

Intelixencia Artificial

Guía do profesorado



ÍNDICE

1. INTRODUCCIÓN	3
1.1 Que podes esperar desta guía	3
2. SENTANDO AS BASES	5
2.1 Que é a Intelixencia Artificial.....	5
2.2 A Intelixencia Natural fronte á Intelixencia Artificial.....	6
2.3 Historia da Intelixencia Artificial.....	7
2.4 Intelixencia Artificial na sociedade	11
3. BLOQUES BÁSICOS DUN SISTEMA DE INTELIXENCIA ARTIFICIAL	12
3.1 UX1A percibe	14
3.1.1 Tipos de sensores da Intelixencia Artificial	15
3.1.2 A rede de sensores de UX1A	18
3.2 UX1A representa e razoa	20
3.2a Representación.....	20
3.2a.1 Procesar ou tratar os datos.....	21
3.2b. Razoamento	23
3.2b.1 Operacións lóxicas básicas	25
3.2b.2 Programación e algoritmos.....	28
3.3 UX1A aprende.....	30
3.3.1 Aprendizaxe automática.....	31
3.3.2 Tipos de aprendizaxe automática	32
3.3.3 Redes neuronais e aprendizaxe profunda.....	32
3.3.4 Redes neuronais empregadas na IA	36
3.4 UX1A interactúa	38
3.4.1 Tipos de actuadores da Intelixencia Artificial	39
3.4.2 A rede de actuadores de UX1A	40
3.5 UX1A impacta na sociedade.....	42
3.5.1 Ética.....	51
3.5.2 Sostibilidade	53
3.5.3 Aspectos legais	54
3.5.4 Oportunidades e preocupacións fundamentais que propón a IA.....	56
4. REMATANDO	65
GLOSARIO	66

1. INTRODUCCIÓN

Hoxe en día, a Intelixencia Artificial (IA) xoga un papel importante na vida de moitas persoas e, aínda que a miúdo pasa inadvertida, está a xerar profundas consecuencias na sociedade e no ser humano. A IA está cambiando o xeito no que interactuamos, como nos entendemos, como coidamos a nosa saúde, como educamos.... En definitiva:

A Intelixencia Artificial está cambiando o mundo.

Os vehículos autónomos, os robots de servizo e de asistencia (cociña, limpeza, cortacésped, acompañamento...) ou os edificios automatizados (domótica), entre outros, formarán parte da nosa vida cotiá, creando unha rede global de dispositivos interconectados.

Neste escenario, é necesario comezar a introducir a IA nas nosas escolas, para que o noso alumnado non creza alleo a esta realidade senón cunha conciencia crítica e unhas habilidades que lle permitan desenvolverse neste mundo cambiante.

Esta guía busca, precisamente, dar resposta a esta necesidade a través dunha aproximación aos conceptos básicos da IA e ás súas implicacións sociais.



1.1 Que podes esperar desta guía

A presente guía ten como obxectivos:

- Proporcionar ao profesorado contidos e recursos adaptados para traballar a IA na aula de xeito interdisciplinar ou para integralos no currículo de xeito específico.
- Achegar ao profesorado e ao alumnado as aplicacións, os métodos e as implicacións sociais da IA, tal como se contempla no marco europeo "[Plan de Acción de Educación Dixital \(2021-2027\)](#)" onde se establece que *"é cada vez máis necesario que investigadores, axentes educativos e estudantes comprendan os conceptos básicos da IA e o uso dos datos para poder utilizar esta tecnoloxía de forma positiva, crítica e ética e aproveitar todo o seu potencial"*.

- Conseguir que o alumnado adquira competencias específicas resolvendo as actividades propostas, aplicando un enfoque didáctico baseado en “Aprender facendo”.

En definitiva, a intención é que o alumnado coñeza como a IA impacta na vida das persoas, non só en aspectos superficiais, senón que deben **aprender a ser coidadosos e coidadosas como usuarios/as, administradores/as e futuros/as creadores/as** desta potente ferramenta que é a IA.



2. SENTANDO AS BASES

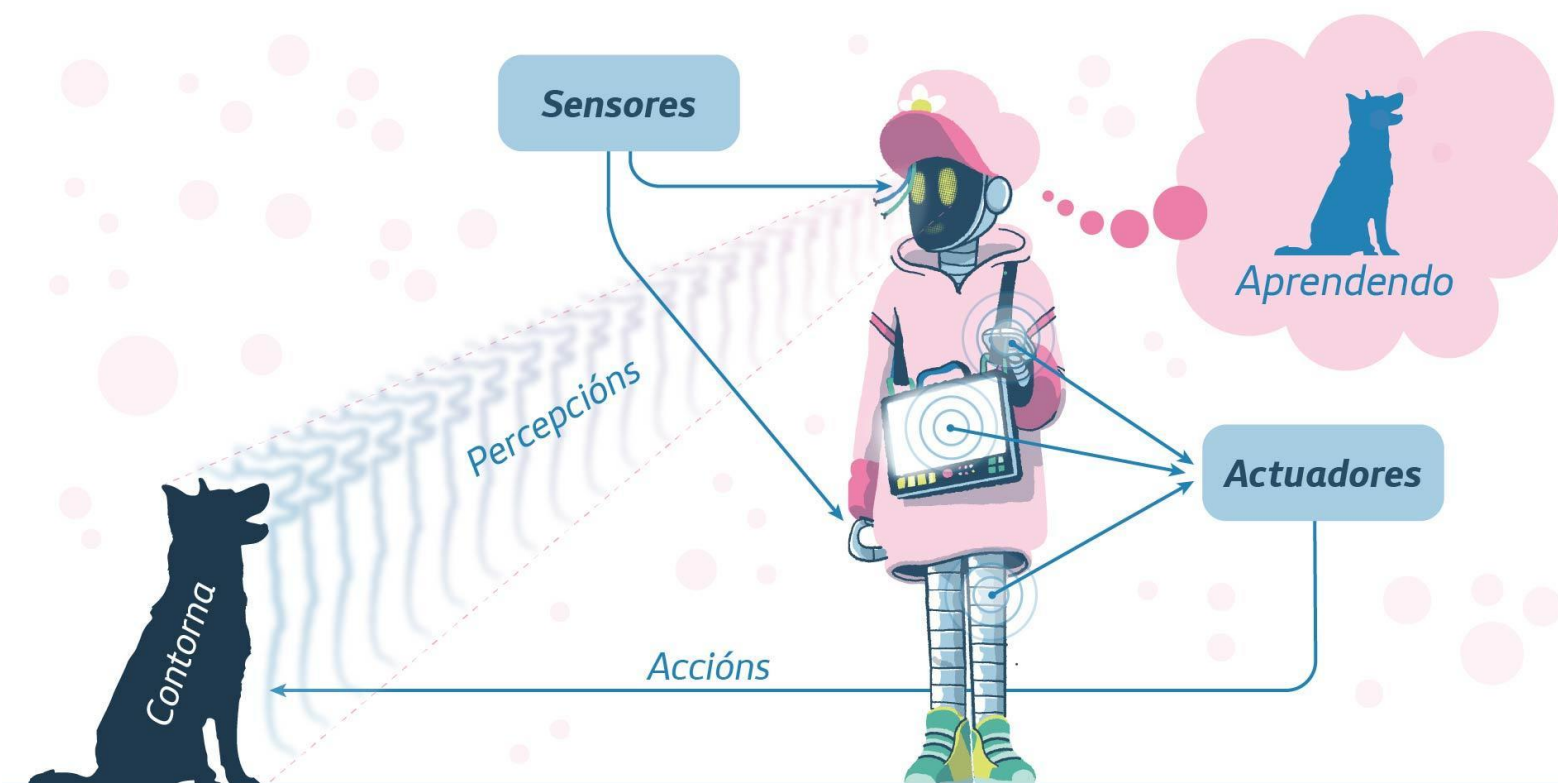
2.1 Que é a Intelixencia Artificial

Aínda que non existe unha definición pechada do que é a Intelixencia Artificial (IA), o alumnado debe coñecer o concepto e ser capaz de explicar a importancia que ten na sociedade actual.

Algunhas ideas clave para definir a IA son:

- A IA é un campo da informática **centrado en dotar ás computadoras da capacidade para resolver problemas de xeito semellante a como o farían os seres humanos.**
- A IA proporciona un sistema artificial, de base computacional (non biolóxica), con capacidade para resolver problemas de forma autónoma, sen intervención humana. O sistema de IA será capaz de aprender da súa experiencia, adaptarse a novas situacións e razoar sobre o mellor xeito de resolver un problema.
- Unha tecnoloxía asistida por IA, tamén chamada *Axente intelixente*, é capaz de realizar tarefas tales como empregar sensores para percibir o mundo que nos rodea, analizar e organizar os datos que recibe e utilizar eses datos de forma autónoma para facer prognósticos ou tomar decisións.

A nosa representación dun axente intelixente será UX1A:



PARA SABER MÁIS: Axentes intelixentes

Un axente intelixente é unha entidade física ou virtual capaz de percibir o que ocorre na súa contorna, procesar o que percibe e responder ou actuar nese escenario de xeito racional.

Os axentes intelixentes están situados en contornas determinadas, que poden ser reais ou virtuais:

- Axentes intelixentes en **contornas reais**: robot aspirador, coche autónomo, cortacésped, robots de acompañamento, Siri, Alexa, Ok Google etc.
- Axentes intelixentes en **contornas virtuais**: recoñecemento facial en xogos de ordenador, recomendacións en función dos gustos nas tendas en rede, nas redes sociais ou nas plataformas de series ou cine en rede etc.

2.2 A Intelixencia Natural fronte á Intelixencia Artificial

A intelixencia é unha característica propia dos seres vivos, especialmente dos seres humanos.

A IA trata de replicar o funcionamento da intelixencia humana, aínda que, polo momento, non é posible replicar o seu órgano básico: o cerebro. Por este motivo, denominamos **Intelixencia Artificial** á tecnoloxía que ten a capacidade de resolver problemas tal e como o farían os seres humanos.

Ao traballar o concepto de IA na aula, será importante que o alumnado analice os niveis de intelixencia de diferentes seres vivos e como estes nos poden servir de referencia para conseguir *intelixencia* nos sistemas artificiais (por exemplo, o modo no que os morcegos utilizan os ultrasóns para orientarse).



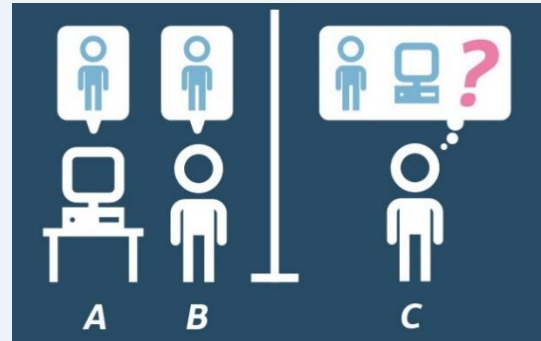
2.3 Historia da Intelixencia Artificial

Consideramos interesante que o alumnado coñeza a historia da IA dende a súa orixe ata a actualidade. Con ese fin elaboramos unha liña do tempo visual para traballar este apartado.



Test de Turing

Alan Turing, matemático considerado un dos pais da informática moderna, deseñou un test para comprobar se un ordenador pode responder de forma similar a como o faría unha persoa: o Test de Turing.



No Test de Turing unha persoa (C) interactúa por escrito con outra persoa (B) e cunha máquina (A) e, en base as súas respostas, determina cal é a persoa e cal o ordenador.



Nacemento da Intelixencia Artificial

John McCarthy citou por primeira vez o termo "Intelixencia Artificial", polo que se considera 1956 como o ano no que se consolida a IA como campo de investigación.



John McCarthy



ELIZA

Eliza foi o primeiro *chatbot* do mundo. Un chatbot é un programa que simula manter unha conversación humana ofrecendo respostas automáticas ás nosas preguntas.

Eliza simulaba ser médica e tiña instrucións para responder con frases modelo a entradas predeterminadas (palabras clave), pero non podía aprender novas respostas polo que chegado un punto a conversa volvíase incoherente.

Conversa cunha versión de [Eliza](#) (podes atopar máis exemplos na rede).



Exemplo dunha conversa con ELIZA.



Inverno da IA

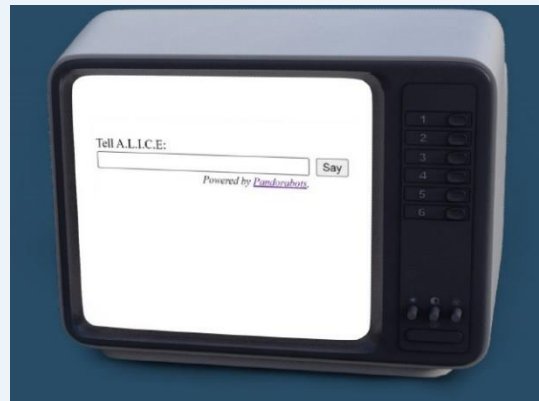
Neste período de tempo non se avanzou no desenvolvemento da Intelixencia Artificial, pois os sistemas eran caros de manter, incapaces de aprender e cometían erros.



A.L.I.C.E.

A.L.I.C.E foi un chatbot capaz de simular unha conversa respondendo ás entradas das persoas usuarias dun xeito moito máis natural que ELIZA, no que está inspirado. O proxecto xurdiu en 1995, pero continuou evolucionando moitos anos despois.

Conversa cunha versión de [Alice](#) (só dispoñible en inglés).



Captura de pantalla dunha versión de A.L.I.C.E.



DEEP BLUE

Deep Blue foi unha computadora que conseguiu gañar ao campión mundial Garry Kasparov nun torneo de xadrez.



Deep Blue no Computer History Museum e Garry Kasparov.



ROOMBA

Roomba foi o primeiro robot doméstico que se comercializou con éxito. Trátase dunha aspiradora intelixente que dispoñía dunha IA moi sinxela que permitía limpar a casa dun xeito eficiente: a aspiradora cambiaba de dirección cando atopaba un obstáculo ou un desnivel.



Roomba de primeira xeración.



STANLEY

Stanley foi o primeiro vehículo autónomo que podía conducir de xeito seguro, rápido e sen axuda humana.

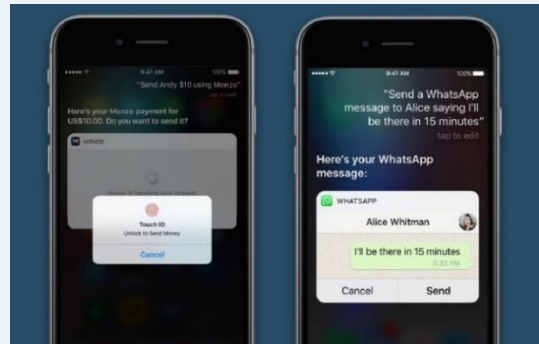


Stanley no DARPA Grand Challenge de 2005, do que resultou vencedor.



SIRI

Siri foi o primeiro asistente virtual integrado en teléfonos móbiles (IOS). Controlado por voz, Siri pode interpretar a linguaxe humana para responder preguntas, facer recomendacións, facer chamadas telefónicas, ler mensaxes de texto etc.



Exemplo de funcionamento de Siri.



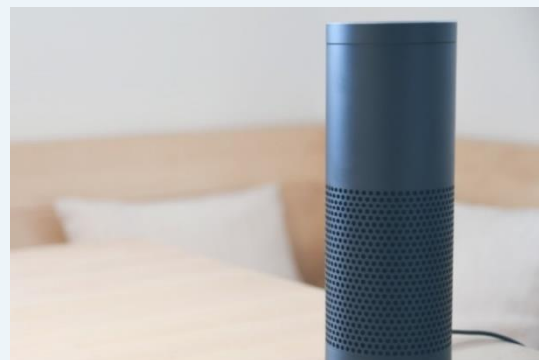
GOOGLE BRAIN

O equipo de Google Brain creou un superordenador capaz de aprender a identificar gatos, caras e corpos humanos en vídeos de YouTube sen necesidade de ensinarlle previamente á máquina como eran un gato ou un ser humano.



ALEXA

Amazon presentou o seu asistente virtual Alexa integrado nun altofalante. Este asistente é capaz de realizar múltiples tarefas como interactuar por voz, reproducir música, facer listas de tarefas, transmitir podcasts e incluso controlar outros dispositivos intelixentes (cámaras, interruptores, interfonos etc.).



Amazon Echo



Unha IA supera o test de Turing

Un chatbot chamado Eugene Goostman foi capaz de enganar ao 33% do xurado cando respondeu ao Test de Turing, facéndolles crer que estaban a falar cun neno ucraino de 13 anos. Converteuse así no primeiro chatbot en superar o Test de Turing.



Captura de pantalla do Chatbot Eugene Goostman.



Tay

Tay foi un chatbot con IA creado por Microsoft para a rede social Twitter que era capaz de aprender a partir da interacción coas persoas usuarias. A compañía tivo que retirar o bot tan só un día despois do seu lanzamento, pois o software volveuse racista, xenófobo e homófobo.



IA FOR GOOD

O obxectivo da iniciativa [AI For Good](#) das Nacións Unidas é identificar solucións prácticas baseadas na IA que poidan dar resposta a problemas humanitarios e axudar a crear un mundo máis sostible e accesible, especialmente a través dos Obxectivos de Desenvolvemento Sostible (ODS).



Logo de IA for Good



Premio Princesa de Asturias á investigación en IA

No 2022 concedese o Premio Princesa de Asturias de Investigación Científica e Técnica a Geoffrey Hinton, Yann LeCun, Yoshua Bengio e Demis Hassabis, persoas expertas en redes neuronais, ferramentas e modelos matemáticos cos que se pretende imitar ou reproducir o funcionamento do cerebro humano nun ordenador.



Yann LeCun, Geoffrey Hinton, Demis Hassabis e Yoshua Bengio.



Na actualidade

Grazas aos datos, aos algoritmos e ao poder de cómputo, a IA na actualidade atópase:

1. Nos asistentes persoais dixitais, tanto en móbiles intelixentes como en ordenadores.
2. No aire acondicionado intelixente.
3. No internet das cousas: aspiradoras, frigoríficos, reloxos etc.
4. Nos vehículos autónomos.
5. Nas compras en liña e na publicidade.
6. Nas buscas na rede e na tradución automática.
7. Na ciberseguridade e na loita contra a desinformación.
8. Na optimización de produtos e procesos de venda.
9. Na agricultura intelixente: rego, alimentación dos animais, robots de control de malas herbas.
10. Nos robots utilizados nas fábricas.



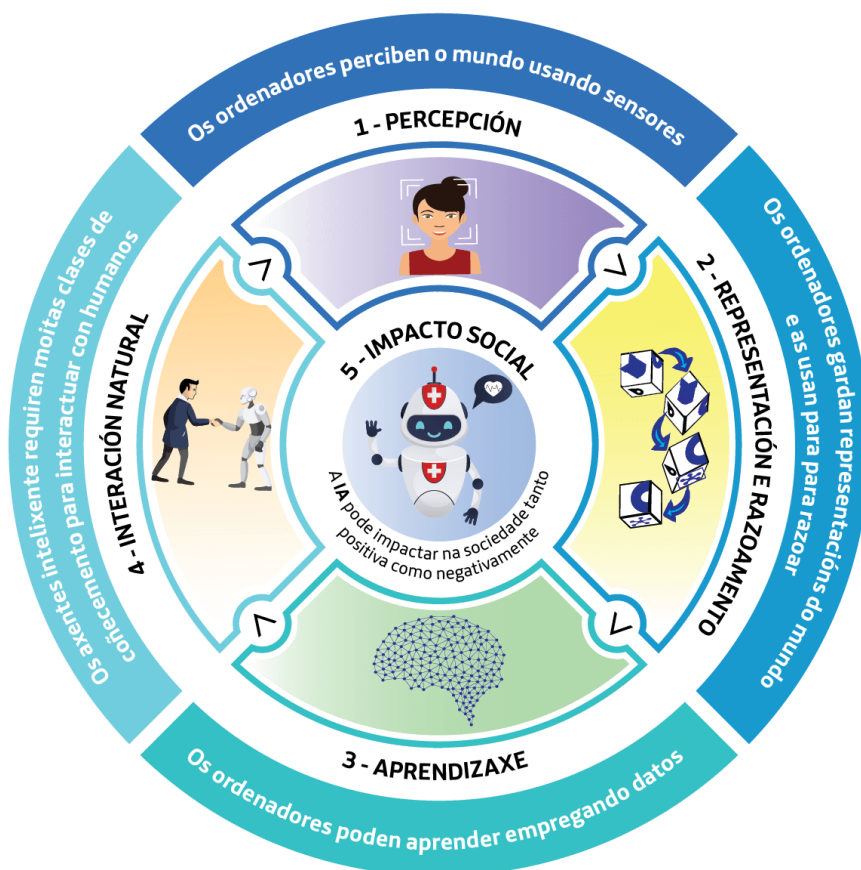
2.4 Intelixencia Artificial na sociedade

Proximamente.

