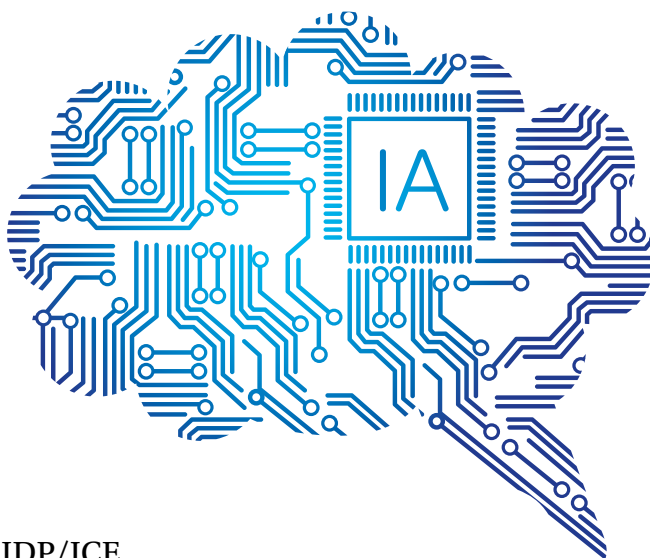


Mireia Ribera, Oliver Díaz Montesdeoca  
(coords.)

# ChatGPT y educación universitaria:

posibilidades y límites  
de ChatGPT como  
herramienta docente





## CONSEJO EDITORIAL IDP/ICE, UB-OCTAEDRO

### **Dirección**

Teresa Pagès Costas (jefa de la Sección Universidad, IDP/ICE, Facultad de Biología, Universidad de Barcelona)

### **Coordinadora**

Anna Forés Miravalles (IDP/ICE, Facultad de Educación, Universidad de Barcelona)

### **Editor**

Juan León Varón (director de la Editorial Octaedro)

### **Consejo Editorial**

Pedro Allueva Torres (Facultad de Educación, Universidad de Zaragoza)

Pilar Ciruelo Rando (Editorial Octaedro)

Mar Cruz Piñol (Facultad de Filología, Universidad de Barcelona)

Carmen Ferrándiz García (Facultad de Psicología, Universidad de Murcia)

Mercè Gracenea Zugarramurdi (Facultad de Farmacia y Ciencias Alimentación, Universidad de Barcelona)

Virginia Larraz Rada (Facultad de Ciencias de la Educación, Universidad de Andorra)

Miquel Martínez Martín (Facultad de Educación, Universidad de Barcelona)

Miquel Oliver Trobat (Facultad de Educación, Universidad de las Islas Baleares)

Joan Carles Ondategui Parra (Facultad de Óptica y Optometría, Universidad Politécnica de Cataluña)

Jordi Ortín Rull (Facultad de Física, Universidad de Barcelona)

Mireia Ribera Turró (Facultad de Matemáticas e Informática, Universidad de Barcelona)

Alicia Rodríguez Álvarez (Facultad de Filología, Universidad de Las Palmas de Gran Canaria)

Antoni Sans Martín (Facultad de Educación, Universidad de Barcelona)

Carmen Saurina Canals (Facultad de Ciencias Económicas y Empresariales, Universidad de Girona)

### **Secretaría Técnica del Consejo Editorial**

Lourdes Marzo Ruiz (IDP/ICE, Universidad de Barcelona), Ana Suárez Albo (Editorial Octaedro)

### **Normas presentación originales:**

[https://www.ub.edu/idp/web/sites/default/files/docs/Normas\\_presenta.pdf](https://www.ub.edu/idp/web/sites/default/files/docs/Normas_presenta.pdf)

### **Revisores:**

[https://www.ub.edu/idp/web/sites/default/files/docs//Revisores\\_Octaedro.pdf](https://www.ub.edu/idp/web/sites/default/files/docs//Revisores_Octaedro.pdf)

### **Criterios de calidad:**

<https://www.ub.edu/idp/web/sites/default/files/docs/criterios.pdf>

**Mireia Ribera**  
**Oliver Díaz Montesdeoca**  
**(coords.)**

# **ChatGPT y educación universitaria**

**Posibilidades y límites de ChatGPT  
como herramienta docente**

**OCTAEDRO - IDP/ICE, UB**

Colección Educación universitaria

Título: *ChatGPT y educación universitaria: posibilidades y límites de ChatGPT como herramienta docente*

Primera edición: enero de 2024

© Mireia Ribera, Oliver Díaz Montesdeoca (coords.)

© De esta edición:

Ediciones Octaedro, S.L.  
Bailén, 5 - 08010 Barcelona  
Tel.: 93 246 40 02  
octaedro@octaedro.com  
www.octaedro.com

Universitat de Barcelona  
Institut de Desenvolupament Professional (IDP/ICE)  
Campus Mundet - 08035 Barcelona  
Tel.: 93 403 51 75  
ice@ub.edu



Esta publicación está sujeta a la Licencia Internacional Pública de Atribución/Reconocimiento-NoComercial 4.0 de Creative Commons. Puede consultar las condiciones de esta licencia si accede a: <https://creativecommons.org/licenses/by-nc/4.0/>

ISBN: 978-84-10054-01-1

Diseño y producción: Servicios Gráficos Octaedro

# SUMARIO

1. Introducción .....	7
— OLIVER DÍAZ	
— MIREIA RIBERA	
2. Indicaciones prácticas para usar ChatGPT .....	21
— ELOI PUERTAS	
3. La tecnología tras ChatGPT .....	51
— DANIEL ORTIZ	
— DAVID BUCHACA	
4. Inteligencia artificial, desinformación y aspectos éticos .....	87
— JAVIER GUALLAR	
— CARLOS LOPEZOSA	
5. Implementación de ChatGPT en el aula .....	97
— MARIONA GRANÉ	
6. ChatGPT, implicaciones para la sociedad y la educación .....	123
— LOS AUTORES	
Autoría .....	127



# 1. INTRODUCCIÓN

— Oliver Díaz  
— Mireia Ribera

Al abrir este libro nos podemos preguntar: ¿por qué otro libro de ChatGPT? Su motivación principal es el deseo de llegar a nuestro público, a nuestros lectores fieles. Octaedro es una editorial de referencia en el entorno universitario, cubre los ámbitos de la innovación docente en muchas disciplinas y se ha convertido en editorial de referencia para el profesorado universitario y de secundaria para innovar y mejorar su docencia. Un libro sobre los avances de la inteligencia artificial (IA) en el ámbito académico no podía faltar.

El libro trata principalmente de ChatGPT porque esta es la herramienta que está en todos los foros, la que ha despertado más debate en los últimos meses. ChatGPT es la primera herramienta de IA que ha cruzado la frontera entre herramientas usadas por especialistas y herramientas usadas por el gran público, llegando a ser la aplicación con mayor crecimiento de la historia (más de un millón de usuarios en cinco días). Pero en realidad, en este texto ChatGPT es una excusa, pues tras ella viene una nueva generación de herramientas de IA generativa que cambia la manera que conocíamos hasta ahora de trabajar, enseñar y aprender. Así, en esta obra se muestra la evolución de los modelos de lenguaje y se debate sobre la autenticidad de la información y sobre la transformación en la educación, temas que trascienden la herramienta en sí.

## 1.1. ¿Por qué ahora?

Según el ciclo de tendencias de Gartner (Gartner, 2022), las tecnologías pasan por diversas fases a lo largo del tiempo que van desde la aparición de la innovación hasta el pico de las expectativas desbordadas, el valle de la desilusión, la rampa de la iluminación y la meseta de la productividad. La IA generativa, que es aquella que tiene como finalidad producir datos sintéticos como texto o imágenes a partir de patrones aprendidos en datos reales, y dentro de esta, ChatGPT, se encontraría ahora con una gran expectativa por parte de la población, entre la innovación y el pico de expectativas desbordadas, tal y como se ve en la figura 1.1.

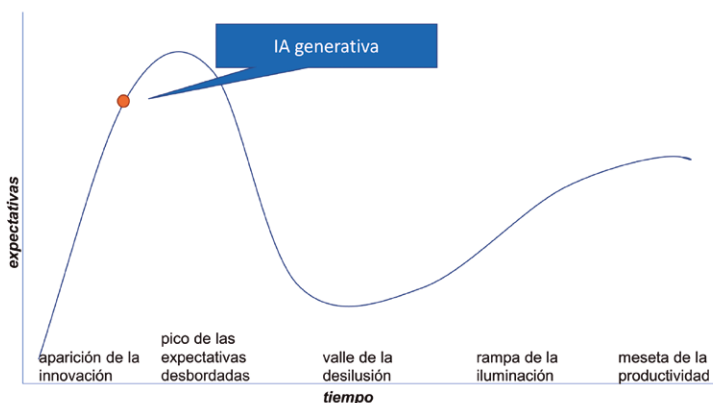


Figura 1.1. Adaptación propia del Hype Cycle for Artificial Intelligence (Gartner, 2022).

Y es en esta fase donde se genera más ruido, tanto a favor de las innovaciones como en su contra. En un entorno sensible como la educación, una postura negacionista o una adopción entusiasta podría tener efectos muy perjudiciales; por ello es importante el trabajo de dilucidar y aclarar conceptos, así como comprender sus limitaciones. Como ya avanza la figura 1.1, tardaremos unos años a aprovechar estas herramientas con plena productividad, pero podemos recorrer este camino con un poco más de calma y conocimiento para conseguir una implementación más satisfactoria de esta tecnología.



## 1.2. ¿Qué es ChatGPT?

ChatGPT es un tipo de IA generativa desarrollado por la empresa OpenAI que utiliza una red neuronal artificial llamada *generative pre-trained transformer* (GPT) para generar texto de forma coherente y relevante en conversaciones de tipo chat con seres humanos. Esta herramienta se puede considerar como un modelo probabilista, que proporciona información plausible de acuerdo con la gran cantidad de datos de internet que ha utilizado para entrenarse. Su éxito radica en la potente infraestructura (granjas de servidores) que hay detrás – financiada por Microsoft –, así como en los desarrollos tecnológicos de los últimos años en el campo del procesamiento del lenguaje natural (*natural language processing*, NLP), juntamente con el acceso a millones de datos en internet. Sin embargo, a pesar de sus grandes logros y expectativas recientes, hemos de tener en cuenta que esta herramienta tiene una capacidad limitada.

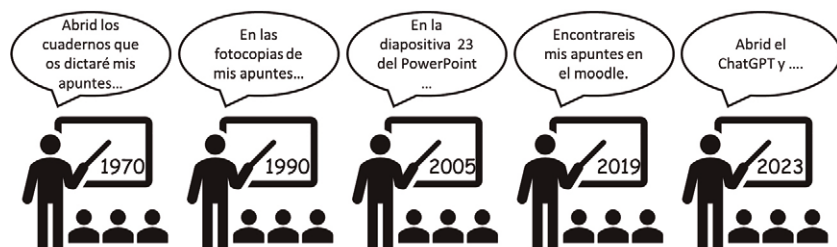
## 1.3. Historia de los *chatbots* y su evolución en la educación

### 1.3.1. El uso de la tecnología en la educación. Avances tecnológicos que han contribuido a su desarrollo

A lo largo de la historia, infinidad de descubrimientos científicos y avances tecnológicos importantes han marcado la evolución del ser humano desde sus inicios hasta la actualidad. Esa evolución tecnológica también ha tenido influencia en el ámbito de la docencia y conocimiento. Por ejemplo, en la Prehistoria, las pinturas rupestres permitían transmitir el arte de cazar de una generación a otra. Esta técnica de divulgación de información ha ido evolucionando a lo largo de los siglos. En épocas más recientes, la producción industrial tanto del papel (sucesor del papiro y el pergamino) como del lápiz marcaron un gran cambio que permitieron a la humanidad compartir ideas y pensamientos en masa de una manera más fácil y sencilla. En el ámbito de la docencia, esto permitió un aumento en la trans-

ferencia del conocimiento entre alumnado y profesorado, que hasta entonces se comunicaban mayoritariamente de forma oral. A partir de aquí, otros avances tecnológicos, como la máquina de escribir, la imprenta, la radio, el ordenador personal o el acceso a internet, han ido derribando barreras, permitiendo poner las bases del aprendizaje a distancia, brindando al alumnado la oportunidad de aprender a su propio ritmo, como los cursos en línea masivos y abiertos (*massive open online courses*, o MOOC). En los últimos años estamos viendo que la transformación digital y las tecnologías basadas en IA están encontrando su hueco en el aula, y se han usado como herramienta docente, como, por ejemplo, la realidad aumentada, la realidad virtual, las redes sociales, las aplicaciones móviles o los *chatbots*, entre otros. Fuera del aula, también encontramos soluciones tecnológicas interesantes de apoyo, como la analítica de aprendizaje (*learning analytics*), que recopila, procesa y analiza datos sobre el alumnado con el fin de comprender y optimizar su aprendizaje.

A medida que evolucionan estas tecnologías educativas, los docentes deben adaptar su forma de enseñanza. El docente hoy en día utiliza menos clases magistrales y trabaja con la tecnología como herramienta de soporte a la docencia para así mejorar la experiencia de aprendizaje del alumnado. El uso de la tecnología para la mejora docente pasa por la búsqueda de alternativas que incrementen la motivación del alumnado e involucrarlos de manera más activa en su proceso de aprendizaje (figura 1.2).



**Figura 1.2.** Ejemplo de evolución tecnológica en el aula. Inspirada en una ilustración de Sergio Duce (@yo\_runner).

### 1.3.2. ¿Qué es un *chatbot*?

Un *chatbot* (o *bot* conversacional) es un programa informático diseñado para simular una conversación en lenguaje natural con seres humanos a través de un canal de texto, voz o incluso imágenes. Esta tecnología utiliza técnicas de IA, como el procesamiento del lenguaje natural y el aprendizaje automático (*machine learning*), para procesar y comprender las preguntas y solicitudes del usuario y proporcionar respuestas relevantes y coherentes. Es decir, los *chatbots* recogen una pregunta, procesan la información y responden en milisegundos.

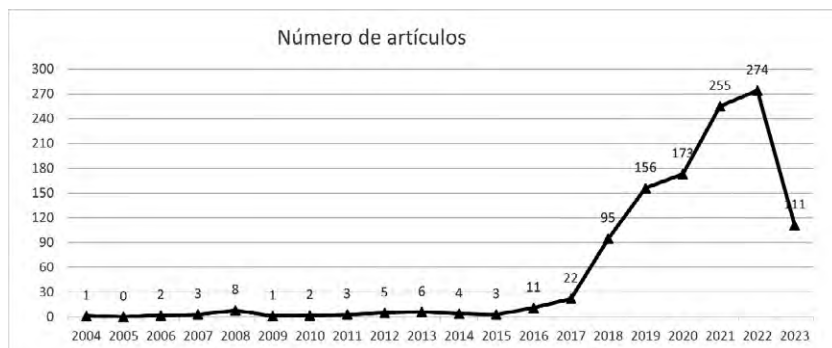
Pueden ser implementados en diversas plataformas, como aplicaciones móviles, sitios web, redes sociales o sistemas de mensajería instantánea. Dependiendo de su nivel de sofisticación, los *chatbots* pueden seguir un conjunto de reglas predefinidas, trabajar según unos patrones de conversación preestablecidos o incluso aprender de interacciones previas para mejorar su capacidad de respuesta. Los más complejos, como ChatGPT, utilizan tecnología de IA con gran cantidad de datos para procesar los mensajes entrantes.

Los *chatbots* pueden ser utilizados para desempeñar diferentes funciones, desde proveer información y asistencia a un usuario hasta realizar tareas más específicas, como podría ser en entornos de educación donde se pretende alcanzar ciertos objetivos educativos y pedagógicos.

### 1.3.3. El uso de *chatbots* en entornos educativos

Los *chatbots* han experimentado una evolución significativa en el entorno educativo durante los últimos años, en términos cuantitativos (la figura 1.3 muestra el número de artículos sobre *chatbots* en educación), de acuerdo con los criterios de búsqueda definidos por Wollny *et al.* (2021) y extraídos del sitio Web of Science, plataforma de acceso a múltiples bases de datos de referencia y citas de revistas académicas, actas de congresos y otros documentos en diversas disciplinas académicas. A nivel cualitativo, los *chatbots* han pasado de ser simples herramientas de asistencia básica a convertirse en soluciones sofisticadas que brindan apoyo personalizado e instantáneo al alumnado en temas

como criterios de evaluación, recordatorios de entrega de tareas o recomendación de lecturas entre otros (Cunningham-Nelson, 2019).



**Figura 1.3.** Número de artículos por año de publicaciones relacionadas con *chatbots* en entornos educativos. El año 2023 solo tiene en cuenta artículos de enero a junio. Fuente: Web of Science.

Los *chatbots* han creado oportunidades únicas en el aprendizaje, donde se involucra a los diferentes actores que participan en todo el ciclo del aprendizaje, como alumnado y profesorado, e incluso personal administrativo de secretarías y otros departamentos que dan servicio al alumnado. Un ejemplo claro podría ser el uso de un *chatbot* integrado en el campus virtual de una universidad donde cualquier miembro del colectivo universitario pueda preguntar sobre los criterios de evaluación de una asignatura, horario de atención al público de la secretaría o la normativa de permanencia en la carrera.

Esta tecnología ayuda al desarrollo y soporte de (nuevas) habilidades, automatización de tareas que permiten una educación más eficiente, aumento de la motivación del alumnado o proveer de repuestas a infinidad de preguntas instantáneamente, en cualquier momento del día, sin interactuar directamente con correos electrónicos o visitas presenciales a los otros actores implicados en el aprendizaje (Wollny *et al.*, 2021).

Los objetivos pedagógicos más comunes que encontramos en los *chatbots* incluyen: (1) el soporte durante las actividades de aprendizaje (p. ej., adquisición de conocimiento), (2) asistencia en el día a día durante la vida estudiantil (p. ej., consulta de horarios o procesos administrativos) o (3) funciones de tutoría que apoya el desarrollo personal de los estudiantes (p. ej., soporte al aprendizaje autorregulado) (Wollny *et al.*, 2021)

Esta tecnología educativa tiene grandes beneficios en el ámbito de la educación. Por ejemplo, en la revisión sistemática del estado del arte realizado por Okonkwo y Ade-Ibijola (2021) se destaca el soporte que dan al usuario (profesorado, alumnado) para subir la información necesaria a las plataformas de aprendizaje (p. ej. Moodle), así como para distribuirla a los usuarios autorizados de forma instantánea y fácil. Además, los *chatbots* fomentan el aprendizaje personalizado, brindan apoyo a los usuarios y permiten que múltiples usuarios accedan a la misma información al mismo tiempo.

Pero quizás lo más interesante es conocer si esta tecnología ayuda a mejorar los resultados de aprendizaje del alumnado. El metaanálisis de Wu y Yu (2023) indica que los *chatbots* basados en IA podrían tener un gran impacto —o moderado— en los resultados de aprendizaje de los estudiantes en términos de rendimiento, motivación, interés, autoeficacia o valor percibido del aprendizaje, sobre todo en el alumnado universitario. Sin embargo, este impacto se ve disminuido (o incluso desaparece) cuando se utiliza con alumnado de primaria y secundaria. A pesar de todo ello, más investigación es necesaria para validar estos resultados.

En la tabla 1.1 se indican algunos ejemplos de *chatbots* utilizados en entornos universitarios que dan soporte al alumnado en diferentes aspectos (Fernández-Ferrer, 2023).

**Tabla 1.1.** *Chatbots* usados en educación

Chatbot	Enlace
Pounce	Estudio de caso en Georgia State University (Estados Unidos), en la que se ofreció a los estudiantes recientemente matriculaos un chat personalizado enviándoles recordatorios, fechas de matrícula, recogiendo datos para cuestionarios. . .
Deakin's Genie	Estudio de caso en Deakin University (Melbourne, Australia), en el que un <i>chatbot</i> funciona como asistente digital para los estudiantes para ayudarles a gestionar su calendario, tareas y otros aspectos de la vida universitaria.
Ash	Un <i>chatbot</i> diseñado como soporte para la salud mental en Monash University (Australia).
Quizbot	Un <i>chatbot</i> para recordar información factual en Stanford (Estados Unidos).
Isidra	Una asistente virtual en la Universidad de Alcalá (España), que orienta y resuelve preguntas frecuentes del alumnado.
Aina	La misma tecnología que Isidra (ver fila anterior) con la información de la Universitat de les Illes Balears.
Lola	La misma tecnología que Isidra (ver dos filas anteriores) con la información de la Universidad de Murcia.

Fuente: Fernández-Ferrer, 2023.

## 1.4. Importancia de la inteligencia artificial en entornos universitarios, aplicaciones en docencia, investigación y gestión

La Universidad de Al Qarawiyyine (Fez, Marruecos) es considerada por la Unesco como la primera universidad del mundo (año 859 d.C.). Ahora sumergidos ya en pleno siglo XXI, los avances de la ciencia y la tecnología han cambiado muchos aspectos de la vida, y la universidad no iba a ser menos.

La IA (generativa) está cambiando –y cambiará mucho más– nuestro día a día, nuestra percepción, nuestro pensamiento y, por supuesto, los modelos de enseñanza-aprendizaje. Tecnologías disruptivas como el ChatGPT tienen gran potencial de cambiar muchos de los procesos actuales dentro del entorno universitario que se ejemplifican a continuación, aunque este volumen se centra en su aplicación en el ámbito docente.

En la tabla 1.2 se muestran algunos posibles ejemplos de uso de IA (incluyendo la IA generativa) en ámbitos de docencia (D) –los abordados por este texto–, así como de investigación (I) y gestión (G). Los diferentes modelos de IA pueden estar programados para mantener conversaciones, generar datos sintéticos u otras tareas específicas.

**Tabla 1.2.** Ejemplos de usos de IA en entornos docentes, de investigación y gestión universitaria en los ámbitos: D, docencia; I, investigación; G, gestión

Ámbito	Aplicación	Descripción
D	Recomendación de contenido	Una IA específica puede analizar perfiles de usuarios y actividades para recomendar material relevante o cursos complementarios.
D	Tutoría virtual	Una IA conversacional puede dar apoyo al alumnado respondiendo a preguntas frecuentes y proporcionando explicaciones claras a dudas recurrentes.
D	Soporte al profesorado	Una IA generativa puede dar ideas al profesorado sobre actividades docentes a realizar dentro del aula, en relación con una materia específica.
I	Análisis de datos	Una IA específica puede ayudar en la investigación científica y el análisis de datos a gran escala.
I	Descubrimiento de conocimiento	Los diferentes ámbitos de la IA (minería de datos, <i>machine learning</i> ) pueden ayudar a identificar patrones emergentes y generar ideas para nuevas áreas de investigación.
I	Asistencia en la diseminación científica	Herramientas basadas en IA generativa pueden permitir analizar más eficientemente el estado del arte y dar soporte en la creación de nuevos documentos científicos.
G	Procesamiento automatizado de documentos	Una IA específica puede ser utilizada para automatizar la revisión y clasificación de documentos administrativos, ahorrando tiempo y reduciendo errores humanos.
G	Sistema de planificación y programación	Algoritmos de IA específicos pueden ser utilizados para programar horarios, asignación de aulas y recursos, teniendo en cuenta profesorado, alumnado y restricciones institucionales.
G	Servicios de atención al alumnado	Asistentes conversacionales basados en IA pueden proporcionar respuestas rápidas a preguntas administrativas comunes, ayudar en la inscripción y orientar a los estudiantes en los procesos de admisión y matrícula.

## 1.5. Retos y oportunidades

Quizá nos preguntemos por qué los docentes debemos conocer ChatGPT. La respuesta para los autores está clara: porque la IA va a ser una realidad en el futuro de nuestro alumnado, una de las habilidades tecnológicas que deberán dominar y porque como posibilidad y peligro debemos conocerla.

La mayor capacidad de ChatGPT es el procesamiento de la lengua, la generación de textos. La herramienta no es científicamente rigurosa ni en la selección de fuentes para la recogida de datos ni mucho menos en su validación (ver capítulo 3), pero sus resultados pueden llevar a confusión, pues crea una argumentación tan bien redactada que parece confiable. Su base de aprendizaje es internet y las respuestas de ChatGPT pueden reproducir sesgos, informaciones tendencias o incluso injusticias existentes en la red (véanse los capítulos 2 y 4).

Una buena síntesis de cuándo y cómo usar ChatGPT la ofrece el árbol de decisión publicado por Sabzalieva y Valentini (2023) (figura 1.4).

Teniendo en cuenta el carácter de esta herramienta, nuestra experiencia y la mayor parte de las fuentes consultadas recomiendan, en cuanto a posibilidades y puntos fuertes, su rol como asistente, como inspiración, como corrector lingüístico o para transformar textos, para iterar y refinar ideas, y para tareas repetitivas. Tareas todas en las que podemos supervisar el trabajo realizado.<sup>1</sup>

Su capacidad de razonar lingüísticamente debe ponerse en equilibrio con su falta de conocimiento real del mundo; ChatGPT puede llegar a presentar hechos falsos o inventados, y tomar al pie de la letra sus recomendaciones sin visión crítica puede incluso ser peligroso.

Otra característica relevante de ChatGPT es que es una herramienta de dominio genérica, es decir, no está especializada en una disciplina específica, y tiene una base de conocimientos generales. Por ello, si trabajamos en un ámbito muy especializado, como podría ser, por ejemplo, medicina o arquitectura, quizá tendrá sentido buscar alternativas o complementar con herramientas de ese dominio si existen.

1. En Sabzalieva y Valentini (2023) se ofrece un listado de roles muy interesante para educación superior.



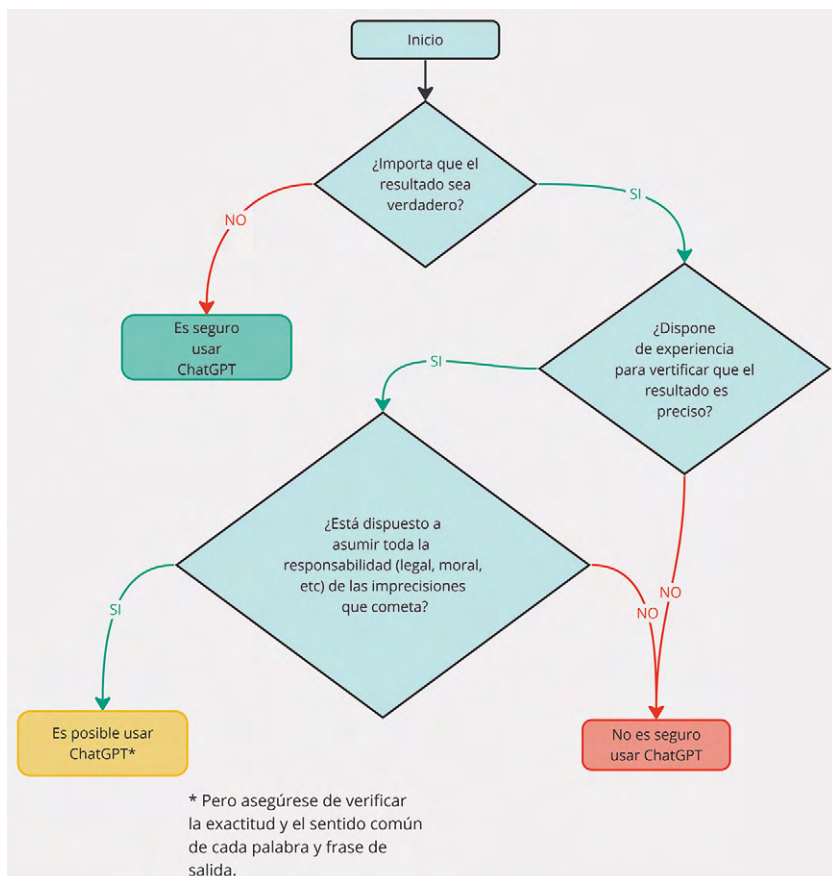


Figura 1.4. Árbol de decisión sobre cuándo usar ChatGPT (Sabzalieva y Valentini, 2023).

Por otra parte, OpenAI es una empresa comercial y ofrece esta herramienta a nivel experimental, y cuando aceptamos usarla estamos aceptando compartir la información que le facilitamos como base de entrenamiento de la herramienta. Por este motivo es prudente no compartir información privada o que pueda afectar a la seguridad de nuestros sistemas informáticos en los diálogos con ChatGPT.

Conociendo sus límites y usando su potencial, esta herramienta supone un cambio fundamental en los procesos de comunicación y, como consecuencia, de la creación de conocimiento, con un efecto

disruptor en la educación. Los efectos de ChatGPT en la educación van a ser muy impactantes. De modo similar a como la irrupción de enciclopedias en línea cuestionaron la importancia de memorizar muchos datos, la irrupción de ChatGPT, de nuevo, plantea la necesidad de repensar dónde está el valor del aprendizaje. Quizá con la IA generativa deje de tener sentido crear contenidos de forma repetitiva y deberemos centrarnos sobre todo en la creatividad y en la capacidad de especificar qué tipo de texto, imagen o vídeo necesitamos en cada ocasión. También parece imprescindible desarrollar el pensamiento crítico.

