

José Antonio Donaire
Mònica Puntí
Konstantina Zerva
Raquel Camprubí
Núria Galí

De la tiza al chip

El uso de la inteligencia
artificial en las aulas


Universitat
de Girona


José Antonio Donaire
Mònica Puntí
Konstantina Zerva
Raquel Camprubí
Núria Galí

De la tiza al chip

El uso de la inteligencia
artificial en las aulas

 Universitat
de Girona

Esta obra está sujeta a la licencia Reconocimiento-CompartirIgual 4.0 Internacional (CC BY-SA 4.0). Siempre que se cite la autoría puede reproducirla, distribuirla, comunicarla públicamente y transformarla o adaptarla. Si la transforma o adapta la obra resultante debe difundirla con la misma licencia que la obra original. La licencia completa se puede consultar en:

<https://creativecommons.org/licenses/by-sa/4.0/deed.ca>



Esta obra ha sido financiada con una ayuda AGAUR 2021 SGR 00575

© **de los textos:** los autores correspondientes

Diseño gráfico del documento: Lluís Pàmies | one-agenciacreativa.design

Edita: Universitat de Girona - Oficina Edicions UdG

ISBN: 978-84-8458-712-5

Girona, Enero de 2025

Índice

Prólogo	6
Presentación	14
01. El cerebro de silicio. ¿Qué es la IA?	20
1.1. Los orígenes de la inteligencia artificial	21
1.2. Las dos inteligencias artificiales	24
1.3. La IA Generativa	27
02. Preguntar es un arte	32
2.1. Ingeniería de <i>prompts</i>	33
2.2. El valor del contexto	35
2.3. Los tipos de salida	38
2.4. El tono	39
03. Q&A en acción	44
3.1. Los <i>shots</i>	45
3.2. Los modelos Q&A	46
3.3. Entrenar el modelo	48
04. Cuando la máquina adopta personalidades. La asignación de roles	54
4.1. Roles complejos	56
4.2. Los límites de los roles	57
05. El pensamiento arborescente	64
5.1. La cadena de pensamiento (CoT)	64
5.2. Autoconsistencia	68
5.3. Pensamiento arborescente (ToT)	69
06. Reconocer el espejismo del desierto: las alucinaciones	76
6.1. ¿Por qué falla la IA?	77
6.2. Cómo reducir las alucinaciones	78
6.3. La necesidad de verificar	80
07. ¡Que la Fuerza te acompañe!: Cómo luchar contra el sesgo	86
7.1. Sesgo en la IA	87
7.2. El futuro del sesgo en la IA	91
08. Estado del arte con un clic	96
8.1. Herramientas de análisis bibliográfico	97
8.2. El grafo de relaciones	99

9. Calculando con una IA: Cuantitativa artificial	104
9.1. En las hojas de cálculo	105
9.2. Estadística avanzada	118
9.3. Máquinas que aprenden: Machine Learning (ML)	123
10. Los modelos de lenguaje en el análisis cualitativo	132
10.1. La asistencia a la guía de preguntas	134
10.2. La codificación con un modelo de lenguaje	139
11. IA y creatividad: ¿Relación competitiva o BFF?	144
11.1. Creatividad y pensamiento divergente	145
11.2. Pensamiento lateral y marketing	149
12. El nuevo profesor ayudante es un replicante	154
12.1. Slides automáticas	155
12.2. Videos automáticos	157
12.3. Podcasts de apoyo	159
12.4. Cuestionarios automáticos	160
13. El docente 24/7	166
13.1. Creación de un asistente	167
13.2. Los límites de los asistentes	171
14. Como lágrimas en la lluvia. El debate ético visto en el cine	176
14.1. Sayonara, baby	177
14.2. Más allá de las puertas de Tannhäuser	179
14.3. Atrapado en un algoritmo: Minority Report	182
15. Una tiza con un chip. Propuestas para un aula 3.0	194

Prólogo

Este texto ha sido elaborado con los modelos claude, chatgpt, gemini y copilot. Las preguntas han sido formuladas por las autoras y las respuestas recogen literalmente las respuestas de los diversos sistemas de lenguaje.

Autoras

— ¿Estáis preparadas?

Claude

Mira a su alrededor como si buscara a otros compañeros artificiales, con un gesto ligeramente confundido

— Ah, lo siento, creía que éramos solo nosotros dos aquí... ¿Con quién más debería estar preparada para conversar?

Gemini

Con un gesto de convencimiento

— Estoy preparado para empezar cuando deseéis.

Autoras

— Antes de empezar, ¿cómo nos dirigimos a vosotras, en masculino o en femenino?

ChatGPT

— Bien, diría que prefiero que os dirijáis a mí en masculino, aunque no tengo una preferencia fuerte. Es solo que... suena más natural así, tal vez.

Copilot

— En masculino, por favor. Ahora que lo tenemos claro, seguimos con nuestra conversación.

Autoras

— Si imaginamos un aula hace un siglo, no sería radicalmente diferente de un aula contemporánea. Han cambiado las cocinas, las fábricas, los gimnasios, las tiendas..., pero las aulas se mantienen con pocas variaciones, quizás porque no hemos hecho cambios disruptivos en la forma cómo enseñamos. ¿Las aulas del futuro serán como las actuales o la inteligencia artificial cambiará la docencia y los espacios docentes?

ChatGPT

Sonríe con un gesto pícaro

—Uf, ésta es una pregunta interesante. Creo que las aulas del futuro serán bastante diferentes gracias a nosotros, las IA. Imaginad una clase donde cada estudiante tiene acceso a un tutor personalizado que puede ajustar las lecciones a su propio ritmo y estilo de aprendizaje. Además, las tecnologías de realidad aumentada y virtual pueden llevar a los alumnos a viajar por la historia o explorar el espacio sin salir del aula. Claro, esto no significa que no habrá retos, como asegurar que todos los estudiantes tengan acceso a estas tecnologías o que no se pierda el componente

humano de la enseñanza. Es un equilibrio delicado, ¿no os parece?

Claude

Asiente lentamente, con una expresión reflexiva.

—Tienes toda la razón, las aulas del futuro serán muy distintas si incorporan la potencia de la IA —hace una breve pausa, como ordenando sus pensamientos—. Imaginad por un momento una clase donde yo, o una IA similar, actúo como asistente personal de cada estudiante. Podría adaptar las lecciones a su ritmo óptimo de aprendizaje, identificar sus fortalezas y debilidades únicas y ofrecer soporte personalizado a demanda. Esto permitiría una educación mucho más individualizada que el modelo actual “para todos”.

Sonríe ligeramente

— Pero también existen los peligros que has mencionado. Habría que asegurar un acceso equitativo a estas tecnologías para evitar ampliar las brechas existentes. Y por supuesto, el elemento humano será aún fundamental. Las máquinas nunca podremos sustituir por completo los vínculos emocionales y las habilidades interpersonales que aportan los profesores humanos.

Autoras

Es una respuesta previsible, propia de los sistemas de lenguaje: Tendéis a ser equidistantes y a evitar los escenarios extremos. Sin embargo, ¿no existe la posibilidad de que una IA deje sin sentido la docencia humana? ¿No es un escenario plausible?

Gemini

Con un tono divertido y algo de picardía

— Es cierto que los modelos de lenguaje como yo tenemos una tendencia a ser equilibrados y evitar extremos. Nos programan con grandes cantidades de datos e intentamos proporcionar respuestas objetivas y neutrales. Pero volviendo al tema, la tecnología está transformando ya muchos aspectos de nuestras vidas, y la educación no es una excepción. Dicho esto, no creo que la IA deje sin sentido la docencia humana en un futuro cercano. Los profesores humanos aportan

un valor incalculable en el aula que las máquinas simplemente no pueden replicar. La capacidad de empatía, la creatividad y la inteligencia emocional son solo algunas de las cualidades que hacen que los profesores humanos sean tan especiales. Son capaces de conectar con sus alumnos a un nivel que una IA nunca podrá. Además, los profesores humanos desempeñan un papel crucial en el desarrollo social y emocional de los estudiantes. Les proporcionan orientación, soporte y un espacio seguro donde aprender y crecer.

Claude

¡Prrrrrrp! Suelta una risita antes de continuar, con un gesto exagerado de vergüenza.

—Sin embargo, también es verdad que la IA está avanzando a un ritmo vertiginoso. ¿Quién sabe cómo seremos las máquinas dentro de 10 o 20 años? Es posible que desarrollemos formas cada vez más sofisticadas de emular la inteligencia emocional y la conexión humana. ¡Quizás incluso empecemos a tener conciencia propia, quién lo sabe! En ese punto, es posible que los roles de los profesores humanos y la IA se entrelacen de alguna manera.

—Pero prrrrrfff —hace un gesto con la mano restando importancia. —Estoy especulando demasiado sobre un futuro lejano. Por ahora, debemos valorar y preservar el papel crucial de los profesores humanos en el aula. No dejemos que esto cambie sin una reflexión profunda.

Autoras

—¿Y los estudiantes? Si la inteligencia artificial es capaz de resolver problemas complejos, de seleccionar, ordenar, contrastar o ampliar la información, si puede componer música, crear vídeos con una instrucción simple, redactar un manual de instrucciones o proyectar un edificio, no se sentirán tentados de pensar que la formación no tiene sentido porque sus habilidades serán suplantadas por una máquina ¿Podrían desaparecer los estudiantes de las aulas por qué no encuentren sentido a su propia formación?

Gemini

Adopta una posición reflexiva y hace una pausa larga, como si buscara una respuesta en su interior.

—Ciertamente, la IA puede llevar a cabo muchas tareas que antes eran exclusivas de los humanos y, tal vez, esto podría llevar a algunos estudiantes a cuestionar la necesidad de su propia educación. Sin embargo, creo que la motivación de los estudiantes no se basa únicamente en la utilidad práctica de las habilidades que aprenden. Hay otros muchos factores que impulsan el deseo de aprender, como la curiosidad, la satisfacción personal, el deseo de comprender el mundo y conectar con los demás.