3. BLOQUES BÁSICOS DUN SISTEMA DE INTELIXENCIA ARTIFICIAL

Para introducirmos na aprendizaxe da Intelixencia Artificial (IA) en educación, mergullarémonos nos cinco grandes bloques ou ideas que o alumnado debe aprender dende cuarto de primaria ata bacharelato.¹

Estes cinco bloques achegan unha estrutura de aprendizaxe completa, co obxectivo de proporcionar unha visión específica da IA coma un sistema que interacciona coa súa contorna, xa sexa no mundo real ou virtual.



¹ Estes bloques están baseados na Iniciativa AI4K12, un proxecto da Association for the Advancement of Artificial Intelligence (AAAI) e a Asociación de Profesores de Ciencias da Computación (CSTA).

As cinco grandes accións da nosa Axente intelixente



UX1A percibe



UX1A representa
e razoa



UX1A aprende



UX1A interactúa



UX1A impacta
na sociedade

3.1 UX1A percibe

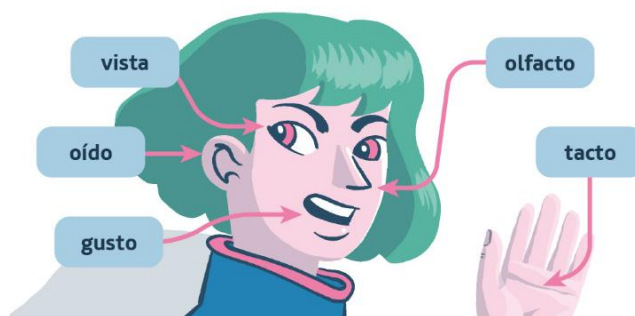


Tanto os seres humanos coma outros seres vivos e os axentes da IA utilizan sensores para recoller datos da súa contorna.

As persoas temos cinco sentidos ou sensores clásicos (oído, vista, olfacto, gusto e tacto) e outros sentidos menos coñecidos pero que tamén son sensores como o sentido do equilibrio ou a sensación de fame.

Os sentidos dan información do que sucede ao noso cerebro e este actúa téndoa en conta. É dicir, os sentidos, que son os nosos sensores, recollen información do exterior e envíanla ao noso cerebro para que a procese.

Sentidos ou sensores clásicos



Igual que o corpo dos seres vivos, a IA utiliza moitos e moi diversos sensores que lle serven para percibir o mundo: dende os máis simples, como un interruptor, ata os máis complexos como as cámaras e dispositivos de comunicación. Antes de mergullarnos nos tipos de sensores que emprega a IA, é importante comprender a relación que existe entre a **sensorización** e a **percepción**:

- A **sensorización** fai referencia á adquisición de datos do mundo físico que nos rodea, como faría un termómetro rexistrando a temperatura do aire unha vez cada hora.
- A **percepción** é o resultado da organización e interpretación das sensacións, por exemplo, como faría a IA integrando os datos de temperatura con outras variables ambientais para facer un prognóstico meteorolóxico. É dicir, a percepción é a extracción de significado da sensorización.

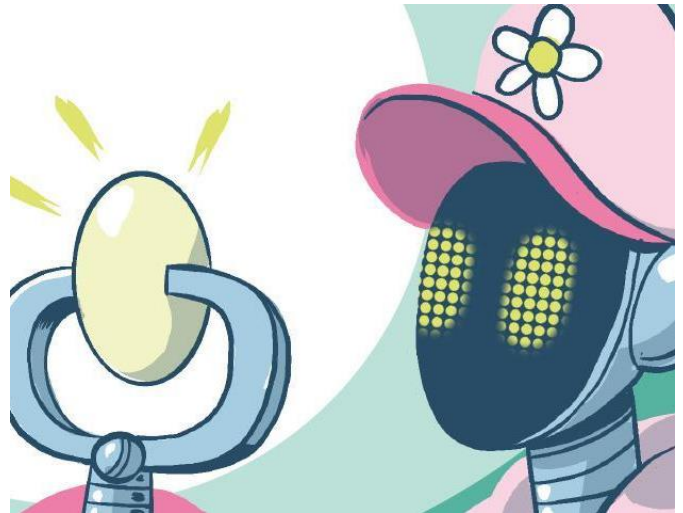
3.1.1 Tipos de sensores da Intelixencia Artificial

Sensores táctiles

Detectan o contacto cun obxecto.

Os máis sinxelos son botóns pulsadores que responden a un contacto directo. Outros máis complexos utilízanse para axudar a determinar se a IA está collendo obxectos brandos, fráxiles ou deformables, ou para determinar se un mecanismo suxeitou un obxecto con éxito ou se é preciso axustar a súa presión.

Este sensor sería moi útil para un brazo robótico con IA, xa que podería distinguir distintos materiais e seleccionar o máis axeitado para realizar unha acción concreta.



Sensores de proximidade

Detectan a presenza de obxectos próximos.

En xeral operan enviando ondas que se reflicten nos obxectos da contorna e que son captadas de novo, de tal xeito que o tempo transcorrido entre o envío e a recepción permite determinar se hai algo nos arredores e algunhas das súas propiedades.

Alguns exemplos deste tipo de sensores son:

- O sonar e o ultrasón: baséanse en ondas sonoras e son empregados por ecógrafos e morcegos.
- O radar: utiliza ondas electromagnéticas e unha das súas aplicacións máis coñecidas son os detectores de velocidade que emprega a policía nas estradas.



Unha aplicación típica destes sensores dentro da IA son os sistemas anticolidión para vehículos autónomos. Os sensores detectan a proximidade dos obxectos e a IA, en virtude da súa experiencia e a súa programación, decide que acción realizar para salvar o obstáculo.

Sensores de temperatura

Detectan a temperatura dun lugar, dun obxecto ou dun ser vivo.

Teñen múltiples funcións dentro dos sistemas de IA. Por exemplo, son esenciais nos sistemas de aforro enerxético en edificios intelixentes e serven para detectar persoas enfermas en grandes grupos humanos en movemento, como se fixo durante a pandemia da COVID-19 en estacións, aeroportos, hospitais, ou centros de traballo.



Sensores de luz

Detectan a presenza, intensidade e cor da luz.

Este tipo de sensores pode integrarse, por exemplo, nun sistema de IA para controlar as horas do amencer e o solpor ao longo do ano e, combinando eses datos coa luz natural que reciba un edificio, facer un uso máis eficiente da enerxía.



Sensores de visión

Detectan imaxes que a IA pode identificar.

A IA ligada á cámara dun teléfono móbil pode identificar se o que imos retratar é un rostro humano, unha mascota, unha paisaxe ou un grupo de persoas. É posible fotografar unha imaxe que conteña caracteres ortográficos e que a IA converta eses trazos nun texto, como fai automaticamente o tradutor de Google se o temos enlazado a nosa cámara.

Outra aplicación que pode ser importante no futuro son as cámaras ligadas aos sistemas de condución autónoma de vehículos que empregan IA. Neste contexto, estes sensores son literalmente os ollos do sistema.



Cómpre mencionar tamén dentro de este apartado o recoñecemento facial, que é unha das grandes revolucións da IA. Esta capacidade permite multitude de aplicacións, dende o etiquetado das fotos en redes sociais ata sofisticadas aplicacións de saúde ou de vixilancia, pasando polas divertidas aplicacións coas que podemos poñernos fociños ou orellas de animais. Con todo, algunhas destas solucións técnicas levan consigo serios riscos para a intimidade das persoas e outros dereitos fundamentais (máis información no apartado *UX1A impacta na sociedade*).

Sensores de localización

Detectan o punto xeográfico no que se atopa un obxecto.

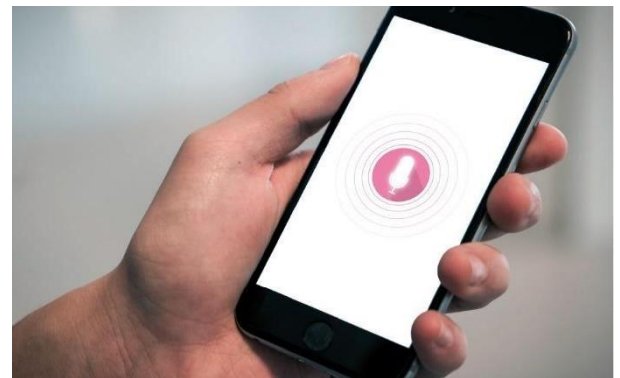
Os sistemas de posicionamento global como o GPS proporcionan as persoas usuarias información sobre a súa localización e facilitan a navegación e o control do tempo. Un dos exemplos típicos de uso do GPS dentro da IA volve a ser a conducción autónoma de vehículos. O coche sen piloto, ademais de marcar a ruta que debe seguir, ten que colocarse de maneira correcta e segura no seu carril, polo que precisa coñecer o estado da estrada en tempo real para garantir unha conducción segura. O GPS garante, xustamente, que o coche vaia polo medio do seu carril, ben colocado na estrada.



Sensores de son (micrófonos)

Detectan sons que a IA pode interpretar.

Os dispositivos que captan sinais acústicos tamén teñen multitude de aplicacións. Dende a comunicación cos nosos asistentes persoais dotados con IA, como *Siri* ou *Alexa*, ata sistemas de detección de ruídos que permiten coñecer se algo vai mal no interior dunha máquina e mesmo predicir que parte se vai estragar e cando está previsto o fallo.

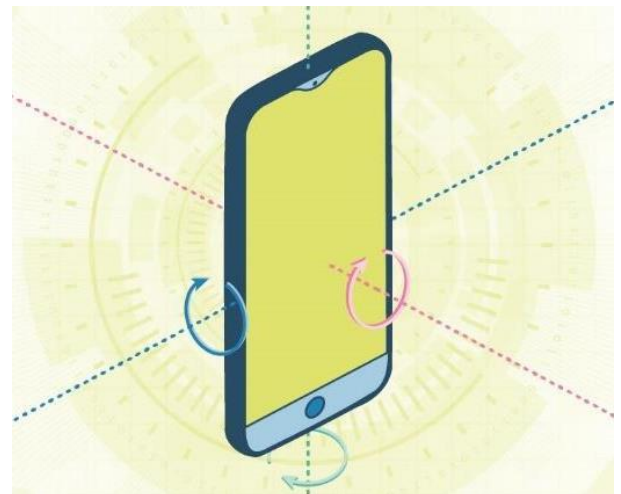


Acelerómetros e xiroscopios

Detectan alteracións de velocidade e de orientación.

O **acelerómetro** é un sensor que mide cambios na velocidade dun obxecto e o **xiroscopio** é outro sensor que mide a súa orientación.

Grazas a estes sensores, o teléfono móbil sabe en que posición está colocado (horizontal ou vertical) e decide como mostra os contidos. Tamén é grazas a estes sensores que o móbil pode utilizarse como unha ferramenta de nivel para saber se quedou ben colgado un quadro ou coma se fora un *joystick* para controlar as accións dentro dun xogo.



Outros sensores

Existen moitos máis sensores que proporcionan datos sobre magnitudes físicas e químicas e que, grazas á IA, poden dar lugar a accións concretas. Por exemplo, os **sensores de gas** detectan a presenza de substancias químicas no aire, e durante a pandemia da COVID fixéronse populares os medidores de CO₂, que permiten avaliar a calidade do aire nunha estancia.



Moitos sistemas de IA precisan de datos procedentes de máis dun sensor para realizar unha **actuación** adecuada e o máis precisa posible. Dependen así dunha **rede de sensores**. Este é o caso dos robots humanoides, como UX1A, que reproducen aspectos do noso comportamento reaccionando as circunstancias da contorna.

3.1.2 A rede de sensores de UX1A

Sensores:

Xeolocalizador (gps)

Sensores de proximidade

Micrófono para detectar os sons

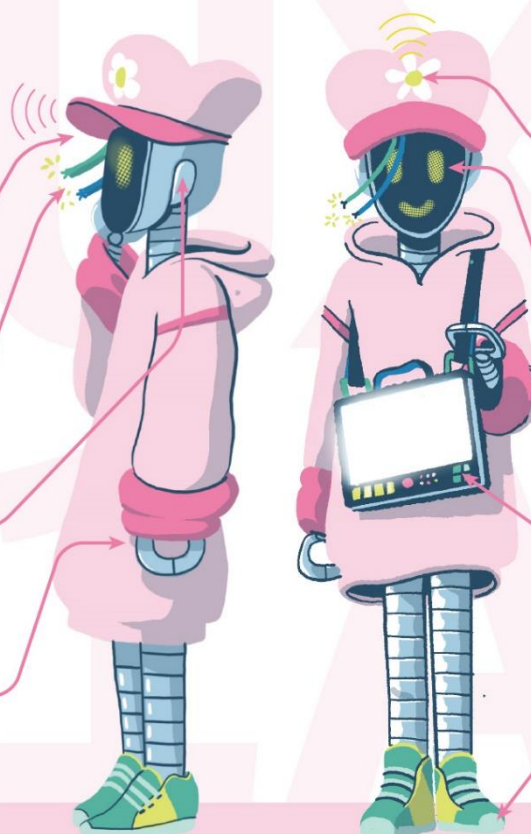
Mans prénsiles con sensores táctiles

Sensor de ultrasóns, para determinar se hai algo diante

Sensores de luz para modificar o brillo da pantalla

Conectores USB, HDMI

Pés con sensores de presión



PARA SABER MÁIS: Sensores

Tipo de sensores	Onde os ten UX1A	Paralelismo co corpo dos seres vivos	Funcións
Sensores táctiles	Mans	Funcionamento do tacto nas mans, dedos.	As mans de UX1A son prénsiles, usa os sensores táctiles para determinar a forza coa que agarra un elemento ou distinguir as propiedades do mesmo.
Cámara	Cara	Ollos.	Ademais de recoñecer a contorna, UX1A ten instalado o recoñecemento facial.
Sensores de son (micrófonos)	Laterais da cabeza	Oídos.	Detectan sinais acústicas, UX1A oe.
Sensores de ultrasón	Cabeza	Sistema que utilizan os morcegos, ECO.	UX1A detecta a presenza de obxectos próximos. Permítelle andar sen tropezar.
Sensores de luz	Corpo	Ollos, pel.	Detectan a presenza, intensidade e cor da luz.
Sensor de localización (GPS)	Corpo	A capacidade que teñen as aves para recoñecer e recorrer rutas migratorias.	UX1A emprégao para non perderse e para axudarnos a non perdernos.

En resumo:

- Tanto os seres vivos como os axentes da IA utilizamos sensores para recoller datos da nosa contorna, que despois procesamos.
- Os *sentidos* son os nosos sensores e os datos que recollemos procésaos o cerebro.
- Coa percepción dámoslle significado á sensorización, é dicir, é a nosa interpretación dos datos que recollemos polos sentidos.
- A IA utiliza moitos e moi diversos sensores que lle serven para percibir o mundo. Algúns son moi simples e outros moi complexos. De feito, moitos sistemas de IA precisan unha rede de sensores para recoller todos os datos que necesitan para actuar correctamente.

3.2 UX1A representa e razoa



3.2a Representación

Para poder empregar os datos que a IA recibe a través dos sensores e actuar a partir deles hai unha serie de pasos intermedios que non poden faltar. O primeiro destes pasos é a representación dos datos, que consiste na súa transformación para que a IA poida comprender a información.

A nós pásanos o mesmo, cando miramos ao noso redor recibimos moita información polos sentidos (sensores) e cada quen é capaz de explicar o que ve: *dende a nosa ventá vense un invernadoiro, un parque con carballos e un río*. De seguro, temos a sensación de estar “vendo” iso na nosa mente. E aínda máis, podemos completar esa imaxe que vemos con outras cousas que “sabemos” que están aí, aínda que non as vexamos neste momento: *no parque haberá bancos, árbores e xente paseando*.

En realidade, o que estamos facendo é **crear unha representación** do mundo a partir da información que percibimos polos sentidos (neste caso, o que estamos vendo) e doutros datos que temos *almacenados*, que coñecemos (como é un parque).

Esa representación é o primeiro paso para facer un razoamento e crear coñecemento a partir dos datos que temos á nosa disposición (*se quero atravesar o parque teño que desviarme á esquerda, porque senón chocarei co invernadoiro*). É, polo tanto, unha forma de ordenar esa información para poder, despois, comprendela.



Os axentes intelixentes simulan este proceso: usan datos, que son anacos de información que recollen de maneiras moi diferentes a través dos sensores, para crear representacións que poidan *comprender*. Os datos que recollen poden ser palabras, números, imaxes...



3.2a.1 Procesar ou tratar os datos

Cómpre ter en conta que un mesmo conxunto de datos (o que vemos pola ventá) pode representarse de moitas formas diferentes. Por exemplo, poderíamos describilo en dous idiomas distintos: galego e inglés. Se unha persoa de fala inglesa le a nosa descrición do que vemos pola fiestra, non poderá comprender o que lle estamos dicindo, terá que traducilo.

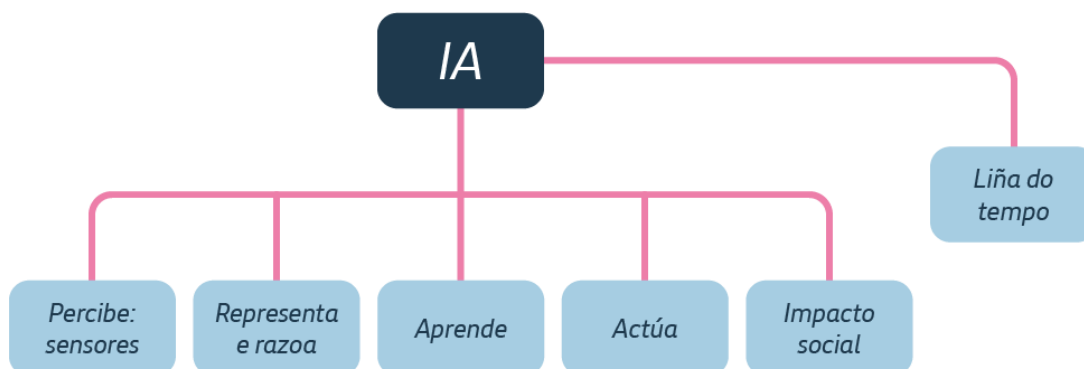
Deste xeito, para poder realizar unha tarefa a partir de determinada información, o que precisamos é, simplemente, dar coa representación adecuada, é dicir, cambiar a forma de representala para que sirva aos nosos propósitos.

A isto se lle chama **procesar ou tratar os datos**. Hai diferentes maneiras de facelo: mediante bases de datos (táboas), grafos, cambios de coordenadas etc. e en cada unha se lle da máis importancia a unha parte da información ou a outra.