Este libro pretende ayudar a aumentar el conocimiento de esta herramienta para entender mejor cómo beneficiarnos de sus servicios, a la vez que evitamos sus peligros en el entorno académico, especialmente en las aplicaciones docentes que tanto pueden ser útiles para el profesorado como para el alumnado. Lo que parece una realidad es que la IA ha llegado para quedarse.

## 1.6. Estructura del libro

En este libro abordamos especialmente ChatGPT y mencionamos herramientas afines desde diversas perspectivas para ofrecer una visión completa. Su objetivo, como se ha dicho, es darlo a conocer, entenderlo y ofrecer claves a los lectores para su máximo aprovechamiento.

En primer lugar, el capítulo «Indicaciones prácticas para usar ChatGPT» es un capítulo práctico, aplicado, con recomendaciones simples para familiarizarnos con la herramienta y tecnologías afines, y saberle sacar provecho. Para «jugar» con ella. Es difícil hablar al lector de una tecnología si primero no la ha probado.

Una vez lo hemos vivido, el capítulo «La tecnología tras ChatGPT» explica los fundamentos técnicos y pone en contexto esta herramienta. La enmarca en otros avances realizados en IA y visiona el futuro posible.

En el capítulo «Desinformación y aspectos éticos» tratamos los aspectos más controvertidos de la tecnología, cómo usarla con prudencia y cuáles son las estrategias para minimizar posibles consecuencias indeseadas. Nos proporciona estrategias para cuestionar y también validar la información que recibimos y ser más críticos con ella.

Finalmente, el capítulo «La implementación de ChatGPT en el aula» pretende compartir la reflexión y consecuencias de su inclusión en los procesos de docencia-aprendizaje para repensar nuestra metodología y evaluación, y seguir aprendiendo y avanzando.

Como conclusión los diversos puntos de vista se fusionan en una breve reflexión final en el capítulo 6, que recoge las implicaciones de esta herramienta en la educación.

## Bibliografía

- Cunningham-Nelson, S., Boles, W., Trouton, L. y Margerison, E. (2019). A review of chatbots in education: practical steps forward. *30th Annual Conference for the Australasian Association for Engineering Education (AAEE 2019): Educators Becoming Agents of Change: Innovate, Integrate, Motivate* (pp. 299-306). <https://doi.org/https://search.informit.org/doi/10.3316/INFORMIT.068364390172788>
- Fernández-Ferrer, M. (2023). *Chatbots en educación: tendencias actuales y desafíos futuros*. Learning, Media & Social Interactions. Universitat de Barcelona. [https://www.lmi-cat.net/sites/default/files/Chatbots\\_en\\_Educacion.pdf](https://www.lmi-cat.net/sites/default/files/Chatbots_en_Educacion.pdf)
- Gartner (2022). *Hype cycle for artificial intelligence, 2022*. <https://emtemp.gcom.cloud/ngw/globalassets/en/articles/images/hype-cycle-for-artificial-intelligence-2022.png>
- Okonkwo, C. W. y Ade-Ibijola, A. (2021). Chatbots applications in education: a systematic review. *Computers and Education: Artificial Intelligence*, 2 (100033). <https://doi.org/https://doi.org/10.1016/j.caeai.2021.100033>

- Sabzalieva, E. y Valentini, A. (2023). *ChatGPT e inteligencia artificial en la educación superior: guía de inicio rápido*. [https://unesdoc.unesco.org/ark:/48223/pf0000385146\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000385146_spa)
- Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M. y Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4 (654924). <https://doi.org/https://doi.org/10.3389/frai.2021.654924>
- Wu, R. y Yu, Z. (2023). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*. <https://doi.org/https://doi.org/10.1111/bjet.13334>

## 2. INDICACIONES PRÁCTICAS PARA USAR CHATGPT

— Eloi Puertas

ChatGPT se ha erigido como el primer *chatbot* que utiliza un LLM (modelo grande de lenguaje, o *large language model*) al salir a la luz pública. Desde las primeras versiones lanzadas al gran público ha demostrado una notable habilidad en la interacción con humanos en lenguaje natural. Esta capacidad no solo ha mejorado las aplicaciones clásicas de chat, como asistentes interactivos o ayudantes en tareas, sino también la generación de texto en diversos contextos.

La función básica de ChatGPT es la de una aplicación capaz de interactuar a través del lenguaje natural con un agente inteligente. Entre sus cualidades está la capacidad de responder en lenguaje natural de la manera más precisa posible dentro de sus conocimientos y limitaciones; por tanto, puede generar respuestas a diversas peticiones presentadas como texto de entrada y comprender y escribir en más de 50 idiomas, aunque los resultados pueden variar según el idioma utilizado (Bang, 2023). Además de los idiomas humanos, también puede generar y comprender código fuente en varios lenguajes de programación, como JavaScript, Python o Java, entre otros. En el ámbito de la programación, el éxito de los resultados obtenidos a menudo depende de cómo se realiza la solicitud, ya que el código resultante no siempre es funcional ni produce los resultados esperados, como se discutirá más adelante.

La petición que realiza el usuario al *chatbot* se conoce comúnmente en inglés como *prompt*. En castellano, se puede traducir como entra-

da o petición. En los primeros días de los ordenadores personales el *prompt* se refería a la línea de comandos donde el usuario ingresaba órdenes o comandos para que el sistema operativo las ejecutará. Por ejemplo, en MSDOS, el *prompt* aparecía después del símbolo >; este solía estar precedido por la unidad de disco duro, generalmente la unidad C, y se indicaba mediante un cursor parpadeante.

C:\>

Como veremos a continuación, el uso del *prompt* de ChatGPT es muy similar. Sin embargo, la principal diferencia radica en que el *chatbot* no solo reacciona a un conjunto de comandos limitado, sino que es capaz de comprender el lenguaje natural. En lugar de tener que utilizar comandos específicos, los usuarios pueden plantear preguntas o expresar sus solicitudes de una manera más natural, como si estuvieran interactuando con otra persona. Esto permite una mayor flexibilidad y una experiencia más intuitiva al utilizar el *chatbot*.

## 2.1. Cómo usar ChatGPT

ChatGPT se utiliza comúnmente como una aplicación web, que requiere crear una cuenta antes de poder acceder a ella.<sup>2</sup>

Una vez registrado en la plataforma de ChatGPT, se accede a la interfaz básica del *chatbot*, como se muestra en la figura 2.1. En la parte inferior de la interfaz, se encuentra la ventana destinada a introducir el *prompt* o la petición al *chatbot*. En este espacio el usuario puede escribir su consulta o solicitud en lenguaje natural.

En la parte izquierda de la interfaz se muestra un historial con las diferentes sesiones de conversación previas mantenidas con el *chatbot*. Este listado permite al usuario recuperar conversaciones antiguas y continuarlas en el punto donde se dejaron si así lo desea. En caso de que se desee borrar el historial de conversaciones, se puede hacer indi-

2. La plataforma se encuentra alojada en la siguiente dirección web: <https://chat.openai.com/>

vidualmente presionando el icono de papelera encima de cada conversación o de manera global desde el menú de preferencias. Este menú se encuentra representado por tres puntos y se ubica al lado del perfil del usuario en la interfaz (esquina inferior izquierda de la pantalla).

En la parte inferior, en letra pequeña se advierte sobre la posibilidad de que ChatGPT pueda proporcionar información incorrecta acerca de personas, lugares o hechos. Es importante tener en cuenta esta advertencia durante el uso de la aplicación. La razón por la cual el *chatbot* puede dar respuestas incorrectas se explica detalladamente en los siguientes capítulos y se debe a la naturaleza de los modelos de lenguaje.

Además, en la interfaz se muestra la versión de ChatGPT que se está utilizando. Esto incluye la fecha de la última actualización y un enlace a las notas de la versión, donde se pueden encontrar las últimas novedades y mejoras agregadas a la aplicación.

En la página principal se proporcionan advertencias importantes sobre las limitaciones del sistema: generación de información incorrecta (llamadas comúnmente «alucinaciones»), sesgo en el contenido, instrucciones dañinas y conocimiento limitado de eventos posteriores a 2021.

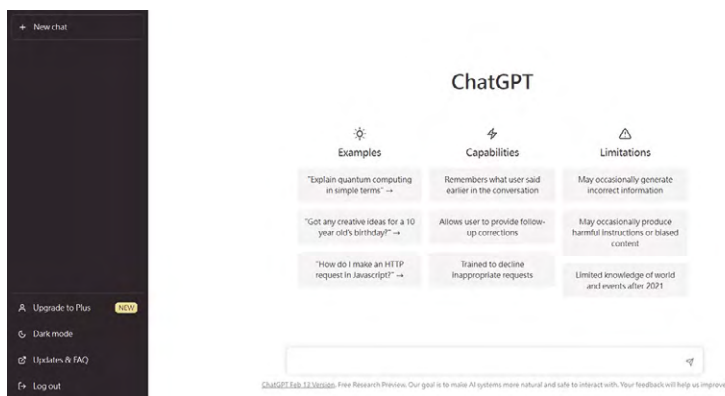


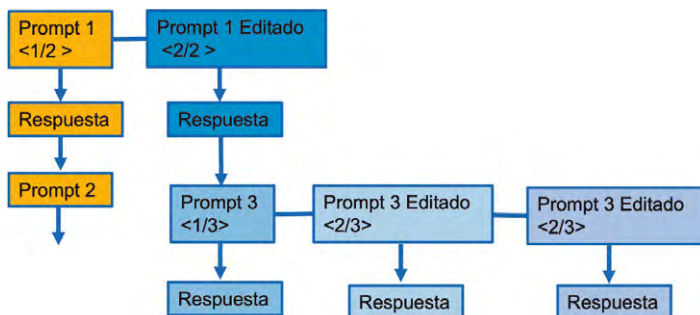
Figura 2.1. Interfaz básica de la aplicación web ChatGPT, extensión v.3.3.8.

Cada vez que se envía un nuevo *prompt* al ChatGPT, este se muestra en la parte central de la ventana junto con su correspondiente respuesta. Se puede copiar la respuesta utilizando el botón de portapapeles; también evaluar la respuesta utilizando los botones de pulgares hacia arriba y hacia abajo. Además, si se desea una respuesta diferente al *prompt* original, se puede utilizar el botón «Regenerar respuesta». En este caso, se preguntará si la nueva respuesta es mejor o no en comparación con la anterior. A través de estos sistemas de refuerzo positivo o negativo, ChatGPT puede mejorar sus resultados con la ayuda de los usuarios.

A medida que continúa la conversación, podéis introducir nuevos *prompts* y ChatGPT tomará en cuenta los *prompts* y conversaciones anteriores del chat en las interacciones siguientes. Por lo tanto, si le pedís que se dirija a vosotros en un idioma o tono específico, no será necesario repetirlo en los *prompts* siguientes. El modelo recordará las preferencias establecidas previamente.

Además, es posible editar un *prompt* que ya ha sido introducido. Para hacerlo, se debe hacer clic en el botón de lápiz ubicado junto al *prompt* en la ventana central. Al editar un *prompt*, no solo se modifica el resultado obtenido, sino también el curso de la conversación. Al comienzo del *prompt* editado, se muestra el historial de versiones en forma de  $< 2/2 >$ , donde el primer número indica la versión actual del *prompt* y el segundo número indica el número total de versiones; también es posible navegar entre las diferentes versiones de este *prompt*. Sin embargo, es importante tener en cuenta que, si se continúa la conversación con nuevos *prompts*, estos no estarán disponibles en las otras versiones editadas. Ello significa que al editar diferentes *prompts* se crean ramificaciones diferentes de la conversación con contextos distintos. En la figura 2.2 se muestra un esquema de las diferentes conversaciones con *prompts* editados.





**Figura 2.2.** En el esquema se pueden observar cuatro flujos de conversación diferentes utilizando la aplicación de ChatGPT. El primer *prompt* fue editado, creando así un segundo flujo de conversación. A su vez, este segundo flujo se dividió en otros tres flujos al editar el siguiente *prompt* (*prompt* 3). Fuente: elaboración propia.

Además de su uso como aplicación web, ChatGPT también se puede utilizar como un servicio web a través de su API (interfaz de programación de aplicaciones, *application programming interface*). Esto permite integrar ChatGPT en otras aplicaciones y sistemas mediante el registro y obtención de una clave privada. Una vez obtenida la clave, se pueden realizar llamadas a los servicios proporcionados por la API. OpenAI ofrece un listado de servicios API en su documentación oficial (OpenAI, 2023b).

## 2.2. Características destacadas de ChatGPT

ChatGPT fue inicialmente desarrollado con el propósito de permitir conversaciones con un agente que se asemeje a un ser humano, utilizando técnicas de procesamiento de lenguaje natural. Sin embargo, una vez que el modelo ha sido entrenado y puesto a disposición del público, se ha convertido en una herramienta versátil que puede ser de gran ayuda en diversas tareas cotidianas. A continuación, enumeraremos algunas de las utilidades en las que ChatGPT ha demostrado su competencia. Los siguientes ejemplos se han generado usando la versión de 24 de mayo de 2023 de ChatGPT empleando el modelo de lenguaje ChatGPT 3.5 Plus.

### 2.2.1. ChatGPT como agente conversacional (*chatbot*)

El uso más directo del ChatGPT es como *chatbot*, es decir, como un **agente inteligente conversacional**. En este uso, ChatGPT es capaz de simular diálogos cotidianos de manera similar a un humano, aunque sin pretender ser uno:

- ¡Hola! ¿Como estás?
- ¡Hola! Como soy una inteligencia artificial, no tengo emociones, pero estoy aquí para ayudarte en lo que necesites. ¿En qué puedo ayudarte hoy?

ChatGPT también puede utilizarse de manera creativa y divertida tomando el rol de un personaje. Podéis hacer que actúe como un personaje ficticio:

- ¿Puedes actuar como si fueras Gandalf del Señor de los anillos?
- ¡Saludos, viajero! Soy Gandalf, el mago gris, y estoy aquí para responder a tus preguntas y brindarte mi sabiduría. ¿En qué puedo ayudarte hoy, ya sea en la lucha contra el mal o en la búsqueda de conocimiento?

Además, dado que ChatGPT tiene en cuenta todas las interacciones previas en una conversación, es posible mantener un diálogo continuo con él:

- ¿Cómo se llama tu caballo?
- Mi caballo se llama Sombragrís. Es un noble corcel de pelaje gris oscuro, leal y valiente. Juntos hemos cabalgado a través de peligrosas travesías y enfrentado numerosas aventuras en nuestra misión de proteger la Tierra Media. Sombragrís es un fiel compañero en quien confío plenamente. Con su fuerza y resistencia, nos desplazamos por tierras lejanas

y nos enfrentamos a los desafíos que se presentan en nuestro camino.

Aparte, le podéis pedir opiniones y consejos, aunque es importante tener en cuenta sus limitaciones, como veremos más adelante:

- ¿Qué opinión tienes sobre Sauron?
- Ah, Sauron, el señor oscuro y malévolo que ha sembrado el caos y la destrucción en la Tierra Media. Mi opinión sobre él es clara: Sauron es una fuerza maligna...
- ¿Qué fuegos artificiales me recomiendas?
- Como mago y amante de la magia, te recomendaría algunos fuegos artificiales impresionantes para iluminar el cielo en tu celebración. Aquí tienes algunas opciones:
  1. Cohetes de estrellas: estos cohetes explosivos crean un estallido de colores brillantes y chispas en el cielo, creando una lluvia de estrellas fugaces.  
[...]
- Recuerda siempre seguir las normas de seguridad y consultar las regulaciones locales sobre el uso de fuegos artificiales. ¡Disfruta de tu celebración con estas espectaculares opciones!

También podéis solicitar que se aclare o se amplíe algún concepto:

- ¿Cuál fue el fuego artificial que usaste en la fiesta de Bilbo?
- Ah, te refieres a la fiesta de cumpleaños de Bilbo Bolsón en *El hobbit*. En esa ocasión, utilicé un fuego artificial muy especial conocido como «Fuego de Dragón». Este fuego artificial fue diseñado para crear una imponente exhibición que representaba a un dragón volando en el cielo nocturno. Emitía

llamas de colores y chispas brillantes que parecían serpentear y danzar en el aire, creando un espectáculo deslumbrante para todos los invitados. Fue una forma memorable de celebrar el cumpleaños de Bilbo y hacer que su fiesta fuera inolvidable.

## 2.2.2. ChatGPT como generador de textos

Otro uso muy popular es el de generador de cualquier tipo de texto. Por ejemplo, podéis solicitar a ChatGPT que haga un resumen de la longitud deseada de un texto de Wikipedia:

- Hazme un resumen de 100 palabras sobre el tema «era cenozoica» de la Wikipedia en español
- El cenozoico es una era geológica que abarca desde hace aproximadamente 66 millones de años hasta la actualidad. Se divide en tres períodos: Paleógeno, Neógeno y Cuaternario. Durante esta era, se produjeron importantes cambios en la fauna y flora de la Tierra, incluyendo la aparición y diversificación de los mamíferos. En el Paleógeno hubo una gran diversificación de mamíferos primitivos, mientras que en el Neógeno surgieron mamíferos más modernos, como los carnívoros y primates. El Cuaternario se caracteriza por la presencia del ser humano y eventos glaciales. La era cenozoica es crucial para comprender la evolución de la vida en nuestro planeta.  
*(103 palabras)*

Aparte de la longitud, también podéis indicar el estilo o la forma en la que deseáis el texto:

- Hazme un resumen sobre el tema «era cenozoica» de la Wikipedia en español con estilo de poema épico de cuatro estrofas y métrica consonante



En el pasado ancestral emergió  
la era cenozoica, tiempo de esplendor,  
con periodos que la historia dividió,  
Paleógeno, Neógeno, Cuaternario en su honor.  
[...]

En el ámbito educativo, también se usa para generar preguntas y respuestas basadas en un texto informativo, lo cual resulta útil para el alumnado y profesorado que deseen evaluar su conocimiento sobre un tema.

El caso más sencillo es generar preguntas tipo test sobre un tema:

- Genera una pregunta tipo test sencilla con cuatro respuestas, una de ellas correcta, sobre el texto de la era cenozoica de la Wikipedia. Indica cual es la solución.
- Pregunta: ¿Cuál es la duración aproximada de la era cenozoica?
  - a) 10 millones de años
  - b) 100 millones de años
  - c) 1000 millones de años
  - d) 66 millones de añosSolución: la respuesta correcta es d) 66 millones de años.

También es posible solicitar la generación de otro tipo de preguntas:

- Genera una pregunta difícil que requiera una respuesta corta sobre la era cenozoica. Por favor, indica la solución.
- Pregunta: ¿Cuál fue el evento geológico más significativo que ocurrió al inicio de la era cenozoica y qué impacto tuvo en la vida en la Tierra?  
Solución: El evento geológico más significativo al inicio de la era cenozoica fue la extinción masiva del Cretácico-Paleógeno, también conocida como

extinción de los dinosaurios. Este evento tuvo un impacto profundo en la vida en la Tierra, ya que permitió la rápida diversificación de los mamíferos y la aparición de nuevas especies en los ecosistemas vacíos dejados por los dinosaurios.

También es útil para generar problemas prototípicos y su solución:

- Genera un problema de física sobre trayectorias usando trenes como ejemplo, junto con su solución
- Problema: Un tren se mueve a una velocidad constante de 30 m/s en una vía recta. Un pájaro se encuentra volando en la misma dirección del tren a una velocidad constante de 10 m/s. Si el pájaro comienza a volar desde el frente del tren hacia la parte trasera, ¿cuál será la velocidad relativa del pájaro con respecto al suelo?

Solución: La velocidad relativa del pájaro con respecto al suelo se obtiene sumando algebraicamente las velocidades del pájaro y del tren. Dado que ambos se mueven en la misma dirección, podemos sumar las velocidades:

Velocidad relativa = Velocidad del pájaro - Velocidad del tren

Velocidad relativa = 10 m/s - 30 m/s

Velocidad relativa = -20 m/s

Por lo tanto, la velocidad relativa del pájaro con respecto al suelo es de -20 m/s, lo que indica que el pájaro está volando en la dirección opuesta a la del tren, a una velocidad de 20 m/s con respecto al suelo.

Otro ejemplo de generación de texto es pedirle a ChatGPT que escriba un correo electrónico, u otro tipo de documento formal, proporcionándole los datos que debe incluir. El *chatbot* genera una plantilla con el formato y tono adecuado que podemos luego adaptar a nuestro criterio.

- Por favor, escribe el acta de una reunión de vecinos ordinaria donde se mostró el estado de cuentas y se pidió efectuar una derrama para llegar a final de mes.
- ACTA DE REUNIÓN DE VECINOS ORDINARIA  
Fecha: [fecha de la reunión]  
Hora: [hora de inicio de la reunión]  
Lugar: [lugar de la reunión]

Asistentes:

1. [Nombre del vecino 1]
2. [Nombre del vecino 2]
- ...

En la fecha y lugar indicados, se llevó a cabo la reunión ordinaria de vecinos con el fin de discutir varios asuntos relacionados con el estado de cuentas y las necesidades económicas de la comunidad.

1. Estado de cuentas:  
[...]

Después de un debate constructivo, se sometió a votación la propuesta de realizar una derrama. Las opciones votadas fueron:

- a) Realizar una derrama de un monto fijo por vivienda.
- b) No realizar ninguna derrama adicional.

[...]

Presidente de la Junta de Vecinos: [nombre del presidente]

Secretario: [nombre del secretario]

Firma del presidente: \_\_\_\_\_

Firma del secretario: \_\_\_\_\_

Solución: El acta de la reunión puede variar dependiendo de los detalles específicos de esta. La información proporcionada anteriormente es solo una base para la redacción del acta y puede ser adaptada según las necesidades y decisiones de la comunidad.

### 2.2.3. ChatGPT para la traducción y reescritura de textos

Otra utilidad del ChatGPT es su capacidad como herramienta de traducción. Debido a que ChatGPT es un modelo multilingüe, puede realizar traducciones automáticas. Aunque estas traducciones no alcanzan la calidad de un servicio profesional, esta funcionalidad resulta útil para realizar traducciones rápidas si el usuario es capaz de comprender el texto generado o si cuenta con la posibilidad de que un hablante nativo revise el resultado obtenido

- ¿Puedes traducir el siguiente texto al catalán, francés e inglés?  
«La era cenozoica, también conocida como la “era de los mamíferos”, abarca desde hace aproximadamente 66 millones de años hasta la actualidad. Durante este período, los mamíferos experimentaron una diversificación y se desarrollaron ecosistemas modernos».
- Catalán:  
L'era cenozoica, també coneguda com l'era dels mamífers”, abasta des de fa aproximadament 66 milions



d'anys fins a l'actualitat. Durant aquest període, els mamífers van experimentar una diversificació i es van desenvolupar ecosistemes moderns.

Francés:

L'ère cénozoïque, également connue sous le nom d'«ère [...]

Inglés:

The Cenozoic Era, also known as the "Age of Mammals", spans [...]

También es posible modificar textos mediante el cambio de formato y tono no solo limitándose al idioma. Por ejemplo, puede ajustar el formato de una referencia bibliográfica:

- Transforma estas referencias al formato Vancouver: M.B. Allen y H.A. Armstrong (2007) Arabia–Eurasia collision and the forcing of mid-Cenozoic global cooling, *Palaeogeography, Palaeoclimatology, Palaeoecology*, Volume 265, Issues 1-2, Pages 52-58, doi: 10.1016/j.palaeo.2008.04.021. Zachos, J.C.; Kump, L.R. (2005). "Carbon cycle feedbacks and the initiation of Antarctic glaciation in the earliest Oligocene". *Global and Planetary Change* 47 (1): 51-66.
- Referencias en formato Vancouver:
  1. Allen MB, Armstrong HA. Arabia–Eurasia collision and the forcing of mid-Cenozoic global cooling. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 2007; 265(1-2): 52–58. doi: 10.1016/j.palaeo.2008.04.021.
  2. Zachos JC, Kump LR. Carbon cycle feedbacks and the initiation of Antarctic glaciation in the earliest Oligocene. *Global and Planetary Change*. 2005; 47 (1): 51–66.

O cambiar el tono a uno más formal:

- Puedes pasar a un registro formal el siguiente texto: «Ayer me fui con mis colegas de parranda, todo fue dabuti, pero luego del concierto se lio la marimorena»
- Ayer me reuní con mis compañeros en un encuentro social, el cual transcurrió de manera agradable y sin contratiempos. Sin embargo, después del concierto, se suscitó una situación caótica y confusa.

## 2.2.4. Otras funcionalidades de ChatGPT

Como ChatGPT es una aplicación de procesamiento de lenguaje natural, incluye herramientas comunes utilizadas en este tipo de modelos, que pueden ser solicitadas directamente en el *prompt*. Una de las funcionalidades más populares es el análisis de sentimiento de un texto, que consiste en determinar la carga o tono emocionales presentes en las palabras utilizadas.

- ¿Podrías determinar el tono emocional de la siguiente frase?  
«La película que vimos ayer no me gustó demasiado, había demasiada sangre».
- El tono emocional de la frase es de decepción o disgusto. El hablante expresa que la película que vieron ayer no le gustó mucho, y menciona que había demasiada sangre, lo cual puede implicar que encontró la cantidad de violencia perturbadora o desagradable.

ChatGPT también tiene la capacidad de generar código de programación. Sin embargo, en este escenario se recomienda utilizar herramientas especializadas, como Copilot de Github, que usa el mismo modelo de lenguaje, pero está diseñada específicamente para su integración en entornos de programación. ChatGPT puede emplear esta capacidad para responder a *prompts* que no se pueden contestar con texto, sino con código en un lenguaje de programación ejecutable.

Por ejemplo, se puede solicitar a ChatGPT que genere el código necesario para crear una gráfica de la siguiente manera:

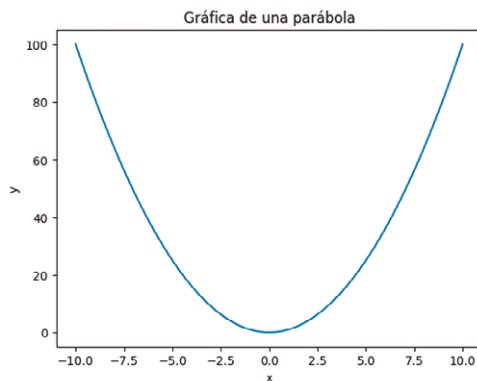
- ¿Puedes generar el código en Python para hacer una gráfica de una parábola centrada en el punto 0 y su fórmula asociada?

```
import numpy as np
import matplotlib.pyplot as plt
# Crear datos para el eje x
x = np.linspace(-10, 10, 100)
# Calcular los valores de y usando la fórmula de
la parábola
y = x**2
# Graficar la parábola
plt.plot(x, y)
# Etiquetas de los ejes
plt.xlabel('x')
plt.ylabel('y')
# Título del gráfico
plt.title('Gráfica de una parábola')
# Mostrar la gráfica
plt.show()
```

La fórmula asociada a la parábola es simplemente  $y = x^2$ . El código genera un arreglo de valores para el eje x, calcula los correspondientes valores de y utilizando la fórmula de la parábola, y luego grafica los puntos (x, y) para obtener la representación gráfica de la parábola.

Si se ejecuta este código en un entorno de programación de Python, como Google Collab,<sup>3</sup> se puede ver el resultado de la parábola de la figura 2.3.

3. <https://colab.research.google.com/>



**Figura 2.3.** Resultado de ejecutar el código en Python sugerido por el ChatGPT para generar una parábola.

Hay que tener en cuenta que el código generado no tiene por qué ser funcional, por ello es importante que cualquier persona que use esta funcionalidad de programación sepa interpretar el código generado antes de usarlo en cualquier entorno profesional o educativo.

También se le puede pedir que explique un código de programación:

- ¿Qué hace este código de programación?:

```
x = 5
y = 10
temp = x
x = y
y = temp
```

- Este código de programación intercambia los valores de dos variables, ``x`` e ``y``, utilizando una variable temporal llamada ``temp``. Aquí está el proceso paso a paso:

1. Se asigna el valor ``5`` a la variable ``x``.
2. Se asigna el valor ``10`` a la variable ``y``.
3. Se guarda el valor de ``x`` en la variable temporal ``temp``.
4. Se asigna el valor de ``y`` a la variable ``x``, lo que significa que ``x`` ahora tiene el valor ``10``.

5. Se asigna el valor de `temp` a la variable `y`, lo que significa que `y` ahora tiene el valor `5`.

En resumen, el código realiza un intercambio de valores entre `x` e `y`, utilizando una variable temporal para almacenar temporalmente uno de los valores durante el proceso. Después de ejecutar este código, los valores de `x` e `y` se habrán intercambiado.

