



MINISTÈRE
DE L'ÉDUCATION
NATIONALE
ET DE LA JEUNESSE

Liberté
Égalité
Fraternité

Inteligencia artificial y educación

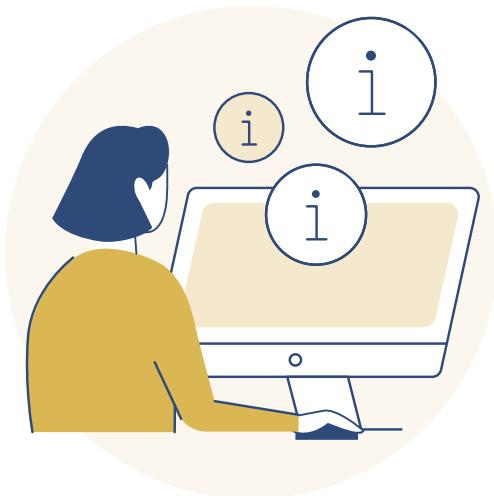
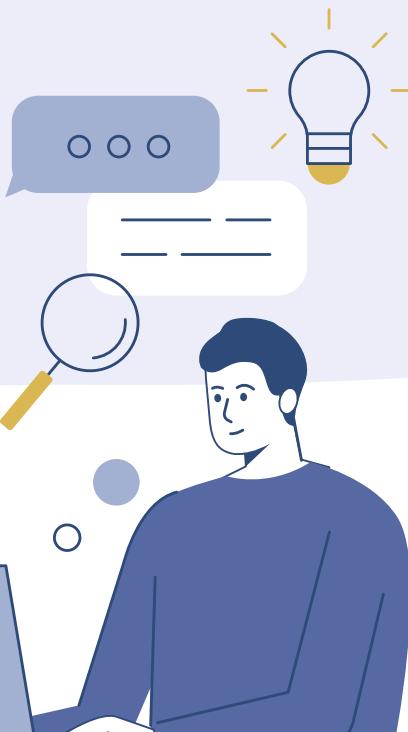
Aportaciones de la
búsqueda y retos para
las políticas públicas

Enero 2024



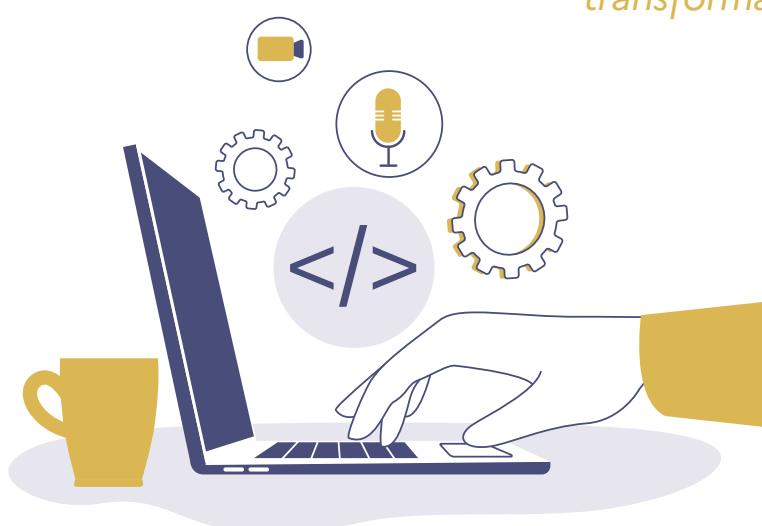
État des lieux sur les apports de la recherche et les lignes directrices des institutions internationales sur l'intelligence artificielle et l'éducation :

définitions et approches, politiques publiques, enjeux éthiques, applications, tournant des IA génératives.



Informe sobre las aportaciones de la investigación y las directrices de las instituciones internacionales en materia de inteligencia artificial y educación:

definiciones y enfoques, políticas públicas, cuestiones éticas, aplicaciones, transformación de las IA generativas.



Índice de contenidos

Resumo	4
Importante	5
1. La inteligencia artificial (IA): una variedad de definiciones y enfoques	6
2. ¿Cuáles son los retos para las políticas públicas?	13
3. ¿Cuáles son los retos éticos?	23
4. ¿Cuáles son los campos de aplicación?	25
5. ¿Cuáles son las pistas de trabajo para formarse en IA y enseñarla?	30
6. El cambio en los sistemas de IA generativa y los grandes modelos de lenguaje.....	38
Referencias.....	50

Resumo

Este documento, producido por el Departamento de Educación Digital (Ministerio de Educación Nacional y Juventud), ofrece un inventario (Enero de 2024) de las contribuciones de la investigación y las directrices de las instituciones internacionales sobre la inteligencia artificial (IA) y la educación: diversidad de definiciones y enfoques, cuestiones para las políticas públicas, cuestiones éticas, campos de aplicación, vías de formación y enseñanza, perspectivas con el giro actual de las IA generativas y los grandes modelos de lenguaje.

Résumé

Ce document, réalisé par la Direction du numérique pour l'éducation / ministère de l'Éducation nationale et de la Jeunesse, propose un état des lieux (janvier 2024) sur les apports de la recherche et les lignes directrices des institutions internationales sur l'intelligence artificielle (IA) et l'éducation : diversité des définitions et des approches, enjeux pour les politiques publiques, enjeux éthiques, domaines d'application, pistes de travail pour former et enseigner, perspectives avec le tournant actuel des systèmes d'IA générative et des grands modèles de langage.

Abstract

This document, produced by the Digital Directorate for Education (Ministry of National Education and Youth), offers an inventory (January 2024) of the contributions of research and the guidelines of international institutions on the artificial intelligence (AI) and education: diversity of definitions and approaches, issues for public policies, ethical issues, fields of application, avenues for training and teaching, perspectives with the turning point of generative AI and large language models.

Importante

La inteligencia artificial: una variedad de definiciones y enfoques

"Sorprendentemente, el término que se ha elegido recientemente para calificar estas increíbles proezas de cálculo es el de *inteligencia artificial* (IA). El regreso a la escena de este vocablo forjado en 1956 por John McCarthy es un enigma interesante para la historia de la ciencia y la técnica."

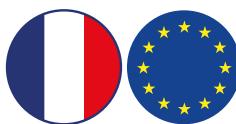
(Cardon et al., 2018)



¿Cuáles son los retos para las políticas públicas?

"La estrategia de la IA en educación debe ofrecer soluciones a los profesionales de la educación y la formación, ayudar al aprendizaje y a la toma de decisiones sin sustituir nunca a las personas."

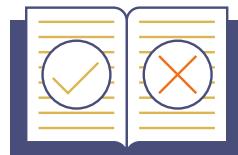
(Gobierno - Dirección General de Empresas, 2023)



¿Cuáles son los retos éticos?

Cuatro consideraciones clave que la Comisión Europea propone incluir: acción humana y control humano, equidad, humanidad, justificación de las elecciones, transparencia y explicabilidad.

Según (Dirección General de Educación, Juventud, Deporte y Cultura y Comisión Europea, 2022)



Los ámbitos de aplicación en la educación

"La inteligencia artificial desempeña un papel importante y creciente en la educación. Un caso importante es el de los sistemas educativos personalizados que ya están bien establecidos, con cada vez más pruebas de su eficacia para mejorar el aprendizaje."

(Khosravi et al., 2022)



¿Cuáles son las pistas de trabajo para formarse en IA y enseñarla?

Una propuesta de trayectoria en cinco pilares

1. Incertidumbre y carácter aleatorio
2. Codificación y cálculo
3. Sensibilización ante los datos
4. Pensamiento crítico
5. Humanismo post-IA

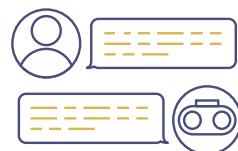
Según (de la Higuera & Bocquet, 2020)



El giro de los sistemas de IA generativa y de los grandes modelos de lenguaje

"En los últimos dos años, se han publicado numerosos modelos generativos importantes como ChatGPT o Stable Diffusion. Son capaces de realizar tareas como un sistema general de preguntas y respuestas o la creación automática de imágenes artísticas que revolucionan múltiples sectores."

(Gozalo-Brizuela & Garrido-Merchan, 2023)



1

La inteligencia artificial (IA): una variedad de definiciones y enfoques

1.1 Origen y actos fundadores

Remontándose a sus orígenes, como nueva ciencia, la IA "se basaba en la conjetaura de que todas las **capacidades cognitivas**, especialmente el razonamiento, el cálculo, la percepción, la memorización e incluso el descubrimiento científico o la creatividad artística, podrían describirse con tanta precisión que debería ser **posible reproducirlas con un ordenador.**" Según (Ganascia, 2022)

"Además de un artículo principal sobre la morfogénesis en 1952, [los] dos artículos más famosos [de Turing] en el mundo de la informática son "On Computable Numbers, With an Application to the Entscheidungsproblem" en 1937, donde establece los cimientos de la **máquina de Turing** y la calculabilidad, y "Computing Machinery and Intelligence" en 1950, donde plantea **la cuestión de la inteligencia para las máquinas** y enuncia el **test de Turing.**" (Le Blanc, 2014)

"Una definición clásica es que la IA es un campo de investigación que desarrolla tecnologías capaces de hacer cosas que requerirían inteligencia si estuvieran por humanos (Minsky, 1969). Este enfoque tiene su origen en **Turing**, que propuso que si una **simulación de ser humano inteligente** no puede distinguirse de una persona real, las cuestiones relacionadas con la inteligencia se vuelven inútiles (Turing, 1950). Muchos especialistas en ciencias cognitivas y algunos investigadores y filósofos especializados en IA han adoptado un punto de vista más firme, que afirma que la investigación sobre la IA puede revelar cómo funciona la mente humana (Gardner, 1985)." (Holmes et al., 2022) traducción de (Bocquet, 2023)

"Si hay que buscar un inicio a la historia, podemos empezar por la **conferencia de Dartmouth**, donde surge el término ahora acuñado de *inteligencia artificial*. Fue organizada en el verano de **1956**, los pioneros **Marvin Minsky y John McCarthy** en el Dartmouth College, cerca de Hanover (New Hampshire)." (Le Cun et al., 2019)

"Sorprendentemente, el término que se ha elegido recientemente para calificar estas increíbles proezas de cálculo es inteligencia artificial (IA). El regreso a la escena de este **lenguaje forjado en 1956 por John McCarthy** es un enigma interesante para la historia de la ciencia y la técnica. La mayoría de los observadores rigurosos subrayan que solo en el ámbito de los métodos de aprendizaje y, especialmente, del aprendizaje profundo (*deep learning*), se están produciendo actualmente avances significativos en la predicción calculada. Sin embargo, la inclusión de estas técnicas en el campo de la IA no siempre ha sido algo natural. (...)"

Las **redes neuronales** tienen su origen en la historia pionera de la informática y la primera cibernetica. Aunque la etiqueta es posterior, se puede decir que es "conectista" y seguirá haciendo referencia a la propuesta del neurofisiólogo **Warren McCulloch** y el lógico **Walter Pitts** de modelizar matemáticamente una red de neuronas **en 1943**. Este artículo fundador se sigue considerando —incluso en las citas actuales de los artículos de *deep learning*— el punto de partida de la aventura conexionista". (Cardon et al., 2018)

Algunas etapas en la historia de la IA

1943	Artículo sobre las neuronas artificiales de McCulloch y Pitts.
1950	Artículo de Turing « <i>Computing Machinery and Intelligence</i> » en la base del «test de Turing».
1956	El término <i>inteligencia artificial</i> se utiliza por primera vez en una conferencia en el Dartmouth College .
1958	John McCarthy inventó el lenguaje de programación LISP.
1964-1966	El programa Eliza , uno de los primeros chatbots , fue desarrollado por Joseph Weizenbaum. Dio nacimiento al «efecto Eliza» (tendencia a asignar involuntariamente rasgos humanos a secuencias de símbolos generados por un ordenador).
1972-1981	Se desarrolla el sistema experto MYCIN para ayudar a los médicos a diagnosticar enfermedades infecciosas.
1974	Paul Werbos desarrolla el primer algoritmo de retropropagación, un método clave para el <i>deep learning</i> o aprendizaje profundo .
1997	Deep Blue , el ordenador de ajedrez desarrollado por IBM, vence al campeón mundial de ajedrez Garry Kasparov.
2011	¡El sistema Watson de IBM gana el juego de televisión Jeopardy! contra dos antiguos campeones humanos.
2012	La red de neuronas convolutivas AlexNet gana el concurso ImageNet, que marca el comienzo de la era del <i>deep learning</i> .
2016	AlphaGo , un programa de IA desarrollado por Google DeepMind, vence al campeón mundial de Go Lee Sedol. Google desarrolla su asistente de voz.
2017	Artículo sobre Transformers de (Vaswani et al.) [Búsqueda en Google].
2018	Google desarrolla su modelo de lenguaje BERT , mejorando el rendimiento del procesamiento automático del lenguaje natural.
2019	Pluribus es un programa desarrollado por Facebook AI que ha logrado un gran avance en el mundo del póquer.
2022	ChatGPT abierto al público por OpenAI.
2023	Rápido desarrollo de sistemas de IA generativa y grandes modelos multimodales (texto, imagen, sonido, vídeo).