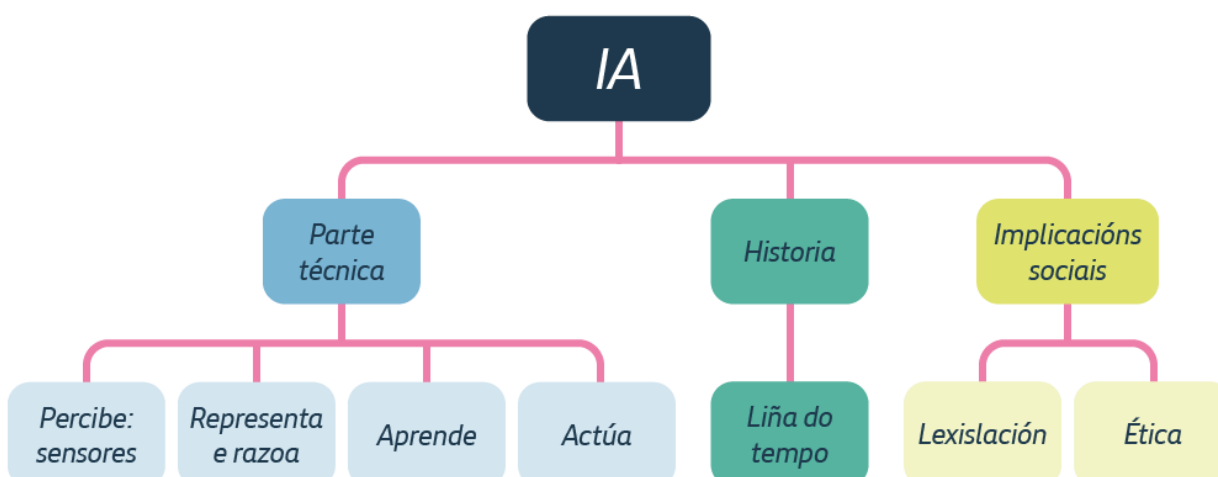
Por exemplo, para representar as relacións existentes entre distintos elementos podemos realizar **un grafo**, no que consideraremos os **nodos** (elementos) e os **camiños** (relacións).

Así, esta guía podería representarse do seguinte xeito:

Representación 1



Representación 2



PARA SABER MÁIS: Información e representación axeitadas

É preciso aprender que representacións de datos nos achegan ao resultado que esperamos, entendendo que unha representación será unha maneira concreta e diferente a outras de mirar os datos, presentalos ou codificalos.

A continuación recóllense algúns exemplos de cambios na representación de datos para conseguir certos obxectivos:

Modificar unha imaxe

Unha imaxe en cor pode representarse de dous modos:

RGB: píxeles en cor vermella (red), verde (green) e azul (blue).

HSV: segundo o seu matiz-saturación-valor.

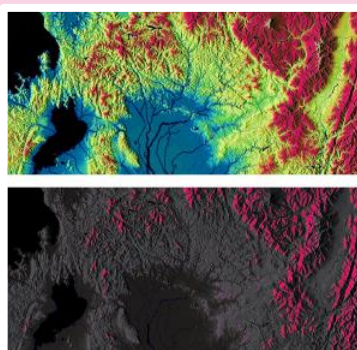
Se nunha imaxe queremos eliminar os píxeles vermellos usaremos a representación RGB, pero se o que queremos é conseguir unha imaxe máis saturada usaremos o modo HSV.



Mellorar o sistema de prevención de incendios

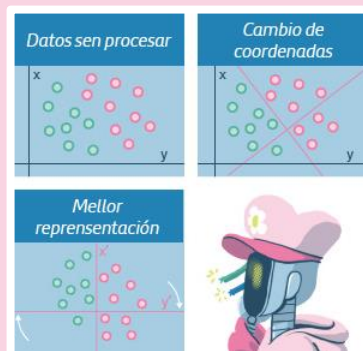
Se o noso obxectivo é predicir en que zona dun monte será máis difícil apagar un lume forestal usaremos unha imaxe de satélite onde a cor vermella dea información sobre a altitude do terreo, a cor verde sobre a inclinación e a cor azul sobre a orientación do terreo.

Para distinguir as zonas de maior altitude, que previsiblemente serán as de máis difícil acceso para o control do lume, escolleremos os píxeles "mais" vermellos. Polo tanto, a representación RGB será a máis útil cumprir o noso obxectivo.



Clasificar puntos pola súa cor

Neste exemplo buscaremos un cambio de representación de datos que facilite a tarefa de clasificar puntos pola súa cor. O conxunto de datos a procesar son puntos verdes e morados situados nuns eixos de coordenadas. Con esta representación é difícil dicir se a cor dun punto depende da súa posición. Pero se cambiamos as coordenadas vemos que, con moita probabilidade, se o punto está a esquerda do eixo vertical y' será verde e se está á dereita será morado. Deste xeito conseguimos unha representación que nos facilita o obxectivo de clasificar os datos.



En resumo:

- O mundo está cheo de información.
- Necesitamos representala para traballar con ela.
- Non hai unha única maneira de representala, debemos buscar a forma máis adecuada para cada obxectivo.
- Debemos, por tanto, representar a información a través do procesamento ou tratamento de datos.

3.2b. Razoamento

Dende o inicio da IA soñouse con máquinas capaces de “pensar”, de razoar, como o facemos nós. Por iso, cando no ano 1997 unha máquina chamada Deep Blue (sistema experto) foi capaz de gañar a Kasparov, naquel momento campión de xadrez, xerouse unha gran conmoción.

Parecía un fito moi importante para acadar a “intelixencia” nas máquinas, pero pronto se descubriu que algo que é complexo para o ser humano pode ser moi fácil para unha computadora, e ao revés.

O xadrez réxese por unhas normas moi estritas e un entorno moi delimitado e perfectamente especificado, o cal é ideal para unha computadora. Pola contra, identificar a alguén pola rúa, como fai un neno de catro anos (ao que seguramente lle custaría xogar o xadrez pero que reconece a súa avoa entre unha multitude) é unha tarefa ben complexa para un ordenador. Nese caso, non é sinxelo poñer normas nin delimitar os conceptos involucrados no escenario.

Cómpre recordar que, como xa mencionamos, aínda non se conseguiu replicar o órgano básico da intelixencia, que tamén é o que razoa: o cerebro. Por iso, cando falamos de razoamento artificial nos referimos a un pseudo “pensamento” levado a cabo polos axentes intelixentes, de acordo cunha programación, un deseño, feito por seres humanos. Ese é o motivo polo que, para un ordenador, será máis sinxelo gañar ao xadrez que recoñecer caras, porque é máis fácil darlle as regras para facelo.

Para que un axente intelixente poida *razoar* a forma de resolver un problema necesitamos “formalizar” dito problema.

Formalizar quiere dicir atender unicamente a forma do razoamento, prescindindo do seu contido.

É dicir, consiste en “traducir” ou mostrar a estrutura interna dun problema, identificando os seus elementos e as conexións que existen entre eles e outorgándolle unha forma concreta para comprender como funciona. Unha vez identificada esa estrutura, unha persoa ou unha computadora poderán empregala como modelo para dar resposta a problemas similares, aplicando sempre as mesmas regras formais: dadas unhas condicións iniciais poderá determinar o resultado que se obtén.

Por exemplo, estas frases teñen unha forma similar, aínda que o seu significado sexa diferente, polo que seguramente podemos rematar a segunda tomando como modelo a primeira:

- Os gatos son encantadores e Tip é un gato, polo tanto Tip é encantador.
- As mazás son froita e as froitas son saudables, polo tanto...

O que fixemos foi *formalizar* o enunciado:

- Condición 1: un conxunto (gatos/froitas) ten unha propiedade (encantadores/saudables).
- Condición 2: un elemento (Tip/Mazás) forma parte do conxunto.
- Resultado: o elemento ten a propiedade do conxunto.

Deste xeito, para formalizar necesitamos crear regras a partir da información do problema que queremos resolver. E, a través dun programa que utilice esas regras e datos, o ordenador daranos un resultado.



É dicir, para que un axente intelixente poida “razoar”, necesitamos establecerlle a forma de facelo. Para facer isto o primeiro que debemos saber é que un ordenador está composto de circuítos eléctricos e electrónicos. Estes circuítos só teñen dous valores posibles: funcionando ou non funcionando. É o único que a máquina entende.

Estes estados represéntanse con 0s e 1s:

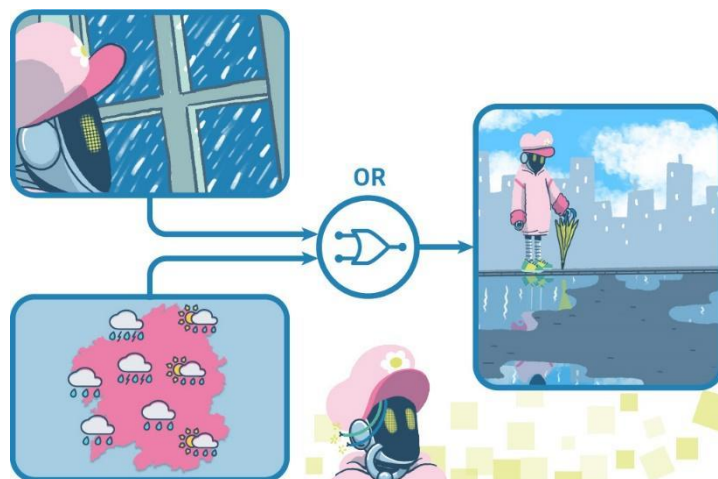
- 0: apagado, baixo (low), 0 voltios.
- 1: aceso, alto (high), 5 voltios.

Para operar cos 0s e 1s temos que usar a lóxica formal, que nos di mediante regras fixas se algo é verdadeiro (1) ou falso (0). Usando 0s e 1s con operacións lóxicas conseguimos que os ordenadores teñan todas as aplicacións que coñecemos.

3.2b.1 Operacións lóxicas básicas

- **"ou (or)" (suma lóxica)**

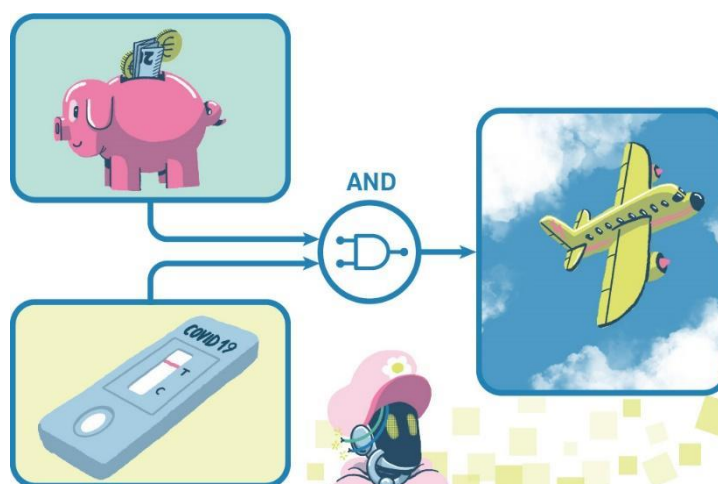
Se temos varias entradas e polo menos unha delas é verdadeira (ten valor 1) a función nos dirá que a saída é verdadeira (1).



Exemplo: se digo *saco o paraugas de casa se Meteoroloxía di que chove* "ou" se vexo que chove pola ventá, sacarei o paraugas no caso de que suceda algunha das dúas cousas (polo menos unha verdadeira).

- **"e (and)" (multiplicación lóxica)**

Se temos varias entradas todas elas terán que ser verdadeiras (ter valor 1) para que a función nos diga que a saída é verdadeira (1).



Exemplo: se digo *vou ir de viaxe se teño diñeiro* "e" se a Covid mo permite, isto implica que só viaxarei se ambas son verdadeiras, no caso de que algunha sexa falsa, non viaxarei.

PARA SABER MÁIS: Táboas da verdade das funcións “ou”, “e”, “non”

Para representar as funcións lóxicas usamos unha ferramenta que se chama táboa da verdade. Nela aparecen todas as posibles combinacións de entradas dunha función e as súas saídas correspondentes. Cada variable de entrada pode ter dous valores, 0 ou 1, polo que se temos dúas variables imos ter catro combinacións: 0,0; 0,1; 1,0; 1,1.

Función OR

Pensemos nun circuío eléctrico formado por dous interruptores, unha lámpada e unha pila na que calquera interruptor acende a lámpada:

- Cando non prememos ningún interruptor (ambas entradas son 0) a lámpada non ilumina (a súa saída é 0 tamén).
- No caso de que premamos algún interruptor (unha das entradas é 1) a lámpada ilumina.
- E se prememos os dous interruptores (ambas entradas son 1) a lámpada tamén ilumina.

Iso escribímolo na táboa como as entradas e as saídas correspondentes á función OR:

FUNCIÓN OR		
Entradas (interruptores)		Saída (lámpada)
Interruptor A	Interruptor B	
0	0	0
0	1	1
1	0	1
1	1	1

Función AND

Neste caso pensemos nun circuío similar formado por dous interruptores, unha lámpada e unha pila pero no que necesitamos premer ambos interruptores para que a lámpada funcione. A táboa da verdade da función AND sería:

FUNCIÓN AND		
Entradas (interruptores)		Saída (lámpada)
Interruptor A	Interruptor B	
0	0	0
0	1	0
1	0	0
1	1	1

Función NOT

A terceira función lóxica básica é a función NOT ou negación. Esta función dá como saída o valor contrario ao dunha variable. Así:

- Se a variable vale 0 o seu valor “negado” (saída) é o 1.
- Se ten como valor 1 o valor de saída da función será 0.

Pensaremos agora nun circuío formado só por un *pulsador* e unha lámpada que se acende cando o *pulsador* está aberto e se apaga cando está pechado:

FUNCIÓN NOT	
Entrada (<i>pulsador</i>)	Saída (lámpada)
0	1
1	0

EXERCICIOS: Codificación binaria

Proximamente.

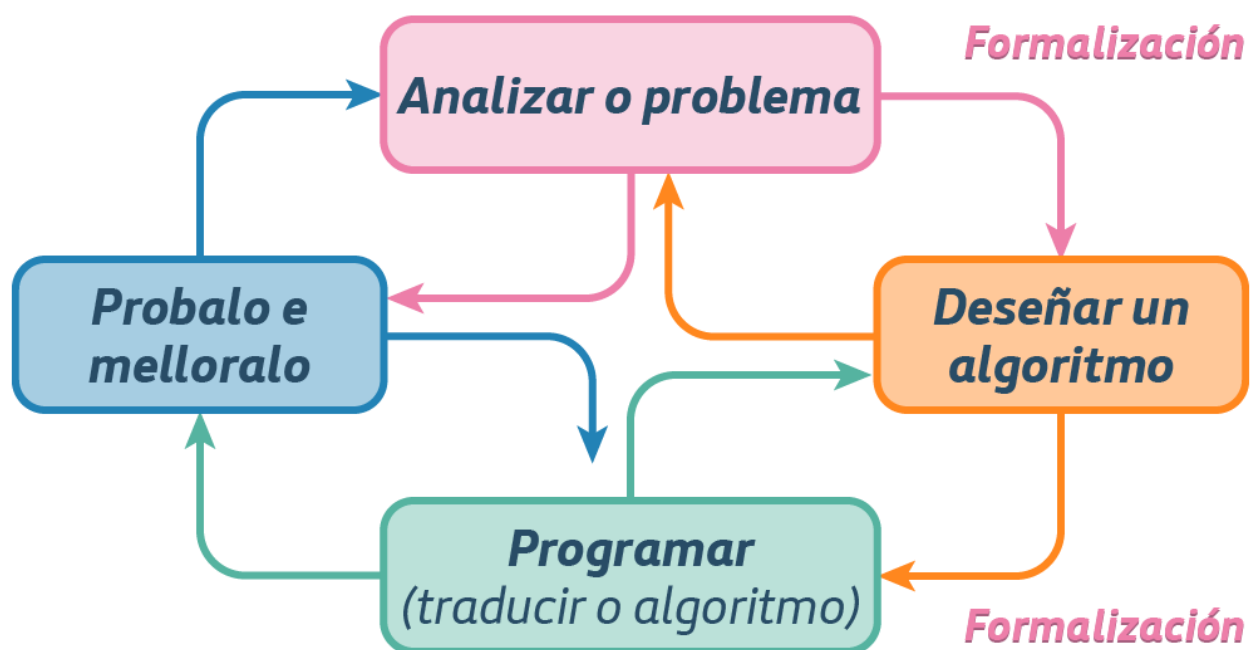
3.2b.2 Programación e algoritmos

Como para nós traballar só con 0s e 1s é difícil, creamos unhas linguaxes chamadas *de programación* que se parecen máis ao idioma que usamos habitualmente e que as propias máquinas son capaces de transformar en 0s e 1s. Nesas linguaxes creamos os programas ou aplicacións.

Para crear un programa para un ordenador necesitamos coñecer o algoritmo que nos vai permitir resolver o problema.

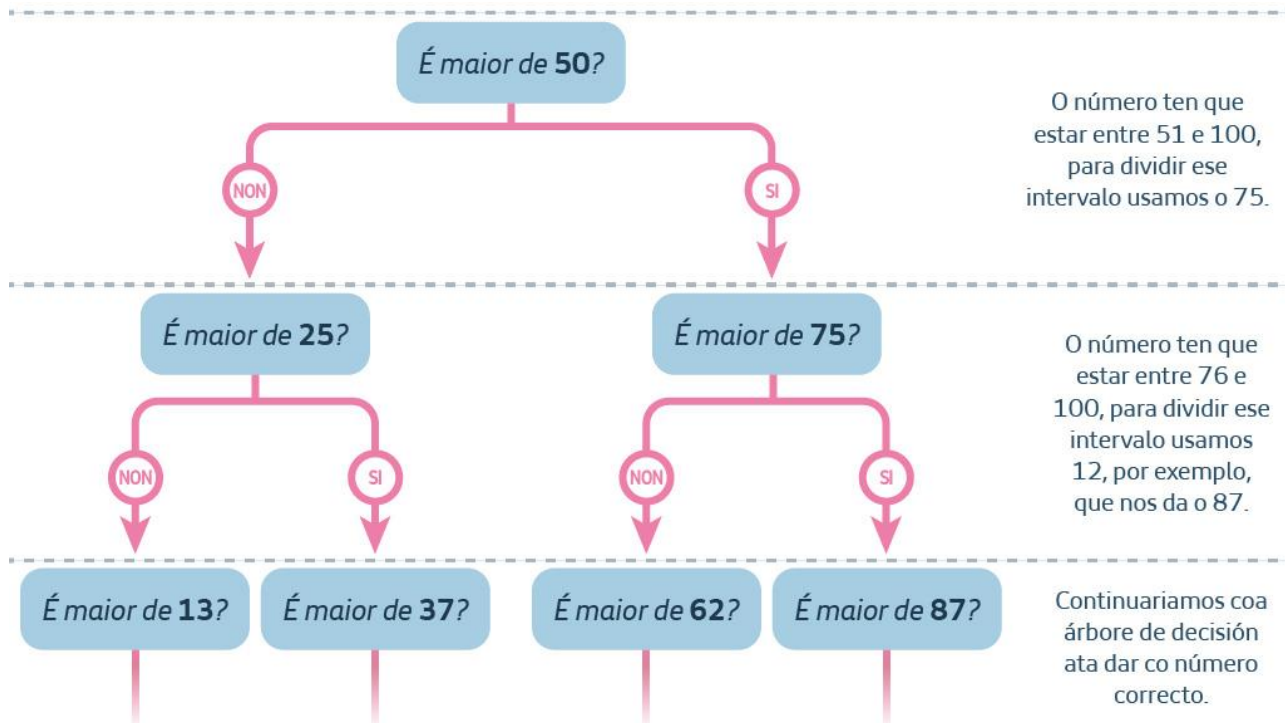
Un algoritmo é un conxunto de instrucións ordenadas e finitas que nos permite resolver un problema, partindo dunha entrada e dando un resultado. Unha receita é un algoritmo, os ingredientes serán a entrada e o prato será a saída ou resultado.

Construímos o algoritmo a partir da formalización do problema, é dicir, tras coñecer a súa estrutura expresámola ou representámola nun algoritmo. Logo, este algoritmo convértese en instrucións para a máquina utilizando a linguaxe de programación. A isto se lle chama **programar**.