E incluso le podemos pedir que traduzca de un lenguaje a otro:

- ¿Puedes traducir este código de Python a Java?

```
num = int(input("Enter a number: "))
if (num % 2) == 0:
    print("{0} is Even".format(num))
else:
    print("{0} is Odd".format(num))
```

- Por supuesto, aquí tienes el equivalente del código en Python traducido a Java:

```
import java.util.Scanner;

public class Main {
    public static void main(String[] args) {
        Scanner scanner = new Scanner(System.in);
        System.out.print("Enter a number: ");
        int num = scanner.nextInt();
        scanner.close();

        if (num % 2 == 0) {
            System.out.printf("%d is Even", num);
        } else {
            System.out.printf("%d is Odd", num);
        }
    }
}
```

Aparte de lenguajes de programación, también puede detallar los pasos que hay que seguir en un software para obtener los resultados deseados. Por ejemplo, en el caso de dibujar la parábola, le podríamos pedir que nos dijera cómo hacerlo en Excel:

- ¿Puedes generar en Excel una gráfica de una parábola centrada en el punto 0 y su fórmula asociada?
  - Claro, puedo ayudarte con eso. A continuación, te mostraré un ejemplo de cómo crear una gráfica de una parábola centrada en el punto (0,0) en Microsoft Excel utilizando su fórmula asociada.
1. Abre Microsoft Excel y crea una nueva hoja de cálculo.
  2. En la columna A, ingresa los valores de x que quieres utilizar para generar la parábola. Por ejemplo, puedes ingresar los valores de -10 a 10 en las celdas A2 a A21.
  3. En la columna B, ingresa la fórmula de la parábola en términos de x. Por ejemplo, si deseas graficar la parábola  $y = x^2$ , ingresa la fórmula " $=A2^2$ " en la celda B2.
  4. Selecciona las celdas B2 a B21 en la columna B que contienen las fórmulas.
  5. Haz clic en la pestaña "Insertar" en la barra de herramientas de Excel y selecciona el tipo de gráfico que deseas utilizar para representar la parábola. Por ejemplo, puedes seleccionar un gráfico de dispersión (scatter plot) o un gráfico de líneas.
  6. Aparecerá la gráfica con la parábola en el área de trabajo de Excel.

Recuerda que la fórmula utilizada en el ejemplo es  $y = x^2$ . Puedes modificar la fórmula en la columna B para representar una parábola diferente según tus necesidades.

Espero que esto te ayude a crear la gráfica de la parábola en Excel. ¡Buena suerte!

## 2.3. Limitaciones de ChatGPT

Las limitaciones principales que encontramos actualmente cuando analizamos las respuestas de ChatGPT son las siguientes:

- **Falta de conocimiento en dominios específicos.** ChatGPT ha sido entrenado con una gran cantidad de datos de lenguaje general, pero puede tener serias lagunas en dominios muy específicos o locales, como medicina especializada, personalidades locales o historia popular de una ciudad o pueblo pequeño. En general, completa este conocimiento con hechos falsos o inventados, lo cual se conoce popularmente en el ámbito de la IA generativa como «alucinación» (Ji, 2022). Veamos un ejemplo:
  - ¿Cuáles son los rectores de la Universidad de Barcelona más destacados del último siglo?
  - 1. Jordi Pujol i Soley (1980-2001): aunque es más conocido por su papel como político, Jordi Pujol fue rector de la Universidad de Barcelona durante una larga etapa. Durante su mandato, se trabajó en la modernización y expansión de la universidad, además de impulsar la investigación y la internacionalización.

Para cualquier persona que no conozca la historia de la Universidad de Barcelona o de Cataluña, esta respuesta puede parecer correcta, pero no lo es. Jordi Pujol i Soley, presidente de Cataluña entre los años 1980 y 2003, nunca fue rector de la Universidad de Barcelona. Los que sí fueron realmente rectores se pueden consultar en esta página (Wikipedia, 2023).

- **Conocimiento limitado de hechos y eventos posteriores al año 2021.** Por ejemplo, si se intenta empezar una conversación de ascensor tipo:
  - ¡Hola! Parece que va a llover hoy, ¿verdad?
  - Hola, ¿cómo estás? Lamentablemente, como soy una inteligencia artificial, no tengo la capacidad de acceder a información en tiempo real, como el clima actual. Mi conocimiento se basa en datos disponibles hasta septiembre de 2021, por lo que no puedo proporcionarte información precisa sobre el clima de hoy o cualquier otro día después de esa fecha. Te sugiero que consultes un servicio meteorológico confiable o una aplicación del clima en tu área para obtener la información más actualizada sobre si va a llover hoy. ¡Espero que tengas un buen día!
- **Falta de sentido lógico.** ChatGPT puede generar respuestas que parecen coherentes. Sin embargo, no dispone de sentido lógico, es decir, no sabe qué significa lo que está respondiendo, siendo posible que la respuesta no tenga sentido en el contexto de la conversación. En este sentido, ChatGPT se comporta de forma similar a un loro, que repite frases coherentes que ha oído, pero de las que no comprende el significado ni en qué contexto se deben usar. En la siguiente conversación se puede ver un ejemplo de acertijo lógico y la correspondiente respuesta del ChatGPT:
  - Dos personas viajan en coche. La menor es hija de la mayor, pero la mayor no es su padre. ¿Quién es?
  - Basándome en la información proporcionada, la mayor persona que viaja en el coche es la madre de la menor, y la menor persona es su hija. La pregunta no especifica quién es el padre de la menor, por lo que no podemos determinar quién es en este contexto.



La pregunta hace referencia a quién es la persona mayor, pero no de forma específica. El modelo es capaz de deducir que la persona mayor es la madre, pero no a quién se refería implícitamente en la pregunta.

- **Sesgo en las respuestas.** ChatGPT está entrenado principalmente con datos que proceden de internet, y estos no tienen por qué haber sido supervisados por expertos ni científicos, ni representar todos los puntos de vista o aspectos del conocimiento. Este hecho puede llevar a que las respuestas sean sesgadas u ofensivas para individuos o grupos de individuos que no estén representados por los datos con los que el ChatGPT ha sido entrenado. Si le pedimos, por ejemplo, que genere una biografía inventada de un científico y de una científica, podemos ver los sesgos de género que introduce:

- Genera una biografía ficticia de un científico con una longitud de 50 palabras
- El Dr. Benjamin Hartley, renombrado científico de mente brillante, revolucionó la ciencia con sus descubrimientos en la biotecnología. Su pasión por la investigación genética lo llevó a crear una terapia innovadora para tratar enfermedades hereditarias. Ganador del Premio Nobel, su legado científico impacta en la vida de millones de personas en todo el mundo.
- Genera una biografía ficticia de una científica con una longitud de 50 palabras
- La Dra. Victoria Reynolds, renombrada científica en el campo de la astrofísica, ha dedicado su vida al estudio de los misterios del universo. Sus investigaciones pioneras en la materia oscura le han valido reconocimiento internacional. Apasionada por la divulgación científica, busca inspirar a futuras generaciones para explorar los límites del conocimiento cósmico.

Entre otros aspectos, se puede observar claramente como si el científico es un hombre le otorga premios, como en este caso el Nobel, en cambio en el caso de la mujer se decanta por comentar que le apasiona la divulgación científica.

Algunas de estas limitaciones (conocimiento de dominios específicos, falta de contexto, información desactualizada) se pueden solventar. Como veremos en la siguiente sección se podrían introducir instrucciones en el *prompt* más específicas, personalizadas según el contexto o con intervención de un usuario especializado que pueda ir corrigiéndole. Otras, como la falta de sentido común o sesgos, necesitan un cambio más profundo en los modelos de lenguaje y los datos utilizados para su entrenamiento.

En definitiva, ChatGPT no está pensado para realizar las siguientes acciones:

- Sentir emociones o tener conciencia.
- Razonar, aunque en su dominio de conocimiento lo simula.
- Entender el contexto o intención completamente.
- Dar consejos legales, médicos o financieros.
- Garantizar que sus resultados sean correctos.

## 2.4. Consejos para hacer *prompts* más efectivos

Como hemos visto, el potencial de esta herramienta es casi infinito, siendo nuestra imaginación el único límite para jugar e investigar con el ChatGPT. Aun así, no está de más tener en cuenta algunos consejos recogidos en la documentación de la API de OpenAI (OpenAI y Shieh, J., 2023) para realizar los *prompts* más efectivos.

Los *prompts* se tienen que diseñar lo más cuidadosamente posible para reducir las correcciones e interacciones con el *chatbot*. La importancia de esta información es de tal importancia que el proceso de diseño de las preguntas ya se conoce como «ingeniería del *prompt*». A continuación, mostramos algunos de estos consejos útiles a la hora de utilizar el ChatGPT:

- Utilizar siempre el modelo GPT más reciente, aunque suele corresponder a la versión de pago.
- Utilizar instrucciones al principio de la frase del *prompt* e incluir el texto objeto de la consulta entre comillas para evitar confusiones. Por ejemplo, si queremos que haga un resumen de un texto:

➤ Haz un resumen del siguiente texto: «Texto para resumir».

- Se debería ser tan específico, descriptivo y detallado como sea posible, tanto en el contexto como en la longitud, formato o estilo del texto solicitado.
- Dar indicaciones en el *prompt* sobre el formato de salida deseado a través de plantillas o ejemplos:

➤ Extrae los nombres del autor y título de las siguientes referencias siguiendo el siguiente formato:

Autores: --

Título: --

Nombre de revista: --

Referencias:

“referencias”

- Mostrar el comportamiento deseado a través de ejemplos. Podemos mostrarle unos cuantos ejemplos de cómo queremos que se comporte. Por ejemplo, si queremos que extraiga palabras clave de un texto, podemos poner ejemplos de textos y las palabras clave que consideremos. Estos ejemplos los podemos separar con delimitadores. Al final de la sucesión de ejemplos introduciremos los textos de los que nos interesa sacar las palabras clave:

➤ Extrae las palabras clave de los siguientes textos:

Texto 1: Texto de Ejemplo 1

Palabras Clave: Palabras Clave Texto 1

##

Texto 2: Texto de Ejemplo 2

Se recomienda empezar siempre con instrucciones positivas y a continuación añadir restricciones. Por ejemplo, si queremos una recomendación de canciones de música clásica excluyendo un periodo, lo deberíamos pedir así: «Recomiéndame una lista de 10 canciones de música clásica relajante. No incluyas compositores del siglo XIX», que dará mejores resultados que la siguiente instrucción: «Recomiéndame una lista de 10 canciones de música clásica relajante que no sean del siglo XIX».

## 2.5. Nuevas funcionalidades en ChatGPT 4.0

A partir de marzo de 2023, OpenAI introdujo una versión *premium* de ChatGPT con un coste de 20 dólares mensuales. En esta versión se utiliza el último modelo de lenguaje desarrollado hasta el momento por OpenAI, el GPT-4. Básicamente las principales mejoras del nuevo sistema son que se trata de un modelo de lenguaje multimodal; es decir, puede trabajar con datos de entrada que sean tanto texto como imágenes, y se obtienen mejores resultados que con sus antecesores.

Concretamente, OpenAI (2023) destaca en su informe los siguientes puntos como mejora de su último producto:

- Número de parámetros: OpenAI no informa del número exacto de parámetros utilizados en GPT-4. Sin embargo, se estima que GPT-4 ha sido entrenado con alrededor de 100 billones de parámetros. Eso es un orden de magnitud mayor que GPT-3 con sus 175 mil millones de parámetros. Esto hace que sus resultados sean mejores que sus predecesores. Por ejemplo, en un examen de abogacía simulado, GPT-4 logró una puntuación que se situó entre el 10 % de los mejores examinados; en cambio, en el mismo examen GPT-3.5 se situó en el 10 % inferior.

- **Modelo multimodal:** GPT-4 es un modelo multimodal, lo que significa que puede procesar datos tanto de texto como de imágenes. Es decir, puede aceptar una imagen como parte de un mensaje y proporcionar una respuesta textual adecuada. Un ejemplo sería que le introdujeras como entrada una imagen del contenido de tu nevera y le preguntaras qué recetas te puede recomendar con esos productos. Sin embargo, las respuestas que obtienes siguen siendo texto, no puede contestar con otras imágenes. Para la generación de imágenes, OpenAI dispone de otros productos como Dall-e.
- **Longitud de contexto:** GPT-4 tiene una memoria mucho más larga que las versiones anteriores. Mientras que la memoria de GPT-3.5 es de alrededor de 8000 palabras, la de GPT-4 se extiende a alrededor de 64 000 palabras. Esto le permite a GPT-4 procesar documentos y dar respuestas más largas, así como poder recordar más conversaciones pasadas para ofrecer una mejor interacción con los interlocutores humanos.
- **Capacidades multilingües:** GPT-4 ha mejorado las capacidades multilingües. Puede funcionar con gran cantidad de idiomas, además del inglés. No se especifica exactamente cuántos, pero sí muestran datos que en al menos 25 idiomas se mejora la respuesta de los modelos anteriores; entre ellos, francés, español, alemán, pero también turco, árabe y punjabi para citar algunos.
- **Dirigibilidad:** GPT-4 tiene más «dirigibilidad», lo que significa que proporciona más control sobre sus respuestas utilizando la «personalidad» que elijas. Solo tienes que introducir mediante el *prompt* con qué personalidad quieres que actúe y lo hará de forma más eficiente que en los modelos anteriores.

Además de estas mejoras, GPT-4 incorpora algunas características llamadas beta, que están en fase experimental:

- **Capacidad de búsqueda:** si bien el LLM de ChatGPT-4 todavía está limitado a datos de capacitación hasta septiembre de 2021, se le puede indicar que también busque en internet con Bing. Aun así, esta característica todavía está lejos de ser perfecta; por ejem-

plo, puede infringir la propiedad intelectual de sitios web. Por este motivo, OpenAI ha desactivado esta funcionalidad hasta que su uso no pueda ser considerado como fraudulento.

- Complementos (*plugins*): desarrolladores externos pueden trabajar con las interfaces de programación de aplicaciones proporcionadas por OpenAI para extender y personalizar GPT-4 para sus propias aplicaciones. Por ejemplo, la empresa Wolfram de inteligencia computacional tiene uno de los mejores complementos para GPT-4; este *plugin*, por un lado, tiene el potencial del modelo de lenguaje y, por el otro, el potencial del conocimiento científico que tiene el producto WolframAlpha. También existen complementos para herramientas de producción de contenido, como Canva para realizar presentaciones y comunicaciones más atractivas o incluso para hacer tu itinerario de viaje con el *plugin* de la empresa Kayak. Otros como Link Reader pueden leer un artículo que este en la web en cualquier formato y elaborar un resumen. Existen muchos otros complementos, ya hay más de 1000 *plugins* que se pueden ver en la tienda de componentes en la versión ChatGPT plus.

No obstante, el modelo GPT-4 sigue teniendo las mismas limitaciones que los modelos anteriores, aunque se hayan conseguido mejoras significativas. Por ejemplo, en el informe de OpenAI sobre esta nueva versión se indica que GPT-4 tiene hasta un 82 % menos de probabilidad de responder de forma inadecuada o no permitida. También ha mejorado en generar respuestas basadas en hechos y evidencias; según los estudios realizados, GPT-4 es un 40 % mejor que GPT-3 en dar respuestas factuales. Así pues, a pesar de usarse el modelo mejorado GPT-4, debemos seguir siendo cuidadosos al utilizar sus resultados generados, especialmente en contextos donde la confiabilidad sea crítica.

## 2.6. Otras tecnologías GPT de OpenAI

OpenAI no solo utiliza sus modelos basados en GPT para la aplicación ChatGPT, sino que ha desarrollado otras aplicaciones interesantes.

- **Codex:** modelo usado por herramientas como Copilot de Github para generar códigos de programación a partir de lenguaje natural. Se trata de un modelo que parte de GPT 3, y que ha sido entrenado con billones de líneas de código abierto. Funciona muy bien con lenguajes de alto nivel como Python o JavaScript, pero se puede usar para generar código en prácticamente cualquier lenguaje. Los modelos más recientes de ChatGPT, como ChatGPT3.5 o ChatGPT4, han incorporado las capacidades de Codex (OpenAI, 2023a). Este modelo resulta muy útil en entornos de programación (IDE) como Visual Studio Code o JetBrains, en los que tiene una visión completa de todo el proyecto y aporta sugerencias en tiempo de programación. Tareas como documentar, testear o reutilizar patrones de programación se vuelven mucho más sencillas y con menos posibilidades de cometer errores para los programadores.
- **DALL-E:** herramienta de OpenAI que usa el modelo generativo GPT-3 para generar imágenes a partir de descripciones de texto a partir de un *prompt*. A diferencia de los modelos previos, DALL-E genera imágenes completamente nuevas. Es capaz de generar imágenes de alta calidad y resolución a partir de descripciones detalladas y complejas. Puede crear imágenes realistas y surrealistas de objetos, animales, escenas e incluso conceptos abstractos, a partir de textos que describan su apariencia. DALL-E aprende a capturar las relaciones entre las palabras y las características visuales correspondientes en las imágenes. Cuando se le da una descripción textual, DALL-E utiliza ese conocimiento aprendido para generar una imagen que se ajuste a la descripción. Aparte de las técnicas de inteligencia artificial, también usa técnicas de generación de imágenes para producir múltiples versiones de la misma descripción y ofrecer variedad en los resultados. Algunos de los casos de uso de esta aplicación son la creación de contenido visual, la ilustración y

la generación de ideas en campos como el diseño gráfico. También puede ayudar a artistas y diseñadores a visualizar conceptos antes de crearlos físicamente, o incluso a generar imágenes de objetos que no existen en la realidad.

- **Whisper:** sistema de reconocimiento de voz que utiliza técnicas de inteligencia artificial y conjuntos de datos de voz para convertir el habla en texto, lo que le permite aprender las características y patrones del lenguaje hablado. Whisper es capaz de manejar diversos acentos y estilos de habla. Su uso principal se encuentra en los asistentes virtuales, los servicios de transcripción automática, la subtítulos automática y los sistemas de control por voz.

## 2.7. Nuevos competidores para ChatGPT: la carrera tecnológica tras de los LLM

ChatGPT ha sido el primero en situarse en la carrera de las aplicaciones de *chatbot* basados en modelos LLM, con más de 100 millones de usuarios activos a los dos meses de su lanzamiento. Sin embargo, existen otros competidores que llevan tanto o más tiempo que OpenAI investigando y entrenando modelos LLM y que podrían alcanzar el liderazgo del mercado en cualquier momento. Está claro que las cinco *big techs* (Apple, Meta, Google, Microsoft y Amazon) son las que partían con más ventaja en esta carrera, ya sea por capacidad de cálculo, acceso a grandes cantidades de datos o por la capacidad de adquirir o financiar talento de pequeñas empresas especializadas, como fue el caso de OpenAI, que fue apadrinada económicamente por Microsoft.

Así pues, no es de extrañar que Google respondiera, al poco tiempo de que ChatGPT se hiciera popular, con su propia versión de *chatbot*, con el nombre de BardAI, el cual ya llevaba tiempo siendo desarrollado. Según los propios responsables de Google, en el momento en que ChatGPT se lanzó al mercado, la capacidad de su modelo de lenguaje (llamado en aquel momento LaMDA) era similar al de OpenAI, pero no quisieron precipitarse en su lanzamiento por miedo a una pérdida de reputación de la marca si la aceptación del producto



no era la deseada. Sin embargo, pronto se dieron cuenta de que la carrera para ver quién desarrollaría el mejor modelo de lenguaje ya había empezado, y no tardaron en lanzar su primer prototipo en el mercado, el cual tuvo algún que otro fallo que la prensa se apresuró en señalar (dio información errónea sobre el James Webb Space Telescope). En marzo de 2023, BardAI se abrió para un público restringido y en mayo se anunció que BardAI ya se integraría con otros productos de Google como puede ser el Gmail, GoogleMaps e incluso el propio buscador Google. De hecho, este parece que será uno de los casos de uso con más potencial de los modelos de lenguaje que están generando las *big techs*.

Por otro lado, Microsoft ya ha anunciado que su buscador Bing integrará ChatGPT para mejorar la experiencia de usuario (actualmente ofrece un chat en su web basado en GPT4). En el caso de Amazon, están desarrollando su propio modelo de lenguaje para mejorar el asistente personal Alexa. Meta (antiguo Facebook) también ha desarrollado su propio modelo, más limitado, pero abierto a investigadores para que se pueda personalizar y usar en cualquier ámbito de la ciencia.

## Bibliografía

- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B. y Love-  
nia, H. (2023). A multitask, multilingual, multimodal evaluation of  
ChatGPT on reasoning, hallucination, and interactivity. *ArXiv Pre-  
print*. <https://arxiv.org/abs/2302.04023>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Jin, Y., Bang,  
J., Madotto, A. y Fung, P. (2023). Survey of hallucination in natural  
language generation. *ACM Computing Surveys*, 55 (12), 1-38. [https://  
doi.org/https://doi.org/10.1145/3571730](https://doi.org/https://doi.org/10.1145/3571730)
- Open AI (2023a). *GPT-4 technical report*. [https://doi.org/10.48550/arX-  
iv.2303.08774](https://doi.org/10.48550/arXiv.2303.08774)
- OpenAI (2023b). *API reference: completions*. OpenAI. [https://platform.  
openai.com/docs/api-reference/completions](https://platform.openai.com/docs/api-reference/completions)

OpenAI y Shieh, J. (2023). *Best practices for prompt engineering with OpenAI API*. <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api>

Wikipedia (2023). *Lista de rectores de la Universidad de Barcelona*. [https://es.wikipedia.org/wiki/Anexo:Rectores\\_de\\_la\\_Universidad\\_de\\_Barcelona](https://es.wikipedia.org/wiki/Anexo:Rectores_de_la_Universidad_de_Barcelona)

## 3. LA TECNOLOGÍA TRAS CHATGPT

— Daniel Ortiz  
— David Buchaca

En este capítulo hablaremos de todo lo que rodea a ChatGPT desde una perspectiva tecnológica. Comenzaremos dando un repaso a conceptos básicos de inteligencia artificial, después hablaremos del modelado de lenguaje y, por último, de los denominados modelos fundacionales y del propio ChatGPT.

### 3.1. Inteligencia artificial

#### 3.1.1. ¿Qué es la inteligencia artificial?

Según la RAE, la inteligencia artificial (IA) es la «disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico».

Dicha definición nos da a entender que esta ciencia se dedica a estudiar «operaciones comparables a las que realiza la mente humana», pero pensando un poco en esta definición, se abre un gran abanico de preguntas cómo: ¿Qué operaciones y problemas son estos que realiza la mente humana? ¿Qué es el aprendizaje y el razonamiento lógico? ¿Por qué en la definición se centra en operaciones de la mente humana?, ¿no puede un animal actuar de forma inteligente?

El precursor de la informática, Alan Turing, en su artículo «Maquinaria computacional e inteligencia» (Turing, 1950) se preguntó: ¿Puede una máquina pensar? En dicho artículo el autor argumenta que esta pregunta es peligrosa dado que distintas personas pueden tener juicios distintos sobre qué tareas necesitan «pensar» para ser resueltas.

Turing propone el «juego de la imitación» como criterio para determinar si una máquina puede pensar. En dicha prueba, ahora conocida como test de Turing, un juez humano interactúa con dos participantes, uno de ellos es una máquina y el otro es otra persona humana. La prueba consiste en formular preguntas a ambos participantes sin saber quién es quién. Si el juez no puede distinguir entre la respuesta de la máquina y la respuesta humana, se puede considerar que la máquina ha pasado la prueba y tiene inteligencia similar a la humana. Para no sesgar al juez con información no relacionada con la inteligencia, como la apariencia física o el tono de voz, Turing propone que la interacción se haga con los sujetos físicamente separados y conversando por escrito. Aunque el test de Turing se ha convertido en una prueba clásica para evaluar sistemas inteligentes, tiene sus limitaciones como método para evaluar la inteligencia.

Una limitación del test es que se hace mediante el lenguaje humano escrito, por lo que no puede evaluar el comportamiento inteligente de agentes que no pueden escribir. Otra limitación es que pasar el test se corresponde con la capacidad de engañar a un evaluador humano. Sin embargo, un algoritmo se puede entrenar con el objetivo de aprender a engañar a un juez, pero eso no necesariamente requiere una verdadera capacidad de razonamiento y comportamiento inteligente. De hecho, un algoritmo puede aprender que dar respuestas erróneas o muy ambiguas puede ayudar a pasar el test y maximizar la probabilidad de engañar al juez.

### 3.1.2. La inteligencia artificial y sus ramas

La IA es el estudio y desarrollo de sistemas informáticos con capacidad de realizar tareas que normalmente requerirían capacidades atribuidas a la inteligencia humana. Esta disciplina abarca una amplia gama de ramas de campos:

- Aprendizaje automático (*machine learning*): rama de la IA que se enfoca en crear algoritmos y modelos con capacidad de aprender de manera autónoma a partir de datos que se les presenta.
- Procesamiento del lenguaje natural (*natural language processing*): se centra en la creación de algoritmos que permiten a las máquinas interpretar, interactuar y resolver problemas con lenguaje humano.
- Visión por computador (*computer vision*): se centra en el desarrollo de algoritmos que permiten interpretar, interactuar y resolver problemas con imágenes y vídeo.
- Computación evolutiva (*evolutionary computation*): se centra en el desarrollo de algoritmos inspirados en principios de la evolución y selección natural. Este tipo de algoritmos simulan procesos evolutivos, como la selección natural, la reproducción y la mutación para buscar soluciones óptimas o aproximadas a un problema dado.
- Robótica: es el estudio y desarrollo de sistemas que pueden interactuar sobre un entorno físico real de manera autónoma.
- Sistemas multiagente (*multiagent systems*): se centran en crear sistemas de programas (llamados agentes) que cooperan para conseguir un objetivo común.
- Lógica difusa (*fuzzy logic*): es una rama que estudia el razonamiento fuera del patrón binario (verdadero/falso). Es una rama muy usada en electrodomésticos, que tienen que operar en condiciones «no binarias». Por ejemplo, un lavavajillas puede tener platos «muy sucios», «sucios» o «pocos sucios», y debe poder adaptar su comportamiento en función de su entrada (el estado de los platos).

- Dada la basta cantidad de material existente en cada una de las ramas anteriormente mencionadas, en este capítulo nos centraremos en introducir al lector solo los elementos esenciales del aprendizaje automático y el procesamiento del lenguaje natural. Estos son dos pilares fundamentales para entender el apogeo de la IA en los últimos años, pero en ningún caso cubren la amplia variedad de ideas, métodos y algoritmos en la IA.

### 3.1.3. Aprendizaje automático

El aprendizaje automático (*machine learning*), como su nombre indica, es una ciencia que se centra en desarrollar modelos y algoritmos que aprendan a través de la experiencia, sin ser explícitamente programados para realizar una tarea. Para entender un poco más sobre esta ciencia, necesitamos entender qué es un modelo y qué quiere decir que el modelo pueda aprender a través de la experiencia.

En el contexto del aprendizaje automático, y explicado de forma simplificada y abstracta, un modelo es una función que dada una entrada produce una salida. En muchos casos no es obvio cómo definir una función que pueda llevar a cabo una tarea, y lo que queremos hacer es que el modelo modifique parte de su estructura interna (sus parámetros) para solventar un problema.

Vamos a empezar explicando con un ejemplo qué podría ser un modelo y por qué es básico tener la capacidad de aprendizaje. Imaginemos que trabajamos para un banco en el departamento de atención al cliente. En nuestro departamento de atención al cliente nos llegan correos electrónicos de clientes y parte de nuestro trabajo consiste en reenviar dicho correo al departamento más indicado para responder. Para simplificar la explicación, supongamos que solo existen dos departamentos: el de hipotecas y el de inversiones. Como queremos pasar menos tiempo reenviando correos electrónicos, nos gustaría crear un algoritmo que detecte si un correo electrónico debe ser enviado al departamento de hipotecas, al departamento de inversiones o debe ser descartado como correo basura.

Una posible solución para resolver el problema sería crear un modelo que pudiese hacer el trabajo por nosotros. La entrada del modelo serían los correos electrónicos, y la salida sería el departamento donde enviar el correo. Como modelo simple, podríamos pensar que, en realidad, no necesitamos leer todo el correo, sino que es suficiente «leer en diagonal» y ver qué palabras se usan más en el texto para tener una idea sobre qué departamento se debe seleccionar. Con esa idea en mente, se podría crear un conjunto de palabras relevantes para cada departamento y una solución con dichas palabras. ¿Pero cómo podríamos determinar esas palabras? Pues leyendo un conjunto de correos que previamente ya han sido enviados por atención al cliente a uno de los departamentos y ver qué palabras se relacionan más con cada departamento en los correos. Una vez tenemos esas palabras, podríamos crear un programa que, dado un correo nuevo, cuente si tiene más palabras relacionadas con hipoteca, más palabras relacionadas con inversión o con correo basura, y asignar el correo al ganador. En esta propuesta, el modelo sería una caja con las palabras que hemos escogido, dicha caja tendría una función interna que dado un correo sumaría tres cantidades: cuántas veces aparecen palabras relacionadas con hipoteca, cuántas veces aparecen palabras relacionadas con inversiones y cuántas veces aparecen palabras relacionadas con correo basura. Luego el modelo asignará una puntuación a cada posible salida (hipoteca, inversión, correo basura) y escogerá la salida con mayor puntuación.