1.2 Decisiones humanas, modelización y automatización de tareas cognitivas

"La expresión *inteligencia artificial*, creada en la década de 1950, es el **campo de investigación que estudia los mecanismos de la inteligencia** modelizándolos con algoritmos y experimentándolos con máquinas. Estos mecanismos incluyen, por ejemplo, la capacidad de encontrar automáticamente soluciones a problemas, lo que puede implementar capacidades de planificación, predicción, control, memoria o aprendizaje. Por extensión, el término inteligencia artificial se utiliza a menudo para referirse a **algoritmos** [cadenas de acciones elementales para resolver un problema] **que simulan o tienen puntos en común con algunas de las capacidades de inteligencia de los seres vivos**". (Inria & Class'Code, 2020)

La IA se refiere a "un **sistema automatizado** que, para un conjunto dado de objetivos definidos por el ser humano, **es capaz de hacer previsiones, formular recomendaciones o tomar decisiones** que influyen en entornos reales o virtuales." Los sistemas de IA interactúan con nosotros e influyen directa o indirectamente en nuestro entorno. A menudo parecen funcionar de forma autónoma y pueden adaptar su comportamiento en función del contexto." Según (UNICEF, 2021) tomado por (Holmes & Tuomi, 2022) traducido por (Bocquet, 2023)

También es "un **campo interdisciplinario teórico y práctico** que combina tecnologías digitales, matemáticas, estadísticas y algorítmicas." (Gran léxico español de la inteligencia artificial, 2021)

"Podemos definir la inteligencia artificial (IA) como **sistemas informáticos capaces de participar en procesos de tipo humano**, como el aprendizaje, la adaptación, la síntesis, la autocorrección y el uso de datos para tareas de procesamiento complejo." (Popenici & Kerr, 2017) "[Esta definición] permite **superar la oposición entre la inteligencia humana y la IA y contemplar interacciones complejas entre ambas.**" (Romero et al., 2023)

1.3 Los tipos de IA

Existen varios tipos de IA:

Dependiendo de su tecnología:

"Una **IA basada en los datos**, o aprendizaje automático (*machine learning* en inglés), y otra IA basada en los conocimientos, o **IA simbólica**. Un tercer modelo conceptual, la **IA híbrida**, combina los enfoques basados en los datos y los conocimientos con la cognición humana." [Traducción] Según (Holmes et al., 2022) traducción por (Bocquet, 2023)

Dependiendo de su alcance:

"**Distinguimos entre la IA débil y la IA fuerte** en función del alcance de aplicación de estos programas informáticos. La inteligencia artificial débil (*weak AI*) es una inteligencia artificial que se centra en una tarea concreta, mientras que la inteligencia artificial fuerte (*strong AI*) o la inteligencia artificial general (*AGI*: *artificial general intelligence*) se aplica a cualquier problema. Actualmente, todos los sistemas de IA existentes se consideran inteligencias artificiales débiles." (*Gran léxico español de la inteligencia artificial*, 2021)

“Inteligencia artificial completa o fuerte: Una inteligencia artificial que sería capaz de copiar las habilidades humanas (aprender, comprender, aprehender, razonar, tomar decisiones, tener una conciencia, emociones...). **Hasta la fecha, la inteligencia artificial fuerte no existe, es una creencia.**

Inteligencia artificial débil: Es la inteligencia artificial que se conoce actualmente: es un algoritmo que "aprende", adaptando sus parámetros a datos de aprendizaje, desprovisto de capacidades mentales y cognitivas, pero capaz de realizar con mucha más eficiencia, a veces más que el ser humano, una tarea precisa." (Inria y Class'Code, 2020)

IA clásica y aprendizaje automático:

IA clásica: La IA más antigua o 'clásica', también conocida como '**IA simbólica**', '**IA basada en reglas**' o '**IA a la antigua**' (GOFAI, por sus siglas en inglés), consiste en escribir secuencias de IF... (SI) THEN... (ENTONCES) y otras reglas de lógica condicional, pasos que la computadora seguirá para completar una tarea. Durante décadas, los '**sistemas expertos**' de IA basados en reglas se han desarrollado para una amplia gama de aplicaciones, como el diagnóstico médico, la calificación crediticia y la manufactura. Los sistemas expertos se basan en un enfoque conocido como 'ingeniería del conocimiento', que implica la obtención y el modelado de los conocimientos de los expertos en un dominio específico, una tarea que requiere muchos recursos y que no está exenta de complicaciones(...)

Aprendizaje automático (AA): Muchos de los avances recientes de la IA –como el procesamiento del lenguaje natural, el reconocimiento facial y los vehículos autónomos– han sido posibles gracias a los avances en los enfoques computacionales basados en el AA. En lugar de utilizar reglas, **el aprendizaje automático analiza grandes cantidades de datos para identificar patrones** y construir un modelo que luego se utiliza para predecir valores futuros. En este sentido, se dice que los algoritmos, en lugar de estar preprogramados, 'aprenden'.

Hay **tres enfoques principales de AA:**

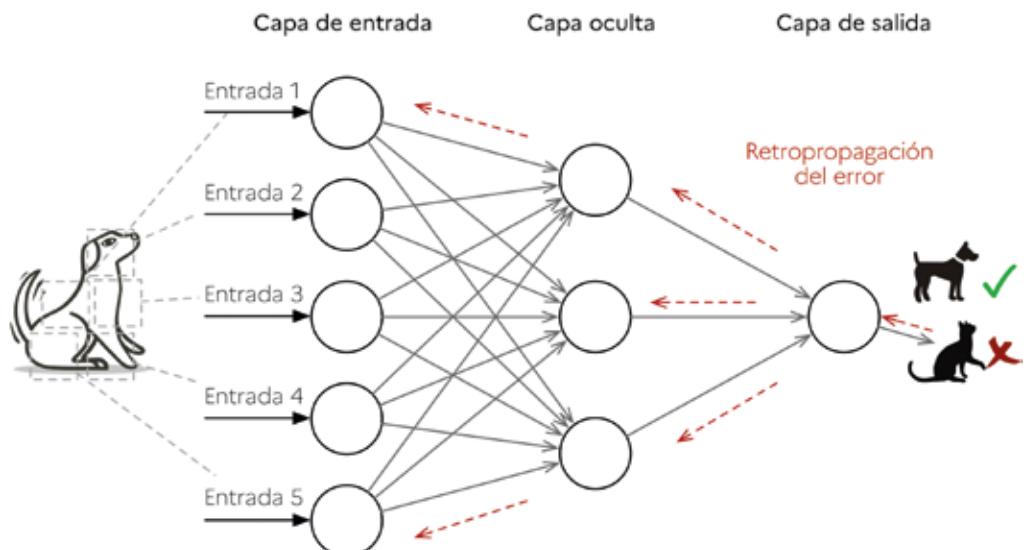
- **supervisado**,
- **no supervisado**,
- **de refuerzo**.

- **El aprendizaje supervisado** utiliza datos que ya han sido etiquetados, como varios miles de fotografías de personas etiquetadas por humanos. El aprendizaje supervisado relaciona los datos con las etiquetas, para construir un modelo que pueda aplicarse a datos similares – por ejemplo, para identificar automáticamente a las personas en nuevas fotografías. En el aprendizaje no supervisado, la IA dispone de cantidades de datos aún mayores, pero esta vez los datos no han sido categorizados ni etiquetados.
- **El aprendizaje no supervisado** pretende descubrir patrones ocultos en los datos, agrupamientos que pueden utilizarse para clasificar nuevos datos. Por ejemplo, puede identificar automáticamente letras y números en la escritura a mano buscando patrones en miles de ejemplos. Tanto en el aprendizaje supervisado como en el no supervisado, el modelo derivado de los datos es fijo y, si los datos cambian, hay que volver a realizar el análisis.
- Sin embargo, el tercer enfoque del AA, **el aprendizaje por refuerzo**, implica una mejora continua del modelo basada en la retroalimentación, es decir, se trata de un aprendizaje automático en el sentido de que este es continuo. La IA recibe unos datos iniciales a partir de los cuales obtiene un modelo que se evalúa como correcto o incorrecto y se premia o castiga en consecuencia.

Una red neuronal artificial (RNA) es un enfoque de la IA que se inspira en la estructura de las redes neuronales biológicas (es decir, los cerebros de los animales). Las RNA consisten en tres tipos de capas interconectadas de neuronas artificiales: una capa de entrada, una o varias capas computacionales intermedias y una capa de salida que proporciona el resultado. (...)

El aprendizaje profundo se refiere a las RNA que comprenden múltiples capas intermedias. Es este enfoque el que ha dado lugar a muchas de las recientes y notables aplicaciones de la IA (por ejemplo, en el procesamiento del lenguaje natural, el reconocimiento del habla, la visión por computadora, la creación de imágenes, el desarrollo de fármacos y la genómica). (UNESCO, 2021)

Funcionamiento de una red neuronal simple



(Cardon et al., 2018)

IA, aprendizaje automático, red neuronal y aprendizaje profundo



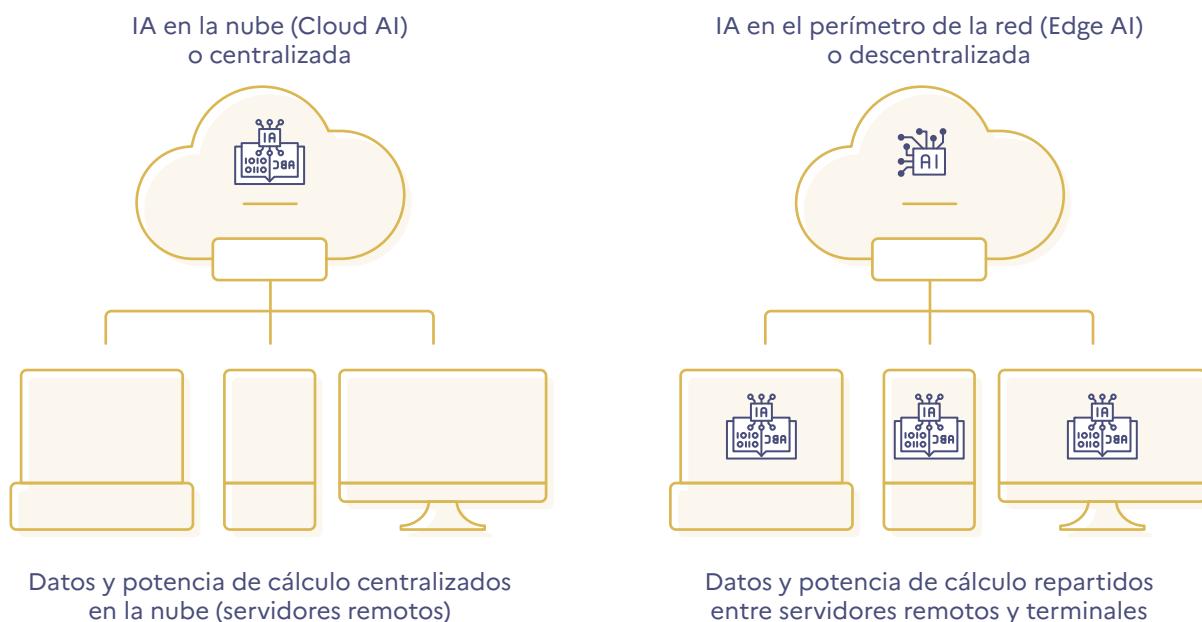
(UNESCO, 2021)

¿Qué es la IA en la periferia de la red (*Edge AI*)?

¿Qué es Edge AI?

"Edge AI es el uso de la IA en dispositivos del mundo real. Edge AI se refiere a la práctica de realizar cálculos de IA cerca de los usuarios en el borde de la red en lugar de una ubicación centralizada como el centro de datos de un proveedor de servicios en la nube o el almacén de datos privado de una empresa." (Singh & Gill, 2023)

IA en el perímetro de la red (*Edge AI*)



2.

¿Cuáles son los retos para las políticas públicas?

2.1 El consenso de Pekín

"El Consenso de Pekín recomienda a los gobiernos y otras partes interesadas de los Estados miembros de la UNESCO que:

- **Planificar la IA en las políticas educativas** para aprovechar las oportunidades y hacer frente a los desafíos que presentan las tecnologías de IA, adoptar enfoques que involucren a todo el gobierno, intersectoriales y multipartidistas que también permitan establecer prioridades estratégicas locales para alcanzar los objetivos del ODS 4 [Objetivo de Desarrollo Sostenible 4: Garantizar que todos tengan acceso a una educación de calidad e igualitaria y promover las oportunidades de aprendizaje a lo largo de toda la vida].
- **Apoyar el desarrollo de nuevos modelos** posibles gracias a las tecnologías de IA para la prestación de servicios de educación y formación donde los beneficios superan claramente los riesgos, y utilizar herramientas de IA para ofrecer sistemas de aprendizaje permanente que permitan un aprendizaje personalizado en cualquier momento, en cualquier lugar y para todos.
- **Considerar el uso de datos relevantes**, si procede, para fomentar la planificación de políticas basadas en datos concluyentes.
- Asegurarse de que las tecnologías de IA sirven para **dar autonomía a los profesores en lugar de reemplazarlos** y desarrollar programas adecuados para el refuerzo de las capacidades para que los profesores trabajen junto a los sistemas de IA.
- Preparar a la próxima generación de trabajadores **equipándola con los valores y las competencias necesarios para la vida y el trabajo** más relevantes en la era de la IA.
- Promover un **uso justo e inclusivo** de la IA, independientemente de la discapacidad, la situación social o económica, el origen étnico o cultural o la ubicación geográfica, haciendo hincapié en la igualdad de género, al tiempo que se garantizan usos éticos, transparentes y verificables de los datos educativos."

(UNESCO, 2019)

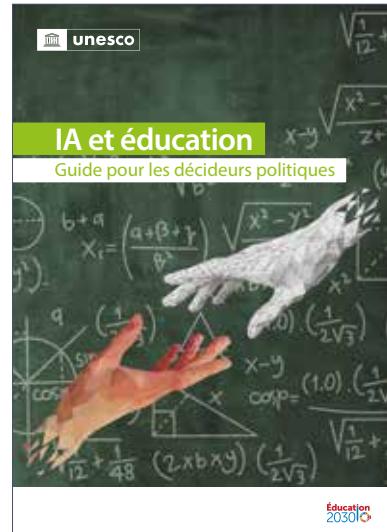


2.2 Recomendaciones y previsiones de la UNESCO y la OCDE

Recomendaciones para los responsables políticos

1. Visión y prioridades estratégicas integrales del sistema

- Definir una visión sistémica de las políticas de IA y educación
 - Garantizar el uso inclusivo y equitativo de la IA en la educación;
 - Hacer uso de la IA para mejorar la educación y el aprendizaje;
 - Promover el desarrollo de habilidades para la vida en la era de la IA, incluyendo la enseñanza del funcionamiento de la IA y sus implicancias para la humanidad;
 - Salvaguardar el uso transparente y auditável de los datos educativos.
- Evaluar la preparación integral del sistema y seleccionar las prioridades estratégicas



2. Principio general para las políticas de IA y educación

- Adoptar un enfoque humanista como principio general de las políticas de IA y educación

3. Planificación interdisciplinaria y gobernanza intersectorial

- Movilizar los conocimientos interdisciplinarios y de las múltiples partes interesadas para fundamentar la planificación de políticas y desarrollar las capacidades de quienes formulan políticas
- Establecer mecanismos de gobernanza y coordinación intersectoriales

4. Políticas y normativas para un uso equitativo, inclusivo y ético de la IA

- Establecer objetivos estratégicos transversales y planificar normativas y programas, para garantizar el uso equitativo de la IA en la educación

5. Planes maestros para utilizar la IA en la gestión de la educación, la enseñanza, el aprendizaje y la evaluación

- Hacer uso de la IA para impulsar y mejorar la gestión e impartición de la educación
- Cultivar el uso de la IA centrado en el estudiante para mejorar el aprendizaje y la evaluación
- Garantizar que la IA se utilice para empoderar a los docentes
- Planificar el uso de la IA para apoyar el aprendizaje a lo largo de la vida en todas las edades, lugares y entornos
- Desarrollar valores y habilidades para la vida y el trabajo en la era de la IA

6. Pruebas piloto, monitoreo y evaluación, y creación de una base de evidencias

- Crear una base de evidencias fiable para apoyar el uso de la IA en la educación
- Reforzar la investigación y la evaluación en el ámbito de la IA y la educación

7. Promoción de las innovaciones locales en materia de IA para la educación

- Promover el desarrollo local de las tecnologías de la IA para la educación

(Referencia : UNESCO, 2021)

Los nuevos retos de la IA generativa

"Las implicaciones de la IA generativa para la evaluación van mucho más allá de las preocupaciones inmediatas sobre las trampas que los alumnos puedan hacer al redactar sus deberes. Debemos tener en cuenta que la **IA generativa puede producir artículos y ensayos relativamente bien organizados y obras de arte impresionantes**, y puede aprobar ciertos exámenes basados en conocimientos en algunas áreas. **Por lo tanto, debemos replantearnos lo que se debe aprender exactamente y para qué fin**, así como la forma de evaluar y validar el aprendizaje. (...)"

La perspectiva más fundamental de las implicaciones a largo plazo de la IA generativa para la educación y la investigación sigue englobando la **relación complementaria entre la acción humana y las máquinas**. Una de las cuestiones clave es si los seres humanos podrían ceder niveles fundamentales de reflexión y de procesos de adquisición de competencias a la IA para centrarse, más bien, en capacidades de reflexión de nivel superior basadas en los resultados proporcionados por la IA."

