



Trustworthy **AI**

Hacia una IA Fiable

Guía del profesor



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

Índice de contenidos

Índice de contenidos.....	2
Enseñanza sobre la IA Fiable	5
¿Qué es la IA?	6
Problemas técnicos, jurídicos, éticos y sociales	7
La IA Fiable	9
¿Por qué enseñar sobre la IA Fiable?	10
IA Fiable para Estudiantes de STEM	11
IA Fiable para Estudiantes de Administración de Empresas	11
IA de confianza para Estudiantes de Ciencias Políticas	11
Recursos Educativos	13
Clips de Conocimiento	13
Cartas de la AI Fiable	14
Ejercicios discusión	15
Las Partes Interesadas	16
7 pasos hacia una IA Fiable	17
Cómo utilizar el ejercicio	17

Nota sobre los derechos de autor:

Este material se presenta para garantizar la difusión de trabajos académicos y técnicos. Los derechos de autor y todos los derechos sobre los mismos pertenecen a los autores o a otros titulares de derechos de autor. Se espera que toda persona que copie esta información se adhiera a los términos y restricciones invocados por los derechos de autor de cada autor. En la mayoría de los casos, estas obras no pueden volver a publicarse sin el permiso explícito del titular de los derechos de autor.

Titulares de la producción intelectual: Catelijne Muller, (ALLAI)

Revisión del documento: todos los socios de Trustworthy AI

Versión	Por	Cambios
0.1	Catelijne Muller	Versión inicial

Sección 1

Introducción

Enseñanza sobre la IA Fiable

Esta guía pretende ayudar a los profesores a familiarizarse con la Inteligencia Artificial (IA) Fiable, basándose en las Directrices Éticas de la UE para la IA Fiable, e implementar Recursos Educativos sobre la IA Fiable en sus planes de estudio. La guía explica la IA como tecnología, su impacto ético, legal y social, el concepto de IA Fiable y la importancia de enseñar a los estudiantes sobre la IA Fiable, incluso cuando el ámbito educativo pueda parecer ajeno al principio.

La guía incluye tres recursos educativos dedicados a la IA Fiable. Estos recursos pretenden ayudar a los estudiantes a familiarizarse con las Directrices Éticas para una IA Fiable y a desarrollar la capacidad de comprender, evaluar críticamente, debatir y aplicar la IA Fiable de forma práctica.

Los recursos educativos abiertos incluyen: 8 videoclips cortos de conocimiento, una baraja de cartas, y un ejercicio de 7 pasos hacia la IA Fiable. Esta guía explicará cómo utilizar estos recursos, sus objetivos de aprendizaje y su importancia para los estudiantes de diferentes campos de la educación.

En esta guía te encontrarás con:

- ◇ ¿Qué es la IA y cuáles son sus implicaciones sociales y éticas?
- ◇ ¿Qué es la IA fiable y cómo podemos realizarla?
- ◇ ¿Por qué debería enseñar esto a sus alumnos y qué relevancia tiene este tema en su campo?
- ◇ ¿Cómo utilizar los diferentes recursos educativos?
- ◇ Más información sobre la IA con material de referencia adicional y nuestro glosario

Esta guía y los Recursos Educativos tienen como objetivo apoyar la educación práctica de las implicaciones éticas y sociales de la IA basada en los "7 requisitos para una IA Fiable" que forman parte de las Directrices Éticas para una IA Fiable. Van más allá de la educación ética general, sino que se limitan a ofrecer un enfoque práctico de la ética de la IA. Como tal, la guía y los Recursos Educativos pretenden ayudar a los educadores a tener más confianza en la enseñanza de la IA Fiable y estimular un debate significativo y ético entre los estudiantes.



¿Qué es la IA?

No existe una única definición aceptada y rígida de la IA. La IA es un término que engloba un gran número de subcampos, como la computación cognitiva, el aprendizaje automático, el aprendizaje profundo, el procesamiento del lenguaje natural, el reconocimiento de patrones, etc. El objetivo central de la investigación y el desarrollo de la IA es, sin embargo, automatizar el comportamiento inteligente, como el reconocimiento de imágenes, la recopilación de información, la planificación, la comunicación, la manipulación, la detección y la predicción. En 2019, el Grupo de Expertos de Alto Nivel de la Comisión Europea definió la IA como:

"Sistemas que muestran un comportamiento inteligente analizando su entorno y emprendiendo acciones -con cierto grado de autonomía- para alcanzar objetivos específicos".