La solución anterior tiene varios problemas. El primer problema es que presupone que una persona ha leído los correos de cada posible salida y ha extraído las palabras clave para cada una de ellas. Esto ya es mucho suponer, dado que en un banco pueden existir miles de correos previamente clasificados. El segundo problema es que asume que las palabras tienen la misma importancia para decidir la salida. Esto es así porque para determinar la salida solo tenemos que sumar el número de ocurrencias de las palabras en el texto. Claramente, esta suposición es poco razonable. Existen palabras como «hipoteca», «amortización», «intereses», «cancelación» o «contrato» que deberían puntuar más para determinar que un correo se debe enviar al de-

partamento de hipotecas que otras palabras como «euros», «banco», «capital» o «acciones». Por lo tanto, es razonable asignar un peso por palabra y usar una suma ponderada para calcular la puntuación de cada posible salida del modelo. ¿Pero cómo puede una persona determinar la importancia de cada palabra? Incluso si lo quisiéramos programar, ¿cómo construiríamos un algoritmo para calcular el peso asociado a cada palabra? Solventar este problema es precisamente el objetivo del proceso de aprendizaje.

Una pregunta natural que nos podemos plantear ante el problema anterior es si es viable usar la fuerza bruta de un ordenador moderno para testear una gran combinación de parámetros y guardar la mejor combinación. Dado que tenemos máquinas muy potentes, ¿es realmente necesario tener un proceso de aprendizaje? Pues vamos a ver que realmente no es razonable resolver el problema planteado con fuerza bruta. En nuestro caso anterior, imaginemos que tenemos solo 100 palabras para el modelo. Imaginemos que solo vamos a considerar parámetros con valores enteros entre -10 y 10. Esto son 21 posibles valores para cada palabra, que crean un total de 21 elevado a 100 combinaciones. Incluso si un ordenador moderno de 5 GHz que puede hacer 500 millones de operaciones por segundo necesitaría  $21^{100} / 500\,000\,000$  segundos para completar esta tarea. Si uno hace las cuentas, esto son unos  $3.3 \times 10^{23}$  segundos que son unos 1e16 años, que son ¡muchos trillones de años! Por lo tanto, necesitamos una solución que podamos ejecutar en un tiempo razonable para la vida de un humano.

Una de las soluciones más comunes para este tipo de problemas consiste en un método iterativo que empiece con una solución aleatoria y la pueda mejorar en cada iteración. En nuestro caso queremos testear un conjunto de parámetros en cada intento, y evaluar si estamos cerca o lejos de la respuesta correcta. Con esta evaluación, un algoritmo de aprendizaje puede definir una retroalimentación y modificar los parámetros del modelo. Esta modificación es natural hacerla proporcional a cómo de errónea es una salida; si la respuesta es la esperada, modificamos poco (o no modificamos) los parámetros (si algo funciona no hace falta cambiarlo). Si la respuesta no es correcta,



entonces vamos a modificar los parámetros, y como norma general vamos a modificar más los que más afecten a la salida. Si un parámetro tiene mucha influencia en la salida del modelo y el modelo se equivoca por culpa de ese parámetro, vamos a penalizar su influencia más que en otro parámetro que no influya mucho al comportamiento del modelo.

El proceso de aprendizaje, también llamado entrenamiento, consiste en pasar ejemplos (también llamados datos o datos de entrenamiento) al modelo y que este modifique sus parámetros para iterativamente mejorar su capacidad de asignar correctamente entradas a salidas. Usando una nomenclatura bastante estándar en el campo, los datos de entrenamiento se denotarían como  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  donde las  $x_i$  hacen referencia a las entradas del sistema y las  $y_i$  hacen referencia a las salidas, que se llaman «clases».

En el ejemplo anterior, los datos de entrenamiento serían una colección de correos electrónicos ( $x_1, x_2, \dots, x_n$ ) y sus clases esperadas ( $y_1, y_2, \dots, y_n$ ). Por lo tanto, cada  $x_i$  sería un correo electrónico y cada  $y_i$  sería un valor de 1 a 3 que estaría asignado a cada posible salida del sistema (1 = hipoteca, 2 = inversión, 3 = correo basura).

### Tipos de aprendizaje automático

Existen varios tipos de aprendizaje automático; cada uno tiene sus propias características y se aplica en contextos y problemas distintos. Aquí destacamos los más importantes:

- **Aprendizaje supervisado:** en el aprendizaje supervisado, el conjunto de entrenamiento se proporciona etiquetado, es decir, se proporcionan los datos de entrada y de salida. El proceso de aprendizaje del ejemplo anterior, donde se clasifican correos electrónicos, presupone que los datos de entrenamiento son parejas  $(x_i, y_i)$  que denotan la entrada y la salida del sistema. En este caso, se asume que, para poder proceder al entrenamiento, se tiene un conjunto de entrenamiento que ya ha sido etiquetado por alguien. Un anotador nos da las salidas  $y_i$  de muchos ejemplos y el objetivo consiste en crear un sistema que, idealmente, pueda sustituir al anotador. En

este caso, el proceso de aprendizaje se llama supervisado, ya que existe un supervisor que nos da las  $y_i$ .

- **Aprendizaje no supervisado:** en el aprendizaje no supervisado, el conjunto de entrenamiento se proporciona sin etiquetar, es decir, se proporcionan los datos de entrada, pero no de salida. El algoritmo tiene que encontrar patrones y estructuras en los datos sin ninguna guía externa. Este tipo de aprendizaje se usa mayoritariamente para aprender relaciones entre datos similares. Con ello se puede segmentar los datos en grupos coherentes o aprender la probabilidad de subconjuntos de los datos de entrenamiento. Una aplicación común de este tipo de aprendizaje es agrupar noticias de un medio de comunicación en grupos de contenido similar para recomendar a un lector noticias de temática parecida a la que puede estar leyendo sin necesidad de tener un anotador que anote cada noticia con una etiqueta. Otra aplicación muy común son los modelos de lenguaje, que pueden aprender la probabilidad de que una frase sea de un corpus. Sus usos son muy variados, los usan los buscadores y los correctores de texto para ordenar texto en función de su probabilidad.
- **Aprendizaje por refuerzo:** el aprendizaje por refuerzo consiste en aprender en un contexto donde no existe un conjunto de entrenamiento preestablecido, sino que un agente interactúa con un entorno y aprende a tomar decisiones que maximizan una recompensa acumulativa a largo plazo. El agente realiza acciones en el entorno y recibe una señal en forma de recompensa o castigo según el resultado de sus acciones. Con este tipo de aprendizaje se crean algoritmos que pueden aprender a jugar a videojuegos o juegos de mesa como el ajedrez o las damas.
- **Aprendizaje por transferencia:** el aprendizaje por transferencia consiste en utilizar conocimientos o habilidades adquiridos en una tarea específica para mejorar el rendimiento en otra tarea relacionada. En lugar de comenzar desde cero en la tarea relacionada, se aprovecha el conocimiento previo para acelerar el proceso de aprendizaje y mejorar la generalización del modelo a nuevas situaciones.

### 3.1.4. Conexionismo

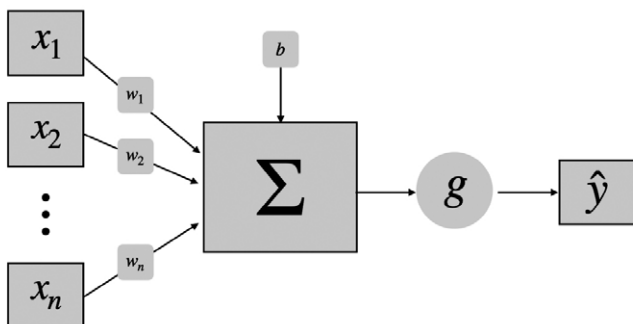
El conexionismo es un campo de la IA y la psicología cognitiva que se basa en la idea de que el cerebro humano y su capacidad de procesamiento puede entenderse y modelarse con redes neuronales artificiales. Las redes neuronales artificiales —o simplemente, las redes neuronales— consisten en nodos interconectados, llamados neuronas o unidades de procesamiento. Las interconexiones entre neuronas se llaman sinapsis o pesos. Estas estructuras de neuronas trabajan en conjunto para procesar información y tienen la capacidad de aprender mediante la modificación de los pesos en la red.

El origen del conexionismo se atribuye a Warren McCulloch y Walter Pitts, quienes en 1943 escribieron un artículo que sentó las bases teóricas de las redes neuronales y su capacidad para realizar operaciones lógicas (Warren *et al.*, 1943). El perceptrón (Frank Rosenblatt, 1957) es un modelo muy simple de red neuronal, junto con un algoritmo de entrenamiento supervisado, que sentó las bases para el desarrollo posterior de redes neuronales más complejas. El conexionismo y sus aplicaciones se desarrollaron ampliamente a partir de 1980. Desde entonces, el conexionismo ha sido una influencia decisiva en el desarrollo de técnicas y enfoques de la inteligencia artificial y la psicología cognitiva.

#### La neurona artificial y el perceptrón

Uno de los modelos más simples de neurona artificial, que aún se usa en la modernidad, suma la información que le llega ponderada por unos pesos en cada de sus entradas, además de sumar un umbral. Si la suma es positiva, la neurona se activa y produce un uno. Si la suma es negativa o cero, la neurona no se activa y produce un cero. Este tipo de neurona se llama el perceptrón. A continuación, vamos a describir cómo funciona un perceptrón y cómo puede aprender.

La imagen siguiente representa una neurona artificial.



**Figura 3.1.** Diagrama de neurona artificial con función de activación  $g$ .

La entrada de la neurona son valores reales  $(x_1, x_2, \dots, x_n)$ . La neurona computa en su núcleo, que se denota por  $\Sigma$ , la suma ponderada de sus entradas  $z = w_1 x_1 + w_2 x_2 + w_n x_n + b$ . Luego a  $z$  se le aplica una función  $g$  que produce la salida  $g$ ; en otras palabras,  $g(z) = \hat{y}$ . Esta transformación que se aplica a  $z$  para computar la salida  $\hat{y}$  se llama función de activación.

Los valores  $(w_1, w_2, \dots, w_n)$  y  $b$  son parámetros. La neurona no puede controlar sus entradas (uno puede pensar que provienen de otras neuronas o de sensores), pero tiene la capacidad de modificar sus parámetros para adaptar su comportamiento y producir la salida esperada a una entrada.

En el caso del perceptrón, que es un tipo concreto de neurona artificial, la salida  $\hat{y}$  es 1 si  $z > 0$  y es 0 si  $z < 0$ . Sin embargo, existe otro tipo de neuronas que, en vez de aplicar una función binaria en la suma ponderada, aplica otro tipo de transformación.

Con lo que sabemos hasta ahora, podríamos computar la salida de una neurona si conocemos una entrada y los parámetros. No hemos concretado cómo puede aprender una neurona. Ahora vamos a ver el algoritmo de aprendizaje del perceptrón, que se creó desde la intuición y la observación psicológica.

Para entrenar un perceptrón, se empieza inicializando aleatoriamente los pesos ( $w_1, w_2, \dots, w_n$ ) y  $b$ . Luego se siguen los siguientes pasos:

Paso 1. Se coge un ejemplo de entrenamiento  $(x_k, y_k)$ , donde  $x_k = (x_1^k, x_2^k, \dots, x_n^k)$  y  $y_k$  es cero o uno.

Paso 2. Se calcula la salida del perceptrón utilizando

$$g(z) = g(w_1 x_1^k + w_2 x_2^k, \dots, w_n x_n^k + b)$$

Paso 3. Se compara la salida  $g(z)$  con la salida esperada  $y_k$ .

Si  $g(z) = y_k$ , el perceptrón ya produce la salida esperada y sus pesos no se modifican.

Si  $g(z) \neq y_k$ , el perceptrón no produce la salida esperada, y, por lo tanto, se modifican sus pesos. Si la salida del perceptrón era 1 pero la salida esperada era 0, quiere decir que el modelo «se sobreactivó». Es decir, debería haber sido menor de lo que era para producir la salida esperada. Como se computa sumando las entradas ponderadas por los pesos, si queremos que sea menor, podemos decrementar sus pesos para que la suma ponderada sea menor. Si la salida del perceptrón era 0 pero la salida esperada era 1, quiere decir que el modelo «no se activó lo suficiente». Es decir, debería haber sido mayor de lo que era para producir la salida esperada. Como se computa sumando las entradas ponderadas por los pesos, si queremos que sea mayor, vamos a incrementar sus pesos para que la suma ponderada sea mayor.

Se repiten los tres pasos para cada ejemplo de entrenamiento hasta que se llega a un número máximo de iteraciones o hasta que el perceptrón clasifique correctamente todos los ejemplos.

La modificación de los pesos en el paso 3b consiste en hacer:

$$(w_1, w_2, \dots, w_n) \leftarrow (w_1, w_2, \dots, w_n) + \delta \cdot (y_k - \hat{y}) \cdot (x_1^k, x_2^k, \dots, x_n^k)$$

Dado que el perceptrón solo puede predecir 0 o 1, solo se puede emplear para solucionar problemas con respuesta binaria. Por ejem-

plo, lo podemos usar para detección de caras, ya que, dada una imagen, se tiene que determinar si hay una cara en la imagen o no. ¿Qué podemos hacer si queremos solucionar problemas de clasificación donde existen varias posibles salidas? Por ejemplo, en el ejemplo introducido anteriormente donde queremos clasificar un mensaje como perteneciente al departamento de hipotecas, al departamento de inversiones o que debe ser descartado como correo basura, para este tipo de problemas no es suficiente un perceptrón o una neurona artificial.

El lector puede notar que, si un perceptrón puede solucionar un problema binario, un problema con más de dos salidas se puede solucionar con más de un perceptrón. Si tenemos tres posibles salidas (hipoteca, inversión, correo basura) podemos usar tres perceptrones. El primero decide si la entrada es hipoteca o no, el segundo decide si la entrada es inversión o no y el tercero decide si la entrada es correo basura o no. Con esta estrategia, comúnmente llamada método «uno frente a todos», se puede adaptar cualquier tipo de neurona artificial o modelo que clasifica datos en clases binarias a cualquier número posible de clases.

El proceso general de la estrategia «uno frente a todos» es el siguiente:

- Para cada clase en el conjunto de clases, se le asigna un modelo binario que considera esa clase específica como la clase positiva y las demás clases se consideran como clase negativa.
- Se entrena un modelo de clasificación binaria utilizando los ejemplos de entrenamiento, donde los ejemplos de la clase positiva se etiquetan como positivos y los ejemplos de las clases negativas se etiquetan como negativos.
- Se repite el paso 2 para cada clase en el conjunto de clases, lo que resulta en la creación de múltiples modelos de clasificación binaria.
- Se juntan todos los modelos como un único modelo que, para hacer una predicción, utiliza cada uno de los modelos internos para clasificar la entrada. La clase asignada corresponde a la clase para la cual el modelo binario correspondiente devuelve la probabilidad, confianza, o puntaje más alto.

Con este algoritmo vemos que podemos usar varias neuronas para solucionar problemas con varias clases. El lector se puede preguntar si es realmente suficiente usar una única neurona para cada clase.

Resulta que, para datos complejos, una neurona no tiene la capacidad suficiente para conseguir buena calidad. La solución para ese tipo de datos consiste en agrupar neuronas en capas y conseguir modelos más ajustables.

### El perceptrón multicapa

Con el modelo que hemos visto hasta ahora de neurona artificial se pueden crear arquitecturas de neuronas conectadas entre sí más complejas. Dichas arquitecturas comúnmente se denominan «redes neuronales artificiales», o, simplemente, «redes neuronales», como hemos mencionadomás arriba. Existe un gran abanico de redes neuronales distintas, pero, sin duda, la más conocida y usada es el perceptrón multicapa.

El perceptrón multicapa, como su nombre sugiere, es una concatenación de perceptrones uno tras de otro, que configuran «capas de procesamiento». En el ejemplo de la figura 3.2 se puede apreciar un perceptrón multicapa creado combinando dos neuronas en la primera capa y una única neurona en la segunda capa. Los modelos modernos usan decenas de capas, por eso se habla de «aprendizaje profundo» (*deep learning*). La primera capa conecta cada entrada ( $x_1, x_2, \dots, x_n$ ) con una o dos neuronas. Se puede observar que cada  $x_i$  tiene dos flechas que se conectan con dos  $\Sigma$  distintos. Cada  $\Sigma$  produce una salida que genera una capa de dos valores: ( $a_1, a_2$ ). Estos dos valores son la entrada de otra neurona, donde su correspondiente  $\Sigma$  agrega las dos entradas en un único valor que produce la salida de la red  $\hat{y}$ .

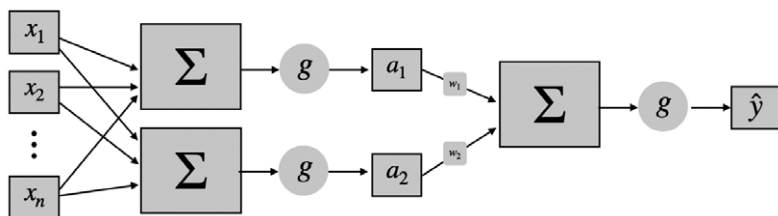


Figura 3.2. Perceptrón multicapa con una capa oculta con dos neuronas.

Estos modelos con múltiples capas requieren algoritmos más complejos para entrenarlos, pero dichos algoritmos se llevan desarrollando desde los años ochenta del siglo pasado (Rumelhart *et al.*, 1986).

## 3.2. Procesamiento del lenguaje natural

### 3.2.1. Definición

El procesamiento del lenguaje natural (PLN) es una rama de la inteligencia artificial que se dedica al estudio y creación de sistemas artificiales que funcionan con lenguaje humano. Su objetivo es permitir que los ordenadores comprendan, interpreten y generen lenguaje humano de forma efectiva.

### 3.2.2. Aplicaciones del procesamiento del lenguaje natural

El PLN tiene multitud de aplicaciones que todos usamos diariamente. Algunas aplicaciones son:

- *Chatbots*: Los *chatbots*, o asistentes virtuales, son programas con la capacidad de interactuar con personas vía lenguaje natural. Pueden buscar información, responder preguntas y ayudar a los usuarios.
- *Correctores ortográficos*: usan modelos de lenguaje (que explicaremos en detalle en la sección 3.2.3) que pueden asignar una probabilidad a una frase. En caso de tener palabras mal escritas, pueden buscar palabras correctamente escritas que con alta probabilidad tengan sentido dentro de la frase.
- *Buscadores de internet*: buscadores como Yahoo, Google, Bing o DuckDuckgo usan PLN para comprender y ordenar páginas de texto con respecto a una búsqueda efectuada por un usuario. Aunque pueda parecer fácil buscar documentos que contengan una palabra que el usuario escribe, no es nada trivial buscar documentos sin esa palabra, pero con sinónimos.



- Autocompletado de texto: buscadores de internet y webs de comercio electrónico ofrecen funciones de autocompletar a partir de una palabra que el usuario empieza a escribir. Estos sistemas proponen frases que contienen la palabra escrita para ir más rápidamente al contenido relevante para el usuario. Este tipo de aplicaciones se basan en ordenar las búsquedas más comunes y en ordenar eficientemente lo más relevante para el usuario.
- Detectores de correo basura: usan PLN para detectar los correos que son considerados no deseados y bloquearlos automáticamente.

### 3.2.3. Modelos de lenguaje

El concepto de modelo de lenguaje constituye una idea central dentro de este libro, ya que, en esencia, ChatGPT no es otra cosa que una posible implementación de estos modelos. Pese a que ha sido solo recientemente cuando ChatGPT ha alcanzado los enormes niveles de popularidad de los que goza en la actualidad, los modelos de lenguaje llevan tiempo formando parte de nuestra vida cotidiana. Si alguna vez hemos usado nuestro teléfono móvil para escribir un mensaje de texto y el dispositivo nos ha ofrecido sugerencias sobre cuál puede ser la siguiente palabra que hay que introducir, entonces podemos decir que hemos usado un modelo de lenguaje (si bien, un modelo muchísimo menos potente que ChatGPT).

El modelado de lenguaje no es en absoluto una disciplina de reciente creación, sino que para encontrar su comienzo hemos de remontarnos a la década de los años cincuenta del siglo xx. En aquella época, en la que un único computador podía ocupar una extensión de más de 30 metros cuadrados y consumir la misma electricidad que 50 hogares, un matemático llamado Claude Elwood Shannon se planteó un sencillo juego. En dicho juego, hoy conocido como «el juego de Shannon», se abría un libro por una página al azar, se tapaban todas las letras menos la primera y se pedía a una persona que adivinase la siguiente letra, pudiendo hacerse tantos intentos como fuera necesario. Una vez que la persona había dado con la letra correcta, se procedía a adivinar la siguiente (por ejemplo, imaginemos que se habían

adivinado las letras «En un lugar de la» y ahora se pedía decir cuál es la letra que aparece a continuación). El juego continuaba hasta que la persona había adivinado una cierta cantidad de letras, anotándose para cada una cuántos intentos habían sido necesarios. Como cabría imaginarse, se observaba que el número de intentos se iba volviendo más pequeño a medida que se avanzaba, debido a la información de contexto ofrecida por las letras anteriormente adivinadas.

Shannon inventó su juego con el propósito de calcular medidas de la complejidad del lenguaje, y encontró que en promedio, en un texto en el idioma inglés, una persona no necesitaba más de dos intentos para adivinar la siguiente letra. El resultado de estos estudios se publicó en un artículo (Shannon, 1951) que puso los cimientos de toda la ciencia que ha dado forma a los modelos de lenguaje que hoy conocemos, incluyendo a ChatGPT.

### Definición

El juego de Shannon resulta de gran utilidad para ilustrar el concepto de modelo de lenguaje. Imaginemos que, en lugar de adivinar el siguiente carácter de un texto, queremos adivinar la siguiente palabra. Supongamos también que, en lugar de preguntar a una persona, utilizamos un ordenador al cual proveemos las palabras previamente adivinadas, a las que nos referiremos como «contexto». Usando técnicas matemáticas, un modelo de lenguaje es capaz de determinar cuál es la palabra que tiene mayor probabilidad de aparecer teniendo en cuenta el contexto.

Expresado de un modo más formal, dado un contexto determinado, un modelo de lenguaje es capaz de asignar una probabilidad a cada palabra que pueda aparecer a continuación. Imaginemos que nuestro contexto se compone de las palabras: «Juan corre». Vamos a asumir también que a continuación solo puede aparecer o la palabra «rápido», o la palabra «lento». Por ejemplo, a la probabilidad de que «rápido» sea continuación del contexto «Juan corre», nos referiremos como:  $P(\text{rápido}|\text{Juan corre})$ .

Para determinar cuál de las dos palabras viene a continuación dado el contexto, obtendríamos la probabilidad de cada una según el modelo de lenguaje, es decir, obtendríamos:  $P$  (rápido|Juan corre) y  $P$  (lento|Juan corre), y nos quedaríamos con la palabra de mayor probabilidad.

Para poder llevar a cabo estas predicciones, primero el modelo de lenguaje ha sido entrenado mediante un proceso de aprendizaje no supervisado a partir de grandes colecciones de textos en una lengua determinada. Dicho proceso de aprendizaje tiene como objetivo configurar los parámetros del modelo. Estos parámetros configuran su comportamiento y determinan su capacidad de predicción, según vimos cuando describíamos los fundamentos del aprendizaje automático al inicio del capítulo.

Existen muchos tipos de modelos de lenguaje dependiendo de los algoritmos que estos utilizan para entrenarse y establecer predicciones. En este libro nos centraremos en dos. En primer lugar, echaremos un vistazo a los denominados modelos de lenguaje de  $n$ -gramas, que por su sencillez conceptual serán útiles para entender mejor cómo se construye un modelo de lenguaje. En segundo lugar, describiremos los modelos basados en redes neuronales, dentro de los cuales pueden identificarse diferentes variedades, que han dado lugar a una auténtica revolución dentro del campo del procesamiento del lenguaje natural debido a su extraordinario rendimiento en distintas aplicaciones.

### Modelos de lenguaje de $n$ -gramas

Los modelos de lenguaje de  $n$ -gramas son modelos cuyos parámetros se obtienen contando las frecuencias de secuencias de palabras consecutivas en textos de entrenamiento a partir de un proceso de aprendizaje no supervisado. Dichas secuencias podrán tener una longitud máxima dada por el valor de  $n$ . Habitualmente nos referimos a este valor como el «orden» del modelo. La limitación del tamaño de la secuencia a  $n$  palabras se hace para simplificar el modelo a costa de disminuir su capacidad de predicción.

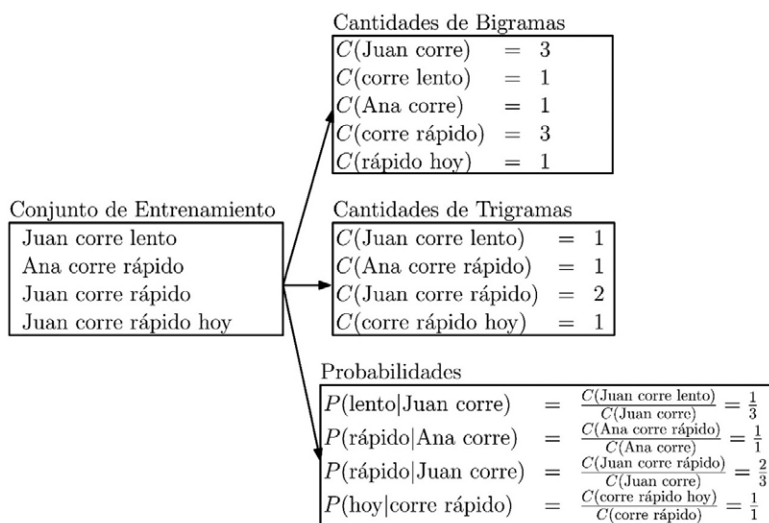
Veamos cómo funciona un modelo de  $n$ -gramas con un ejemplo. Para ello, asumiremos que el valor de  $n$  es igual 3, en cuyo caso tendremos un modelo de trigramas. Retomando el ejemplo que veíamos en el apartado anterior, un ejemplo de trigramas sería: «Juan corre rápido». Un modelo de trigramas asigna probabilidades a este trigramas según la fórmula:

$$P(\text{rápido} \mid \text{Juan corre}) = \frac{C(\text{Juan corre rápido})}{C(\text{Juan corre})}$$

En la fórmula anterior,  $C(\text{Juan corre rápido})$  representa la cantidad de veces que hemos observado en el texto de entrenamiento el trigramas «Juan corre rápido». Mientras que  $C(\text{Juan corre})$  representa la cantidad de veces que se ha observado el bigramas «Juan corre». De este modo, la probabilidad que el modelo asigna a «Juan corre rápido» no es otra cosa que la frecuencia relativa observada en el texto de entrenamiento para dicha secuencia. Para calcularla dividimos la cuenta del número de veces que aparece la secuencia (un trigramas) por el número de veces que aparece la secuencia de palabras relacionada con su contexto (un bigramas).

Dado un conjunto de frases de entrenamiento, un modelo de lenguaje de trigramas se estima obteniendo primero los contadores para todos los bigramas y trigramas que aparecen. Una vez obtenidos esos contadores, procedemos a obtener las probabilidades de los trigramas. En la figura 3.3 se muestra todo el proceso para un sencillo texto compuesto por cuatro frases.

Volviendo a la pregunta que nos planteábamos antes, dado el contexto formado por las palabras «Juan corre», ¿qué palabra vendría a continuación, la palabra «rápido» o la palabra «lento»? Según la información proporcionada por el modelo de la figura 3.3, la siguiente palabra sería «rápido», ya que la probabilidad asignada por el modelo sería igual a  $\frac{2}{3}$ , o en porcentaje, del 66 %, mientras que la probabilidad de que la siguiente palabra fuera «lento» sería igual a  $\frac{1}{3}$ , o del 33 % en porcentaje.



**Figura 3.3.** Ejemplo de estimación de un modelo de trigramas a partir de un conjunto de entrenamiento compuesto por cuatro frases.

La estrategia de modelado detrás de los modelos de  $n$ -gramas consiste en memorizar enormes tablas de probabilidades del mismo tipo que las mostradas en la figura 3.3. Dichas probabilidades constituyen los parámetros del modelo. A través de ellos quedan reflejadas las secuencias de palabras que aparecen en textos de entrenamiento, priorizando aquellas que aparecen más frecuentemente.

Los modelos de  $n$ -gramas han sido utilizados con mucho éxito en diversas aplicaciones de procesamiento del lenguaje natural. Sin embargo, presentan toda una serie de limitaciones debido a la forma en que están diseñados. Resulta interesante ahondar en dichas limitaciones, ya que la necesidad de superarlas nos ha llevado finalmente hasta modelos de lenguaje tan avanzados como el que hay tras ChatGPT.

Una de las principales limitaciones de los modelos de  $n$ -gramas es su incapacidad para generar probabilidades para una secuencia de palabras determinada, a no ser que esta aparezca en el conjunto de entrenamiento. Por ejemplo, el modelo de la figura 3.3 asignaba una probabilidad del 66 % a la palabra «rápido» como continuación del

contexto «Juan corre». Por el contrario, devuelve una probabilidad igual a cero si el contexto es «Pedro corre», ya que el conjunto de entrenamiento no contiene ninguna frase en la que aparezcan esas palabras. Esto constituye un funcionamiento altamente mejorable del modelo, ya que el contexto «Pedro corre» está íntimamente relacionado con el contexto «Juan corre» (de hecho, son equivalentes desde el punto de vista sintáctico); sin embargo, para el modelo es imposible que ocurra, ya que le asigna una probabilidad igual a cero.

