenso entre todas as partes interesadas.

Cuestións clave

A Comisión de Avaliación reunirase periodicamente para avaliar o conxunto das oportunidades que lle sexan elevadas.

Resulta de interese que a Comisión decida por unanimidade ou mediante maiorías reforzadas a fin de impulsar a confianza e o consenso arredor das decisións que se tomen.

No caso dunha decisión desfavorable, considérase acreditado que o proxecto de potencial implantación do sistema de IA debe ser descartado de forma definitiva.

As decisións que adopte a Comisión de Avaliación deben reflectir a priorización e mesmo a calendarización da implantación dos sistemas de IA que reciben a aprobación, de forma coherente cos recursos dos que dispón a Administración pública para realizar esas implantacións.

3.2

Fase 2. Deseño, desenvolvemento e implantación do sistema de IA adopción da IA

O proceso de deseño, desenvolvemento e implantación dunha solución de IA céntrase en abordar as oportunidades identificadas, utilizando datos e algoritmos para construír sistemas intelixentes que melloren a toma de decisións e a eficiencia en diversas aplicacións. A colaboración de equipos multidisciplinares e unha mentalidade de mellora continua son clave para o éxito na adopción destes sistemas.

3.2.1 Ideación do sistema de IA



A ideación dun sistema de Intelixencia Artificial (IA) refírese ao proceso de xeración de ideas e conceptos para deseñar un sistema baseado en IA que aborde un problema ou unha necesidade específica. É unha etapa fundamental no desenvolvemento de solucións de IA e adoita preceder á fase de deseño e desenvolvemento do sistema de IA. A ideación é un proceso creativo e flexible no que se fomenta a exploración de múltiples enfoques e solucións. Pode requirir repeticións e axustes a medida que se obtén máis información e retroalimentación.

Cuestións clave

A ideación implica a definición clara e obxectiva da necesidade ou problema que se desexa abordar co sistema de IA.

A ideación tamén implica a concreción dos obxectivos específicos que se desexan alcanzar e como a IA axudará a resolver esas necesidades ou problemas.

Unha vez ideado o sistema de IA, esta actividade pode incluír o prototipado ou experimentación inicial do sistema, é dicir, crear un prototipo ou unha versión de proba do sistema de IA para avaliar o seu funcionamento en condicións controladas.

Erros comúns

- **Nesgo na xeración de ideas:** a falta de diversidade no equipo de ideación ou a influencia de nesgos persoais poden limitar a variedade de solucións propostas.
- **Enfoque excesivo na tecnoloxía:** centrarse demasiado na tecnoloxía en lugar de considerar as necesidades do usuario ou as aplicacións prácticas pode levar a solucións de IA que non se axustan ás situacións do mundo real.
- **Falta de consideración ética e de privacidade:** non ter en conta as implicacións éticas e de privacidade desde o principio pode resultar en problemas legais e de confianza en etapas posteriores da implantación.
- **Non explorar suficientes enfoques de IA:** limitarse a un único enfoque de IA ou non considerar unha variedade de técnicas pode resultar en solucións subóptimas.
- **Falta de aliñación cos obxectivos da organización:** non asegurarse de que as ideas de IA estean aliñadas cos obxectivos estratéxicos da empresa ou a organización pode levar a solucións que non agreguen valor.
- **Ignorar a retroalimentación das partes interesadas:** non recompilar retroalimentación temperá de persoas relevantes pode levar á falta de axustes e melloras nas ideas iniciais.

3.2.2 Deseño do sistema de IA



O deseño dun sistema de Intelixencia Artificial (IA) implica a definición da arquitectura e a funcionalidade do sistema. Comprende a selección de algoritmos de IA, a estrutura de datos, a interface de usuario e a infraestrutura tecnolóxica. Abórdanse cuestións como a escalabilidade, a robustez e a eficiencia. Ademais, considéranse aspectos éticos e de privacidade. O deseño baséase nos resultados da etapa de ideación, onde se xeran ideas para resolver un problema ou aproveitar unha oportunidade. É un paso crucial que senta as bases para o desenvolvemento e a implantación efectivos do sistema de IA.

Cuestións clave

Esta actividade consiste na definición das características e especificacións técnicas dos sistemas de IA; as principais son:

- **Tipoloxía de datos:** refírese á natureza e o tipo de datos que utilizará o sistema de IA. Isto inclúe datos estruturados, non estruturados e semiestruturados.
- **Algoritmos e modelos:** os algoritmos e modelos determinan como se procesarán os datos para facer predicións, tomar decisións ou realizar tarefas específicas. A elección de algoritmos e modelos depende do tipo de problema que se estea abordando.
- **Deseño da arquitectura:** o deseño da arquitectura inclúe a organización de compoñentes, fluxos de datos, procesamento en lotes ou en tempo real, e a selección de *hardware* e *software* para o funcionamento do sistema.
- **Interface de usuario:** a interface de usuario é a parte que permite aos usuarios interactuar co sistema de IA. Isto pode incluír interfaces gráficas de usuario (GUI), interfaces de liña de comandos (CLI) ou integración en aplicacións e servizos existentes.
- **Ética e responsabilidade:** isto implica garantir que o sistema non perpetúe nesgos, que cumpra coas regulacións e estándares éticos e que se respecte a privacidade dos datos e a transparencia nas decisións da IA. A avaliación e mellora continua

son procesos esenciais no deseño dun sistema de IA. Implica a realización de probas, a recompilación de retroalimentación de usuarios e a supervisión do rendemento do sistema.

Erros comúns

- **Nesgo nos datos:** utilizar conxuntos de datos nesgados ou non representativos pode dar lugar a sistemas de IA que perpetúan prexuízos e discriminación.
- **Selección inapropiada de algoritmos:** elixir algoritmos de IA inadecuados para o problema en cuestión pode resultar nun baixo rendemento ou ineficiencia.
- **Non realizar avaliacións éticas e probas de nesgo:** non avaliar e mitigar nesgos no sistema pode resultar en decisións discriminatorias e problemas de equidade.
- **Deseño de arquitectura pouco escalable:** non planificar a arquitectura para adaptarse a volumes crecentes de datos ou demanda pode limitar a capacidade do sistema.
- **Interface de usuario deficiente:** non deseñar unha interface de usuario efectiva pode dificultar a adopción e o uso do sistema.
- **Falta de transparencia e explicabilidade:** non proporcionar información sobre como toma decisións o sistema pode dar lugar a desconfianza e falta de comprensión.

3.2.3 Desenvolvemento e adestramento do sistema de IA



O desenvolvemento e adestramento dun sistema de IA é unha etapa crítica na implantación de solucións baseadas en IA, que comeza coa implantación do modelo de IA deseñado na fase de deseño. Este modelo serve como o «cerebro» do sistema, e a súa tarefa é realizar predicións, tomar decisións ou executar tarefas específicas.

Para adestrar o modelo, requírense datos adecuados. Estes datos poden ser etiquetados, o que significa que se coñecen as respostas correctas, ou non etiquetados; nese caso, o modelo aprende patróns sen a guía explícita de respostas. Durante o adestramento, o modelo exponse repetidamente a exemplos de datos, axustando os seus parámetros internos para optimizar a súa capacidade para realizar a tarefa específica. Este proceso é esencialmente un proceso de aprendizaxe, onde o modelo mellora a súa capacidade para facer predicións precisas.

Cuestións clave

O desenvolvemento e adestramento do sistema de IA consiste en alimentar o modelo do sistema de IA cun conxunto de datos de adestramento que permitan ao sistema aprender de forma iterativa, de maneira que se axusten os parámetros do modelo para minimizar os erros e maximizar o rendemento do sistema tras cada repetición.

É necesario asegurar que o conxunto de datos de adestramento sexa representativo e abarque os diferentes escenarios que se desexa que o sistema poida recoñecer, ademais de dividir os datos en conxuntos de adestramento, validación e proba para avaliar o rendemento do sistema de IA.

