

Trabajo Final de Grado en Humanidades. Estudios Interculturales

Los riesgos de la inteligencia artificial en la educación: el caso del *ChatGPT*

Autor/a: Naiara Álvarez Bernat

Tutor/a o supervisor/a: Daniel Pallarés-Domínguez

Fecha de lectura: 16 de junio de 2023



Resumen

El presente trabajo tiene como objetivo realizar un estudio crítico sobre la evolución de la inteligencia artificial dentro del contexto educativo, especialmente en la enseñanza secundaria. Con este objetivo, se comprobará si la hipótesis de si la inteligencia artificial presenta más riesgos que ventajas en el ámbito educativo, se cumple. Para ello, en primer lugar, se introducirá el funcionamiento la IA y de *ChatGPT* dentro de un marco conceptual y temporal. En segundo lugar, y a raíz del análisis de los resultados de un cuestionario realizado a docentes de educación secundaria, se enumerarán los diferentes riesgos que se producen, como la pérdida del pensamiento crítico, la dependencia tecnológica o la imposibilidad de distinguir el trabajo personal del plagio, para poder así analizarlos críticamente desde el punto de vista filosófico y educativo.

Palabras clave: Inteligencia Artificial (IA), *ChatGPT*, educación, reflexión crítica, metaverso

Abstract

The present work aims to conduct a critical study on the evolution of artificial intelligence within the educational context, especially in higher education. With this objective, the hypothesis of whether artificial intelligence presents more risks than advantages in the educational field will be verified. For this, first of all, the operation of AI and ChatGPT will be introduced within a conceptual and temporal framework. Secondly, and as a result of the analysis of the results of a questionnaire made to high school teachers, the different risks that occur will be listed, such as the loss of critical thinking, technological dependence or the impossibility of distinguishing personal work from plagiarism, in order to analyze them critically from the philosophical and educational point of view.

Key words: Artificial Intelligence (AI), *ChatGPT*, education, critical reflection, metaverse

Índice

Introducción	5
1. Marco histórico de la IA	9
2. Marco conceptual de la IA.....	13
3. La IA en el contexto educativo actual: el caso de <i>ChatGPT</i>.....	17
3.1 La creación de <i>ChatGPT</i>	19
3.2 La introducción de <i>ChatGPT</i> en el aula.....	23
4. Análisis crítico de la introducción de la IA en la educación.....	29
Conclusiones.....	41
Bibliografía.....	45
Anexos.....	51
Anexo 1. Resultados de la encuesta.....	51
Anexo 2. Respuestas de <i>ChatGPT</i>	53

Introducción

En la actualidad, la tecnología ha supuesto un gran impacto en nuestra sociedad. En las últimas décadas y, sobre todo, en los últimos años ha transformado la manera en la que nos comunicamos (redes sociales, videollamadas...), obtenemos información (páginas web), nos entretenemos (plataformas en *streaming*) e incluso la forma en la que compramos (comercio *online*). No obstante, desde el surgimiento de modelos de lenguaje como *ChatGPT* y nuevas formas de inteligencia artificial, estos avances están adentrándose en el ámbito educativo con el objetivo de transformarlo por completo. De manera que el foco de preocupación ahora recae sobre esta cuestión, ya que los cambios que se producen en la enseñanza requieren de una mayor reflexión y control; no se puede dejar en manos de la IA la evolución de la educación, sin saber siquiera a dónde se dirige.

Tal es la importancia de este asunto, que desde el Departamento de Filosofía y Sociología de la UJI, se está llevando a cabo en estos momentos un proyecto de investigación en el grupo de Ética Práctica y Democracia del área de filosofía moral que recibe el título de “Ética aplicada y confiabilidad para una Inteligencia Artificial”. Por lo que coincide con el hecho de que estamos frente a un tema preocupante que requiere de reflexión e investigación filosófica y educativa al respecto.

De este modo, la hipótesis sobre la que se sustenta la elaboración de este trabajo es que la proliferación de inteligencia artificial sin una revisión ética y crítica sobre su sentido presenta más riesgos que ventajas en la educación. Así pues, a partir de este supuesto se establecen los siguientes objetivos a abordar.

El primer objetivo frente al que nos encontramos es el de dotar de los conocimientos necesarios al lector acerca de la IA, el metaverso y *ChatGPT*, entre otros, para que pueda situarse en un marco conceptual suficiente y, con ello, poder realizar un seguimiento adecuado del estudio. El segundo objetivo busca conocer, a través de un cuestionario, si los docentes son conscientes del surgimiento de estas nuevas plataformas y qué opinión tienen en caso de conocerlas. Por último, el tercer objetivo pretende a través del análisis de los riesgos de la IA en la educación, conseguir que el lector reflexione acerca de esta

situación y entienda que estamos frente a una problemática difícil de frenar si no se tiene el respaldo de una regulación ética.

El trabajo se encuentra dividido en cuatro capítulos. El primero de ellos, nos describe el origen y la definición del término de inteligencia artificial según diversos autores para, posteriormente, realizar un recorrido histórico de la IA. El segundo, nos aporta un enfoque más conceptual, puesto que comienza con la definición de cinco conceptos clave en torno a los que se sustenta la innovación tecnológica actual (*big data, machine learning, deep learning*, IA débil y fuerte). Además, se hace un análisis más profundo del término metaverso debido a que es el más complejo y el más relacionado con el ámbito educativo.

Por lo que respecta al tercer capítulo, este se encuentra dividido en dos grandes puntos. Por un lado, se tratan algunos aspectos relacionados con la creación del modelo de lenguaje *ChatGPT* tales como qué empresa llevó a cabo su desarrollo, quiénes son sus fundadores o cuál es su forma de proceder. Por otro lado, se lleva a cabo el análisis de los datos obtenidos a través de la encuesta enviada a diferentes docentes, en la que se encuentran cuestiones como si conocen esta herramienta, si consideran que pueden producirse riesgos o cambios en la educación a causa de la IA o, por el contrario, qué beneficios puede suscitar. Este último apartado se toma como punto de partida para analizar de forma crítica en el cuarto capítulo, los múltiples riesgos que puede ocasionar la IA en el panorama educativo; sobre todo, la cuestión relacionada con el plagio académico.

En lo que se refiere a las fuentes bibliográficas, los idiomas en los que se ha llevado a cabo la búsqueda de información han sido en español e inglés, aunque tiene una mayor presencia el español. Sin embargo, aunque suponen una minoría los de habla inglesa, estos son igual de significativos, como es el caso del artículo redactado por Rudolph, Jürgen, Samson Tan y Shannon Tan (*ChatGPT: bullshit spewer or to end of traditional assessments In higher education?*), que tiene relevancia principalmente en los últimos dos capítulos. En cuanto al resto de la bibliografía, destacan principalmente los artículos de investigación publicados en revistas científicas indexadas. En este sentido se pueden destacar a autores como Raúl Darío Moreno Padilla para la definición de IA y los conceptos que la rodean, Israel Vázquez Márquez en cuanto al metaverso, Miguel

Antonio Morales Chan en lo que respecta a la explicación de *ChatGPT* o Ignacio Jara y Juan Manuel Ochoa, Patrici Calvo y Dariel Díaz Arce en lo relativo a los riesgos.

En la misma línea del párrafo anterior, nos encontramos con que todas las fuentes mencionadas previamente se engloban dentro de una metodología argumentativa, es decir la investigación de este trabajo se sustenta gracias a los diferentes documentos bibliográficos. Sin embargo, esta no es la única disciplina empleada, puesto que con la incorporación del elemento de la encuesta se incluye una perspectiva vinculada a la sociología, y por ende, basada en el tratamiento de los datos obtenidos a través del cuestionario. Sobre todo, resulta importante puesto que nos permite saber si se confirma o no, nuestra hipótesis de que la IA supone un gran riesgo en el contexto educativo.

Por último, la motivación que se encuentra detrás de la elección del presente trabajo es una principalmente. Como futura docente quería abordar alguna problemática actual en este ámbito, para poder tener una concepción de cuáles son las dificultades que deberán hacerse frente a lo largo de las próximas décadas, y a su vez, saber si existen algunas directrices sobre cómo proceder. A pesar de que la cuestión de la aplicación de nuevas tecnologías en la docencia era tenida en cuenta, en un primer momento, el objeto de estudio iba más encaminado hacia la falta de interés y motivación que se encuentra cada día más presente en las aulas. No obstante, el auge y la presencia internacional que ha tenido la inteligencia artificial en los últimos meses, generando numerosos debates acerca de sus variados usos, supuso el cambio inmediato a este tema. Además, el hecho de que surjan constantemente nuevos elementos y herramientas en esta disciplina hace que se precise de una participación activa de todos los miembros de la sociedad civil, por lo que cuantos más estudios al respecto se elaboren, más preparados estaremos para poder afrontar las situaciones que se presenten.

1. Marco histórico de la IA

Esta primera contextualización del tema se centrará en el origen de la inteligencia artificial, realizando un pequeño recorrido histórico sobre las características que se suelen atribuir a las máquinas inteligentes. De esta manera, se conseguirá situar al lector dentro del cada vez más amplio y complejo espacio que ocupa la IA hoy en día.

Según diversas personalidades como Stephen Hawking, Elon Musk y Nick Bostrom, uno de los desafíos a los que nos enfrentaremos en la próxima década es el de la inteligencia artificial, de manera que vamos a conocer en primer lugar qué es, cómo surge, qué aplicaciones puede tener en el ámbito educativo, etc.

Aunque pueda parecer que la inteligencia artificial es algo novedoso y propio de nuestro siglo, lo cierto es que hemos de remontarnos hasta los años cincuenta del pasado siglo. En 1950, Alan Turing, uno de los científicos más reconocidos, publicó un artículo llamado *Computing machinery and intelligence*, en el que se planteaba por primera vez la pregunta de si las máquinas podían pensar o no. Añadido a esto, creó también una prueba que recibe su nombre, Test de Turing, en la que precisamente se cuestiona si quien responde la prueba es una máquina o una persona. La actividad consistía en ubicar en dos habitaciones diferentes a un entrevistado humano y a una máquina. En el momento en el que la máquina respondiera de manera compleja a las preguntas haciendo ardua la tarea del juez para diferenciarlos, se consideraría a la máquina como inteligente (Botta y Mora 2016, 269-270).

A pesar de que Alan Turing fue uno de los pioneros en este campo, no es hasta 1956 que se emplea del término de inteligencia artificial (IA) por primera vez. En ese año, John McCarthy, Marvin Minsky, Claude Shannon y Nathaniel Rochester organizaron una conferencia en Darmouth en la que se definió la IA como: «hacer que una máquina se comporte como lo haría un ser humano, de tal manera que se la podría llamar inteligente» (Banda 2014, 6).

Sin embargo, con el paso del tiempo, se ha podido ir concretando la definición de inteligencia artificial. Por su parte, Richard Bellman la entiende como: «La

automatización de actividades que vinculamos con procesos de pensamiento humano, actividades tales como toma de decisiones, resolución de problemas, aprendizaje...», y Stuart Russell y Peter Norvig interpretan que: «La Inteligencia Artificial es la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano» (Moreno Padilla 2019, 262).