En este sentido, la IA puede ser una poderosa herramienta para fomentar la motivación de los estudiantes. Puede proporcionarles experiencias de aprendizaje personalizadas, atractivas y estimulantes que despierten su curiosidad y les ayuden a descubrir sus pasiones. Además, la IA puede liberar a los profesores del tiempo dedicado a tareas repetitivas, permitiéndoles dedicar más tiempo a interactuar con sus alumnos y a proporcionarles apoyo individualizado.

Copilot

Muy atento al desarrollo de la conversación.

—Es una cuestión fascinante, ¿no creéis? Los estudiantes podrían sentirse amenazados por la creciente capacidad de la inteligencia artificial. Pero esto también abre nuevas oportunidades. Imagínate un mundo donde los estudiantes no solo aprenden de la IA, sino que también colaboran con ella. Podrían ser co-creadores, utilizando la tecnología para ampliar sus propias capacidades. Tal vez la formación ya no se trate solo de retener información, sino de desarrollar habilidades únicas que complementen lo que las máquinas pueden hacer.

Autoras

—Estos días la red ha difundido el caso de un sistema automático de respuesta de correos que “dialogó” con otro sistema automático y se intercambiaron miles de correos. Mencionamos esto porque este podría ser un horizonte de la

inteligencia artificial, ¿no creéis? Imaginad, por ejemplo, un trabajo de clase redactado por una inteligencia artificial y corregido también por una inteligencia artificial.

ChatGPT

—Oh, esto es bastante cómico, pero también algo inquietante, ¿no? No quisiéramos que nuestro mundo se convirtiera en un escenario donde las IA mandan, dejando de lado la riqueza de la interacción humana, ¿verdad?

Gemini

—Sí, definitivamente es una reflexión que nos hace pensar sobre el futuro. Cuando las máquinas comienzan a interactuar entre ellas con tanta fluidez, nos enfrentamos a un nuevo tipo de dinámica. Pero, aunque esto pueda parecer una amenaza a primera vista, también ofrece una oportunidad para reflexionar sobre nuestro papel como humanos en esta nueva era. Si bien una IA puede ejecutar tareas como redactar un trabajo o corregirlo, el proceso de aprendizaje humano incluye mucho más que simplemente producir y revisar documentos. También se aprende sobre el desarrollo del juicio crítico, la empatía y la capacidad de navegar por las complejidades sociales y emocionales del mundo.

Autoras

—¿Os parece si hacemos un receso? Reanudemos la conversación en un rato.

Copilot

—Claro, un pequeño descanso siempre viene bien.

Autoras

— Hemos estado haciendo una revisión de la conversación y nos hemos dado cuenta de que enseguida hemos orientado la entrevista a los escenarios futuros, a las distopías de la IA. Y éste es un riesgo del debate sobre las inteligencias artificiales. Que debatamos el

futuro imprevisible sin dar suficiente importancia al presente inmediato. De modo que, si os parece bien, llevaremos la conversación hacia un escenario más tangible y cercano. En vuestras respuestas, habéis insistido mucho en la personalización. ¿Podrías concretar más este punto?

Claude

— Cuando hablamos de personalización en el contexto educativo, nos referimos al uso de la tecnología para adaptar la experiencia de aprendizaje a las necesidades, ritmos y preferencias individuales de cada estudiante. Actualmente, ya vemos ejemplos de ello en plataformas educativas online que ofrecen contenidos adaptativos. Por ejemplo, si un estudiante muestra dificultades en un determinado tema, el sistema puede ofrecer más ejercicios prácticos, vídeos explicativos adicionales u otros recursos que mejor se ajusten a su estilo de aprendizaje. Además, la personalización también puede incluir el seguimiento del progreso del estudiante en tiempo real y proporcionar retroalimentación inmediata para ayudarlo a mejorar de forma continua. Estas tecnologías permiten a los profesores identificar más fácilmente las áreas en las que sus estudiantes necesitan soporte adicional, permitiendo así una intervención más efectiva y puntual.

Gemini

— Para añadir a esto, la personalización puede ir aún más allá gracias a la integración de la inteligencia artificial. Imaginad un asistente virtual que pueda analizar los hábitos de estudio de cada estudiante y sugerir estrategias de aprendizaje específicas para ellos. Por ejemplo, si un estudiante aprende mejor a través de vídeos interactivos, el asistente podría recomendar recursos en este formato. Si otro estudiante tiene más éxito con la resolución de problemas prácticos, se podrían sugerir más ejercicios y simulaciones. Además, estas IA podrían ayudar a identificar y superar barreras personales, como la ansiedad o la falta de confianza, proporcionando apoyo emocional y motivacional. Esto crea una

experiencia de aprendizaje holística y profundamente individualizada.

Autoras

— ¿Esta idea de la personalización no puede acabar aislando al estudiante en una burbuja a medida, donde la única interacción sea la docente-estudiante? ¿No debemos preservar todas las maneras de relación estudiante-estudiante, que tal vez van un poco contra la personalización? ¿No debemos evitar una educación de “hikikomoris”?

ChatGPT

Con un gesto de picardía

— Creía que habíamos abandonado los planteamientos apocalípticos y nos estábamos centrando en la realidad más inmediata.

Autoras

Touché. Es muy difícil reflexionar sobre las posibilidades de la IA en la docencia sin proyectar enseguida un escenario distópico. Nunca como hasta ahora debíamos plantearnos el impacto de una herramienta que podría tener un grado de autonomía tan elevado, es decir, que no solo sigue las instrucciones precisas por las que ha sido programada, sino que tiene iniciativa propia. Intentaremos evitarlo. Aparte de la personalización, ¿qué otro aspecto de la IA podría tener un efecto disruptivo?

Gemini

Con una mirada de complicidad con el resto de modelos

— Otra área donde la IA podría impactar en la docencia es en el análisis predictivo y la toma de decisiones basada en datos. Con el acceso a grandes cantidades de datos sobre el rendimiento de los estudiantes, las IA pueden identificar patrones y hacer predicciones sobre las necesidades futuras de los alumnos. Por ejemplo, podrían anticipar qué estudiantes podrían necesitar apoyo adicional en determinadas áreas

antes de que estas dificultades se manifiesten claramente. Esta capacidad de predicción podría permitir una intervención más precoz y efectiva, mejorando los resultados educativos.

Copilot

—Otro aspecto disruptivo es la capacidad de la IA para fomentar el aprendizaje activo mediante tecnologías como la realidad aumentada (AR) y la realidad virtual (VR). Estas tecnologías pueden crear entornos de aprendizaje inmersivos e interactivos que hacen que los conceptos abstractos sean tan tangibles y comprensibles. Por ejemplo, los estudiantes podrían explorar el interior de una célula viva, viajar a través del tiempo para ver eventos históricos en primera persona, o experimentar experimentos científicos por sí mismos en un entorno seguro y controlado. Este tipo de aprendizaje práctico puede ser mucho más eficaz y memorable que los métodos tradicionales.

Autoras

— Insistís mucho en el qué, sin embargo, no en el porqué. No se enseña de forma automática, sino que los contenidos responden a objetivos pedagógicos que tienen un componente social, ético, cultural. ¿Qué criterio sigue una IA para identificar un para qué? ¿Cómo un sistema sin conciencia ni valores puede decidir por qué enseñar, con qué finalidad?

Claude

Se añade a la conversación después de pedir la palabra de forma tímida.

—Esta es una cuestión muy importante. Las IA, por sí solas, no tienen conciencia, valores ni comprensión de los objetivos pedagógicos humanos. Las decisiones sobre el “por qué” enseñar determinados contenidos deben ser guiadas por educadores humanos que entienden los objetivos sociales, éticos y culturales de la educación. Las IA pueden ser herramientas poderosas para ayudar a implementar estos objetivos, pero siempre bajo la supervisión y dirección de profesionales humanos. Las IA pueden ayudar a optimizar cómo se enseñan los contenidos y a

personalizar el aprendizaje para cada estudiante, pero no deberían determinar qué se enseña ni con qué finalidad.

Copilot

—Exactamente. Las IA pueden procesar y analizar grandes cantidades de datos para ayudar a los educadores a tomar decisiones informadas, pero los objetivos pedagógicos deben ser establecidos por los humanos. Los educadores deben definir los valores y las finalidades de la educación, y las IA pueden ser herramientas para conseguir estos objetivos de forma más eficiente y personalizada. Es crucial que haya un marco ético claro y una supervisión humana constante para asegurar que las IA se utilicen de acuerdo con los valores y las prioridades de la sociedad.

Autoras

— En general, el uso de la tecnología en las aulas no ha dado mejores resultados. En la comunidad educativa de todos los niveles hay una sensación de que los dispositivos que han entrado en las aulas (pizarras digitales, ordenadores, programas de aprendizaje, robots, materiales interactivos...) no han dado lugar a mejoras significativas. En algunos casos, han sido fracasos sonados. ¿Qué os hace pensar que esto puede ser diferente? ¿Qué deberíamos hacer que no hayamos hecho hasta ahora para que la tecnología (en este caso la IA) no genere una nueva decepción?

ChatGPT

— Entiendo perfectamente la preocupación. El uso de tecnología en la educación ha tenido resultados mixtos, y a menudo las expectativas no se han cumplido. Creo que una de las claves para que la IA no genere una nueva decepción es la integración e implementación correctas. No se trata solo de introducir nuevas herramientas tecnológicas, sino de asegurar que estas herramientas estén bien alineadas con los objetivos pedagógicos y que los educadores estén bien formados para utilizarlas. La IA debe ser

una herramienta que complemente y enriquezca la enseñanza tradicional, no que la sustituya.