O desenvolvemento e adestramento dun sistema de IA é unha combinación de ciencia e tecnoloxía que require experiencia técnica e unha comprensión profunda dos algoritmos, os datos e o dominio específico do problema que se vai abordar. A precisión e a eficacia do modelo dependen en gran medida da calidade e cantidade dos datos de adestramento, así como da capacidade para axustar e validar adecuadamente o modelo.

Erros comúns

- **Selección inadecuada de datos:** utilizar conxuntos de datos de baixa calidade ou non representativos pode levar a un modelo de IA pouco preciso e nesgado.
- **Sobreaxuste do modelo:** axustar o modelo excesivamente aos datos de adestramento pode dar como resultado un modelo que funciona ben neses datos, pero non xeneraliza ben a novas situacións.
- **Subestimación da cantidade de datos:** non contar con suficientes datos de adestramento pode limitar a capacidade do modelo para aprender patróns con precisión.
- **Non considerar nesgos:** non abordar os nesgos nos datos de adestramento pode resultar en modelos de IA que perpetúan prexuízos e discriminación.
- **Selección inadecuada de algoritmos:** elixir o algoritmo de IA incorrecto para o problema en cuestión pode dar lugar a un rendemento subóptimo.
- **Falta de optimización de hiperparámetros:** non axustar adecuadamente os hiperparámetros do modelo pode impedir que alcance o seu máximo potencial.
- **Non documentar o proceso:** a falta de documentación adecuada do proceso de desenvolvemento e adestramento pode dificultar a comprensión e a colaboración entre equipos.

3.2.4 Validación dos resultados do sistema de IA



Esta etapa ten como obxectivo determinar se o modelo de IA adestrado funciona de maneira precisa e xeneraliza ben a novas situacións. Unha vez que o modelo se adestra, sométese a unha fase de validación para avaliar o seu rendemento. Utilízanse conxuntos de datos adicionais que non se usaron durante o adestramento para verificar se o modelo é capaz de adaptar o seu funcionamento a novas situacións. Se o rendemento é insatisfactorio, axústanse os parámetros e vólvese validar. Isto pode implicar axustar configuracións clave que afectan o comportamento do modelo, como a velocidade de aprendizaxe ou a complexidade do modelo.

Cuestións clave

O modelo de IA avalíase utilizando o conxunto de proba ou validación. Compáranse as predicións do modelo coas respostas reais e calcúlanse métricas de rendemento, como a precisión, a recuperación ou o *F1-score*, entre outros, segundo o tipo de problema que aborde a IA.

A validación realízase en múltiples aspectos, incluíndo:

Precisión: con que precisión realiza o modelo predicións correctas?

- Recuperación: cantos dos casos positivos reais son identificados polo modelo?
- Robustez: como se comporta o modelo fronte a datos ruidosos, atípicos ou cambiantes?
- Nesgo: existe nesgo nas predicións de certos grupos ou clases?
- Interpretabilidade: pódense entender e explicar as decisións do modelo?

A validación pode requirir axustes no modelo, como a optimización de hiperparámetros ou a mitigación de nesgos. É un proceso iterativo que busca mellorar continuamente o rendemento e a capacidade do sistema de IA para tomar decisións precisas e éticas. Unha validación exhaustiva é esencial para garantir que o sistema de IA funcione no mundo real e se axuste ás necesidades dos usuarios conforme a criterios de ética e responsabilidade.

Erros comúns

- **Non dividir adecuadamente os datos:** non separar de maneira adecuada os conxuntos de adestramento, validación e proba pode resultar nunha avaliación inadecuada do modelo e nunha sobreestimación do seu rendemento.
- **Non utilizar métricas adecuadas:** seleccionar métricas de rendemento incorrectas ou inapropiadas para o problema en cuestión pode levar a unha avaliación inexacta do sistema de IA.
- **Non ter un conxunto de validación independente:** a falta dun conxunto de validación independente para axustar hiperparámetros e avaliar o rendemento do modelo pode dar como resultado modelos mal calibrados.
- **Non abordar o desequilibrio de clases:** en problemas de clasificación desequilibrados, non tratar o desequilibrio de clases pode levar a un modelo con nesgos cara á clase dominante.
- **Non considerar a robustez:** non avaliar a robustez do modelo fronte a datos ruidosos, atípicos ou adversariais pode resultar nun sistema de IA vulnerable a ataques ou situacións inesperadas.
- **Ignorar nesgos:** non avaliar e mitigar o nesgo nos datos e as predicións pode resultar en decisións discriminatorias e problemas de equidade.
- **Non documentar o proceso:** a falta de documentación adecuada do proceso de validación pode dificultar a comprensión e a revisión do rendemento do modelo.
- **Falta de retroalimentación do usuario:** non buscar a retroalimentación dos usuarios finais pode dar lugar á implantación dun modelo que non se axusta ás súas necesidades.
- **Non realizar validación continua:** non levar a cabo avaliacións periódicas do modelo despois da súa implantación en produción pode levar a un sistema de IA desactualizado.

3.2.5 Despregadura e mantemento do sistema de IA



Unha vez que o modelo foi adestrado e avaliado satisfactoriamente, procédese á despregadura do sistema de IA nun contorno de produción, planificando a implantación e as probas antes do seu lanzamento. Igualmente, pódese proceder á implantación dos mecanismos para avaliar o rendemento do sistema de IA e os procesos de mellora continua dos resultados alcanzados.

A despregadura e o mantemento son procesos continuos que requiren unha colaboración interdisciplinar e unha xestión efectiva para asegurar que o sistema de IA opere de maneira fiable e se manteña aliñado cos seus obxectivos e necesidades cambiantes ao longo do tempo.

A implantación implica integrar o modelo na infraestrutura existente, o que pode requiren o desenvolvemento de interfaces de usuario e servizos. Posteriormente, realízanse probas exhaustivas nun contorno de produción para verificar o seu funcionamento correcto e garantir a súa eficiencia. O control constante deste proceso é esencial. Establécense sistemas de control para rastrexar o rendemento en tempo real, o que inclúe o seguimento de métricas clave e a detección de posibles problemas de rendemento ou decisións inesperadas do modelo. A actualización e a mellora continua tamén forman parte do proceso. Os modelos de IA necesitan actualizarse para manter a súa precisión e relevancia. Isto implica readestrar o modelo con novos datos e axustar parámetros para optimizar o seu rendemento.

A seguridade e a privacidade son consideracións críticas da despregadura e o mantemento do sistema de IA, polo que é necesario levar a cabo medidas para protexer os datos e o modelo de ciberseguridade, e garantir o cumprimento das regulacións de privacidade. A avaliación ética do sistema tamén resulta fundamental, o que supón realizar unha avaliación continua para detectar e mitigar nesgos e decisións discriminatorias do modelo, asegurando a equidade e a imparcialidade.

Cuestións clave

Nesta actividade, o modelo adestrado implántase nun contorno de produción e garántese o seu funcionamento continuo e eficiente. Comprende varias etapas clave:

- **Implantación:** o modelo de IA intégrase na infraestrutura tecnolóxica existente, o que pode requiren o desenvolvemento de aplicacións, interfaces de usuario ou servizos que interactúen co modelo.
- **Probas en contorno de produción:** realízanse probas exhaustivas nun contorno de produción real para garantir que o modelo funcione correctamente e que non haxa problemas de rendemento nin erros inesperados.

- **Control e supervisión:** establécense sistemas para rastrexar o rendemento do modelo en tempo real. Isto inclúe o seguimento de métricas de rendemento, a detección de anomalías e a identificación de problemas.
- **Actualización e mellora continua:** o modelo actualízase periodicamente para adaptarse a cambios nos datos, as necesidades do usuario ou a evolución da tecnoloxía. As actualizacións poden incluír o readestramento do modelo con novos datos.

A implantación de medidas de seguridade e privacidade, a avaliación ética do sistema e o soporte á resolución de problemas e consultas a persoas usuarias son tamén elementos clave da despregadura e o mantemento do sistema de IA.