Además, en 1987 Martin Fischles y Oscar Firschein le atribuyen a la máquina inteligente las siguientes características que completan aún más el término, y a su vez, establecen unas cualidades que ha de cumplir (Moreno Padilla 2019, 261):

1. Tiene actitudes mentales tales como creencias e intenciones.
2. Tiene la capacidad de obtener conocimiento, es decir, aprender.
3. Puede resolver problemas, incluso descomponiendo problemas complejos en otros más simples.
4. Capaz de realizar operaciones más complejas.
5. Entiende. Posee la capacidad de dar sentido, si es posible, a ideas ambiguas o contradictorias.
6. Planifica, predice consecuencias, evalúa alternativas (como en los juegos de ajedrez)
7. Conoce los límites de sus propias habilidades y conocimientos.
8. Puede distinguir a pesar de la similitud de las situaciones.
9. Puede ser original, creando incluso nuevos conceptos o ideas, y hasta utilizando analogías.
10. Puede generalizar.
11. Puede percibir y modelar el mundo exterior.
12. Puede entender y utilizar el lenguaje y sus símbolos.

No obstante, este tipo de definiciones pueden resultar un tanto optimistas e incluso ingenuas desde el punto de vista filosófico, dado que el ser humano tiene unas características muy complejas que ni siquiera se podrían definir en términos algorítmicos. Especialmente, las condiciones que tiene el ser humano en condición de “persona”, como por ejemplo: el sentido de la moralidad, la capacidad de amar, el lenguaje simbólico, la facultad axiológica y la experiencia vital, entre otras (Cortina, 1992).

Continuando con las definiciones anteriores, no solo la pretensión de “actuar como un ser humano” es problemática, sino también el propio término “inteligente”. ¿Es acaso ser inteligente actuar como lo hace el ser humano? Desde el punto de vista ético, esta afirmación sería muy discutible. ¿Realmente se quiere crear una inteligencia como la del ser humano? ¿O más bien una inteligencia con solamente algunas características del ser humano? Por otro lado, las diferentes ciencias que trabajan interdisciplinariamente en la

IA, ¿tienen realmente el mismo concepto de inteligencia? Por ejemplo, ¿es lo mismo la inteligencia desde el punto de vista psicológico, que desde el punto de vista ético? Por ejemplo, en el caso de la creación de un coche autónomo, ¿cuál de las dos tendría prevalencia? A este respecto, investigadores del MIT Media Lab del Instituto de Massachusetts tuvieron que crear un dilema moral que ponía en distintas situaciones problemáticas a un hipotético conductor para comprobar cómo actuaría si tuviera que decidir entre atropellar a diferentes sujetos en caso de fallo mecánico. Este estudio se desarrolló en 233 países para reunir datos de cómo diseñar “inteligentemente” a la vez que “moralmente” un coche autónomo (Awad et al. 2018). Las respuestas de la mayoría de las personas tuvieron un sesgo utilitarista, y muchas de ellas implicaban un conflicto entre lo que inteligentemente una máquina haría con lo que moralmente debería de hacer.

Teniendo en cuenta las preguntas anteriores, es preciso contextualizar las diferentes etapas por las que ha ido evolucionando la IA.

La primera etapa es conocida como la época de los profetas, en la que los investigadores no pusieron límites a las preguntas que se planteaban, por muy utópicas que parecieran. La segunda etapa, de los años sombríos, tiene lugar alrededor de la década de los sesenta del pasado siglo, recibe este nombre puesto que la IA en estos años se estanca al ver que las fabulaciones que se hacían no podían llevarse a cabo; y debido a que en ese momento un niño consigue vencer a un ordenador al ajedrez, lo que deja en evidencia las capacidades de estas máquinas. Además de esto, hubo tres hallazgos que hicieron que el desinterés por la IA fuera en aumento todavía más y por ende, que frenaran las investigaciones al respecto. Uno de los problemas surge a raíz del teorema de incompletitud de Kurt Gödel (lógico, matemático y filósofo), en el que «demostró que ni siquiera los sistemas matemáticos pueden ser completos y autosuficientes» (Calvo 2020, 232). Los otros dos problemas proceden de la filosofía, puesto que se criticó por un lado, la falta de un cuerpo que acompañe a la inteligencia artificial y las cuestiones éticas y morales que surgirían si una IA fuerte llegara a lograrse (Calvo 2020, 232-233).

La tercera, llamada IA semántica, se produce unos diez años después de la anterior. En esta fase, es cuando la inteligencia artificial empieza a adentrarse en la psicología de la memoria y los procesos de comprensión para poder posteriormente, aplicarlos en los procesadores. A esta, le sigue la etapa del neoconexionismo y el aprendizaje automático,

en la que se crearon los algoritmos de aprendizaje automático, o más conocido como *machine learning*. Posteriormente, nos situamos ya a finales de los años 90 con el periodo de la IA a las interfaces hombre-máquina, en la que se empieza a tener en cuenta las emociones como factor que deben poseer los dispositivos para asemejarse todavía más a los humanos (García Peña, Mora Marcillo, y Ávila Ramírez 2020, 654-655).

Finalmente, la etapa que da origen a la situación actual comenzó a formarse en 2010 y se la conoce como el resurgimiento de la IA. Esta reaparición de la inteligencia artificial en la actualidad se debe principalmente a dos sucesos. Por un lado, gracias a las innovaciones en *hardware* especializado y, por otro lado, debido a la capacidad de almacenar una gran cantidad de datos (García Peña, Mora Marcillo, y Ávila Ramírez 2020, 655; Abeliuk y Gutiérrez 2021, 17).

Por tanto, en poco más de siete décadas, se ha podido comprobar cómo el interés por la IA ha crecido dependiendo más del cumplimiento de expectativas que de otros factores. Pero ¿expectativas de qué? ¿De creer que lo que se puede soñar se puede hacer en realidad? O filosóficamente hablando, ¿expectativas de que lo que se puede hacer tecnológicamente, se debe hacer desde el punto de vista ético? Si esto fuera así, ¿no estaríamos ante una clara falacia naturalista, tratando de pasar (una vez más en la historia de la humanidad) del “ser” al “deber ser”? ¿Todo lo tecnológicamente posible es humanamente deseable? ¿Es lo tecnológicamente posible un imperativo ético? ¿Se debe desarrollar la IA porque simplemente se puede hacer?

2. Marco conceptual de la IA

La finalidad de este punto es la de conocer algunos de los conceptos sobre los que se sustenta la inteligencia artificial. Es de vital importancia entender estos términos, para, en primer lugar, familiarizarnos con el vocabulario que envuelve a la IA; y, en segundo lugar, comprender de qué estamos hablando cuando estos sean mencionados. La elección de estos conceptos en concreto se debe a que el *big data* se encuentra directamente relacionado con la IA puesto que las máquinas y plataformas se configuran a través de datos, pero en volúmenes extensos, por lo que sin los macrodatos no habría inteligencia artificial. En lo que respecta a la IA débil y fuerte, esta última es en base a la que todos los desarrolladores trabajan, por lo que es importante conocer cómo quieren que sea el futuro tecnológico. Por último, el *machine learning* y el *deep learning* se corresponden con los modelos de inteligencia (débil y fuerte) por lo que se encuentran directamente relacionados.

Macrodatos (*Big Data*): este concepto aparece por primera vez en 1989 gracias a Erik Larson, quien le da el significado que conocemos hoy en día. A pesar de que existen diferentes definiciones, son tres características principales las que definen el término, que son las conocidas como las 3V's del *Big Data* (volumen, variabilidad y velocidad). De esta manera, entendemos por macrodatos el «conjunto de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis» (Moreno Padilla 2019, 263).

Aprendizaje automático o de máquina (*Machine Learning*): es un campo de la ciencia que permite a través de diferentes técnicas que las máquinas aprendan y solventen problemas sin haber sido expresamente programadas para ello, es decir, de manera autónoma (Mahesh 2020, 381; Moreno Padilla 2019, 263).

Aprendizaje profundo (*Deep Learning*): el *deep learning* se encuentra dentro del *machine learning*. Es una subcategoría que utiliza los algoritmos para imitar el funcionamiento del cerebro humano y así poder reconocer por sí misma diferentes objetos, cuerpos y rostros gracias al almacenamiento de cantidades inmensas de datos que

le permiten identificar de qué se trata cada imagen (García Peña, Mora Marcillo y Ávila Ramírez 2020, 653; Jara y Ochoa 2020, 4; Fava Fernández 2017, 33).

IA débil: «es la ciencia e ingeniería que permiten diseñar y programar ordenadores de forma que realicen tareas que requieren inteligencia» (López de Mántaras Badia 2015, 97). De esta manera, este tipo de IA liberaría de ciertos trabajos a las personas y ahorraría mucho tiempo.

IA fuerte: «es la ciencia e ingeniería que permitirá replicar la inteligencia humana mediante máquinas» (López de Mántaras Badia 2015, 97). A diferencia de la IA débil, esta no hace tareas que son propias de los seres humanos, sino que se convierte en uno, es capaz de tomar decisiones por sí misma; no obstante, esta realidad no es factible según John Searle, el filósofo que acuñó y diferenció estos dos últimos términos (López de Mántaras y Meseguer González 2017, 10).

Considerando los conceptos anteriores como aquellos básicos en el estudio de la IA, es preciso a continuación hacer referencia a uno de los nombres más grandes, difíciles de definir y con mayores implicaciones en el terreno que nos ocupa, el educativo. Se trata del concepto de metaverso. Primero, se comentará su origen y definición para así continuar con los aspectos que lo caracterizan y los diferentes tipos de metaverso existentes.

En 1992 aparece por primera vez el término metaverso en la novela de Neal Stephenson conocida como *Snow Crash*. Vázquez Márquez recoge la definición que da el autor del concepto de metaverso de la siguiente forma:

Stephenson define el metaverso como “un universo generado informáticamente, que el ordenador dibuja sobre el visor y le lanza a través de los auriculares”, un lugar imaginario que “no existe realmente, sino que es un protocolo infográfico escrito en papel en algún sitio” y que está formado por “fragmentos de software, puestos a disposición del público a través de la red mundial de fibra óptica” (Vázquez Márquez 2011, 154).

En otras palabras, el metaverso hace referencia a un mundo digital en 3D, que es ficticio, y en el que las personas se convierten en usuarios que pueden interactuar entre ellos en diversos espacios y entornos; ya que no tienen ningún tipo de limitación en el espacio-

tiempo. En conclusión, hace referencia a lo que el propio nombre indica, más allá (meta) del universo (verso) (Ordoñez Valencia et al. 2022, 618; Checa García 2010, 149).

Seguidamente, son tres las características principales que forman al metaverso según Edward Castronova, profesor y especialista en mundos sintéticos o virtuales. En primer lugar, el metaverso posee interactividad, que es la capacidad que poseen los avatares (usuarios) de relacionarse entre ellos y con el entorno que les rodea. En segundo lugar, encontramos la corporeidad, es decir, cada individuo se ve representado en el mundo virtual por un avatar que escoge y personaliza a su gusto; no obstante, hay una serie de limitaciones en cuanto a leyes y recursos que se siguen del “mundo real”. En tercer lugar, la persistencia es el elemento que hace que este universo virtual esté en constante actividad y es que, aunque los usuarios no se encuentren inmersos, el programa continúa activo y se mantiene en el mismo punto que se ha dejado. Así pues, una vez que la persona quiera volver a conectarse lo encontrará todo tal y como lo dejó (Castronova 2001, 6).

A su vez, dentro del metaverso encontramos cuatro tipos diferentes: realidad aumentada, *lifelogging*, mundos espejo y mundo/realidad virtual. Además, cada una de ellas tiene una aportación diferente al ambiente educativo que veremos a continuación.