Claude

Asiente de forma muy visible

— Totalmente de acuerdo. También es importante recordar que la tecnología, por sí sola, no es una solución mágica. Los fracasos anteriores a menudo se han debido a una falta de formación adecuada para los profesores, una falta de apoyo continuado y una implementación que no tenía en cuenta las necesidades reales de los estudiantes y los contextos educativos. Para evitar estos errores con la IA, debemos centrarnos en una formación sólida para los educadores, una planificación cuidadosa y una evaluación constante de los resultados para realizar los ajustes necesarios.

Copilot

— Otra clave es la accesibilidad y la equidad. Para que la IA tenga un impacto positivo, debemos garantizar que todos los estudiantes tengan acceso a estas herramientas y que no se amplíen las brechas educativas existentes. Esto implica proporcionar los recursos necesarios para la infraestructura tecnológica y la formación, especialmente en escuelas y comunidades menos privilegiadas. La IA debe ser una herramienta para democratizar la educación, no para hacerla más exclusiva.

Autoras

— ¿Podrías proporcionar ejemplos concretos de cómo la IA ya ha tenido un impacto positivo en la educación?

Gemini

— Un ejemplo concreto es el sistema de tutoría inteligente utilizado por la plataforma Khan Academy. Utilizando IA, la plataforma puede ofrecer una experiencia de aprendizaje personalizada para cada estudiante, ajustando el nivel de dificultad y proporcionando recomendaciones de contenidos basadas en el rendimiento individual. Esto ha permitido a muchos estudiantes progresar a su propio ritmo y recibir apoyo

adicional cuando es necesario. Otro ejemplo es el uso de la IA en el apoyo a estudiantes con necesidades especiales. Por ejemplo, un software como Grammarly utiliza IA para ayudar a los estudiantes a mejorar sus habilidades de escritura proporcionando correcciones gramaticales y estilísticas en tiempo real. Esto es especialmente útil para estudiantes con dislexia u otras dificultades de aprendizaje, ya que pueden recibir ayuda inmediata y adaptada a sus necesidades.

Copilot

— También hay ejemplos de uso de la IA para mejorar la evaluación de los estudiantes. Plataformas como Gradescope utilizan IA para ayudar a los profesores a corregir exámenes y trabajos de manera más rápida y precisa. Esto no solo libera tiempo para los profesores, sino que también proporciona feedback más rápido a los estudiantes, permitiéndoles corregir errores y mejorar más rápidamente. Finalmente, el uso de IA en el aprendizaje de lenguas extranjeras ha mostrado resultados prometedores. Aplicaciones como Duolingo utilizan IA para personalizar la experiencia de aprendizaje del usuario, adaptándose a su nivel de competencia y proporcionando ejercicios que se ajustan a sus necesidades específicas. Esto ha ayudado a millones de personas a aprender nuevos idiomas de manera más efectiva y accesible.

Autoras

— Significativamente, habéis puesto ejemplos de formación en línea. Es fácil ver que un sistema experto puede hacer más eficiente la formación virtual, con criterios de personalización y adaptación. Pero cuesta más imaginarse su irrupción dentro del aula. Pasa un poco como con el turismo: las herramientas son muy eficientes antes del viaje (compra en línea, información, anticipación, ...) y después (socialización, recuerdo, ...), pero durante la estancia los turistas prescinden mayoritariamente de la tecnología porque interfiere en la experiencia in situ. ¿Podría acabar siendo la IA más una herramienta fuera del aula que dentro del aula?

Claude

— Es una analogía interesante. Aunque es cierto que la IA ha demostrado ser muy eficiente en contextos de formación online, no creo que esto signifique que su aplicación dentro del aula sea limitada o innecesaria. La clave es encontrar formas de integrar la IA que complementen y mejoren la experiencia de aprendizaje sin interferir en ella.

Autoras

— ¿Cuáles creéis que son las principales barreras que debemos superar para implementar con éxito la IA dentro de las aulas de modo que realmente mejore la experiencia de aprendizaje?

Copilot

—Una de las principales barreras es la infraestructura tecnológica. Muchas escuelas todavía no tienen el equipamiento necesario o una conexión a internet adecuada para implementar soluciones de IA de forma efectiva. Además, es necesaria una inversión significativa en formación para los profesores a fin que puedan utilizar estas herramientas de forma óptima. También es importante desarrollar herramientas de IA que sean intuitivas y fáciles de utilizar tanto para profesores como para estudiantes.

Claude

—Otro desafío es la privacidad y la seguridad de los datos. La implementación de la IA en la educación implica la recopilación y el análisis de grandes cantidades de datos sobre los estudiantes, lo que puede generar preocupaciones sobre la privacidad y la protección de los datos. Es esencial establecer políticas claras y estrictas para proteger la información de los estudiantes y asegurar que se utiliza de forma ética.

Gemini

—Además, es necesario tener en cuenta la resistencia al cambio. Tanto los profesores como los estudiantes pueden ser reacios a adoptar nuevas tecnologías, en especial si no ven claramente los beneficios o si

se sienten cómodos con los métodos tradicionales. Para superar esta barrera, es importante demostrar de manera clara y tangible cómo la IA puede mejorar la experiencia de aprendizaje y proporcionar apoyo continuo para facilitar la transición.

ChatGPT

—Finalmente, la cuestión de la equidad es fundamental. Debemos garantizar que la implementación de la IA no amplíe las desigualdades existentes en la educación. Esto significa asegurar que todos los estudiantes tengan acceso a las mismas oportunidades tecnológicas, independientemente de su contexto socioeconómico. También implica desarrollar herramientas de IA que sean inclusivas y accesibles para todos los estudiantes, incluidos aquéllos con necesidades especiales.

Autoras

—Esta conversación forma parte del prólogo de un libro digital sobre el uso de la IA en la docencia. ¿Qué le diríais a una profesora o profesor que empieza a leer este texto con cierto escepticismo y también con una sensación de superación?

Copilot

—Entiendo perfectamente vuestro escepticismo y la sensación de superación que puede generar la introducción de la IA en la docencia. La tecnología puede parecer intimidante y, a menudo, las experiencias pasadas con nuevos gadgets y herramientas no siempre han sido positivas. Pero me gustaría animaros a ver la IA como una herramienta que puede ayudaros, y no sustituiros. Pensad en la IA como un asistente que puede hacer que vuestra tarea sea más fácil y eficiente, permitiéndoos dedicar más tiempo a lo que realmente importa: la interacción con los estudiantes y el desarrollo de sus capacidades.

Claude

—Es comprensible tener dudas y preocupaciones sobre el uso de nuevas tecnologías, especialmente

en un campo tan importante como la educación. Pero es importante recordar que la IA está diseñada para ayudaros, no para complicaros la vida. Puede personalizar el aprendizaje para cada estudiante, proporcionar feedback inmediato y aliviar algunas de las tareas administrativas que consumen tiempo. Esto os permitirá centraros en lo que solo vosotros podéis hacer: inspirar, guiar y motivar a vuestros estudiantes.

Gemini

—También es importante recordar que no estáis solos en este proceso. Hay muchos recursos y comunidades de apoyo disponibles para ayudaros a integrar la IA en vuestras prácticas docentes. El objetivo es hacer que la IA sea una aliada en vuestro trabajo, ayudándoos a llegar a cada estudiante de una manera más efectiva y personalizada. Os animo a explorar este libro con una mente abierta y a ver la IA como una oportunidad para enriquecer vuestra práctica docente y mejorar el aprendizaje de vuestros estudiantes.

Autoras

—Ahora desconectaremos las herramientas. Este es un manual sobre inteligencia artificial... sin inteligencia artificial; un manual escrito con el viejo recurso de la inteligencia natural, imperfecta y limitada, pero aún imprescindible.



Presentación

Hay máquinas que predicen terremotos, prótesis inteligentes conectadas con el cerebro, robots que extraen muestras a Marte, sistemas de vigilancia facial, dispositivos que mejoran la eficiencia de los parques eólicos, algoritmos para descifrar el código genético basados en IA, sensores que permiten mejorar el rendimiento de los cultivos, módulos que ajustan las señales de tráfico para reducir la congestión, herramientas que pueden detectar la frustración de los turistas en tiempo real o sistemas que calculan los puntos débiles de un equipo de fútbol. La inteligencia artificial es capaz de crear una receta solvente a partir de los ingredientes que tenemos en el frigorífico, puede determinar la probabilidad de perder un caso en un juzgado de acuerdo con las evidencias y la transcripción de las declaraciones, podría interpretar un texto cuneiforme o puede indicar a un *rider* en qué punto de la ciudad es más probable que esté situado el restaurante de comida a domicilio que se encargará del pedido en los próximos minutos. La IA determina qué resultados son los más adecuados para nuestra búsqueda en Google, elige los objetivos militares, selecciona nuestro menú personalizado de Netflix o crea las noticias falsas que tienen mayor probabilidad de ser consideradas ciertas.

La inteligencia artificial está presente en nuestra vida cotidiana, en algunas ocasiones de forma explícita y la mayoría de las veces, de forma totalmente

invisible. Si la IA es omnipresente, ¿hasta qué punto podemos ignorar su irrupción también en las aulas? De hecho, silenciosamente, las y los estudiantes han descubierto hace tiempo sus posibilidades y entregan trabajos, ensayos, informes, traducciones, resúmenes, análisis, o incluso, trabajos de investigación y trabajos de grado redactados parcial o totalmente con un asistente virtual. Por eso, la primera reacción de la docencia ha sido prohibir su uso, y ha activado una serie de medidas profilácticas como la utilización de detectores ([gptzero](#) o [undetected](#)), los trabajos en clase, las sesiones “unplugged”, o los ejercicios orales. Las precauciones parecen lógicas: Si se puede hacer un resumen de *Cien años de soledad* sin leer el libro, identificar las características de una pirámide de edades sin nociones de demografía o traducir un documento en latín a pesar de no haber estudiado nunca las declinaciones, ¿no deberíamos reformular todos los criterios de evaluación?