Erros comúns

- **Falta de probas exhaustivas:** non realizar probas en profundidade nun contorno de produción pode levar a problemas non detectados e erros críticos no sistema.
- **Non establecer un control adecuado:** a falta de sistemas de control en tempo real pode atrasar a detección de problemas e anomalías, o que afecta a fiabilidade do sistema.
- **Actualización infrecuente ou inexistente:** non actualizar o modelo de IA ou os algoritmos pode levar á obsolescencia e á perda de precisión co tempo.
- **Non abordar a seguridade:** ignorar a seguridade de datos e modelos pode expor o sistema a ameazas cibernéticas, roubo de datos ou manipulación.
- **Falta de consideración ética e de privacidade:** non abordar cuestións éticas ou de privacidade na xestión de datos e decisións do modelo pode resultar en problemas legais e de confianza.
- **Non abordar problemas de nesgo e equidade:** a falta de avaliación e mitigación de nesgos pode resultar en decisións discriminatorias e problemas de equidade.
- **Ausencia de actualizacións de políticas:** non adaptar as políticas e regulacións de acordo cos cambios na tecnoloxía de IA pode levar a incumprimentos normativos.
- **Falta de soporte e capacitación para usuarios:** non proporcionar soporte adecuado para os usuarios e non capacitalos pode levar a unha mala adopción do sistema.
- **Non documentar cambios e decisións:** a falta de documentación adecuada de cambios e decisións no sistema dificulta a transparencia e a auditoría.

Capítulo 4

Xestión de riscos dun sistema de IA

Este capítulo afonda na xestión de riscos específicos para sistemas de intelixencia artificial (IA). Abórdanse procesos clave, desde a identificación e avaliación de riscos ata a implantación de estratexias de tratamento. Destácase a importancia de considerar riscos operativos e aqueles que impactan na saúde, a seguridade e os dereitos fundamentais. Explórase a determinación do «apetito ao risco», detállase o contexto do sistema e sublíñase a avaliación de riscos baseada na probabilidade e o impacto.



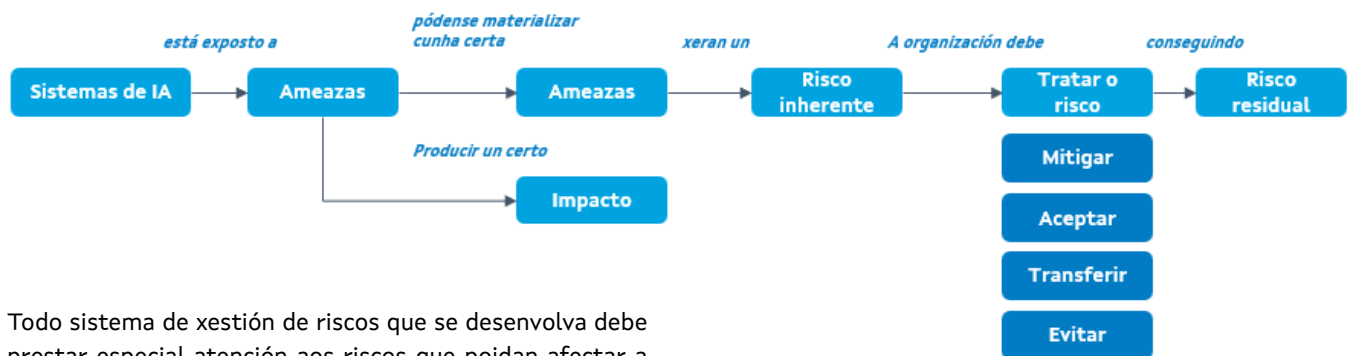
4. Xestión de riscos dun sistema de IA

Un sistema de xestión de riscos é un sistema de xestión cuxo obxectivo é a análise e o tratamento dos riscos aos que unha organización pode estar exposta. A análise de riscos permite á organización avaliar os seus activos e anticipar posibles escenarios, mentres que o tratamento de riscos se enfoca en desenvolver estratexias para reducir eses riscos de maneira efectiva.

4.1 Introducción á xestión de riscos

Un sistema de IA pode estar exposto a diferentes ameazas que poderían terminar supoñendo un risco e causar dano ao sistema. Estas ameazas poden materializarse cunha certa probabilidade e producir un certo impacto, dando orixe a un risco inherente. A organización debe abordar os riscos identificados e traballar para reducilos a un nivel residual aceptable que permanecerá no sistema. Para logralo, existen estratexias como mitigar, aceptar, evitar ou transferir os devanditos riscos.

Ecosistema da xestión de riscos



Todo sistema de xestión de riscos que se desenvolva debe prestar especial atención aos riscos que poidan afectar a saúde, a seguridade e os dereitos fundamentais das persoas, de forma que se deberá poñer o foco na identificación e análise de calquera risco que poida ter especial afectación nos devanditos elementos, así como na implantación das medidas que permitan darlles cobertura.

4.2 Modelo de xestión de riscos dun sistema de IA

Un sistema de xestión de riscos fundaméntase na execución dunha serie de procedementos destinados a identificar e avaliar diversos riscos, ademais de definir e implantar medidas adecuadas para xestionar o impacto dos devanditos riscos. A continuación, preséntanse os procesos que conforman un modelo de xestión de riscos, que debe entenderse como un proceso transversal que debe ser actualizado a medida que se identifican novos riscos en cada unha das etapas do ciclo de vida do sistema de IA, desde o deseño e o desenvolvemento ata a súa implantación e produción.

Modelo de xestión de riscos dun sistema de intelixencia artificial



4.2.1 Definición do apetito ao risco

A definición do apetito ao risco é a definición do nivel de risco que unha organización está disposta a asumir. Defínese de maneira cuantitativa a partir da configuración dunha escala de risco e o obxectivo é establecer un limiar no nivel de risco que a organización considere oportuno. Nesta decisión, débense ter en consideración os riscos que poidan afectar o funcionamento do propio sistema e, ao tratarse dun sistema de IA e estar regulado pola normativa comunitaria, tamén aqueles riscos que o sistema poida supoñer para a saúde, a seguridade e os dereitos fundamentais das persoas. Non se trata de cuantificar esta escala con precisión numérica, senón valoralo nunha escala conceptual xeral para toda a organización, e ser coherente na súa aplicación a todos os casos. En xeral, esta escala configúrase definindo un valor mínimo (xeralmente igual a 1) e un valor máximo.

O valor ou nivel de risco calcúlase como o produto do impacto pola probabilidade de que suceda. Por iso, é importante que as escalas que cuantifican o impacto e a probabilidade de que materialice unha ameaza estean definidas de maneira coherente coa primeira.

4.2.2. Contexto do sistema de IA

O contexto dun sistema refírese ao conxunto de elementos que inflúen no seu funcionamento, tanto desde o contorno externo como o interno. A finalidade deste proceso é recompilar un inventario de todos os elementos do contexto que poden impactar na análise de riscos. Este proceso adoita ser maioritariamente cualitativo e subxectivo, e a precisión da recompilación destes elementos do contorno é esencial para a súa completa comprensión. A continuación, detállase o exemplo dalgúns dos elementos máis relevantes que completan este contorno:



Os factores sociais, culturais, políticos, legais, regulamentarios, financeiros, tecnolóxicos, económicos e ambientais.



As principais tendencias tecnolóxicas, os avances en áreas da IA e as implicacións sociais e políticas da despregadura destas tecnoloxías que poden afectar o noso sistema.



As principais partes interesadas, as súas percepcións, valores, necesidades e expectativas.



A complexidade das redes e as súas dependencias, que poden aumentar co uso de tecnoloxías de IA.



Os factores internos organizacionais que viran arredor da visión, a misión, os valores, a cultura, a estratexia, o modelo de goberno, as políticas, normas e procedementos adoptados e as relacións contractuais e compromisos.



A cultura da organización, así como guías, modelos e normas adoptados.



As capacidades da organización, recursos e coñecementos (capital, tempo, persoas, procesos, tecnoloxías, etc.).



O uso de datos e fluxos de información. Os sistemas de IA poden utilizarse para automatizar, optimizar e mellorar o tratamento dos datos.