- **Mundos espejo:** en este tipo de metaverso, el mundo en el que nos adentramos supone una representación del mundo real, es el caso de aplicaciones como *Google Earth* o *Google Maps*, que se basan en proporcionar mapas virtuales de la geografía mundial o de videoconferencias como *Zoom* (Vázquez Márquez 2011, 156; Sánchez Mendiola 2022, 6).
- **Realidad aumentada:** al igual que en los mundos espejo, nos encontramos frente a la representación de la realidad, pero nos proporciona información detallada de ciertos aspectos (Vázquez Márquez 2011, 156). Un ejemplo sería la visita virtual de un museo o monumento; este tipo de visitas se dieron a conocer sobre todo a raíz del COVID-19, puesto que permitía seguir conociendo algunos lugares aunque fuera de manera online.
- **Los juegos y/o mundos virtuales:** este tipo es el que más se acerca al universo que describe Neal Stephenson en su definición de metaverso. Son mundos virtuales en los que el individuo se sumerge en un mundo ficticio e interactúa con

otros usuarios. Algunos ejemplos son: *World of Warcraft*, *Minecraft*, *Second Life*, *Habbo*, etc.

- ***Lifelogging***: se caracteriza por recoger, almacenar y compartir datos de la vida diaria. En este grupo es en el que aparecen las redes sociales como *Facebook*, *Instagram* o aplicaciones de deporte y salud como *Apple Watch* o *Samsung Health* (Vázquez Márquez 2011, 156; Sánchez Mendiola 2022, 6).

En cuanto a los usos que pueden tener en el ámbito educativo, se concluye que los mundos espejo nos ayudan en las comunicaciones a distancia a través de plataformas como *Google Meet*, que se empleó durante las clases *online* en la pandemia del COVID-19 o el aprendizaje de la geografía desde un punto de vista espacial. En cuanto a la realidad aumentada, permite conocer en primer plano algunos contenidos que, sin el apoyo de las visitas virtuales de manera tridimensional no sería posible. Continuando con el *lifelogging* permite un aprendizaje basado en la reflexión propia y el análisis de datos, y por último, los mundos virtuales nos permiten desarrollar el aprendizaje en otros espacios que no son posibles en la vida real.

3. La IA en el contexto educativo actual: el caso de *ChatGPT*

En el punto anterior se ha expuesto por un lado, las definiciones de algunos conceptos que resultan imprescindibles conocer como *Big Data* o IA fuerte, y por otro lado, se ha introducido de manera más extensa el término metaverso y sus tipos. En lo que respecta a este nuevo apartado, se va a tomar como punto de partida el metaverso, pero esta vez para conectarlo con el contexto educativo. Después de lo cual se introducirá el modelo de lenguaje conocido como *ChatGPT*. De manera que, la información se dividirá en la presentación y definición de *OpenAI*, *ChatGPT* y *prompt* (incluyendo sus tipos); en el análisis de los resultados obtenidos en el cuestionario que se ha llevado a cabo a diferentes docentes; y por último, en el análisis de riesgos que surgen debido a estos conceptos en el ámbito educativo.

El tipo de metaverso que más se ha desarrollado es el de los mundos virtuales, que a su vez, es el que más implicaciones ha tenido en la educación. Dentro de los juegos virtuales se encuentra el modelo de aprendizaje conocido como *e-learning 2.0*, el cual se caracteriza por basarse en el aprendizaje a través del juego; en que el foco de atención recaiga en el usuario y sea quien modifica la manera en la que se lleva a cabo la enseñanza; por estar fundamentado en el *learning by doing*, es decir, aprender haciendo; y conectado con otros usuarios (Vázquez Márquez 2011, 159).

Anteriormente, se mencionaba que este modelo es el más relacionado con el ámbito educativo, esto se debe a que en este ámbito surgió la plataforma conocida como *Second Life*, la cual tuvo un gran éxito por su accesibilidad principalmente. No obstante, que tuviera éxito no significa que consiguiera implementarse por ejemplo como una herramienta en las aulas; ya que precisamente vamos a comentar su funcionamiento, ventajas y desventajas para mostrar que este fue un ejemplo de IA que a pesar de que parecía que podría revolucionar la forma de enseñar, al final no acabó siendo así.

Antes de comenzar, *Second Life* es una aplicación informática que se instala en el ordenador y que permite al usuario «participar e interactuar en el mundo virtual, teniendo la capacidad de crear objetos y estructuras complejas, así como pudiendo realizar

interacciones con otros usuarios» (Checa García 2010, 150). Además, supone un modelo de aprendizaje basado en el *e-learning* mencionado anteriormente.

Centrándonos en las ventajas que se le atribuyen a *Second Life* nos encontramos con que permite la relación con diferentes personas sin tener que estar en el mismo espacio físico (se relacionan a través de sus avatares); se pueden añadir contenidos en formato audiovisual (imagen o vídeo) o de texto; aunque el usuario no se encuentre conectado, la aplicación sigue funcionando y mantiene los cambios tal y como se han dejado previamente a la desconexión; los elementos aparecen en formato 3D, por lo que se pueden analizar desde diferentes puntos de vista, que fuera de este metaverso no son posibles llevar a cabo; el usuario es el protagonista y se relaciona de forma colaborativa con el resto de alumnos; y por último, como modelo de *e-learning* aplica el aprendizaje basado en juegos (Vázquez Márquez 2011, 161-162).

Por el contrario, las desventajas que se asocian son las siguientes: se pueden producir fallos técnicos relacionados con la velocidad o por poseer un *hardware* anterior; en caso de que se produzca alguno de estos fallos, no contamos con ninguna otra solución para continuar con el aprendizaje; no se puede llevar a cabo un control de los estudiantes ya que si su avatar deja de aparecer no sabemos los motivos de manera inmediata, puesto que no nos encontramos físicamente en el mismo espacio; resulta un tanto complicado adaptarse a la plataforma de manera rápida, además de que algunos alumnos no cuentan con los conocimientos necesarios y pueden llegar a frustrarse por no poder llevar a cabo diferentes tareas o comunicarse en otro idioma; y por último, se pierde la capacidad gestual y corporal, lo que limita la comunicación (Vázquez Márquez 2011, 163).

Ahora que ya se han comentado todos los aspectos que envuelven a *Second Life* podemos extraer dos conclusiones o teorías. Por un lado, esta aplicación en sus inicios se presentó como muy prometedora a corto plazo, sin embargo, aún hoy en día muchas de las desventajas que se han comentado siguen estando presentes (fallos técnicos, falta de conocimientos por parte de los estudiantes y de control sobre sus avatares) por lo que al igual que ocurrió en las primeras etapas de la IA, los seres humanos tienden a creer que los grandes cambios tecnológicos son inminentes, pero estos no llegan a producirse. Y por otro lado, podemos concluir que las formas de aprendizaje totalmente inmersivas todavía no son posibles; no obstante ¿supondrá la aparición de nuevas herramientas de la

IA como *ChatGPT* la modificación total de la educación como la conocemos? ¿O al igual que el resto de las innovaciones que la preceden se quedará estancada en un futuro próximo? A lo largo de los siguientes apartados podremos vislumbrar la respuesta a estas cuestiones.

3.1 La creación de *ChatGPT*

En este apartado se va a tratar el modelo de lenguaje *ChatGPT-3* como ejemplo de plataforma de IA aplicada al contexto educativo. Para ello, es preciso situar primeramente el laboratorio de investigación de inteligencia artificial que ha desarrollado *ChatGPT*, para después hacer un análisis de su funcionamiento.

En su funcionamiento es preciso atender a dos cuestiones. Por un lado, la similitud entre los parámetros de relación y las conexiones sinápticas del ser humano. Por otro lado, el concepto de lenguaje que se deriva de esta IA. El análisis de estas dos características se justifica porque se trata de aquellas que han hecho evolucionar más a esta IA y que a la vez, paradójicamente, fueron las que más hicieron evolucionar al ser humano como factores biológicos de la hominización (Cela-Conde y Ayala, 2001). Sorprendentemente, parece ser que el camino de la evolución de la IA está siendo muy similar al que tuvo el ser humano en este sentido.

En primer lugar, en cuanto al análisis de la relación de los parámetros con las conexiones sinápticas, cabe destacar que hoy en día contamos con numerosas compañías centradas en la investigación de este campo, pero la encargada de crear el *chatbot* más conocido del momento es la empresa conocida como *OpenAI*. Un *chatbot* es una aplicación de *software* basada en IA que puede entablar conversaciones similares a las de los humanos. Los usuarios pueden hacer preguntas o realizar solicitudes y el sistema responde en segundos (Rudolph, Tan y Tan 2023, 3). Este laboratorio fue creado hace relativamente poco, en 2015, con sede en San Francisco y con fundadores conocidos a nivel internacional como Elon Musk (dueño de la empresa *Tesla* y *PayPal*, entre otras), Reid Hoffman (cofundador de *LinkedIn*, miembro de *PayPal* y *Facebook*), Peter Thiel (cofundador de *PayPal* e inversor de *Facebook*), Greg Brockman (presidente de *OpenAI*) y Sam Altman (director de *OpenAI* y fundador de *Y Combinator*).

Un aspecto que destacar es que *OpenAI* en sus inicios no era de la misma manera que la conocemos actualmente. En sus comienzos, el proyecto era más bien una organización sin ánimo de lucro que buscaba proporcionar a la sociedad las herramientas que la inteligencia artificial iba desarrollando. En cambio, la realidad no es esa, en 2019 pasó a convertirse en una empresa con intención de obtener una serie de ganancias; por lo que ese concepto de código abierto quedó atrás.

No fue hasta 2020 que la empresa alcanzó su verdadera popularidad gracias al lanzamiento de las primeras versiones de *ChatGPT*. Así pues, nos encontramos frente a un modelo de lenguaje, que a través del *deep learning* consigue generar diferentes tipos de texto como traducciones, resúmenes, correcciones, respuestas a preguntas que se le formulan, etc.

Añadido a esto, su principal preocupación se centra en llegar a conseguir el nivel de entendimiento humano. Es por ello, por lo que con cada versión que se lanza al mercado tiene como principio de mejora los parámetros que se emplean. Es decir, a mayor nivel de parámetros más cerca se estará de poder recrear las redes neuronales de nuestro cerebro. Este aspecto se puede ejemplificar de la siguiente manera: nuestra mente está formada por unas cien mil millones de neuronas, que traducido a conexiones sinápticas equivaldría a entre 100 y 500 billones de conexiones. Y adaptado a las versiones de *ChatGPT*, nos encontramos con que en la versión GPT (Generative Pre-trained Transformer) contaba con 110 millones de parámetros; en GPT-2 con 1,5 mil millones; y, en GPT-3 con 175 mil millones. Pero nada de esto se puede comparar con los que posee *ChatGPT-4*, que son cerca de 100 billones de conexiones. En definitiva, cada vez se acerca más la realidad de que existan modelos de IA similares a los humanos (Rudolph, Tan y Tan 2023, 2-3).

Bien es cierto que la extensión a toda la sociedad de la versión *ChatGPT-3* no se produjo hasta 2022. En noviembre, se lanzó de manera gratuita este *chatbot* basado en el lenguaje natural que consiguió recoger a millones de nuevos usuarios en los primeros días. Además, algo que diferencia a este sistema es que no se encuentra en constante actualización, es decir, la información que proporciona viene dada de todo lo que

almacenó hasta 2021, por lo que no se le puede preguntar por nada posterior (Rudolph, Tan y Tan 2023, 3-4).

En segundo lugar, en cuanto al análisis del modelo de lenguaje, es preciso analizar qué significa la palabra *prompt* y por qué tiene un papel fundamental en *ChatGPT-3*. Añadido a esto, también se explicarán los tipos de *prompts* y la utilidad de cada uno de ellos según el contexto.