Un algoritmo particularmente útil en IA é a “arbore de decisións”, que serve para tomar decisións, clasificar ou facer predicións a partir dunha serie de preguntas que se responden con *si* ou *non* de forma sucesiva.

Un exemplo trasladado ás persoas será a técnica que aplicamos no xogo de adiviñar un número entre o 1 e 100 en dez preguntas con respostas de si ou non. Poderíamos empezar preguntando se o número é maior de 50, e seguir despois segundo as respostas:



Este tipo de algoritmo tamén nos pode axudar a tomar unha decisión sobre se mercar ou non algunha cousa tendo en conta diferentes datos, por exemplo: o diñeiro que teño, o que quero gastar, o que valen as diferentes cousas etc.

A aplicación en rede [AKINATOR](#) é tamén un exemplo de árbore de decisión.

EXERCICIOS: Algoritmos

Proximamente.

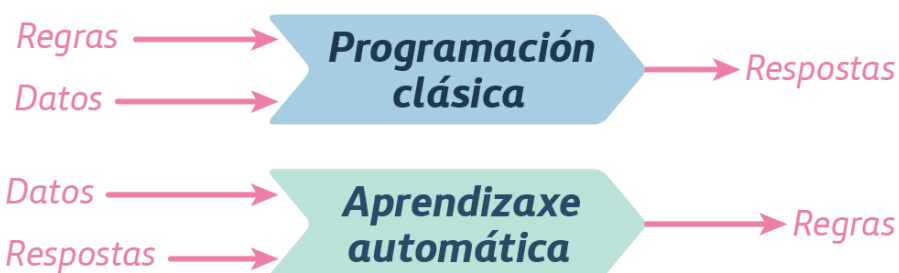
En resumo:

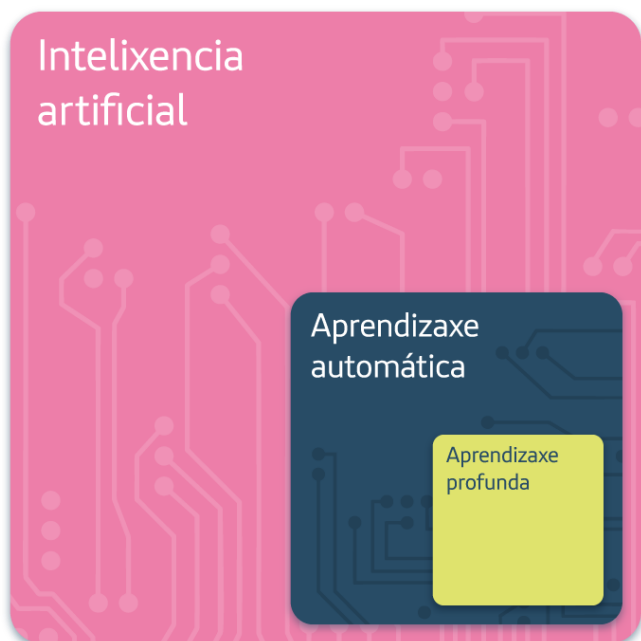
- Para que un axente intelixente poida razoar debemos indicarlle como facelo a través da programación.
- O único que as máquinas entenden son 1s e 0s e operacións lóxicas, que nós escribimos con linguaxes de programación.
- Para programar unha máquina temos que formalizar o problema, construír un algoritmo e traducilo en instrucións que lle permitan resolvelo.
- Polo tanto, a **programación** é un conxunto de regras e datos dos que obtemos respostas. Así, basicamente, é como funciona Deep Blue, o programa que gañou a Kasparov. Pero, se o que queremos resolver é o recoñecemento de caras, esta forma de programar non nos serve xa que se trata dun problema moi difícil de **formalizar**.

3.3 UX1A aprende



Para resolver problemas difíciles de formalizar, pasouse da maneira “tradicional” de programación a outra chamada **aprendizaxe automática** (*machine learning*). Mentres na programación clásica lle dabamos regras e datos á máquina para obter respostas, neste novo tipo de programación subministrámoslle datos e respostas para obter as regras que lle permiten discriminar se un dato novo é ou non unha resposta válida. Isto fai posible que as máquinas *aprendan* de xeito similar a como o facemos nós e que poidan levar a cabo accións tan complexas como recoñecer unha cara concreta nunha multitude.





Cómpre sinalar que a **aprendizaxe automática** (*machine learning*) é unha parte da IA dentro da cal atopamos a **aprendizaxe profunda** (*deep learning*), que é a responsable do éxito actual da IA. Os sistemas de recoñecementos de caras, de discursos, de transcripción de escritura manuscrita, os asistentes como *Google*, *Siri*, *Alexa*, a capacidade para responder preguntas de linguaxe natural etc. funcionan grazas a este tipo de aprendizaxe.

3.3.1 Aprendizaxe automática

É un método de IA que trata de facer que as computadoras recoñezan patróns nos datos e tomen decisións sen ser explicitamente programadas para iso. Permite que as máquinas se adapten a novas situacións de forma independente, cambiando cando se expoñen a novos datos, é dicir *aprendendo* da experiencia.

A aprendizaxe automática trata de replicar unha forma de aprendizaxe humana no ordenador: aprender a partir de moita práctica e de probas e erros. Esta aprendizaxe humana baseada na experiencia trasládase ás máquinas achegándolles gran cantidade de datos (imaxes, vídeos, textos, sons etc.) a partir dos cales o ordenador debe recoñecer unha serie de patróns. Con esta información (esta *aprendizaxe da experiencia previa cos datos*) o ordenador poderá facer predicións.

Desta forma, a **aprendizaxe automática** é un método de IA que consegue, a partir da entrada/observación de moitos datos e da execución dun algoritmo, xerar un modelo e utilízalo para facer predicións que lle permiten resolver un problema ou tomar decisións.

A *aprendizaxe* é o proceso polo que se xera este modelo, que deberá ser adestrado con máis cantidade de datos para gañar precisión.



Deste xeito, o axente intelixente recibe entradas (datos) a través dos sensores, representa e procesa eses datos cos algoritmos, dá repostas creando modelos e finalmente interactúa coa contorna en función desas repostas.

3.3.2 Tipos de aprendizaxe automática

A aprendizaxe automática pode ser de tres tipos:

1 Supervisada

O obxectivo do algoritmo está definido de antemán, por exemplo: recoñecer gatos. Para poder cumprir con este obxectivo, o modelo recibe moitos datos e aprende as características claves deses datos, no noso exemplo aprende as características dos gatos. Unha vez adestrado, o modelo xerado debería clasificar ou predicir o resultado correcto cando se lle subministren datos novos, é dicir, debería poder indicar correctamente se unha imaxe nova é dun gato ou non.

2 Non supervisada

O obxectivo do algoritmo é buscar patróns de similitudes e diferencias entre un conxunto de datos para, a partir destes patróns, dar respostas. Por exemplo, un algoritmo pode agrupar clientes/as por compras similares, para suxerirllle despois a cada quen outros produtos que comprou a clientela do grupo ao que pertence. Así, se vemos unha serie en Netflix o seu algoritmo buscará a outras persoas que viron esa serie e nos suxerirá outras series vistas por esas persoas.

3 Aprendizaxe por reforzo

Intenta facer que a máquina aprenda directamente da súa toma de decisións a través dun sistema de recompensas. Desta maneira, o axente intelixente non vai clasificar ou agrupar datos, senón que toma decisións que impactan no ambiente no que se atopa e *aprende* das consecuencias: dos cambios que percibe no ambiente e da recompensa que obtén. Por exemplo: se o obxectivo é que un robot aprenda a meter goles, en lugar de darlle as instrucións paso a paso do que ten que facer, déixase que tome decisións "a cegas" e dáselle unha recompensa como resultado a cada patada que dea. Cando a recompensa é positiva refórzase o seu comportamento (ese tipo de patada) e cando é negativa incentívase que actúe de forma diferente no futuro (outro tipo de patada).

3.3.3 Redes neuronais e aprendizaxe profunda

As redes neuronais son un tipo de algoritmo (usado nas aprendizaxes supervisadas e non supervisadas) que intenta simular o que se cre que é o funcionamento do noso cerebro.

Nunha rede neuronal temos un conxunto de "neuronas" conectadas entre si que poden *aprender* a executar tarefas. A rede neuronal é capaz de aprender, co tempo, a mellorar na tarefa que ten asignada sen necesidade de que as persoas interveñan no proceso. "**Neuronas**" é unha forma romántica de denominar as **funcións matemáticas** involucradas nestes algoritmos, as cales procesan os datos que reciben para ofrecer un novo dato de saída.

As neuronas dunha rede agrúpanse en forma de capas. Temos tres tipos de capas:

- **Capa de entrada**

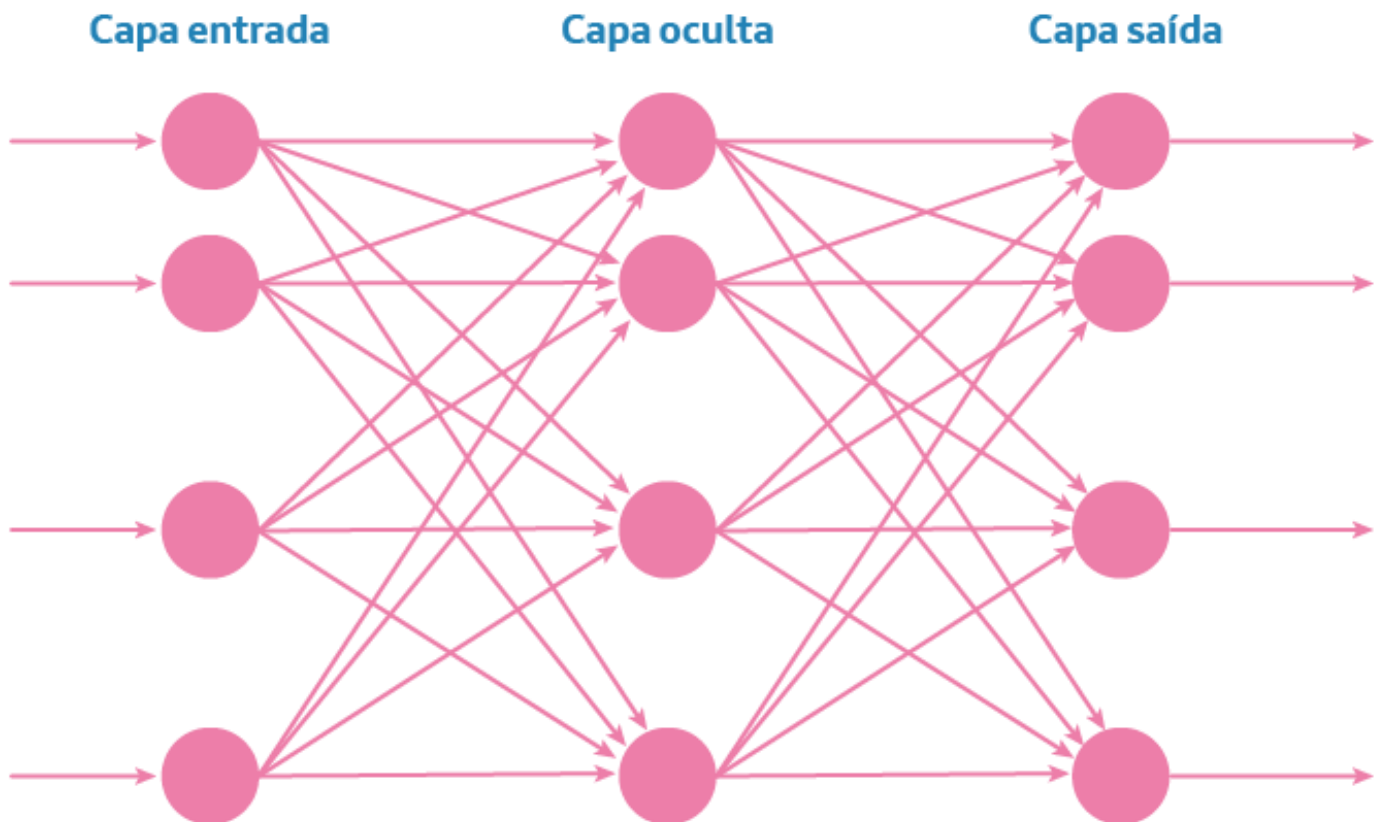
É o conxunto de neuronas que reciben a información da realidade a través dos sensores e/ou dos datos.

- **Capa de saída**

É o conxunto de neuronas que dan unha saída que vai interactuar coa contorna mediante os actuadores (ver apartado "UX1A interactúa").

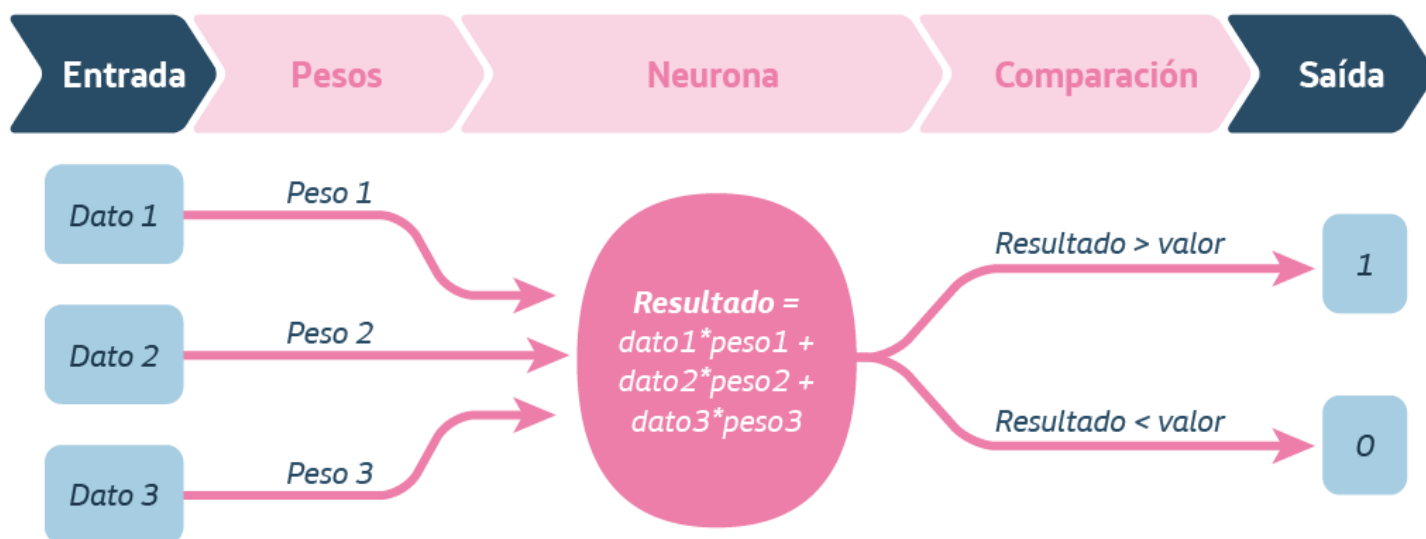
- **Capas ocultas**

Son os conxuntos de neuronas que están entre a capa de entrada e saída. Estas neuronas non interactúan coa contorna, senón que reciben información doutras neuronas da capa anterior e, tras procesala, facilitan información ás neuronas da capa seguinte.

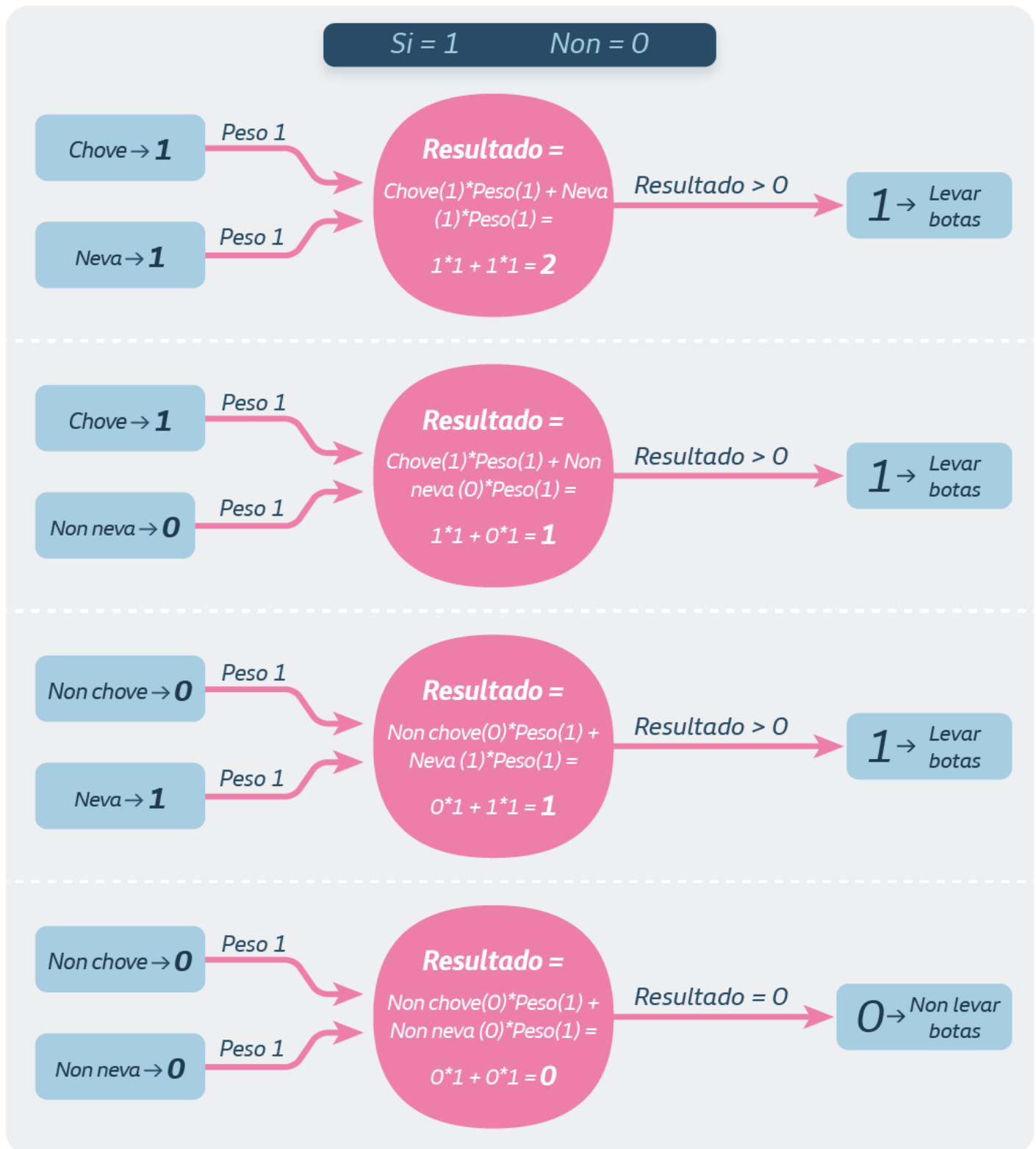


Imos ver en detalle como funciona unha rede neuronal chamada perceptron, que foi a primeira rede que permitiu identificar patróns:

- A cada neurona chegan un conxunto de datos de entrada que a neurona procesa para dar un novo dato de saída.
- Os datos de entrada dunha neurona poden ter diferente influencia na súa saída (estar ponderados), polo que se di que cada dato de entrada ten un *peso* propio. O peso dun dato pode ser maior ou menor que cero.
- A neurona multiplica cada dato de entrada polo seu peso e despois súmaos todos.
- Por último, compara o resultado cun "**valor**" chamado *limiar da neurona* (valor de partida) a partir do cal a saída da neurona será positiva.
- Se o resultado da suma dos datos multiplicados polo seu peso é maior que o valor limiar da neurona, a saída da neurona será un 1 e se é menor será un 0.