(Miao y Holmes, 2023)

"Los modelos de IA generativa que generan texto, imágenes, videos y contenido de audio (música, voz, etc.) están avanzando a una velocidad vertiginosa. Esto abre **posibilidades infinitas, demostradas en un número creciente de áreas**. Sin embargo, la tecnología también plantea muchos desafíos y riesgos para las personas, las empresas, las economías, las sociedades y los responsables políticos de todo el mundo, desde la perturbación del mercado laboral y la desinformación a corto plazo hasta los posibles desafíos a largo plazo en el control de las acciones de las máquinas. Las trayectorias futuras de la IA generativa son difíciles de predecir, pero los gobiernos deben explorarlas para poder darles forma.

El desarrollo tecnológico de la IA generativa está en sus inicios, con pioneros como actores tecnológicos establecidos como Microsoft, Google y Meta, y laboratorios de investigación privados como OpenAI, Midjourney y Stability.AI. Estas empresas siguen múltiples estrategias para sacar partido de la IA generativa y, en cierta medida, mitigar sus inconvenientes.

El debate público sobre la IA generativa surgió hace menos de un año. A medida que las empresas tecnológicas lanzan al mercado aplicaciones de IA generativa, los responsables políticos de todo el mundo se enfrentan a sus implicaciones. Los investigadores de ciencias aplicadas y los universitarios están participando en un intenso debate sobre cómo gestionar la IA generativa, desde las medidas de mitigación en el diseño y el desarrollo de los modelos hasta el lanzamiento al mercado y más allá.

El camino a seguir no está claro y las perspectivas divergen. Uno de los extremos aboga por una moratoria sobre las experiencias de IA generativa más avanzadas que el GPT-4 (Future of Life Institute, 2023), mientras que el otro considera que los supuestos riesgos existenciales de la IA son exagerados (LeCun y Ng, 2023).²⁸ Otros, quizás la mayoría, se sitúan en algún punto intermedio. Sea cual sea la postura ideológica sobre estas **cuestiones, es urgente seguir investigando para prepararse para los diferentes escenarios posibles de la IA generativa**. Dada la gran incertidumbre y el impacto potencialmente significativo que la tecnología podría tener a nivel micro y macro, los responsables políticos deben mantenerse informados y preparados para tomar las medidas adecuadas a través de políticas de IA orientadas al futuro.

La OCDE desea que este documento sirva de trampolín para ayudar a los gobiernos a avanzar en este ámbito. El Observatorio de Políticas de IA de la OCDE y su nuevo Grupo de Expertos de la OCDE sobre el Futuro de la IA servirán, junto con otros organismos relevantes, como foro de diálogo sobre estos temas, generando ideas y recomendaciones explotables para los gobiernos. También se está llevando a cabo un trabajo complementario en el marco de otras iniciativas de la OCDE, como el trabajo del Comité de Empleo, Trabajo y Asuntos Sociales, el Centro de Recursos DIS/MIS de la OCDE y la iniciativa horizontal "Going Digital" de la OCDE."

(Lorenz et al., 2023)

2.3 Una estrategia nacional

"La inteligencia artificial (IA) es una prioridad para Francia en términos de investigación, economía, modernización de la acción pública, regulación y ética. En 2018, el Gobierno inició una estrategia nacional para la IA.

A raíz del informe "Dar sentido a la inteligencia artificial: creación de una estrategia nacional y europea" (2018) y en el marco del plan Francia 2030, el Gobierno lanzó en 2018 una **estrategia nacional para la inteligencia artificial (SNIA)**. Esta estrategia tiene como objetivo posicionar a Francia como uno de los líderes europeos y mundiales de la inteligencia artificial (IA).

La estrategia nacional para la inteligencia artificial tiene como objetivo preservar y consolidar la soberanía económica, tecnológica y política de Francia y poner la IA al servicio de la economía y la sociedad. Está vinculada al componente de dominio de las tecnologías digitales soberanas y seguras de Francia 2030.

La estrategia nacional de IA se divide en dos fases:

FASE 1 (2018-2022)

Dotar a Francia de capacidades de investigación competitivas. Este primer paso se financió por un importe de 1850 millones de euros. En particular, ha financiado la creación y el desarrollo de una red de institutos interdisciplinarios de inteligencia artificial, la creación de cátedras de excelencia y programas de doctorado, así como el despliegue de la supercomputadora Jean Zay.

FASE 2 (2021-2025)

Difundir las tecnologías de inteligencia artificial en la economía y apoyar el desarrollo y la innovación en áreas prioritarias como la IA integrada, la IA de confianza, la IA frugal y la IA generativa. Esta segunda fase cuenta con 1500 millones de euros en el marco de Francia 2030. Esta nueva fase de la SNIA se basa en tres pilares estratégicos: el apoyo a la oferta de *deep tech*, la formación y la atracción de talento, y la aproximación de la oferta y la demanda a soluciones de IA."

(Gobierno - Dirección General de Empresas, 2023)

Cronología e implantación geográfica

Etapas del desarrollo de la estrategia nacional de IA



Implantación de los institutos temáticos en IA (3IA PR[AI]RIE, MIAI, 3IA Côte d'Azur y ANITI) y de los centros de excelencia (SCAI, DATAIA y Hi! PARÍS) frente a las zonas históricamente activas en este ámbito



Fuente: Tribunal de Cuentas

Nota de lectura: La actividad histórica en IA se mide a través de la distribución departamental de las tesis en IA sostenidas entre 1989 y 2019 en los centros franceses de educación superior a partir de los datos abiertos de theses.fr (ABES).

Las tesis se dan por año de defensa. La detección de las tesis en IA se realiza según el método semántico desarrollado por el Tribunal. Los logotipos de 3IA y de los centros de excelencia en IA proceden de sus sitios web oficiales.

2.4 Una Ley Europea de IA (AI Act)

"Los diputados [del Parlamento Europeo] han llegado a un acuerdo político con el Consejo sobre un proyecto de ley destinado a garantizar que la IA en Europa sea segura, respete los derechos fundamentales y la democrática, a la vez que las empresas puedan prosperar y desarrollarse.

El viernes [08/12/23], los negociadores del Parlamento y del Consejo llegaron a un acuerdo provisional sobre la ley de inteligencia artificial. El objetivo de este Reglamento es garantizar que los derechos fundamentales, la democrática, el Estado de derecho y la sostenibilidad medioambiental estén **protegidos contra la IA de alto riesgo**, al tiempo que se estimula la innovación y se hace de Europa un líder en este campo. Las normas establecen obligaciones para la IA en función de sus riesgos potenciales y su nivel de impacto. (...)

Entre las aplicaciones prohibidas se incluyen: sistemas de clasificación biométrica que utilizan características sensibles, recuperación no dirigida de imágenes faciales de Internet o secuencias de videovigilancia para crear bases de datos de reconocimiento facial, reconocimiento de emociones en el lugar de trabajo y en las instituciones

educativas, calificación social basada en comportamiento social o características personales, sistemas de IA que manipulan el comportamiento humano para eludir su libre elección, explotación de las vulnerabilidades de las personas (debido a su edad, discapacidad, situación social o económica).

Medidas de apoyo a la innovación y a las PYME: Los diputados querían garantizar que las empresas, especialmente las pymes, pudieran desarrollar soluciones de IA sin la presión indebida de los gigantes industriales que controlan la cadena de valor. Con este fin, el acuerdo promueve lo que se conoce como "entornos aislados (o de pruebas) reglamentarios y tests en condiciones reales", establecidos por las autoridades nacionales para desarrollar y formar una IA innovadora antes de su comercialización".

Clasificación de los sistemas de IA

"El acuerdo de compromiso proporciona una capa horizontal de protección, incluida una clasificación de alto riesgo, para no incluir sistemas de IA que no puedan causar graves violaciones de los derechos fundamentales u otros riesgos significativos. Los **sistemas de IA de riesgo limitado** estarán sujetos a obligaciones de transparencia muy bajas, por ejemplo, informar de que el contenido ha sido generado por IA, para que los usuarios puedan tomar decisiones informadas sobre su uso posterior.

Se permitiría una amplia gama de **sistemas de IA de alto riesgo**, pero estaría sujeto a un conjunto de requisitos y obligaciones para acceder al mercado de la UE. Estos requisitos han sido aclarados y adaptados por los organismos reguladores para que cumplirlos sea técnicamente más factible y menos limitativo para las partes interesadas, por ejemplo, en lo que respecta a la calidad de los datos o a la documentación técnica que las pymes deberían elaborar para demostrar que sus sistemas de IA de alto riesgo cumplen con los requisitos.

Dado que los sistemas de IA se desarrollan y distribuyen a través de cadenas de valor complejas, el acuerdo de compromiso incluye **cambios que aclaran la distribución de las responsabilidades y los roles de los diferentes actores** de estas cadenas, especialmente los proveedores y usuarios de sistemas de IA. También aclara la relación entre las responsabilidades en virtud de la legislación sobre IA y las responsabilidades que ya existen en virtud de otros actos legislativos, como la legislación relevante de la UE en materia de protección de datos o la legislación sectorial.

Para algunos usos de IA, el riesgo se considera inaceptable y, por lo tanto, estos sistemas estarán prohibidos en la UE. El acuerdo provisional prohíbe, por ejemplo, la manipulación conductual cognitiva, la recogida no selectiva de imágenes faciales de Internet o de videovigilancia, el reconocimiento de emociones en el lugar de trabajo y en las instituciones educativas, la calificación social, la categorización biométrica para deducir datos sensibles, como la orientación sexual o las creencias religiosas, y algunos casos de policía predictiva aplicada a la persona."

Según (Parlamento Europeo, 2023) y (Consejo de Europa, 2023)

2.5 Acciones y producciones de los grupos temáticos digitales (#GTnum)

Acciones de GTnum IA y educación abierta

L'IMPACT DE L'INTELLIGENCE ARTIFICIELLE
à travers l'Éducation Ouverte

L'originalité de notre GTnum repose sur le fait d'avoir travaillé simultanément sur deux sujets contemporains, d'importance croissante et en utilisant l'éducation ouverte comme vecteur de recherche pour une intelligence artificielle plus éthique, au service de l'enseignement.

Le groupe thématique numérique #IA_EO a été co-animé par Nantes Université et IMT-Atlantique Brest, avec le soutien du ministère de l'Éducation nationale et de la Jeunesse (Direction du numérique pour l'Éducation), avec la contribution des délégations régionales académiques au numérique éducatif (DRANE) de Nantes et de Rennes.

CHIFFRES CLÉS

- 20 événements
- 35 articles du blog de la Chaire étiquetés #GTnum #IA_EO
- 35 brèves ou références au GTnum sur le carnet Hypothèses « Éducation, numérique et recherche »
- 13 vidéos de rediffusion hébergées sur le mediaserver de la Chaire UNESCO RELIA
- +200 tweets et mentions #GTnum #IA_EO

RESSOURCES ACCESSIBLES À TOUTES ET TOUS

- atelier de création de Ressources Éducatives Libres
- vidéos explicatives sur l'IA et les REL
- ressources de formation à l'IA
- foire aux questions en lien avec le GTnum

Nos Contributeurs

Nos Partenaires

Retrouvez-nous :

- chaireunescorelia.univ-nantes.fr
- twitter.com/Chaire_RELIA
- mediaserver.univ-nantes.fr/channels/#chaire-rel
- edunumrech.hypotheses.org

(DNE-TN2, 2023a)

Carteras de GTnum en la IA



Detalles y contenido de la cartera: (DNE-TN2, 2023c)



Detalles y contenido de la cartera: (DNE-TN2, 2023b)

2.6 El rendimiento de la IA en lectura y matemáticas

"El estudio forma parte de un proyecto integral para evaluar las capacidades de la IA y sus implicaciones para el trabajo y la educación, dirigido por el Centro de Investigación e Innovación Educativa de la OCDE (CERI). El **proyecto AI and the Future of Skills (AIFS)** tiene como objetivo desarrollar medidas de capacidades de IA que sean comprensibles, completas, repetibles y relevantes para las políticas. Mediante el uso de diversas fuentes de información sobre la IA, incluidas evaluaciones de expertos, el proyecto tiene como objetivo proporcionar a los responsables políticos los conocimientos que necesitan para dar forma a políticas educativas y del mercado laboral orientadas al futuro. El informe muestra significativas mejoras en la habilidad lectora de la IA desde 2016, lo que refleja **los avances en el procesamiento del lenguaje natural (NLP)** en los últimos años. La habilidad de la IA para resolver tareas matemáticas no ha mejorado al mismo ritmo. Sin embargo, los expertos **predicen que el aumento de las inversiones en investigación y desarrollo de la IA conducirá, en los próximos años, a avances significativos en la IA, tanto en lectura como en matemáticas**. El informe también demuestra el potencial de la IA para superar a gran parte de la población en lo que a la habilidad lectora y matemática se refiere. Esto tiene grandes implicaciones para el empleo y la educación, ya que, en el futuro, es probable que las máquinas se conviertan en una competencia cada vez mayor para los trabajadores en estos campos de habilidades. También destaca la necesidad de **reforzar las competencias básicas de la mano de obra y prepararla para trabajar junto con la IA en ámbitos clave**. Este estudio enfatiza **la importancia de supervisar, de forma periódica y sistemática, la evolución de las capacidades de la IA y de compararlas con las habilidades humanas**. Para ello, proporciona un ejemplo de cómo mejoran las capacidades de la IA con respecto a dos habilidades cognitivas clave de los seres humanos. **Esto resultará útil para los responsables políticos, educadores e investigadores que intenten comprender las implicaciones de los avances tecnológicos para el futuro del trabajo y la educación.**"

(OCDE, 2023)

3. • ¿Cuáles son los retos éticos?

3.1 Directrices de las instituciones internacionales

"Es fundamental comprender que los sistemas de IA no son magia. **Las personas diseñan, entrenan y dirigen la IA**, ya sean los que definen las políticas y estrategias de IA, los programadores de software que construyen los sistemas de IA, las personas que recopilan y clasifican los datos utilizados por estos sistemas o las personas que interactúan con ellos. Esto significa **que todas las partes interesadas en el ecosistema de desarrollo de tecnologías de IA deben comprender los principales retos** a los que se enfrentan para contribuir al uso responsable de la IA. Esto puede incluir **saber por qué y cómo se diseñó un sistema de IA, quién lo diseñó y con qué propósito.**"

(UNICEF, 2021)

Para un uso ético de la IA y los datos en la enseñanza, el aprendizaje y la evaluación, **la Comisión Europea propone cuatro consideraciones clave** que deben integrarse:

- **acción humana y control humano:** capacidad de acción de los actores de la educación (profesores, alumnos, jefes de institución);
- **equidad:** imparcialidad, inclusión, no discriminación y distribución justa de los derechos y las responsabilidades;
- **humanidad:** consideración por las personas, su identidad, integridad y dignidad;
- **justificación de las decisiones, transparencia y explicabilidad:** uso de conocimientos, hechos y datos para justificar las elecciones colectivas en el entorno escolar, a través de modelos participativos y colaborativos de toma de decisiones.