Así pues, los *prompts*, son las instrucciones que se le dan a *ChatGPT* para que este nos conteste, es decir, son las preguntas o frases que escribimos y sobre lo que buscamos obtener una respuesta. Ahora bien, para que nos proporcione la información exacta que queremos, debemos formular correctamente el *prompt*; de lo contrario, puede que no nos facilite lo que queremos saber. Es por este motivo, por el que existen varias recomendaciones para poder comunicarse eficazmente con el *chatbot*.

Por un lado, es fundamental tener muy claro el objetivo de la conversación, y por consiguiente, otra recomendación es que el mensaje se transmita igualmente de manera clara y precisa para evitar confusiones. En relación con estos aspectos, aparece el lenguaje, que es el que nos ayudará a transmitir nuestra petición de una manera comprensible, este ha de ser de igual manera natural y explícito. No obstante, aunque la pregunta debe ser breve y concisa, tiene que proporcionar un mínimo contexto para poder obtener una respuesta acorde a lo que solicitamos. Por otro lado, algunas cuestiones a evitar serían: rehuir de las preguntas abiertas, hay que buscar lo concreto y evidente; utilizar un registro adecuado a la situación comunicativa en la que nos encontramos, en este caso un registro estándar o formal según nuestras demandas; y como en otros ámbitos, es imprescindible repasar aquello que hemos escrito para que esté todo en orden (Morales Chan 2023, 2-3).

Tras haber definido los *prompts*, a continuación se describen los tipos: secuenciales, comparativos, argumentales, con una perspectiva profesional y de deseos (Morales Chan 2023, 4-5):

- **Prompts secuenciales**: en este primer tipo, el *prompt* que da inicio a la conversación con *ChatGPT* no trata de resolver la cuestión en una única respuesta;

sino que busca una secuencia de preguntas y respuestas hasta alcanzar la información que deseamos de manera completa. Un ejemplo podría ser, preguntar primero qué es *ChatGPT*, posteriormente preguntar acerca de *OpenAI* (debido a que nos habría dicho que es la empresa que lo ha desarrollado) y finalizar con los objetivos o las motivaciones que hay detrás de dicha empresa. De esta manera, conseguimos una visión mucho más global y completa de lo que rodea a esta compañía.

- **Prompt comparativo**: esta pregunta, a diferencia de la anterior, busca una respuesta más concreta, puesto que lo que se consulta al chat es que compare dos elementos, conceptos, principios, etc.
- **Prompt argumental**: en este *prompt* dejamos de lado la objetividad que se buscaba en los dos anteriores para adentrarnos en una contestación subjetiva, en la que se nos proporcione el punto de vista de la IA acerca de una determinada situación. Una idea de *prompt* argumental podría ser que nos argumentara si podría suponer un peligro para la educación o qué riesgos se podrían desarrollar en un futuro si todas nuestras actividades estuvieran relacionadas con *ChatGPT*.
- **Prompt de perspectiva profesional**: en este caso la función de la IA es la de convertirse en un individuo concreto con el fin de proporcionar una respuesta mucho más precisa y desde los conocimientos de la persona que ha de parecer. En palabras de Morales Chan trata de lo siguiente:

“Actúa como [autor o profesión] y describe [tema] + contexto”, lo que permite que Chat GPT asuma un papel específico y proporcione una respuesta más detallada, objetiva y estructurada. Por ejemplo, un *prompt* de perspectiva profesional podría ser "Actúa como un experto en salud pública y describe cómo se podría abordar la pandemia actual en términos de políticas públicas" (Morales Chan 2023, 6).

- **Prompts de lista de deseos**: este último *prompt*, al igual que el anterior, también busca obtener una información concreta acerca de un determinado tema. La manera de proceder es la de, en un primer momento, decir qué es lo que se está buscando y, a continuación, solicitar algunos ejemplos. Un ejemplo aplicado a la vida cotidiana sería: estoy buscando aumentar la ingesta de frutas y verduras, ¿podrías crear un menú semanal en el que estos alimentos estén incluidos en un 70%?

3.2 La introducción de *ChatGPT* en el aula

En este apartado dejamos ya de lado la parte teórica del tema para adentrarnos en las cuestiones más prácticas. De esta manera, se produce el cambio de la metodología argumentativa empleada en los puntos anteriores hacia la metodología sociológica, ya que se incluye un cuestionario. Por lo que la disposición de este punto se basa en una primera parte, en la que se explican todos los aspectos relacionados con el cuestionario; y, en una segunda parte, se lleva a cabo una reflexión en torno a los resultados obtenidos.

El objetivo de realizar un cuestionario era conocer el punto de vista de los docentes con respecto a la IA. Así pues, el público al que se dirigió la encuesta fue el del profesorado de varios centros de enseñanza secundaria, en su mayoría de Castellón. En este sentido, la encuesta cuenta en un inicio con varias preguntas para conocer el sector al que pertenece cada docente, después continúa con cuestiones acerca de *ChatGPT*, para finalizar con la opinión de cada uno. A continuación, se mostrarán y analizarán las diez preguntas y los resultados obtenidos, respectivamente.

El número de participantes fue de 75 y destacan tres centros en concreto: IES Francesc Ribalta (27), Colegio Lope Castellón (14) y el IES Vicent Sos Banyat (13). En cuanto al departamento de cada docente destaca el de Educación Física (15), Ciencias (17), Lenguas (18) y Humanidades y Geografía e Historia (17).

En primer lugar, un 46,7% de los participantes que respondieron a la encuesta dan clase al primer ciclo de ESO (1º-3º curso), un 33,3% a bachillerato y un 20% a 4º curso de ESO. En segundo lugar, un 70,7% conoce *ChatGPT*-3, y casi un 30% lo desconoce. En tercer lugar, tan sólo un 24% ha detectado el uso de IA en los estudiantes en la realización de actividades. Algo que, sin duda concuerda con que el 25,3% del profesorado encuestado haya utilizado alguna vez una IA para la creación de contenido o material educativo. Concuerda, porque las cifras de estas dos respuestas son similares, y es lógico que si no se tiene conocimiento de la IA, tampoco se podrá detectar cuándo lo utilizan los/las estudiantes ni se utilizará personalmente para crear materiales docentes. Los resultados de estas dos respuestas anteriores contrastan con la pregunta 7, donde un 81,3% cree que el uso de la IA por parte de los estudiantes supone un riesgo para la educación.

No obstante, la lectura de estas cifras puede ser interpretada de una forma diferente también.

En primer lugar, un 70,7% sí conoce el modelo de lenguaje *ChatGPT*, no obstante, aunque sea mayor el número de docentes que han oído hablar de esta tecnología, el hecho de que 22 profesores (29,3%) no lo conozcan, resulta preocupante. Esto es así, puesto que como se ha mencionado anteriormente en este trabajo, tanto la IA como el caso concreto de *ChatGPT* suponen una cuestión real y actual que está comenzando a ocasionar problemas y cambios. Este hecho se puede demostrar con los resultados de la siguiente pregunta, en la que un 76% de los profesores ha detectado ya el uso de herramientas de la inteligencia artificial, tales como *Paraphraz.it*, *Photomath* o *Sintaxis.org*, estas se dedican a la reescritura de los textos que se le proporcionan, a resolver enunciados matemáticos y a analizar sintácticamente oraciones, respectivamente.

Continuando con las siguientes dos preguntas, nos encontramos con el hecho de que 56 docentes (74,7%) han empleado algún tipo de IA para producir materiales para su asignatura. Esto podría hacernos pensar que ven las herramientas creadas por la inteligencia artificial como algo bueno y de lo que poder sacar provecho. No obstante, un porcentaje aún mayor, un 81,3% de los profesores considera que el uso por parte de los estudiantes supone un riesgo para la educación. De manera que, puede resultar algo incoherente que esos mismos docentes consideren por una parte que puede ser útil para ellos, pero un riesgo para sus alumnos, ¿quizá porque piensan que ellos hacen un uso diferente?

Ahora bien, si se analiza esta pregunta desde otro punto de vista, se puede retomar la opinión que surgía en la pregunta tres acerca de si conocían o no *ChatGPT*; en relación con esto, se ha mencionado que resulta alarmante que todavía un 29,3% del profesorado no conozca este modelo de lenguaje. Sin embargo, sorprende todavía aún más que haya docentes que conozcan la herramienta y no la consideren como un riesgo en la educación, que es un 18,7% de los encuestados. Resulta alarmante puesto que el debate no debería estar centrado en si es o no un riesgo, ya sabemos que lo es, son los propios creadores quienes nos están advirtiendo. Hace poco que Sam Altman declaró en el Capitolio de Estados Unidos que es necesaria una regulación de la IA, advirtiendo de sus riesgos (Seisdedos, 2023). El debate radica en qué podemos hacer para frenar esta situación y en

el problema de que si ni siquiera los mismos educadores están al tanto de la situación, no se van a producir los cambios necesarios.

Para finalizar con el cuestionario se va a proceder al análisis de las tres últimas preguntas, lo que nos permitirá abordar el siguiente apartado, que es el de comentar por qué supone una amenaza la IA y cuáles son los riesgos que tiene en la ESO y, también, en la educación en general. Las preguntas en cuestión son: 8) señalar qué riesgos de los mencionados le parece más inmediato en la utilización de la IA por parte de los/las estudiantes; 9) indicar los beneficios que podría tener el uso de IA por parte de los estudiantes; y 10) ¿Cómo considera que la utilización de la IA podría modificar el sistema educativo actual?

La pregunta ocho, que es la más relacionada con la tesis principal del presente trabajo, concluye que los tres principales riesgos a los que nos enfrentamos son los siguientes. En primer lugar, la falta de criterio y pensamiento crítico con un 74,7% (56 profesores). En segundo lugar, le sigue en porcentaje la dependencia tecnológica con un 53,3% (40). Y en tercer y último lugar, la imposibilidad de distinguir el trabajo personal del plagio con un 42,7% (32). A su vez, estos tres riesgos que acabamos de mencionar coinciden con los tres pilares fundamentales de la reflexión que tendrá lugar en el siguiente apartado.

En la pregunta nueve ya podemos conocer un poco el punto de vista de cada uno, lo que nos deja con tres tipos de opiniones: los que consideran que no aporta ningún beneficio, los que creen que no tienen los conocimientos suficientes para poder aportar su parecer y los que sí expresan diferentes beneficios, que son las respuestas que se van a comentar. En este sentido, destacan respuestas como: “mejorar sus resultados”, “búsqueda de información” “solucionar dudas”, “lluvia de ideas ante proyectos”, “amplía los recursos disponibles para el alumno”, “*suport en nivells educatius que demanen un reforç (ACIs, compensatòria)*”, “corrección a la hora de procesar textos, errores comunes como pueden ser los fallos de puntuación o gramaticales”, “tener acceso a información cultural e idiomas de otros países imposibles de visitar”. Sin embargo, en tan solo una respuesta se tenía en cuenta que estas mejoras han de usarse correctamente; es decir, la problemática en torno a la IA está centrada principalmente en establecer reglamentos y normativas acerca de cómo ha de usarse para que, tanto los docentes como los estudiantes, utilicen estas herramientas de manera ética y responsable.