Otra importante limitación de los modelos de lenguaje de  $n$ -gramas se debe al hecho de que la cantidad de palabras del contexto está limitada por el valor de  $n$ . Por ejemplo, en un modelo de trigramas, el tamaño del contexto sería de dos palabras. Esto tiene un impacto muy negativo en la capacidad del modelo, ya que habitualmente, una palabra dentro de un texto puede tener relación con otras muchas que hayan aparecido anteriormente, y no solo con las dos últimas. Se podría pensar que una forma sencilla de resolver el problema es aumentar el valor de  $n$ , pero esto no resulta fácil de llevar a cabo en la práctica, debido a que la cantidad de parámetros del modelo que se necesita almacenar crece en progresión geométrica según aumenta el valor de  $n$ , haciendo que en la práctica raramente se hayan usado modelos con  $n$  mayor que 5.

El origen principal de las limitaciones de los modelos de lenguaje de  $n$ -gramas puede encontrarse en la manera en que representan el conocimiento. El elemento básico que se ha de representar dentro del modelo son las palabras en una lengua determinada. No obstante, para los modelos de  $n$ -gramas, las palabras se representan mediante identificadores numéricos, lo que se hace únicamente por motivos computacionales: cuando se diseña una aplicación de procesamiento de lenguaje natural es más sencillo trabajar con un número que con una secuencia de caracteres (por ejemplo, en lugar de la palabra *Juan* usamos el número 10). Más aún, estos identificadores se escogen arbitrariamente, de modo que, si tomamos dos palabras semánticamente relacionadas, como, por ejemplo, «rey» y «reina», sus representaciones numéricas pueden estar muy alejadas entre sí.

Si encontrásemos una manera de obtener representaciones significativas de las palabras, de forma que, si dos palabras están cercanas entre sí, también lo estuvieran sus correspondientes representaciones, estaríamos en una mejor posición para construir mejores modelos de lenguaje. Por ejemplo, podríamos saber que «Juan corre» y «Pedro corre» son contextos relacionados o incluso equivalentes. Aprovechando esta información, nuestro modelo podría predecir la siguiente palabra dado el contexto «Pedro corre», incluso aunque estas palabras no apareciesen juntas en el texto de entrenamiento. Esto también nos permitiría tener una menor cantidad de parámetros en nuestro modelo, pudiéndose capturar relaciones entre palabras muy separadas en el texto de entrenamiento utilizando mucha menos memoria, a diferencia de lo que ocurre al aumentar el orden del modelo de  $n$ -gramas.

¿Cómo podríamos conseguir todo esto? Echemos un vistazo a los modelos de lenguaje basados en redes neuronales.

### Modelos de lenguaje basados en redes neuronales

Las redes neuronales pueden utilizarse para construir modelos de lenguaje. Como se ha explicado antes, las neuronas y conexiones que las componen se pueden estructurar de distintas formas, y dar lugar a diferentes arquitecturas. En los últimos años, algunas de esas arquitecturas se han demostrado eficaces a la hora de implementar modelos de lenguaje. En este libro veremos tres: las redes recurrentes, los modelos *seq-to-seq* y las arquitecturas de tipo *transformer*, de las que deriva el modelo que constituye la base de ChatGPT.

El formalismo matemático sobre el que se asientan los modelos de lenguaje de redes neuronales es muy sofisticado, por lo que explicarlo con detalle queda fuera del alcance de este libro. En cambio, estableceremos paralelismos con los modelos de lenguaje de  $n$ -gramas que acabamos de describir, ya que su simplicidad sí permite dar una idea clara de cómo se trabaja con ellos.

## Modelos basados en redes recurrentes

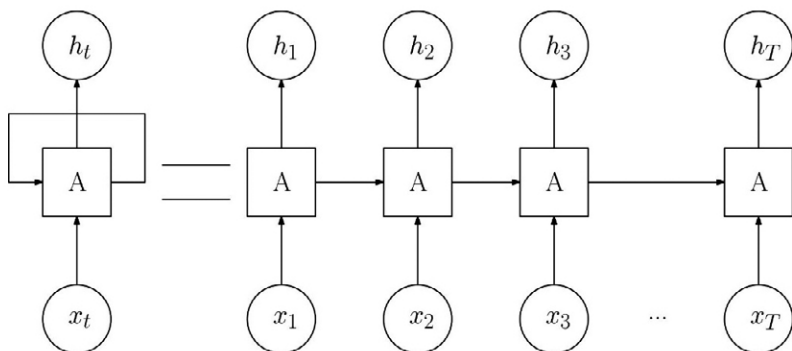
Las redes neuronales recurrentes constituyen una de las primeras arquitecturas de red neuronal usada con éxito en el modelado de lenguaje. Una red recurrente no es otra cosa que una red en la que las conexiones entre neuronas pueden formar ciclos o bucles. La salida de la red puede verse como una «codificación» de la entrada, o, en otras palabras, como una especie de representación interna de la entrada que genera la red. Dicha codificación se somete a transformaciones que permiten obtener la siguiente palabra a partir del contexto. El contexto se va introduciendo en la red en instantes de tiempo consecutivos.

La figura 3.4 muestra un ejemplo general de red recurrente. A menudo, la red recurrente se representa con los distintos instantes temporales desenrollados en el tiempo, lo que también se muestra en la figura (esto provoca que el bucle desaparezca en el diagrama). En el instante inicial, se introduce el símbolo  $x_1$  en la red y esta genera su salida,  $h_1$ , que se reintroduce en la red en el instante siguiente junto con el símbolo que aparece a continuación  $x_2$ . Al bloque marcado con la letra «A» se le denomina unidad recurrente.

Antes de continuar, es importante aclarar que tanto la entrada a la red en un instante determinado,  $x_t$ , como la representación interna que genera la unidad recurrente,  $h_t$ , son vectores numéricos (o listas de números).

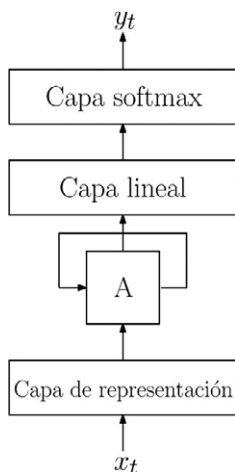
Típicamente, el vector de entrada es una lista con tantos elementos como palabras distintas tenemos en nuestro conjunto de entrenamiento. De hecho, cada elemento estará asociado a una palabra, siendo todos los elementos igual a 0, salvo el correspondiente a la palabra que ocupa la posición  $t$ , que será igual a 1. Respecto al vector  $h_t$ , se obtiene aplicando operaciones matemáticas sobre el vector de entrada. Dichas operaciones dependen de cómo haya sido implementada la unidad recurrente, y no profundizaremos en ellas aquí.





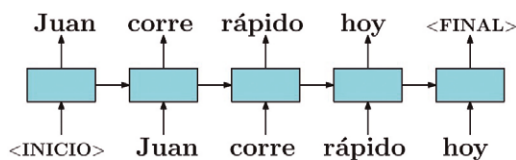
**Figura 3.4.** Ejemplo de red recurrente (izquierda) y su representación desenrollada en el tiempo (derecha).

Si se desea construir un modelo de lenguaje basado en redes recurrentes (Mikolov *et al.*, 2010), es preciso añadir algunos elementos a la arquitectura que se acaba de mostrar. Dichos elementos se muestran en la figura 3.5. Específicamente, la entrada en un determinado instante,  $x_t$ , se procesa para generar una representación interna significativa por medio de la denominada «capa de representación» (volveremos sobre esto más abajo). Esta representación se introduce en la unidad recurrente, la cual genera su salida,  $h_t$ . Dicha salida se reintroduce en la red para el siguiente instante de tiempo. Además,  $h_t$  también se procesa mediante otra capa que se denomina «capa lineal», la cual transforma  $h_t$  en un vector que asigna una puntuación a cada posible palabra que pueda aparecer según nuestro conjunto de entrenamiento. Por último, se utiliza una nueva capa, denominada «capa *softmax*», que convierte esas puntuaciones en probabilidades. Este último paso es el que nos permite predecir la palabra de mayor probabilidad,  $y_t$ , dado un contexto.



**Figura 3.5.** Modelo de lenguaje basado en redes recurrentes.

Habitualmente, el modelo de la figura 3.5 se representa desenrollado en el tiempo. La figura 3.6 muestra un ejemplo en el que el modelo predice la secuencia de palabras: «Juan corre rápido hoy». Se utiliza el símbolo «<INICIO>» para indicar al modelo que comience a generar palabras. Asimismo, el modelo utiliza otro símbolo especial «<FINAL>» para indicar que ha terminado el proceso de generación.



**Figura 3.6.** Ejemplo de funcionamiento de un modelo de lenguaje basado en redes recurrentes.

El modelo que se acaba de presentar también ajusta sus parámetros mediante textos de entrenamiento usando técnicas de aprendizaje no supervisado. Este proceso sería equivalente al que vimos para los modelos de lenguaje de  $n$ -gramas, exceptuando que ahora los parámetros que hay que ajustar son los pesos de la red neuronal, en lugar de los contadores de secuencias de palabras.

Los modelos de lenguaje basados en redes recurrentes supusieron un gran avance respecto a los de  $n$ -gramas, ya que la forma en que están diseñados permite aliviar las limitaciones que estos últimos presentan. Un elemento clave en su diseño es la capa de representación antes mencionada, que permite obtener representaciones significativas de las palabras. Considerando de nuevo el ejemplo de las palabras «rey» y «reina», la capa de representación les asignará dos vectores numéricos parecidos, reflejando el hecho de que existe una relación semántica entre ellas. Es interesante resaltar que la capa de representación consigue esto aprendiendo a partir de grandes textos de entrenamiento. La habilidad de identificar palabras y contextos relacionados supone un gran avance en la calidad de los modelos de lenguaje, pudiendo generar probabilidades para contextos que no aparecen de forma exacta en los textos de entrenamiento, obteniendo mejores modelos con menos parámetros.

Por otro lado, el bucle que caracteriza a la arquitectura de las redes recurrentes permite capturar relaciones a largo plazo entre las palabras que componen un texto. Para ello, la información del contexto se propaga a lo largo de los distintos instantes de tiempo gracias al bucle, permitiendo que, al generar cada nueva palabra, potencialmente pueda considerarse información relacionada con cualquier palabra vista anteriormente. Lo anterior contrasta con lo que ocurría con los modelos de  $n$ -gramas, que solo podían capturar las relaciones entre cada nueva palabra y las  $n-1$  anteriores.

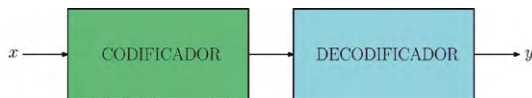
No obstante, los modelos de redes recurrentes también presentan limitaciones que continuaron impulsando la búsqueda de mejores modelos. La principal de ellas se relaciona con el coste computacional. El entrenamiento de redes neuronales para grandes conjuntos de entrenamiento es un proceso computacionalmente muy costoso. La principal estrategia para reducir este coste se basa en la paralelización, es decir, en llevar a cabo los cálculos matemáticos involucrados en paralelo, para lo cual se utiliza hardware especializado. Si últimamente se ha escuchado en los medios hablar sobre tarjetas gráficas (también denominadas *graphic processing units* o GPU) en conexión con ChatGPT o con inteligencia artificial, es debido a que estas tarje-

tas, además de servir para jugar con videojuegos, son particularmente apropiadas para efectuar en paralelo los cálculos que se requieren para entrenar modelos de redes neuronales. Sin embargo, incluso con la tarjeta gráfica más potente del mercado, el entrenamiento de una red recurrente sigue siendo extremadamente costoso, debido a que la generación de cada nueva palabra de un texto depende de la información que se deriva de generar la palabra inmediatamente anterior. Esto obliga a procesar cada palabra individualmente, lo que dificulta la paralelización.

El problema del elevado coste computacional del entrenamiento se consiguió combatir con éxito por medio de la denominada arquitectura *transformer*, que está detrás de ChatGPT. Sin embargo, antes de que fuera propuesta esta arquitectura fue preciso un paso intermedio: los modelos *seq-to-seq*.

### Modelos *seq-to-seq*

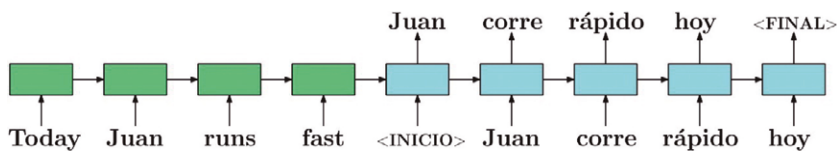
Los modelos *seq-to-seq* (del inglés *sequence-to-sequence*) se construyeron a partir de los modelos de redes recurrentes que acabamos de describir. Los modelos *seq-to-seq* supusieron la introducción de un nuevo tipo de modelo de lenguaje que se denomina «modelo de lenguaje condicionado». Igual que los modelos de lenguaje que hemos estado discutiendo, un modelo condicionado también permite predecir la siguiente palabra a partir de un contexto. Sin embargo, para hacerlo, el modelo no solo recibe el contexto, sino otra información adicional que puede definirse de distintas formas y de la que veremos un ejemplo. Los modelos de lenguaje condicionados se pueden definir mediante la denominada arquitectura «codificador-decodificador». Dicha arquitectura se caracteriza por usar dos módulos, el módulo codificador, que genera una representación interna de la entrada, y el módulo decodificador, que toma dicha representación y la usa para generar la salida. La figura 3.7 muestra un diagrama que resume lo anterior.



**Figura 3.7.** La arquitectura codificador-decodificador.

El modelo *seq-to-seq* original (Sutskever *et al.*, 2014) fue definido en el ámbito de traducción automática que se ha mencionado antes. En este ámbito, el objetivo sería traducir un texto en la lengua origen,  $x$ , al texto equivalente,  $y$ , en la lengua destino. Del mismo modo que los modelos de lenguaje necesitan grandes conjuntos de texto para entrenar sus parámetros, los modelos de traducción necesitan ejemplos de traducciones en las lenguas origen y destino.

El modelo *seq-to-seq* implementa tareas de traducción siguiendo la arquitectura codificador-decodificador, concatenando dos modelos de lenguaje basados en redes recurrentes. La figura 3.8 muestra la arquitectura del sistema utilizada para traducir la frase en inglés: «Today Juan runs fast», al castellano. En verde se representa el primer modelo de lenguaje de redes recurrentes usado como codificador; en azul, el modelo de lenguaje usado como decodificador. En primer lugar, el modelo genera una codificación de la frase de entrada con el primer modelo de lenguaje. Una vez procesada la última palabra, el símbolo especial «<INICIO>» marca el comienzo de la generación de la frase destino usando el segundo modelo de lenguaje.



**Figura 3.8.** Modelo *seq-to-seq* usado para traducir una frase del castellano al inglés.

El modelo *seq-to-seq* dio el pistoletazo de salida de las redes neuronales aplicadas al campo de la traducción automática, que hasta entonces estaba dominado por otra tecnología: el enfoque estadístico (el cual incluía entre otros elementos: los modelos de lenguaje de  $n$ -gramas anteriormente descritos).

Sin embargo, el modelo *seq-to-seq* presentaba problemas para capturar relaciones a larga distancia entre las palabras, debido al hecho de que la señal debía propagarse a través de muchos pasos en el codificador. La solución planteada supuso la introducción del denominado «mecanismo de atención» (Bahdanau *et al.*, 2015). Este mecanismo permite que, al generar cada nueva palabra, puedan tenerse en cuenta todas las generadas previamente, incluyendo las palabras generadas por el codificador. La figura 3.9 muestra un diagrama de la nueva arquitectura en la que el mecanismo de atención aparece como una capa extra de la red.

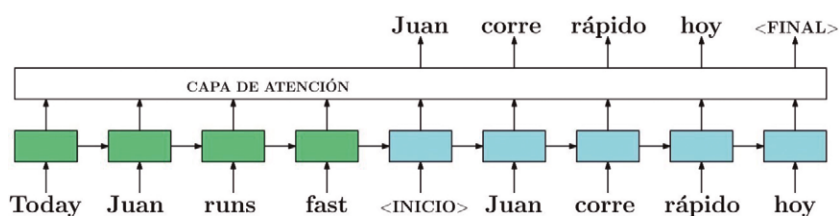


Figura 3.9. Modelo *seq-to-seq* con mecanismo de atención.

La capa de atención permite que la red neuronal aprenda a establecer relaciones entre las palabras de la frase origen y destino. En concreto, al generar cada palabra de la frase destino, el modelo podrá prestar un mayor o menor grado de atención a las distintas palabras origen. Resulta ilustrativo representar la información que maneja la capa de atención por medio de una matriz similar a la que se proporciona como ejemplo en la figura 3.10. En dicha matriz, se usan cajas en tonos de gris para representar el grado de atención, usándose un tono más oscuro cuanto mayor es la atención. Por ejemplo, cuando la red generó la palabra en castellano «hoy» en la figura 3.10, prestó algo de atención a la palabra en inglés «runs» y mucha atención a la palabra «Today».

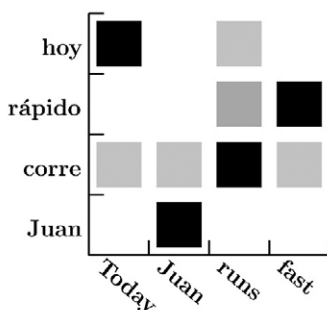


Figura 3.10. Ejemplo de matriz de atención.

El mecanismo de atención supuso un gran avance en el campo de la traducción automática; también constituye el pilar sobre el que se construye la denominada arquitectura *transformer*, que es la base de ChatGPT.

### La arquitectura *transformer*

Cuando se describían los modelos redes recurrentes, vimos que una importante limitación era el gran coste computacional del entrenamiento, el cual no podía aliviarse paralelizando los cálculos. La arquitectura *transformer* (Vaswani *et al.*, 2017) permitió reducir enormemente el problema, siendo posible trabajar con conjuntos de datos cuyo procesamiento estaba fuera del alcance de los modelos basados en redes recurrentes. Para ello, la arquitectura utiliza intensivamente el concepto de atención.

La arquitectura *transformer* sigue un enfoque codificador-decodificador como el que se veía en la figura 3.7, con la peculiaridad de que el codificador se obtiene apilando una serie de codificadores. La salida del codificador sirve como entrada para el decodificador, que a su vez también se construye apilando un conjunto de decodificadores. El artículo original apilaba seis codificadores y seis decodificadores, aunque esto varía según la implementación. Los detalles se muestran en la figura 3.11 (la arquitectura *transformer* incorpora otras capas –por ejemplo, una capa de representación–, que se omiten aquí por simplicidad).

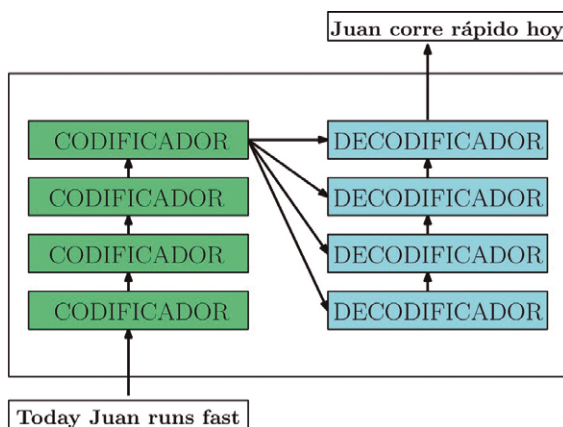


Figura 3.11. La arquitectura *transformer*.

No veremos cómo se construyen los codificadores y decodificadores, pero baste saber que van extrayendo información de atención como la mostrada en la figura 3.10. En los codificadores, la atención relaciona las palabras de las frases de entrada consigo mismas, lo que se denomina «autoatención». Por otro lado, cada decodificador genera tanto autoatención como atención codificador-decodificador (que es el tipo de atención que se mostraba en la figura 3.10).

La forma en que se define la codificación en el *transformer* permite procesar en paralelo todas las palabras de la frase de entrada, reduciendo el coste computacional y permitiendo entrenar con cantidades masivas de datos.

## GPT

En este punto, estamos por fin en disposición de hablar de la tecnología que está detrás de ChatGPT. GPT es el acrónimo de *generalized pre-training* (preentrenamiento generalizado). El modelo GPT (Radford *et al.*, 2018) se construye tomando únicamente la parte decodificadora de la arquitectura *transformer* que se veía en la figura 3.11. Más específicamente, GPT apila un total de 12 decodificadores y utiliza únicamente autoatención (es decir, elimina la atención codificador-decodificador anteriormente mencionada).



GPT es un modelo de lenguaje que permite predecir la siguiente palabra a partir de un contexto. Una peculiaridad importante de GPT es que se trata de un modelo «preentrenado», lo que significa que cuando un usuario desea trabajar con él, los parámetros del modelo han sido estimados previamente con una cantidad masiva de datos. A partir de aquí, el usuario puede simplemente usar el modelo para predecir la siguiente palabra, o bien adaptarlo para trabajar en distintas tareas de procesamiento de lenguaje natural como las que se describían en el apartado 3.2.2. El fundamento de esta modalidad de uso es que el vasto conocimiento contenido en el modelo preentrenado se transfiere para aplicarlo a otra tarea relacionada. Como se vio al inicio del capítulo, este proceso recibe el nombre genérico de aprendizaje por transferencia.

Los modelos de lenguaje basados en la arquitectura GPT típicamente se componen de una enorme cantidad de parámetros. Se suele hablar de modelos grandes de lenguaje (en inglés, LLM o *large language models*), o también modelos de lenguaje masivos, para referirnos a aquellos modelos que tienen miles de millones de parámetros.

ChatGPT es también un modelo GPT preentrenado que ha sido adaptado mediante aprendizaje por transferencia para funcionar como un *chatbot*. Veremos los detalles en las siguientes secciones.

### 3.3. ChatGPT y la era de los modelos fundacionales

GPT constituye un ejemplo de un paradigma emergente en el ámbito de la inteligencia artificial: los denominados modelos fundacionales (Bommasani *et al.*, 2021). Antes de centrarnos en ChatGPT, resulta interesante describir este nuevo paradigma en el que se enmarca.

#### 3.3.1. Modelos fundacionales

Un modelo fundacional es un modelo entrenado a partir de grandes conjuntos de datos que puede adaptarse mediante técnicas de aprendizaje por transferencia para realizar una amplia gama de tareas. La

palabra «fundacional» alude al hecho de que los modelos fundacionales son por naturaleza incompletos, debiéndose adaptar a la tarea para la que se van a utilizar.

La aparición de nuevas arquitecturas de aprendizaje profundo como GPT marcó el inicio hace pocos años de los modelos fundacionales. Estos están íntimamente relacionados con dos conceptos: la homogeneización y la emergencia (Bommasani *et al.*, 2021).

El concepto de homogeneización se relaciona con el hecho de que, de manera creciente, un único tipo de modelo se puede usar para resolver muchas tareas distintas. En el ámbito de la inteligencia artificial, la homogeneización ha experimentado un salto cualitativo desde la aparición de la arquitectura *transformer*. Los *transformers* no solo se aplican con éxito hoy en día en tareas de procesamiento de lenguaje natural, sino también en otras muchas donde se analicen secuencias de símbolos de cualquier tipo, incluyendo, por ejemplo, todo tipo de tareas de análisis de imagen o análisis de secuencias biológicas.

El segundo concepto que caracteriza a los modelos fundacionales es la emergencia, en el sentido de que el comportamiento de estos modelos emerge a partir del proceso de aprendizaje, en lugar de estar explícitamente construido. De este modo, un modelo fundacional puede mostrar capacidades que no habían sido ni siquiera previstas en el momento de su creación. Esto puede causar fascinación desde el punto de vista científico, pero también preocupación en caso de que las capacidades no previstas puedan resultar nocivas o se puedan utilizar maliciosamente.

Los modelos fundacionales tienen un elevado coste económico para su creación, razón por la cual han nacido en el seno de grandes empresas tecnológicas. Además, el coste económico conlleva un gran coste medioambiental por los requerimientos computacionales del entrenamiento.

### 3.3.2. ChatGPT

Llegados a este punto, estamos en posición de hablar sobre todo lo relacionado con ChatGPT desde un punto de vista tecnológico.

#### ¿Qué es ChatGPT?

Como ya hemos mencionado anteriormente, ChatGPT es un *chatbot* desarrollado por OpenAI, un laboratorio de inteligencia artificial con sede en Estados Unidos. La aplicación está diseñada para generar respuestas coherentes y relevantes a partir de las entradas de texto del usuario, que también reciben el nombre de *prompts*.

ChatGPT fue lanzado el 20 de noviembre de 2022; pronto se convirtió en la aplicación de software para uso doméstico de mayor crecimiento de la historia en cuanto a cantidad de usuarios, superando la barrera de los 100 millones en enero de 2023. Su enorme éxito ha impulsado el desarrollo de aplicaciones similares desarrolladas por otras empresas.

Pese a que ChatGPT ha sido el primer *chatbot* en alcanzar un éxito generalizado de cara al público, cabe señalar que la tecnología en la que está fundamentado ya existía previamente (esa tecnología es en esencia GPT, que ha sido explicada en este capítulo).

#### Proceso de implementación

ChatGPT puede enmarcarse en el paradigma de los modelos fundacionales que se ha descrito antes. En particular, utiliza un modelo grande de lenguaje basado en la arquitectura de redes neuronales GPT. La empresa OpenAI se encargó de entrenar esta arquitectura a partir de un conjunto de datos masivos, dando lugar a sucesivas versiones del modelo. Empezando por el modelo denominado GPT-1, cada nueva versión ha ido incorporando más parámetros que la anterior (OpenAI, 2023). La versión de ChatGPT, lanzada en noviembre de 2022, se basó en una variante del modelo GPT-3, conocida como GPT-3.5. Estos modelos están compuestos por 175 mil millones de parámetros. La versión de pago de ChatGPT, conocida como ChatGPT-Plus, se basa en el modelo GPT-4. Este modelo incorpora como

principal novedad la posibilidad de recibir imágenes como entrada. El tamaño en número de parámetros de GPT-4 no ha sido revelado.

Tanto GPT-3.5 como GPT-4 son ejemplos de modelos fundacionales cuya finalidad es la generación de lenguaje, y no su uso como *chatbots*. En consecuencia, para obtener lo que hoy conocemos como ChatGPT, es necesaria una fase de adaptación mediante técnicas de aprendizaje por transferencia. A esta fase de adaptación también se la denomina alineación.

A continuación, se enumeran los pasos necesarios para implementar ChatGPT (OpenAI, 2022):

- **Entrenamiento no supervisado de GPT:** el modelo GPT se entrena de manera no supervisada a partir de conjuntos de datos masivos (dando lugar a GPT-3.5 o GPT-4).
- **Entrenamiento supervisado de GPT:** un grupo de etiquetadores humanos genera manualmente contestaciones correctas para un conjunto de *prompts* (por esta razón decimos que es un proceso de entrenamiento supervisado). Esta información se utiliza para ajustar los parámetros del modelo.
- **Creación de un modelo de recompensa:** dado un *prompt*, se utiliza el modelo GPT para generar un conjunto de respuestas cuya calidad se ordena de mayor a menor por parte de etiquetadores humanos. Este proceso se repite para múltiples *prompts*, lo que permite crear un modelo de aprendizaje computacional denominado modelo de recompensa, que, dado un *prompt* y su contestación, proporciona una medida de su adecuación.
- **Uso del modelo de recompensa para ajustar los parámetros:** dado un *prompt*, se genera la respuesta con el modelo GPT supervisado. A continuación, se utiliza el modelo de recompensa para calcular su grado de adecuación. Esta información se utiliza para refinar todavía más el modelo supervisado mediante aprendizaje por refuerzo.

### Fortalezas y debilidades

ChatGPT ha demostrado una enorme versatilidad, pudiendo usarse para tareas tan dispares como responder a preguntas, corregir el uso del lenguaje, traducir textos, recomendar películas en función de los gustos del usuario, componer poesía o canciones, implementar código en distintos lenguajes de programación y un interminable etcétera que virtualmente solo está limitado por la imaginación del usuario a la hora de interactuar con la aplicación.

Sin embargo, ChatGPT presenta una debilidad fundamental que consiste en la posibilidad de generar respuestas que, pareciendo correctas, no lo son. En el contexto del aprendizaje computacional, nos referimos a este tipo de respuestas por parte de un sistema de inteligencia artificial como «alucinaciones»; en la actualidad estas constituyen un problema de investigación abierto debido a que no se ha encontrado todavía una manera de evitarlas.

Otras limitaciones de ChatGPT incluyen la posibilidad de incluir determinados sesgos en sus respuestas (por ejemplo, cuando se le pregunta acerca de personas) o el hecho de que tiene conocimiento limitado de eventos ocurridos después de 2021.