Según (Dirección General de Educación, Juventud, Deporte y Cultura y Comisión Europea, 2022)

El Consejo de Europa destaca varios principios para un uso de la IA en la educación respetuosa con **los derechos fundamentales**:

- *"Derecho a la dignidad humana.* La enseñanza, la evaluación y la acreditación no deben delegarse a un sistema de IA.
- *Derecho a la autonomía.* Los niños deberían tener derecho a evitar ser perfilados individualmente, a evitar trayectorias de aprendizaje dictadas y a proteger su desarrollo y su vida futura.
- *El derecho a ser escuchado.* Los niños deberían tener derecho a no participar en el uso de un sistema de IA, sin que esto afecte negativamente a su educación.

- *Derecho a no ser discriminado.* Todos los niños deberían tener la oportunidad de beneficiarse del uso de la tecnología, no solo aquellos de los grupos socioeconómicos que puedan permitírselo.
- *Derecho a la confidencialidad y a la protección de datos.* Los niños deben tener derecho a que sus datos no se agreguen y se utilicen con fines comerciales sin su beneficio directo.
- *Derecho a la transparencia y la explicabilidad.* Los niños y sus padres deben ser capaces de comprender y rechazar cualquier decisión tomada por un sistema de IA.”

Según (Holmes et al., 2022)

3.2 La diversidad de los retos éticos observada por la investigación

“Los retos éticos y críticos que plantea la IA en el ámbito educativo son múltiples y tienen diversos orígenes: uso **de datos masivos**, que pueden tener un impacto en **la vida privada** de los estudiantes y el personal escolar, producción por parte de empresas privadas en lugar de instituciones escolares, por la investigación en informática o en ciencias, tecnologías, ingeniería y matemáticas en lugar de ciencias educativas, **representatividad sesgada** por falta de diversidad en los equipos de diseño.” Según (Collin & Marceau, 2021)

3.3 El reto de la explicabilidad

“En el ámbito educativo, hay dos categorías de usuarios que deben dominar el uso del entorno digital: los alumnos y los profesores. **La cuestión de la apropiación de los EIAH [entornos informáticos para el aprendizaje humano] por parte de los profesores es fundamental para que estas herramientas se utilicen más en la enseñanza.** Por lo tanto, es necesario diseñar sistemas que los profesores puedan adaptar para satisfacer sus necesidades. Para **ello, el profesor debe ser capaz de comprender las decisiones o recomendaciones del sistema de IA sobre sus alumnos**, para tener confianza en el sistema y, de alguna manera, “ser uno” con él para poder explicar el comportamiento combinado del dispositivo tal y como se ha implementado. Esto requiere una representación explícita de los conocimientos y los procesos de decisión. La cuestión de la explicabilidad debe tenerse en cuenta desde el diseño.”

(Asociación francesa para la inteligencia artificial, AFIA, 2022)

4. ¿Cuáles son los campos de aplicación?

4.1 Categorización y taxonomía

Categorización propuesta por la Comisión Europea

Enseñanza	Apoyo al alumno	Apoyo al profesorado	Apoyo al sistema educativo
<ul style="list-style-type: none">Entorno informático para el aprendizaje humanoSistemas de tutoriales basados en el diálogoAplicaciones de ayuda al aprendizaje de idiomas	<ul style="list-style-type: none">Entornos de aprendizaje exploratorioEvaluación formativa de los escritosAprendizaje colaborativo asistido por IA	<ul style="list-style-type: none">Evaluación sumativa de los escritos, calificaciónIdentificación de palabras clave en los foros de alumnosAsistentes pedagógicos basados en IARecomendación de recursos educativos	<ul style="list-style-type: none">Exploración de datos educativos para la asignación de recursosDiagnóstico de las dificultades de aprendizajeServicios de orientación

Según (Comisión Europea. Dirección General de Educación, Juventud, Deporte y Cultura, 2022)

Otra clasificación realizada por los grupos temáticos digitales:

- Sistemas adaptativos y personalizados (recomendación y selección de contenido, recursos), evaluación adaptativa y corrección automática
- Robots de conversación y sistemas de tutoría inteligente, por ejemplo para la enseñanza de la iniciativa investigadora y científica (Cisel & Baron, 2019)
- Anticipación temprana, predicción y prevención del abandono escolar
- Detección y corrección de problemas de aprendizaje
- Uso de conocimientos generados por la investigación con IA (análisis de imágenes, anotación, clasificación, resumen/síntesis automáticas, procesamiento de datos masivos en corpus textuales o iconográficos)
- Usos administrativos

Según (DNE-TN2, 2021) (DNE-TN2, 2022) (Lepage Alexandre, 2022) (Romero et al., 2021)

Taxonomía según la revisión de la literatura (Holmes et al., 2022)

Ayuda dirigida al estudiante

- Sistemas inteligentes de tutoría (ITS)
- Aplicaciones asistidas por IA (por ejemplo, matemáticas, texto a voz, aprendizaje de idiomas)
- Simulaciones asistidas por IA (por ejemplo, aprendizaje basado en juegos, realidad virtual, realidad aumentada)
- IA para ayudar a los alumnos con discapacidades
- Escritura automática de ensayos (AEW)
- Chatbots
- Evaluación formativa automática (AFA)
- Orquestadores de redes de aprendizaje
- Sistemas de tutoría basados en el diálogo (DBTS)
- Entornos de aprendizaje exploratorio (ELE)
- Asistente de aprendizaje permanente asistido por IA

Ayuda dirigida al profesorado

- Detección de plagios
- Tratamiento inteligente de materiales de aprendizaje
- Supervisión en el aula
- Evaluación sumatoria automática
- Asistente de enseñanza de IA (incluido el asistente de evaluación)
- Orquestación en el aula

Ayuda dirigida a la institución

- Admisiones (p. ej., selección de estudiantes)
- Planificación de cursos, programación, horarios
- Seguridad escolar
- Identificación de abandonos y estudiantes en riesgo
- Supervisión electrónica

Según (Holmes et al., 2022) traducción por (Bocquet, 2023)

Por último, (Ouyang & Jiao, 2021) proponen una categorización según tres paradigmas:

	Fundamentos teóricos	Implementaciones	Técnicas de IA	Ejemplos
Paradigma 1: dirigido por la IA, el aprendiz como destinatario	Behaviorismo <i>[La IA representa el conocimiento del ámbito y dirige los procesos de aprendizaje.]</i>	Estudios previos sobre los sistemas de tutoría inteligentes	IA basada en técnicas relacionales estadísticas	Tutor de programación ACT (Anderson et al., 1990); Stat Lady (Shute, 1995)
Paradigma 2: respaldado por la IA, el aprendiz como empleado	Constructivismo cognitivo y social <i>[El sistema de IA como herramienta de apoyo, mientras que el alumno colabora para centrarse en el proceso de aprendizaje.]</i>	<ul style="list-style-type: none"> Sistemas de tutoría basados en el diálogo Entornos de aprendizaje exploratorio 	Red bayesiana, procesamiento del lenguaje natural, árboles de decisiones de Markov	Entorno exploratorio QUE (Metzler & Martincic, 1998)
Paradigma 3: la IA parte de un sistema en el que el alumno es piloto	Conectivismo, sistema adaptativo complejo <i>[La IA para aumentar la inteligencia humana, coloca al aprendiz en el centro del dispositivo.]</i>	<ul style="list-style-type: none"> Colaboración hombre-ordenador Aprendizaje personalizado/adaptativo 	Interfaz cerebro-ordenador - Aprendizaje automático, aprendizaje profundo	Modelización predictiva MOOC en tiempo real (Le et al., 2018)

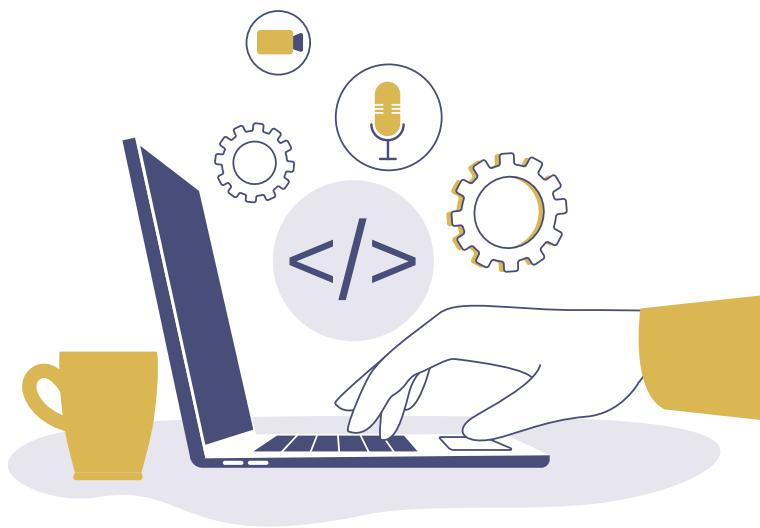
Tabla de (Ouyang & Jiao, 2021)

4.2 Qué aportaciones hay para la experiencia de aprendizaje?

"La inteligencia artificial (IA) cada vez desempeña un papel más importante en la educación. Un caso importante son **los sistemas de enseñanza personalizados** que ya están bien establecidos y **cada vez hay más pruebas de su eficacia para mejorar el aprendizaje** (VanLehn, 2011; Kulik & Fletcher, 2016; Steenbergen-Hu & Cooper, 2014, 2013; Ma, Adesope, Nesbit, & Liu, 2014; du Boulay, 2016). Los sistemas de IA en educación (AIED) también pueden hacer **un uso diverso y sofisticado de la IA para crear la interfaz que es tan importante para la experiencia de aprendizaje**. Por ejemplo, la interfaz puede usar **procesamiento y generación de lenguaje natural, interfaces de voz, avatares, análisis de vídeo del alumno** para juzgar su atención y emoción."

(Khosravi et al., 2022)

"Usando algoritmos, **el aprendizaje puede ser adaptativo**. Al analizar los rastros de aprendizaje del alumno, como los resultados de los cuestionarios o de los datos de uso de un software, el sistema puede modificar su funcionamiento para adaptarse a la persona, sobre todo mediante **la selección del contenido y el nivel de dificultad**. Está empezando a ser posible analizar su comportamiento mediante sensores, algunos externos como una cámara y otros más intrusivos como una interfaz cerebro-ordenador. Este principio de adaptación es una de las claves de la pedagogía digital y se encuentra con mayor frecuencia en un contexto en el que también se persiguen



objetivos de ludificación; el alumno entra, entonces, en un juego pedagógico con la máquina, a veces en colaboración con otros alumnos (Giraudon et al., 2020). El proyecto KidLearn (...) ofrece una actividad de aprendizaje cuyas múltiples variantes, que implican la adición o resta de números enteros o decimales, han sido diseñadas e implementadas por expertos en didáctica. Estas variantes se organizan en forma de un gráfico de dificultades crecientes, respetando el concepto de zona proximal de desarrollo (Vygotsky, 1978). Este concepto se basa en la idea de que, entre un ejercicio demasiado difícil que desalienta y un ejercicio demasiado fácil que aburre, existe un área óptima que maximiza el progreso del aprendizaje, medido, aquí, observando el rendimiento del alumno a lo largo del juego. Estos son los elementos que se integran en el algoritmo, que se adaptará automáticamente al aprendiz (Oudeyer et al., 2020)."

(Romero et al., 2023)

Título: Ejemplo de enseñanza adaptativa para el aprendizaje de la lectura y el francés

"Lalilo está desarrollando un asistente pedagógico para profesores de CP, CE1 y CE2 (educación primaria) para ayudarlos a **diferenciar su enseñanza de la lectura en función del progreso de cada alumno**, gracias a tecnologías de inteligencia artificial. El asistente es una plataforma en línea, con una interfaz de alumno y una interfaz de profesor. Se construye en colaboración con los/as profesores/as de primaria: la mejora de la plataforma es continua, teniendo en cuenta los comentarios de los profesores. (...)

Un algoritmo de aprendizaje adaptativo: El objetivo de Lalilo —ayudar a los niños a aprender a leer a su propio ritmo— se basa en un progreso pedagógico fijo, dividido en lecciones. Cada lección contiene varios "objetivos de aprendizaje", que a su vez contienen una amplia gama de preguntas. Por ejemplo, la lección "vocal [a]" contiene dos objetivos de aprendizaje: "reconocer el grafema a" y "descodificar el fonema [a]". (...) El sistema de aprendizaje adaptativo permite elegir el mejor ejercicio para un niño en un momento dado en función de su progreso y de los ejercicios realizados."

(Gelin, 2022)

[Lalilo es uno de los proyectos ganadores de la asociación para la innovación en inteligencia artificial (P2IA) apoyada por el Ministerio Nacional de Educación y Juventud.]

4.3 IA y producción de conocimientos: el ejemplo de las humanidades digitales

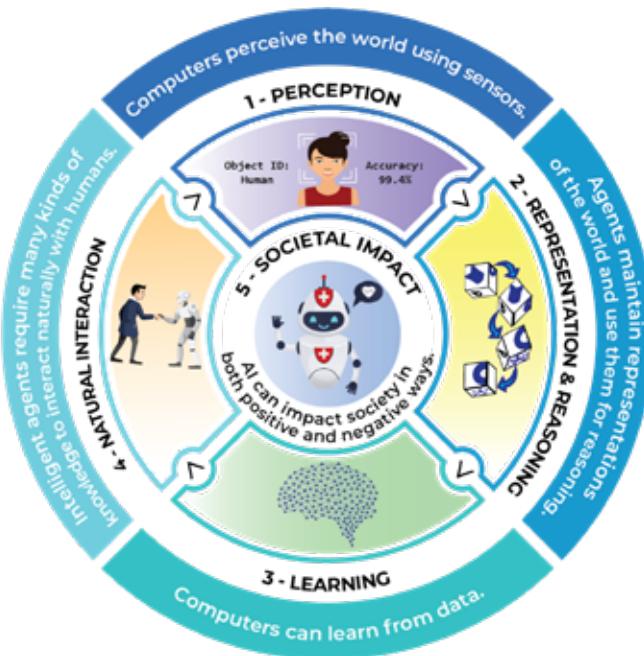
"La lingüística de corpus y la semántica han evolucionado considerablemente gracias a los análisis de IA sobre la representación del conocimiento y el aprendizaje automático. Del mismo modo, muchos desarrollos han abierto **nuevas perspectivas para los estudios literarios**. A principios del milenio, bajo la presión de los editores (y no de los científicos), el término inglés *Humanities Computing*, que caracterizaba la investigación en la frontera de las humanidades en el sentido anglosajón, es decir, las **ciencias de la cultura**, se transformó en *Digital Humanities* ("humanidades digitales"). Este cambio de nombre significaba que las humanidades ya no estaban simplemente instrumentadas por el uso de ordenadores, sino que se transformaban convirtiéndose en digitales. Se ha hablado de un "**giro de cálculo**" para caracterizar esta evolución considerada como fundamental. La renovación que siguió fue muy exitosa. En todos los países, muchos equipos de investigación que trabajan en el campo de las humanidades clásicas, ya sea en historia, arqueología o literatura, han renovado sus enfoques en colaboración con informáticos."

(Ganascia, 2022)

5

¿Cuáles son las pistas de trabajo para formarse en IA y enseñarla?