Recientemente se han hecho muchos progresos en el campo de la llamada IA estrecha ("Narrow AI"). La IA estrecha es capaz de realizar tareas únicamente específicas, al contrario que la IA general ("General AI"), que sería capaz de realizar cualquier tarea mental que pueda llevar a cabo un ser humano. La IA general no existe y los científicos no se ponen de acuerdo sobre si se conseguirá algún día. Sobre todo con el crecimiento de la potencia de procesamiento de los ordenadores, la disponibilidad de grandes volúmenes de datos y el desarrollo del Machine Learning (ML). Machine Learning se refiere a la IA que se basa en el proceso de "datos de entrenamiento" a partir de los cuales un algoritmo aprende a reconocer patrones y a elaborar reglas.

Los algoritmos pueden entenderse como una colección de instrucciones que llegan a un resultado específico. Podemos comparar un algoritmo con una receta para una pastel de manzana: una lista de instrucciones sobre cómo hornear una pastel de manzana. Sin los ingredientes y sin que alguien que siga las instrucciones, la receta nunca se convertirá en una pastel. Lo mismo ocurre con los algoritmos. Sin datos (ingredientes) y sin modelo (alguien que siga la receta), el algoritmo no producirá ningún resultado. Y al igual que con un pastel, el tipo de ingredientes y las habilidades del cocinero pueden marcar la diferencia en el resultado final, con la IA, el tipo de datos y las decisiones tomadas en torno al propósito del sistema de IA marcan la diferencia en el resultado final.

Muchos sistemas de IA se basan en software y actúan en el mundo virtual como, por ejemplo, los asistentes de voz, los sistemas de recomendación, o softwares de análisis de imágenes. Pero la IA también puede integrarse en dispositivos de hardware, por ejemplo, en robots, coches autónomos, drones o en el Internet of Things (IoT) (Comisión Europea, 2019).

Aunque la IA se puede emplear a través de diferentes técnicas, su principal objetivo y fuerza hoy en día reside en las técnicas relacionadas con el reconocimiento de patrones y el análisis de datos. Al entrenar el sistema con grandes conjuntos de datos, adquiere "conocimiento" de los patrones que se pueden encontrar en los datos. Este conocimiento se puede aplicar a nuevos datos que el sistema no ha visto antes y se puede hacer una predicción sobre los datos desconocidos.



Problemas técnicos, jurídicos, éticos y sociales

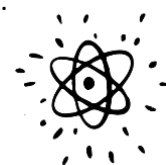
La capacidad de la IA para realizar tareas de forma automatizada y su potencial para ser aplicada en todos los ámbitos de la sociedad hacen que la tecnología sea muy atractiva, pero también conlleva retos. Técnicamente, la IA tiene todavía muchos puntos débiles. Como se puede ver en su método, la IA no tiene sentido común como los humanos. No entiende el contexto ni el significado de los datos que procesa, solo utiliza técnicas de estadística para hacer predicciones. Hay que entrenar a un sistema de IA con muchos ejemplos antes de que pueda reconocer algo tan simple como un gato. Si se le muestra una imagen en el que el gato está al revés, es probable que no la reconozca como tal. El concepto general de gato no es algo que la IA pueda aprender. Además, el rendimiento de un sistema de IA sigue dependiendo en gran medida de los humanos. Los humanos tienen que procesar y etiquetar los datos de forma que la IA pueda aprender de ellos, lo que da lugar a resultados que reflejan la forma en que interpretamos los datos.

El potencial de la IA tiene doble cara: tiene la capacidad de beneficiar a la sociedad tanto afectarla negativamente. Cuando se desarrolla y se emplea de forma responsable, la IA puede hacerse cargo de trabajos peligrosos, pesados o monótonos y aumentar la eficiencia de los humanos. También puede apoyarnos a la hora de afrontar grandes problemas como el cambio climático, la injusticia social, o descubrir información oculta en todo tipo de datos que nos permite tomar decisiones informadas y seguir innovando. Sin embargo, la IA también conlleva retos relacionados con la seguridad, equidad, privacidad, autonomía, responsabilidad y legislación. Por eso, con cada sistema de IA que diseñemos o desplaguemos debemos asegurarnos de que su uso es seguro (incluso en situaciones imprevisibles), inclusivo, ético, que garantiza nuestros derechos fundamentales y mucho más.