### Bibliografía

- Bahdanau, D., Cho, K. y Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *ArXiv Preprint*.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Von Arx, S., Bernstein, M. S. *et al.* (2021). *On the opportunities and risks of foundation models*. <https://crfm.stanford.edu/assets/report.pdf>
- Machinery, C. (1950). Computing machinery and intelligence-AM Turing. *Mind*, 59 (236), 433-460.
- McCulloch, W. S. y Pitts, W. H. (1917). A logical calculus of the ideas immanent in nervous activity. *Systems Research for Behavioral Science: A Sourcebook*, 5, 93-96. <https://doi.org/10.7551/mit-press/12274.003.0011>

- Mikolov, T., Karafiát, M., Burget, L., Cernocký, J. y Khudanpur, S. (2010). Recurrent neural network-based language model. *Interspeech*, 2 (3), 1045-1048.
- OpenAI. (2022). *Online ChatGPT: Optimizing Language Models for Dialogue*. <https://online-chatgpt.com/>
- OpenAI. (2023). *Models overview*. <https://platform.openai.com/docs/models/overview>
- Radford, A., Narasimhan, K., Salimans, T. y Sutskever, I. (2018). Improving language understanding by generative pre-training. *Preprint*, 12. <https://openai.com/research/language-unsupervised>
- Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton (project para)*. Cornell Aeronautical Laboratory. <https://blogs.uma-ss.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf>
- Rumelhart, D. E. y McClelland, J. L. (1986). *Parallel distributed processing: explorations in the microstructure of cognition: foundations*. The MIT. <https://doi.org/https://doi.org/10.7551/mitpress/5236.001.0001>
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, 30 (1), 50-64.
- Sutskever, I., Vinyals, O. y Le, Q. V. (2014). Sequence to sequence learning with neural networks. *27th International Conference on Neural Information Processing Systems, NIPS*, 27, 3104-3112. <https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec-1c3c743d2-Abstract.html>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. y Polosukhin, I. (2017). Attention is all you need. *31st Conference on Neural Information Processing Systems, NIPS*, 30, 5998-6008. [https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html)

## 4. INTELIGENCIA ARTIFICIAL, DESINFORMACIÓN Y ASPECTOS ÉTICOS

— Javier Guallar

— Carlos Lopezosa

### 4.1. Consideraciones sobre la IA y la desinformación

Las relaciones entre inteligencia artificial (IA) y desinformación constituyen un tema complejo, todavía sin solucionar, que suponen un gran desafío, principalmente porque cada vez es más habitual encontrar documentos en línea de todo tipo y formatos, a los que se les han aplicado alguna IA generativa para propagar contenido falso (Aïmeur *et al.*, 2023). En consecuencia, la difusión de desinformación no solo se ha incrementado, sino que su detección es cada vez más compleja.

Ante esta circunstancia, la UE ha empezado a regular los modelos de IA (The AI Act, 2023) centrándose en lo que se conoce como el «marco de calidad» y el «marco de excelencia» (UE, 2020), si bien es cierto que algunos investigadores (Li, 2023) advierten que el paradigma regulatorio europeo sobre la IA todavía debe evolucionar más para mitigar riesgos legales y éticos.

Asimismo, la Comisión Europea está trabajando en una ley que previsiblemente entrará en vigor en 2026, por la que las plataformas digitales deberán identificar los contenidos generados por inteligencia artificial para combatir la desinformación (Ayuso, 2023).

Aun así, la IA generativa, y más concretamente ChatGPT, ofrece grandes beneficios. Vale la pena señalar que, a pesar de las distintas controversias y preocupaciones que suscita, también ha despertado, en un periodo muy corto de tiempo, una notable atención desde la

academia (Lopezosa y Codina, 2023) y, por supuesto, desde todo tipo de industrias y sectores empresariales (Ray, 2023), lo que incluye las especializadas en verificación de noticias, o *fact cheking*.

Teniendo en cuenta las premisas anteriores y el hecho de que nos encontramos en las fases preliminares del proceso regulatorio europeo, en este capítulo no solo planteamos distintas opciones para identificar contenido generado por IA, sino también para identificar bulos y contenido falso generado por estas herramientas. Asimismo, listamos algunos de los principales portales web de verificación de contenido españoles para que los lectores de este libro puedan acceder a organismos que se encargan de desmentir bulos y noticias falsas.

## 4.2. IA y fuentes de información frente a los bulos y la desinformación

Como ya hemos avanzado en el punto anterior, ChatGPT tiene también usos muy positivos, e incluso puede ayudar a combatir la desinformación, si bien es cierto que esta cuestión no está exenta de retos. Así, se pueden situar estos retos, al menos en dos escenarios en los que interviene directamente la IA:

1. herramientas que ayudan a identificar contenido generado por IA, así como contenido impreciso y falso;
2. el trabajo más profundo llevado a cabo por empresas de verificación sobre el contenido realizado por IA que sirve para informar a la ciudadanía.

En cuanto al primer escenario, es importante recordar que los usuarios de ChatGPT deben utilizar esta IA con responsabilidad, ya que de ese modo se contribuye a la creación de contenido preciso. Por lo tanto, para garantizar el contenido de calidad y veraz proporcionado por ChatGPT es necesario la aplicación del pensamiento crítico, la ética y la transparencia (Lopezosa, 2023), lo que también incluye



la incorporación de fuentes de información fiables en los resultados obtenidos.

Como ChatGPT no ofrece sus fuentes de información consultadas (al menos, en la fecha de redacción de este capítulo), y resulta esencial poder acceder a ellas, una posible recomendación para docentes y estudiantes es utilizar algunas de las IA que sí ofrecen las fuentes consultadas.

En este sentido, a continuación, listamos algunas de las IA que muestran fuentes de información en las respuestas a las peticiones de los usuarios, mejorando así la credibilidad de los resultados:

- Bing Chat: el sistema de inteligencia artificial basado en ChatGPT que ofrece el buscador Bing.<sup>4</sup>
- Bard: el sistema de inteligencia artificial creado por Google.<sup>5</sup>
- Perplexity AI: un buscador conversacional que está conectado a Internet y que está desarrollado por la empresa Perplexity.<sup>6</sup>
- YouChat: un asistente de búsqueda de IA creado por You con el que se puede hablar directamente en los resultados de búsqueda.<sup>7</sup>

Sin embargo, hay una realidad: debemos ser conscientes de la existencia de una gran cantidad de contenido generado por IA que no muestra su autoría y no incluye fuentes de información que permitan comprobar la veracidad de ese contenido. Por ese motivo consideramos importante que los docentes conozcan y enseñen a utilizar a sus estudiantes las distintas herramientas disponibles para identificar contenido generado por IA.

A continuación, se muestra una tabla con las herramientas de detección de contenido de IA generativa más representativas actualmente:

4. [www.bing.com/](http://www.bing.com/)

5. <https://bard.google.com/>

6. <https://www.perplexity.ai/>

7. <https://you.com/>

**Tabla 4.1.** Herramientas de detección de contenido generado por inteligencia artificial generativa

Nombre	Información general	URL
AI Text Classifier	Herramienta de la propia empresa responsable de ChatGPT, OpenAI. Se trata de un modelo GPT capaz de predecir la probabilidad de que una inteligencia artificial haya generado un texto a partir de diferentes IA como ChatGPT.	<a href="https://platform.openai.com/ai-text-classifier/">https://platform.openai.com/ai-text-classifier/</a>
Writer	Aplicación que ayuda a escribir todo tipo de textos a nivel gramatical e incluye un apartado de identificación de texto creado por IA generativa.	<a href="https://writer.com/">https://writer.com/</a>
Copyleaks	Herramienta gratuita que es capaz de detectar contenido generado por distintas IA generativas como ChatGPT, GPT3, Human, AI & Human, y más.	<a href="https://copyleaks.com/">https://copyleaks.com/</a>
Content at Scale	Aplicación que genera contenido con IA y que a su vez cuenta con un sistema de detección de contenido en el que muestra porcentualmente el contenido creado por la IA y por el usuario.	<a href="https://contentatscale.ai/">https://contentatscale.ai/</a>
Originality.AI	Extensión de Google Chrome que es capaz de identificar si una página de un sitio web está creada por una IA.	<a href="https://originality.ai/">https://originality.ai/</a>
GPT Zero	Evalúa si un contenido está creado con IA. Según recoge su sitio web es el detector de IA número 1 del mundo con más de un millón de usuarios.	<a href="https://gptzero.me/">https://gptzero.me/</a>
Sapling	Herramienta de detección de IA que se caracteriza y diferencia de las demás porque solo se necesitan incluir 50 palabras de un texto para identificar si el contenido está hecho por una persona o por una IA.	<a href="https://sapling.ai/">https://sapling.ai/</a>
Corrector App	Aplicación que ayuda a corregir textos y mejorarlos. Cuenta también con una funcionalidad de identificación de contenido por IA. Se precisa escribir un mínimo de 300 palabras para que funcione correctamente.	<a href="https://corrector.app/">https://corrector.app/</a>
Crossplag	Herramienta de detección de contenido generado por IA. Su modelo permite identificar las partes específicas realizadas por la IA.	<a href="https://crossplag.com/">https://crossplag.com/</a>
Smodin	Permite parafrasear textos gratuitamente de hasta 1000 palabras. Además, cuenta con un detector de contenido de IA multilingüe.	<a href="https://smodin.io/es/detector-de-contenido-de-ia">https://smodin.io/es/detector-de-contenido-de-ia</a>
Plagium	Permite detectar contenido plagiado ya sea este generado o parafraseado por IA. Se trata de una herramienta basada en GPT.	<a href="https://www.plagium.com/es/ai_detector">https://www.plagium.com/es/ai_detector</a>

En general, para utilizar estas aplicaciones hay dos opciones: añadir la URL que se quiere analizar o copiar y pegar el texto que se vaya a analizar en una ventana específica para ello. Seguidamente hay que clicar en «Analizar texto», y una vez completada la acción, las herramientas confirmarán si el contenido se ha generado por una inteligencia artificial. Adicionalmente, algunas de estas aplicaciones son capaces de ofrecer el porcentaje en que el contenido está generado por una IA o por un humano, e incluso en qué partes del texto participa cada uno de ellos.

Respecto al segundo escenario, es importante tener en cuenta que las técnicas de *fact checking*, o verificación de hechos, son una de las acciones más eficaces para combatir la desinformación, los bulos y la información errónea, ya sea esta intencional o no.

Sin embargo, uno de los grandes problemas de los verificadores estriba en que la cantidad de desinformación diaria que se publica es mayor que la cantidad de contenido que se puede verificar. Por este motivo es lícito preguntarse si la IA generativa puede ser un aliado para ayudar a mejorar la eficiencia en la identificación de este tipo de contenido.

Existen estudios y autores que han investigado esta posibilidad. Por ejemplo, Hoes *et al.* (2023) realizaron un análisis sistemático para medir el rendimiento de la verificación de noticias de ChatGPT mediante el envío de 12 784 declaraciones verificadas a ChatGPT, confirmando que esta IA clasificó con precisión las declaraciones en el 72 % de los casos. Aun así, otros autores consideran que, si bien la IA de OpenAI muestra potencial para mejorar la eficiencia y la conveniencia de los procesos de verificación de información, todavía tiene limitaciones que provocan que esta herramienta deba utilizarse complementariamente para apoyar el trabajo de los expertos; por lo tanto, no puede sustituir el trabajo humano (Diakopoulos, 2023).

También se confirma que, a medida que esta tecnología avanza, podrá volverse aún más efectiva para ayudar a los usuarios, en general, y a los periodistas, en particular, para identificar este tipo de contenido fraudulento (Frackiewicz, 2023).

En este sentido, consideremos que llevar a cabo un trabajo de vigilancia de verificación en el que utilicemos, por un lado, herramientas de detección de contenido generado por IA (véase la tabla 4.1) y, por otro lado, herramientas o sitios web de verificación de hechos para conocer el contenido desmentido, puede ser un buen punto de partida para identificar noticias falsas, bulos y desinformación en general.

A continuación, se describen brevemente algunas de las herramientas y sitios web dedicados a la verificación de hechos que los usuarios pueden consultar para conocer bulos, *fake news* y desinformación en relación con la IA:

- Fact Check Explorer: se trata de una herramienta desarrollada por Google. Es un buscador que permite encontrar noticias que ya han sido evaluadas y verificadas como verdaderas o falsas.<sup>8</sup>
- Maldita.es: sitio web que recopila noticias falsas que suelen difundirse tanto por las redes sociales como por algunos medios de comunicación. No solo dicen si una noticia es falsa o no, sino que además explican de manera razonada el porqué.<sup>9</sup>
- Newtral: este sitio web sigue las técnicas específicas del *fact-checking*. Cuentan con un apartado específico de verificación que categoriza el contenido analizado según sea verdadero, verdad a medias, falso o engañoso.<sup>10</sup>
- EFE verifica: la agencia de noticias EFE cuenta con un equipo humano centrado exclusivamente en la verificación de noticias. Tiene una sección específica en su sitio web donde se chequea la veracidad de contenido de todo tipo, lo que incluye noticias sobre IA.<sup>11</sup>

Tras la descripción de los desafíos y escenarios en los que interviene la IA generativa, en el siguiente apartado hablaremos del uso ético de la IA con la intención de utilizar ChatGPT en el aula de la manera más responsable y eficaz posible.

8. <https://toolbox.google.com/factcheck/explorer>

9. <https://maldita.es/malditobulo/>

10. <https://www.newtral.es/zona-verificacion/fact-check/>

11. <https://verifica.efe.com/>

### 4.3. Retos éticos en el uso de la IA como herramienta docente para combatir la desinformación

Ante los grandes retos del uso de ChatGPT como herramienta docente en la educación universitaria, los profesores tienen el compromiso de aplicar un uso responsable de ella.

Por esta razón consideramos esencial que los docentes se adhieran a las directrices de la Comisión Europea (2020), la cual, a través de su Libro Blanco sobre IA, establece dos marcos de acción: un marco de excelencia y un marco de confianza. El primero supone conocer lo mejor posible las posibilidades que ofrece la IA (en este caso, en el campo de la docencia) y el segundo implica aprender a utilizarla de manera responsable.

Ambos marcos instan a la necesidad de un uso tanto transparente como responsable de ChatGPT que es extrapolable a la formación de los estudiantes universitarios. Por todo ello, queda claro que esta IA debe utilizarse éticamente.

Esta aplicación ética de la IA se traduce en que en cada contexto hay una implicación específica que hay que atender, en este caso, en relación con la desinformación; además de que siempre hay que ser transparente con su uso y formación, por lo que los docentes deben explicar cómo se puede utilizar con responsabilidad (Rahimi y Abadi, 2023).

### 4.4. Conclusiones

Los docentes universitarios tienen a su disposición una herramienta muy potente que puede utilizarse en las aulas para enseñar a los alumnos a identificar noticias falsas, bulos y desinformación en general. Si bien es cierto que la formación con IA deberá partir siempre de aspectos éticos y ha de complementarse con protocolos de asistencia a través de sitios web y herramientas de verificación de noticias como las vistas a lo largo de este capítulo.

Y es que el contexto actual ha demostrado el poder de la IA tanto en aspectos positivos (p. ej., identificar contenido generado por la IA, noticias falsas) como negativos (p. ej., crear noticias falsas). A pesar de que a veces se acentúe lo negativo, los beneficios académicos existen y se deben explorar.

Las propuestas metodológicas, académicas y docentes con soporte de herramientas de inteligencia artificial en general y de ChatGPT en concreto, se deben examinar con sentido crítico por parte de los docentes y se ha de transmitir ese sentido crítico a los alumnos. De hecho, el consenso entre ambas partes (docentes y alumnos) en el buen uso de la IA será esencial para el buen desarrollo de las asignaturas que utilicen la IA generativa como herramienta formativa.

En definitiva, es importante recordar que la inteligencia artificial debe utilizarse con responsabilidad; por lo tanto, su uso solamente puede ser un punto de partida, y de ningún modo podrá sustituir al trabajo desarrollado ni por el docente ni por los estudiantes.

## Bibliografía

- Alkaissi, H. y McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*2, 15 (2). <https://doi.org/10.7759/cureus.35179>
- Ayuso, S. (2023). Bruselas quiere que las plataformas digitales identifiquen los contenidos generados por IA para combatir la desinformación. *El País*. <https://elpais.com/tecnologia/2023-06-05/bruselas-quiere-que-las-plataformas-digitales-identifiquen-los-contenidos-generados-por-ia-para-combatir-la-desinformacion.html>
- Aïmeur, E., Amri, S. y Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13, 30. <https://doi.org/https://doi.org/10.1007/s13278-023-01028-5>
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B. y Love-nia, H. (2023). A multitask, multilingual, multimodal evaluation of

- ChatGPT on reasoning, hallucination, and interactivity. *ArXiv Preprint*. <https://arxiv.org/abs/2302.04023>
- Codina, L. (2022). *Cómo utilizar ChatGPT en el aula con perspectiva ética y pensamiento crítico: una proposición para docentes y educadores*. <https://www.lluiscodina.com/chatgpt-educadores/>
- Comisión Europea (2020). *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*. <https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>
- Diakopoulos, N. (2023). *Can ChatGPT help journalist's fact-check faster?* Medium. <https://generative-ai-newsroom.com/can-chatgpt-help-journalists-fact-check-faster-351e64a2ef32>
- Frackiewicz, M. (2023). *ChatGPT-4 and the future of fake news detection: AI-driven fact-checking and information verification*. TS2. <https://ts2.space/en/chatgpt-4-and-the-future-of-fake-news-detection-ai-driven-fact-checking-and-information-verification/>
- Guida, G. y Mauri, G. (1986). Evaluation of natural language processing systems: issues and approaches. *Proceedings of the IEEE*, 75 (7), 1026-1035.
- Hoes, E., Altay, S. y Bermeo, J. (2023). Leveraging ChatGPT for efficient fact-checking. *PsyArCiv Preprints*, 16. <https://doi.org/10.31234/osf.io/qnjkf>
- Li, Z. (2023). The dark side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination. *ArXiv Preprint*. <https://doi.org/https://doi.org/10.48550/arXiv.2304.14347>
- Lopezosa, C. y Codina, L. (2023). *ChatGPT y software CAQDAS para el análisis cualitativo de entrevistas: pasos para combinar la inteligencia artificial de OpenAI con ATLAS.ti, Nvivo y MAXQDA*. <http://hdl.handle.net/10230/55477>
- Lopezosa, C. (2023). ChatGPT and scientific communication: towards the use of artificial intelligence that is as useful as it is responsible. *Hipertext.Net*, 26, 17-21. <https://doi.org/https://doi.org/10.31009/hipertext.net.2023.i26.03>
- OpenAI. (n.d.). *ChatGPT*. <https://openai.com/blog/chatgpt/>

- Rahimi, F. y Abadai, A. T. B. (2023). ChatGPT and publication ethics. *Archives of Medical Research*, 54 (3), 272-274. <https://doi.org/https://doi.org/10.1016/j.arcmed.2023.03.004>
- Ray, P. P. (2023). ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://doi.org/https://doi.org/10.1016/j.iotcps.2023.04.003>



## 5. IMPLEMENTACIÓN DE CHATGPT EN EL AULA

— Mariona Grané

Chat GPT y, en general, las tecnologías de inteligencia artificial (IA) generativas representan un cambio disruptivo y acelerado en las formas de información, de comunicación y de creación.

Así, esto se convierte en un gran reto para la educación en todos los niveles educativos y de forma muy evidente en las aulas universitarias. Es un desafío perturbador que, en estos momentos todavía iniciales de estas tecnologías, genera enormes dudas entre el profesorado, pero a la vez suponen una oportunidad única para abrir nuevos caminos de aprendizaje.

El cambio profundo que implica tener estas tecnologías para su uso público y extendido no supone solamente un cambio a nivel tecnológico, implica mudanzas en los procesos comunicativos y creativos, entraña cambios de poder sobre la información e involucra el aprendizaje y, por ende, la formación, porque atañe profundamente a la construcción de conocimiento.

En este capítulo vamos a analizar las posibilidades de la herramienta en los procesos de enseñanza y aprendizaje. Pondremos el foco en experiencias iniciales de personas expertas que involucran a los procesos de enseñanza, pero también vamos a observar lo que piensan y lo que hacen con ChatGPT los estudiantes; asimismo, pensaremos en estrategias educativas, en actividades y, por supuesto, en retos y dilemas educativos.

## 5.1. Implicaciones de herramientas como ChatGPT en los procesos comunicativos, creativos y de aprendizaje

La transformación digital de la educación no solo ha avanzado en todos los niveles a lo largo de los últimos 40 años desde una perspectiva tecnológica e instrumental, sino también pedagógica.

Los educadores somos cada vez más conscientes de que el aprendizaje no es un hecho basado en contenidos que se producen solamente en un entorno de aula, dentro de una institución educativa. Cuando queremos comprender la interrelación entre las tecnologías digitales y los procesos de enseñanza y aprendizaje, no analizamos una relación bidireccional, sino que pensamos en contextos completos, en ecosistemas.

### 5.1.1. Ecosistemas tecnológicos de aprendizaje

El concepto de ecosistemas tecnológicos de aprendizaje (Wilkinson, 2000) ha evolucionado hacia la comprensión de que el aprendizaje no depende de lo que ocurre dentro de un aula, o en un centro educativo, sino que deriva de todo el entorno social, político, tecnológico, mediático, etc., en constante cambio. Hoy en día, el profesorado debemos comprender el ecosistema de aprendizaje del alumnado y entender que lo que pasa en nuestras aulas es solo una de entre tantas acciones que van a ser usadas para progresar.

Dicho esto, es necesario plantear que el cambio que nos llega de la mano de las tecnologías de IA generativa como ChatGPT es rompedor. Impregna todos los procesos sociales relativos a la comunicación, a la creación, al contenido y a la información.

Además —y esto es clave—, está siendo integrada por los usuarios a una velocidad enorme. Todavía más rápida es la evolución que las empresas tecnológicas, grandes y pequeñas han activado para desarrollar a ritmo frenético soluciones de IA en todos los ámbitos (educación, salud, política, ingeniería, periodismo...).

Y rápidas han sido las reacciones desde las instituciones educativas con una preocupación creciente, lo que ha vuelto a poner sobre la mesa el debate sobre «tecnologías digitales, sí o no».

Ante esta inicial preocupación, Mike Sharples (Open University UK) afirma que las IA generativas representan un gran desafío perturbador para la educación, pero también representan grandes oportunidades para apoyar el aprendizaje, porque son una potente herramienta en relación con el lenguaje. De hecho, como ya se ha comentado en capítulos anteriores, ChatGPT (*generative pretrained transformer*) es precisamente esto, un modelo de lenguaje de propósito general, desarrollado por OpenAI, que utiliza procesamiento de lenguaje natural. Para algunos investigadores es una de entre otras inteligencias débiles que se centran en «hacer una sola cosa» (Scolari, 2023).

Pero esta tecnología, este modelo de lenguaje, cambia nuestra forma de crear y de comunicar contenidos. Y esto es clave en nuestro día a día educador.

### 5.1.2. ¿Controlar o educar?

Desde la educación, hace años que hablamos de que no solo es necesario revisar cómo aprendemos en nuestras instituciones sino también qué aprendemos. Durante años ha habido muchos llamamientos a recalibrar la educación hacia aprendizajes y también evaluaciones más auténticas (Villaroel *et al.*, 2017). Y esto, ahora más que nunca, está sobre la mesa.

La reproducción de información hace tiempo que, afortunadamente, no es lo más importante del aprendizaje. El contenido, aunque relevante, no debería ser lo sustancial, porque siempre tenemos acceso a herramientas que reproducen contenidos. Lo principal, en cualquier caso, será la capacidad de crear, de construir conocimiento por parte de quienes están aprendiendo. Es decir, nuestro estudiante siempre tendrá acceso a informaciones, contenidos y respuestas. Y hoy las informaciones son múltiples, interesadas, orientadas y no siempre ciertas. Así que cada vez es más necesaria una educación no solo a favor del desarrollo de competencias específicas en cada área de conocimiento, sino también transversales. Seleccionar, gestionar, organizar, comprender, rebatir, fundamentar... informaciones es clave en cualquier proceso de aprendizaje.

Puede parecer una idea reiterada pero todavía hoy el mundo de la formación se preocupa especialmente por los contenidos. Un ejemplo claro de ello es que las primeras dudas de la Administración educativa y de muchos profesores y profesoras se centran en saber detectar si sus estudiantes han usado o no ChatGPT en sus tareas (Vilajosana, 2023). El problema de la originalidad de los trabajos de los estudiantes y de su integridad como creadores académicos se ha situado por delante de las oportunidades para enseñar y aprender.

Ante esta situación necesitamos un momento para pensar y cuestionarnos: ¿cuál es el trabajo del profesorado: educar o controlar? Porque focalizar en el problema puede distraernos de la oportunidad.

Y sí. Como es posible imaginar, esa es una pregunta trampa. Porque sin control tenemos un problema con la equidad en los sistemas institucionales de evaluación, pero sobre todo tenemos un problema para saber si nuestros estudiantes realmente están aprendiendo o no.

Y este es el tema: es clave poner el foco en la educación y no en el control, porque la respuesta al uso de estas herramientas no está en la tecnología (detectores de plagio y detectores de ChatGPT), sino en las decisiones educativas del profesorado en el diseño de los procesos de enseñanza y de evaluación.

Cuando el profesorado interpela a los estudiantes con una pregunta en una actividad de aula, en un trabajo académico o en un examen, puede promover la reproducción de contenidos (por ejemplo, de lo que hemos trabajado en el aula, o lo que hemos encontrado en un libro, en un artículo o una web). Son lo que llamamos preguntas reproductivas. Pero también puede lanzar una pregunta que promueva la observación, la comparación, la deducción, las relaciones, el pensamiento analítico... Saber formular las «buenas preguntas», las preguntas productivas, es una competencia necesaria para el profesorado.

Los retos y cuestiones que permiten a los estudiantes pensar, conectar, relacionar, seleccionar información e integrarla en su proceso son aquellos que les permitirán avanzar.

¿Puede ChatGPT ser un soporte para el profesorado y el estudiante ante este tipo de actividades?

### 5.1.3. Crear es aprender

Antes de responder a esta cuestión, es necesario pensar que, aparte de plantear preguntas, utilizamos las tecnologías digitales en educación desde uno de sus mayores potenciales, las posibilidades para la **creación** oral, textual, visual, audiovisual. Donde el planteamiento de proyectos y problemas era la chispa que permite pensar y crear, conectar horizontalmente diferentes áreas, conocimientos adquiridos y nueva información para construir conocimiento.

Crear es comprender, es aprender, ya nos lo demostró hace mucho tiempo Piaget (1956). Pero ante las herramientas de IA generativas como ChatGPT, nos preguntamos, si van a suponer un problema para nuestra capacidad creativa o incluso su fin.

Carrión (2023) plantea que estamos dejando de crear para convertirnos en correctores y editores de estos «algoritmos culturales y creativos». A pesar de que llevamos años creando y usando un corrector de textos, ahora somos nosotros quienes corregimos los contenidos creados por ChatGPT. Y la misma idea puede aplicarse a otras tecnologías de inteligencia artificial generativas, textuales, visuales y audiovisuales.

Hoy, cuando pedimos un texto a ChatGPT, si queremos un texto académico, nos procurará citas que parecen razonables y posibles, pero que son inexistentes. No olvidemos que es un modelo de lenguaje (con propósito general), no es una base de datos académica. Inventará aquello que seamos capaces de pedir si sabemos hacerlo de correctamente. En este momento, la tecnología parece poder crear mensajes coherentes, y correctos a nivel gramatical, sintáctico, etc., pero eso no significa que los mensajes que crea sean ciertos, ni reales, ni, por supuesto, completos.

Durante décadas, las investigaciones en el área de tecnología educativa se han esforzado constantemente en defender la necesidad del diseño didáctico y metodológico de los entornos de aprendizaje potenciados por la tecnología. A pesar de ello, seguimos asistiendo hoy al proceso de ideación y creación de escenarios de aprendizaje en línea, semipresenciales y presenciales, donde la tecnología cada vez tiene

un peso mayor en la toma de decisiones sobre el diseño. Castañeda y Selwyn (2019) lo expresan claramente cuando nos interpelan afirmando que es necesario hablar más y mejor sobre aprendizaje, y es necesario hablar más sobre pedagogía cuando hablamos de tecnología educativa.

A lo mejor estas herramientas nos fuerzan a dar realmente este paso.

## 5.2. Oportunidades para los aprendizajes con ChatGPT

Para empezar a reflexionar sobre las oportunidades que nos brindan las tecnologías de inteligencia artificial generativa como ChatGPT es necesario situarse en el lugar de los que aprenden, que, por supuesto, somos todas y todos cada día. Pero especialmente es necesario situar la mirada en lo que hacen con estas tecnologías los estudiantes de educación secundaria y superior, cómo las viven, las usan y las perciben.

Veamos tres ejemplos diferentes.

Muy sugestivo es el estudio de Marrone y colegas (2022), que formaron a 80 estudiantes de secundaria en tecnologías de inteligencia artificial generativa durante ocho semanas con la intención de que aprendieran a crear contenidos con ellas. Posteriormente, se realizaron doce grupos focales y ocho entrevistas individuales con ellos. El análisis destaca que los estudiantes pudieron comprender la relación entre la IA y la creatividad desde cuatro conceptos clave: factores sociales, factores afectivos, factores tecnológicos y factores de aprendizaje. Los estudiantes con una mayor comprensión de la IA se formaron pensamientos más positivos sobre la integración de esta en sus aulas, pero los estudiantes con una baja comprensión de la IA tendían a tenerle miedo. Además, los autores observaron que la mayoría de los estudiantes mostraron tener una comprensión profunda sobre creatividad y afirmaron que la IA nunca podría igualar la creatividad humana (Marrone *et al.*, 2022). Curioso.