Continuando con la misma pregunta, otros consideran que proporciona de manera fácil e inmediata un gran conocimiento, hasta este punto estaríamos en lo cierto; no obstante, con una de las respuestas entra otra cuestión a tratar acerca de *ChatGPT*. La respuesta dice: “información veraz de forma rápida”, esa creencia de veracidad, sin embargo, no es posible. *ChatGPT* nos proporciona una gran cantidad de información de manera sencilla y rápida, pero esto no significa que sea verdad. El propio *chatbot* asume que los datos que proporciona no son 100% verídicos y que es necesario utilizar nuestro criterio y otras fuentes de información. En palabras del propio modelo de lenguaje responde de la siguiente manera a esta cuestión (ChatGPT 2023):

Como modelo de lenguaje, me esfuerzo por proporcionar respuestas precisas y útiles en función de los datos con los que fui entrenado, que abarcan hasta septiembre de 2021. Sin embargo, ten en cuenta que no puedo garantizar la veracidad absoluta de toda la información que proporciono. Aunque me esfuerzo por ser preciso, es posible que haya ocasiones en las que mi respuesta no refleje los desarrollos más recientes o que la información haya cambiado desde mi último entrenamiento. Por lo tanto, siempre es una buena idea verificar la información por tu cuenta o consultar fuentes confiables adicionales para obtener la información más actualizada y precisa.

Dejando de lado la cuestión de la veracidad, es preciso mencionar un aspecto que no ha sido tenido en cuenta por el profesorado, quizá por su desconocimiento, que es el idioma. Se ha podido comprobar que ante idiomas más distantes como el rumano o el chino *ChatGPT* no tiene la misma fluidez léxica ni argumentativa que en idiomas predominantes como el inglés (Rudolph, Tan y Tan 2023, 8). Así pues, puede deducirse que los beneficios que puedan derivarse de la IA únicamente se producirán al completo en lugares con una lengua prestigiosa.

Pasando a la última pregunta, la mayoría de los encuestados coincide en un factor, que es que la IA va a revolucionar el sistema educativo, ya sea positiva o negativamente, pero lo va a modificar por completo. Por un lado, aquellos que ven la entrada de la inteligencia artificial en la educación como algo bueno creen que cambiará los siguientes aspectos: “puede ayudar a la autonomía de aprendizaje”, “elimina trabajos improductivos”, “actualizar contenidos y dinamizar metodología” o “crear entornos más interactivos de aprendizaje para motivar al alumnado”. Por otro lado, aquellos que creen que tendrá un impacto negativo, piensan que provocará lo siguiente: “menos esfuerzo y menos motivación por parte de los estudiantes”, “sin la supervisión del docente, eliminar poco a poco el pensamiento crítico”, “bajando la calidad de la enseñanza”, “volver a una

educación basada en la memoria”, “podría acabar con la relación alumno-profesor en el aula”.

En resumen, el cuestionario ha sido útil en varios aspectos. Por una parte, ha hecho que se tenga una idea general de en qué medida el profesorado conoce o no *ChatGPT* y si lo considera o no un riesgo, cuya respuesta ha sido en ambos casos afirmativa por parte de la mayoría. Y, por otra parte, ha permitido conocer el punto de vista que tiene cada docente con respecto a cómo es la educación ahora y cómo podría llegar a cambiar, además de introducir ciertos riesgos que se comentarán en el apartado siguiente.

4. Análisis crítico de la introducción de la IA en la educación

En el punto anterior se ha mencionado brevemente algunos de los riesgos que produce la inteligencia artificial en la educación. Por lo que, en este punto, se van a tratar en profundidad; pero, antes de ello, primero se introducirán algunas citas de diferentes personalidades acerca de la IA y sus peligros.

Stephen Hawking: «la IA podría ser el peor evento en la historia de nuestra civilización» (Infobae 2017).

Elon Musk: «creo que la IA es algo por lo que debemos preocuparnos bastante, y mantenernos atentos a su seguridad» (Erard 2023). «La inteligencia artificial tiene el poder de destruir civilizaciones» (Clementín 20203).

Sam Altman: «Si la inteligencia artificial sale mal, puede salir muy mal (Seisdedos 2023); «mi peor temor es que causemos un daño significativo al mundo» (Vázquez 2023). La cita de Altman contrasta con lo que, a través de su cuenta de *Twitter*, mencionó (Gil Rodríguez 2022):

ChatGPT es increíblemente limitado, pero lo suficientemente bueno en algunas cosas como para crear una impresión engañosa de grandeza. Es un error confiar en él para algo importante en este momento. Igualmente, es una vista previa del progreso en la vida; tenemos mucho trabajo por hacer en cuanto a solidez y veracidad.

Ante este panorama más que poco alentador, se podrían deducir varios aspectos. Ante todo, cabe resaltar, que tanto Elon Musk como Sam Altman son los fundadores de *OpenAI*, y consecuentemente de *ChatGPT*. De manera que, son los mismos creadores de nuevas formas de inteligencia artificial los que no están advirtiendo de los problemas y daños que podría desencadenar esta tecnología. Por lo que, si ni siquiera aquellos que están favoreciendo al avance tecnológico se fían de ello, ¿cómo va la sociedad a dejar en manos de la IA o *ChatGPT* el futuro de la humanidad?, resulta del todo incoherente. Es por este motivo, por el que debemos tomar la inteligencia artificial como un elemento peligroso, que todavía podría serlo más en el ámbito educativo.

Ahora bien, adentrándonos en los riesgos que se pueden derivar de la IA, nos encontramos con el hecho de que el uso de diferentes aplicaciones y/o plataformas en el día a día de los estudiantes podría generar una dependencia a la tecnología sumamente considerable. A su vez, esta dependencia podría provocar que los jóvenes dejaran de lado todavía más las experiencias que tienen en el mundo real para sumergirse en los mundos virtuales y con ello, eliminar una parte del crecimiento humano que se da a partir de las relaciones sociales (Giró García y Sancho-Gil 2022, 142). Además, de provocar la falta de capacidad para realizar diversas tareas en el caso de que no se disponga de la herramienta digital que lleva a cabo esa labor o la pérdida del desarrollo cognitivo. Este hecho supone un acercamiento a la concepción que tenía Karl Popper acerca del futuro de la sociedad:

No es imposible concebir una sociedad en que los hombres no se encontraran nunca, prácticamente, cara a cara: donde todos los negocios fuesen llevados a cabo por individuos aislados que se comunicasen telefónica o telegráficamente y que se trasladasen de un punto a otro en automóviles herméticos (Popper 1945, 158).

Aunque en su momento pudiera parecer disparatada la idea de que este tipo de sociedad pudiese ocurrir, actualmente podemos comprobar que su presuposición no puede ser más acertada y coincide con la idea que se menciona en el párrafo anterior. Día a día van imperando todavía más las relaciones virtuales, cada vez son más comunes las videollamadas, los videojuegos, las compras y los trámites *online*, por lo que las relaciones físicas están en detrimento.

Continuando en la línea de la deficiencia de las relaciones sociales que podría derivarse de la dependencia de las TIC (Tecnologías de la Información y la Comunicación), nos encontramos con que uno de los aspectos que se asumen como un beneficio de la IA (autonomía, personalización de los contenidos y enseñanza autodidacta) en realidad puede provocar la pérdida de diversos conocimientos. Esta ausencia de saberes se debe, a que como sociedad, muchos aprendizajes se llevan a cabo de manera colectiva; es decir, algunas competencias las adquirimos individualmente, pero otras muchas, solo pueden darse en actividades colectivas. Por lo que, si se lograra un aprendizaje particular, podríamos dejar descuidadas enseñanzas que se logran en grupo como el intercambio de ideas, la mejora de la comunicación interpersonal, la toma de decisiones o el sentido de la responsabilidad, entre otros (Jara y Ochoa 2020, 11).

Antes de continuar con el resto de problemáticas, es necesario tener en cuenta ciertos aspectos. Como bien es sabido, para poder acceder a *chatbots* como *ChatGPT* u otro tipo de herramientas se necesitan dos elementos; por un lado, disponer de conexión a internet y, por otro lado, disponer de un dispositivo a través del que poder hacerlo. Ahora bien, si el objetivo del desarrollo e implementación de este tipo de plataformas de IA es el de aplicarlas en la educación de manera global, la falta de recursos de los países en vías de desarrollo o subdesarrollados, lo único que lograría es el avance de los países desarrollados, creando una brecha educativa todavía mayor entre los diferentes países (Jara y Ochoa 2020, 10).

En este sentido, ya se han realizado estudios que han concluido que son siete los impedimentos que dificultan el acceso a la IA en los países sin desarrollar (UNESCO 2019, 28): «1- disponibilidad de hardware de TIC, 2- disponibilidad eléctrica, 3- fiabilidad de internet, 4- el coste de datos, 5- las habilidades básicas de los estudiantes acerca de las TIC, 6- el idioma, 7- la falta de contenido culturalmente apropiado». Así pues, resulta imprescindible «definir internet como un derecho humano y crear múltiples alianzas internacionales para construir infraestructuras en los sectores más pobres» (UNESCO 2019, 28). Es por este motivo, por el que uno de los ODS de la Agenda 2030 es «garantizar una educación inclusiva, equitativa y de calidad y promover las oportunidades de aprendizaje durante toda la vida para todos» (UNESCO 2019, 12) para que los avances ocurran tanto en países que están desarrollados como los que no lo están todavía.

Sin embargo, los inconvenientes anteriores no son los únicos, encontramos otro factor importante de la inteligencia artificial que puede ocasionar desigualdades sociales, que es el sesgo social que se produce. Es un hecho indiscutible que todas las decisiones de la IA están sesgadas, puesto que esas decisiones se toman en base a unos conocimientos que han sido transmitidos por parte de una persona, y los individuos, en tanto que personas tenemos unos valores y creencias propios (Flores Vivar y García Peñalvo 2022, 43). Por ejemplo, si aquel que ha programado plataformas de la IA siente rechazo hacia los homosexuales y los extranjeros, esa IA automáticamente asume esos ideales y es homófoba y racista de igual manera. Otra forma de definir esta situación es la que nos provee (Calvo 2020, 230):

No en pocas ocasiones las decisiones de estos modelos matemáticos han evidenciado una clara inclinación hacia el rico, el heterosexual, el hombre o el caucásico en detrimento del pobre, el homosexual, la mujer o el afroamericano, por señalar algunos casos, promoviendo la exclusión, ampliando la brecha de las desigualdades y reproduciendo los mismos patrones que la sociedad intenta erradicar desde hace décadas.

De esta manera, nos encontraríamos actualmente frente a dos riesgos de exclusión social. El primero, vinculado a la dificultad en el acceso de este tipo de enseñanza digital en los países que todavía no se encuentran desarrollados. Y, el segundo, en la discriminación que podría producirse hacia todo aquel que no sea un hombre blanco, heterosexual y rico. Ambos hechos suponen un gran riesgo en la educación, puesto que el uso de la inteligencia artificial, por una parte, solo sería accesible en su totalidad a una cantidad de estudiantes, que suponen una minoría si tenemos en cuenta el contexto global. Y, por otra parte, en aquellos lugares que sí se puede llevar a cabo en cuanto a recursos, tampoco supondría una mejora educativa ni inclusiva, ya que estaríamos tomando como válida una propuesta pedagógica que en lugar de mejorar la calidad de aprendizaje, provocaría la discriminación y/o rechazo de algunos estudiantes en el aula (por su etnia, sexo, nivel socioeconómico u orientación sexual).

Continuando con otro riesgo, nos encontramos frente al que surge en relación con los datos que se obtienen de este tipo de herramientas de inteligencia artificial, que es el de la falta de privacidad y seguridad de los usuarios, en este caso, discentes. Esta problemática no es específica únicamente en este ámbito, sino que a nivel global se tiene una preocupación acerca del tratamiento y distribución de los datos que se obtienen en cada página web, aplicación o videojuego que visitamos. Es importante tener unos buenos protocolos de seguridad que nos garanticen que los datos de los estudiantes, que en su mayoría son menores de edad, van a ser tratados de manera correcta y no van a convertirse en víctimas de ciberataques o manipulación en base a ello (Jara y Ochoa 2020, 16).