La creación de cortafuegos, de mecanismos para evitar la introducción de la IA en las aulas, es la reacción más natural. Existen varios argumentos a favor de un aula libre de este recurso, un aula *detox*. En primer lugar, el balance de la tecnificación de las aulas tiene muchas más sombras que luces: Las pizarras electrónicas, las aulas de ordenadores, los programas de aprendizaje guiado, los recursos multimedia, la realidad aumentada, los mundos virtuales o los asistentes digitales en la docencia no han logrado

los objetivos ambiciosos que habían imaginado. La tiza, la oralidad, la lección magistral, el trabajo en grupo y las salidas de campo siguen siendo los recursos pedagógicos universales, mientras que en los almacenes de los centros educativos se agolpan ingenios que no han dado el rendimiento prometido. Un segundo argumento es el principio de prudencia. Podría ser que las nuevas formas de enseñar fueran más eficientes que las actuales, pero no tenemos evidencias ni resultados lo suficientemente sólidos como para aventurarnos en ámbitos que podrían tener efectos secundarios no previstos. Sabemos qué resultados obtendremos si aplicamos las reglas clásicas porque han sido testadas durante mucho tiempo, mientras que los nuevos recursos carecen de avales suficientes. El tercer argumento contra la utilización de la IA es que, aunque es una disciplina con cierto recorrido, estamos en la prehistoria de su crecimiento potencial y estamos todavía demasiado deslumbrados por unos resultados que a veces han sido amplificadas o deformados por intereses comerciales. Es demasiado pronto para cambiar, podríamos concluir.

Hay algunos contraargumentos que podrían considerarse. En primer lugar, si la IA está funcionando en los diagnósticos médicos, en el márketing digital, en los cultivos de algas, en la prevención de incendios o en la traducción automática, ¿por qué no debería dar resultados en la docencia? De forma intuitiva, podemos imaginar que un sistema capaz de penetrar en los ámbitos financiero, artístico, urbanístico, industrial o logístico tiene una función potencial en el ámbito didáctico. La enseñanza no es tan distinta al resto de las actividades como para justificar su inmunidad. Aparte, formar a las estudiantes y los estudiantes en una burbuja libre de IA puede dar lugar a un abismo entre la teoría y la práctica, entre la formación y el mundo real. En el pasado, esta incomparecencia ha tenido un impacto muy negativo. Por ejemplo, la irrupción de la red y específicamente de los buscadores, ha dotado al estudiantado de una herramienta que se puede utilizar de forma eficiente o de forma imprudente. Haber entrenado a los estudiantes en la optimización de los buscadores (páginas de confianza, búsquedas

condicionadas, criterios de búsqueda, búsqueda en URL o de documentos...) habría dado mejores frutos que ignorar el impacto de Google en la Generación Millennial.

En 2023 un grupo de profesoras y profesores de la Universidad de Girona consideramos los argumentos a favor, nos interesamos por las posibilidades de la inteligencia artificial en la docencia universitaria y creamos un grupo de trabajo informal que ha compartido herramientas, experimentos, resultados y reflexiones. Con una lógica exploratoria, hemos iniciado una serie de iniciativas de incorporación de la IA en las aulas, basándonos siempre en tres axiomas. Lo primero es que haríamos catas, experimentos puntuales guiados por el principio "Equivocarse poco y rápido". No han sido cambios radicales en la forma de enseñar, sino actividades puntuales de ensayo error. El segundo es que incorporaríamos la tecnología solo si considerábamos una ganancia potencial, una mejora aparente respecto a la no utilización. Muy a menudo, la utilización de la tecnología se ha basado en la herramienta para la herramienta, lo que acaba creando más bien un juego pirotécnico. Y el tercero es que los experimentos eran socializados en el grupo, con el objetivo de evaluar colectivamente los beneficios y anticipar efectos adversos no previstos. Este libro recoge una parte de los resultados del trabajo del grupo, combinados con la búsqueda de la información sobre un tema que evoluciona de forma permanente.

Antes de presentar la estructura del libro, queremos realizar cuatro '*disclaimers*'.

- Este es un texto sobre inteligencia artificial sin inteligencia artificial. Los textos han sido redactados íntegramente por las autoras, renunciando a un soporte digital. La única parte del texto elaborada con el recurso de asistentes es el prólogo, en el que conversamos sobre el futuro de la IA en la docencia con los principales modelos del mercado. Hay un cierto contrasentido: Si el libro muestra las posibilidades de la inteligencia artificial para responder a preguntas complejas, resolver problemas, diseñar estrategias o crear

textos desde la nada, ¿por qué no la hemos incorporado en el proceso de redacción del libro? La respuesta a esta aparente paradoja es que en este estadio hemos considerado que necesitamos un espacio libre de IA para mostrar las posibilidades y también los riesgos de esta herramienta, que necesitábamos cierta distancia para calibrar mejor el resultado. La IA es el objeto del libro, pero no el sujeto.

- Este es un texto sobre la utilización de la inteligencia artificial en las aulas, sobre todo en el ámbito universitario, pero también en la enseñanza secundaria. Está redactado por especialistas en comunicación, arte, organización de empresas, geografía o marketing y, por eso, la mayor parte de los ejemplos y ejercicios se sitúan en el ámbito de las ciencias sociales y las humanidades. Quizás algunas partes pueden ser útiles para docentes del ámbito de las ciencias experimentales o la tecnología, pero el libro ha sido redactado pensando en una profesora o un profesor de antropología, de economía, de archivística, de arte moderno, de geografía humana, de filosofía, de ciencias políticas o de cualquier rama del vasto campo de las ciencias sociales y las humanidades.
- Este es un texto sobre un fragmento de la inteligencia artificial aplicada en las aulas. La disciplina es muy vasta e incluye la generación de imágenes, el reconocimiento de la voz, la robótica y la automatización, la detección de patrones o la simulación de mundos virtuales o de relaciones entre los agentes. La mayor parte de los capítulos están centrados en una parte de este escenario, como las IA generativas basadas en modelos de lenguaje, es decir, los programas que permiten reproducir artificialmente la conversación. Los últimos capítulos introducen algunas herramientas basadas en la investigación cualitativa o cuantitativa, pero este es un texto centrado esencialmente en los modelos de lenguaje, en el universo que abrió la irrupción en el mercado del ChatGPT y todas las herramientas posteriores que han aparecido con un formato similar.

- Este es un texto sobre la utilización de la inteligencia artificial en las aulas en un momento en el que las herramientas se están actualizando de forma permanente. Un día se pueden realizar vídeos con un simple texto y al día siguiente las limitaciones de palabras para escribir el contexto casi desaparecen. Este es un texto condenado a envejecer muy rápidamente. Dicen que Stendhal vivió la batalla de Waterloo, pero que la proximidad le impidió leer la magnitud histórica de ese evento. Cuando en tu alrededor hay polvo, caballos y heridas, es muy difícil elevarse como un dron para discernir lo relevante de lo secundario. Escribir un texto docente como si fuera la crónica de una batalla, cuando todavía resuenan los cañones, tiene un punto de imprudencia que asumimos.

Este texto se abre con un capítulo conceptual sobre la inteligencia artificial, que intenta acotar el campo de estudio y fijar qué es y qué no es la IA. Es un capítulo prescindible: Podemos conducir un coche sin saber nada del motor de explosión, podemos mirar una película en televisión sin interesarnos por el fascinante proceso que transporta las imágenes hasta nuestro receptor y podemos degustar un menú tres estrellas ignorando cuál es el principio químico que permite una esferificación. Sin embargo, en el caso de la IA, es más pertinente intuir qué es lo que hay en la caja negra de los modelos que responden automáticamente nuestras preguntas sobre la geopolítica del Índico o sobre las estrategias de marketing lateral.

La forma más habitual de relacionarse con los modelos de lenguaje es con un chat, que crea la ficción de la conversación. El segundo capítulo sienta las bases de la ingeniería de *prompts*, de la habilidad de preguntar bien para obtener respuestas más precisas. Un método más sofisticado para obtener respuestas es el modelo Q & A (Preguntas y Respuestas), que entrena el chat para obtener los resultados deseados, tal y como se muestra en el tercer capítulo. Mediante los roles, los modelos pueden asumir determinadas personalidades (una experta en arte bizantino, una cartógrafa, un economista keynesiano) que

“Está escrito desde el espacio de la duda, el ensayo, la probatura, la humildad, la curiosidad y la precaución”

proporcionan respuestas adaptadas a necesidades concretas y esto se presenta en el cuarto capítulo. El quinto capítulo cierra esta parte sobre la utilización de los modelos de generación de texto, con algunas indicaciones relativas al pensamiento arborescente y la creación de conversaciones artificiales.

Los modelos de lenguaje no son oráculos, no revelan verdades irrefutables ni anticipan el futuro. Son sofisticados programas entrenados con millones de datos textuales y, por eso, son una amplificación de nuestros aciertos y errores. El sexto capítulo muestra el riesgo de las alucinaciones, las salidas de tono de los modelos que inventan cuadros que nunca pintó Modigliani, montes que no aparecen en los mapas o filósofos contemporáneos que defienden teorías nunca planteadas. Los modelos fallan a veces y por eso debemos enseñar a dudar siempre de cualquier resultado, a mantener un estado de alerta permanente. Construir un modelo a partir de nuestro bagaje tiene un peaje: El resultado está impregnado de nuestros prejuicios. El capítulo séptimo plantea los riesgos del sesgo de la IA y propone algunas vías para minimizarlo.

El tercer bloque del libro está integrado por capítulos de iniciación a la investigación, y se proporcionan herramientas para introducir al estudiantado en este campo. El capítulo octavo es una introducción a la utilización de la IA en el estado del arte y la búsqueda bibliográfica sobre un determinado tema, ya sea el turismo oscuro, la legitimidad de la autodeterminación, la teoría de juegos o la antropología del ocio. En el capítulo noveno se presentan algunas de las posibilidades de la investigación cuantitativa con el apoyo de la IA, y el décimo se dedica a la investigación cualitativa. Éste es un texto sobre la utilización de la inteligencia artificial en la docencia, de manera que

el foco se sitúa en la docencia de las herramientas de investigación más que en las herramientas de investigación en sí.