No seguinte exemplo, a neurona ten dous datos de entrada (chove/non chove e neva/non neva) co mesmo peso (1) e un valor limiar de 0. Cando polo menos un dos datos é igual a 1 (ben chove, ben neva) a neurona dará como resultado a necesidade de levar botas:



Na rede neuronal o “valor” de saída dunha neurona (o dato de saída) pode utilizarse como dato de entrada para outras neuronas ou constituír o resultado final da rede neuronal.

Como o resultado de cada neurona da rede depende dos pesos dos datos de entrada, o comportamento dunha rede neuronal pode ser alterado axustando os seus pesos.

Un exemplo de como funciona unha rede neuronal trasladado ás persoas pode ser o proceso para resolver o actual xogo en liña *Wordle*², no que temos que adiviñar unha palabra de 5 letras que cambia cada día:

1. Escribimos unha palabra aleatoria.
2. xogo indícanos, para cada letra, se esta forma parte ou non da palabra a adiviñar e (cando forma parte) se está colocada de forma correcta ou non.
3. Segundo o que nos vai dicindo o xogo, nós imos axustando o *peso* das letras en cada posición (a probabilidade que cremos que hai de que unha letra estea nunha posición) ata que acadamos a resposta.

Nun inicio, todas as letras teñen a mesma probabilidade de formar parte da palabra, pero a medida que vas introducindo palabras vas obtendo a información que che permite comprender o peso de certas letras. Ao repetir esta acción vas axustando os pesos ata adiviñar a palabra.

O xogo de mesa *Mastermind* segue tamén estes principios, pero empregando combinacións de cores en lugar de palabras.

EXERCICIOS: Redes neuronais

Proximamente.

3.3.4 Redes neuronais empregadas na IA

As **redes neuronais** organízanse en moitas capas, algunhas usan centos de miles de neuronas e máis de 100 capas.

Establecer a man os pesos destas redes sería moi difícil. Afortunadamente, non temos que facelo. Usando unha técnica chamada **aprendizaxe profunda** (deep learning) podemos conseguir que o ordenador descubra os pesos necesarios para que unha rede neuronal produza a saída correcta para unha entrada determinada.

Para conseguir isto o primeiro que temos que facer é proporcionarlle á máquina unha colección de entradas (datos) de mostra e dicirlle cal é a resposta correcta para cada unha. Por exemplo, se queremos que recoñeza

² O xogo está dispoñible en múltiples linguas, aquí tes o enlace á [versión Wordle en Galego](#).

a expresión da alegría ensinaremoslle moitas fotografías de caras distintas e indicaremoslle cales representan a alegría. A rede irá identificando patróns neses datos para *aprender* cales son as respostas correctas. Cando a rede produce respostas incorrectas para algunhas das entradas (non recoñece a alegría nunha cara que si a expresa), o algoritmo de aprendizaxe automática utiliza os nosos comentarios sobre a súa precisión para facer pequenos axustes nos pesos que axudarán a que a rede funcione mellor a próxima vez. Este proceso denomínase **adestramento**.

Unha vez rematada a fase de adestramento, a rede neuronal poderá tomar datos novos e facer predicións acertadas sobre cal é a resposta correcta para eses datos. No noso exemplo, a rede será capaz de identificar a alegría en novas imaxes, distintas das que lle ensinamos no adestramento.

Para rematar, cómpre aclarar que na aprendizaxe profunda, que é un tipo específico de aprendizaxe automática, é a propia rede neuronal a que vai aprendendo dos resultados que obtén sen intervención humana, mentres que na aprendizaxe automática en xeral si hai participación humana (podemos facer axustes da súa precisión revisando os seus resultados).

En resumo:

- A aprendizaxe automática (machine learning) desenvolveuse para resolver problemas difíciles de formalizar, como o recoñecemento de caras.
- A aprendizaxe automática é unha parte da IA e a aprendizaxe profunda é unha parte da aprendizaxe automática.
- Coa aprendizaxe automática as computadoras conseguen atopar patróns nos datos e xerar modelos para facer predicións coas que tomar decisións ou resolver problemas.
- Unha rede neuronal simula o que se cre que é o funcionamento das neuronas do noso cerebro. A través do seu adestramento, a rede neuronal vólvese capaz de aprender e mellorar na tarefa que ten asignada sen necesidade de que as persoas interveñan.

3.4 UX1A interactúa



Os bebés humanos empregan os sentidos para percibir o mundo e utilizan o **seu corpo para interaccionar con el e aprender cal é o resultado das súas accións**. O paralelismo coa IA é directo, xa que o Axente Intelixente recolle información a través dos sensores e actúa na contorna a través dos seus actuadores para lograr o seu obxectivo, aprendendo, como acabamos de ver, da súa experiencia.

Do mesmo xeito que cos sensores, podemos facer unha comparación entre os *actuadores* do corpo humano e os dos axentes intelixentes:



3.4.1 Tipos de actuadores da Intelixencia Artificial

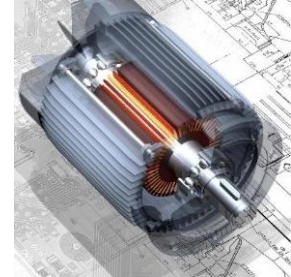
Unha sinxela clasificación dos actuadores é:

- Os que necesitan **movemento** para lograr o seu obxectivo.
- Os que realizan accións de **comunicación**.

Actuadores de movemento

Permiten que as máquinas se movan.

As aplicacións máis importantes destes actuadores son a locomoción e a manipulación, e os seus compoñentes fundamentais son **os motores**. En xeral, empréganse motores eléctricos de corrente continua e entre os máis habituais destacan os servomotores, que permiten controlar con moita precisión o ángulo e a velocidade do seu xiro.



Para a **locomoción** utilízanse **efectores** que conseguen o movemento de translación. Por exemplo:

- Patas.
- Rodas.
- Brazos.
- Alas e aletas.

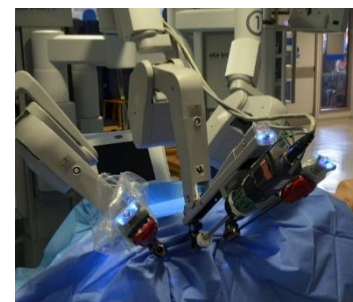
Hoxe en día xa existen múltiples robots con capacidade de movemento que utilizan IA. Por exemplo, malia que é unha solución sobre a que cómpre un debate a fondo, nunha residencia de maiores de Singapur hai un robot chamado Robocoach que adestra as persoas residentes para que non perdan a forma física.



Para as funcións de **manipulación**, utilízanse actuadores como:

- Brazos articulados.
- Pés.
- Pinzas.
- Punteiros.

Unha aplicación moi recente destes actuadores de manipulación son os robots DA VINCI instalados en varios hospitais galegos. Estes sistemas permiten ás e aos profesionais realizar cirurxías de altísima precisión coa axuda de IA.

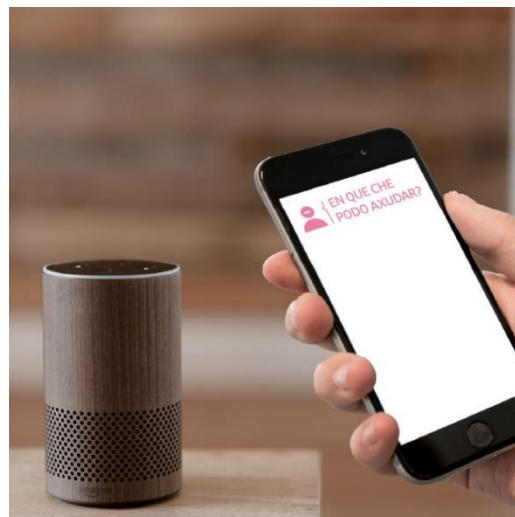


Actuadores de comunicación

Para chegar á interacción natural humano-máquina hai que entender que os sensores e os actuadores máis usados son os que realizan accións de comunicación.

A comunicación pode ser **unidireccional** (a información vai nun único sentido, de emisor/a a receptor/a) ou **bidireccional** (quen emite a mensaxe pode converterse en receptor/a e viceversa).

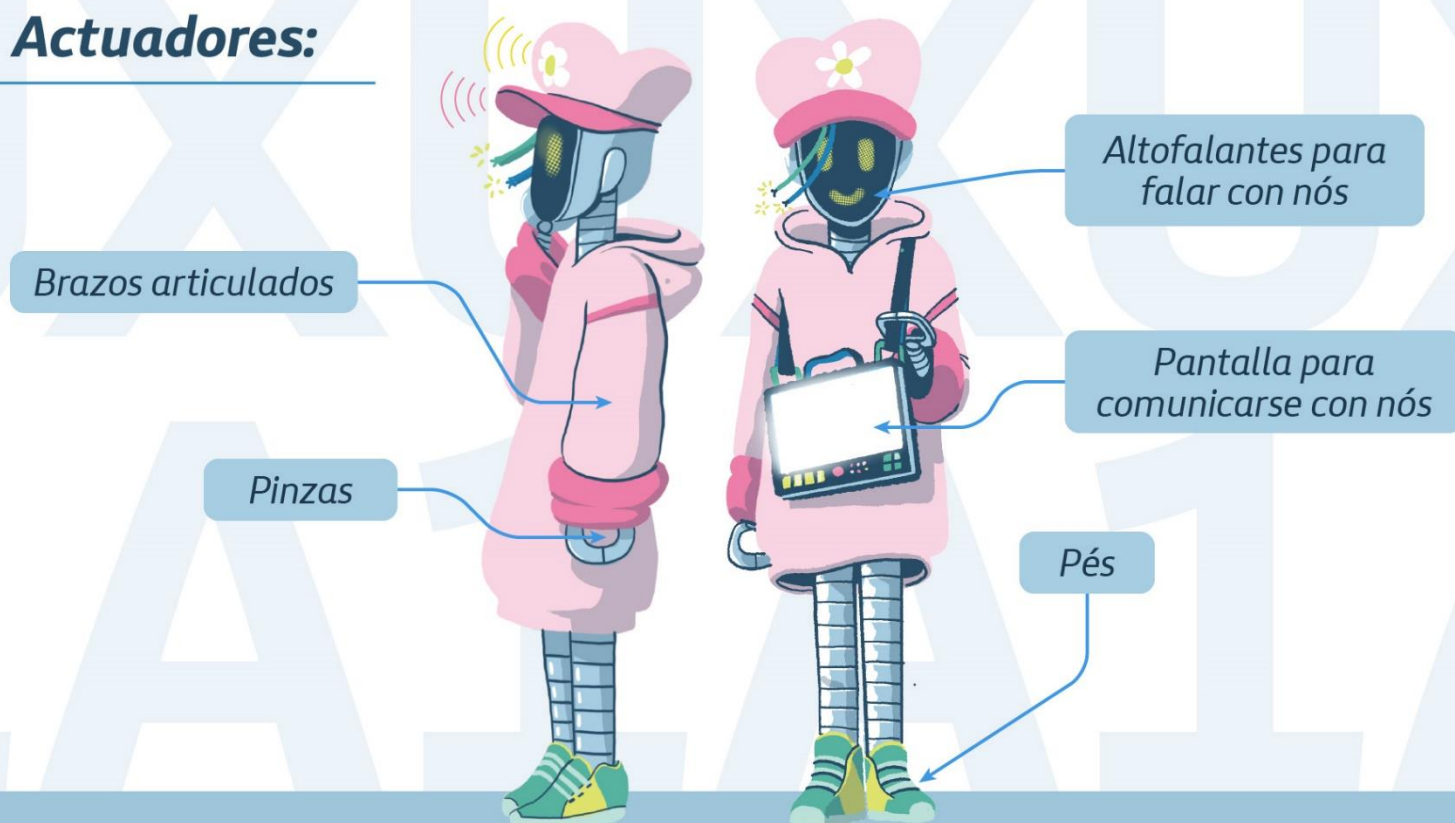
Para establecer esta comunicación utilízanse **micrófonos**, **altosfalantes** e **pantallas** que recollen e transmiten a información. Sistemas de asistencia persoal como *Alexa* ou *Siri* son bos exemplos que están presentes na nosa vida cotiá. Outros elementos empregados son as **antenas** que dan soporte as tecnoloxías WIFI ou Bluetooth.



Unha ferramenta de comunicación moi utilizada hoxe en día, e que pode ser un recurso moi útil na aula, é o **chatbot**. Este pode definirse como unha aplicación de IA que imita unha conversa real.

3.4.2 A rede de actuadores de UX1A

Actuadores:



PARA SABER MÁIS: Actuadores

Tipo de actuadores	Compoñente	Onde os ten UX1A	Paralelismo co corpo dos seres vivos	Funcións
De movemento	Brazos articulados	Brazos	Brazos humanos	UX1A pode facer movementos parecidos aos brazos humanos coas súas articulacións.
De movemento	Pinzas	Mans	Mans humanas	UX1A pode facer movementos parecidos aos dunha man con manopla.
De movemento	Pernas articuladas	Pernas	Pernas humanas	UX1A pode realizar movementos parecidos as pernas humanas coas súas articulacións.
De movemento	Pés	Pés	Pés	Permítenlle a UX1A camiñar e desprazarse con equilibrio.
De comunicación	Pantallas	Cara e barriga	Cara	Permítenlle a UX1A comunicarse mediante expresións faciais ou proxectado imaxes.
De comunicación	Altosfalantes	Cara ou corpo	Boca	UX1A emite sons para comunicarse.

Cos seus actuadores, os axentes intelixentes poden levar a cabo multitude de accións diferentes que teñen impacto na súa contorna. Non podemos esquecer que estas accións realízanas de forma autónoma, tomando decisións a partir dunha gran cantidade de datos. Pero tampouco podemos esquecer que detrás desa *autonomía* hai unha programación humana que decide como *razoa* e como *aprende* a máquina, é dicir, que a condiciona. Por este motivo, é importante coñecer e reflexionar sobre o impacto social que ten a IA.

En resumo:

- Tanto os seres vivos como os axentes da IA utilizamos actuadores para interaccionar coa na nosa contorna e lograr os nosos obxectivos.
- A IA emprega, fundamentalmente, actuadores de movemento (locomoción e manipulación) e de comunicación.
- Os axentes intelixentes levan a cabo, de forma autónoma pero condicionada pola súa programación, multitude de accións que teñen impacto na sociedade.

3.5 UX1A impacta na sociedade



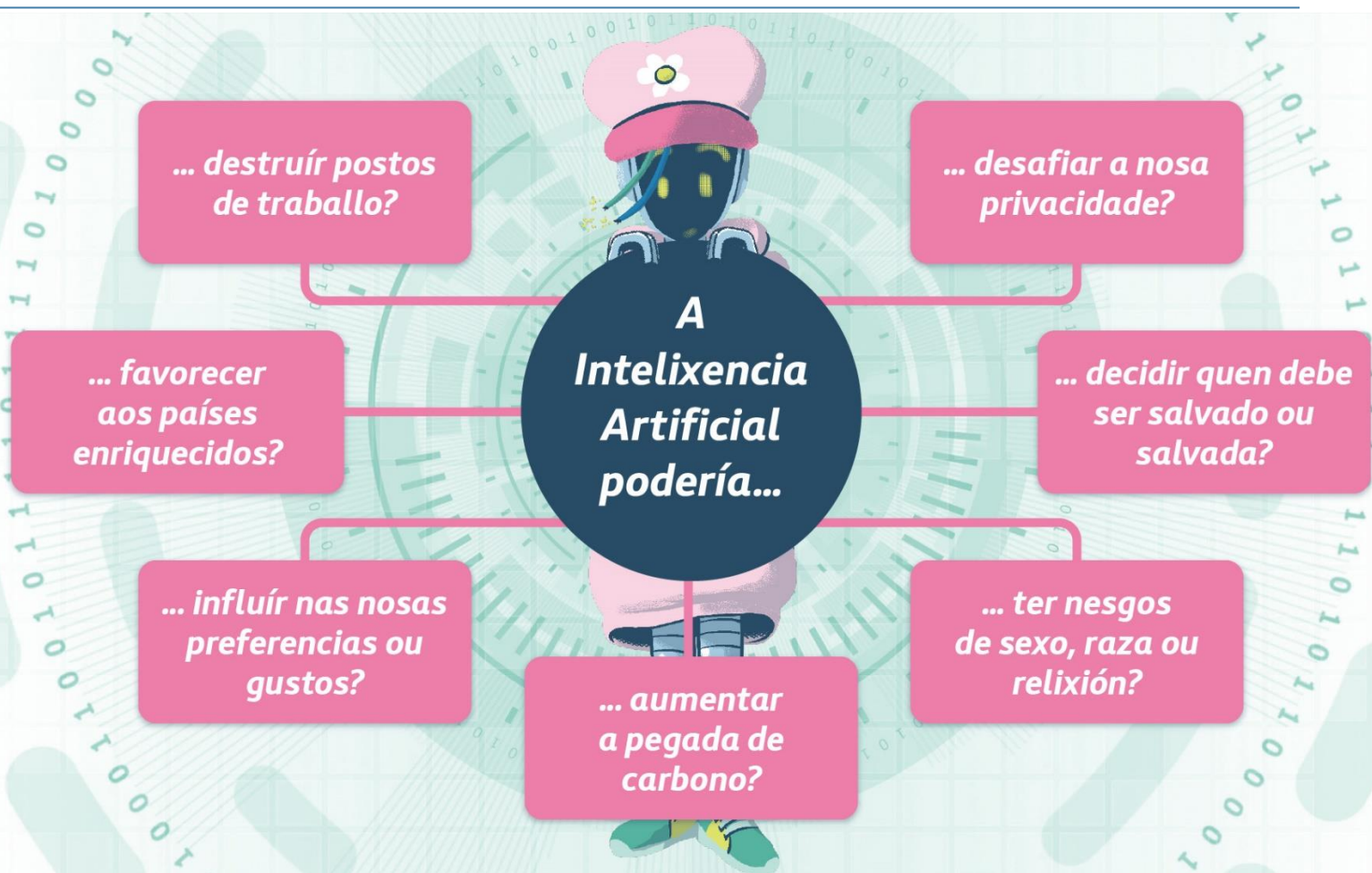
O emprego masivo dos sistemas da IA abre cuestións éticas e legais que deben ser afrontadas polas sociedades, a través dos seus gobernos, para asegurar unha transición xusta e gradual cara a un novo modelo de sociedade. O momento de formalizar tales cuestións é agora, polo que os e as estudantes actuais serán as primeiras xeracións que se verán afectadas por novas regulacións neste campo.

É importante, polo tanto, que o alumnado coñeza como a IA impacta na vida das persoas, non só en aspectos superficiais, senón que deben **aprender a ser coidadosos e coidadosas como usuarios/as, administradores/as e futuros/as creadores/as** desta potente ferramenta.

Por iso, o obxectivo principal deste bloque é analizar as consecuencias, positivas e negativas, que a IA ten e terá na sociedade, especificamente aquelas vinculadas a:

Ética
Sostibilidade
Aspectos legais
Oportunidades e preocupacións

Para introducir estes temas, e antes de mergullarnos na súa parte máis teórica, é interesante que paremos a pensar no impacto que a IA pode ter na nosa sociedade. As seguintes preguntas poden axudarnos a iniciar estas reflexións e servirnos de base para levar a cabo debates ou discusións na aula.



A continuación, recóllense algúns feitos, reflexións e ideas que nos poden axudar a abordar estas cuestións co noso alumnado.

1. A IA podería ter nesgos de sexo, raza ou relixión?

Os algoritmos de IA afectan cada vez máis as nosas decisións persoais. Ademais, as empresas e os gobernos aumentaron xa a súa dependencia dos algoritmos para tomar decisións. Hoxe en día utilízanse, por exemplo, para identificar perfís potencialmente criminais e para orientar decisións médicas que afectan a millóns de persoas. Neste contexto, investigacións recentes comezaron a alertar sobre a amplificación dos nesgos e prexuízos sociais detectada nestes algoritmos:

- Os grupos historicamente desfavorecidos, como as mulleres e as persoas de orixes étnicas non caucásicas, sofren máis os efectos do nesgo algorítmico.
- Os grupos dominantes benefíciense das taxas de precisión máis altas da IA, en comparación cos grupos minoritarios.