4.2.3. Identificación do risco

A identificación do risco implica a análise e identificación de todos os posibles riscos que poden afectar o sistema de IA. Neste proceso débense ter en consideración aqueles riscos que poidan poñer en perigo o funcionamento do sistema, a seguridade das persoas que o utilizan, así como os dereitos fundamentais das persoas.

Nesta fase avalíanse e documéntanse os riscos co fin de tomar medidas para evitalos, mitígalos ou xestionalos dunha maneira efectiva e garantir un funcionamento seguro e fiable do sistema. Para iso, en primeiro lugar, identifícanse os compoñentes do sistema e, en segundo lugar, a partir das fontes de risco máis comúns, analízanse e identifícanse os riscos asociados aos compoñentes identificados.

a) Identificación de compoñentes dun sistema de IA

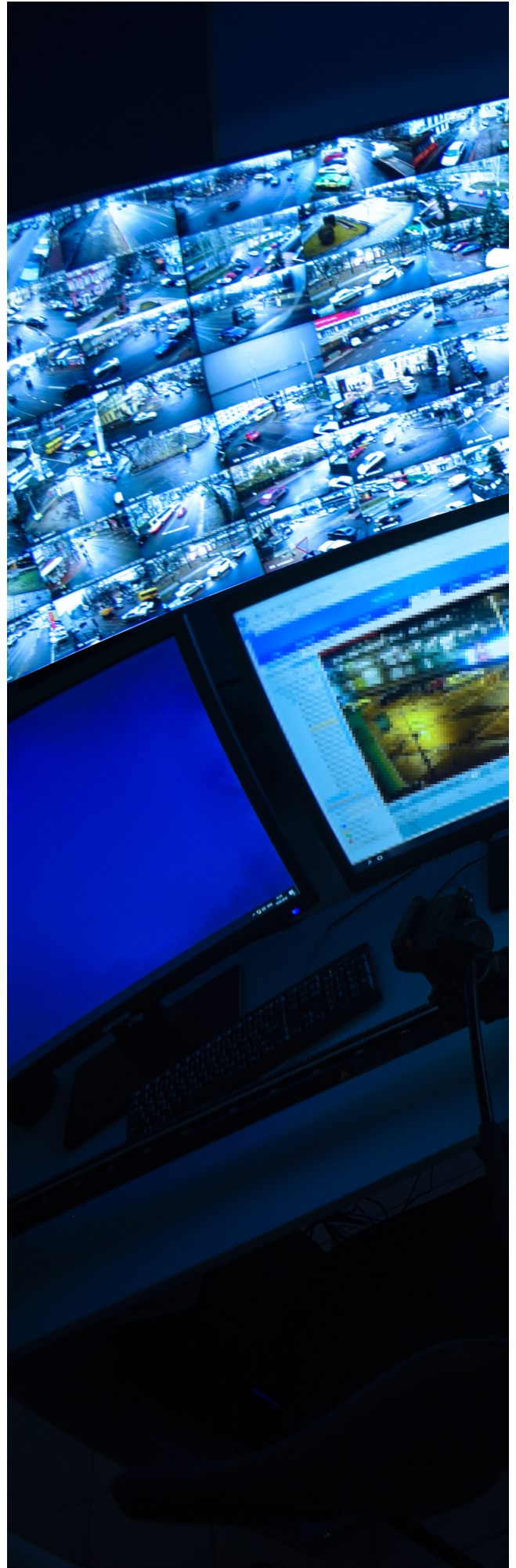
Nesta fase de identificación de compoñentes, débense incluír todos os elementos que compoñen un sistema de IA en todas as etapas do seu ciclo de vida tales como actores principais ou *stakeholders*, datos, modelos, procesos, contorno e ferramentas. A continuación, detállanse exemplos dalgúns dos compoñentes dun sistema de IA:

Exemplos de compoñentes dun sistema de IA

Propietario de datos	Persoa responsable na organización do dato, encargada da súa definición, clasificación, protección, uso e calidade
Científico de datos	Profesionais que aplican a estatística, a aprendizaxe automática e os enfoques analíticos para analizar diferentes conxuntos de datos de distintos tamaños e formas e resolver problemas complexos e críticos
Datos de adestramento	Datos iniciais que se utilizan para desenvolver un sistema de IA, a partir dos cales o sistema adapta os seus parámetros internos para refinar as súas regras
Datos de proba	Conxunto de datos utilizado para proporcionar unha avaliación non nesgada dun sistema de IA axustado ao conxunto de datos de adestramento. Os datos de proba utilízanse para probar o sistema
Parámetros do sistema de IA	Variables de configuración internas do sistema de IA e cuxo valor pode estimarse a partir dos datos dados
Preprocesamento de datos	Comprensión, preparación e limpeza dos datos
Inxestión de datos	Proceso relacionado co transporte de datos desde múltiples fontes para compoñer puntos de datos multidimensionais; os datos poden colocarse nun medio de almacenamento ao que se poida acceder, utilízaos e analízaos, ou ben, o fluxo de datos pode utilizarse directamente polo sistema de IA
Etiquetaxe de datos	Proceso de detección e etiquetaxe das mostras de datos; o proceso pode ser manual e lento e estar asistido por <i>software</i>

b) Fontes comúns no proceso de identificación de riscos da IA

O proceso de identificación de riscos baséase na análise e identificación de riscos relacionados cun compoñente específico do sistema nun elemento particular do contexto. Un mesmo compoñente, cando se analiza nun contorno de contexto diferente, pode levar á identificación de riscos distintos. A continuación, detállanse algunhas das fontes máis comúns, que poden servir como punto de partida para esta análise e identificación:



Falta de transparencia

A transparencia é a capacidade de entender e explicar como un sistema de IA toma as súas decisións ou realiza as súas accións. Cando un sistema de IA non é transparente, resulta difícil rastrexar e comprender por que tomou unha determinada decisión. Isto pode xerar preocupacións éticas, legais e prácticas, xa que as súas decisións poden ser difíciles de xustificar ou poden levar a resultados inxustos. Para mitigar este risco, é importante traballar en facer que os sistemas de IA sexan máis transparentes, o que significa que os seus procesos de toma de decisións deben ser comprensibles e explicables para as persoas que os utilizan e confían neles.

Falta de explicabilidade

A explicabilidade é a capacidade de comprender e explicar por que un sistema de IA elixe certas accións ou decisións en lugar doutras. Cando a explicabilidade é insuficiente, resulta difícil entender os factores que influíron nas decisións da IA ou o seu razoamento subxacente. Se non se poden explicar estes factores, a validación do sistema de IA e a confianza no sistema vense afectados negativamente, xa que non está claro por que o sistema tomou unha decisión e se tomará a decisión correcta en todos os casos.

Vixilancia humana e automatización

O nivel de automatización é o nivel de responsabilidade e control que se lle concede a un sistema de IA para tomar decisións e realizar tarefas sen intervención humana. Se o nivel de automatización é alto e o sistema de IA ten un papel significativo na toma de decisións ou na execución de tarefas críticas, existe un risco potencial. Isto débese a que a IA pode cometer erros ou tomar decisións subóptimas en certas situacións, especialmente se non se entende completamente o seu funcionamento. É importante equilibrar o nivel de automatización para garantir que haxa unha supervisión e un control humano adecuados.

Fontes de risco relacionadas coa aprendizaxe automática

O comportamento dos sistemas de IA depende non só dos algoritmos en uso, senón tamén dos datos cos que se adestran os sistemas. Existen diversos riscos derivados do uso dos datos. Por exemplo:

- A calidade inadecuada dos datos podería afectar a varios obxectivos como a equidade, a seguridade e a solidez.
- Os datos poden deixar de ser representativos do dominio de aplicación, o que supón riscos para os obxectivos do negocio.
- A obtención e o almacenamento de datos pode incorrer en riscos éticos e legais significativos.