Es obvio que la IA puede causar daños cuando se utiliza para prácticas ilegales. Puede verse como una nueva herramienta para cometer delitos. Sin embargo, también presenta riesgos cuando se utiliza para fines aparentemente inocentes. Un ejemplo del riesgo al que probablemente todos hemos sido sometidos es el aumento de las noticias falsas y la clasificación de las personas en burbujas de filtros. Las plataformas de medios de comunicación, como Facebook, utilizan sistemas de recomendación diseñados para maximizar el tiempo que uno pasa en las plataformas. De este modo, sólo se nos muestran artículos parecidos a las publicaciones que han captado previamente nuestra atención, y que confirman o intentan cambiar nuestras creencias sobre determinados temas. Por lo tanto, las personas están expuestas a un flujo de información más estrecho que distorsiona la realidad y aumenta aún más la división social. Estos sistemas de IA influyen lo que la gente ve en Internet, lo que a su vez influye en sus decisiones en la vida. De este modo, la IA supone una amenaza para nuestro objetivo de mantener una sociedad democrática, abierta y justa.

Las burbujas de filtros son el resultado de organizaciones comerciales que quieren maximizar sus beneficios. Pero incluso cuando pensamos que estamos utilizando la IA para el bien, existen grandes riesgos de los que debemos ser conscientes. Por ejemplo, la IA se utiliza cada vez más con fines de vigilancia y control para prevenir delitos y hacer del mundo un lugar más seguro y justo. Sin embargo, la vigilancia (desproporcionada) provoca el llamado "efecto escalofriante", donde las personas se sienten constantemente vigiladas y causan adaptaciones del comportamiento. Esto socava, entre otras cosas, nuestro derecho a la libertad y el respeto a la dignidad humana.

Otro ejemplo del daño accidental es que algunos sistemas acaban produciendo sesgos o inexactitudes. Los fallos de un sistema o el sesgo de un conjunto de datos utilizado para el entrenamiento de un sistema pueden conducir a resultados injustos y discriminatorios. Aunque los desarrolladores no tengan la intención de discriminar, un sistema puede ampliar el sesgo subyacente hacia ciertos grupos de personas presentes en los datos que se utilizan. Otros usos potencialmente problemáticos de la IA son, el reconocimiento biométrico, la puntuación social o el uso de armas autónomas. Todas estas aplicaciones de la IA suponen grandes riesgos para nuestra seguridad y derechos fundamentales.



Sección 2

IA Fiable

La IA Fiable

La IA tiene el potencial de fortalecer la sociedad y apoyar a los seres humanos en diferentes frentes. Sin embargo, debemos implementar cuidadosamente los sistemas de IA para proteger nuestros derechos fundamentales, nuestra salud y nuestra seguridad. Desarrollar, desplegar y utilizar la IA de forma centrada en el ser humano puede aportarnos una IA segura, protegida, justa e inclusiva. Para lograrlo, el Grupo de Expertos de Alto Nivel en Inteligencia Artificial elaboró las Directrices Éticas para una IA Fiable¹. Como base para estas Directrices, el grupo eligió tres pilares que deben sustentar la IA para que sea Fiable:

1. debe ser legal, lo que debe cumplir con todas las leyes y reglamentos aplicables
2. debe adherirse a los principios y valores éticos, y
3. debe ser sólida desde el punto de vista técnico y social y no causar daños involuntarios

Para los pilares 2 y 3, el grupo elaboró **7 requisitos para una IA Fiable** que debería guiar el desarrollo, el despliegue y el uso de la IA en Europa:



1. **Agencia Humana y Supervisión:** Los sistemas de IA deben apoyar la autonomía humana y permitirles tomar decisiones informadas. Para lograrlo, los sistemas de IA deben actuar como facilitadores de una sociedad democrática y equitativa, apoyando la agencia del usuario, fomentando los derechos fundamentales y permitiendo la supervisión humana. Hay que enseñar a los estudiantes los niveles adecuados de agencia y autonomía humanas, el control humano, y la dignidad humana en general.