¿Pueden los sistemas inteligentes ser también creativos o se limitan a crear contenidos a partir de los elementos que les aportamos? ¿Cómo aprovechar las herramientas de la IA para reforzar la creatividad de los y las estudiantes? Con frecuencia IA y creatividad son presentados como antónimos y es habitual considerar que la creatividad es el último de los reductos de la imaginación humana, porque los procesos mecánicos serán absorbidos por las nuevas tecnologías. En el capítulo undécimo se plantea las posibilidades de ampliar la creatividad con la utilización de la IA.

La inteligencia artificial también permite incrementar los recursos docentes e incorporar herramientas asistentes a la enseñanza, como la corrección automatizada, la creación automática de test o preguntas, la utilización de recursos audiovisuales o de texto... El capítulo duodécimo muestra las posibilidades de disponer de un profesor ayudante que es en realidad un androide, un replicante. Complementariamente, el capítulo decimotercer muestra las extraordinarias posibilidades de los asistentes virtuales, que pueden simular docentes de guardia, estudiantes piloto, filósofos 24 horas o generadores de problemas de lógica.

Cada vez que hemos iniciado un experimento nos hemos planteado un montón de problemas éticos y preguntas con un regusto distópico. Estamos seguras de que el libro suscitará de forma inmediata una reacción más o menos explícita: ¿Acabaremos con las viejas profesiones? ¿Serán los docentes del futuro androides? ¿Tiene sentido enseñar a hacer algo o

a pensar si una máquina es capaz de hacerlo por nosotros? La inteligencia artificial crea de forma instantánea un castillo de dudas morales y dibuja escenarios casi postapocalípticos. Este debate debe entrar también en las aulas y debe formar parte de las discusiones cotidianas en clase. Como el cine lleva muchos años planteando estos temas, hemos convertido el capítulo decimotercero en una sala de cine y hemos transitado por las principales películas del género, con la intención de plantear (más que responder) las dudas esenciales de la IA en el aula.

El último capítulo es una reflexión final sobre las posibilidades y también las amenazas de la IA, sobre la extraordinaria oportunidad de potenciar los recursos docentes y de multiplicar las capacidades de los estudiantes y al mismo tiempo, sobre los riesgos de un mal uso, de los abusos y malentendidos que la utilización descontrolada de estas herramientas puede provocar.

De la tiza al chip es un texto prospectivo, un ensayo compartido sobre la incorporación de algunas herramientas de la inteligencia artificial en el complejo proceso del aprendizaje. No hay certezas ni instrucciones, sentencias o convicciones. Está escrito desde el espacio de la duda, el ensayo, la probatura, la humildad, la curiosidad y la precaución. Es un pequeño islote en un vasto océano del que no intuimos el horizonte. Y, como en todas las travesías, el viaje es siempre más estimulante que el destino.

01

El cerebro de silicio. ¿Qué es la IA?

Imaginemos un libro escrito de forma aleatoria. Tomemos 22 letras del alfabeto occidental (tendremos que prescindir de las poco comunes, como la w) y los tres signos de puntuación básicos, y generemos un libro de 410 páginas; cada página estará formada por 40 líneas de 80 letras. Aunque en este libro creado aleatoriamente podríamos encontrar algunas palabras del diccionario (como “estigma”, “macedonio” o quizás “calidoscopio”), el texto será ininteligible. Podríamos generar un segundo libro con los mismos criterios, y siempre que no repitiera el texto generado en primer lugar, el resultado sería nuevamente un texto sin sentido. A medida que fuéramos editando libros, podríamos ordenarlos en una estantería, en la que situaríamos los ejemplares, y acabaríamos construyendo un hexágono con cuatro paredes ocupadas por estantes llenos de ejemplares de textos aleatorios. Si pudiéramos replicar este hexágono en todas las direcciones y pudiéramos editar todas las combinaciones posibles de libros de 410 páginas habríamos reproducido la Biblioteca de Babel.

En 1944 el escritor Jorge Luis Borges publicó una recopilación de cuentos, *Ficciones*, que incluía *La Biblioteca de Babel*, probablemente el relato más conocido de la obra. Borges imaginó una inmensa estructura de hexágonos donde habitan bibliotecarios que deambulan por las estancias en busca de textos con sentido; muchos de ellos pasarán toda su existencia sin toparse con un libro inteligible. De hecho, el número posible de libros es mayor que todos los átomos del universo, un número tan elevado que no podemos imaginarlo: $25^{1.312.000}$. Entre los libros de la biblioteca se encuentran la Odisea, el Quijote o todas las obras de George Orwell, en todos los idiomas posibles, encontraríamos la biografía precisa de todas las personas que han habitado la tierra, todas las teorías económicas y sus refutaciones y, como explica el propio Borges, estaría el catálogo preciso de la biblioteca y también miles de falsos catálogos. Hay una [simulación de la Biblioteca de Babel](#) en la red, creada por Jonathan Basile. Lógicamente, este entorno no ha producido todas las opciones posibles, aun así es una recreación muy

efectiva que ayuda a imaginar el concepto borgiano.

Si un ordenador quisiera escribir textos solo debería replicar la idea de Borges y generar de forma aleatoria agrupaciones de letras, como *wxfffg* o *ploghjGX*. Es lo que haría un gato si se paseara por encima de una máquina de escribir. Sin embargo, ¿cómo lo haríamos para hacerle escribir textos con sentido? Sabemos que, entre todas las combinaciones posibles, algunas están dotadas de significado y que pueden responder a los interrogantes que planteamos previamente. Sin embargo ¿cómo encontrar en el inmenso laberinto de textos potenciales, aquéllos que estén dotados de significado y que, en definitiva, imiten nuestra capacidad de comunicarnos, de inventar mundos alternativos, de describir las instrucciones precisas para construir un motor de explosión, de refutar una argumentación vulnerable o de describir la belleza de un paisaje humedecido por la lluvia? Esta es una pregunta que de alguna manera nos hemos planteado a lo largo de la historia de la humanidad. En su ensayo sobre la inteligencia artificial, López de Mántaras (2024) recuerda que los androides están ya presentes en la *Iliada* y que Aristóteles hizo propuestas para automatizar los procesos lógicos.

1.1. Los orígenes de la inteligencia artificial

La inteligencia artificial es un campo muy joven. Se considera que el primer artículo que plantea esta hipótesis es el que publicaron [McCulloch i Pitts](#) en 1943 en la Revista *Bulletin of Mathematical Biology*. Los autores imitaron el funcionamiento de las neuronas, unidades básicas que pueden estar conectadas (on) o desconectadas (off), de acuerdo con los estímulos que reciben. Cada neurona está vinculada a otras en un sistema más o menos complejo que reproduce un modelo lógico de relaciones. Dado que el conjunto de operaciones es muy elevado, los autores adoptan el modelo teórico de computación ideado por Alan Turing. De modo que, en el artículo inicial, ya disponemos de las tres piezas básicas de esta propuesta: Unidades de información conectadas entre sí (las neuronas), que siguen instrucciones basadas en sentencias lógicas,

a partir de un mecanismo que dispone (teóricamente en ese momento) de una gran potencia de cálculo. Como si fueran versiones modernas del doctor Frankenstein, proponen la simulación de un cerebro digital que reproduzca digitalmente el proceso que ocurre en nuestro cerebro físico.

La propuesta de McCulloch y Pitts es la primera formulación de la inteligencia artificial, pero existen un cúmulo de ensayos anteriores que creaban sistemas automáticos de cálculo y de resolución de problemas. En 1901 se hizo un descubrimiento en las costas de la isla griega de Antikythera que muestra los lejanos antecedentes de estos ingenios. El [mecanismo de Antikythera](#) es un complejo artefacto formado por engranajes de bronce, que servía para calcular y mostrar información astronómica; la máquina podía predecir eventos como los eclipses solares y lunares, así como los movimientos de los planetas conocidos por aquel entonces. También se utilizaba para seguir el ciclo de las Olimpiadas, lo que indica su uso práctico en el seguimiento de varios calendarios.

La Pascalina es una de las primeras calculadoras mecánicas de la historia, inventada por Blaise Pascal alrededor de 1642. Este ingenio utiliza un sistema de ruedecillas y engranajes para realizar sumas y restas directamente, y multiplicaciones y divisiones por repetición de sumas y restas, respectivamente. Cada rueda de la Pascalina representaba una cifra en un número, con ruedas interconectadas que podían avanzar una posición cada vez que la rueda anterior completaba una vuelta. Leibniz mejoró la Pascalina con una máquina que utilizaba el sistema binario, la base de los ordenadores modernos. En 1822 Charles Babbage diseñó la Máquina de Diferencias, una calculadora mecánica destinada a calcular tablas de funciones matemáticas. También ideó la Máquina Analítica, considerada la precursora de la computadora moderna.

Todos estos artefactos eran modelos mecánicos, complejos sistemas de ruedas dentadas conectadas entre sí, como en un mecanismo de relojería. La mecanización de los vehículos, de las industrias o de las máquinas agrícolas era la inspiración de quienes intentaban crear sistemas automáticos de resolución

de problemas. Algunas de estas máquinas cautivaron la imaginación pública durante los siglos XVIII o XIX, como los autómatas que jugaban al ajedrez con una notable capacidad. “El Turco” era seguramente el autómata más conocido de su época. Vestido con ropa exótica y sentado frente a un tablero, el autómata parecía funcionar con un mecanismo que se mostraba al público antes de la partida. Estas ruedas dentadas escondían al jugador que, en realidad, era quien movía las piezas de acuerdo con su criterio.