5.1 Las pistas propuestas por AI4K12



1. Percepción: Los ordenadores (y los terminales digitales) perciben el mundo a través de sensores.

2. Representación y razonamiento: Los agentes de IA mantienen representaciones del mundo, a través de estructuras de datos, y las utilizan para razonar, a través de algoritmos, pero sin pensar como un ser humano.

3. Aprendizaje: Los ordenadores (y los terminales digitales) pueden aprender de los datos a través de algoritmos de aprendizaje.

4. Interacción natural: Los agentes de IA requieren muchos tipos de conocimientos para interactuar "de forma natural" con los seres humanos (conversación, reconocimiento facial, etc.), pero no pueden razonar.

5. Impacto social: La IA puede tener un impacto positivo y negativo en la sociedad y debemos ser conscientes de ello, ya que estas tecnologías están cambiando la forma en que trabajamos, viajamos, nos comunicamos, etc.

Según (Association for the Advancement of Artificial Intelligence & Computer Science Teachers Association, 2020)

5.2 Marco de competencias de la alfabetización informacional de la IA

1. Reconocer la IA

Distinguir entre artefactos tecnológicos que utilizan o no utilizan IA

2. Comprender la inteligencia

Analizar y debatir de forma crítica las características que hacen a un ente "inteligente". Debatir sobre diferencias entre la inteligencia humana, animal y de las máquinas.

3. Interdisciplinariedad

Reconocer que existen varias formas de concebir y desarrollar máquinas "inteligentes". Identificar una variedad de tecnologías que utilizan IA, incluida la tecnología que comprende los sistemas cognitivos, la robótica y el AA.

4. IA general versus IA débil

Distinguir entre IA general y débil.

5. Debilidades y fortalezas de la IA

Identificar los tipos de problemas en los que destaca o no la IA. Determinar cuándo es apropiado utilizar IA y cuándo aprovechar las habilidades humanas.

6. Imaginar la IA del futuro

Imaginar posibles aplicaciones futuras de la IA y considerar los efectos de dichas aplicaciones para con el mundo.

7. Representaciones

Comprender qué es una representación del conocimiento y describir algunos ejemplos.

8. Toma de decisiones

Reconocer y describir ejemplos de cómo razonan y toman decisiones las computadoras.

9. Pasos del AA

Comprender los pasos involucrados en el aprendizaje automático y las prácticas y desafíos que conlleva cada paso.

10. El rol humano en la IA

Reconocer que los humanos desempeñan un papel importante en la programación, la elección de modelos y el ajuste preciso de los sistemas de IA.

11. Alfabetización en datos

Comprender los conceptos básicos de la alfabetización en datos.

12. Aprender de los datos

Reconocer que las computadoras frecuentemente aprenden de los datos (incluidos los datos propios).

13. Interpretación crítica de los datos

Comprender que los datos requieren de interpretación. Describir de qué modo los ejemplos de instrucción incluidos en un conjunto inicial pueden afectar los resultados de un algoritmo.

14. Acción y reacción

Comprender que algunos sistemas de IA cuentan con la capacidad de actuar físicamente en el mundo. Esta acción puede estar orientada por un razonamiento d'alto nivel (como caminar a lo largo de una ruta planificada) o por impulsos reactivos (como saltar hacia atrás para evitar un obstáculo detectado).

15. Sensores

Comprender qué son los sensores, entendiendo que las computadoras perciben el mundo a través de ellos. Identificar sensores en diversos dispositivos. Reconocer que diferentes sensores soportan distintos tipos de representación y razonamiento respecto del mundo.

16. Ética

Identificar y describir diferentes perspectivas sobre las cuestiones éticas clave en torno a la IA: privacidad, empleo, desinformación, "singularidad",¹¹ toma de decisiones, diversidad, parcialidad, transparencia y responsabilidad.

17. Programabilidad

Comprender que los agentes son programables.

Según (Long y Magerko, 2020) citado por (UNESCO, 2022)

5.3 Para la formación de la ciudadanía en IA

"La IA afecta a todos los ciudadanos, en primer lugar porque cada vez está más presente en todos los sectores de la sociedad y, en segundo lugar, **porque es necesario comprenderla para participar en los debates éticos** que genera. **Es normal plantearse la cuestión de si es conveniente confiar a algoritmos tareas que conduzcan a decisiones cruciales**, por ejemplo, en materia de justicia, de contratación u otras situaciones con grandes consecuencias humanas. (...) **El MOOC ["inteligencia artificial con inteligencia" de Inria] adopta un enfoque pedagógico lúdico y experimental.** (...) Los participantes se ponen en acción en actividades concretas: manipulan redes neuronales, intentan hacer que sus dibujos sean reconocidos por una IA, se les pide que practiquen ellos mismos modelos de aprendizaje automático. Se les invita a reflexionar colectivamente, a través de foros de debate, sobre cuestiones planteadas por el desarrollo de la IA".

(Romero et al., 2023)

Mensaje de respuesta de ChatGPT - módulo de formación (adjunto)
Actualizado con nuevo título: Producción de un módulo de iniciación a la IA por parte de tres agentes de conversación

Ilustración a insertar en 5.3 en un cuadro

Igual que la edición de mayo de 2023, pero actualizada (pruebas del 06/12/23 y 07/12/23)

Leyenda:

Producción de un módulo de iniciación a la IA por parte de tres agentes de conversación
Respuesta de ChatGPT (06/12/23)

Instrucciones (prompt)

En el contexto de la enseñanza escolar en Francia. Ofrecer, en forma de tabla, un módulo de formación de aproximadamente una hora de iniciación a la inteligencia artificial para profesores (primaria y secundaria). Este módulo debe presentar los orígenes de la IA, las definiciones básicas, los retos para la educación, los retos éticos y ejemplos de aplicación pedagógica.

Respuesta

"Por supuesto, he aquí una propuesta de módulo de iniciación a la inteligencia artificial para profesores de primaria y secundaria en Francia, presentada en forma de tabla:

Tiempo	Contenido	Objetivo
0-5 min	Introducción	Sensibilizar a los profesores sobre la importancia de la inteligencia artificial en el mundo moderno. Presentar brevemente los orígenes de la IA.
5-15 min	Definiciones básicas	Explicar los conceptos básicos de la inteligencia artificial, como algoritmos, aprendizaje automático y redes neuronales. Utilizar términos sencillos para garantizar una comprensión básica.
15-25 min	Retos para la educación	Hablar sobre el posible impacto de la IA en la educación. Destacar los beneficios, como la personalización del aprendizaje, y los desafíos, como la necesidad de adquirir nuevas habilidades.
25-35 min	Retos éticos	Revisar las cuestiones éticas relacionadas con el uso de la IA en la educación, haciendo hincapié en la privacidad de los datos, la discriminación algorítmica y la transparencia. Fomentar la reflexión sobre la responsabilidad ética de los profesores.
35-45 min	Ejemplos de aplicación pedagógica	Presentar ejemplos reales de uso de la IA en la educación, como tutoriales inteligentes, análisis automatizado del rendimiento de los alumnos y herramientas de apoyo al aprendizaje. Animar a los profesores a pensar en cómo podrían integrar estas tecnologías en su enseñanza.
45-55 min	Preguntas y debate	Invitar a los profesores a hacer preguntas y compartir sus preocupaciones. Facilitar un debate sobre cómo la IA puede integrarse positivamente en el entorno educativo.
55-60 min	Conclusión	Resumir los puntos clave del módulo. Proporcionar recursos adicionales para ampliar los conocimientos. Animar a los profesores a mantenerse informados sobre los futuros desarrollos de la IA en la educación.

El objetivo de este módulo es ofrecer una introducción equilibrada y accesible a la inteligencia artificial, centrándose al mismo tiempo en los aspectos relevantes para los profesores de primaria y secundaria en Francia".

Fuente: <https://chat.openai.com/chat>

Respuesta de Llama (07/12/23)

Instrucciones (prompt)

En el contexto de la enseñanza escolar en Francia. Ofrecer, en forma de tabla, un módulo de formación de aproximadamente una hora de iniciación a la inteligencia artificial para profesores (primaria y secundaria). Este módulo debe presentar los orígenes de la IA, las definiciones básicas, los retos para la educación, los retos éticos y ejemplos de aplicación pedagógica.

Respuesta

A continuación se muestra un ejemplo de un módulo de formación de inteligencia artificial de aproximadamente una hora para profesores:

Duración	Parte	Descripción
10 min	Introducción	Presentación del módulo y del objetivo: introducir a los profesores a la inteligencia artificial y sus aplicaciones pedagógicas.
20 min	Origen de la IA	Presentación de los orígenes de la inteligencia artificial, los primeros conceptos y los desarrollos históricos.
20 min	Definiciones básicas	Explicación de los términos clave como aprendizaje automático, <i>deep learning</i> , <i>big data</i> , etc.
15 min	Retos para la educación	Debate sobre los retos de la inteligencia artificial para la educación, incluidos la personalización del aprendizaje, la eficacia de los métodos pedagógicos y la accesibilidad de los recursos educativos.
20 min	Retos éticos	Presentación de los retos éticos relacionados con la inteligencia artificial, como la privacidad, los sesgos, la responsabilidad, etc.
25 min	Ejemplos de aplicación pedagógica	Presentación de ejemplos reales de aplicaciones de inteligencia artificial en la educación primaria y secundaria, como el uso de chatbots para ayudar a los estudiantes a aprender, el análisis de datos para mejorar la gestión del aula, etc.
10 min	Conclusión	Repaso de la sesión y propuesta de acciones para avanzar en la exploración de la inteligencia artificial en la enseñanza.

Este módulo se puede adaptar a las necesidades y niveles de los profesores, y se puede complementar con recursos adicionales como vídeos, artículos y ejercicios prácticos.

Fuente: <https://poe.com/>

Respuesta de Claude (07/12/23)

"A continuación se muestra una tabla con un módulo de formación de iniciación a la inteligencia artificial de aproximadamente una hora de duración para profesores:

Duración	Contenido	Objetivo
5 min	Introducción: presentación del módulo y su objetivo de sensibilización ante la IA	Situar el marco del curso de formación.
10 min	Historial y definiciones básicas de la IA	Comprender los orígenes y conceptos clave de la IA.
10 min	Retos de la IA para la educación	Identificar el potencial y los retos de la IA para la enseñanza y el aprendizaje.
10 min	Retos éticos de la IA	Tomar conciencia de las cuestiones éticas planteadas por el desarrollo de la IA
15 min	Ejemplos concretos de aplicaciones pedagógicas de la IA	Ilustrar el potencial de la IA para enriquecer las prácticas pedagógicas.
10 min	Resumen y perspectivas	Resumir los puntos clave tratados y pasar a lo siguiente.

Este módulo aborda de forma resumida, pero concreta, los fundamentos de la IA y su impacto en la educación para concienciar a los profesores sobre las oportunidades y los retos que plantea esta tecnología. Se prevén tiempos de debate para responder a las preguntas."

Fuente: <https://poe.com/>

5.4 Una propuesta de itinerario en cinco pilares

"La inteligencia artificial no ha alcanzado su madurez. El tema tal y como se definió en 1956, estudiado durante 40 años, ha alcanzado sus espectaculares resultados desde 2012 y aún es difícil de comprender. Es aún más difícil predecir cómo evolucionarán las tecnologías, incluso en un futuro próximo. Aunque la construcción de un programa de estudios completo está fuera del alcance de este documento, se pueden proponer cinco pilares y apoyarse en ellos. (...)

1. Incertidumbre y aleatoriedad

Los datos son incoherentes. No demuestran una naturaleza estrictamente causal.

2. Codificación y cálculo

La codificación y el pensamiento informático forman parte hoy en día de los programas de estudio de muchos países, de acuerdo con las recomendaciones de los expertos.

3. Sensibilización sobre los datos

La educación basada en datos (ciencia de los datos o *data science*) se basará en actividades en las que los datos se recopilan y visualizan, manipulan y analizan.

4. Pensamiento crítico

Las ciencias sociales pueden y deben contribuir a muchas de las cuestiones éticas planteadas por la IA. El pensamiento crítico es un aspecto importante, pero es esencial que se base en una comprensión real de cómo funciona la tecnología.

5. Humanismo post-IA

La idea clave es que los avances de la IA nos hacen reconsiderar ciertas verdades básicas como seres humanos."

Según (de la Higuera & Bocquet, 2020)

5.5 Formación en IA y prevención del abandono escolar

"La **Scientoteque** es una asociación sin ánimo de lucro que se encuentra en el Experimentarium, museo de Física de la Universidad Libre de Bruselas. Desde su fundación en 2001, su principal misión ha sido contribuir a reducir las desigualdades sociales. Dos hechos llevaron a la creación de la asociación: por un lado, el mayor riesgo de abandono escolar y la dificultad de acceso a la educación superior para los jóvenes de entornos desfavorecidos (Coslin, 2012) y, por otro, el desinterés por los estudios científicos por parte del público femenino. Además, se ha demostrado que es menos obvio atraer a las niñas a proyectos científicos o tecnológicos (Blanchard, 2021). (...) Se ofreció una selección de actividades a dos grupos de jóvenes de 10-12 años durante el año escolar 2020-2021 en **talleres extraescolares semanales** en el Institut des Ursulines de Koekelberg. Este proyecto, financiado por la región de Bruselas-Capital, tenía como objetivo reducir el abandono escolar de estudiantes que vivían en barrios desfavorecidos. A lo largo de las sesiones, los jóvenes pudieron, a través de un **enfoque lúdico y colaborativo**, descubrir la historia de las máquinas, establecer conexiones con la inteligencia biológica, comprender lo que hay detrás de la palabra "algoritmo", observar y programar robots Thymio, aprender a programar en la plataforma Scratch y debatir las consecuencias de la IA en la sociedad."

(Romero et al., 2023)

5.6 El proyecto europeo "AI4T": Mooc "Inteligencia artificial para y por los profesores"

"El objetivo de este curso es ayudar a los profesores y al conjunto de la comunidad educativa a comprender la IA en general, sus usos, tecnologías y limitaciones, y a guiarlos hacia un uso inteligente de las herramientas basadas en la IA en el contexto educativo.

El Mooc está organizado en cuatro módulos:

- Módulo 1: Uso de la IA en la educación
- Módulo 2: ¿Qué es la IA?
- Módulo 3: ¿Cómo funciona la IA?
- Módulo 4: ¿La IA a nuestro servicio como profesores?

Cada módulo incluye vídeos, textos, actividades, cuestionarios y acceso al foro para compartir sus experiencias e interactuar con el equipo de aprendizaje y los demás participantes. El recorrido completo puede completarse en dos o tres horas.

Antes y después de estos 4 módulos del curso, la sección "Presentación general" presenta los objetivos de la formación y le guía en el uso de la plataforma del curso.

La sección “Conclusión” aborda los aspectos éticos relacionados con el uso de la IA en la enseñanza. (...)



Este curso se ha desarrollado y experimentado con profesores y jefes de centros educativos como parte del **proyecto europeo AI4T, que reúne a cinco países: Francia, Italia, Irlanda, Luxemburgo y Eslovenia**.

En este contexto, los contenidos se desarrollaron con los cinco ministerios de educación involucrados en el proyecto y, en particular, con la MENJ (Dirección de las tecnologías digitales para la educación)."