2. **Robustez técnica y seguridad:** Este principio exige que los sistemas de IA se desarrollen con un enfoque preventivo de los riesgos y de manera que se comporten de forma fiable según lo previsto, minimizando los daños involuntarios e inesperados y evitando los daños inaceptables. Además, debe garantizarse la integridad física y mental de los seres humanos. Los estudiantes deben aprender a reconocer y garantizar la precisión y fiabilidad de los sistemas de IA. Además, los estudiantes deben saber cómo equilibrar la solidez técnica y las limitaciones éticas.



3. **Privacidad y gobernanza de los datos:** La privacidad es un derecho fundamental especialmente afectado por los sistemas de IA. La prevención del daño a la privacidad debe ser una prioridad. Este requisito requiere, lógicamente, una gobernanza de datos adecuada. Esto abarca la calidad e integridad de los datos utilizados, su relevancia considerando el ámbito en el que se desplegarán los sistemas de IA, sus protocolos de acceso y la capacidad de procesar los datos de manera que se proteja la privacidad. Hay que enseñar a los estudiantes a recopilar y reconocer datos de alta calidad, a manejarlos con sensibilidad, a mantener la privacidad y a evitar sesgos en los datos y en los modelos contruidos a partir de ellos.



4. **Transparencia:** Este requisito tiene dos partes. Abarca la transparencia de los elementos relevantes para los sistemas de IA; los datos recogidos, la formación y el funcionamiento del sistema, explicaciones de resultados y los modelos comerciales pertinentes. También abarca la obligación de ser transparente sobre el uso de los sistemas de IA y no utilizarlos de forma encubierta. Los estudiantes deben reconocer sistemas transparentes, y adquirir habilidades para desarrollar una IA explicable. Esto implica enseñar a los estudiantes a documentar y comunicar adecuadamente el uso de los datos, así como las decisiones tomadas en el proceso de diseño.

¹ Directrices éticas para una IA digna de confianza, Grupo de Expertos de Alto Nivel en Inteligencia Artificial, 2019



5. **Diversidad, no discriminación y equidad:** Los resultados de los sistemas de IA deben ser no discriminatorios y estar libres de prejuicios inaceptables. Otra parte importante de este requisito es la igualdad de acceso mediante procesos de diseño inclusivos. La inclusión y la diversidad deben estar presentes durante todo el ciclo de vida del sistema de IA. Esto incluye la consideración y la participación de todas las partes interesadas afectadas a lo largo del proceso. Los estudiantes deben aprender sobre la importancia y el valor añadido de la experiencia interdisciplinaria a la hora de desarrollar, desplegar y utilizar los sistemas de IA. También se les debe enseñar los posibles efectos discriminatorios de las decisiones tomadas a lo largo del proceso de desarrollo.



6. **Bienestar medioambiental y social:** El medio ambiente y la sociedad en su conjunto deben considerarse "partes interesadas" durante el ciclo de vida del sistema. Este requisito incluye el fomento de la sostenibilidad y la responsabilidad ecológica. Implica tanto la investigación de soluciones de IA que aborden el cambio climático u otras preocupaciones sociales, como ser conscientes de la huella ecológica de la formación y el despliegue de un sistema de IA. También implica comprender y mitigar los efectos sociales, democráticos o sistémicos más amplios que puede tener la IA.



7. **Responsabilidad:** Este requisito requiere que se establezcan mecanismos para garantizar la responsabilidad y la rendición de cuentas de los sistemas de IA y sus resultados, tanto antes como después de su desarrollo, despliegue y uso. Se debe enseñar a los estudiantes sobre la auditoría y el mantenimiento de registros, así como los marcos legales para la responsabilidad y poder demostrar la minimización de los efectos negativos.

Estos requisitos deben abordarse y evaluarse continuamente a lo largo del ciclo de vida del sistema de IA -desde la fase de diseño y desarrollo hasta el final de su uso-, teniendo en cuenta métodos tanto técnicos como no técnicos para garantizar su cumplimiento. De este modo, se puede fomentar y garantizar una IA ética y sólida.

¿Por qué enseñar sobre la IA Fiable?