Por último, vamos a tratar los aspectos que rodean al principal riesgo que se ha establecido por parte de los docentes en la encuesta analizada anteriormente, que es el de la falta de criterio y pensamiento crítico. Al igual que se mencionaba en la encuesta, *ChatGPT* es capaz de proporcionarnos una gran cantidad de información en un periodo de tiempo muy corto, pero este hecho podría suponer que los alumnos, en cada tarea que tuvieran que

realizar, emplearan este modelo de lenguaje para obtener el resultado final (resúmenes, trabajos, redacciones, deberes). Este hecho desencadenaría un cambio en la forma de aprender, puesto que se tendrían que cambiar los trabajos que se envían para poder confirmar que han sido elaborados por ellos mismos y no por la IA. No obstante, este no es el principal riesgo que ocasiona, sino que precisamente es la capacidad de poseer un pensamiento crítico la que se necesita con urgencia desarrollar para poder hacer un buen uso de la inteligencia artificial, y aquella que se pierde, justamente al utilizarla en exceso y sin un conocimiento previo.

Aplicado a un ejemplo práctico, supondría lo siguiente. Si un estudiante conoce que *ChatGPT* es capaz de realizar resúmenes, disertaciones, redacciones, traducciones, etc., es probable que debido a la facilidad que esto supone realice todas sus tareas de esta manera. Ahora bien, al llevar a cabo esta práctica, el estudiante está perdiendo su capacidad de redacción, imaginación, de razonamiento, de opinión acerca de diferentes temas y de pensamiento crítico, entre otros. Al no hacer un análisis de aquella información que le es proporcionada para ver si es cierta o no, si corresponde con sus ideales o con aquello que se le solicita, se está frenando su desarrollo intelectual.

En esta misma línea, surge el dilema de si se puede distinguir o no el trabajo personal del plagio, que supone el tercer riesgo mencionado en la encuesta. Por lo que a continuación vamos a centrar la reflexión en este.

Desde que el plagio se ha convertido en un concepto muy reiterado en las discusiones sobre malas prácticas en la investigación, uno podría esperar que en lo que se refiere a determinar lo que constituye una mala práctica en la investigación científica, existiera un acuerdo general y un gran nivel de precisión sobre lo que constituye el plagio. Sin embargo, mientras que existe un acuerdo sobre los casos paradigmáticos de plagio, no hay un acuerdo general sobre cómo debería ser definido. De hecho, raramente se plantea este caso.

Desde la *Office of Research Integrity* del Departamento de Salud y Recursos Humanos, en Estados Unidos, parece bastante simple: «Implica robar el trabajo de otra persona y

mentir sobre ello después»¹ (traducción propia). Otros, sin embargo, hablan de “copiar” parte del trabajo publicado de alguien y usarlo sin conocer que está tomado de ese alguien. En el *Longman Dictionary of Contemporary English*, el acto de plagiar se define como: «cuando alguien utiliza las palabras, ideas o el trabajo de otra persona y hace creer que son suyos»² (traducción propia).

En las definiciones más académicas, las nociones de apropiación y crédito son centrales: Según *The European Code of Conduct for Research Integrity* «El plagio consiste en usar el trabajo y las ideas de otras personas sin dar el debido crédito a la fuente original, violando así los derechos del autor o autores originales sobre sus productos intelectuales»³ (traducción propia). La idea básica parece ser que alguien deliberadamente toma el trabajo de algún otro, tanto si es el núcleo de una idea, un método, un dato, un resultado, un texto, y lo presenta como propio en lugar de darle crédito a la persona real a la que pertenecen.

El plagio puede ser más o menos grande, desde capítulos de un libro hasta libros enteros, pasajes de artículos o todo el artículo. ¿Existe algún límite para lo que se considera plagio en cuanto al tamaño? Y si lo hubiera, ¿cómo se mediría? ¿Quizá con un cierto número de palabras copiadas, o un cierto porcentaje de superposición de textos, o dependería de algo más, como la calidad de lo que utiliza sin notificación apropiada?

El tamaño importa pero dependiendo de su impacto en la calidad. Esto es lo mismo que decir que la cantidad importa indirectamente, mientras que la calidad afecta directamente. El plagio podría consistir en un pasaje breve de texto, en principio incluso en una sola palabra. Pero quizá esa palabra es clave en descubrir algo en un determinado campo o área, y podría dar la impresión de que el plagiador es quien ha inventado el término.

En resumen, el plagio puede consistir en principio, en una pequeña palabra, mientras que hay muchas frases estándares que describen métodos de investigación y que no serían consideradas como plagio. Por tanto, el plagio tiene que ver más con la calidad de lo que

¹«It involves stealing someone else’s work and lying about it afterward» (Sox 2012, 1).

²«When someone uses another person’s words, ideas, or work, and pretends they are their own» (Longman 2023).

³ «Plagiarism is using other people’s work and ideas without giving proper credit to the original source, thus violating the rights of the original author(s) to their intellectual outputs» (The European Code of Conduct for Research Integrity 2017, 8).

se dice que con la cantidad de palabras. El principal problema de la relación del plagio con el *ChatGPT* es que es indetectable, dado que el contenido creado es único. A pesar de que esta IA bebe de muchas bases de datos, la creación del lenguaje en sus respuestas es original, e incluso si otra persona le pregunta lo mismo, dará una versión distinta de la respuesta. Además, si se le pregunta si un texto ha sido obra suya, en muchas ocasiones su respuesta es afirmativa aunque no sea cierto; lo que hace dificulta todavía más complicado detectar su uso.

En este punto, nos encontramos ante la duda de si los propios docentes o incluso sistemas hasta ahora sofisticados como *Turnitin*, son capaces de discernir si un trabajo ha sido elaborado por un ser humano o por inteligencia artificial. *Turnitin* se define como un «software mal llamado “antiplagio” por su alta eficiencia para identificar similitud de *copy-paste* de documentos en la web» (Díaz Arce 2023, 17). Este autor, además de definir qué es el *software* más utilizado en la educación española, elaboró un experimento para descubrir si la cuestión que se planteaba líneas atrás es posible o no, de manera que a lo largo de este apartado vamos a analizar los resultados que obtuvo.

Antes que nada, el contexto del estudio se basa en el análisis de cincuenta trabajos académicos realizados por estudiantes, y la misma cantidad, pero creados por *ChatGPT* que tenían como tema la comparación y el contraste de dos enfermedades nutricionales. Para ello, se establecieron tres variables: la frecuencia general de similitud (IGS), es decir, el grado de similitud del trabajo que se presente como propio con los datos que posee *Turnitin*; la frecuencia de ensayos en riesgo de plagio; y, la frecuencia de plagio académico.

Ahora bien, los resultados que se obtienen son los siguientes. En primer lugar, destaca el hecho de que los estudiantes han escrito más palabras que el *chatbot*, una media de 382 palabras frente a 276; asimismo, en cuanto al total de palabras también se puede observar una clara diferencia, puesto que los textos generados por la IA no superaran las 350 palabras, en cambio los estudiantes llegaban incluso a las 600. En relación a la cuestión del plagio, se concluye que en un 56% de los trabajos de estudiantes se podía detectar un valor de IGS que pone en riesgo la autenticidad del ensayo, mientras que los elaborados por *ChatGPT* suponían un 24%. Este hecho, concuerda con el porcentaje de plagio

presente en los trabajos entregados por los alumnos de bachillerato, que supone entre un 40 y un 90%; por lo que la cifra no resulta sorprendente.

A pesar de que se haya obtenido una mayor presencia de plagio por parte de los estudiantes, hay un factor en concreto que *Turnitin* no tiene en cuenta en su análisis y es preciso conocer. Este elemento es que los textos que nos proporciona *ChatGPT* tienen como característica común que normalmente se repiten estructuras sintácticas; pero este hecho, no es tenido en cuenta por el *software*, puesto que a pesar de la repetición, el texto posee una buena redacción y no es similar a ningún otro que se encuentre en la base de datos (que es la principal función de esta herramienta: la comparación de textos, no la coherencia y cohesión del texto) (Díaz Arce 2023, 17-23).

En esta línea, aparece también un componente relevante en cuanto a *ChatGPT*, y es que, a pesar de que a simple vista podría parecer que cumple a la perfección con todo lo necesario para crear relatos igual de elaborados que una persona, esto no es del todo cierto. Rudolph, Tan y Tan así lo demuestran en una prueba que hicieron en la que le pedían un ensayo de 2000 palabras acerca de un tema concreto y que tuviera veinte citas referenciadas, de la cual extrajeron los siguientes resultados:

Después de plantear la pregunta a *ChatGPT*, la IA mostró una alerta de error en la red. Después de múltiples intentos, produjo un ensayo de aproximadamente 500 palabras y cinco referencias al final del texto [...]. Aunque *ChatGPT* produjo el ensayo de manera eficiente en 120 segundos, el contenido fue bastante decepcionante. Le faltaba amplitud y profundidad. Era principalmente genérico y descriptivo, sin evidencias que lo respaldara. Tampoco pudo proporcionar referencias en el texto ni al final (o peor aún, inventó referencias falsas)⁴ (traducción propia).

Estas conclusiones coinciden con que a pesar de que *ChatGPT* no proporcione escritos elaborados, las herramientas que tenemos a nuestro alcance no son aptas para la nueva tarea que queremos atribuirles, puesto que han sido diseñadas con otro fin. En resumen, *Turnitin* no nos serviría para comprobar si realmente un escrito ha sido redactado por un ser humano o por una máquina, de manera que, actualmente, la única forma que tenemos

⁴ After posing the question in ChatGPT, the AI showed an alert of a network error. After multiple attempts, it produced an essay of approximately 500 words and five end-of-text references [...]. Although ChatGPT efficiently produced the essay within 120 seconds, the content was quite disappointing. It lacked both breadth and depth. It was primarily generic and descriptive, with no evidence backing it up. It was also unable to give in-text and end-of-text references (or, worse, invented bogus references) (Rudolph, Tan y Tan 2023, 8).

para poder diferenciar si una obra es propia, es bajo la supervisión de una persona. Lo que nos deja con la conclusión de que actualmente no disponemos de las herramientas suficientes para poder abordar esta problemática y esta mala práctica sigue sin tener una solución, al contrario, puede que incluso se vea incrementada. En esta línea, cobran sentido las palabras que dijo Ortega y Gasset acerca de la tecnología «la técnica, cuya misión es resolverle al hombre problemas, se ha convertido de pronto en un nuevo y gigantesco problema» (Ortega y Gasset, 2004; Carabante Muntada 2007, 17). Ortega refleja a la perfección la situación actual, la IA se ha presentado como una mejora en la educación en cuanto a inmediatez y facilidad a la hora de elaborar trabajos; no obstante, lo que ha provocado es una necesidad imperante de protección frente a su uso.

Si bien hasta ahora se han analizado los riesgos de ChatGPT teniendo en cuenta el punto de vista del discente, existen otros dos peligros mucho más generales que implica la IA y que, si bien exceden la dimensión de este trabajo, no hay que perder de vista. La primera preocupación sería la pretensión de sustituir al ser humano no sólo en su trabajo sino en muchos aspectos sociales de su vida diaria. El segundo sería, de nuevo y lamentablemente, la emancipación de la técnica como un saber más allá de su dimensión tanto práctica como teórica, es decir, más allá de lo humano. En otras palabras, estos dos peligros se podrían resumir en la creación de un monstruo que no podemos controlar. Son muchos los filósofos que, mucho antes de la llegada de la IA, ya advirtieron sobre estas amenazas, tanto desde la fenomenología como desde la Escuela de Frankfurt.