El MOOC va acompañado de un manual abierto AI4T (*Open TextBook*) a través de la cátedra UNESCO RELIA:
<https://chaireunescorelia.univ-nantes.fr/2023/11/16/le-manuel-ouvert-ai4t-en-route-pour-la-deuxieme-edition/>
<https://pressbooks.pub/iapourlesenseignants>

Según (Inria & AI4T, 2023)



6. El cambio en los sistemas de IA generativa y los grandes modelos de lenguaje

6.1 El desarrollo de la IA generativa

Presentación general

"Antes de 2014, todos los modelos de aprendizaje profundo existentes eran principalmente descriptivos, centrándose en el resumen o los esquemas y las relaciones de datos existentes. El objetivo de estos modelos era explicar los patrones de datos y predicciones basados en la información presente. Sin embargo, (Goodfellow et al., 2014) introdujeron la **Generative Adversarial Network (GAN)**, lo que abrió una nueva era de la inteligencia artificial generativa (GAI). A diferencia de sus homólogos descriptivos, los modelos generativos, como los GAN, están diseñados para aprender la distribución de probabilidad subyacente de los datos. Su objetivo principal es **generar nuevas muestras de datos que se asemejen estrechamente a los modelos observados en los datos de aprendizaje**. El avance de la GAN marcó una disrupción significativa de los métodos tradicionales de aprendizaje profundo, abriendo interesantes posibilidades para la inteligencia artificial generativa. Desde entonces, la inteligencia artificial generativa ha atraído mucha atención debido a su **impacto transformador** en diversas áreas de la vida. Ofrece soluciones elegantes a problemas complejos, lo que permite crear datos sintéticos, contenido artístico e imágenes realistas. Este cambio de paradigma en la tecnología de IA ha influido profundamente en la nueva percepción, la implementación y uso de la inteligencia artificial, impulsando la innovación y nuevas posibilidades de aplicación en todos los sectores." (...) El trabajo innovador de Vaswani et al. "**Attention Is All You Need**" [2017] del equipo de Google Brain presentó **un modelo de transformador [Transformer]** capaz de analizar conjuntos de datos a gran escala. [Este modelo] se desarrolló inicialmente para el procesamiento natural (NLP), pero después se adaptó a otras áreas del aprendizaje automático, como la visión por ordenador. (...) Por lo tanto, la solución propuesta ofrecía una nueva forma de gestionar tareas como la traducción, la generación de contenido y el análisis emocional."

(Bengesi et al., 2023)

"En los últimos dos años, se han publicado **numerosos modelos generativos importantes** como ChatGPT o Stable Diffusion. En concreto, estos modelos son capaces de realizar tareas como un sistema general de preguntas y respuestas o la creación automática de imágenes artísticas **que revolucionan varios sectores**. Por lo tanto, las implicaciones que estos modelos generativos tienen en la industria y la sociedad son considerables, ya que se pueden transformar varias tareas laborales.

Por ejemplo, la IA generativa puede convertir de forma eficaz y creativa texto en imágenes, como el modelo DALLE-2; texto en imágenes 3D, como el modelo Dreamfusion; imágenes en texto, como el modelo Flamingo; texto en vídeo, como el modelo Phenaki; texto en audio, como el modelo AudioLM; texto en otro texto, como ChatGPT; texto en código, como el modelo Codex; textos en textos científicos, como el modelo Galactica, o incluso crear algoritmos como AlphaTensor."

La **IA generativa** se refiere a las tecnologías de IA basadas en datos masivos (un conjunto de Wikipedia, Github, redes sociales, imágenes de Google, etc.) **que pueden generar contenido nuevo** (texto, imagen, audio y vídeo) de forma aleatoria a partir de información de entrada (o "prompt") introducida por el usuario.

Los **modelos de IA generativa** (redes neuronales profundas, *transformers* y otros modelos, como las redes antagonistas generativas [GAN: *generative adversarial networks*]) que se han multiplicado principalmente desde 2022, se pueden dividir de acuerdo con la siguiente taxonomía:

- Texto a imagen (p. ej., DALL-E 2 de OpenAI)
- Texto con representaciones en 3D (p. ej. Dreamfusion de Google Research)
- Imagen a texto (p. ej., Flamenco de Deepmind)
- Texto en vídeo (p. ej., Phenaki de Google Research)
- Texto a audio (Jukebox de OpenAI)
- Texto a texto (por ejemplo, ChatGPT de OpenAI)
- Texto a código (p. ej., OpenAI Codex)
- Texto con fórmula científica (p. ej., Galactica de Meta AI)
- Otros modelos

Según (Gozalo-Brizuela y Garrido-Merchan, 2023) [Traducción]

Traducción — utilizar el texto original en inglés como base

<https://arxiv.org/pdf/2301.04655.pdf>

pág. 1 (resumen), 2 y a partir de la pág. 5 (taxonomía)

Los modelos de fundación

"Un modelo de fundación (*Foundation Model*), denominación introducida por la Universidad de Stanford, es un modelo de gran tamaño basado en una arquitectura de red neuronal profunda, entrenado en una gran cantidad de datos sin anotaciones (normalmente mediante aprendizaje autosupervisado). Los grandes modelos de lenguaje (LLM, por *Large Language Model*) son casos particulares de modelos de fundación que se entranan con un corpus de textos. Estos modelos de fundación abren **nuevas perspectivas e introducen un nuevo paradigma en el procesamiento del lenguaje, pero también en el procesamiento de señales multimodales** (sonido, imagen, vídeo, etc.). Estos modelos pre-entrenados con corpus voluminosos se pueden optimizar para realizar una nueva aplicación utilizando pocos datos adicionales específicos para dicha tarea."

(CNPEN, 2023)

6.2 Los grandes modelos de lenguaje y la nueva generación de agentes de conversación

¿Cómo funcionan los grandes modelos de lenguaje?

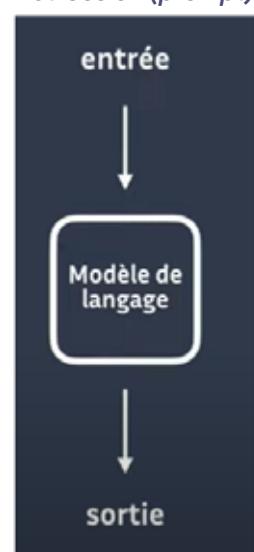
El principio de los modelos de lenguaje (o LLM: *Large Language Models*) **no es nuevo**: este tipo de algoritmo ya está presente a diario con **los asistentes de redacción de mensajes instantáneos**. Su función consiste **en predecir la siguiente palabra** a partir de una instrucción ("prompt") introducida por el usuario. La **fase de entrenamiento** permite ajustar los parámetros internos a partir de datos textuales procedentes de la web (no verificados). Una gran cantidad de datos y parámetros permite mejorar el rendimiento de la predicción. Sin embargo, no se puede hablar de "comprensión" ni de "personalidad" del algoritmo, se trata en primer lugar de **respuestas producidas a partir de regularidades estadísticas**. Además, estos modelos pueden producir "**alucinaciones**", inventando información falsa, lo que requiere la mayor atención por parte del usuario. Actualmente, los modelos de lenguaje más potentes son propiedad de grandes empresas, que comunican muy poco sobre sus datos de entrenamiento.

Según (Inria Flowers, 2023a)

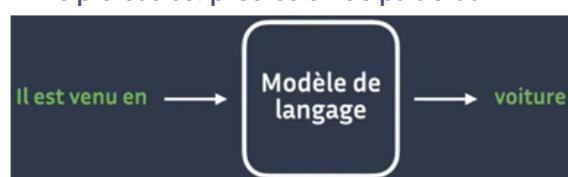
Véase también (Inria Flowers, 2023b) sobre el "prompting": "hacer que un modelo de lenguaje realice una determinada tarea (...) explicándole esta tarea mediante frases en lenguaje natural." Se presentan dos métodos: "explicar la tarea en lenguaje natural y dar ejemplos de realización de la tarea."

Funcionamiento de un gran modelo de lenguaje y *prompting*

Instrucción (*prompt*)



Principio básico: predicción de palabras



Comment faire réaliser une tâche
à un modèle de langage ?

Inria
flowers

Según (Inria Flowers, 2023)

El “cambio de ChatGPT”: orígenes y base tecnológica

En esta categoría de IA generativas, la aplicación ChatGPT, abierta al público por la empresa OpenAI desde el 30 de noviembre de 2022, representa una **nueva generación de agentes de conversación** (chatbot en inglés), “agente que dialoga con un usuario con el fin de realizar una búsqueda, encontrar una respuesta a su pregunta o incluso practicar en una disciplina determinada.” (CSEN y DNE-TN2, 2022). Debido al rendimiento mostrado, es obviamente un **“salto cualitativo”** en las tecnologías de procesamiento automático del lenguaje (TAL) — o **análisis del lenguaje natural (NLP)** (Boisson y DNE-TN2, 2023).

Este **“prototipo de agente conversacional (...)”**, un modelo de lenguaje perfeccionado utilizando técnicas de aprendizaje supervisado y aprendizaje mediante refuerzo” (“ChatGPT”, 2023), basado en una inteligencia artificial generativa, capaz de generar textos a la demanda o código informático a partir de preguntas de lenguaje natural, es una **“aplicación patentada desarrollada”** a partir de una optimización del modelo algorítmico GPT 3.5.

“El GPT-3 es más importante que el GPT-2, con 175 000 millones de parámetros. Para tener una idea más precisa de la dimensión de este algoritmo, los datos de Wikipedia, que incluyen aproximadamente 6 millones de artículos, solo representan el 0,06 % de los datos generados por el GPT-3”.

(Boa Sorte et al., 2021)

Las tecnologías que conforman ChatGPT se pueden resumir de la siguiente manera:

- se trata de un **“modelo de lenguaje ampliado”**, un tipo de inteligencia artificial que utiliza el aprendizaje profundo (una forma de aprendizaje automático) para procesar y generar textos en lenguaje natural (...) [Este tipo de modelo está] formado en **grandes cantidades de datos textuales, lo que le permite aprender los matices y las complejidades del lenguaje humano**” (según (Susnjak, 2022) [traducción]);
- su denominación resulta de una contracción entre “chat” (conversación) y “GPT” [*Generative Pre-trained Transformer*], que es un “modelo de transducción de lenguaje predictivo” (transducción: transformación de un dato representado en una forma en un dato representado en otra forma);
- su algoritmo basado en un **“transformer”** (Vaswani et al., 2017) corresponde a un modelo de arquitectura de red de neuronas que utiliza **el mecanismo de autoatención que permite establecer relaciones entre las palabras y determinar el significado de una palabra en función del contexto**, centrándose en las palabras más relevantes, permitiendo así acercarse de forma creíble al lenguaje natural;
- funciona según **un modelo de predicción de palabras**, a partir de instrucciones/ consignas o “*prompts*” del usuario en lenguaje natural, combinando en su fase de entrenamiento aprendizaje por refuerzo y aprendizaje supervisado a partir de retroalimentaciones humanas;
- por lo tanto, es importante recordar que, como cualquier tecnología de IA, los textos producidos se generan y **refuerzan a partir de acciones y trabajo humanos** (desde las personas que diseñan el algoritmo hasta las que identifican, verifican los datos o interactúan con el programa).

La base tecnológica de ChatGPT es el resultado de: “**una combinación de dos modelos**: un **modelo lingüístico** (cuyas referencias se detienen en 2021), GPT, un modelo gigante de predicción de texto y un modelo enciclopédico, y un **modelo conversacional** (que se afina en función de los comentarios de los usuarios), InstructGPT, una versión “rectificada” de GPT-3 a partir de la anotación de textos generados”.

También es el resultado de “una serie **de capas de investigación lingüística e informática**:

- **la estadística semántica**, con el concepto de “ventana contextual” (el sentido procede de la posición relativa de las palabras entre sí);
- **la vectorización de palabras** (*word embedding* o inmersión lexical) que permite caracterizar una palabra dentro de espacios contextuales en función de las palabras vecinas;
- **las redes de neuronas** para reconocer las relaciones sintácticas y semánticas precisas entre las palabras y gestionar las transferencias lingüísticas (las respuestas a las instrucciones de los usuarios pueden basarse en un espacio semántico multilingüe).
- los **modelos transformers** (2017) basados en el mecanismo de atención y que modelizan las fórmulas sintácticas;
- el desarrollo **de los grandes modelos de lenguaje**, a partir de 2018 con BERT (Google) y después GPT-2 y 3 (OpenAI).“

Según (Langlais, 2023)

El 14 de marzo de 2023, OpenAI anunció la actualización de su modelo con el lanzamiento de **GPT-4** (servicio de pago hasta la fecha o disponible en la nueva versión del motor de búsqueda de Bing): “**modelo multimodal** a gran escala que puede aceptar entradas de imagen y texto y producir salidas de texto. Aunque es menos capaz que los humanos en muchos escenarios del mundo real, GPT-4 tiene un rendimiento de nivel humano en varios criterios profesionales y académicos, como aprobar un examen simulado del colegio de abogados con una puntuación próxima al 10 % de los mejores candidatos.” (OpenAI, 2023)

Los otros modelos de lenguaje importantes

Otros modelos de lenguaje importantes son, entre otros, de código abierto:

- **Bert** (*Bidirectional Encoder Representations from Transformers*), desarrollado por Google en 2018 (Devlin et al., 2019);
- **Bloom**, iniciado por la start-up francoamericana Hugging Face, entrenado en la máquina “Jean Zay”, “una de las supercomputadoras más potentes de Europa” (CNRS, 2022);
- **LLaMA** de Meta AI (Touvron et al., 2023).

Para obtener más información, consulte también (Departamento de especialización en regulación digital, 2023).

Ejemplos de grandes modelos de lenguaje de última generación y políticas de apertura asociadas

	Año	Nº Máximo de parámetros (en mds)	Arquitectura pública	Modelo entrenado abierto	Datos abiertos	Conversacional (RLHF)	Accesible a los usuarios (UI ou API)
BigScience BLOOM	2022	175	✓	✓	✓	✗	✓
Google GLaM/PaLM	2021/2022	1200/540	✓	✗	✗	✗	✓ API de pago
Google LaMDA/Bard	2022	137/?	✓/✗	✗	✗	✓	✓ RU / EE.UU.
Meta OPT	2022	175	✓	✓	✓	✗	✗
Meta BlenderBot3	2022	175	✓	✓	✓	✓	✓ EE.UU.
Meta LLaMA	2023	65	✓	✓	✓	✗	✗
OpenAI GPT-3	2020	175	✓	✗	✓ 1ª versión solo	✗	✓ opción de pago
OpenAI GPT-3.5 (InstructGPT /ChatGPT)	2022	175/?	✓/✗	✗	✗	✓	✓ opción de pago
OpenAI GPT-4	2023	?	✗	✗	✗	✓	✓ API y UI de pago

(Departamento de especialización de la regulación digital, 2023)

En diciembre de 2023, Google presenta **Gemini** (Red Generalizada de Inteligencia Multimodal) su nueva inteligencia artificial multimodal (Hsiao, 2023).