Es probable que muchos estudiantes participen en el desarrollo, la implantación, la adquisición o el uso de la IA en sus futuros trabajos. Por lo tanto, es importante que adquieran las habilidades necesarias para desarrollar e implementar la IA de manera responsable y fiable. Por ello, hemos transformado los 7 requisitos para una IA Fiable en ejercicios específicos y recursos educativos que ayudan a enseñar habilidades específicas, entre ellas la capacidad de:

1. Identificar la aplicabilidad de los 7 requisitos en diferentes contextos y sus diferentes dimensiones para las diferentes partes interesadas
2. Deliberar sobre las posibles aplicaciones de los 7 requisitos
3. Seleccionar y aplicar un curso de acción en respuesta a un análisis ético sobre los requisitos



Al adquirir estas habilidades y desarrollar la capacidad de comprender y actuar de acuerdo con los valores éticos y sociales, los estudiantes pueden garantizar un enfoque "humano al mando" de la IA, en el que sigue dependiendo de ellos decidir si, cuándo y cómo deben desarrollarse, desplegarse y utilizarse los sistemas de IA. Dado que la responsabilidad de una IA Fiable no recae sólo en los desarrolladores, sino en diferentes partes interesadas y en expertos de otros campos, la enseñanza a los estudiantes de diferentes ámbitos es de gran importancia.



IA Fiable para Estudiantes de STEM

Es fácil imaginar que los alumnos que estudian asignaturas STEM (matemáticas, biología, física, química, etc.) pueden acabar formando parte del desarrollo de la IA. A menudo surgen debates en torno a las consecuencias no deseadas de la IA, el sesgo algorítmico, la recogida y protección de datos, etc. Muchos de estos problemas surgen durante el desarrollo. Por eso es importante que los estudiantes de STEM, los "futuros desarrolladores de IA", sepan apreciar los valores éticos y que los apliquen en las diferentes etapas del desarrollo. Esto les permitirá identificar en una fase temprana las consecuencias imprevistas o relaciones causales entre determinadas opciones de desarrollo y sus problemas éticos. Por ejemplo, pueden asegurarse de que la recogida de datos se haga de forma adecuada y responsable, minimizando el riesgo de resultados sesgados.

Todo desarrollador/a de IA o investigador/a de IA estudiará una o más materias STEM en algún momento de su vida. Por lo tanto, es importante que la ética esté bien integrada en su carrera educativa. No sólo para prevenir las consecuencias imprevistas del desarrollo, sino para adquirir conocimientos fiables que permitan tomar decisiones informadas.

IA Fiable para Estudiantes de Administración de Empresas



La enseñanza de la IA de confianza en el ámbito de la Administración de Empresas es importante dado el crecimiento exponencial de la IA en entornos empresariales, su uso comercial, y el rápido crecimiento de las empresas tecnológicas. Muchos consumidores están expuestos a la IA a través de chatbots de atención al cliente, sistemas de predicción/recomendaciones que toman decisiones sobre ellos, o dispositivos IoT o IoB (Internet of Bodies) que utilizan la IA para su propio funcionamiento. Con mayor frecuencia, el reclutamiento, la contratación y la evaluación de los trabajadores dentro de las organizaciones se realizan con IA. Además, las empresas pueden beneficiarse de las del análisis de datos que la IA puede ofrecer. La IA puede ser muy beneficiosa para la logística, por ejemplo, en ayudar a determinar el mejor sistema organizativo para los vuelos. También puede ayudar en el sector sanitario revisando los historiales médicos, los enfoques de los tratamientos y/o complementar los conocimientos de los médicos. El aumento del uso de la IA en la vida cotidiana de las personas también hace que sea atractiva para añadir la IA a la línea de productos de las empresas.

Aunque es evidente que la IA ofrece muchas oportunidades, las empresas deben tener cuidado con los riesgos que llevan falta de seguridad, los sesgos e injusticias, pero también el desplazamiento masivo de puestos de trabajo o el impacto negativamente en la sociedad o la democracia. No hay que pasar por alto estos riesgos sólo por las posibilidades económicas que puede ofrecer la IA.

IA de confianza para Estudiantes de Ciencias Políticas



Como hemos comentado hasta ahora, no basta con que la IA se desarrolle de forma fiable. También es crucial que se despliegue de forma responsable. Por desgracia, no podemos depositar toda nuestra confianza en las empresas y los desarrolladores para que esto ocurra. Las políticas y los marcos legislativos establecen los límites dentro de los cuales la IA puede desarrollarse y utilizarse de forma aceptable. Por eso es importante que los estudiantes de ciencias políticas reciban también formación sobre la IA fiable. Si asumimos que muchos de estos estudiantes se convertirán en responsables políticos o tendrán un papel en la gobernanza, es esencial que entiendan las verdaderas capacidades y limitaciones de la IA y cómo la política y la legislación pueden establecer los límites adecuados para la IA.