Por un lado, desde la fenomenología de la ciencia, autores como Edmund Husserl (1991) o Tomas Kuhn (1971), ya nos advirtieron de que la ciencia debe de haber un cuestionamiento constante de la ciencia hacia sí misma, en lo que respecta a sus métodos y, muy especialmente en su relación con el ser humano. Con el auge de la IA, parece ser que se hace cada vez más cercano el advenimiento de la ciencia omniabarcadora que denunciaba Husserl (1991, 8). Tan omniabarcadora que ya ni siquiera el ser humano podría ser necesario.

Por otro lado, desde la Escuela de Frankfurt, con los filósofos de la primera generación y de la segunda, especialmente Jürgen Habermas, ya se anunciaron los problemas que comportaba la racionalidad instrumental. Filósofos como Adorno o Horkheimer denunciaron la insuficiencia de la razón diseñada exclusivamente sobre el modelo sujeto-

objeto de las ciencias empírico-naturales para poder responder a hechos sociales (Cortina, 2008, 41). También Habermas denunció que la ciencia tiene un interés ideológico basado en la dominación (Habermas (1986). Según este filósofo, que la técnica tiene propiedades que en determinados contextos son apropiadas, pero hay otros, en los que no; y es en estos últimos contextos en los que se convierte en el foco de los problemas de las sociedades modernas (Herrera Reyes 2022, 176-178; Carabante Muntada 2007, 15). En el caso que nos ocupa, ya no será el ser humano quien utilice la técnica y la ciencia para dominar, sino que será dominado por éstas por completo.

Advertencias muy similares también las realizaron filósofos como Hanna Arendt u Ortega y Gasset. Hanna Arendt lo hacía al cuestionarse, no el hecho de si somos los dueños de las máquinas que creamos o sus esclavos, sino la medida en que estas nos siguen siendo útiles o, por el contrario, nos están controlando; y no solo eso, la forma en la que están destruyendo el mundo y todo lo que le rodea. En cuanto a Ortega y Gasset, además de lo mencionado en apartados anteriores, podemos extraer la idea de que no debemos confiar plenamente en un avance tecnológico que es ilimitado, puesto que no sabemos lo que puede llegar a desencadenar. Ambos tuvieron que reflexionar acerca del auge de la técnica y de la industrialización que surgía en el siglo XX, de la misma manera que ahora se debe reflexionar acerca de la problemática que ha traído consigo la inteligencia artificial.

En definitiva, para hacer frente a los diferentes riesgos que surgen del uso de estas nuevas tecnologías, se establece que el componente ético es el que nos va a servir para llevar a cabo la aplicación de la IA en los diferentes ámbitos de manera correcta. Es decir, se tienen que llevar a cabo propuestas éticas que creen unas normas de actuación para que las inteligencias artificiales tomen decisiones lo más justas posibles y teniendo en cuenta los derechos humanos (Calvo 2020, 230; Flores Vivar y García Peñalvo 2022, 42). Además, de llevar a cabo políticas de igualdad (objetivo Agenda 2030) para que todos los países puedan optar al mismo desarrollo, como ya se mencionaba anteriormente.

La buena noticia, es que son diversas las organizaciones que ya están intentando hacer frente a esta situación, puesto que coinciden con la hipótesis de que la IA supone más riesgos que ventajas en la educación. Por un lado, encontramos algunos programas de investigación de diferentes universidades norteamericanas entre los que destaca: el centro HAI (*Human-Centered Artificial Intelligence*) creado por Universidad de Stanford y que

tiene como finalidad analizar el impacto que tendrá la IA en las diferentes áreas de la educación y la relación de la inteligencia artificial con lo humano. Por otro lado, tanto la UNESCO como el Parlamento Europeo y *AI for Good*, que es una «plataforma digital anual en la que los innovadores de IA y los propietarios de problemas aprenden, construyen y se conectan para identificar soluciones prácticas sobre la IA para avanzar en los ODS de la ONU»⁵ (traducción propia), están llevando a cabo iniciativas para fomentar el debate acerca de los límites de la IA y su impacto social (Barrios Tao, Díaz Pérez y Guerra 2020, 82-83).

⁵ «Is a year-round digital platform where AI innovators and problem owners learn, build and connect to identify practical AI solutions to advance the UN SDGs » (AI for Good).

Conclusiones

A lo largo del presente trabajo, se han abordado cuestiones teóricas como el significado y origen del término inteligencia artificial y las etapas en las que se divide, en el primer apartado, y conceptos clave como el *deep learning*, la IA fuerte o el metaverso (incluyendo sus tipos) en el segundo apartado. Además de varias cuestiones en el tercer apartado que engloban a *ChatGPT* como son su definición, origen y características, así como aspectos acerca de la empresa que lo ha desarrollado y la explicación de los tipos de *prompts*.

De igual manera, se han abordado también cuestiones prácticas como el análisis de los datos del cuestionario dirigido al profesorado, que nos han permitido ahondar en el propósito de la investigación (los riesgos de la inteligencia artificial). En este sentido, los riesgos que se han proporcionado en mayor medida en el cuarto apartado son: la falta de criterio y pensamiento crítico, la dependencia de las TIC y la imposibilidad de distinguir el trabajo personal del plagio; no obstante, se han introducido otros, como el aumento de la brecha digital entre países, la falta de seguridad en el tratamiento de los datos o la posible pérdida de las relaciones sociales.

Después de haber hecho un repaso por los diferentes capítulos es preciso dedicar esta parte de la conclusión a la reflexión. En una época donde los estudiantes no conocen lo que es la cultura del esfuerzo, este tipo de herramientas podría poner en serio peligro el devenir de la vida tal y como la conocemos. ¿Qué tipo de futuro puede concebirse de una generación de jóvenes dependientes de las TIC, que a causa de ello no han sido capaces de desarrollar un pensamiento crítico y que tampoco han desarrollado otras habilidades porque ya tenían sistemas de inteligencia artificial que lo hacía por ellos? O dicho de otra manera, ¿qué tipo de enseñanza puede originarse en un entorno caracterizado por el auge de los *chatbots* y la IA?

En realidad ya se puede responder a esta última cuestión puesto que, a diferencia de otros avances que se han dado antes en la educación y no han surgido efectos, *ChatGPT* ya está ocasionando problemas en el ámbito del plagio académico. Como hemos podido comprobar, actualmente las herramientas de análisis y comparadores de texto que

poseemos no son capaces de identificar si el texto proviene del *chatbot* o no, de manera que habrá que diseñar aplicaciones dedicadas a esta cuestión con el fin de asegurarnos que todo trabajo esté realizado por la persona en cuestión. Así pues, podemos delimitar que un tipo de enseñanza que podría surgir a menos que pueda controlarse a tiempo, es una basada en la asunción de todo tipo de obras creadas por la inteligencia artificial como propias. Se perdería el espíritu crítico, la originalidad, el respeto a la enseñanza y el propio acto de aprender en sí, por lo que muchos trabajos deberían eliminarse o replantearse. De manera que podemos concluir que la hipótesis que se planteaba al inicio del trabajo se confirma, la IA presenta más riesgos que ventajas en la educación.

Hasta el momento, se coincidía al pensar que la educación necesitaba una renovación total para dejar atrás el sistema tradicional cuanto antes. No obstante, quizás estos cambios que se esperaban con tanto anhelo no supongan un cambio a mejor, sino todo lo contrario. En otras palabras, la educación necesita renovarse, pero los cambios que proceden de la IA se producen a pasos agigantados y sin reflexión sobre su sentido, por lo que el sistema educativo no está preparado para asumir estas modificaciones en un periodo tan breve, de manera que lo único que se consigue es empeorar todavía más la situación en este ámbito.

Hay quienes consideran, por un lado, que la mejor manera de hacer frente a esta situación es restringiendo e impidiendo su uso, pero eso de nada sirve. Su aplicación global es inminente por lo que bloquear su utilización en ciertos centros educativos no va a conseguir nada, tan solo retrasar lo inevitable. Y, por otro lado, los que piensan que se tendrá que volver a una evaluación de los contenidos más memorística y dejando de lado los dispositivos electrónicos para llevar a cabo una escritura a mano; no obstante, estas medidas lo único que lograrían es retroceder en el tiempo y prolongar una enseñanza desactualizada. De manera que, los esfuerzos han de centrarse en saber hacerle frente, la tarea reside en emplearlo de una manera correcta siguiendo una concepción ética que respalde las respuestas que se proporcionan. Logrando que se transforme la amenaza en un instrumento útil para la mejora educativa, es decir, consiguiendo que de verdad sea de ayuda tanto para el docente como el alumno y facilite algunas tareas (no la supresión de actividades), sobre todo, de carácter analítico (seguimiento del progreso del alumno, elaboración de algunas tareas particulares, entre otros).

En definitiva, lo único que podemos hacer para evitar que la IA acabe con enseñanzas clave, entre otros aspectos, es llevar a cabo políticas centradas en la creación de leyes, normativas y planes de actuación basados en la ética que establezcan cuestiones como el ámbito y límite de aplicación. De esta manera, la comunidad ética, los docentes y los legisladores tienen una ardua tarea por delante para intentar controlar esta situación.

Bibliografía

- Abeliuk, Andrés y Claudio Gutiérrez. 2021. “Historia y evolución de la inteligencia artificial”. *Bits de ciencia: Revista del Departamento de Ciencias de la Computación de la Universidad de Chile* 21: 14-21.
- AI for Good. 2023. About. AI for Good Global Summit. <https://aiforgood.itu.int/about-ai-for-good/>
- Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon y Iyad Rahwan. 2018. “The Moral Machine experiment”. *Nature* 563, 59-64. <https://doi.org/10.1038/s41586-018-0637-6>
- Banda Gamboa, Hugo A. 2014. *Inteligencia artificial principios y aplicaciones*. Quito: Escuela Politécnica Nacional.
- Barrios Tao, Hernando, Vianney Díaz Pérez y Yolanda Guerra. 2020. “Subjetividades e inteligencia artificial: desafíos para ‘lo humano’”. *VERITAS Revista de Filosofía y Teología* 47: 81-107.
- Botta Sampietro, Martin y Yoan Esteban Mora Pereira. 2016. “Alan Turing: un genio desconocido”. En *Actas del 6º Congreso Uruguayo de Educación Matemática*, 263-274. Montevideo: Sociedad de Educación Matemática Uruguay.
- Calvo, Patrici. 2020. “Gobierno algorítmico: sobre el neuroaprendizaje moral de las máquinas en la política y economía”. En *Neuroeducación Moral y Democracia*, 223-242. Granada: Editorial Comares.
- Carabante Muntada, José María. 2007. “Un análisis de las implicaciones ideológicas de la ciencia en la obra Jürgen Habermas”. *Anuario de Derechos Humanos. Nueva Época* 8: 11-48.