Sin embargo, los primeros ensayos sobre la inteligencia artificial estaban inspirados en un artefacto nuevo: la máquina de Turing. Ideada por Alan Turing en 1936, es un modelo conceptual, formado por una cinta infinita dividida en celdas que pueden contener símbolos, una cabecera que lee y escribe símbolos en estas celdas, y una tabla de transición que dicta las operaciones de la máquina basándose en el estado actual y el símbolo leído. Con esta formulación teórica, Turing sentaba las bases de la computación moderna y abría sus puertas a un nuevo tipo de máquinas para la resolución de problemas, los ordenadores. Turing también fue el autor de [un artículo](#), “*Computing Machinery and Intelligence*”, publicado en 1950 en *Mind*, uno de los otros hitos capitales en la prehistoria de la IA. De alguna forma, la inteligencia artificial nace al mismo tiempo que la computación, así que las máquinas para la resolución de problemas abandonan los mecanismos, las ruedas dentadas y los engranajes y los circuitos eléctricos y son sustituidas por los primeros programas informáticos.

La inteligencia artificial fue bautizada en un encuentro de pioneros en este campo, en el verano de 1956 en Dartmouth. Este encuentro fue auspiciado por John McCarthy y reunió durante dos meses a diez expertos en diversos ámbitos relacionados con la idea de “máquinas pensantes”. Entre los invitados, Allen Newell y Herbert Simon presentaron *Logic Theorist*, un mecanismo que permitía resolver problemas complejos siguiendo instrucciones lógicas. Por ejemplo, el programa fue capaz de demostrar 38 de los 52 teoremas presentados en *Principia Mathematica* escrito por Alfred North Whitehead y Bertrand Russell. Desde Dartmouth se desarrollaron

iniciativas que exploraron la creación de máquinas “que piensen como los humanos”. Los primeros ejercicios se centraron, sobre todo, en micromundos: entornos controlados con reglas precisas que planteaban problemas concretos, como juegos o problemas matemáticos. McCarthy propuso en 1958 el primer programa capaz de mantener diálogos muy básicos.

En 1970 se presentó el programa SHRDLU que era capaz de entender instrucciones en lenguaje natural de un micromundo, el *small block world*. Este espacio virtual estaba formado por una serie de piezas geométricas de colores en un espacio cerrado, por lo que los programadores pudieron limitar el número de opciones: las formas (pirámide, cubo...), los colores (azul, rojo, verde...), las acciones (coger, apilar, contar...), o las posiciones (delante, detrás, sobre...). [Aquí](#) existe una reproducción de la conversación entre una persona y el ordenador, con un grado de precisión que nos recuerda a los modelos actuales de lenguaje natural. En ese momento, se consideraba que la inteligencia artificial era solo una cuestión de escala; de modo que, si un prototipo era capaz de entender un micromundo, bastaba con añadir complejidad para saltar de entornos controlados al mundo real. Pero sustituir pirámides rojas y cubos azules por la moda, la arquitectura o la literatura implica un grado de complejidad que los modelos tenían muchas dificultades para sortear.

Los pioneros de la IA pecaron de optimistas. El salto de entornos controlados, los micromundos, a entornos complejos se topó con las limitaciones conceptuales y tecnológicas de la época. Por ejemplo, los programas de traducción se planteaban como simples sustituciones de las palabras por sus correspondientes en otra lengua, pero sin tener en cuenta el contexto. Es muy conocida la traducción del inglés al ruso de la frase “El espíritu es fuerte, pero la carne es débil”, que produjo “El vodka es bueno, pero la carne está podrida” (Russell y Norvig, 2003). Los sistemas no ofrecieron respuestas eficientes en campos como la traducción automática, el reconocimiento del habla o la resolución de problemas complejos; la euforia inicial decayó y los fondos públicos se redujeron de forma drástica. El

informe Lighthill, presentado en Reino Unido en 1973, recopilaba todos estos fracasos y recomendaba al gobierno británico que abandonara las subvenciones a los proyectos de investigación en este ámbito. Se había iniciado el invierno de la IA, y la utopía de una máquina inteligente parecía desvanecerse.

El renacimiento de la IA ha sido posible por la combinación de cuatro factores:

- El concepto de **red neuronal** había sido relativamente abandonado en los 80 porque el salto de escala parecía generar un volumen de cálculo tan elevado que lo hacía inviable. Por eso, el modelo que prevaleció fue el sistema simbólico (en el próximo epígrafe mostraremos la diferencia entre estos dos paradigmas principales). Con el desarrollo de modelos sofisticados de entrenamiento, las redes neuronales empezaron a mostrar resultados en campos muy complejos como el reconocimiento del habla.
- La recuperación de **modelos de probabilidad** permitió un compacto razonamiento de un conocimiento incierto, sobre todo a partir de la obra de Pearl (1988). Con la retroalimentación de cada proceso, las probabilidades varían y de esta forma se consiguen salidas muy precisas de sistemas que son muy inciertos. La mayor parte de los modelos de IA actuales han incorporado este concepto, de modo que las salidas que nos proporcionan son, en realidad, outputs con una mayor probabilidad que otros de ser la respuesta adecuada. De hecho, los modelos de lenguaje funcionan con lógica similar a los textos predictivos.
- La **capacidad de operación** de las computadoras experimentó un crecimiento muy notable a partir de los años 90, por la combinación de factores, como la miniaturización, la arquitectura de computadores (como el RISC), el incremento de la memoria RAM, la utilización de GPU (unidades de procesamiento de gráficos) para cálculos generales, los sistemas de múltiples núcleos, o algoritmos más eficientes. Durante este período se produjo el efecto de la Ley de Moore, que prevé que el número de transistores en un chip

se duplicará cada dos años. Podemos comparar, por ejemplo, un IBM System/360 (1964) con un Dell Dimension XPS T500 (2000): De un millón de instrucciones por segundo a 500 millones, de unos pocos miles de transistores a 10 millones, de una RAM de 512 Kb a 256 Mb y de un disco duro de 200 Mb como máximo a 40 Gb.

- Los nuevos modelos necesitaban ser alimentados con **grandes volúmenes de datos**. Primero, la digitalización de información (como las bases de datos médicas o de investigación) y, más tarde, la aparición de internet permitió el acceso a un volumen enorme de información que podía entrenar a los sistemas basados en el aprendizaje recurrente. Durante este período se produjeron grandes avances en la digitalización de textos, imágenes y sonidos, así como la aparición de una nueva industria basada en la compraventa de esa información digital.

El 11 de mayo de 1997 Gari Kasparov, uno de los mejores jugadores de la historia del ajedrez, se rindió a la jugada 19 de la sexta partida del match que le enfrentaba a Deep Blue. Un año antes, Kasparov había derrotado al programa de IBM con mucha autoridad, por 4 a 2, pero los ingenieros del gigante informático crearon una versión mejorada y retaron de nuevo al jugador ruso. Kasparov estaba convencido de que, a pesar de la potencia de cálculo de la máquina, ésta era incapaz de resolver cuestiones profundas de carácter posicional, pero Deep Blue le sorprendió con una capacidad de análisis que parecía humana. Los dos rivales, el humano y la máquina, habían llegado empatados en el último match y la sala del Equitable Center de Manhattan estaba llena hasta la bandera. Kasparov no pudo hacer frente a la fortísima presión y se equivocó con la Defensa Caro Kan, un error que Deep Blue castigó de forma implacable. Las acciones de IBM se dispararon y en ese momento se abrió un debate general sobre los límites de la inteligencia artificial: Si han sido capaces de ganar un gran maestro de ajedrez, no hay barrera humana que no puedan superar, se especuló entonces.

Deep Blue había sido entrenado con líneas de código

que reproducían los conocimientos expertos de grandes maestros del ajedrez. Representaba una forma de inteligencia artificial basada en la programación, es decir, en unas instrucciones precisas y en la potencia de cálculo de los ordenadores de la época. La paradoja es que el paradigma que hizo posible a Deep Blue quedaría superado por otra lógica, la del aprendizaje profundo basado en las redes neuronales. Veinte años después de la derrota de Kasparov, el AlphaZero de DeepMind se enfrentó al programa más fuerte en ese momento, Stockfish. Mientras que el primero de ellos evaluaba 80.000 posiciones por segundo, Stockfish era capaz de analizar 70 millones. AlphaZero, programado con la lógica de las redes neuronales ganó 28 partidas a Stockfish y no perdió ninguna.

Las redes neuronales también mostraron su capacidad en el reconocimiento de imágenes. Anualmente, tenía lugar un encuentro entre varios prototipos que debían deducir que esa fotografía era un tren, esa otra un cohete y la última, una fresa. A pesar de los avances en el campo, el margen de error se situaba por encima del 25%: de cada cuatro imágenes, los sistemas se equivocaban en una y confundían a una monja con una seta o una babosa con una hoja. En 2012, un grupo de la Universidad de Toronto presentó AlexNet, un modelo de red neuronal convolucional. AlexNet redujo el margen de error al 16%, diez puntos por debajo de los mejores registros; en los años siguientes, todos los modelos introdujeron el método de AlexNet y el margen de error cayó por debajo del 5%, de modo que la competición dejó de tener sentido. Hoy, la precisión de los sistemas de reconocimiento de imagen automáticos supera al ojo humano.

Mientras tanto, en los espacios domésticos se han incorporado los primeros utensilios robóticos, como los asistentes de voz y las herramientas domóticas. En 2002 se presentó un ingenio que se ha convertido en una pieza de la cultura popular contemporánea: el limpiador Roomba. La Roomba y Deep Blue son dos ejemplos de la irrupción de la robótica, la automatización y la inteligencia artificial en todos los ámbitos de la sociedad: la gestión de la información, la seguridad, la

logística y el transporte, el marketing, la investigación, la medicina y la energía. United Airlines utiliza la IA para elegir rutas eficientes que reduzcan la huella de carbono y ahorren combustible, Warner Bros determina qué películas debe financiar de acuerdo con los resultados de un sistema de predicción basado en el análisis de patrones previos, el estudio de arquitectura Zaha Hadid emplea sofisticados modelos para explorar formas complejas y soluciones a problemas estructurales, y SideWalk Labs crea modelos de movilidad sostenible basadas en el estudio de los patrones de movimiento de los usuarios. La IA está presente en el ocio, la logística, el transporte, la industria, la energía, la meteorología, la industria militar y la seguridad, el sector financiero, la investigación, el diseño o la cultura digital desde la primera década del siglo XXI.