Sección 3

Recursos Educativos

Recursos Educativos

Los dilemas éticos suelen llevar a la confusión sobre cómo debe desarrollarse y utilizarse la IA de forma fiable. Para ayudar a los estudiantes a adquirir conocimientos al respecto, se han desarrollado los siguientes recursos educativos abiertos.

- ◇ Clips de conocimiento que introducen la IA fiable y sus 7 requisitos
- ◇ Cartas IA Fiable
- ◇ Ejercicio de 7 pasos para IA Fiable

Clips de Conocimiento

El objetivo de los clips de conocimiento es ofrecer una visión general de las Directrices Éticas de la IA Fiable y sus siete requisitos. Los clips pueden utilizarse en una conferencia, como recurso educativo independiente o integrado como parte de un curso.

Los clips constan de 8 vídeos (un vídeo introductorio y 7 vídeos de requisitos) que duran entre 2 y 4 minutos cada uno. Cada vídeo de requisitos incluye una explicación del requisito y de los principios que conlleva, un ejemplo real de aplicación de dicho requisito y termina con preguntas pertinentes sobre el mismo.



Cartas de la AI Fiable

El objetivo de esta baraja de cartas es crear un debate ético significativo en clase y comprender la diversidad de aplicaciones de la IA en diferentes ámbitos. Con esta baraja, los alumnos comprenderán la complejidad de algunos dilemas éticos y su relación con los puntos de vista de las diferentes partes interesadas. Las tarjetas ofrecen una explicación general de las técnicas de IA que pueden mezclarse y combinarse en diferentes ámbitos para crear un caso de uso. La falta de familiaridad con las técnicas de IA puede provocar ciertas dificultades con algunos ejercicios, por lo que también hemos creado una baraja de tarjetas de casos de uso preseleccionados para facilitarlos. También somos conscientes de que no todas las técnicas de IA están presentes en esta baraja. Las técnicas y los enfoques de la IA (y sus nombres) evolucionan constantemente y se están desarrollando conforme hablamos. Por lo tanto, por ahora, simplemente hemos presentado un conjunto de técnicas de IA fáciles de entender para que puedan utilizar por personas de diferentes ámbitos.

Antes de utilizar la baraja de cartas, aconsejamos que los alumnos vean los clips de conocimiento para familiarizarse con el concepto de IA Fiable y sus 7 requisitos. No obstante, cualquier conocimiento adicional sobre la IA será útil para plantear los casos de uso necesarios para algunos de los ejercicios. Puede obtener una visión de las técnicas generales de IA en el siguiente enlace: [¿Qué es la Inteligencia Artificial? En 5 minutos](#) (en inglés)

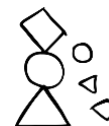
Composición de la baraja

La baraja consta de 5 conjuntos de cartas:

Con estas cartas los estudiantes pueden inspirarse de casos de uso para pensar en cómo utilizar diferentes técnicas de IA en diferentes ámbitos. Utilizando las cartas de requisitos, pueden analizar críticamente las técnicas y los casos de uso desde el punto de vista de diferentes partes interesadas. Todo esto puede hacerse a través de una serie de ejercicios de clase que se explica a continuación.

- ◇ **Técnica de IA:** Una técnica de IA genérica que puede utilizarse en muchos ámbitos de diversas maneras
- ◇ **Ámbito:** Un sector en el que se puede aplicar una técnica de IA que contiene múltiples subsectores pertenecientes al mismo.
- ◇ **Requisito:** Un requisito ético que evalúa la fiabilidad de una técnica de IA aplicada en un ámbito específico. Estas tarjetas también pueden utilizarse para evaluar las cartas de casos de uso.
- ◇ **Parte interesada:** Papel de una persona en el desarrollo o despliegue de un sistema de IA que tiene su propio "interés competitivo", por ejemplo, el dinero, la eficiencia, la seguridad, la equidad, la privacidad, la autonomía, etc.
- ◇ **Caso de uso:** Un ejemplo del funcionamiento y los objetivos de una técnica de IA aplicada a un ámbito en el que intervienen varias partes interesadas.