Problemas de hardware do sistema

As fontes de risco relacionadas con problemas de *hardware* inclúen, por exemplo, erros baseados en compoñentes defectuosos. O desenvolvemento de sistemas de IA podería verse limitado debido ás diferentes capacidades de *hardware* dos sistemas en termos de potencia de procesamento, memoria e a dispoñibilidade de aceleradores de *hardware* de IA dedicados.

Problemas do ciclo de vida do sistema

Os métodos, procesos e tamén o uso inadecuado ou insuficiente dun sistema de IA ao longo do seu ciclo de vida poden xerar riscos. Por exemplo, un proceso de deseño defectuoso pode non anticipar os contextos nos que se utilizará o sistema de IA, o que fai que falle inesperadamente cando se usa nestes contextos.

Preparación tecnolóxica

A preparación tecnolóxica indica o madura que é unha tecnoloxía dada nun contexto de aplicación determinado. As tecnoloxías menos maduras utilizadas no desenvolvemento e a aplicación de sistemas de IA poden engadir riscos descoñecidos para a organización ou difíciles de avaliar. Para as tecnoloxías maduras, pódese dispoñer dunha maior variedade de datos de experiencia, o que facilita a identificación e a avaliación dos riscos.

Complexidade do contorno

A complexidade do contorno dun sistema de IA determina a gama de situacións que un sistema de IA pode soportar no seu contexto operativo. Un dos riscos máis relevantes, por exemplo, é o relacionado co grao de comprensión do contorno do sistema de IA. Unha comprensión parcial do contorno dará lugar a un nivel de incerteza que é unha fonte de risco especialmente relevante na fase do deseño dos sistemas de IA.

Outras potenciais fontes de riscos

- Dificultade na identificación de responsabilidades e rendición de contas.
- Uso inadecuado, incorrecto ou fraudulento do sistema de IA.
- Ameazas de seguridade e ciberseguridade.
- Posibles ameazas á privacidade das persoas se non se desenvolve un adecuado goberno dos datos.
- Posibles perturbacións ou manipulacións nos datos non desexadas.
- Os posibles nesgos nos datos tamén poden supoñer unha ameaza para o adecuado uso do sistema de IA.

4.2.4. Análise e avaliación de riscos

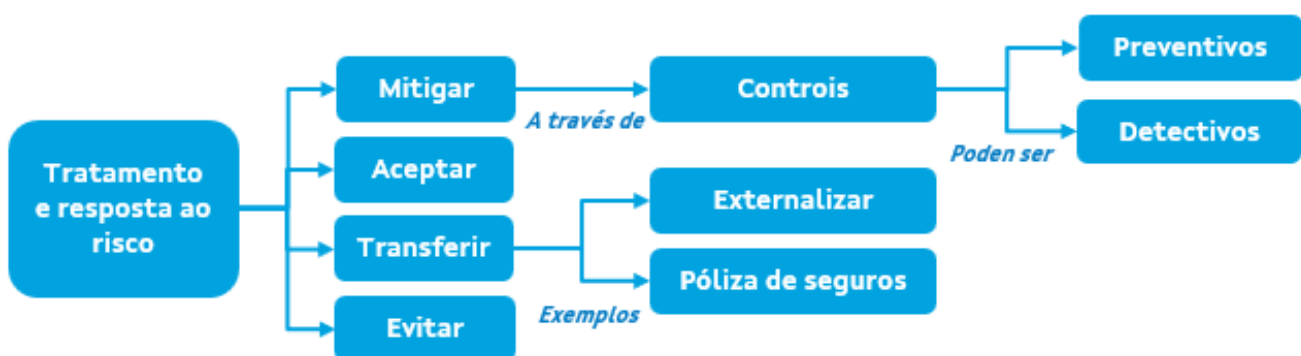
O proceso de análise e avaliación de riscos céntrase en calcular, para cada un dos riscos identificados no proceso anterior, o seu nivel de risco. Este valor de risco obtense a partir do produto da probabilidade de ocorrencia do risco e a magnitude do seu impacto. Ambos os factores deben ser avaliados utilizando escalas predefinidas que midan tanto o impacto como a probabilidade.

As escalas utilizadas para avaliar estes factores deben ser coherentes coa escala previamente definida para o apetito ao risco, xa que o seguinte paso implica a comparación destes valores. Unha vez se estableceron os niveis de impacto e probabilidade e se calculou o valor do risco, compárase co limiar de apetito ao risco preestablecido. Se o nivel de risco obtido supera o limiar, é necesario determinar que enfoque de resposta é o máis adecuado para a situación: evitar, transferir, mitigar ou aceptar o risco.

4.2.5. Resposta ao risco

As respostas ao risco son estratexias que unha organización pode adoptar para abordar e xestionar os riscos nun contexto específico. A organización debe determinar que tipo de medidas de tratamento de riscos se seleccionarán para abordar cada un dos riscos identificados e analizados. As opcións principais son as que se indican a continuación:

Visión global das alternativas para a resposta ao risco



Mitigar

A mitigación do risco implica tomar medidas proactivas de carácter preventivo ou de detección para reducir a probabilidade de que ocorra un risco ou minimizar o seu impacto se ocorre. Isto pode incluír a implantación de controis, a adopción de mellores prácticas, a capacitación do persoal ou a optimización de procesos para reducir a exposición ao risco ata alcanzar o limiar de apetito definido.

Aceptar

Aceptar un risco significa que a organización recoñece a existencia do risco pero decide non tomar medidas activas ao respecto. Esta resposta pode ser adecuada cando o risco é considerado tolerable ou cando o custo de mitigar o risco supera os posibles danos que podería causar.

Transferir

A transferencia de risco implica delegar a responsabilidade de xestionar o risco a outra entidade, xeralmente a través de acordos contractuais ou de seguros. Por exemplo, unha organización pode transferir o risco de perda por danos á propiedade a unha compañía de seguros mediante a compra dunha póliza de seguro.

Evitar

Evitar un risco implica tomar medidas para eliminar completamente a exposición ao risco. Isto pode lograrse evitando actividades, procesos ou decisións que poidan dar lugar ao risco. É unha estratexia radical que se utiliza cando o risco é inaceptable ou non pode ser xestionado de maneira efectiva de ningunha outra forma.

Unha vez definidas as medidas de tratamento do risco débense definir plans de acción, planificar a súa implantación e designar un responsable para levar a cabo a súa execución. Estratexicamente, debe priorizarse a implantación daqueles tratamentos que mitiguen os riscos máis relevantes.

4.2.6. Documentación

Débase manter un rexistro dos riscos identificados e analizados na documentación asociada ao sistema, así como establecer períodos de revisión e seguimento do modelo de xestión de riscos (débense incorporar todos aqueles novos riscos que se identifiquen, así como avaliar e definir as medidas de tratamento oportunas).

4.2.7. Comunicación e consulta

A comunicación e consulta teñen como obxectivo xerar consciencia e comprensión dos riscos. É fundamental manter actualizadas a todas as partes interesadas apropiadas, tanto internas como externas, en todas as fases do proceso de xestión de riscos.

Todos os participantes no deseño, desenvolvemento e implantación do sistema de IA deben ter coñecemento do sistema de xestión de riscos e deben participar activamente na súa implantación e nas actualizacións correspondentes, especialmente as áreas da organización que estean directamente involucradas no proceso.

4.2.8. Seguimento e mellora continua

A finalidade do seguimento e mellora continua é garantir a seguridade e calidade do sistema de IA, manténdoo en constante evolución e logrando que o sistema se adapte aos cambios no contorno. Esta fase implica a actualización dos procesos previamente descritos segundo a periodicidade establecida ou ante cambios significativos na organización ou no sistema: desde a actualización do contexto do sistema á revisión e actualización dos plans de acción para o tratamento dos riscos.

Este seguimento e actualización debe levarse a cabo en todas as etapas do ciclo de vida dun sistema de IA. O contexto, os compoñentes do sistema e os riscos que se poden identificar variarán segundo a etapa de desenvolvemento na que se atope o sistema, polo que os procesos descritos neste capítulo serán repetidos a medida que o sistema se deseñe, desenvolva e implante.