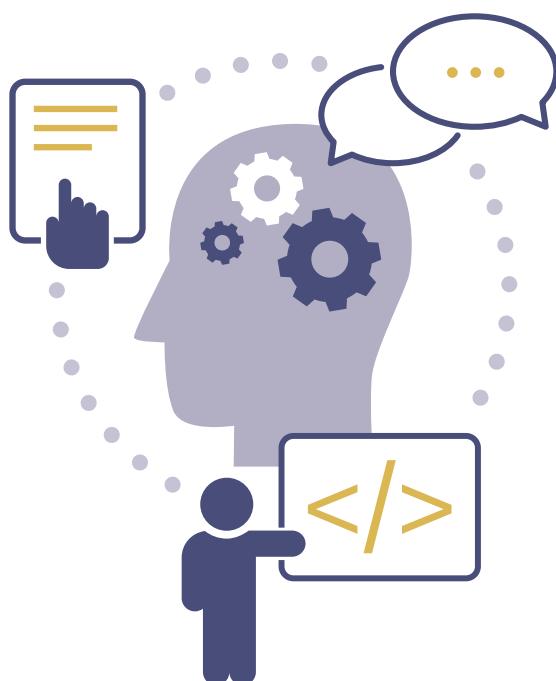
Las reacciones y preguntas

Desde que se abrió la aplicación ChatGPT, ha habido muchas reacciones y debates en los medios de comunicación y en muchos entornos profesionales, entre los que se incluyen la investigación y la educación.

"Estas IA generativas multilingües, multitareas e interactivas plantean **muchas incógnitas**: los datos elegidos para entrenarlos, la distribución de los idiomas en un sistema multilingüe, los parámetros de optimización de los sistemas, la propiedad del contenido generado, etc. Además, el poder generativo de la IA a menudo se amplifica con **filtros que permiten censurar determinados temas** y módulos de deducción lógica para verificar la veracidad de las afirmaciones. Un grupo reducido de personas (ingenieros, transcritores y evaluadores) han creado este tipo de sistema que utilizan millones de personas. Por lo tanto, estos sistemas de inteligencia artificial utilizados masivamente plantean **grandes desafíos éticos**, entre los que se incluyen la transformación de la noción de producción de información, la relación con la verdad y los enormes riesgos relacionados con la desinformación y la manipulación."

(Devillers, 2023)

Las reacciones fueron **de fascinación**, frente al carácter creíble y plausible de las "conversaciones" producidas y a la pertinencia —al menos aparente— de algunas respuestas, **a la preocupación** por los posibles malos usos (posibilidad de engaño para realizar deberes escolares, fraude en los exámenes, duda sobre la autoría de los escritos producidos en un determinado contexto personal o profesional) y **a las incertidumbres sobre las tecnologías, los datos** movilizados y la fiabilidad de las respuestas producidas.



¿Cuál es su potencial pedagógico?

Este tipo de aplicación ofrece varios potenciales pedagógicos que la comunidad de profesores está explorando activamente y compartiendo actualmente:

- una asistencia **en ingeniería pedagógica** (diseño de secuencias de enseñanza y aprendizaje, véase la propuesta de matriz a continuación);
- **un apoyo a la creatividad** proponiendo, a partir de las consultas y preguntas planteadas, una primera base de redacción o de código;
- un **apoyo al aprendizaje del francés y de idiomas extranjeros** mediante la función de seguimiento conversacional;
- una **asistencia** para iniciar u optimizar una tarea de redacción;
- un **nuevo reto para la educación** en los medios de comunicación y la información (**EMI**) y un marco para la formación de mente crítica, que requiere que los profesores y sus alumnos comprueben sistemáticamente el contenido y la pertinencia de las respuestas proporcionadas.

Estas capacidades ya han sido objeto de varias publicaciones, entre las que (Kasneci et al., 2023) destaca las siguientes posibilidades: aprendizaje personalizado (análisis del escritos y sugerencia de materiales educativos adaptados), planificación de cursos (ayuda a la creación de planes y actividades inclusivas), aprendizaje de idiomas (resúmenes y traducciones).

Además, hay varias **guías pedagógicas** que ofrecen pistas de uso para el profesor. Así, la guía producida por (Herft, 2023), profesor australiano, y traducida al francés, propone una serie de ejemplos de instrucciones que se deben introducir para convertirla en un asistente pedagógico y didáctico. Por ejemplo: "Cree un plan formativo para una clase... que cubra el concepto de... e incluya una variedad de actividades y evaluaciones".

Prompting e ingeniería pedagógica: ejemplo de matriz para una hoja de instrucciones

Información	Situación pedagógica	Campo por completar
Identidad Contexto Disciplina/área de aprendizaje	Soy profesor de [...] en [tal tipo de institución]	
Marco de acción Limitaciones Dificultades	La institución o los alumnos tiene(n) [tal perfil], han de frente a [tal dificultad]	
Finalidades Objetivos Público objetivo/nivel Acciones objetivo	Enseño a estudiantes de [tal nivel, tal edad] - quiero ayudarles a progresar o enseñarles [tal temática, tal punto del programa, tal competencia]	
Pregunta Solicitud	Quiero construir [tal secuencia], producir [tal recurso, tal situación de aprendizaje] para/con mis alumnos.	
Estilo Formato Palabras clave Etapas	Deseo la respuesta en forma de [lista/tabla...], para producir una imagen, un esquema, en un lenguaje adaptado a [tal nivel], utilizando las siguientes palabras clave:... Eventualmente, completar con un ejemplo de la producción esperada. En caso necesario, detallar las etapas a seguir para mejorar el resultado (prompting por "cadena de pensamiento").	

(Allouche, 2023b)

Funciones de ChatGPT para ayudar a los profesores en sus prácticas.

Aspecto	Función	Citas representativas
Preparación para la enseñanza	Generación de materiales del curso	"Pedimos a ChatGPT que creara uno de los diálogos anteriores en un formato que pudiera importarse a DialogFlow. ChatGPT fue capaz de generarlo correctamente"
	Hacer sugerencias	"Le dije a ChatGPT que el alumno tenía dislexia y, finalmente, le recomendó materiales de aprendizaje específicos"
	Realizar traducciones	"ChatGPT se puede utilizar para traducir material educativo a diferentes idiomas"
Evaluación	Generación de tareas de evaluación	"Otro uso de ChatGPT es su capacidad para generar ejercicios, cuestionarios y escenarios que se pueden utilizar en el aula para ayudar en la práctica y la evaluación"
	Evaluación del rendimiento de los alumnos	"Se puede entrenar a ChatGPT para corregir las redacciones de los estudiantes y así, los profesores disponen de más tiempo para centrarse en otros aspectos de la enseñanza"

Según (Lo, 2023)

Las perspectivas de los grandes modelos de lenguaje para las personas con discapacidad

"Para empoderar a los alumnos con discapacidades, se pueden utilizar modelos de lenguaje en combinación con soluciones de voz a texto o de texto a voz para ayudar a las personas con discapacidad visual. En combinación con las oportunidades de tutoría en grupo y a distancia mencionadas anteriormente, los modelos lingüísticos se pueden utilizar **para desarrollar estrategias de aprendizaje inclusivas** con el apoyo adecuado en tareas como la redacción adaptativa, la traducción y el resalte de contenido importante en varios formatos. Sin embargo, es importante tener en cuenta que el uso de modelos lingüísticos grandes **debe ir acompañado de la ayuda de** profesionales como logopedas, educadores y otros especialistas que puedan adaptar la tecnología a las necesidades específicas de las discapacidades del alumno."

(Kasneci et al., 2023)

¿Cuáles son los límites y los puntos de vigilancia (dimensiones éticas, pedagógicas y científicas)?

Se realizaron varias pruebas para evaluar el rendimiento y los límites de ChatGPT (GPT 3.5) (Allouche, 2023a). El siguiente diagrama propone una modelización de un recorrido de usuario crítico, que integra instrucciones que pueden ser sencillas o complejas, desde una simple solicitud de información hasta una problematización.

Modelización de recorrido de usuario crítico con un agente de conversación



(Allouche, 2023a) actualizado

Al final de estas pruebas, se pueden señalar varios límites y puntos de vigilancia (teniendo en cuenta que, en el marco **del RGPD** la aplicación no se puede utilizar en un entorno escolar, solo son posibles usos a título individual, bajo el régimen del contrato privado):

- por el momento, las respuestas proporcionadas **no se basan en datos en tiempo real** a través de una conexión web (hay que comprobarlo, pero se trata de datos procedentes de la web, anteriores a 2022);
- **resultan** sesgadas debido a la naturaleza de los datos tratados, aunque se apliquen filtros para limitar los sesgos más evidentes;
- los **errores factuales** y posibles “alucinaciones” requieren casi siempre una verificación del contenido por parte del usuario, lo que puede constituir tanto un punto débil como uno fuerte en un marco pedagógico para desarrollar el espíritu crítico y establecer una distancia con el contenido generado;
- el riesgo de producir y difundir infox o información truncada, errónea o no verificable;
- una tecnología de este tipo, debido incluso a su rendimiento y a la masa de los datos procesados (sin transparencia sobre la naturaleza de los mismos), plantea inevitablemente dudas sobre las referencias (¿podemos referirnos a las respuestas producidas cuando no son reproducibles y su fuente no se indica sistemáticamente?), sobre la autoría de los contenidos (¿a quién pertenecen?, ¿se puede atribuir la autoría de una obra a una IA?) (Stokel-Walker, 2023) o sobre los **derechos de autor** (¿las respuestas generadas a partir de estos datos masivos respetan los derechos de autor cuando es aplicable? por ejemplo, sobre una obra científica o literaria que no es de licencia libre);
- **la ausencia de fuentes** - o el posible carácter erróneo de las citadas - constituye una verdadera dificultad para su uso en un marco académico (educación, investigación y documentación), a menos que el usuario lo solicite expresamente, y con la condición de que este verifique una a una las referencias citadas.

En respuesta a este problema, otras aplicaciones complementarias o de la competencia, como Bing Chat (el motor de búsqueda de Microsoft que ahora integra GPT-4) o Perplexity (<https://www.perplexity.ai/>), ofrecen fuentes (pero no garantizan su pertinencia o integridad).

También cabe destacar el **impacto medioambiental** de este tipo de modelo, no específico de ChatGPT:

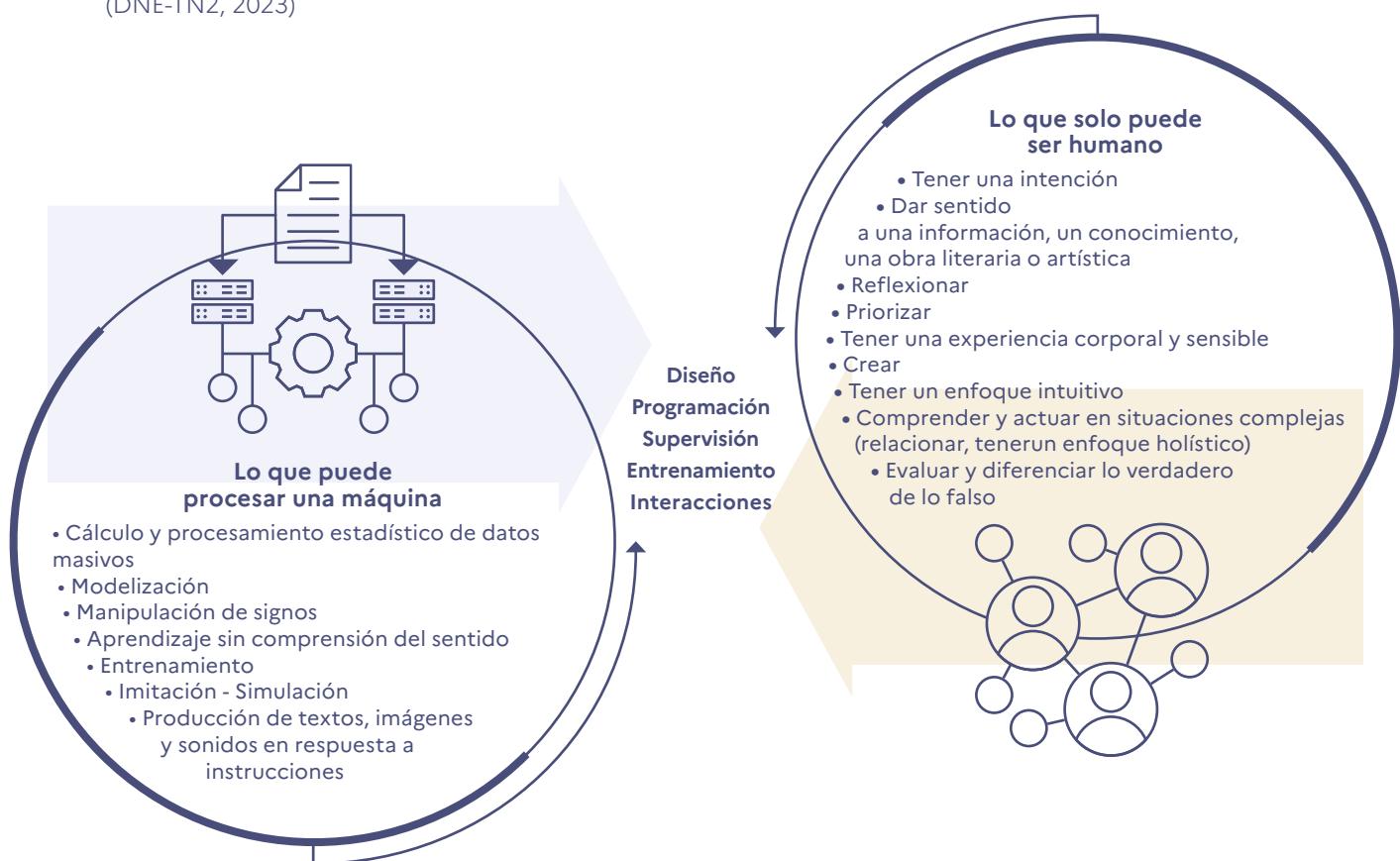
“Training OpenAI's GPT-3 has been estimated to have required as much energy as driving a car to the moon and back, thus generating an equivalent of 85,000 kg of CO₂ emissions (Quach, 2020). Energy consumption is now widely understood to be a major challenge for data-driven AI (Strubell et al., 2019).” (Holmes & Tuomi, 2022) traduction par (Bocquet, 2023)

En resumen, las principales vías a explorar son principalmente la asistencia (ver ejemplos en las páginas 33-35), la inspiración, el complemento o la profundización. De hecho, este salto tecnológico invita sobre todo a los investigadores y a los pedagogos a innovar, colaborar, informarse y **formarse, a tomar distancia y a ser críticos** para acompañar a los alumnos en su aprendizaje y formación.

Como sugiere el siguiente esquema, entre los retos que hay que afrontar se encuentra la distinción que hay que establecer con los alumnos entre lo que es exclusivamente humano: dar sentido a la información, el conocimiento, la obra literaria o artística, y lo que realmente se puede confiar a la IA: el procesamiento y cálculo de datos masivos, más allá de las capacidades humanas, con fines de predicción, probabilidad u optimización, pero sin intención ni comprensión de lo que se produce.