Ejercicios discusión



Ejercicio 1 - Defensa su caso de uso

Materiales: Cartas de ámbito, cartas de técnica, cartas de partes interesadas, cartas de requisitos

6 jugadores: 1 juez, 5 desarrolladores

1. Se roba una carta del mazo de ámbito. Este será el ámbito en esa ronda.
2. Cada jugador roba y muestra una carta de interesado.
3. Cada jugador roba una carta de técnica sin mostrarla.
4. Cada jugador diseña un caso de uso de la IA (sobre el papel) dentro del ámbito desde el punto de vista de su parte interesada.
5. **NB: para una versión más sencilla: que cada jugador coja una tarjeta de caso de uso y diseña el sistema de IA descrito en ese caso de uso**
6. Una vez que todos los jugadores hayan diseñado su caso de uso, el juez sacará una tarjeta de requisitos por estudiante.
7. El juez pregunta al primer desarrollador si su caso de uso cumple con el requisito y, en caso de que si, cómo.
8. Si la respuesta es aceptable, el desarrollador recibe 2 puntos.
9. Si no es aceptable, o si el estudiante no puede responder, cualquiera de los otros desarrolladores puede argumentar por qué su caso cumple con el requisito.
10. Si esta respuesta es aceptable, el otro desarrollador recibe 2 puntos.
11. El juez repite los pasos 7-10 para cada uno de los promotores.
12. El diseño con más puntos es discutido por todos los desarrolladores para evaluar los resultados.

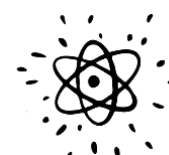
Ejercicio 2 - Diseño y debate en grupo

Materiales: Cartas de técnica, cartas de ámbito, cartas de partes interesadas, cartas de requisitos

5 jugadores



1. En grupo, elijan una tarjeta de ámbito y otra de técnica.
2. Juntos, inventen un caso de uso para ese ámbito y esa técnica.
3. Cada jugador coge una carta de parte interesada.
4. Tome una tarjeta de requisito a la vez y comience a discutir ese requisito desde la perspectiva de su parte interesada.
5. Repita el paso 4 para cada requisito.
6. Opcional: Ordene las tarjetas de requisitos por orden de prioridad desde el punto de vista de sus propias partes interesadas y comente por qué ha elegido esa orden.



Ejercicio 3 - Discutir un caso de uso

Materiales: Cartas de caso de uso, cartas de partes interesadas, cartas de requisitos

5 jugadores

1. En pequeños grupos, elija una tarjeta de caso de uso.
2. Cada jugador coge una carta de parte interesada.
3. Tome una tarjeta de requisito a la vez (5 en total) y comience a discutir ese requisito desde la perspectiva de su parte interesada.
4. Añade: Ordene las tarjetas de requisitos por orden de prioridad desde el punto de vista de sus propias partes interesadas y comente por qué ha elegido esa orden.

Las Partes Interesadas

Una parte interesada es el papel de una persona en el desarrollo o la implantación de un sistema de IA que tiene sus propios "intereses contrapuestos", por ejemplo, el dinero, la eficiencia, la seguridad, la equidad, la privacidad, la autonomía, etc. Entendemos que la representación de algunas de estas partes interesadas puede resultar difícil para algunos, por lo que ofrecemos algunos ejemplos para que sirvan de inspiración.

Gobernanza: Los diferentes responsables gubernamentales pretenden garantizar que la IA sea transparente, sostenible y cumpla con los diferentes requisitos éticos y legales. Interés: cumplimiento legal

Autoridad/supervisor: Las autoridades supervisoras son autoridades independientes que velan por la aplicación y el cumplimiento uniforme y coherente de las normas y leyes (dentro de su ámbito específico de competencia). Interés: seguridad

Implantador: El implantador de un sistema de IA tiene como objetivo garantizar que el sistema funcione de forma rentable y cumpla con sus objetivos. Interésese: eficiencia, dinero

Afectado: Los afectados se preocupan por su propio bienestar (físico y mental) y por la protección de sus derechos humanos fundamentales. Intereses: equidad, privacidad, autonomía

Experto en el ámbito: los expertos en el campo tienen como objetivo informar a los desarrolladores, a los implantadores y a cualquier otra parte interesada con información correcta y fiable sobre su campo de experiencia para garantizar que un sistema funcione correctamente tanto a nivel técnico como social. Intereses: equidad, seguridad, información correcta