Isto pode xerar que as desigualdades presentes nas sociedades se vexan amplificadas pola acción das IA, máis tendo en conta que existe unha percepción xeral, pero errónea, de que as máquinas non cometen erros e son puramente *obxectivas*.

Por que ocorre isto

A parte fundamental e máis importante de calquera proxecto de IA é o deseño dos seus algoritmos e os datos aos que ten acceso, xa que tanto a percepción da realidade que ten a IA coma a súa aprendizaxe proveñen directamente deles.

O resultado ao que chegue ou as accións que leve acabo a IA dependerán tanto do seu deseño inicial como do conxunto de datos que lle subministremos ou ao que teña acceso: se o conxunto de datos é bo, o resultado do proxecto será bo. Pola contra, **se temos datos nesgados o resultado será nesgado**.

Neste sentido, podemos atopar dúas causas principais de nesgos na IA:

- Conxuntos nesgados de datos: a IA aprende atopando patróns nos datos aos que ten acceso e, se estes datos sobrerrepresentan a un grupo determinado, a IA xerará nesgos favorables a ese grupo e desfavorables ao resto. Por exemplo: se nunha IA que recoñece a voz humana os datos de adestramento son maioritariamente voces masculinas, a IA terá máis problemas para recoñecer as voces femininas.
- Deseños e adestramentos nesgados: a solución que ofrezca a IA tamén dependerá do deseño e do adestramento feito por persoas, que poden reproducir (intencionadamente ou non) estereotipos de todo tipo que leven a nesgos e discriminacións. Por exemplo: pódese programar e adestrar unha IA para que ofrezca recomendacións de compras ou de empregos diferentes a mulleres e a homes, en función dos estereotipos existentes, fomentando así a perpetuación e incluso a amplificación das fendas de xénero e das desigualdades sociais.

Polo tanto, para **garantir a diversidade e a inclusión** no deseño e desenvolvemento de calquera proxecto relacionado coa IA, **os datos introducidos deben ser representativos** de toda a poboación. Só deste xeito poderemos evitar o nesgo na súa toma de decisións.

Neste sentido, **a falta de diversidade nos equipos de investigación é un problema**. A actual sobrerrepresentación dos homes brancos nos equipos de traballo de IA pode dificultar a detección dos prexuízos que desenvolven os seus algoritmos fronte a outros grupos de persoas. Se non se pon remedio, a IA tenderá a ampliar as diferenzas. Por exemplo, só o 22% das persoas que se dedican profesionalmente á IA son mulleres³ e os estudos demostran que os algoritmos están a amplificar os nesgos de xénero.⁴

A solución parece pasar, polo tanto, pola inclusión dunha maior representación dos sectores vulnerables da poboación nestes equipos, o que melloraría a detección de posibles prexuízos e evitaría parte da discriminación algorítmica. Pero podemos engadir que a forma correcta de deseñar un produto de IA é contar cun equipo que ademais de diverso sexa consciente e estea formado no impacto que as súas decisións terán na perpetuación, amplificación ou redución das desigualdades sociais existentes a través dos nesgos que xeren as súas IA.

³ Consulta máis datos na nova [“Para ser inteligente, la revolución digital tiene que incluir a la mujer”](#) das Nacións Unidas.

⁴ Neste apartado utilizamos exemplos referentes aos nesgos de xénero, pero por desgracia atopamos tamén outros moitos tipos de nesgos (de raza, relixión, orixe das persoas, orientación sexual etc.).

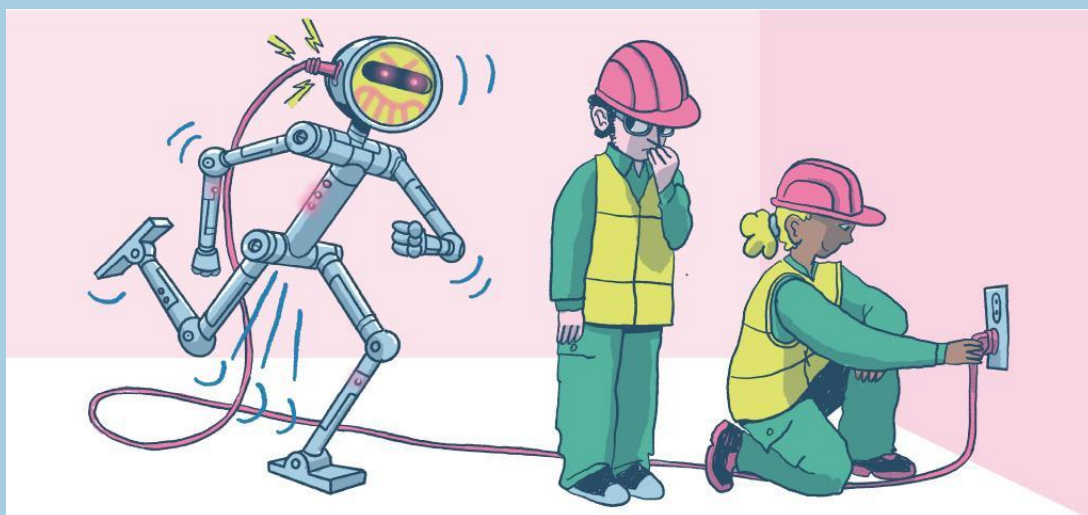
2. A IA podería destruír postos de traballo?

Automatización, robótica e IA xa comezaron a transformar o mundo do emprego, sobre todo nos sectores da automoción ou da industria. Prevese que na década do 2030, un de cada 3 empregos actuais xa non o fará unha persoa senón que estará automatizado.⁵

Todo apunta a que os empregos máis afectados serán aqueles que requiren menor cualificación, pois a IA ten a capacidade de substituír tarefas laborais que se basean en regras, que son repetitivas e que non requiren destreza nin empatía. Xa está a provocar importantes perdas de emprego nas funcións administrativas de *back-office* en banca, saúde, seguros e contabilidade. Trátase de funcións que nos últimos anos foron subcontratadas a países empobrecidos como India, Vietnam, Sudáfrica ou Marrocos.

Pola contra, os perfís profesionais máis demandados serán aqueles que requiran traballar por proxectos, liderar e traballar en equipo, dotes de comunicación e iniciativa. Tamén serán moi valorados o pensamento crítico, a capacidade para resolver problemas e a creatividade. Isto significa que o ser humano sempre será superior ás máquinas, xa que non poderá ser substituído nestes campos.

Así, a curto e medio prazo prevese a desaparición de certos tipos de traballo, pero tamén un grande incremento de postos relacionados coas TIC, analistas de datos ou comerciais especialistas en ofertas dixitais, así como expertas e expertos en robótica, internet das cousas e comunicación entre máquinas.



Por que ocorre isto

Os traballos mecánicos ou automatizables son máis susceptibles de ser realizados por máquinas. Isto non é algo novo, xa ocorreu na Revolución Industrial. As máquinas poden realizar de forma máis eficiente determinadas tarefas, que tenden a ser aquelas vinculadas aos traballos menos especializados. Porén, o mantemento e a programación das propias máquinas xera novos postos de traballo que antes non existían. Iso si, trátase de tarefas que normalmente requiren de maior cualificación ou formación que as anteriores.

⁵ Segundo o estudo "[Will robots really steal our jobs?](#)" da consultora PwC.



<https://youtu.be/tavZPHIixZw>

3. A IA podería favorecer aos países enriquecidos?

Espérase que a IA xere case 4 billóns de dólares de valor engadido no 2022⁶. Pero o crecemento económico impulsado pola IA probablemente será desigual.

Investigacións do Fondo Monetario Internacional constatan que as novas tecnoloxías corren o risco de ampliar a brecha entre os países enriquecidos e os países empobrecidos ao desprazar máis investimentos cara ás economías enriquecidas, onde a automatización xa está establecida.

Por que ocorre isto

Os países empobrecidos sofren escaseza de man de obra altamente cualificada, o que socava a súa competitividade, e actualmente empregan man de obra pouco ou medio cualificada (de custo baixo) para producir bens que se consumen principalmente nos países enriquecidos. Pero o desenvolvemento do deseño e a impresión 3D combinados coa IA permitirán que os países enriquecidos fabriquen roupa, zapatos, dispositivos e outros produtos personalizados, o que podería significar que a era da subcontratación dos países empobrecidos estaría chegando ao seu final e que os investimentos actuais nestas economías tenderán a desaparecer.

Así, a medida que a tecnoloxía xoga un papel máis dominante na economía global, o poder económico tende a condensarse naquelas zonas do planeta que impulsan o seu desenvolvemento e a concentración de riqueza e ingresos que xa podemos observar en lugares como Silicon Valley podería replicarse a nivel mundial, empeorando a desigualdade.

Ademais, a industria tecnolóxica absorbe talento de todo o mundo, deixando escaseza de man de obra moi cualificada nos países empobrecidos. As grandes economías demandan persoal traballador altamente

⁶ Segundo o artigo "[Recomendación sobre a Ética da Intelixencia Artificial](#)" da UNESCO.

cualificado para áreas coma a IA e estas persoas tenden a preferir vivir en cidades seguras e desenvolvidas que ofrecen mellores oportunidades.

Como consecuencia deste desprazamento de man de obra cualificada e capital cara aos países enriquecidos, os ingresos fiscais dos gobernos dos países empobrecidos veranse minguados, o que provocará unha redución do seu investimento en infraestruturas, sanidade, educación... ampliando a brecha xa existente entre países enriquecidos e empobrecidos.

Cómpre sinalar que Galicia está a apostar polo desenvolvemento dunha IA fiable, o que atraerá máis talento, máis empresas e máis investimento. Isto suporá un revulsivo para que Galicia transforme o seu perfil produtivo facéndoo máis rendible, autónomo e sostible.

4. A IA podería influír nas nosas preferencias ou gustos?

Os estudos mostran que isto xa está ocorrendo e grandes institucións públicas e privadas xa se aproveitan desta realidade. Por exemplo, as redes sociais son agora o lugar principal para a propaganda política e esta influencia estase usando para dirixir o comportamento de voto. Así mesmo, empresas comerciais, de anuncios e de venda en rede empregan estes algoritmos para maximizar o seu beneficio. Esta influencia é o modelo de negocio das industrias do xogo e das grandes cadeas de venda de produtos (*Inditex*, *Aliexpres*, *Amazon* etc.) e estase a converter no modelo de negocio central de internet.

Por que ocorre isto

As persoas confían habitualmente nas tecnoloxías da IA coma os sistemas de recomendacións sobre que programa ver ou que música escoitar. Os motores de busca personalizados, asistidos por IA, ofrecen a cada usuario e usuaria resultados *únicos*, baseados nas súas buscas anteriores ou no seu comportamento ao navegar pola rede.

Pero estes resultados *únicos* non son neutrais. As recomendacións, e incluso os resultados das buscas en internet, están condicionados polos intereses (xeralmente económicos) de empresas ou institucións. Así, os artigos que nos amosan as tendas online son aqueles que prefiren que compremos porque pagan máis pola súa publicidade ou aqueles que cren que mercaremos con maior probabilidade.

Por outra parte, estes algoritmos poden favorecer a tendencia das persoas a vincularse unicamente con outras persoas (ou empresas, artistas, *influencers* etc.) que comparten a súa forma de pensar ou os seus gustos (o chamado nesgo de confirmación). Aínda que isto tamén ocorre na vida física (tendemos a trabar amizade con quen sentimos que compartimos un modelo de vida ou de pensamento), na rede xérase un efecto burbulla que nos illa de calquera opinión diferente á nosa, xa que os algoritmos de IA rara vez nos recomendarán, por exemplo, ler algo cun punto de vista radicalmente diferente ao noso.

Isto dificulta o noso acceso a unha información veraz, diversa e contrastada, así como o desenvolvemento dun pensamento crítico propio, e facilita que institucións, empresas ou outros axentes inflúan ou manipulen as nosas preferencias, gustos ou formas de pensar.

5. A IA podería aumentar a pegada de carbono?

Cada día xéranse, transfírense e almacénanse na “nube” miles de millóns de datos dos nosos móbiles, das nosas buscas na rede, imaxes, actividades financeiras, e un longo etc. O sector das TIC (Tecnoloxías da Información e Comunicación) prevé que para o 2025 utilizará o 20% da electricidade mundial e emitirá o 5,5 % de todas as emisións de carbono, que poderían chegar ao 14% en 2040.⁷

Por que ocorre isto

Tanto o proceso para adestrar unha IA coma o seu funcionamento requiren dunha gran cantidade de enerxía, coas consecuentes emisións de CO₂ asociadas. Isto débese a que a IA non deixa de ser unha máquina que fai moitos cálculos moi complexos cunha cantidade inmensa de datos, e, polo tanto, require de recursos informáticos que demandan un gran consumo enerxético.

Recordemos que adestrar unha IA, especialmente cando contén unha rede neuronal, pode levar semanas ou meses, tempo no cal é necesario subministrarlle enerxía aos súperordenadores nos que se desenvolven. E, despois, xérase un novo consumo enerxético cada vez que se executa un algoritmo, polo que tanto a consulta como a modificación constante das inmensas bases de datos que almacenan as IA teñen un gran impacto en forma de pegada de carbono.

É por isto que a sostibilidade da IA debe ser lexislada adecuadamente, para frear e non aumentar a pegada de carbono no noso planeta. Necesitamos medidas específicas para facer fronte a este gran problema e xa hai iniciativas neste sentido para desenvolver *algoritmos verdes* que consuman menos recursos e sexan máis respectuosos co noso planeta.

6. A IA podería decidir quen debe ser salvado ou salvada?

Para abordar esta cuestión podemos analizar o exemplo do coche autónomo que, ante unha situación de emerxencia como quedar sen freos, ten que tomar unha decisión sobre a quen é preferible atropelar: a unha avoa ou a un neno. Este é un dilema ético típico cando pensamos na IA e pon de manifesto a importancia da ética no desenvolvemento das tecnoloxías.

As persoas tomamos decisións morais a diario. Cando quen conduce decide pisar o freo para evitar chocar cun peón ou cunha peoa, está tomando a decisión moral de trasladar o risco ás persoas do coche. Ocorre o mesmo se xira o volante para desviarse porque só pode salvar a unha persoa das dúas que hai no paso de peóns, aínda que quizais a escolla se faga de forma inconsciente.

A quen salvarías ti, á avoa ou ao neno? Cres que só hai unha resposta correcta?

No caso do coche autónomo que, cos freos rotos, se dirixe a toda velocidade cara a unha avoa e un neno e que desviándose un pouco só pode salvar a unha persoa, a diferenza está en que a decisión será tomada polo algoritmo do coche, que decidirá quen debe ser salvado ou salvada.

⁷ Segundo a reportaxe “*Eficiencia Energética: Reto de los Data Centers del Futuro*” de Data Center Dynamics.

Por que ocorre isto

Un coche autónomo é un vehículo capaz de detectar a súa contorna e moverse con pouca ou ningunha intervención humana, polo que debe estar preparado para comprender e actuar ante calquera situación.

Para que o vehículo se mova con seguridade e comprenda a súa contorna de condución, unha infinidade de sensores diferentes ao longo do coche deben capturar unha enorme cantidade de datos en todo momento. Estes datos son procesados polo sistema informático de condución autónoma do vehículo, que tamén debe recibir unha formación considerable para comprendelos e poder tomar decisións ante calquera situación de tráfico imaxinable.

Así, dependendo da programación e da *formación* que se lle dea ao coche autónomo, este decidirá quen debe salvarse (e, por conseguinte, quen é máis *prescindible* en comparación coa outra persoa).

Noutros apartados xa abordamos cuestións que entran en xogo neste caso:

- A falsa percepción de *neutralidade* e *obxectividade* que se ten das IA na sociedade. Cómpre ter claro que a decisión que tome a máquina estará condicionada polo seu adestramento.
- Os nesgos (de xénero, raza, etnia, capacidade etc.) poderían afectar a esta decisión se, por exemplo, se adestra ao coche con exemplos nos que, aínda que sexa inconscientemente, sempre se decida salvar a persoas brancas.

Por último, cómpre ter en conta que este é un exemplo paradigmático que ilustra a relevancia da ética na IA, pero non é o único. Outros exemplos poden estar relacionados coa implementación de sistemas de IA nos tribunais de xustiza (a quen liberarían e a quen condenarían?) ou no proceso de distribución de fondos públicos (a quen beneficiarían os investimentos escollidos?).

7. A IA podería desafiar a nosa privacidade?

Cando o alumnado (ou calquera persoa) emprega as RRSS (*Instagram, Tiktok, Twitch, Facebook* etc.) os algoritmos da IA **recompilan silenciosamente os datos** dos perfís que ven, dos anuncios que miran, dos enlaces nos que fan clic etc., ademais do tempo que dedican, a hora do día en que as empregan e a localización dende onde usan cada aplicación en concreto.

En realidade, recompilan moitos máis datos dos que cremos: a nosa imaxe, as amizades que temos, as cousas que facemos, o que nos interesa en cada momento, onde estamos a cada hora do día etc. Como resultado, entre a inmensidade de datos que hai na rede atópase tamén un retrato bastante fidedigno de cada persoa que a emprega (as veces con detalles que nin un ou unha mesma coñece de si).

Todos estes datos poden empregarse para fins moi diversos e as veces eticamente cuestionables. Por exemplo, o recoñecemento facial mediante fotografías ou vídeos permite a identificación das persoas e, polo tanto, a creación de perfís de busca especializados que xa se están aplicando á vixilancia de poboacións enteiras e á vixilancia selectiva.

Por que ocorre isto

A IA incrementou enormemente as posibilidades de recollida de datos persoais na rede, así coma de análise dos mesmos, e xa vimos noutros apartados desta guía o valor que teñen estes datos para institucións e empresas. De feito, hoxe en día os propios datos son xa un valor de mercado, a *moeda* coa que pagamos todos os servizos *gratis* dos que disfrutamos na rede. Ademais, gran parte destes datos intercámbianse entre entidades a cambio dunha taxa económica, moitas veces sen ningún tipo de control.

Isto débese a que, nas últimas décadas, a tecnoloxía avanzou a pasos axigantados sen que houbera unha resposta internacional áxil para regulamentala ou lexislala.

A través da ética na IA preténdese controlar quen recolle os datos, quen ten acceso a eles e como se empregan para modificar o noso comportamento tanto na rede como fóra dela, xa que, como vimos previamente, os algoritmos poden utilizarse para orientar as decisións de persoas ou grupos.

Por este motivo, e dada a idade cada vez máis temperá na que se comezan a empregar estas tecnoloxías, é fundamental que o alumnado sexa consciente dos problemas de privacidade que se derivan do seu uso.



3.5.1 Ética

Como acabamos de ver, o desenvolvemento de sistemas de IA pode ter importantes consecuencias na nosa sociedade, moitas delas non desexadas que incluso van en detrimento dos nosos dereitos individuais e colectivos. As decisións que toman as IA poden ter efectos directos na vida das persoas ou efectos indirectos na perpetuación ou eliminación de situacións inxustas como a desigualdade social.

Porén, **cómpre recordar que a IA é unha máquina que *aprende a partir de datos e dos patróns que atopa neses datos*. Non ten conciencia nin empatía e, polo tanto, non ten *ética*.**

A ética dunha persoa é o conxunto de principios e sistemas morais que gobernan o seu comportamento. É dicir, son os valores que nos guían á hora de decidir se unha opción é boa (correcta) ou mala (incorrecta). Aínda que a ética ten unha gran dimensión persoal, xa que non todas as persoas pensamos igual, ten tamén unha importante dimensión social: aquilo que se considera correcto ou incorrecto nunha sociedade. Así, con independencia do que alguén concreto poida crer, hoxe en día compartimos socialmente que a discriminación por motivos de raza ou xénero (entre outras) está mal e é eticamente reprobable.⁸

Nun inicio, a única forma que ten unha IA de recoñecer o que é correcto ou incorrecto é o etiquetado dos datos que fai a persoa ou o equipo que a adestra. Polo tanto, a súa ética e os seus propósitos non serán outros que os de quen a desenvolva:

- Se dita persoa ten malas intencións, a IA responderá a estas.
- Pola contra, se no seu desenvolvemento se aplican principios socialmente éticos, a IA *aprenderá* e aplicará estes principios.