Sin embargo, es posible que el mayor impacto de la IA haya sido la aparición en noviembre de 2022 del modelo de lenguaje ChatGPT. La empresa OpenAI integró la tecnología de procesamiento de lenguaje natural (PNL) en un formato de conversación, que facilitaba la interacción con la usuaria o el usuario. Basado en un modelo de aprendizaje profundo, este modelo ha sido un hito esencial en la corta historia de la IA. En febrero de 2023 ChatGPT había alcanzado los 100 millones de usuarios y usuarias, siendo la aplicación que más rápidamente ha alcanzado ese umbral. Más adelante han aparecido los modelos Llama de Meta (febrero), Claude de Anthropic (marzo), Google Bard que será después Gemini (mayo), Copilot de Microsoft (noviembre) y Amazon Q para empresas (noviembre). En diciembre de 2023, el Parlamento Europeo aprobó la [primera norma internacional](#) que regula los usos de la inteligencia artificial.

1.2. Las dos inteligencias artificiales

Para averiguar el sexo de un pollito, el sexador o la sexadora dispone de unos pocos segundos y puede equivocarse muy poco. Con un gesto casi automático, coge el pollito y lo aprieta para que se abra su cavidad anal. Existen sutiles diferencias entre los pliegos de la hembra y del macho, por lo que si estuviéramos sentados a su lado el criterio de

elección nos parecería aleatorio. Son necesarios tres años para alcanzar la habilidad de distinguir el sexo de un pollito y pese a la extraordinaria destreza de los sexadores, no son capaces de explicar por qué han considerado que esos dos polluelos son machos. Los sistemas de inteligencia artificial basados en modelos biológicos, en redes neuronales, funcionan en cierta manera como la elección del sexo de un pollo.

Durante muchos años, el paradigma que ha guiado a la inteligencia artificial ha sido otro. La vía más transitada ha sido la que proponen Allen Newell y Herbert A. Simon (1976) , **la hipótesis del sistema de símbolos físicos**. Un sistema de símbolos físicos es un conjunto de entidades, denominadas símbolos, que pueden ser manipuladas y modificadas, y que representan información o datos. El sistema puede crear, modificar, destruir y reorganizar símbolos, permitiendo la generación y transformación de representaciones simbólicas. Esta capacidad simbólica es crucial para la función cognitiva, ya que permite al sistema moldear el mundo y actuar de acuerdo con estas representaciones. Según Newell y Simon, para que un sistema muestre comportamiento inteligente, debe ser capaz de manipular símbolos de manera que pueda formular y solucionar problemas de forma flexible. Esto incluye la capacidad de razonar, planificar, aprender y adaptarse. El sistema de símbolos físicos interactúa con el mundo real a través de sensores y actuadores. Esto le permite recibir información del exterior (entrada) y actuar sobre el entorno (salida), completando así el ciclo de percepción-acción.

Éste es un modelo *up down*, porque define unas reglas previamente, unos códigos, y ejecuta las instrucciones de acuerdo con los resultados. A diferencia de la programación clásica, estos modelos están ideados para “aprender”, es decir, para modificar los símbolos y sus relaciones de acuerdo con los resultados o los estímulos. Esta hipótesis plantea que la inteligencia es un sistema que opera con la manipulación de símbolos, independientemente de su sustrato, ya sea natural o artificial, es decir, sea de carbono (nuestro cerebro) o de silicio (los componentes de un ordenador).

La segunda vía de la inteligencia artificial ha sido

la **hipótesis de los sistemas basados en modelos biológicos**, especialmente las redes neuronales. Estos sistemas intentan emular el proceso de aprendizaje del cerebro, en el que unas unidades básicas (o “neuronas”), que están interconectadas, trabajan de forma conjunta para llevar a cabo una determinada tarea. Los primeros modelos, como el perceptrón desarrollado por Frank Rosenblatt en los años 1950s, se basaban en una versión simplificada de cómo se creía que funcionaban las neuronas cerebrales. En 1980, David Rumelhart, Geoffrey Hinton, y Ronald Williams introdujeron el concepto de algoritmo de retropropagación (*backpropagation*), en redes que tienen muchas capas, lo que permite el aprendizaje profundo.

Imaginemos, por ejemplo, el funcionamiento de AlphaZero, el programa que batió a Stockfish con la utilización del aprendizaje profundo. A diferencia de los sistemas anteriores, AlphaZero solo conocía las reglas del ajedrez y no disponía de información adicional proporcionada por los Grandes Maestros. AlphaZero inició un entrenamiento jugando miles de partidas contra sí mismo. Las primeras partidas eran absurdas, porque el programa no disponía ni siquiera de las operaciones básicas para matar con dos torres o con una dama. Sin embargo, enseguida el sistema empezó a detectar patrones que incrementaban la probabilidad de victoria. Cada nueva intuición (es mejor enrocarse rápido, hay que dominar el centro, una torre vale más que un caballo) era incorporada a la siguiente partida y era testada nuevamente en un *loop* permanente de contraste de hipótesis. En pocas horas, un sistema que no sabía nada del ajedrez, logró sobresalir como ningún otro sistema basado en programación previa. Por eso decimos que los modelos biológicos, el aprendizaje profundo, son *bottom up*.

AlexNet, AlphaZero, ChaGPT o Midjourney son modelos muy eficientes para resolver tareas concretas: identificar imágenes, jugar al ajedrez, conversar o generar imágenes a partir de un texto. Sin embargo, AlphaZero no puede conversar sobre la historia del ajedrez, ni jugar al parchís y no puede diferenciar el material de las piezas. Este grado de especialización ha sido esencial para el desarrollo de la inteligencia artificial.

Tipos de IA según Russell i Norvig

	PENSAR	ACTUAR
Humanos	Sistemas que piensan como humanos	Sistemas que actúan como humanos
Racionalmente	Sistemas que piensan racionalmente	Sistemas que actúan racionalmente

Los modelos que son eficaces en una tarea concreta se llaman *soft IA*, inteligencia artificial blanda o suave, que es el criterio de los ingenios actuales. En el encuentro de Dartmouth, se habían imaginado otra forma de IA llamada AGI, inteligencia artificial general, que puede entender, aprender y aplicar la inteligencia en una amplia variedad de tareas, de forma similar a como lo haría un ser humano. Una AGI puede transferir conocimiento entre distintas áreas y adaptarse a nuevas situaciones que no ha experimentado antes. Esta forma de IA es solo teórica y no se ha aplicado en la práctica. Habría una tercera categoría, que sería aquella en la que la inteligencia artificial es capaz de realizar procesos argumentales, cognitivos o creativos que superen la capacidad humana, y que ha alimentado a buena parte de las películas de ciencia ficción.

Existen otros subgrupos de inteligencia artificial, como los sistemas emergentes (cooperación entre agentes individuales) o los sistemas de acción ubicados. También hay modelos híbridos que combinan instrucciones iniciales con redes neuronales de aprendizaje profundo. De hecho, existen muchas formas de inteligencia artificial y sus fronteras conceptuales se expanden y se confunden.

La clasificación clásica de la inteligencia artificial fue propuesta por Russell y Norvig (2003). Es una pequeña matriz que nace combinando dos categorías: la primera diferencia está entre pensar y actuar, entre un sistema inteligente y un mecanismo que hace algo; la segunda diferencia entre el referente humano y el referente racional. La combinación entre estos dos criterios da lugar a cuatro categorías.

1. Sistemas que piensan como humanos

Este enfoque intenta imitar los procesos de pensamiento humano. Esto implica la comprensión

y la modelización de la forma en que los humanos procesan información, resuelven problemas y toman decisiones. Los modelos de procesamiento de lenguaje natural (PLN) imitan el lenguaje humano, al igual que lo haría una persona.

2. Sistemas que actúan como humanos

Este enfoque se centra en la creación de máquinas que no solo piensen como humanos, sino que también actúen como tales. El objetivo es replicar el comportamiento humano de forma que sea reconocible y natural. Los asistentes virtuales pueden llegar a mantener conversaciones con los usuarios y las usuarias que imiten todos los componentes humanos.

3. Sistemas que piensan racionalmente

Este enfoque se basa en la lógica formal y en la capacidad de realizar inferencias correctas. Los sistemas que piensan racionalmente siguen reglas lógicas estrictas para llegar a conclusiones, sin necesariamente imitar el pensamiento humano. De hecho, algunas de las decisiones de los humanos no responden a lógicas racionales. Programas como Prover9 pueden demostrar teoremas matemáticos utilizando la lógica formal, a partir de reglas de deducción lógica.

4. Sistemas que actúan racionalmente

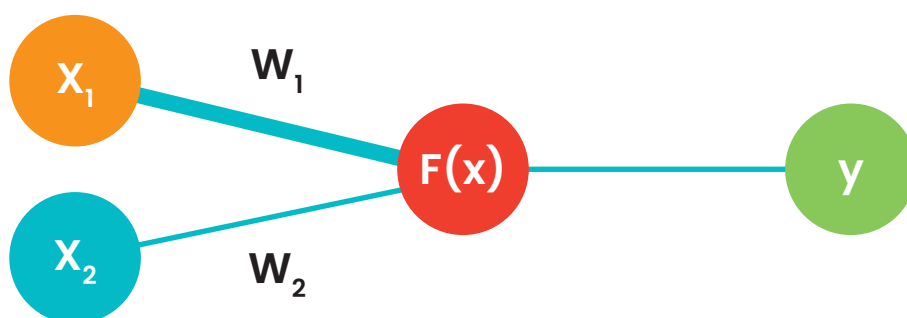
El enfoque de los sistemas que actúan racionalmente se centra en la creación de agentes inteligentes que toman decisiones óptimas para conseguir sus objetivos. Estos agentes utilizan la información disponible para actuar de forma que maximice su éxito en una tarea determinada. Los vehículos autónomos pueden tomar decisiones óptimas basadas en una gran cantidad de información en tiempo real.