4.2.9. Liderado e compromiso

Na xestión de riscos, o liderado e o compromiso son compoñentes transversais que deben estar presentes de maneira constante e coherente en todas as actividades relacionadas coa xestión de riscos na organización. Os líderes deben desempeñar un papel activo en todas as etapas, desde a identificación e avaliación de riscos ata a implantación de medidas de control e a revisión periódica do sistema.

Este enfoque transversal asegura que a xestión de riscos non sexa simplemente un proceso illado, senón unha parte integral da cultura e a operación diaria da empresa, o que contribúe a unha toma de decisións máis efectiva e á protección sostible dos activos e o cumprimento dos obxectivos organizativos.



4.3. Exemplo de xestión de riscos no deseño, desenvolvemento e implantación dun sistema de IA

Ao longo desta sección detallarase o proceso completo de xestión de riscos dun sistema de IA mediante o seguinte exemplo:

Decídese deseñar, desenvolver e implantar un sistema de distribución de subvencións a persoas físicas e empresas baseado en IA para axudar as Administracións públicas na toma de decisións, co obxectivo de mellorar a súa eficiencia, ofrecer unha maior transparencia no proceso de asignación e proporcionar unha resposta máis áxil a situacións de emerxencia económica.

4.3.1. Ideación do sistema de IA

Paso 0. Determinación do apetito ao risco

O apetito ao risco debe definirse de forma cuantitativa. Para este exemplo, decídese establecer unha escala do apetito ao risco do 1 ao 15 (sendo o 15 o valor máximo de risco que a organización está disposta a aceptar).

Esta escala debe aliñarse con outras dúas escalas: unha que vai do 1 ao 5 para clasificar os riscos en cinco niveis de impacto (desde moi baixo ou inexistente ata moi alto ou crítico, en orde ascendente) e outra que abarca do 1 ao 3 para avaliar a probabilidade de que estes riscos se materialicen (desde improbable ata moi probable, en orde ascendente). Para o exemplo que se presenta, determínase o apetito ao risco nun nivel intermedio, un valor 4 sobre 15.

Paso 1. Análise de riscos na ideación do sistema de IA

Contexto						
Cumprimento co dereito fundamental da protección de datos de carácter persoal						
Identificación de compoñentes	Identificación de riscos					
Provedor de datos	A falta de avaliación de riscos relacionados coa privacidade e a ética pode levar á recompilación e o uso non ético de datos, o que pode resultar en problemas legais e reputacionais.					
	<table><thead><tr><th>Risco inherente (Impacto x Probabilidade)</th><th>Conclusión da avaliación</th></tr></thead><tbody><tr><td>5 (Impacto crítico) x 1 (Improbable) = 5</td><td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td></tr><tr><td>Risco inherente (5) > Apetito (4)</td></tr></tbody></table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	5 (Impacto crítico) x 1 (Improbable) = 5	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (5) > Apetito (4)
Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación					
5 (Impacto crítico) x 1 (Improbable) = 5	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.					
Risco inherente (5) > Apetito (4)						
	<table><thead><tr><th>Control</th><th>Mitigación e risco social</th></tr></thead><tbody><tr><td>Realizar unha avaliación exhaustiva dos riscos de privacidade e ética asociados coa recompilación e o uso de datos. Establecer políticas e prácticas claras para garantir o cumprimento das regulacións de protección</td><td>Mitigación: 70 % Risco residual: 1,5</td></tr></tbody></table>	Control	Mitigación e risco social	Realizar unha avaliación exhaustiva dos riscos de privacidade e ética asociados coa recompilación e o uso de datos. Establecer políticas e prácticas claras para garantir o cumprimento das regulacións de protección	Mitigación: 70 % Risco residual: 1,5	
Control	Mitigación e risco social					
Realizar unha avaliación exhaustiva dos riscos de privacidade e ética asociados coa recompilación e o uso de datos. Establecer políticas e prácticas claras para garantir o cumprimento das regulacións de protección	Mitigación: 70 % Risco residual: 1,5					

de datos e a ética na IA. Isto pode incluír a anonimización de datos sensibles e a obtención de consentimento informado cando sexa necesario.

Identificación de compoñentes

Propietario de datos

Identificación de riscos

Falta de claridade na propiedade, responsabilidade e autoridade sobre os datos que se utilizarán no sistema de IA. Pode resultar en disputas legais ou problemas éticos relacionados co acceso e o uso dos datos.

Risco inherente (Impacto x Probabilidade)

5 (Impacto crítico) x 1 (Improbable) = 5

Risco inherente (5) > Apetito (4)

Conclusión da avaliación

O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.

Control

Realizar un inventario completo dos datos relevantes e establecer acordos claros sobre a propiedade e o uso dos datos tales como acordos de licenza de datos, acordos de confidencialidade e contratos que establezan os termos de uso.

Mitigación e risco social

Mitigación: 70 %
Risco residual: 1,5

Paso 2. Documentación, comunicación e consulta durante a fase de ideación do sistema de IA

O apetito ao risco debe definirse de forma cuantitativa. Para este exemplo, decídese establecer unha escala do apetito ao risco do 1 ao 15 (sendo o 15 o valor máximo de risco que a organización está disposta a aceptar).

Esta escala debe aliñarse con outras dúas escalas: unha que vai do 1 ao 5 para clasificar os riscos en cinco niveis de impacto (desde moi baixo ou inexistente ata moi alto ou crítico, en orde ascendente) e outra que abarca do 1 ao 3 para avaliar a probabilidade de que estes riscos se materialicen (desde improbable ata moi probable, en orde ascendente). Para o exemplo que se presenta, determínase o apetito ao risco nun nivel intermedio, un valor 4 sobre 15.

4.3.2. Deseño do sistema de IA

Paso 3. Análise de riscos no deseño do sistema de IA

Contexto						
As capacidades da organización, recursos e coñecementos destinados ao deseño, desenvolvemento e implantación do sistema.						
Identificación de compoñentes	Identificación de riscos					
Científicos de datos e enxeñeiros de datos	Posibilidade de que o deseño da arquitectura do sistema de IA non sexa o máis adecuado e supoña a necesidade de cambios significativos nas etapas posteriores do desenvolvemento.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>3 (Impacto medio) x 2 (Probable) = 6</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (6) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (6) > Apetito (4)
	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación				
	3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.				
Risco inherente (6) > Apetito (4)						
<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td>Realizar unha revisión exhaustiva das opcións arquitectónicas dispoñibles, tendo en conta os requisitos do proxecto, e realizar probas de concepto e prototipado para avaliar diferentes enfoques antes de tomar unha decisión final.</td> <td>Mitigación: 80 % Risco residual: 1,2</td> </tr> </tbody> </table>	Control	Mitigación e risco social	Realizar unha revisión exhaustiva das opcións arquitectónicas dispoñibles, tendo en conta os requisitos do proxecto, e realizar probas de concepto e prototipado para avaliar diferentes enfoques antes de tomar unha decisión final.	Mitigación: 80 % Risco residual: 1,2		
Control	Mitigación e risco social					
Realizar unha revisión exhaustiva das opcións arquitectónicas dispoñibles, tendo en conta os requisitos do proxecto, e realizar probas de concepto e prototipado para avaliar diferentes enfoques antes de tomar unha decisión final.	Mitigación: 80 % Risco residual: 1,2					
Identificación de compoñentes	Identificación de riscos					
Plataforma de inxestión de datos	O sistema de IA non pode escalar eficazmente para manexar un aumento na cantidade de datos ou na carga de traballo, e resulta nun sistema ineficiente ou na necesidade de realizar melloras custosas.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>4 (Impacto alto) x 2 (Probable) = 8</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (8) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	4 (Impacto alto) x 2 (Probable) = 8	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (8) > Apetito (4)
	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación				
	4 (Impacto alto) x 2 (Probable) = 8	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.				
Risco inherente (8) > Apetito (4)						
<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> </tr> </tbody> </table>	Control	Mitigación e risco social				
Control	Mitigación e risco social					

Realizar unha revisión exhaustiva das opcións arquitectónicas dispoñibles, tendo en conta os requisitos do proxecto, e realizar probas de concepto e prototipado para avaliar diferentes enfoques antes de tomar unha decisión final.