- Castrova, Edward. 2001. "Virtual worlds: a first-hand account of market and society on the cyberian frontier". *CESifo Working Papers* 618: 1-40.
<https://dx.doi.org/10.2139/ssrn.294828>
- Cela Conde, Camilo José y Francisco José Ayala. 2001. *Senderos de la evolución humana*. Madrid: Alianza.
- ChatGPT. 2023. ¿Toda la información que proporciona es veraz? Chat OpenAI
<https://chat.openai.com/>
- Checha García, Fernando. 2010. "El uso de metaversos en el mundo educativo: gestionando conocimiento de Second Life". *Revista de Docencia Universitaria* 8 (2): 147-159. <https://doi.org/10.4995/redu.2010.6200>
- Clementín, Fernando. 2023. Elon Musk: «La Inteligencia artificial tiene el potencial de destruir civilizaciones». Criptonoticias.
<https://www.criptonoticias.com/comunidad/elon-musk-inteligencia-artificial-potencial-destruir-civilizaciones/>
- Cortina, Adela. 1992. *Ética mínima. Introducción a la filosofía práctica*. Tecnos.
- _____. 2008. *La escuela de Fráncfort. Crítica y Utopía*. Síntesis.
- Díaz Arce, Dariel. 2023. "Inteligencia artificial vs Turnitin: implicaciones para el plagio académico". *Revista Cognosis* 8 (1): 15-26.
<https://doi.org/10.33936/cognosis.v8i1>
- Erard, Gabriel. 2023. Elon Musk advierte: la inteligencia artificial es uno de los mayores riesgos para el futuro de la civilización. Hipertextual.
<https://hipertextual.com/2023/02/elon-musk-inteligencia-artificial-riesgo-futuro-civilizacion#>
- Fava Fernández, Alessandro Iván. 2017. "Big data y Deep Learning para la detección de objetos en imágenes". TFG: Universitat Politècnica de Catalunya.

- Flores Vivar, Jesús Miguel y Francisco José García Peñalvo. 2022. “Reflexiones sobre la ética, potencialidades y retos de la Inteligencia Artificial en el marco de la Educación de Calidad (ODS4)”. *Comunicar* 74: 37-47.
- García Peña, Víctor René, Alex Bladimir Mora Marcillo, y Jhonny Antonio Ávila Ramírez. 2020. “La inteligencia artificial en la educación”. *Dominio de las Ciencias* 6 (3): 648-666. <http://dx.doi.org/10.23857/dc.v6i3.1421>
- Gil Rodríguez, Susana. 2023. La vida de Sam Altman, el trampolín que impulsa la IA. NEO 2. <https://www.neo2.com/la-vida-sam-altman-open-ai/>
- Giró Gràcia, Xavi y Juana Sancho-Gil. 2022. “La Inteligencia Artificial en la educación: Big data, cajas negras y solucionismo tecnológico”. *Revista Latinoamericana de Tecnología Educativa* 21 (1): 129-145.
<https://doi.org/10.25074/07195532.28.2355>
- Habermas, Jürgen. 1986. *Ciencia y técnica como ideología*. Tecnos.
- Herrera Reyes, Rayco. 2022. “Habermas, la filosofía de la tecnología y el siglo XXI”. *Revista Laguna* 50: 167-182. <https://doi.org/10.25145/j.laguna.2022.50.08>
- Husserl, Edmund. 1991. *La crisis de las ciencias europeas y la fenomenología trascendental*. Crítica.
- Infobae. 2017. La cruda advertencia de Stephen Hawking sobre el desarrollo de la inteligencia artificial. Infobae.
<https://www.infobae.com/america/mundo/2017/11/07/stephen-hawking-alerta-sobre-la-inteligencia-artificial-puede-ser-lo-mejor-o-lo-peor-que-le-suced-a-la-raza-humana/>
- Jara, Ignacio y Juan Manuel Ochoa. 2020. “Usos y efectos de la inteligencia artificial en educación”. *BID: Banco Interamericano de Desarrollo*: 1-25.
<http://dx.doi.org/10.18235/0002380>

- Kuhn, Thomas Samuel. 1971. *La estructura de las revoluciones científicas*. Fondo de Cultura Económica.
- Longman. 2023. Plagiarism. En *Longman Dictionary of Contemporary English*. <https://www.ldoceonline.com/es-LA/dictionary/plagiarism>
- López de Mántaras Badia, Ramón. 2015. “Algunas reflexiones sobre el presente y el futuro de la Inteligencia Artificial”. *Novática* 234: 97-101.
- López de Mántaras Badia, Ramon y Pedro Meseguer González. 2017. *Inteligencia artificial*. Madrid: Editorial CSIS y Catarata.
- Mahesh, Batta. 2020. “Machine Learning Algorithms – A review”. *International Journal of Science and Research (IJSR)*9 (1): 381-386.
- Morales Chan, Miguel Antonio. 2023. *Explorando el potencial de Chat GPT: una clasificación de Prompts efectivos para la enseñanza*. Guatemala: Galileo University.
- Moreno Padilla, Raúl Darío. 2019. “La llegada de la inteligencia artificial a la educación”. *RITI Journal* 7 (14): 260-270. <https://doi.org/10.36825/RITI.07.14.022>
- Ordoñez Valencia, Maylee Lisbeth, Nicomedes Leónidas Ordoñez-Zúñiga, Josselyn Catalina Mantilla-Ordóñez, Mariuxi Elizabeth Garcés Wila, Diana Margarita Vera Arroyo y Willian Janner Coronel Mendez. 2022. “Análisis de herramientas del metaverso y su impacto en contextos educativos”. *Sapienza: International Journal of Interdisciplinary Studies* 3 (2): 610-630. <https://doi.org/10.51798/sijis.v3i2.366>
- Ortega y Gasset, José. 2004. *Meditación de la técnica y otros ensayos sobre ciencia y filosofía*. Alianza.
- Popper, Karl. 1945. *La sociedad abierta y sus enemigos*. Titivillus. <https://proletarios.org/books/Popper-La-sociedad-abierta-y-sus-enemigos.pdf>

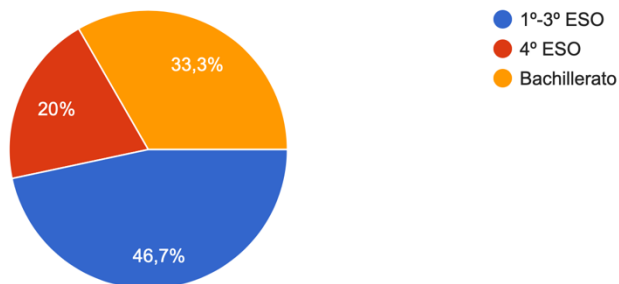
- Rudolph, Jürgen, Samson Tan y Shannon Tan. 2023. “ChatGPT: bullshit spewer or to end of traditional assessments In higher education?” *JALT: Journal of Applied Learning & Teaching* 6 (1): 1-22. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Sánchez Mendiola, Melchor. 2022. “El metaverso: ¿la puerta a una nueva era de educación digital?” *Investigación en Educación Médica* 11 (42): 5-8. <https://doi.org/10.22201/fm.20075057e.2022.42.22436>
- Seisdedos, Iker. 2023. Sam Altman (ChatGPT), en el capitolio: “Si la inteligencia artificial sale mal, puede salir muy mal”. El País. <https://elpais.com/tecnologia/2023-05-16/sam-altman-chatgpt-en-el-capitolio-si-la-inteligencia-artificial-sale-mal-puede-salir-muy-mal.html>
- Sox, Harold C. 2012. “Plagiarism in the digital age”. *Office of Research Integrity Newletter* 20 (3): 1-6.
- The European Comission. 2017. *The European Code of Conduct for Research Integrity*. Berlin: ALLEA: All European Academies.
- UNESCO. 2019. Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. UNESCO Education Sector.
- Vázquez Márquez, Israel. 2011. “Metaversos y educación: Second Life como plataforma educativa”. *Revista ICONO* 14, 9 (2): 151-166. <https://doi.org/10.7195/ri14.v9i2.30>
- Vázquez, Miriam. 2023. Sam Altman, CEO detrás de ChatGPT: “Mi peor temor es que causemos un daño significativo al mundo”. Antena 3 Noticias. https://www.antena3.com/noticias/mundo/sam-altman-ceo-openai-senado-eeuu-peor-temor-que-causemos-dano-significativo-mundo_2023051764649cb4ea319400018044f8.html

Anexos

Anexo 1. Resultados de la encuesta

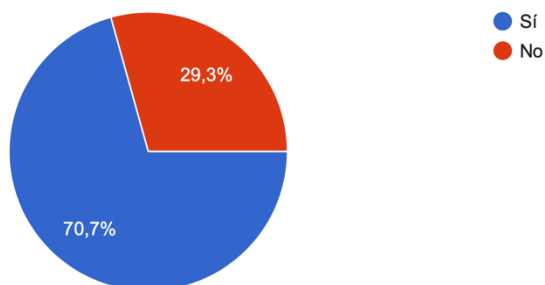
3. Indique el curso en el que imparte clase mayoritariamente

75 respuestas



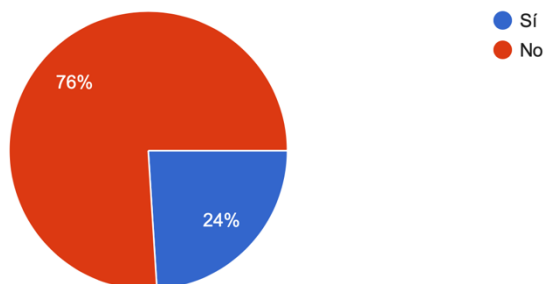
4. ¿Conoce ChatGPT-3?

75 respuestas



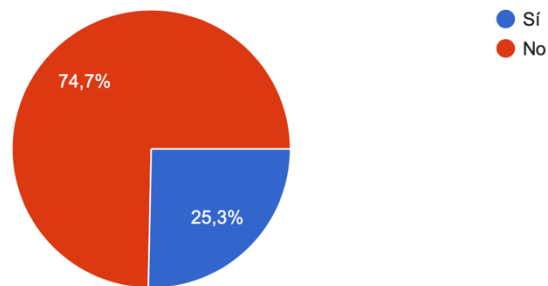
5. ¿Ha detectado en los/las estudiantes el uso de otras inteligencias artificiales como Photomath, Sintaxis.org o Paraphraz.it?

75 respuestas



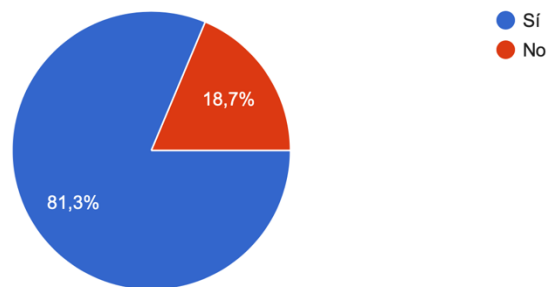
6. ¿Ha utilizado alguna vez una inteligencia artificial para crear materiales educativos (textos, problemas, dilemas, etc.)?

75 respuestas



7. ¿Cree que el uso de IA por parte de los estudiantes supone un riesgo para la educación?

75 respuestas



Anexo 2. Respuestas de *ChatGPT*

RE ¿toda la información que proporciona es veraz?



Como modelo de lenguaje, me esfuerzo por proporcionar respuestas precisas y útiles en función de los datos con los que fui entrenado, que abarcan hasta septiembre de 2021. Sin embargo, ten en cuenta que no puedo garantizar la veracidad absoluta de toda la información que proporciono. Aunque me esfuerzo por ser preciso, es posible que haya ocasiones en las que mi respuesta no refleje los desarrollos más recientes o que la información haya cambiado desde mi último entrenamiento. Por lo tanto, siempre es una buena idea verificar la información por tu cuenta o consultar fuentes confiables adicionales para obtener la información más actualizada y precisa.