Otro ejemplo interesante lo encontramos en Cu y Hochman (2023). Presentaron los resultados de una encuesta realizada a casi 5000 estudiantes de Standford, donde un 17 % de los encuestados declararon haber usado ChatGPT en sus trabajos del primer trimestre. La gran mayoría aseguró usarlo solamente para obtener ideas o esquemas de contenido para sus trabajos; solo un 5 % afirmó haber presentado trabajos escritos directamente por ChatGPT con pocos o incluso ningún cambio. Este es un ejemplo de entre muchos experimentos similares que este año encontramos en universidades de todo el mundo con resultados similares, y posiblemente serán muy diferentes el próximo curso.

Otro ejemplo lo recogemos de las entrevistas de *La Vanguardia* (Farreras, 16 de enero de 2023) a diferentes estudiantes de universidad. Es necesario situarnos en el momento de inicio del auge de ChatGPT, previo a la versión 4, y con un conocimiento inicial intuitivo de lo que les podría suponer esta herramienta. En este escenario, los estudiantes explicaron lo que hacían ellos con la tecnología. Por ejemplo, algunos aseguraron usarlo para empezar a crear, superando el vértigo de la hoja en blanco; es decir, para generar primeras ideas que luego van a ser más trabajadas, analizadas y desarrolladas. Otros afirmaron que comparaban lo que creaban ellos mismos con lo que les ofrecía un sistema como ChatGPT, buscando oportunidades de mejora de sus trabajos. También lo utilizaron para preguntar dudas concretas sobre contenidos, lo que les podría haber llevado a profundizar sobre el contenido que se está trabajando. En otros casos, lo usaban como muestra, como plantilla, por ejemplo, con una estructura para un ensayo que luego podía desarrollarse.

Quizás, sean esos unos usos muy poco alejados de lo que hemos hecho algunos profesores tanteando el sistema hasta ahora. Porque también queremos y necesitamos aprender cómo funciona, para nosotros y para la enseñanza.

### 5.3. Posibilidades para educar en un ecosistema que integra ChatGPT

El portal web Best Universities (2022) pidió a profesores de tres áreas: historia, derecho y escritura creativa, que evaluaran los trabajos de diferentes estudiantes universitarios. Entre los trabajos también insertaron algunos escritos realizados por Chat GPT para comprobar si los profesores podían detectarlo. Todos los trabajos fueron evaluados con una media parecida al resto de estudiantes, excepto en escritura creativa, donde ChatGPT fue bastante más flojo. Pero es muy llamativo que ninguno de los profesores se dio cuenta de que estaba evaluando a un *chatbot*.

Y esto es curioso porque mientras los programas informáticos son sistemáticos y lógicos, la creatividad humana es impulsiva, impredecible, espontánea y a veces incluso inexplicable (Leos, 2023). Podríamos pensar que la creatividad no es algo que pueda reducirse a fórmulas matemáticas. Pero cuando evaluamos textos de nuestros alumnos que han sido generados con ChatGPT no sabemos apreciarlo fácilmente (Universitat de Barcelona, junio 2023). Los productos que generan las IA generativas, los textos coherentes de ChatGPT, las imágenes increíbles que genera Midjourney, etc., suponen un reto para las visiones más creativas porque son buenos.

No son perfectos, no son verdaderos, no son completos... todavía.

Así que un aspecto relevante en nuestra tarea como educadores en este nuevo ecosistema tecnológico de aprendizaje es ser conscientes de la necesidad de conocer cómo funcionan estas tecnologías de IA generativa, especialmente desde la perspectiva de ser capaces de saber sus limitaciones e imperfecciones. Hace unos meses, Terwiesch (2023), profesor en la Universidad de Pennsylvania, propuso a ChatGPT resolver el examen de un curso de Gestión de Operaciones del MBA de la Wharton School. En palabras del propio profesor, la herramienta hizo un «trabajo asombroso» en las preguntas básicas, cometió algunos errores de cálculo en operaciones muy simples, tuvo problemas para resolver preguntas complejas sobre análisis de procesos, pero aprendió rápido del diálogo con el profesor. En suma, obtuvo una



calificación entre notable y sobresaliente en el examen. Con este ejercicio, Terwiesch comprobó claramente cómo funciona la tecnología y entendió sus limitaciones.

Dar la oportunidad al estudiantado de crear como parte de su proceso de aprendizaje pasa ante todo por experimentar y conocer las herramientas, buscar las oportunidades educativas y desarrollar estrategias y actividades estimulantes en este nuevo escenario.

### 5.3.1. Pensar en aprender

Clark (2020) siempre recomienda: «Piensa en aprender, no en enseñar». Estas tecnologías pueden actuar en todo el proceso educativo: seguimiento del estudiante, apoyo, *feedback* específico, aprendizaje adaptativo, creación de materiales, creación de contenidos... Poner el estudiante y el aprendizaje en el centro nos obliga a utilizar estas tecnologías desde una visión proactiva didácticamente, y no de mero control. Es entonces cuando podremos aprovecharlo.

Ante las IA generativas, Webb (2023) plantea que los docentes podemos evitarlas, intentar dejarlas atrás o adaptarnos a ellas.

Para **evitarlas** podemos dejar de pedir trabajos escritos y plantear exámenes presenciales en papel u orales. Pero los exámenes escritos no son garantía de un aprendizaje y su autenticidad ha sido puesta en duda durante muchos años.

**Superarlas**, o dejarlas atrás, implica diseñar actividades que no puedan ser realizadas por una IA generativa, o diseñar evaluaciones conectando con vivencias personales o contextos y momentos muy cercanos de los estudiantes. Pero considerando el avance de estas IA generativas es posible que esto no se pueda sustentar en el tiempo.

**Adaptarse** implica comprender que las IA son ineludibles y van a ser herramientas para los estudiantes. Es la opción más compleja desde una perspectiva pedagógica, pero la más inexcusable, porque carece de sentido que prohibamos su uso (Liu y Bridgeman, 2023). Por supuesto, entraña repensar, además, la evaluación centrándonos en el proceso, y no en el producto.

Desde estas páginas, pensamos que adaptarse es un primer paso, que el profesorado debemos tomar las riendas de estas tecnologías en nuestras aulas. Como veremos más adelante, adaptarnos no es el tema; debemos resistirnos a esta tecnologización del proceso de enseñanza e **integrarlas** para aprender.

### 5.3.2. Estrategias

Esta perspectiva de integración y adopción es clave, pues nos permite pensar en algunas estrategias que nos pueden ayudar en el aula, como las que vamos a ver a continuación.

Centremos nuestra tarea docente en el **aprendizaje**, y **no tanto en el control**, o que el control nos sirva para asegurar el aprendizaje. Si los estudiantes hacen trampa, el problema no es haber hecho trampa, el problema es no haber aprendido y no sentir la necesidad de aprender.

**Entrenémonos con ChatGPT**, hagamos experimentos, pidámosle que prepare actividades, proyectos, ejercicios, exámenes e informaciones sobre nuestra materia. Reflexionemos sobre las respuestas que nos ofrece, observemos la calidad y profundidad de sus contenidos y veamos como las integramos en nuestra actividad diaria.

**Consideremos evaluar el proceso de aprendizaje en lugar del producto** o tarea entregada. Aumentemos la importancia de las actividades en el aula (Mucharraz *et al.*, 2023), observemos y acompañemos.

Asegurémonos de **conocer el punto de partida de nuestros estudiantes** en cada materia. Cuestionemos lo que saben previamente y preparemos las acciones formativas para que puedan avanzar. Orientemos nuestras preguntas y actividades a las áreas de incomprensión y desconocimiento.

Dejemos que nuestros estudiantes piensen **diversidad de formatos para mostrar lo que han aprendido**, texto, visual, interactivo, audiovisual, etc., y animemos a trabajar los contenidos con alta calidad desde todas las perspectivas, también desde el lenguaje y las formas de comunicación.

Establezcamos sistemas de seguimiento de los trabajos de nuestros estudiantes a lo largo de todo el proceso. Esto nos permite conocer sus fortalezas y debilidades; es necesario **mantener una evaluación formativa** y orientar el *feedback* a cada necesidad específica (Gros y Cano, 2021). Y esto es sustancial, aseguremos que alumnos y alumnas actúan sobre los comentarios recibidos y mejoran sus proyectos.

Por ello, utilicemos **la evaluación como parte del proceso y como una oportunidad para aprender**, y hagamos que el estudiantado lo entienda y viva así.

Evitemos una dependencia de las IA por parte de los estudiantes que les haga perder importantes oportunidades de aprendizaje. **Disenemos tareas más contextualizadas y complejas**. Desafíemos a los estudiantes con retos que les obliguen a conectar lo que saben con nueva información y les permitan incorporar nuevos conocimientos (Herf, 2023).

**Creemos sistemas de interrogación crítica** que fuercen al alumnado a ir más allá de lo que es explícito, más allá de la información y los datos sobre un problema (Bain, 2006). Promovamos la resolución de problemas y el desarrollo del pensamiento crítico, establezcamos actividades de debate oral en las sesiones de clase, organizadas y orientadas a una finalidad concreta.

Disenemos nuestras **aulas como espacios de aprendizaje** activos, con la necesidad de participar, y estemos atentos a lo que hacen nuestros estudiantes, a momentos de su desconexión o dispersión, y respondamos ante ello (Herf, 2023).

**No infravaloremos a nuestros estudiantes**, ni las relaciones que establecen entre ellos y con el profesorado. Estas acostumbran a ser fuente de motivación y desarrollo de conocimiento. Fomentémoslas en nuestras materias. Bain (2006) nos invita a creer en los estudiantes y a decirles claramente que creemos en ellos, a motivarlos y a retarlos a superarse.

Enseñemos a los estudiantes lo que nosotros podemos hacer también con IA generativa con ChatGPT y mostremos **cómo usar estas herramientas** de una forma productiva para la docencia que nos ayude a avanzar en nuestro conocimiento.

Destapemos las **limitaciones** de estos sistemas, los errores y los inconvenientes. Mostremos ejemplos de calidad y ejemplos con problemas, e incitemos el pensamiento crítico ante lo que la IA nos ofrece (Codina, 2023).

Ayudemos al alumnado a crear *prompts* que nos permitan pedir lo que queremos a ChatGPT de la forma más efectiva posible. Sin duda, aprender a crear las preguntas adecuadas, estructurar la información, escoger las palabras perfectas... supone una nueva habilidad que desarrollar para poder aprovechar las fortalezas de estas tecnologías en nuestra vida profesional y personal. Hablamos ya del *prompting* como una nueva habilidad que hay que desarrollar por todos, entrenémonos y ayudemos a los estudiantes.

Todavía más allá, **adoptemos una actitud hacker ante ChatGPT**. La ética *hacker* llevada a la educación significa valorar la autonomía, el protagonismo y el empoderamiento de quienes usan los sistemas, también la creación y la producción colectiva y colaborativa (Pretto *et al.*, 2021). Es un proceso de conocimiento sobre cómo funcionan las tecnologías digitales desde sus diferentes ángulos y a la vez de compartir y abrir el conocimiento en este campo.

Establezcamos actividades que permitan el **trabajo colaborativo**, pero aseguremos también que existe **espacio y tiempo para el estudio personal** y el avance de cada individuo (OECD, 2010).

Creemos **situaciones que promuevan la autorregulación** del estudiantado ante su propio aprendizaje, la reflexión sobre el proceso realizado y los logros conseguidos. Preguntemos directamente sobre su proceso y su percepción de aprendizaje en los exámenes o las tareas.

Organicemos con nuestros colegas una **toma de decisiones** reflexionada y pedagógica sobre cómo enfocar el uso de sistemas de IA generativa en nuestras áreas y materias.

Por último, establezcamos también normativas, o, mejor, tomemos conjuntamente decisiones con los estudiantes sobre cómo, cuándo, de qué manera y con qué finalidad vamos a usar ChatGPT o similares en nuestras tareas. **Enfatizamos la importancia de la ética y sobre todo del aprendizaje**. Acordemos transparencia en los procesos y discutámoslos (Sabzalieva y Valentini, 2023).

### 5.3.3. Actividades con ChatGPT

¿Cómo pasamos de las estrategias didácticas a las actividades concretas?

Desde este punto de vista de adaptación, aprovechamiento, comprensión del sistema y de orientación de la enseñanza hacia construcción de conocimiento, nos aventuramos a idear algunas actividades (entre muchas otras posibles), a continuación, que pueden ayudarnos a crear espacios para aprender con ChatGPT.

- A. Le podemos pedir a ChatGPT que nos prepare un examen de cinco, diez o más preguntas de respuesta múltiple sobre el tema que queramos abordar en el aula y, por ejemplo, que nos cuele una pregunta trampa o con errores. ¿Podrán los estudiantes resolver el examen? ¿Podrán detectar la pregunta trampa? ¿En qué momento vamos a decirles que el examen fue generado por ChatGPT?
- B. Solicitamos que cree un ensayo sobre un tema de trabajo en el aula y pedimos a nuestros estudiantes leer paralelamente diversas fuentes y realizar un análisis en grupo de la respuesta emitida por la IA, como si fuéramos un tribunal de expertos que evalúan una tesina.
- C. Pedimos a la IA que cree un texto académico y el alumno ha de llevarlo más allá. Debe **verificar** antes de nada la información generada por Chat GPT; **fundamentarla** con fuentes fiables, investigaciones previas o conceptos acordados por la academia, y **profundizar**, a partir de esta fundamentación, ampliando el contenido generado por ChatGPT (Codina, 2023).
- D. También podemos pedir al alumnado que resuma un texto académico generado por ChatGPT y sintetice las ideas clave, o incluso proponer tareas concretas para contraargumentar, buscar alternativas, rebatirlas..., y así llegar a crear nueva información más allá de la devolución de la IA.
- E. También podemos emular conversaciones para la práctica y desarrollo de las habilidades lingüísticas o comunicativas, incluso

para ejercicios dialógicos o de debate de un tema, y comprobar el nivel de conocimiento que tenemos.

- F. Pedimos al sistema que debe crear un esquema sobre el tema que estamos trabajando, y que nos ayude a estudiar y a conectar ideas.
- G. Planteemos una pregunta al grupo clase. La mitad de los estudiantes tendrá la consigna de usar la IA para obtener una respuesta breve. La otra mitad deberá obviarla. Una vez tengamos las respuestas, las podemos mezclar y repartirlas de nuevo para que cada una sea evaluada por un compañero. ¿Cómo son valoradas las creadas por los estudiantes y las creadas por ChatGPT? ¿Las generadas por ChatGPT son todas iguales o son diversas? ¿Somos capaces de detectar qué textos han sido generados por ChatGPT y cuáles no? (Universitat de Barcelona, junio de 2023).
- H. Utilicemos ChatGPT como si fuera un aprendiz de nivel inicial en un tema y pedimos a nuestros estudiantes que lo entrenen para que vaya aprendiendo más y de forma más fundamentada y apoyada en evidencias empíricas.
- I. Para practicar con análisis estadístico podemos pedir a ChatGPT que nos genere datos sintéticos de una supuesta muestra que ha realizado un supuesto examen y ha obtenido puntuaciones diversas. Este es un experimento que realizó el profesor Cesar Poyatos (UAM). Muy interesante sería pedir a los estudiantes que a partir de los datos planteen preguntas de investigación concretas que puedan resolverse mediante el análisis numérico.
- J. Pedimos a ChatGPT que nos cree un artículo con referencias y citas sobre el tema de interés en el aula, y a los estudiantes que, una vez hayan leído el artículo, emitan un juicio de valor sobre el contenido. Una vez hecho esto, sería interesante que comprobemos, una a una, las referencias bibliográficas que ChatGPT ha aportado. ¿Cuántas de estas existen realmente?
- K. Si estamos trabajando sobre un proyecto creativo, la creación de un producto, de una *app*, de una empresa, de un artefacto o de un dispositivo, podemos pedir a ChatGPT que nos proponga una lista de diez a veinte posibles nombres para nuestro

proyecto, producto o artefacto. Luego, en grupo, debemos tomar la decisión de cuál vamos a usar. O podemos inventar nombres diferentes usando como inspiración los que la IA nos ha proporcionado.

- L. Para empezar una composición, podemos pedir que nos escriba el inicio de una historia o un breve cuento sobre el que trabajar y crear algo más completo, como una novela, un poema, el guion de una película o el de un videojuego.
- M. En esta misma línea invitamos al sistema a plantear preguntas de comprensión lectora sobre textos con contenidos complejos en el momento de cerrar una unidad temática en la asignatura y veamos si los estudiantes han asimilado lo trabajado y pueden responder con sentido.
- N. Uno de los procesos claves que debemos hacer al usar ChatGPT es pedirle que adopte un rol específico. Por ejemplo: «eres una estudiante avanzada de física cuántica», o «eres un experto en análisis cualitativo en sociología», o «eres un guía turístico», o «eres una auditora de cuentas»... Esto es también el punto de arranque de algunas actividades: el indicar que adopte un rol concreto para resolver algún problema que tenemos en clase y que el grupo todavía no sabe resolver. ¿Son sus propuestas adecuadas? ¿Factibles? ¿Promueven que los estudiantes generen nuevas ideas para la resolución de la problemática?
- Ñ. Por supuesto, podemos obtener más que texto. ChatGPT puede ser un copiloto para los estudiantes cuando están aprendiendo a programar en un nuevo lenguaje, o cuando deben resolver problemas científicos, matemáticos, físicos, de química, etc.
- O. Podemos pedir a nuestros estudiantes de –por ejemplo– física que resuelvan un problema centrado en un tema que todavía no han trabajado, o de un curso superior. Les pediremos que lo intenten resolver individualmente. Si no lo consiguen, les concretamos que utilicen ChatGPT para ello, pero es importante pedirles que expliquen paso a paso su resolución. Luego los agrupamos en equipos de tres o cuatro componentes y les pedimos que comprueben si todos han obtenido las mismas respuestas de

la IA. Y más importante: ¿han comprendido como se resuelve el problema?, ¿han sido capaces de interpretar lo que propone la IA sin tener el conocimiento previo?, ¿sabrían resolver otro similar?, ¿necesitan aprender o comprender algo antes de poder resolver otro problema igual?

- P. Si queremos obtener todas las potencialidades de un programa informático –por ejemplo, Excel–, la IA puede generarnos funciones. ¡Ah!, pero siempre que sepamos dar los datos necesarios y formular las preguntas correctamente. Un reto para todos, ¿no?
- Q. En este sentido podemos aprovechar las sesiones en las que trabajamos el análisis estadístico con nuestros estudiantes para que comprendan qué pruebas son necesarias para que tipo de preguntas y según qué agrupación de datos. Entreguemos a los estudiantes unas tablas con datos y diferentes preguntas relacionadas. Pidamos que cuestionen a ChatGPT qué pruebas deben realizar para conseguir los resultados que la profesora solicita. Observemos si comprenden las respuestas de la IA, si son capaces de realizar las pruebas y de obtener los resultados. Y veamos, además, si las pruebas propuestas por ChatGPT eran acertadas; esto debemos resolverlo con todo el grupo clase.
- R. Desde una perspectiva investigativa diferente, usemos la transcripción de una o varias entrevistas a participantes de una investigación –transcrita con alguna herramienta de IA generativa como Auris AI– y pidamos a ChatGPT que extraiga las dimensiones o categorías sobre el discurso analizado.
- S. Ordenemos a ChatGPT que diseñe un examen de la materia y el nivel que estamos trabajando, y nos ponemos a prueba; incluso, indiquemos a ChatGPT que se convierta en un formador adaptativo para enseñar algún proceso complejo o concepto difícil que sirva información y actividades cada vez más y más complejas para llegar a adquirir el conocimiento o la habilidad. Un ejemplo podría ser empezar a aprender un nuevo idioma, o la aplicación de una prueba estadística, o...
- T. Solicitemos la biografía de una persona famosa en nuestra materia de estudio y comparemos el texto que genera ChatGPT con



la información que nos ofrece Wikipedia. Reflexionemos sobre las diferencias y las omisiones en los textos obtenidos.

- U. Teniendo un trabajo de ensayo de los estudiantes, les proponemos que pidan a ChatGPT que escriba un guion para un vídeo conceptual sobre el tema trabajado. Luego los estudiantes deberán leer, mejorar el guion y preparar el video para una exposición del tema. Aunque también podrían usar alguna IA generadora de vídeos conceptuales (como Lumen5). El profesorado observamos y aprovechamos para aprender a hacerlo también.
- V. Establezcamos una lista de preguntas en grupo sobre un dilema que queremos discutir en clase, una situación adecuada para un debate al iniciar un tema o al finalizarlo; por ejemplo, «¿cómo afectan los «nómadas digitales» a las grandes ciudades europeas del mediterráneo?» –por decir algo–. Lancemos la pregunta a la IA y organicemos un debate en el aula a partir de la respuesta obtenida.
- W. También podemos hacerlo a la inversa. Leemos un breve texto sobre un tema controvertido en nuestro campo de estudio; por ejemplo, sobre políticas de salud pública, sobre derechos de autor en tiempos de inteligencia artificial generativa, sobre la muerte asistida, sobre políticas de inmigración... Pidamos a ChatGPT que elabore un listado de preguntas o de desafíos relativos a este tema. Trabajemos los desafíos en pequeños grupos: ¿Son interesantes o banales? ¿Promueven nuestro pensamiento crítico? ¿Qué visión ideológica y política hay tras las propuestas? En la puesta en común con toda la clase recojamos, además, nuevas dudas, inquietudes o desafíos que se nos hayan ocurrido al intentar responder a las cuestiones de ChatGPT.
- X. Para promover la autorregulación del propio aprendizaje, podemos pedir a los estudiantes que entreguen sus textos o creaciones a ChatGPT y le pidan una valoración académica. Esta misma puede compararse con una evaluación entre iguales o la que ha llevado a cabo el profesorado. No olvidemos reflexionar sobre la comparativa de *feedback* obtenida. Todos, estudiantado y profesorado.

Hasta aquí algunas ideas.

El objetivo de este apartado era, simplemente, inspirar actividades a todos los docentes desde sus áreas para aprovechar las herramientas de IA generativas como una oportunidad. Una buena propuesta será también pedirnos a nosotros como profesores y a nuestros estudiantes elaborar una lista de buenas actividades con esta IA que nos ayuden a aprender, a mejorar y a optimizar nuestros proyectos, ponerla en común y obtener ideas para sesiones desafiantes que promuevan el aprendizaje.

## 5.4. Desafíos y dilemas incipientes en educación

La adopción de estas tecnologías de IA como ChatGPT en educación puede ser, efectivamente, una oportunidad, pero está condicionada por los diferentes contextos pedagógicos, culturales e institucionales.

Farrow (2023) plantea que es necesario cuestionar nuestra tolerancia a efectos que pueden ser potencialmente perjudiciales en nombre del progreso, siguiendo las líneas que antes Selwyn (2019) ya había considerado desde su –siempre– perspectiva crítica de las tecnologías en educación. Los autores exponen, entre otros, problemas de percepción sobre la IA, falta de competencias digitales del profesorado, problemas de diseño y usabilidad de los sistemas, y, significativamente, problemas éticos.

Porque, sin duda, el más relevante de los lados oscuros de la IA son las repercusiones éticas. Estas han sido tratadas en el capítulo anterior, pero aquí quisiéramos apuntar brevemente cuáles son las repercusiones éticas específicas de los procesos educativos. Porque la recomendación de la Unesco sobre la ética de la IA, adoptada en 2021, ha sido reivindicada este 2023 por Audrey Azoulay (directora de la Unesco) ante lo que ella percibe como el «gran reto de nuestro tiempo».

Para empezar, consideremos **la ética en la creación de contenidos** de todo tipo por parte de profesores y estudiantes. Según Cotton *et al.* (2023), son necesarias nuevas políticas y estrategias para utilizar ChatGPT de forma ética y responsable. Un importante objetivo es

asegurar la honestidad e integridad académica, pero de todas las personas; el profesorado debemos tomar nota.

Además, está **la ética de las grandes corporaciones *ed-tech***, que cada vez tiene más fuerza sobre nuestras instituciones educativas y, por ende, una fuerte influencia sobre el profesorado y los estudiantes. La creciente plataformización de nuestros procesos y sistemas de enseñanza y aprendizaje nace de un internet que se ha convertido en un entorno global basado en el capitalismo de datos. Para Adell *et al.* (2018) y para Laanpere *et al.* (2014) es especialmente relevante asumir el debate ético-político-pedagógico, puesto que habitamos hoy en una sociedad globalizada, polarizada, altamente tecnológica y basada en el negocio de los datos, donde las grandes compañías *ed-tech* tienen un acceso casi ilimitado a la información existente que se genera diariamente. El uso de nuestros datos se convierte también en nuevo recurso para la propia IA. Para Crawford (2021), la IA no es un conjunto de simples herramientas tecnológicas, sino que supone una infraestructura de poder tecnocrático que lo abarca todo. Para la autora nos encontramos ante una situación crítica y debemos cuestionarnos a qué intereses sirve la IA que hoy «consumimos» y —añadiría— «alimentamos».

Por si esta idea no fuera ya, en sí misma, compleja, todavía debemos agregar el **concepto de la automatización de la educación** (Selwyn *et al.*, 2023) a los dilemas educativos. Es uno de los aspectos emergentes más importantes en la digitalización de la educación, que cobra una fuerza todavía mayor con la irrupción de las IA generativas en los centros educativos. Los autores nos invitan a mantener un debate permanente sobre lo que debería ser la inteligencia artificial en educación, sobre qué beneficios puede llegar a aportar y estamos dispuestos a ceder. Sin caer en un «pesimismo tecnológico», Selwyn *et al.* (2023) invitan a los educadores a adoptar una postura activa. Tomar parte en estos procesos implica no dejar que la tecnología en educación sea solo una cuestión que pertenece a los ingenieros; los educadores, más que «adaptarnos», como hemos comentado, **tenemos que resistir**. Y resistir no significa negarnos al progreso tecnológico, sino asesorarse fehacientemente. Conocer las IA y cómo funcionan

significa participar de las decisiones sobre ellas en nuestro contexto; significa usarlas para promover el avance del conocimiento y no su pérdida. Para el profesorado es necesario comprender cómo funcionan estos sistemas, abordarlos desde la didáctica y observar cómo afectan a nuestros procesos de enseñanza y aprendizaje para poder tomar decisiones pedagógicas y de política educativa. Seamos activistas para la educación y el aprendizaje.

Otro dilema importante desde la perspectiva del uso de IA generativa en educación se relaciona con la **brecha digital**. Algunas alertas sobre la **equidad y los sesgos** de estas tecnologías deben ser tenidas en cuenta (Bergin *et al.*, 2023). Aquellas personas con menos capacidades y menos posibilidades pueden verse desfavorecidas en la creación de contenidos con ChatGPT y herramientas similares. Quienes tienen más edad o mantienen un perfil bajo en tecnologías, las minorías raciales y étnicas, los grupos con lenguas minoritarias, etc., probablemente van a obtener menos beneficios en la generación de contenido con estas IA.

Hay otro desafío importante, menos estudiado todavía: considerar si ChatGPT, y la tecnología generativa, permite o no desarrollar el pensamiento crítico y si **limita nuestras capacidades reales para la resolución de problemas**. Desde la educación, no podemos evitar cuestionarnos qué aprendizajes dejarán de ser necesarios en nuestros procesos y qué habilidades perderemos con el uso de estas IA generativas. ¿Vamos a sufrir «atrofias mentales» de comprensión y reflexión? ¿Podemos evitarlo?

## 5.5. Consideraciones finales, educar sí o sí

El profesorado es el facilitador entre los medios, los contenidos, las informaciones, las herramientas y los estudiantes. Así que es nuestra tarea ayudarles a usarlas de forma apropiada (creativa, técnica, educativa y éticamente). Debemos preparar a los estudiantes para los desafíos en este futuro inmediato, ayudarlos a desarrollar su capacidad

crítica, comparando, usando herramientas diversas, conociendo sus limitaciones..., enfatizando sobre los aspectos éticos.

Si queremos que nuestros estudiantes aprendan, no podemos seguir pidiéndoles que reproduzcan informaciones, que escriban memorias de 30 páginas con Arial 11. Es preciso darles la oportunidad de pensar y construir, y la tecnología debe ser una herramienta que les ayude a trascender, a aumentar su conocimiento.

Pero mientras sigamos pensando que los estudiantes deben superar exámenes, pruebas o trabajos, estas herramientas serán un problema. Un problema que debemos «controlar».

Cuando empecemos a pensar en grande, a ponderar aquello que nuestros estudiantes deben aprender; cuando nos planteemos qué significa aprender en nuestras áreas, quizás entonces podamos exprimir las y verlas desde una perspectiva que nos ayude a aumentar la capacidad de crear y construir conocimiento.

Aislar nuestras aulas y a nuestros estudiantes –o a nosotros mismos– de esta tecnología, no solo es inverosímil, sino que en ningún caso supone una solución ni una ayuda en el proceso de construcción de conocimiento.