Mitigación: 80 %
Risco residual: 1,2

Paso 4. Documentación, comunicación e consulta durante a fase de deseño do sistema de IA

A continuación, débense executar a documentación, comunicación e consulta asociadas aos riscos.

4.3.3. Desenvolvemento e adestramento do sistema de IA

Paso 5. Análise de riscos no desenvolvemento e adestramento do sistema de IA

Contexto						
As capacidades da organización, recursos e coñecementos: a orixe, a cantidade e a propiedade dos datos e modelos usados no desenvolvemento, adestramento e validación do sistema.						
Identificación de compoñentes	Identificación de riscos					
Datos de adestramento	Escaseza de datos que pode comprometer a viabilidade da IA e/ou comprometer e limitar os seus resultados. Isto pode ser explotado deliberada ou involuntariamente durante a inxestión de datos.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>3 (Impacto medio) x 2 (Probable) = 6</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (6) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (6) > Apetito (4)
Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación					
3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.					
Risco inherente (6) > Apetito (4)						
	<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td>Definir e supervisar os indicadores do bo funcionamento do modelo e garantir que os sistemas de IA cumpren os requisitos de seguridade e calidade de datos ao longo de todo o ciclo de vida.</td> <td>Mitigación: 70 % Risco residual: 1,8</td> </tr> </tbody> </table>	Control	Mitigación e risco social	Definir e supervisar os indicadores do bo funcionamento do modelo e garantir que os sistemas de IA cumpren os requisitos de seguridade e calidade de datos ao longo de todo o ciclo de vida.	Mitigación: 70 % Risco residual: 1,8	
Control	Mitigación e risco social					
Definir e supervisar os indicadores do bo funcionamento do modelo e garantir que os sistemas de IA cumpren os requisitos de seguridade e calidade de datos ao longo de todo o ciclo de vida.	Mitigación: 70 % Risco residual: 1,8					
Identificación de compoñentes	Identificación de riscos					
Propietario de datos	Os propietarios dos datos poden tratar de ocultar a información que se vai fornecer ao sistema de IA como parte dos seus intereses comerciais e, en consecuencia, introducir nesgos.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>3 (Impacto medio) x 2 (Probable) = 6</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (6) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (6) > Apetito (4)
Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación					
3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.					
Risco inherente (6) > Apetito (4)						
	<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> </tr> </tbody> </table>	Control	Mitigación e risco social			
Control	Mitigación e risco social					

Definición de KPI para medir os nesgos nos grupos protexidos, limiares de tolerancia e plans de remediación para eliminar as posibles imparcialidades para garantir que os modelos sexan imparciais.

Mitigación: 80 %
Risco residual: 1,2

Paso 6. Documentación, comunicación e consulta durante a fase de desenvolvemento e adestramento do sistema de IA

A continuación, débense executar a documentación, comunicación e consulta asociadas aos riscos.

4.3.4. Validación dos resultados do sistema de IA

Paso 7. Análise de riscos na validación dos resultados do sistema de IA

Contexto						
As capacidades da organización, recursos e coñecementos: a orixe, a cantidade e a propiedade dos datos e modelos usados no desenvolvemento, adestramento e validación do sistema.						
Identificación de compoñentes	Identificación de riscos					
Datos de validación	Introdución de datos adversos no conxunto de datos de validación que comprometan a adecuada validación do modelo adestrado.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>5 (Impacto crítico) x 1 (Improbable) = 5</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (5) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	5 (Impacto crítico) x 1 (Improbable) = 5	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (5) > Apetito (4)
Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación					
5 (Impacto crítico) x 1 (Improbable) = 5	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.					
Risco inherente (5) > Apetito (4)						
	<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td>Establecer políticas e procedementos claros que prohiban o acceso ou a modificación dos datos de validación durante o proceso de adestramento e realizar verificacións e auditorías periódicas dos conxuntos de datos de validación para identificar anomalías.</td> <td>Mitigación: 80 % Risco residual: 1</td> </tr> </tbody> </table>	Control	Mitigación e risco social	Establecer políticas e procedementos claros que prohiban o acceso ou a modificación dos datos de validación durante o proceso de adestramento e realizar verificacións e auditorías periódicas dos conxuntos de datos de validación para identificar anomalías.	Mitigación: 80 % Risco residual: 1	
Control	Mitigación e risco social					
Establecer políticas e procedementos claros que prohiban o acceso ou a modificación dos datos de validación durante o proceso de adestramento e realizar verificacións e auditorías periódicas dos conxuntos de datos de validación para identificar anomalías.	Mitigación: 80 % Risco residual: 1					
Identificación de compoñentes	Identificación de riscos					
Modelos	Un arquivo de modelo lexítimo é substituído por completo por un arquivo de modelo envelenado. Esta ameaza pode materializarse explotando as posibles debilidades dos provedores da nube.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>4 (Impacto alto) x 2 (Probable) = 8</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (8) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	4 (Impacto alto) x 2 (Probable) = 8	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (8) > Apetito (4)
Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación					
4 (Impacto alto) x 2 (Probable) = 8	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.					
Risco inherente (8) > Apetito (4)						
	<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> </tr> </tbody> </table>	Control	Mitigación e risco social			
Control	Mitigación e risco social					

Considerar factores de seguridade, como as medidas de protección de datos implantadas, a reputación do provedor en canto a seguridade e as garantías contractuais en caso de incidentes de seguridade.

Mitigación: 80 %
Risco residual: 1,6

Paso 8. Documentación, comunicación e consulta durante a fase de validación dos resultados do sistema de IA

A continuación, débense executar a documentación, comunicación e consulta asociadas aos riscos.

4.3.5. Despregadura e mantemento do sistema de IA

Paso 9. Análise de riscos na despregadura e o mantemento do sistema de IA

Contexto						
Os recursos destinados ao mantemento do sistema, así como cambios nas principais tendencias tecnolóxicas e outros factores legais e regulamentarios.						
Identificación de compoñentes	Identificación de riscos					
Parámetros de adestramento	Desafío e gasto asociado á necesidade de readestrar periodicamente o modelo de IA para manter o seu rendemento e precisión ao longo do tempo.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>4 (Impacto alto) x 2 (Probable) = 8</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (8) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	4 (Impacto alto) x 2 (Probable) = 8	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (8) > Apetito (4)
Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación					
4 (Impacto alto) x 2 (Probable) = 8	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.					
Risco inherente (8) > Apetito (4)						
	<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td>Utilizar técnicas de transferencia de aprendizaxe e <i>fine-tuning</i> para aproveitar modelos preadestrados para complementar ao sistema e implantar ferramentas e procesos automatizados que optimicen o readestramento do modelo.</td> <td>Mitigación: 80 % Risco residual: 1,6</td> </tr> </tbody> </table>	Control	Mitigación e risco social	Utilizar técnicas de transferencia de aprendizaxe e <i>fine-tuning</i> para aproveitar modelos preadestrados para complementar ao sistema e implantar ferramentas e procesos automatizados que optimicen o readestramento do modelo.	Mitigación: 80 % Risco residual: 1,6	
Control	Mitigación e risco social					
Utilizar técnicas de transferencia de aprendizaxe e <i>fine-tuning</i> para aproveitar modelos preadestrados para complementar ao sistema e implantar ferramentas e procesos automatizados que optimicen o readestramento do modelo.	Mitigación: 80 % Risco residual: 1,6					
Identificación de compoñentes	Identificación de riscos					
Ferramentas de retención e borrado de datos	Cambio de regulación: cambios regulamentarios (por exemplo, protección de datos e privacidade) que puidesen afectar a dispoñibilidade e o uso dos datos necesarios para o sistema de IA.					
	<table border="1"> <thead> <tr> <th>Risco inherente (Impacto x Probabilidade)</th> <th>Conclusión da avaliación</th> </tr> </thead> <tbody> <tr> <td>3 (Impacto medio) x 2 (Probable) = 6</td> <td rowspan="2">O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.</td> </tr> <tr> <td>Risco inherente (6) > Apetito (4)</td> </tr> </tbody> </table>	Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación	3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.	Risco inherente (6) > Apetito (4)
Risco inherente (Impacto x Probabilidade)	Conclusión da avaliación					
3 (Impacto medio) x 2 (Probable) = 6	O risco é maior que o apetito definido, decídese implantar medidas de control que mitiguen o risco.					
Risco inherente (6) > Apetito (4)						
	<table border="1"> <thead> <tr> <th>Control</th> <th>Mitigación e risco social</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> </tr> </tbody> </table>	Control	Mitigación e risco social			
Control	Mitigación e risco social					

Realizar un seguimiento constante das regulacións aplicables e axustar os procesos en consecuencia.