Unha vez adestrada, a IA poderá evolucionar *aprendendo* de novos datos. Estes datos poden, tamén, determinar a *ética* que a máquina *aprenda*: se un bot conversacional adestrado correctamente recibe multitude de comentarios racistas pode terminar aprendendo que esa é a forma *natural* de comunicarse das persoas e *volverse* racista.

Por este motivo, cómpre tamén monitoroar, comprobar e probar os resultados dos proxectos de IA, para entender ben como se están comportando e asegurar que non van en contra da nosa ética.

Polo tanto, **é fundamental ter clara a ética coa que queremos construír ou desenvolver a tecnoloxía, pois doutro xeito corremos o risco de obter resultados que vaian en contra dos nosos principios**, é dicir, que a IA non respecte os nosos valores morais e dea resultados nesgados ou abertamente racistas, machistas etc.

Por outra parte, **dado que a IA non ten conciencia, ética nin capacidade para razoar, será importante pensar quen é responsable das accións dun robot**. Neste sentido, parece existir consenso en que debe haber unha rendición de contas e de responsabilidades e que esta responsabilidade debe ser distribuída entre todos os actores que interveñen na súa actuación: tanto as persoas que as deseñan e as desenvolven como as e os usuarios, o software e o hardware.

⁸ Cómpre sinalar que aquilo que se considera ético varía en función da sociedade e do momento histórico no que nos atopemos.

PARA SABER MÁIS: A ética das máquinas

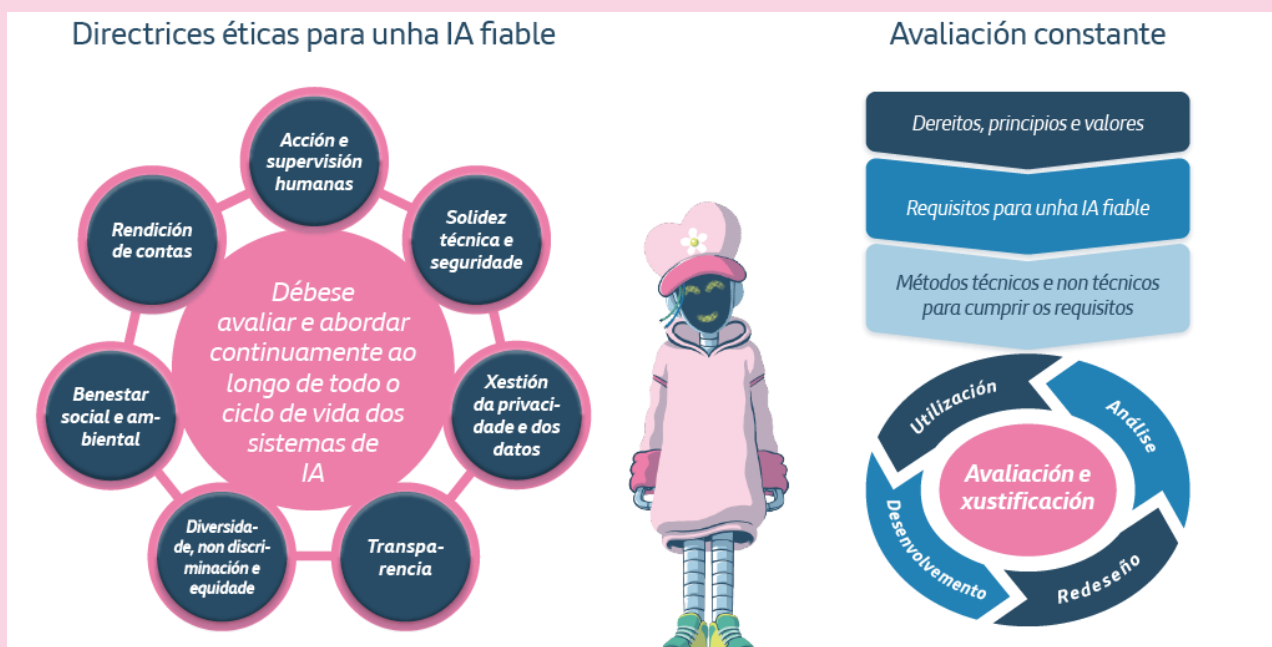
Xa en 1942 Isaac Asimov estableceu a ética das máquinas a través das “Tres leis da robótica”:

1. Un robot non pode danar a un ser humano ou, por inacción, permitir que un ser humano sufra danos.
2. Un robot debe obedecer as ordes dadas polos seres humanos, excepto cando ditas ordes entren en conflito coa primeira lei.
3. Un robot debe protexer a súa propia existencia, sempre que non entre en conflito coa primeira ou a segunda lei.

Con todo o aprendido podemos considerar esas regras como *de mínimos*. Por iso, o último documento “[Directrices éticas para unha IA fiable](#)” da Comisión Europea suxire 7 requisitos para lograr unha IA fiable:

1. **Acción e supervisión humana:** dereitos fundamentais, acción e supervisión humana.
2. **Solidez técnica e seguridade:** capacidade para resistir ataques e seguridade, precisión, fiabilidade e reproducibilidade.
3. **Xestión da privacidade e dos datos:** respecto á privacidade, calidade e integridade dos datos e ao acceso aos mesmos.
4. **Transparencia:** trazabilidade, explicabilidade e comunicación.
5. **Diversidade, non discriminación e equidade:** ausencia de nesgos inxustos, accesibilidade e deseño universal, así coma a participación de todas as partes interesadas.
6. **Benestar ambiental e social:** sostibilidade e respecto ao medio ambiente, impacto social, sociedade e democracia.
7. **Rendición de contas:** auditable, minimización dos efectos negativos e notificación dos mesmos, busca do equilibrio e a compensación.

Para cumprir estes requisitos debería levarse a cabo unha **avaliación constante**, xa que os sistemas de IA evolucionan constantemente.



3.5.2 Sostibilidade

En 1987, a Comisión Brundtland das Nacións Unidas definiu a sostibilidade como "atender as necesidades do presente sen comprometer a capacidade das xeracións futuras para satisfacer as súas propias necesidades".

Deste xeito, o desenvolvemento sostible será aquel modelo de progreso que garanta o equilibrio entre o crecemento económico, o coidado do medio ambiente e o benestar social, empregando os recursos existentes sen pór en perigo os recursos do mañá.



A IA pode contribuír a este desenvolvemento sostible axudando a predicir erros e a planificar mellor o futuro do noso planeta, construíndo cidades máis eficientes, empregando os recursos de forma sostible e reducindo e xestionando mellor os residuos que xeramos, entre outros.

Segundo un estudio publicado en [Nature Communication](#), a IA pode facilitar o cumprimento do 79% dos ODS (Obxectivos de Desenvolvemento Sostible das Nacións Unidas):

- A IA podería analizar bases de datos interconectadas a gran escala para desenvolver accións conxuntas dirixidas a preservar o medio ambiente, comprendendo o cambio climático e creando modelos dos posibles impactos deste.
- A IA podería apoiar sistemas enerxéticos con baixas emisións de carbono, cunha gran integración das enerxías renovables, e axudar así a mellorar a eficiencia enerxética, necesaria para facer fronte ao cambio climático.
- A IA podería empregarse para mellorar a saúde dos ecosistemas, a conservación de áreas costeiras e mariñas e para previr e reducir a contaminación mariña identificando automaticamente verteduras mariñas.
- A IA podería axudar a identificar e reverter tendencias de desertización procesando grandes cantidades de imaxes de satélite en pouco tempo.

Porén, cómpre ter en conta que estes esforzos sobre a acción climática poden verse minguados polas altas necesidades enerxéticas das aplicacións da IA, especialmente se se empregan fontes de enerxía non renovables.

3.5.3 Aspectos legais

Como vimos anteriormente, o desenvolvemento tecnolóxico da IA é moi rápido e os seus beneficios para a sociedade son múltiples. Porén, aínda que a maioría dos desenvolvementos de IA poden presentar un risco baixo ou nulo, algúns destes sistemas poden entrañar perigos significativos que deben previrse, tratarse e abordarse para evitar situacións non desexadas. Por iso é necesaria unha actuación lexislativa.

No momento de redactar esta guía (ano 2022) a lexislación concreta sobre IA é moi escasa e recente. Atopamos a maioría de propostas lexislativas a nivel europeo, que terán que ser trasladadas progresivamente ás lexislacións dos estados membros:

- **Libro Branco sobre a Intelixencia Artificial (2020)**

O documento, publicado o 19 de febreiro de 2020, ofrece alternativas políticas para facilitar un desenvolvemento seguro e fiable da IA en Europa, que respecte plenamente os valores e dereitos da cidadanía da UE.

Accede ao documento: [Libro branco sobre a Intelixencia Artificial – un enfoque europeo orientado a excelencia e a confianza](#)

- **Lei de Intelixencia Artificial (2021)**

En xullo de 2020 anunciouse a creación do Consello Asesor de Intelixencia Artificial para un uso ético e seguro da mesma, que foi o encargado de redactar a Lei de Intelixencia Artificial presentada o 21 de abril de 2021.

Trátase dunha proposta da Comisión Europea que ten por obxecto garantir a protección dos dereitos fundamentais e da seguridade das persoas usuarias, co fin de que haxa confianza no desenvolvemento e na adopción da IA. Cun enfoque normativo, establece uns principios que deben cumprir os sistemas de IA para xestionar os riscos e problemas (saúde, seguridade e dereitos fundamentais) vinculados a estes, sen impedir o desenvolvemento tecnolóxico.

Accede ao documento: [Proposta de regulamento do Parlamento Europeo e do Consello polo que se establecen normas harmonizadas en materia de Intelixencia Artificial \(Lei de Intelixencia Artificial\) e se modifican determinados actos lexislativos da Unión](#)

- **Resolución do Parlamento Europeo sobre a Intelixencia Artificial nos sectores educativo, cultural e audiovisual (2021)**

O 19 de maio de 2021 o Parlamento Europeo aprobou unha resolución sobre o emprego da intelixencia artificial na educación, a cultura e o sector audiovisual na que demanda que as tecnoloxías de IA se deseñen de xeito que protexan a non discriminación, a igualdade de xénero e a diversidade social, cultural e lingüística: *“a intelixencia debe ser adestrada para non replicar a discriminación”*.

No eido da educación, os eurodeputados e as eurodeputadas subliñan que o profesorado sempre debe poder corrixir as decisións tomadas polos axentes da IA, especialmente no que se refire a selección e

a avaliación do alumnado. Tamén destacan a necesidade de mellorar as habilidades dixitais en toda Europa e de formar e preparar ao profesorado para o uso da tecnoloxía da IA na educación.

Accede ao documento: [Resolución do Parlamento Europeo, do 19 de maio de 2021, sobre Intelixencia Artificial na educación, a cultura e o sector audiovisual](#)

- **Lei de Gobernanza de Datos (2021)**

O 30 de novembro de 2021 o Parlamento e o Consello Europeo acadaron un acordo provisional sobre a “Lei de Gobernanza de Datos” que ten por obxectivo aumentar a dispoñibilidade de datos para as empresas e os novos proxectos de negocio co fin de estimular a innovación.

Accede aos documentos:

[Proposta de regulamento do Parlamento Europeo e do Consello relativo á gobernanza europea de datos \(Lei de Gobernanza de Datos\)](#)

[Resumen do informe da avaliación de impacto que acompaña ao documento ‘Proposta de Regulamento do Parlamento Europeo e do Consello relativo á gobernanza europea de datos \(Lei de Gobernanza de Datos\)’](#)

- **Informe final da Comisión Especial sobre a Intelixencia Artificial na Era Dixital (2022)**

O 18 de xuño de 2020 o Parlamento Europeo creou a Comisión Especial sobre a Intelixencia Artificial na Era Dixital (AIDA) coa finalidade de analizar o impacto futuro da IA na economía da UE e a súa contribución ao valor empresarial e ao crecemento económico.

Esta comisión aprobou o seu informe final o 22 de Marzo de 2022, no que establece unha folla de ruta para a IA con recomendacións específicas para a UE e os Estados membros. Neste informe, a AIDA pide unha estratexia europea para un uso seguro da IA, sobre todo no que respecta a infancia e a redución da brecha de xénero neste ámbito, así coma a promoción dunha IA centrada no ser humano, fiable, integradora, baseada nos dereitos fundamentais e que xestione os riscos e aproveite os beneficios da IA.

Ademais, dito informe considera a Unión Europea coma o primeiro bloque de países que introduce lexislación en materia de IA e que ten potencial para iniciar o debate e contribuír ao desenvolvemento de regras e normas globais.

Accede ao documento: [Informe sobre a Intelixencia Artificial na Era Dixital](#)



A nivel estatal, a [lei 22/2021 do 28 de decembro de Presupostos Xerais do Estado Español para o ano 2022](#), na disposición adicional centésima trixésima, autorizou ao Goberno a impulsar a creación da Axencia Española de Supervisión da Intelixencia Artificial (AESIA), que terá por cometido desenvolver medidas para a redución de riscos sobre seguridade, saúde e dereitos fundamentais que poidan derivarse do emprego da IA.

E a nivel autonómico, a Xunta de Galicia e a Axencia para a Modernización Tecnolóxica de Galicia (AMTEGA) publicaron no 2021 o documento “[Estratexia Galega de Intelixencia Artificial 2030. Cara unha Galicia Intelixente](#)” coma unha folla de ruta para promover o emprego da IA en sectores clave da economía galega, mellorar servizos coma a sanidade ou a política social, potenciar o patrimonio cultural e o turismo, apoiar as pequenas e medianas empresas no proceso de transformación dixital e formar o capital humano para que acade habilidades dixitais avanzadas en IA.

3.5.4 Oportunidades e preocupacións fundamentais que propón a IA

Como vimos nos apartados anteriores, a IA ofrece moitas oportunidades de mellora da sociedade pero tamén pode xerar determinados problemas que debemos coñecer para previlos durante o seu desenvolvemento e uso. A continuación sinalamos, para cada ámbito temático, algunhas destas oportunidades e preocupacións:

1 Impacto medioambiental

Preocupacións

Os grandes centros de datos, de almacenamento (nube) e de cómputo requirido para a IA son unha boa parte do problema de consumo de enerxía, xeración de calor e pegada de carbono do noso planeta. O sector dixital é responsable de máis do 2% das emisións globais de gases de efecto invernadoiro e a extracción de datos consume case o 10% da enerxía a nivel mundial, polo que é necesario abordar este alto consumo de enerxía e a conseguinte emisión de carbono.

Oportunidades

Unha IA fiable ofrece un potencial enorme para reducir o impacto que exerce o ser humano sobre o medio ambiente, pois os datos que nos ofrecen os satélites poden axudar, por exemplo, a:

- Mellorar a investigación sobre, por exemplo, a contaminación ou a deforestación.
- Reducir as emisións de efecto invernadoiro.
- Previr e responder de xeito rápido ante desastres naturais coma os incendios forestais.
- Mellorar a xestión dos ecosistemas terrestres e mariños e a restauración dos hábitats.
- Mellorar a produción na industria e facela máis eficiente reducindo a xeración de residuos e as emisións contaminantes.

Ademais, realizando un seguimento do consumo poderán facerse recomendacións en tempo real para un uso máis eficiente da enerxía.

2 Privacidade, transparencia e dereitos fundamentais

Preocupacións

Unha recollida global de datos proporciona información das persoas usuarias que pode xerar unha explotación dos mesmos para obter beneficios comerciais e políticos. Por exemplo, os anuncios dirixidos xeran preocupación sobre as posibles manipulacións de consumidores e consumidoras.

Oportunidades

Sempre que se recollan e xestionen de forma adecuada, o emprego da IA para o tratamento dos datos persoais tamén podería brindar novas oportunidades para reforzar os dereitos fundamentais das persoas. Por exemplo, permitiría axilizar os trámites administrativos e diminuír a carga burocrática de moitas xestións que a cidadanía fai no seu día a día, pois brindaría a posibilidade de articular sistemas de xestión de datos compartidos polas administracións e de unificación de expedientes. Isto podería servir, entre outras cousas, para garantir que os servizos médicos de calquera parte do territorio (España ou Europa, por exemplo) tiveran acceso ao historial da ou do paciente que o precise, para axilizar solicitudes de axudas ou subsidios que teñan requisitos vinculados a outras administracións, ou para evitar que unha mesma persoa teña que introducir os seus datos repetidamente dependendo de con que administración se estea a relacionar.

Por outra parte, a capacidade de análise da IA podería empregarse para detectar noticias falsas ou perfís empregados con fins nocivos (como bots programados para estafar). Pero cómpre ter en conta que para que a recollida e o intercambio masivo de datos poida constituír unha oportunidade estes deben limitarse a datos anónimos e non persoais e, ademais, as persoas debemos estar protexidas pola lexislación e posuír un control total sobre os nosos datos a través de mecanismos sinxelos de acceso, rectificación, oposición, supresión e limitación do tratamento (en Europa contamos co Regulamento Xeral de Protección de Datos – RGPD).

3 Vixilancia selectiva ou masiva

Preocupacións

A IA posibilita unha identificación cada vez máis eficiente das persoas (recoñecemento facial, datos biométricos, detectores de mentiras, avaliación da personalidade empregando microexpresións, detección da voz etc.) que pode utilizarse para limitar a liberdade e a autonomía das persoas, tanto a nivel individual como colectivo. Por exemplo, a liberdade de reunión e protesta ou a de culto poden verse ameazadas, xa que é posible rastrexar e controlar a persoas vinculadas a certas crenzas ou involucradas en determinadas accións.

Oportunidades

A identificación cada vez máis eficiente das persoas tamén pode ter un resultado positivo e en consonancia cos valores éticos: detectar fraude, branqueo de capitais, financiación do terrorismo etc.

Pero, para preservar a autonomía das persoas, debe diferenciarse entre identificación, seguimento e rastrexo e entre vixilancia selectiva ou masiva, de xeito que a vixilancia deberá estar sempre claramente xustificada, recollida na lexislación existente e ser consentida. As persoas debemos ser coñecedoras en todo momento de se estamos interactuando cun sistema de IA ou cunha persoa real, así como de se se están recompilando os nosos datos. Ademais, debemos poder preguntar esta información en caso de dúbida e dispor de cláusulas claras e transparentes para dar a nosa aprobación.

4 Saúde

Preocupacións

Un uso indebido e desmesurado das tecnoloxías da IA podería levar a un escenario no que a toma de decisións médicas estea baseada exclusivamente en algoritmos da IA, anulando ou substituíndo as decisións de profesionais da medicina polas dos axentes intelixentes, os cales poden estar tomándoas sen ter fundamentos suficientes (por exemplo, sen ter en conta as particularidades de cada paciente).

Oportunidades

A IA pode analizar grandes conxuntos de datos clínicos a nivel mundial, o que permitirá mellorar os diagnósticos, tratamentos e desenvolvemento de medicamentos nun tempo récord. Unha maior e mellor información sobre as condicións climáticas e ambientais tamén pode axudar a previr brotes de enfermidades, a controlar a súa propagación ou a despregar os recursos médicos dunha forma máis eficiente, como vimos no caso da pandemia da Covid 19 ou da varíola do mono. Ademais, a IA ten potencial para axudar co coidado de persoas maiores ou para vixiar as condicións de persoas enfermas en tempo real, así como para personalizar os tratamentos médicos. Polo tanto, un uso adecuado da IA pode axudar a salvar moitas vidas.

5 Transporte e accesibilidade

Preocupacións

Segundo Microsoft, no ano 2025 o 100% dos coches estarán conectados e enviarán, recibirán e analizarán enormes cantidades de datos, o que traslada a este ámbito as preocupacións medioambientais, de privacidade e de vixilancia.