Desarrollador: El objetivo de los desarrolladores es desarrollar sistemas que funcionen correctamente teniendo en cuenta los objetivos del implantador. Interés: eficiencia, seguridad

7 pasos hacia una IA Fiable

Los "7 pasos hacia una IA Fiable" funcionan como un marco de pensamiento que enseña a los estudiantes cómo abordar los casos prácticos de una manera práctica y sistemática. En el ejercicio, los estudiantes analizarán un problema o caso utilizando 7 pasos que tienen los requisitos para una IA fiable integrados como parte del proceso. Aprenderán a identificar, aplicar y equilibrar los elementos y dilemas éticos, morales y sociales relacionados con la IA en general. Además, el ejercicio debe enseñar a los estudiantes a comunicar y defender de forma convincente sus posiciones y diseñar ideas para soluciones técnicas y/o procesos organizativos.

Este ejercicio es un método de enseñanza flexible que puede aplicarse a cualquier estudio de caso y puede integrarse de diferentes maneras en la clase; en forma de trabajo en grupo, como tarea escrita individual, actividad de clase o presentación. El análisis de casos proporciona a los estudiantes una forma práctica de analizar situaciones a las que pueden enfrentarse en el trabajo. El ejercicio pretende mejorar la capacidad de análisis y de resolución de problemas de los estudiantes, así como su capacidad para tomar decisiones y enfrentarse a ambigüedades. Al mismo tiempo, mejorarán sus conocimientos de ética y desarrollarán una mentalidad ética para su aplicación práctica. Además, el análisis de casos en clase o en grupos de discusión puede hacer que los estudiantes comprendan mejor que es posible que varias perspectivas sean igual de válidas sobre una cuestión.

En resumen, los pasos que seguirán sus alumnos son:

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

Cómo utilizar el ejercicio

- ◇ **Elija una forma de ejercicio:** este ejercicio puede realizarse de múltiples formas. Los alumnos pueden enfrentarse a los casos en forma de evaluación **individual**, pero también puede optar por convertirlo en un ejercicio de **grupo** o de **clase**. Además, puedes decidir si quieres que los alumnos se limiten a discutir el caso siguiendo los pasos, convirtiéndolo en un "*ejercicio de reflexión*", o si quieres convertirlo en una *evaluación escrita* o incluso en una *presentación*.
- ◇ **Selección del caso:** Dado que el ejercicio ya es un proceso en sí mismo, es preferible que el profesor presente un caso a la clase (o varios casos para que los alumnos elijan). El caso podría estar relacionado con el tema de la clase o con el campo de estudio, presentándoles un caso existente (por ejemplo, de las noticias) o uno realista y simulado. Como alternativa, los alumnos pueden elegir su propio caso. Sin embargo, esto les impide dirigir el ejercicio. Para elegir un caso que sea aplicable al ejercicio, asegúrese de que el caso:

- Es lo suficientemente realista y relacionable para que los estudiantes se comprometan. Debe ser interesante para tus alumnos en concreto y ofrecerles un reto que quieran afrontar.
- Se ajusta al nivel de conocimientos de los alumnos o al conocimiento del tema.
- Cuenta una historia con uno o más aspectos de conflicto.
- Contiene distintas partes interesadas.
- No tiene una solución obvia y tiene múltiples soluciones posibles. Aunque el caso sea sencillo, debe ser lo suficientemente complejo como para generar un debate en grupo y soluciones alternativas al problema planteado.
- Es aplicable a los requisitos.

Los casos pueden ser simples o complejos. Cuando busques un caso más complejo, piensa en dar a los alumnos mucha información, incluyendo cuadros de datos y gráficos sobre el caso, o una historia compleja en la que el problema no sea evidente de inmediato y haya que tener en cuenta muchos factores diferentes.

A continuación, encontrará algunos **casos** más sencillos. Sin embargo, le animamos a que proponga sus propios casos ya que conoce mejor a sus alumnos y sus intereses. También puede consultar las tarjetas de casos de uso de la baraja cartas para inspirarse.

Deja que hagan el ejercicio. Dependiendo de la forma que hayas elegido, los alumnos pueden trabajar en el ejercicio de forma individual o, si decides que sea un ejercicio de clase, puedes guiarles en los pasos.