Qué puede o no ser IA

(DNE-TN2, 2023)



Referencias

- Allouche, E. (2023a, janvier 5). Sens et finalités du numérique en éducation – Hors-série : Tests et simulations d’ “entretien” avec ChatGPT (Open AI) [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/7635>
- Allouche, E. (2023b, octobre 2). IA génératives et ingénierie pédagogique : Le *prompting*, pistes de travail et applications [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/9934>
- Association for the Advancement of Artificial Intelligence, & Computer Science Teachers Association. (2020). AI4K12. AI4K12. <https://ai4k12.org/>
- Association Française pour l’Intelligence Artificielle (AFIA). (2022, janvier 6). Journée Enseignement et Formation en IA : “IA pour l’enseignement”. AfIA. <https://afia.asso.fr/efia-2022/>
- Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2023). *Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers* (arXiv:2311.10242). arXiv. <https://doi.org/10.48550/arXiv.2311.10242>
- Boa Sorte, P., Farias, M. A. de F., dos Santos, A. E., Santos, J. do C. A., & Dias, J. S. dos S. R. (2021). Artificial intelligence in academic writing: What is in store with the GPT-3 algorithm? *Revista EntreLínguas*, 7(00). <https://periodicos.fclar.unesp.br/entrelinguas/article/download/15352/11410/51225>
- Bocquet, F. (2023, janvier 23). État de l’art et de la pratique de l’intelligence artificielle dans l’éducation (Holmes & Tuomi, 2022) [Traduction] [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/8350>
- Boisson, J., & DNE-TN2. (2023, janvier 12). Intelligence artificielle, traitement automatique des langues et agents conversationnels [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/8235>
- Cardon, D., Cointet, J.-P., & Mazieres, A. (2018). La revanche des neurones : L’invention des machines inductives et la controverse de l’intelligence artificielle. *Réseaux*, 5(211). <https://doi.org/10.3917/res.211.0173>
- Cisel, M., & Baron, G.-L. (2019). Vers des intelligences artificielles pour l’enseignement de la démarche d’investigation. *Educational Journal of the University of Patras UNESCO Chair*, 0, Article 0. <https://doi.org/10.26220/une.2990>
- CNPEN. (2023, juin 30). Avis 7 du CNPEN Systèmes d’intelligence artificielle générative : *Enjeux d’éthique*. Comité Consultatif National d’Ethique. <https://www.ccne-ethique.fr/fr/publications/avis-7-du-cnpen-systemes-dintelligence-artificielle-generative-enjeux-dethique>
- CNRS. (2022, juillet 12). *Livraison du plus grand modèle de langue multilingue “open science” jamais entraîné*. CNRS. <https://www.cnrs.fr/fr/livraison-du-plus-grand-modele-de-langue-multilingue-open-science-jamais-entraigne>

Collin, S., & Marceau, E. (2021). L'intelligence artificielle en éducation : Enjeux de justice. *Formation et profession*, 29(2), 1. <https://doi.org/10.18162/fp.2021.a230>

Conseil de l'Europe. (2023, décembre 9). Législation sur l'intelligence artificielle : Le Conseil et le Parlement parviennent à un accord sur les premières règles au monde en matière d'IA. Conseil européen - Conseil de l'Union européenne. <https://www.consilium.europa.eu/fr/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>

Cour des comptes. (2023). *La stratégie nationale de recherche en intelligence artificielle*. Cour des comptes. <https://www.ccomptes.fr/fr/publications/la-strategie-nationale-de-recherche-en-intelligence-artificielle>

CSEN, & DNE-TN2. (2022, novembre 7). Synthèse et recommandations du groupe de travail "Intelligence artificielle et numérique pour l'éducation" du CSEN [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/6707>

De la Higuera, C., & Bocquet, F. (2020, juillet 9). L'éducation, la formation des enseignants et l'apprentissage de l'intelligence artificielle : Un aperçu des questions clés [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/1973>

Devillers, L. (2023, février 20). *ChatGPT et "intelligences" artificielles : Comment déceler le vrai du faux*. The Conversation. <http://theconversation.com/chatgpt-et-intelligences-artificielles-comment-deceler-le-vrai-du-faux-200181>

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>

Directorate-General for Education, Youth, Sport and Culture, Y., & European Commission. (2022). *Lignes directrices pour les enseignants et les éducateurs en matière de lutte contre la désinformation et de promotion de l'habileté numérique par l'éducation et la formation*. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/a224c235-4843-11ed-92ed-01aa75ed71a1/language-fr>

DNE-TN2. (2021, juin 30). Le deep learning, un tournant épistémologique pour les SHS ?(Huma-Num LAB, 2020) [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/3203>

DNE-TN2. (2022, janvier 20). Les enjeux éducatifs à l'ère de l'Intelligence Artificielle : GTnum #Scol_IA – Maison de l'Intelligence artificielle [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/3898>

DNE-TN2. (2023a, mars 9). Intelligence artificielle et éducation ouverte : Retour sur la journée du GTnum #IA_EO (26/01/23) [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/8605>

DNE-TN2. (2023b, juillet 10). Enseigner et apprendre à l'ère de l'intelligence artificielle : Portfolio du GTnum #Scol_IA [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/9593>

DNE-TN2. (2023c, septembre 6). Intelligence artificielle et éducation ouverte : Portfolio du GTnum #IA_EO [Billet]. *Éducation, numérique et recherche*. <https://edunumrech.hypotheses.org/9781>

European Commission. Directorate General for Education, Youth, Sport and Culture. (2022). *Lignes directrices éthiques sur l'utilisation de l'intelligence artificielle (IA) et des données dans l'enseignement et l'apprentissage à l'intention des éducateurs*. Publications Office. <https://data.europa.eu/doi/10.2766/153756>

Ganascia, J.-G. (2022, juin 10). *Intelligence artificielle et épistémologie. Allers-retours indispensables*. JECIS. <https://hal.science/hal-03760357>

Gelin, L. (2022). *Reconnaissance automatique de la parole d'enfants apprenants-tes lecteurs-trices en salle de classe : Modelisation acoustique de phones* [Thèse, Toulouse 3]. <https://hal.science/tel-03715653v1>

Gouvernement - Direction générale des Entreprises. (2023). *Éduquer à l'Intelligence artificielle. Stratégie nationale pour l'intelligence artificielle*. <https://www.entreprises.gouv.fr/fr/numerique/enjeux/developper-ecosysteme-de-talents>

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). *ChatGPT is not all you need. A State of the Art Review of large Generative AI models*. <https://doi.org/10.48550/ARXIV.2301.04655>

Grand lexique français de l'Intelligence artificielle. (2021). DataFranca Wiki. <https://datafranca.org/wiki/Accueil>

Herft, A. (2023). *Guide de l'enseignant—L'usage de ChatGPT "ce qui marche le mieux"* (A. Gagné, Trad.). <https://labua.univ-angers.fr/wp-content/uploads/2023/01/ChatGPT-Guide-de-lenseignant-FR.pdf>

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education—A critical view through the lens of human rights, democracy and the rule of law*. Council of Europe. <https://rm.coe.int/artificial-intelligence-and-education-a-critical-view-through-the-lens/1680a886bd>

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, n/a(n/a). <https://doi.org/10.1111/ejed.12533>

Hsiao, S. (2023, décembre 6). *Bard gets its biggest upgrade yet with Gemini*. Google. <https://blog.google/products/bard/google-bard-try-gemini-ai/>

Inria, & AI4T. (2023). *Intelligence artificielle pour et par les enseignants*. FUN MOOC. <http://www.fun-mooc.fr/fr/cours/intelligence-artificielle-pour-et-par-les-enseignants-ai4t/>

Inria, & Class'Code. (2020). *L'Intelligence Artificielle... avec intelligence !* (MOOC). FUN-MOOC. <https://www.fun-mooc.fr/fr/cours/intelligence-artificielle-avec-intelligence/>

Inria Flowers (Réalisateur). (2023a, mars 8). *Comment fonctionne ChatGPT ? Un tour d'horizon en moins de 5mn*. <https://www.youtube.com/watch?v=K8gOvC8gvB4>

Inria Flowers (Réalisateur). (2023b, mars 13). *Le prompting, ou l'art de se faire comprendre par ChatGPT : Explication en 5mn*. https://www.youtube.com/watch?v=8IQ9i_QoA3A

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*. <https://doi.org/10.35542/osf.io/5er8f>

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Gasevic, D., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Tsai, Y.-S. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>

Langlais, P.-C. (2023, février 7). ChatGPT : Comment ça marche ? [Billet]. Sciences communes. <https://scoms.hypotheses.org/1059>

Le Blanc, B. (2014). Alan Turing : Les machines à calculer et l'intelligence. *Hermès, La Revue*, 68(1), 123-126. <https://doi.org/10.3917/herm.068.0123>

Le Cun, Y., Brizard, C., & Brizard, C. (2019). *Quand la machine apprend : La révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob.

Lepage Alexandre (Réalisateur). (2022, mars 3). *L'intelligence artificielle en éducation*. <https://www.youtube.com/watch?v=ZAtu0xIlcls>

Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), Article 4. <https://doi.org/10.3390/educsci13040410>

Lorenz, P., Perset, K., & Berryhill, J. (2023). *Initial policy considerations for generative artificial intelligence*. OCDE. <https://doi.org/10.1787/fae2d1e6-en>

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO Bibliothèque Numérique. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>

OCDE. (2023, mars 28). *Is Education Losing the Race with Technology?: AI's Progress in Maths and Reading*. OCDE. <https://www.oecd.org/education/is-education-losing-the-race-with-technology-73105f99-en.htm>

OpenAI. (2023). *GPT-4 Technical Report*. OpenAI. <https://cdn.openai.com/papers/gpt-4.pdf>

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education : The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>

Parlement européen. (2023, décembre 9). *Artificial Intelligence Act: Deal on comprehensive rules for trustworthy AI*. News European Parliament. <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

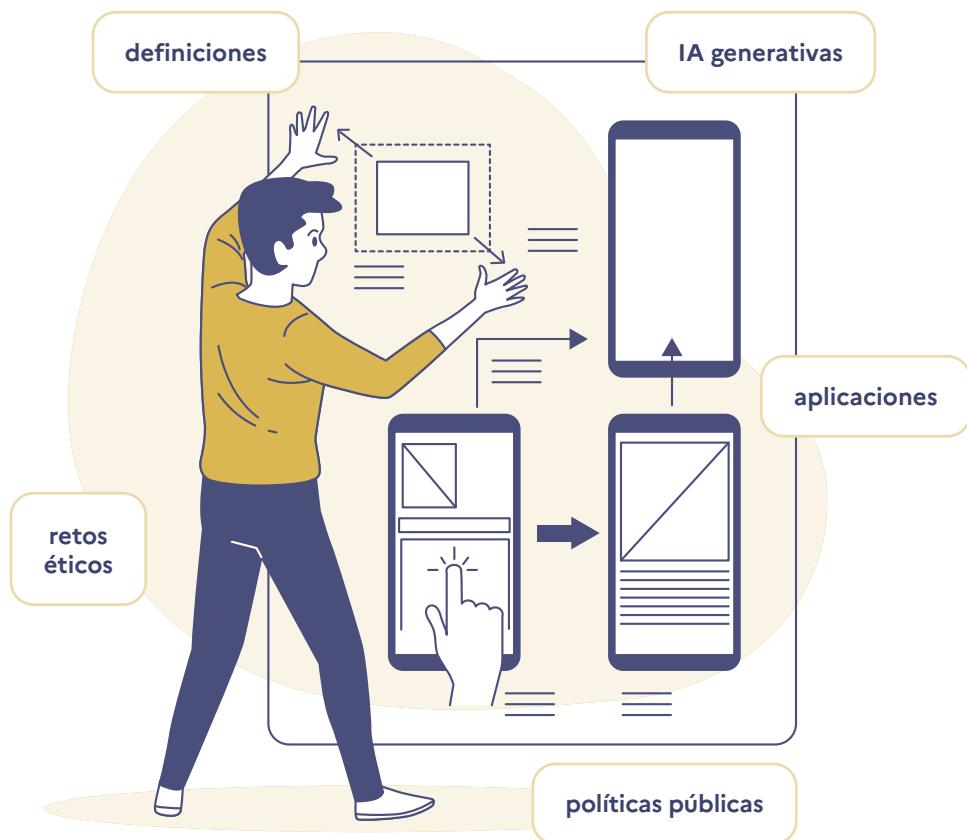
Pôle d'expertise de la régulation numérique. (2023). ChatGPT ou la percée des modèles d'IA conversationnels. *Gouvernement français*, 6. https://www.paren.gouv.fr/rapports/2023-04-06_Eclairage%20sur_CHATGPT_FR.pdf

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22. <https://doi.org/10.1186/s41039-017-0062-8>

Romero, M., Aloui, H., Heiser, L., Galindo, L., & Lepage, A. (2021). *Un bref parcours sur les ressources, pratiques et acteurs en IA et éducation*. <https://hal.archives-ouvertes.fr/hal-03190014>

- Romero, M., Heiser, L., Lepage, A., Gagnebien, A., Bonjour, A., Lagarrigue, A., Palaude, A., Boulard, C., Gagneur, C.-A., Mercier, C., Caucheteux, C., Guidoni-Stoltz, D., Tressols, F., Alexandre, F., Céci, J.-F., Metral, J.-F., Camponovo, J., Henry, J., Fouché, L., ... Borgne, Y.-A. L. (2023). *Enseigner et apprendre à l'ère de l'intelligence artificielle : Livre blanc du GTnum #Scol_IA: Vol. Livre blanc*. <https://hal.science/hal-04013223>
- Singh, R., & Gill, S. S. (2023). Edge AI : A survey. *Internet of Things and Cyber-Physical Systems*, 3, 71-92. <https://doi.org/10.1016/j.iotcps.2023.02.004>
- Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*. <https://doi.org/10.1038/d41586-023-00107-z>
- Susnjak, T. (2022). *ChatGPT: The End of Online Exam Integrity?* (arXiv:2212.09292). arXiv. <https://doi.org/10.48550/arXiv.2212.09292>
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA : Open and Efficient Foundation Language Models (arXiv:2302.13971). arXiv. <https://doi.org/10.48550/arXiv.2302.13971>
- UNESCO. (2019). *Consensus de Beijing sur l'intelligence artificielle et l'éducation*. UNESCO Bibliothèque Numérique. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
- UNESCO. (2021). *IA et éducation : Guide pour les décideurs politiques*. UNESCO Bibliothèque Numérique. <https://unesdoc.unesco.org/ark:/48223/pf0000380006>
- UNESCO. (2022). *K-12 AI curricula: A mapping of government-endorsed AI curricula*. UNESCO Bibliothèque Numérique. <https://unesdoc.unesco.org/ark:/48223/pf0000380602>
- UNICEF. (2021, novembre). *Orientations stratégiques sur l'IA destinée aux enfants 2.0*. UNICEF. <https://www.unicef.org/globalinsight/fr/rapports/orientations-strat%C3%A9giques-sur-lia-destin%C3%A9e-aux-enfants>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. <https://doi.org/10.48550/ARXIV.1706.03762>

Referencia del artículo: DNE-TN2 (2024). Intelligence artificielle et éducation : Apports de la recherche et enjeux pour les politiques publiques (édition 2024) [Billet]. Éducation, numérique et recherche.
<https://edunumrech.hypotheses.org/10764>



Carnet Hypothèses

Educación, digital e investigación



<https://edunumrech.hypotheses.org/>

Contact



Direction du numérique pour l'éducation
(Ministère de l'Éducation nationale et de la Jeunesse)
Bureau du soutien à l'innovation numérique
et à la recherche appliquée (DNE TN2)

@Edu_Num

 Licence Ouverte 2.0
 Open Licence 2.0