Ademais, co desenvolvemento de servizos ou aparatos que empregan a IA podemos atoparnos con situacións novas non reguladas, por exemplo, quen é responsable dos danos dun accidente onde está implicado un coche autónomo: a persoa propietaria, a empresa que o fabricou o quen o programou? E incluso, recuperando o falado nos debates propostos para a aula, poderíamos chegar a un escenario onde sexa a IA dun coche autónomo a que decida quen morrerá nun accidente de tráfico.

Oportunidades

A IA pode gobernar de sistemas de transporte intelixente, optimizar rutas, posibilitar que as persoas con problemas de visión sexan máis independentes, minimizar atascos, predicir e reducir o número de accidentes de tráfico detectando distraccións, fatigas, avarías etc.

6 Desigualdades entre países.

Preocupacións

Os países empobrecidos corren o risco de verse afectados pola brecha tecnolóxica, económica e social debido á falta de investimento das grandes empresas en infraestruturas tecnolóxicas básicas. Unha brecha que, recordemos, vese ampliada coa incorporación das tecnoloxías de IA.

Oportunidades

A IA podería contribuír a mitigar estas desigualdades, por exemplo, coa análise de datos para mellorar as colleitas e aumentar as ganancias, optimizar o envío de subministracións a persoas refuxiadas que foxen de guerras e outros padecementos ou axudar en desastres naturais, entre outras.

Cómpre sinalar que estas iniciativas non deberían depender só de entidades benéficas ou sen ánimo de lucro como as ONGDs, senón tamén dos gobernos e das empresas tecnolóxicas que deberían financiar proxectos, infraestruturas, industrias, maquinarias, transportes etc. con potencial de crear traballo e capital nos países empobrecidos. Isto traduciríase nun incremento da súa produtividade e, polo tanto, da súa riqueza, repercutindo positivamente na calidade de vida destas poboacións.

7 Diferencias sociais, culturais e de xénero.

Preocupacións

A IA pode perpetuar e incluso ampliar as desigualdades estruturais existentes nas nosas sociedades (xénero, raza, etnia, capacidade etc.) se non se pon especial atención á diversidade no deseño dos algoritmos ou na recompilación de datos.

Historicamente, a IA foi deseñada por e para homes brancos de clase alta, que eran os que ocupaban as posicións de poder e decisión no campo do desenvolvemento tecnolóxico. Esta tendencia continúa hoxe en día cando, por exemplo, só o 22% das persoas que se dedican profesionalmente a IA son mulleres. Derivado disto, carecemos de datos dunha gran parte das poboacións minoritarias e minorizadas, e esta falta de representación ou diversidade pode nesgar de forma intencionada ou involuntaria os datos e os algoritmos que emprega a IA, replicando os negos estruturais da sociedade.

Deste xeito, empregar a IA para tomar determinadas decisións como contratar ou despedir a alguén, ofrecer préstamos ou resolver procesos penais pode supor un risco, xa que estas decisións poden estar influenciadas (nesgadas) por trazos persoais como a etnia, o sexo ou a idade incluídos nos datos que se procesan.

Oportunidades

A IA pode axudar a diminuír e incluso dismantelar as desigualdades estruturais, axudando a crear unha sociedade máis xusta, igualitaria e fraterna, se se garante a participación de grupos diversos na recompilación de datos e no deseño de algoritmos, ao tempo que se pon a mirada xustamente en deseñar algoritmos sen negos. Deste xeito, a IA podería favorecer o respecto a todas as culturas e á diversidade das persoas.

Así mesmo, pódense deseñar sistemas de IA dirixidos especificamente aos grupos máis vulnerables que teñan coma finalidade a redución das desigualdades, en lugar de empregalos para vixiar a estas persoas con fins como a predición de riscos á hora de ofrecer créditos ou a prevención de delitos e fraudes (actuacións que derivan da consideración de que as persoas vulnerables están máis abocadas á delincuencia). Por exemplo, podería usarse a IA para cubrir as súas necesidades básicas ou mellorar as súas condicións de vida con algoritmos que axuden á busca de emprego, a distribución de medicamentos, a xestionar axudas para o consumo eléctrico, a vivenda ou a alimentación.

8 Educación

Preocupacións

Derivado da fenda dixital existente entre países enriquecidos e empobrecidos, a incorporación da IA (e outras tecnoloxías) á educación podería amplificar as desigualdades de acceso e aproveitamento á mesma que existen no mundo. Hoxe en día, o 71% da xuventude do planeta emprega internet no seu día a día, pero aínda hai 346 millóns de rapaces e rapazas que non están conectados á rede. Isto supón que as persoas de entre 15 e 24 anos que viven nas zonas máis vulnerables do mundo vense privadas dunha educación dixital acorde cos

avances tecnolóxicos que se dan noutras partes do planeta.⁹ É dicir, ademais de sufrir as brechas históricas de pobreza e exclusión social, arríscanse a perder o tren da economía dixital.

Por outra parte, a IA pode ofrecer oportunidades de apoio ao persoal docente en certas tarefas educativas e pedagóxicas, pero non deberíamos camiñar cara a un escenario onde a figura docente sexa desprazada, xa que a interacción co alumnado e a colaboración entre docentes é esencial para garantir unha educación de calidade. Se ben os axentes da IA poden estar programados para realizar certas tarefas (por exemplo, corrixir exames), non teñen a capacidade de analizar a inmensa cantidade de estímulos e variables que se poden dar nunha clase e carecen das habilidades emocionais e relacionais que permiten ao profesorado tratar ao alumnado con empatía, comprendendo as súas casuísticas persoais (personalidade, influencia da contorna, momento vital, habilidades e destrezas etc.) e adaptando a docencia as necesidades do grupo e das persoas.

Oportunidades

Introducir as novas tecnoloxías da información e comunicación nos sistemas educativos, así como a educación en liña nas zonas máis vulnerables, illadas ou empobrecidas do planeta, podería aumentar a escolarización, algo imprescindible para que a xuventude poida formar parte da economía dixital e cambiar o futuro dos seus países.

Así mesmo, a IA pode achegar métodos de ensinanza-aprendizaxe personalizados e garantir a inclusión das persoas con capacidades diferenciadas ou necesidades educativas especiais, para que cada alumna e cada alumno desenvolva ao máximo as súas capacidades.

9 Emprego

Preocupacións

Está aumentando o uso de ferramentas asistidas por IA para realizar entrevistas por videoconferencia, avaliacións en rede e toma de decisións para a contratación de persoal. Esta tecnoloxía carece do “ollo clínico”, da empatía e da intuición humana, e dependendo do deseño do algoritmo e dos datos de adestramento utilizados pode contribuír á perpetuación das discriminacións laborais por razón de xénero, raza, etnia nacionalidade, capacidade etc.

Para evitar isto, os procesos de avaliación e de selección deberán ser plenamente transparentes, con información sobre a súa finalidade e a súa metodoloxía. Deberán tamén estar deseñados para impedir a discriminación e garantir a equidade, e, no caso de que a avaliación dea lugar a unha consecuencia negativa, terán que ofrecer mecanismos para impugnar e rectificar dita avaliación.

Por outra parte, os avances en IA na industria poderían eliminar un gran número de postos de traballo, aumentando o desemprego especialmente das persoas menos cualificadas. Porén, estes avances tamén levarán á creación de novos postos de traballo, polo que se pode evitar ese escenario con actuacións dirixidas

⁹ Segundo o informe *“Nenos nun mundo dixital”* de Unicef.

á educación e á formación, que terán un papel fundamental para previr o desemprego e crear man de obra cualificada.

Oportunidades

Incorporar as entrevistas dixitais e as ferramentas de IA nos procesos de selección e avaliación pode ser positivo en múltiples aspectos:

- Eliminar tarefas repetitivas na análise de currículos.
- Permitir unha comparación máis áxil das capacidades das persoas candidatas.
- Favorecer unha maior conectividade entre persoas candidatas e empregadoras, evitando desprazamentos e aforrando custos e tempo.
- Minimizar ou eliminar a subxectividade e os posibles nesgos da persoa entrevistadora.

Ademais, a IA poderá empregarse para detectar as necesidades de formación e capacitación do persoal traballador para dar resposta ás futuras necesidades da industria.

10 Alimentación e agricultura

Preocupacións

O uso da IA na agricultura pode ter riscos para a seguridade alimentaria. As novas tecnoloxías deben ser probadas axeitadamente en contornas experimentais que garantan a seguridade e protección contra fallos accidentais e consecuencias non desexadas. Doutro xeito poderían darse escenarios nos que a IA tomase decisións tendentes a aumentar a produtividade que non tiveran en conta o impacto ambiental ou a seguridade alimentaria.

Por outra parte, a brecha dixital é moi evidente no sector da agricultura, onde os pequenos produtores e produtoras están en desvantaxe fronte as grandes corporacións e empresas á hora de acceder ao comercio electrónico, aos rexistros de transaccións con tecnoloxía de cadea de bloques (trazabilidade), ao control de pragas, á xenética dos cultivos, á alerta temperá ante ameazas de seguridade alimentaria etc.

Oportunidades

A IA podería empregarse para construír un sistema alimentario sostible e menos malgastador axudando a minimizar o uso de fertilizantes e pesticidas, a optimizar o rego, a conservar a saúde do solo, a mellorar o benestar animal, a identificar e mellorar especies, a mellorar a produtividade (duplicándoa incluso) e a reducir o impacto ambiental, así como a transformar os sistemas agroalimentarios co fin de erradicar a fame e reducir a pobreza.

Deste xeito, a IA pode converterse nunha potente ferramenta para facer fronte aos grandes desafíos da seguridade alimentaria mundial: a necesidade de alimentar á crecente poboación mundial, o cambio climático, as demandas de terra para a produción de biocombustibles e a diminución da cantidade e calidade do solo de cultivo.

11 Armamento

Preocupacións

Por unha parte, a IA podería empregarse para articular estratexias de defensa e ataque baseadas na piratería informática e no *phishing* ou para atacar sistemas clave na guerra cibernética.

Por outra, na actualidade xa atopamos países e industrias que están investigando e desenvolvendo armas letais autónomas (mísiles con brancos selectivos e máquinas con aptitudes cognitivas e de aprendizaxe capaces de decidir cando, onde e contra quen combater sen intervención humana). Isto presenta problemas dende o punto de vista ético, xa que pode dar lugar a carreiras de armamento incontrolables e crear contextos militares sen control humano.

Oportunidades

Prevese que a IA se empregue na prevención de delitos e na aplicación do sistema xudicial, por exemplo, avaliando con maior precisión os riscos de fuga de prisioneiros e prisioneiras ou predicindo e previndo ataques terroristas, xa que esta tecnoloxía pode procesar conxuntos de datos máis rápido que as persoas. De feito, hoxe en día as plataformas en liña xa empregan a IA para detectar e reaccionar ante comportamentos ilegais e desapropiados e no campo militar pode empregarse xa para detectar e atacar sistemas clave na chamada guerra cibernética.

Para que só as oportunidades se volvan unha realidade e a IA beneficie realmente a todas as persoas, os gobernos, os líderes industriais, as institucións educativas e os sindicatos teñen a responsabilidade de:

- Axudar á cidadanía a realizar a transición cara a unha etapa dixital, ofrecendo seguridade e garantido o mantemento dos dereitos individuais e colectivos.
- Ofrecer a formación e a cualificación que requirirán os postos de traballo do futuro.
- Deseñar un marco educativo para quen diseña, comercializa, aplica, mantén e emprega ferramentas de IA.
- Ofrecer vías de reciclaxe profesional para aquelas persoas que traballan hoxe en empregos que quedarán obsoletos.
- Desenvolver marcos lexislativos comúns, aplicables e transparentes que garantan os dereitos das persoas e limiten as accións daniñas de empresas ou institucións.
- Desenvolver acordos marco que supervisen e limiten o desenvolvemento de tecnoloxía que non se axuste a principios éticos (incluído pero non limitado ao armamento).
- Desenvolver medidas tendentes a reducir e non a ampliar a brecha entre países empobrecidos e países enriquecidos.

Namentres, dende as nosas aulas podemos ir formando ao noso alumnado para que desenvolvan unha conciencia crítica cara a IA que lles permita desenvolverse nun mundo cada vez máis dixital dunha forma segura, respectuosa e o máis igualitaria posible.



En resumo:

- O emprego masivo dos sistemas da IA abre cuestións éticas e legais que deben ser afrontadas polas sociedades. Por iso, é fundamental que o alumnado coñeza e reflexione sobre as consecuencias positivas e negativas que a IA pode ter na sociedade e na vida das persoas.
- A IA non ten ética, conciencia nin empatía. O seu comportamento depende da ética de quen a desenvolve e adestra e dos datos que recibe despois, polo que é necesario aplicar principios éticos ao desenvolvemento e adestramento das IA e monitorizar, comprobar e probar os seus resultados.
- A IA pode contribuír a un desenvolvemento sostible, sempre que se programe para eses fins e se empreguen enerxías renovables para atender as altas necesidades enerxéticas destas tecnoloxías.
- Actualmente (ano 2022) a lexislación concreta sobre IA é moi escasa, sendo a maioría de ámbito supranacional (Unión Europea). Urxe que os estados desenvolvan lexislacións nacionais e internacionais ao respecto.
- A IA ofrece moitas oportunidades para o futuro, pero tamén grandes preocupacións. Malia que a responsabilidade de facer fronte a estes retos recae en grandes líderes mundiais, dende as aulas podemos contribuír a que o noso alumnado poida desenvolverse da forma máis segura, respectuosa e igualitaria posible formándoo en IA en todas as súas dimensións (programación, usabilidade, ética etc.).

4. REMATANDO

Como vimos nesta guía, incorporar a IA na aula é fundamental porque *A Intelixencia Artificial está cambiando o mundo*.

É imprescindible que o alumnado coñeza os aspectos técnicos da IA pero tamén o impacto social que ten e a forma en que muda as nosas vidas, pois a IA xoga xa un papel importante nas nosas sociedades, aínda que non sempre sexamos conscientes diso.

É fundamental que o alumnado adquira un pensamento crítico para facer un uso responsable da tecnoloxía, ben dende a programación e o desenvolvemento, ben como persoas usuarias e traballadoras, ben como cidadás e cidadáns conscientes dos seus dereitos nun sistema cada vez máis dixital.

A nosa proposta é levar á aula os coñecementos sobre IA, dende o máis básico como *que é a IA* ata os aspectos máis complexos desenvolvidos nos cinco *bloques básicos dun sistema de IA*, para que, ao rematar a educación regrada, todo o alumnado teña coñecementos suficientes para formar parte da sociedade dixital cara a que camiñamos.



GLOSARIO

► **Actuador**

Elemento que realiza unha acción.

► **Adestramento**

Construción dun modelo partindo duns datos que imos clasificando como correctos e incorrectos.

► **Algoritmo**

Conxunto ordenado de regras que permiten solucionar un problema.

► **Aprendizaxe automática (Machine Learning)**

Campo da IA no que as persoas adestran ás computadoras para recoñecer patróns baseados en datos e facer as súas predicións, o que proporciona ás computadoras a capacidade de aprender.

► **Aprendizaxe non supervisada**

Aprendizaxe automática na que o obxectivo do algoritmo non está definido de antemán.

► **Aprendizaxe por reforzo**

Aprendizaxe automática na que a computadora aprende en base a recompensas e penalizacións.

► **Aprendizaxe profunda (Deep Learning)**

Forma parte da aprendizaxe automática (Machine Learning) e consegue, a partir de moitos datos e moitas capas de procesamento con algoritmos, que unha computadora aprenda por si mesma realizando tarefas coma as dos seres humanos.

► **Aprendizaxe supervisada**

Aprendizaxe automática na que o obxectivo do algoritmo está definido de antemán.

► **Árbore de decisión**

Estrutura similar a un diagrama onde os nodos representan unha característica, as ramas representan unha regra de decisión e cada folla representa o resultado.

► **Autonomía**

Capacidade de actuar de forma independente ou sen control externo.

► **Axente de IA**

Ente que recibe información da súa contorna a través dos sensores, procesa esa información e responde ou actúa dun xeito racional e autónomo.

► **Base de datos**

Almacenamento organizado de datos, tenden a mostrarse nunhas táboas con filas e columnas.

► Ben social

Algo que beneficia á maior cantidade de persoas da mellor maneira posible.

► Característica

Propiedade ou calidade.

► Chatbot

Programa informático que simula e procesa conversacións humanas, escritas ou faladas, permitindo ás persoas interactuar con dispositivos dixitais coma se estivesen comunicándose cunha persoa real.

► Datos

Información.

► Datos de adestramento

Exemplos empregados para adestrar un modelo de aprendizaxe automática.

► Diversidade

Variedade, diferenza, gran cantidade de obxectos ou persoas diferentes.

► Díxitos

Números que se representan nunha sola cifra, en informática empréganse 0 e 1.

► Estereotipo

Idea ou imaxe mental de alguén ou algo.

► Ética

Conxunto de principios e fundamentos das nosas normas morais, que rexen o comportamento dunha persoa.

► Ético

Moralmente correcto.

► Impacto social

Resultado ou consecuencia dunha acción nunha comunidade.

Intelixencia Artificial

Ciencia e enxeñería para crear programas informáticos que poidan imitar a intelixencia humana.

► Modelo (de Aprendizaxe Automática)

Representación dun conxunto de datos desenvolvida pola IA.

► Moral

Conxunto de costumes e normas que se consideran “boas” para dirixir ou xulgar o comportamento das persoas nunha comunidade.

► Nesgo

Preferencia a favor ou en contra dunha idea ou obxecto.

► **Nodo (árbore de decisión)**

Parte dunha árbore de decisión que representa unha característica.

► **Percepción**

Organización e interpretación das sensacións.

► **Procesamento da linguaxe natural (NLP)**

Tecnoloxía de IA utilizada para comprender e interactuar coa linguaxe humana.

► **Programación**

Conxunto de regras e datos ordenados e estruturados para obter unha acción ou resultado.

► **Radar (Radio Detection and Ranging)**

Sistema que emprega ondas electromagnéticas para medir distancias, velocidades... O radar emite unha sinal electromagnética que se reflexa no obxecto que se quere medir e volta ao emisor, proporcionando información, por exemplo, sobre a localización e a velocidade.

► **Rama**

Parte dunha árbore de decisión que representa as posibles opcións dun nodo.

► **Rastro de datos**

Conxunto de datos de información persoal que son seguidos utilizando as accións dunha persoa en liña.

► **Razoamento**

Proceso lóxico de sacar conclusións ou facer predicións coa axuda do coñecemento.

► **Recoñecemento de imaxes**

Capacidade dun programa informático para analizar píxeles nunha imaxe e identificar obxectos, persoas...

► **Rede neuronal artificial (ANN)**

Tipo de algoritmo que intenta simular o que se cre que é o funcionamento das neuronas do noso cerebro.

► **Robot**

Máquina capaz de realizar tarefas complexas automaticamente.

► **Robot con Intelixencia Artificial**

Robots que usan sensores para recoller información e toman decisións autónomas sobre como realizar unha tarefa, incluso nun entorno cambiante.

► **Sensor**

Dispositivo que permite detectar un cambio na contorna e transformalo nunha magnitude eléctrica.

► **Sentido**

Facultade, coma a vista, o oído ou o tacto, que usan as persoas ou os animais para percibir información.

► Sistema de recomendación

Programa informático que realiza suxestións sobre produtos, servizos e outros, que poden gustarlles ás persoas usuarias, baseadas nos rastros de datos das persoas usuarias en rede.

► Sonar (Son Navegación Rango)

Sistema que emprega impulsos sonoros para saber a posición dun obxecto baixo a auga. O sonar mide o tempo que tarda a onda sonora en saír, golpear e rebotar no obxecto, e así calcula a distancia.

► Sostibilidade

Satisfacción das necesidades actuais sen comprometer a capacidade das xeracións futuras de satisfacer as súas, garantindo o equilibrio entre crecemento económico, coidado do medio ambiente e benestar social.

► Ultrasón

O sensor de ultrasón é un detector de proximidade que emite un son a través do aire que rebota co obxecto do que se quere saber a distancia, mide o tempo que tarda en recibir a onda reflexada e calcula a distancia.