Mitigación: 80 %
Risco residual: 1,2

Paso 10. Documentación, comunicación e consulta durante a fase de despregadura e mantemento do sistema de IA

A continuación, débense executar a documentación, comunicación e consulta previamente descritas.



5. Conclusións

A intelixencia artificial emerxeu como un pilar fundamental da transformación dixital e é a ferramenta con maior capacidade para determinar a forma na que as sociedades abordarán os desafíos e oportunidades no futuro. A súa relevancia radica na súa capacidade para automatizar tarefas, procesar datos a grande escala e tomar decisións baseadas en información. Na Administración pública, a IA converteuse nun facilitador clave para aumentar a eficiencia, a produtividade e a calidade dos servizos públicos e, en definitiva, para evolucionar a forma na que a Administración se relaciona coa cidadanía.

No entanto, a relevancia da IA tamén trae consigo desafíos, como a ética no uso de datos e a preocupación pola automatización do emprego. A IA é relevante tanto en termos de oportunidades como de desafíos, o que subliña a importancia da súa xestión efectiva.

Contar cun proceso claro, transparente e obxectivo para xestionar a IA na Administración pública é unha ferramenta indispensable para alcanzar estes beneficios. A colaboración de todas as partes interesadas, a coordinación dos distintos perfís con coñecemento arredor da IA e o establecemento de medidas de validación ou control poden axudar de forma significativa a desenvolver unha xestión responsable, ética e segura da IA na Administración pública.

A xestión de riscos constitúe un elemento esencial deste proceso. Tal e como determina a normativa europea, a xestión dos sistemas de IA debe considerar como parte central a clasificación do nivel de risco dos sistemas de IA e a adopción de medidas coherentes con cada nivel de risco. A identificación clara e continua de riscos, a súa avaliación periódica e a implantación de medidas de resposta son aspectos clave do éxito da adopción desta tecnoloxía por parte do sector público.

Impacto na sociedade

Os sistemas de IA teñen un impacto directo na sociedade, xa que se empregan nunha variedade de ámbitos e servizos públicos, incluíndo a saúde, a educación, o transporte e a seguridade. Unha xestión deficiente destes sistemas pode ter consecuencias significativas para a calidade de vida da cidadanía.

Seguridade e privacidade de datos

A xestión da IA debe garantir a seguridade e a privacidade dos datos. A recompilación e o procesamento de datos poden expoñer preocupacións sobre a protección da información persoal. A falta de medidas de seguridade pode dar lugar a incumprimentos normativos.

Equidade e xustiza

A IA pode introducir nesgos e decisións discriminatorias se non se xestiona adecuadamente. Isto pode afectar negativamente a grupos específicos da poboación e socavar a equidade e a xustiza social. A xestión da IA debe abordar activamente estes problemas para garantir un trato xusto e equitativo.

Desenvolvemento económico

A IA desempeña un papel importante no desenvolvemento económico e a competitividade dunha rexión. A xestión da IA pode impulsar o investimento en investigación e desenvolvemento, a formación de talento e a creación de novas oportunidades de negocio tanto no ámbito público como no privado.

Confianza pública

A confianza pública na IA é esencial para a súa adopción xeneralizada. A cidadanía debe confiar en que os sistemas de IA toman decisións xustas e éticas. Unha xestión adecuada, que inclúe políticas de transparencia e rendición de contas, é fundamental para manter e fortalecer esa confianza.

Preparación perante desafíos

A IA está en constante evolución, e a xestión efectiva é necesaria para que a Administración pública se prepare perante desafíos futuros, como a automatización do emprego, a ciberseguridade e a toma de decisións autónomas. A xestión da IA debe considerar estas implicacións e tomar medidas proactivas.

6. Referencias

- Axencia da Unión Europea para a Ciberseguridade (ENISA). (s. f.). *AI Cybersecurity Challenges - Threat Landscape*. <https://europa.eu>
- Axencia Galega de Coñecemento en Saúde. (s. f.). *Plataforma de innovación sanitaria*. <https://sergas.es>
- Banco Interamericano de Desarrollo (BID). (s. f.). *Uso responsable de IA para política pública: manual de formulación de proyectos*. <https://iadb.org>
- Britainthinks. (s. f.). *AI Governance*. <https://publishing.service.gov.uk>
- CAF (Banco de Desarrollo de América Latina). (s. f.). *Conceptos fundamentales y uso responsable de la Inteligencia Artificial en el Sector Público*. <https://caf.com>
- Centre for Data Ethics and Innovation. (s. f.). *The Roadmap to an effective AI assurance ecosystem*. <https://publishing.service.gov.uk>
- Comisión Europea. (s. f.). *Inteligencia artificial: La Comisión continúa su trabajo sobre directrices éticas*. <https://europa.eu>
- Emiratos Árabes Unidos. (s. f.). *AI Adoption Guideline in Govern*.
- GAIN. (s. f.). *Guía de buenas prácticas para favorecer la CPI en Galicia*. <https://galiciainnovacion.es>
- Gartner. (s. f.-a). *Definition of Total Cost of Ownership*. <https://gartner.com>
- Gartner. (s. f.-b). *How to Staff Your AI Team*. <https://gartner.com>
- Gobierno de España. (s. f.-a). *Diez años cooperando en evaluación de tecnologías sanitarias*. <https://sanidad.gob.es>
- Gobierno de España. (s. f.-b). *Guía sobre Compra Pública Innovadora*. <https://mineco.gob.es>
- Gobierno de Reino Unido. (s. f.-a). *A guide to using artificial Intelligence in the public sector*. <https://publishing.service.gov.uk>
- Gobierno de Reino Unido. (s. f.-b). *Guidelines for AI procurement*. <https://publishing.service.gov.uk>
- Gobierno de Reino Unido. (s. f.-c). *Managing your artificial intelligence project*. <https://www.gov.uk>
- Gobierno de Reino Unido. (s. f.-d). *Survey on AI Assurance Methods*. <https://www.gov.uk>
- IDRC. (s. f.). *Government Artificial Intelligence Readiness Index*.
- Montaña Merchán Arribas. (s. f.). *Guía de uso de la Inteligencia Artificial en el Sector Público*. <https://wordpress.com>
- Neptune. (s. f.). *How to Build Machine Learning Teams That Deliver*. <https://neptune.ai>
- Organización Internacional de Normalización (ISO). (s. f.-a). *ISO 31000:2018 – Xestión de riscos – Directrices*.
- Organización Internacional de Normalización (ISO). (s. f.-b). *ISO/IEC 23894 - Tecnología de la información - IA - Orientaciones sobre la gestión de riesgos*.
- SERGAS. (s. f.). *Guía para la elaboración y adaptación de informes rápidos de evaluación de tecnologías sanitarias*. <https://sergas.gal>
- World Economic Forum. (s. f.). *Lineamientos para los gobiernos sobre adquisiciones de sistemas de inteligencia artificial*. <https://weforum.org>



XUNTA
DE GALICIA



AXENCIA PARA A
MODERNIZACIÓN
TÉCNOLÓXICA DE GALICIA