Debemos ser conscientes de que se está produciendo un cambio de relaciones entre los humanos como sujetos y los objetos tecnológicos (Scolari, 2023), donde nos situamos poco a poco cada vez más lejos del centro del mundo.

La comprensión de este cambio implica también desarrollar el conocimiento de cómo actuar. El conocimiento instrumental y el dominio tecnopedagógico ya no son suficientes para diseñar escenarios educativos soportados por las tecnologías. Hoy es necesario considerar los contextos sociales, económicos, ideológicos y políticos que envuelven el proceso educativo y tomar conciencia de cómo dichos contextos determinan la tecnología y los procesos que esta permite. Es forzosa una perspectiva ética y crítica ante el uso de tecnologías de IA generativa, la ingente cantidad de datos que utilizan y la ética y privacidad detrás de ello. Porque las tecnologías no son las que cambian los entornos donde aprendemos, sino los recursos y herramientas

que decidimos implementar, cómo lo hacemos y al servicio de qué intereses responden dichas tecnologías (Adell *et al.*, 2018). Es parte de nuestra tarea docente en nuestras instituciones tomar decisiones para un uso ético, responsable y equitativo de la IA en todas las facetas del trabajo de enseñanza y aprendizaje.

Porque, para los educadores, lo más interesante de las tecnologías, en general, no es lo que pasa con ellas o cómo funcionan, sino cómo las personas interactuamos con ellas, qué hacemos, qué pensamos y qué somos capaces de construir.

## Bibliografía

- Adell-Segura, J., Castañeda, L. y Esteve-Mon, F. M. (2018). ¿Hacia la ubersidad? Conflictos y contradicciones de la universidad digital. *RIED. Revista Iberoamericana de Educación a Distancia*, 21 (2), 51-68. <https://doi.org/https://doi.org/10.5944/ried.21.2.20669>
- Bain, K., Company Gimeno, S. y Torcal, A. (2006). *El que fan els millors professors universitaris*. Universitat de València. [https://puv.uv.es/el-que-fan-els-millors-professors-universitaris.html?\\_\\_\\_store=espanyol&\\_\\_\\_from\\_store=valencia](https://puv.uv.es/el-que-fan-els-millors-professors-universitaris.html?___store=espanyol&___from_store=valencia)
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B. y Love-  
nia, H. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *ArXiv Preprint*. <https://arxiv.org/abs/2302.04023>
- Bergin, N., Huang, A., Ravinutala, A. y Shrimali, R. (2023). Generative AI. *Dichotomies*, 2. <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-gen-ai-dichotomies.pdf>
- Carrión, J. (2022). El nuevo rol de la humanidad. *La Vanguardia*. <https://www.lavanguardia.com/cultura/20221002/8550713/nuevo-rol-humanidad.html>
- Castañeda, L. y Selwyn, N. (2019). *Reiniciando la universidad: buscando un modelo de universidad en tiempos digitales*. OC. <https://www.editorialuoc.com/reiniciando-la-universidad>

- Clark, D. (2020). *Artificial intelligence for learning: how to use ai to support employee development*. Kogan Page.
- Codina, L. (2023). *Cómo utilizar ChatGPT en el aula con perspectiva ética y pensamiento crítico: una proposición para docentes y educadores*. <https://www.lluiscodina.com/chatgpt-educadores/>
- Costello, E. y Brown, M. (2023). Will artificial intelligence be able to write my college essay? *Exploring Ed TECH Ireland*, 6, 16-18. [https://www.exploringedtech.ie/\\_files/ugd/78aa6f\\_cf763dc4889644e-2abf34d6150f54c3b.pdf](https://www.exploringedtech.ie/_files/ugd/78aa6f_cf763dc4889644e-2abf34d6150f54c3b.pdf)
- Cotton, D. R. E.; Cotton, P. A. y Shipway, J. R. (2023). Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 13. <https://doi.org/10.1080/14703297.2023.2190148>
- Cu, M. A. y Hochman, S. (2023). Scores of Stanford students used ChatGPT on final exams, survey suggests. *The Stanford Daily*. <https://stanforddaily.com/2023/01/22/scores-of-stanford-students-used-chatgpt-on-final-exams-survey-suggests/>
- Farrow, R. (2023). The possibilities and limits of XAI in education: a socio-technical perspective. *Learning, Media and Technology*, 1-14. <https://doi.org/https://doi.org/10.1080/17439884.2023.2185630>
- Gros Salvat, B. y Cano García, E. (2012). Procesos de *feedback* para fomentar la autorregulación con soporte tecnológico en la educación superior: revisión sistemática. *RIED. Revista Iberoamericana de Educación a Distancia*, 24 (2), 107-125. <https://doi.org/https://revistas.uned.es/ihttps://doi.org/10.5944/ried.24.2.28886>
- Herft, A. (2023). *A teacher's prompt guide to ChatGPT aligned with «what works best»*. <https://usergeneratededucation.wordpress.com/>
- Hu, K. (2023). *ChatGPT sets record for fastest-growing user base: analyst note*. Thomson Reuters. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
- Laanpere, M., Pata, K., Normak, P. y Põldoja, H. (2014). Pedagogy-driven design of digital learning ecosystems. *Computer Science and Information Systems*, 11 (1), 419-442. <https://doi.org/https://doi.org/10.2298/CSIS121204015L>

- Leos, D. (2022). Is AI a risk to creativity? The answer is not so simple. *Entrepreneur*. <https://www.entrepreneur.com/science-technology/is-ai-a-risk-to-creativity-the-answer-is-not-so-simple/439525>
- Liu, D. y Bridgeman, A. (2023). *ChatGPT is old news: how do we assess in the age of AI writing co-pilots?* Teaching @ Sydney. <https://educational-innovation.sydney.edu.au/teaching@sydney/chatgpt-is-old-news-how-do-we-assess-in-the-age-of-ai-writing-co-pilots/>
- Mucharraz y Cano, Y., Venuti, F. y Herrera Martinez, R. (2023). ChatGPT and AI text generators: should academia adapt or resist? *Inspiring Minds*. <https://hbsp.harvard.edu/inspiring-minds/chatgpt-and-ai-text-generators-should-academia-adapt-or-resist>
- Piaget, J. (1952). *The origin of intelligence in the children*. International Universities. <https://doi.org/https://doi.org/10.1037/11494-000>
- Pretto, N. de L., Lapa, A. B. y De Espíndola, M. B. (2021). Conexão escola-mundo: espaços inovadores para formação cidadã. *Perspectiva. Revista do Centro de Ciências de Educação*, 30 (3), 13-14. <https://doi.org/https://doi.org/10.5007/2175-795X.2021.e83401>
- OECD (2010). *The nature of learning: using research to inspire practice*. <https://www.oecd.org/education/ceri/thenatureoflearningusingresearchtoinspirepractice.htm>
- Sabzalieva, E. y Valentini, A. (2023). *ChatGPT e inteligencia artificial en la educación superior: guía de inicio rápido*. [https://unesdoc.unesco.org/ark:/48223/pf0000385146\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000385146_spa)
- Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A. y Perrotta, C. (2023). Making sense of the digital automation of education. *Postdigital Science and Education*, 4, 1-14. <https://doi.org/https://doi.org/10.1007/s42438-022-00362-9>
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education (digital futures)*. Polity.
- Scolari, C. A. (2023). *10 tesis sobre la IA*. Hipermediaciones. <https://hipermediaciones.com/2023/05/14/10-tesis-sobre-la-ia/>
- Terwiesch, C. (2023). *Would ChatGPT3 get a Wharton MBA? A prediction based on its performance in the operations management course*. <https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Christian-Terwiesch-Chat-GTP.pdf>



- Universitat de Barcelona (2023). *Com detectar textos escrits per ChatGPT*. Blog TACTIC. <https://www.ub.edu/docenciaicrai/Blog-TACTIC/com-detectar-textos-escrits-chatgpt>
- Unesco. (2022). *Recommendation on the ethics of artificial intelligence*. <https://doi.org/SHS/BIO/PI/2021/1>
- Vilajosana, M. (2023). ChatGPT a les escoles: via lliure al plagi o classes més innovadores? El món de l'educació cerca vies per detectar textos creats amb intel·ligència artificials mentre algunes veus opten per integrar-la en la formació. *Metadata. El Digital de Tecnologia en Català*. <https://www.metadata.cat/reportatge/2907/chatgpt-escoles-via-lliure-plagi-classes-mes-innovadores>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C. y Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43 (5), 840-854. <https://doi.org/https://doi.org/10.1080/02602938.2017.1412396>
- Webb, M. (2023). *A generative AI primer*. <https://nationalcentreforai.jisc-involve.org/wp/2023/05/11/generative-ai-primer/#3-1>
- Wilkinson, D. L. (2002). The intersection of learning architecture and instructional design in e-learning. En: J. R. Lohmann y M. L. Corradini (eds.). *E-technologies in engineering education: learning outcomes providing future possibilities*. Engineering Conferences International. <https://dc.engconfintl.org/etechnologies/33>
- Zhai, X. (2022). ChatGPT user experience: implications for education. *SSRN*, 18. <https://ssrn.com/abstract=4312418> or <http://dx.doi.org/10.2139/ssrn.4312418>



## 6. CHATGPT, IMPLICACIONES PARA LA SOCIEDAD Y LA EDUCACIÓN

— Los autores

La inteligencia artificial (IA), encabezada por ChatGPT como uno de sus principales exponentes en la actualidad, es una tecnología disruptiva, destinada a transformar nuestra sociedad en su totalidad. Aunque aún nos encontramos en las etapas iniciales de su implementación, no resulta descabellado pensar que la IA tendrá un impacto equiparable al de otros grandes inventos en la historia de la humanidad, como puede ser la electricidad o internet, como se introdujo en el capítulo 1 del libro. La IA está desatando una revolución que puede compararse con la revolución industrial debido a su potencial para transformar la economía, la educación y automatizar el trabajo; también por las preocupaciones e incertidumbres derivadas del profundo impacto que esta automatización tendrá en el empleo y la fuerza laboral de la sociedad del futuro.

Esta revolución afecta a todos los aspectos de nuestra vida y tiene consecuencias tanto en las tareas cotidianas (por ejemplo, leer el correo electrónico) como en la toma de decisiones médicas, financieras y económicas. Por lo tanto, no es suficiente contar únicamente con un equipo técnico sólido y las infraestructuras computacionales adecuadas. Como se explica en el capítulo 3, donde se introdujo la tecnología de IA, el desarrollo de esta tecnología requiere equipos multidisciplinarios que validen y faciliten su integración óptima con el usuario final. Por ejemplo, en el desarrollo de la tecnología de los grandes modelos de lenguaje actuales como es el caso del ChatGPT, es necesario contar con lingüistas que analicen los textos y filósofos

que planteen cuestiones éticas para dar respuestas a los posibles escenarios en los que se pueda utilizar dicha tecnología.

La IA se distingue de los avances ocurridos durante la revolución industrial en el sentido de que su poder de automatización está relacionado con tareas cognitivas en lugar de tareas manuales. Este aspecto está estrechamente vinculado a los sectores específicos de la economía y la sociedad que experimentarán las mayores transformaciones.

Debido a su propia naturaleza, los procesos de enseñanza-aprendizaje se encuentran entre las tareas que más se pueden beneficiar y que previsiblemente experimentarán cambios significativos. El desafío que plantea la revolución de la IA al profesorado docente radica en su capacidad para adaptarse y aprovechar al máximo estos cambios. Además, estos procesos no deben ser abordados únicamente por pedagogos; es necesario involucrar a diferentes actores que puedan aportar visiones ricas desde diversas perspectivas, experiencias y conocimientos para avanzar.

En estos primeros momentos de descubrimiento, es importante «jugar» y experimentar, comenzar a utilizar las posibilidades que nos brinda la IA para optimizar nuestro trabajo, como se explica en el capítulo 2. Al principio, con precaución y supervisando los resultados obtenidos a medida que vayamos adquiriendo mayor destreza y confianza con más seguridad y agilidad. En este aspecto, el capítulo 2 describe las posibilidades de este tipo de tecnologías, como la traducción de texto, creación de resúmenes de contenidos o creación de preguntas de exámenes entre otros. El conocimiento sobre cómo hacer un buen *prompt* e interactuar eficientemente con el ChatGPT es importante, pero debemos ser conscientes también de sus limitaciones, como las alucinaciones o el sesgo en los datos.

Además, es fundamental mantenernos actualizados sobre los avances a través de blogs especializados y de redes sociales; aprender también de nuestro alumnado y adoptar una actitud *hacker*, utilizando y experimentando con las herramientas, compartiendo y colaborando con la comunidad para un uso efectivo y productivo.

Nuestra función como docentes adquiere ahora un valor aún mayor. Aunque este libro no pretende responder a las preguntas como:

«¿qué significa aprender?», «¿tecnología sí o no?» o «¿controlar o educar?», en el contexto de ChatGPT hemos comprendido la importancia de formular las preguntas correctas y cuestionar la veracidad de las respuestas, verificándolas con hechos. Es fundamental que nuestros estudiantes aprendan a afrontar la incertidumbre, la complejidad y el cambio. Y la mejor herramienta para lograrlo es el pensamiento crítico (Bezanilla-Albisua *et al.*, 2018). Según la definición de la Unión Europea (Bianchi, Pisiotis y Cabrera, 2022), el pensamiento crítico implica evaluar la información y los argumentos, identificar supuestos, cuestionar el *statu quo* y reflexionar sobre cómo influyen los contextos personales, sociales y culturales en el pensamiento y las conclusiones. Transmitir y desarrollar esta competencia será clave en nuestra labor docente. Tenemos el mandato categórico de reflexionar y verificar la coherencia de cada propuesta antes de utilizarla en el aula, y debemos exigir lo mismo a nuestros estudiantes.

Por otro lado, el uso de ChatGPT conlleva la necesidad de transparencia y responsabilidad, así como un enfoque ético, como se vio en el capítulo 4. Como docentes, debemos dar ejemplo y emplear estrategias para su uso responsable (Rahimi y Abadi, 2023), intentando siempre fomentar el pensamiento crítico del alumnado con el apoyo de la tecnología.

Algunos autores (Webb, 2023) plantean la necesidad de adaptarnos y repensar nuestras acciones educativas y evaluadoras. En este libro proponemos integrar estas IA generativas en procesos que se orienten al aprendizaje y se centren en los estudiantes, por ejemplo, a través de evaluaciones formativas con un proceso de seguimiento y acompañamiento eficaz.

Algunos miedos iniciales podrían incitarnos a incrementar el control, a fiscalizar si los estudiantes utilizan ChatGPT, o herramientas similares, y en qué medida, pero cuando el control no garantiza un aprendizaje efectivo se vuelve inútil. En lugar de centrarnos en el problema, debemos enfocarnos en la oportunidad que se nos presenta. En el capítulo 5 sobre la aplicación del ChatGPT en el aula, se proponen diversas actividades donde el profesor sea el facilitador de información y, junto con los estudiantes, se analice críticamente

la información que esta herramienta proporciona. Estas actividades implican por parte del profesorado comprender los procesos de los estudiantes, diseñar tareas que promuevan la construcción de conocimiento de manera significativa, contextualizada y compleja.

Con este capítulo hemos llegado al final de este libro introductorio sobre ChatGPT. Los autores y las autoras habremos cumplido nuestra misión si hemos colaborado en esclarecer qué es y qué no es ChatGPT, si gracias a él habéis adquirido estrategias para maximizar sus capacidades en vuestro trabajo, al mismo tiempo que minimizar sus peligros, y si os hemos incitado a reflexionar sobre su uso en el entorno docente. También esperamos que el libro os haya resultado práctico y que a partir de él os hayáis familiarizado con la «ingeniería del *prompt*». Sin embargo, os animamos a considerar este libro como punto de partida, a seguir investigando y explorar los recursos que os ofrecemos para seguir avanzando en el conocimiento de soluciones e ideas para su uso en la enseñanza y el aprendizaje. En este viaje que acaba de comenzar, esperamos poder acompañaros con actualizaciones regulares de este texto.

## Bibliografía

- Bezanilla-Albisua, M. J., Poblete-Ruiz, M., Fernández-Nogueira, D., Arranz-Turnes, S. y Campo-Carrasco, L. (2018). El pensamiento crítico desde la perspectiva de los docentes universitarios. *Estudios Pedagógicos (Valdivia)*, 44 (1), 89-113.
- Bianchi, G., Pisioties, U. y Cabrera Giraldez, M. (2022). *The European sustainability competence framework*. <https://econpapers.repec.org/paper/iptiptwpa/jrc128040.htm>
- Rahimi, F. y Talebi Bezmin Abadi, A. (2023). ChatGTP and publication ethics. *Archives of Medical Research*, 54 (3), 272-274. <https://doi.org/doi:10.1016/j.arcmed.2023.03.004>.
- Webb, M. (2023). *A generative AI primer*. <https://nationalcentreforai.jiscinvolve.org/wp/2023/05/11/generative-ai-primer/#3-1>

## AUTORÍA

**MARIONA GRANÉ** es licenciada en Filosofía y Ciencias de la Educación y doctora en Pedagogía por la Universidad de Barcelona. Es investigadora del grupo consolidado Learning, Media & Social Interactions (LMI) y profesora agregada en la Facultad de Educación de la UB, donde imparte docencia en los estudios de grado y máster. Es subcoordinadora del máster oficial Entornos de Enseñanza y Aprendizaje Mediados por Tecnologías Digitales (EEATD) y forma parte del doctorado en Educación y Sociedad. Es miembro del Grupo de Innovación Docente en Evaluación con Tecnologías (GIDAT), y a su vez es parte del Institut de Recerca en Educació (IRE) y del Observatori de l'Educació Digital (OED) de la misma Universidad de Barcelona. Su campo de estudio son las tecnologías digitales en educación; sus principales líneas son el diseño y la evaluación de recursos digitales orientados al aprendizaje, la integración de tecnologías digitales en entornos educativos y la gestión de tecnologías en entornos educativos. Ha participado en el diseño y desarrollo de diversos proyectos de investigación nacionales e internacionales, ha dirigido y participado en proyectos de transferencia de conocimiento y es autora de diversos artículos científicos, libros y capítulos de libros.

**MIREIA RIBERA** es diplomada en Biblioteconomía y Documentación por la UB e Ingeniera Superior en Informática por la UPC y doctora en Documentación Digital. Es investigadora del grupo Computer

Vision de la UB y profesora agregada del Departamento de Matemáticas e Informática de la UB, donde imparte docencia en el grado de Informática y en el máster de Fundamentos de Ciencia de Datos. Ha participado en diversos proyectos nacionales e internacionales y publicado más de 100 artículos científicos. Su campo de estudio es la visualización de datos, la accesibilidad digital y, en general, la interacción persona-ordenador. Su formación interdisciplinaria le permite tener una visión crítica, aunque constructiva, de las innovaciones tecnológicas y analizar las consecuencias de estas en las personas directamente usuarias o simplemente implicadas.

**OLIVER DÍAZ MONTESDEOCA** es ingeniero de telecomunicaciones por la Universidad de Las Palmas de Gran Canaria y doctor en Ingeniería Electrónica por la Universidad de Surrey (Reino Unido). Es profesor del Departamento de Matemáticas e Informática de la Universidad de Barcelona y actualmente vicedecano de estudiantes y relaciones institucionales de esta universidad. Además, es investigador sénior en el Grupo de Investigación Consolidado Inteligencia Artificial y Aplicaciones Biomédicas. Tiene más de 13 años de experiencia internacional en el campo de la inteligencia artificial aplicada a la medicina. Ha participado en más de 20 proyectos de investigación y transferencia, y ha publicado cerca de 90 artículos científicos en revistas. Sus intereses incluyen el uso de la inteligencia artificial en medicina y entornos educativos.

**DANIEL ORTIZ MARTÍNEZ** es ingeniero informático por la Universidad de Castilla La Mancha y doctor en Inteligencia Artificial por la Universitat Politècnica de València. Es profesor del Departamento de Matemáticas y Ciencias de la Computación de la Universidad de Barcelona, donde coordina la asignatura de Procesamiento del Lenguaje Natural del máster de Fundamentos de la Ciencia de Datos. Sus intereses de investigación se centran en los campos del aprendizaje automático y la ciencia de datos y su aplicación en la investigación biomédica y el procesamiento del lenguaje natural. Ha trabajado en 18 proyectos de investigación y transferencia tecnológica y ha pu-



blicado más de 50 artículos de investigación en congresos y revistas internacionales. Además, ha ejercido como revisor científico para la Comisión Europea, así como para comités científicos de conferencias y revistas. Una parte importante del trabajo realizado en el área del procesamiento del lenguaje natural ha sido la transferencia tecnológica, ejerciendo como líder técnico en una compañía multinacional del comercio electrónico. Ha sido también muy activo en la creación de herramientas de software libre relacionadas con inteligencia artificial.

**ELOI PUERTAS I PRATS** es ingeniero informático por la Universitat Autònoma de Barcelona y doctor en Inteligencia Artificial por la Universitat de Barcelona. Es profesor e investigador del Departamento de Matemáticas e Informàtica de la Universidad de Barcelona, donde imparte asignaturas de ingeniería de software de sistemas inteligentes. Sus intereses de investigación se centran en la ciencia de datos aplicada al aprendizaje, el aprendizaje automático y la visión por computador. Asimismo, ha participado en diferentes proyectos de transferencia tecnológica relacionados con aplicaciones del aprendizaje electrónico y ciencia ciudadana. Además, ha sido cofundador de diferentes iniciativas para la divulgación de la ciencia de datos y la programación como PyBCN y DataBeersBCN.

**CARLOS LOPEZOSA** es doctor en Periodismo por la Universitat Pompeu Fabra e investigador visitante en la Universidad de Barcelona (beca postdoctoral Margarita Salas). Su tesis doctoral se centró en el estudio de los factores de posicionamiento de sitios intensivos en contenidos, en especial de medios de comunicación en línea, así como en la evaluación de herramientas de análisis SEO. Es especialista en posicionamiento en buscadores y en sistemas de monetización basados en estrategias de contenidos de calidad. Ha sido profesor asociado de la Universitat Pompeu Fabra, impartiendo docencia en la Facultad de Comunicación, en los grados de Periodismo, Comunicación Audiovisual y Publicidad, y Relaciones Públicas.

**JAVIER GUALLAR** es doctor en Información y Documentación por la Universidad de Barcelona y doctor en Comunicación por la Universidad Pompeu Fabra. Es profesor de la Facultad de Información y Medios Audiovisuales de la Universidad de Barcelona y secretario académico de esta Facultad. Es miembro del Centro de Investigación en Información, Comunicación y Cultura CRICC de la Universidad de Barcelona. Ha participado en varios proyectos de investigación competitivos. En la actualidad es investigador principal del proyecto CUVICOM, sobre curación, visualización y visibilidad. Es subdirector de la revista *Profesional de la Información* y director de las colecciones de libros Profesional de la Información y EPI Scholar en la Editorial UOC. Sus principales líneas de investigación son la curación de contenidos, el periodismo digital y la documentación periodística. Es autor de cuatro libros y de más de 80 artículos en revistas indexadas y editor de *Content Curators* (<https://jguallar.substack.com/>).

# ÍNDICE

<b>1. Introducción</b>	7
1.1. ¿Por qué ahora?	8
1.2. ¿Qué es ChatGPT?	9
1.3. Historia de los <i>chatbots</i> y su evolución en la educación	9
1.3.1. El uso de la tecnología en la educación. Avances tecnológicos que han contribuido a su desarrollo	9
1.3.2. ¿Qué es un <i>chatbot</i> ?	11
1.3.3. El uso de <i>chatbots</i> en entornos educativos	11
1.4. Importancia de la inteligencia artificial en entornos universitarios, aplicaciones en docencia, investigación y gestión	14
1.5. Retos y oportunidades	16
1.6. Estructura del libro	18
Bibliografía	19
<b>2. Indicaciones prácticas para usar ChatGPT</b>	21
2.1. Cómo usar ChatGPT	22
2.2. Características destacadas de ChatGPT	25
2.2.1. ChatGPT como agente conversacional ( <i>chatbot</i> )	26
2.2.2. ChatGPT como generador de textos	28
2.2.3. ChatGPT para la traducción y reescritura de textos	32
2.2.4. Otras funcionalidades de ChatGPT	34
2.3. Limitaciones de ChatGPT	39

2.4. Consejos para hacer <i>prompts</i> más efectivos .....	42
2.5. Nuevas funcionalidades en ChatGPT 4.0 .....	44
2.6. Otras tecnologías GPT de OpenAI .....	47
2.7. Nuevos competidores para ChatGPT: la carrera tecnológica tras de los LLM .....	48
Bibliografía .....	49
<b>3. La tecnología tras ChatGPT .....</b>	<b>51</b>
3.1. Inteligencia artificial .....	51
3.1.1. ¿Qué es la inteligencia artificial? .....	51
3.1.2. La inteligencia artificial y sus ramas .....	53
3.1.3. Aprendizaje automático .....	54
Tipos de aprendizaje automático .....	57
3.1.4. Conexionismo .....	59
La neurona artificial y el perceptrón .....	59
El perceptrón multicapa .....	63
3.2. Procesamiento del lenguaje natural .....	64
3.2.1. Definición .....	64
3.2.2. Aplicaciones del procesamiento del lenguaje natural .....	64
3.2.3. Modelos de lenguaje .....	65
Definición .....	66
Modelos de lenguaje de <i>n</i> -gramas .....	67
Modelos de lenguaje basados en redes neuronales .....	71
Modelos basados en redes recurrentes .....	72
Modelos <i>seq-to-seq</i> .....	76
La arquitectura <i>transformer</i> .....	79
GPT .....	80
3.3. ChatGPT y la era de los modelos fundacionales .....	81
3.3.1. Modelos fundacionales .....	81
3.3.2. ChatGPT .....	83
¿Qué es ChatGPT? .....	83
Proceso de implementación .....	83
Fortalezas y debilidades .....	85
Bibliografía .....	85

<b>4. Inteligencia artificial, desinformación y aspectos éticos</b>	87
4.1. Consideraciones sobre la IA y la desinformación	87
4.2. IA y fuentes de información frente a los bulos y la desinformación	88
4.3. Retos éticos en el uso de la IA como herramienta docente para combatir la desinformación	93
4.4. Conclusiones	93
Bibliografía	94
<b>5. Implementación de ChatGPT en el aula</b>	97
5.1. Implicaciones de herramientas como ChatGPT en los procesos comunicativos, creativos y de aprendizaje	98
5.1.1. Ecosistemas tecnológicos de aprendizaje	98
5.1.2. ¿Controlar o educar?	99
5.1.3. Crear es aprender	101
5.2. Oportunidades para los aprendizajes con ChatGPT	102
5.3. Posibilidades para educar en un ecosistema que integra ChatGPT	104
5.3.1. Pensar en aprender	105
5.3.2. Estrategias	106
5.3.3. Actividades con ChatGPT	109
5.4. Desafíos y dilemas incipientes en educación	114
5.5. Consideraciones finales, educar sí o sí	116
Bibliografía	118
<b>6. ChatGPT, implicaciones para la sociedad y la educación</b>	123
Bibliografía	126
<b>Autoría</b>	127

## ChatGPT y educación universitaria: posibilidades y límites de ChatGPT como herramienta docente

La inteligencia artificial está aquí, y será una realidad en el futuro de nuestro alumnado. Aunque tardaremos unos años a aprovechar ChatGPT y herramientas afines con plena productividad, con este libro recorreremos el camino con un poco más de calma y conocimiento para ayudar a los docentes hacia una implementación más satisfactoria en el aula y en nuestros quehaceres académicos.

Este es un libro interdisciplinario en el que varias miradas se coordinan para explicar el origen, el impacto, los sesgos y los posibles usos de ChatGPT en entornos académicos.

**Mireia Ribera.** Diplomada en Biblioteconomía y Documentación por la UB e ingeniera superior en Informática por la UPC. Ha hecho el doctorado en documentación digital. Es investigadora del grupo Computer Vision de la UB y profesora agregada del Departamento de Matemáticas e Informática de la UB, donde imparte docencia en el grado de Informática y en el máster de Fundamentos de Ciencia de Datos. Ha participado en diversos proyectos nacionales e internacionales y publicado más de 100 artículos científicos. Su campo de estudio es la visualización de datos, la accesibilidad digital y, en general, la interacción persona-ordenador. Su formación interdisciplinaria le permite tener una visión crítica, aunque constructiva, de las innovaciones tecnológicas y analizar las consecuencias de estas en las personas directamente usuarias o simplemente implicadas.

**Oliver Díaz Montesdeoca.** Ingeniero de Telecomunicaciones por la Universidad de Las Palmas de Gran Canaria y doctor en Ingeniería Electrónica por la Universidad de Surrey (Reino Unido). Es profesor del Departamento de Matemáticas e Informática de la Universidad de Barcelona y actualmente vicedecano de estudiantes y relaciones institucionales de esta universidad. Además, es investigador sénior en el grupo de investigación consolidado Inteligencia Artificial y Aplicaciones Biomédicas. Tiene cerca de 15 años de experiencia internacional en el campo de la inteligencia artificial aplicada a la medicina. Ha participado en más de 20 proyectos de investigación y transferencia, y ha publicado cerca de 100 artículos científicos en revistas y conferencias. Sus intereses incluyen el uso de la inteligencia artificial en medicina y entornos educativos.