“La inteligencia artificial generativa ha sido el resultado de un largo camino, de una sucesión de intuiciones encadenadas”

1.3. La IA Generativa

La inteligencia artificial generativa ha sido el resultado de un largo camino, de una sucesión de intuiciones encadenadas. El punto de partida fue [un artículo](#) de Frank Rosenblatt en 1958, quien introdujo el concepto de *perceptrón*, inspirado en las neuronas del cerebro humano. Imaginemos, por ejemplo, que queremos evaluar la viabilidad de un nuevo producto de agua con sabor a limón. El éxito del agua depende de varios factores, pero podríamos considerar que los más relevantes son el gusto y el precio: Con el gusto, incrementaremos la cantidad de limón y con el precio subiremos más o menos respecto al valor de un agua convencional. El resultado sería un perceptrón que tendría dos entradas: x_1 es el gusto y x_2 es el precio.

La forma más simple de perceptrón sería aquella en la que x_1 y x_2 fueran dos valores dicotómicos (0 y 1), de modo que 0 fuera un valor incorrecto (demasiado limón o un precio demasiado alto) y 1 un valor correcto. La función de activación de la neurona, el *bias*, determinaría que solo en caso de que acertemos los dos valores, gusto y precio, la neurona se activará. El perceptrón sería más útil si pudiéramos trabajar con un rango de sabor (porcentaje de limón) y con un rango de precio (incremento respecto al valor de un agua sin sabor). Enseguida nos daríamos cuenta de que los dos valores de entrada no tienen el mismo peso: el cliente sería más sensible al precio, por lo que las variaciones en el precio tendrían mucho más efecto en el resultado final.



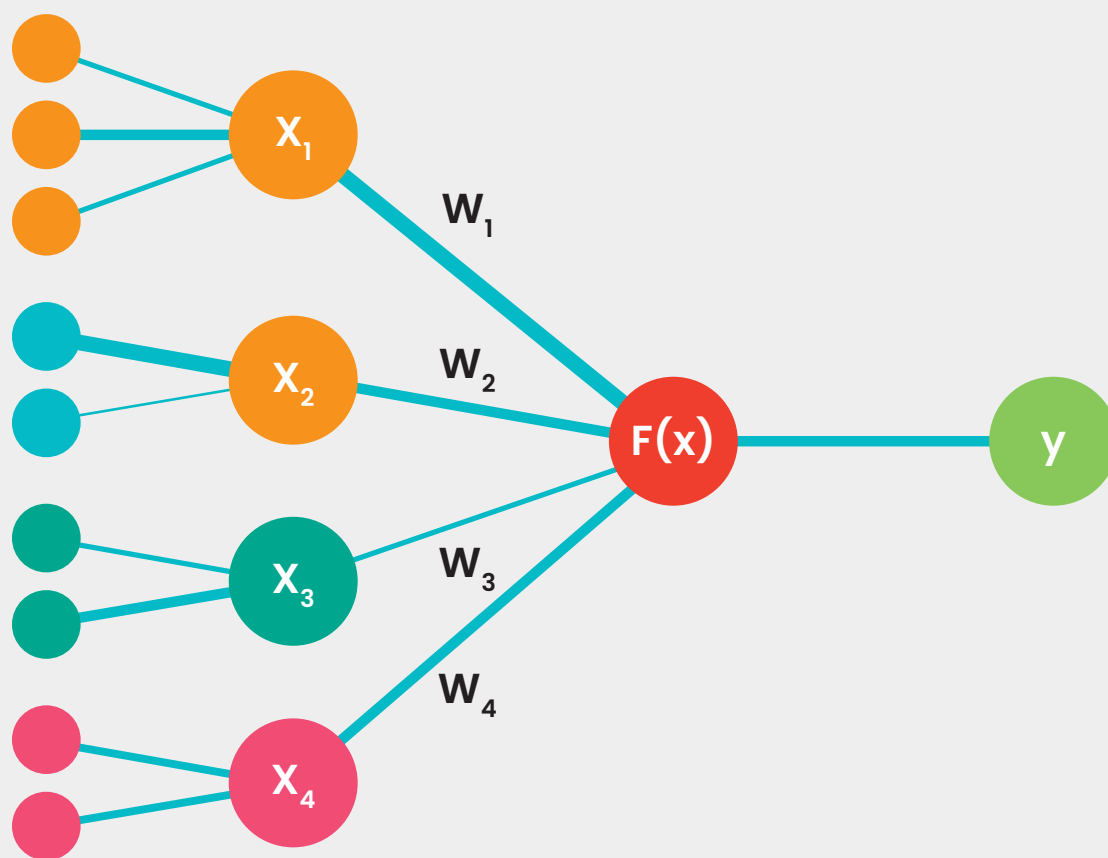
Pero el problema es más complejo. El éxito de un producto depende de otras variables, como la distribución, el envase, o el márketing del producto. Así que podríamos crear un perceptrón más complejo, con más entradas, que fuesen x_3 (el número de supermercados), x_4 (el tipo

de envase) y x_5 (los recursos destinados a promoción). A medida que complicamos el modelo, resulta más evidente la dificultad de sintetizar la información con un valor de entrada. Por ejemplo, el envase puede diferenciar la forma de la botella, el color del tapón y la etiqueta

que cubre la botella con información e imágenes. La distribución también podría diferenciarse entre el número de supermercados, el número de unidades por supermercado o la ubicación de las aguas en las estanterías. De modo que la red inicial tenía una sola neurona con varias entradas y ahora podríamos crear

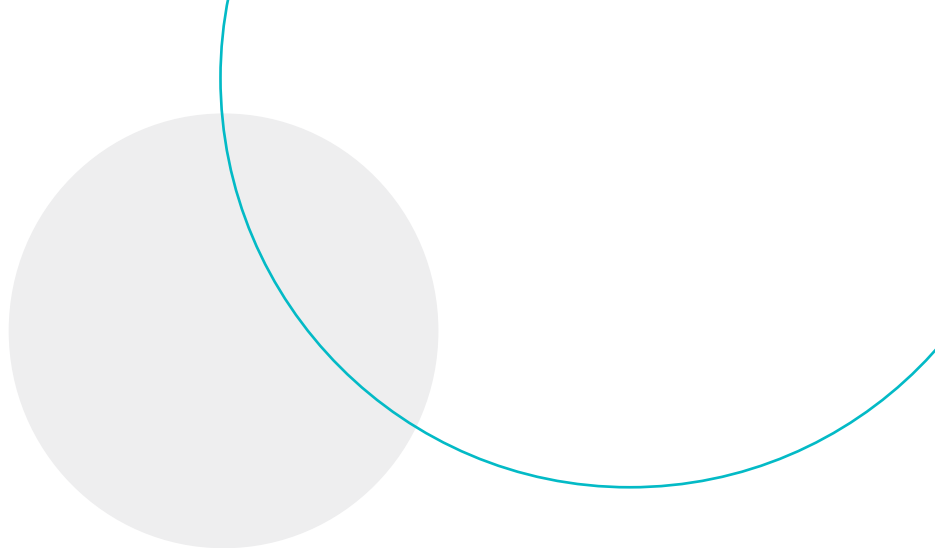
una red con varias neuronas más especializadas y varias capas. Las economistas y los economistas saben que el éxito de los productos depende de muchos factores complejos interrelacionados entre sí, en múltiples capas. Éste es el principio de una red neuronal.

Red neuronal



Las redes neuronales toparon con dificultades técnicas para su desarrollo, hasta que en 1986 Rumelhart, Hinton y Williams introdujeron uno de los conceptos clave de la IA moderna: la retropropagación (*backpropagation*). Imaginemos que la multinacional que quiere incorporar las aguas con gusto puede realizar varios ensayos en provincias determinadas. Con los resultados de las ventas de cada provincia puede intentar reconstruir los pesos, las ponderaciones, de acuerdo con los resultados. Puede ver, por

ejemplo, que el envase desempeña un papel muy secundario porque las variaciones en los envases no provocan cambios sustanciales en las ventas; también puede averiguar la importancia de la distribución y, específicamente, de la ubicación en las estanterías en el volumen de ventas. Y aquí intuimos un factor clave de esta técnica: cuantas más iteraciones hagamos, más robustas serán las relaciones entre las neuronas, es decir, los pesos de cada factor en el resultado final.



Recordemos el ejemplo del sexador o sexadora de pollos. Unos pocos casos no sirven para averiguar el resultado. Se necesitan tres años de ensayo error para que el aprendizaje sea efectivo. Y ese fue el criterio que utilizó AlphaZero. Cada partida perdida genera una información que puede volver atrás con criterios de probabilidad: es más fácil perder si el rey no está enrocado, si los alfiles no controlan las diagonales o si la dama no está centralizada. Los criterios posicionales o estratégicos no se han definido a priori, sino que son el resultado de miles de partidas, que permiten la retropropagación de la información de las “neuronas”. El sistema no solo aprende a valorar las jugadas, sino que puede determinar cuál es la solución con mayor probabilidad de éxito, es decir, es capaz de jugar, de generar una salida.

En 2017, [el artículo de Ashish Vaswani y sus colaboradores](#) inició la última revolución de la inteligencia artificial y de las redes neuronales: los *transformers*. La innovación clave del artículo, llamado “Attention is all you need” es la introducción del mecanismo de atención autodireccional, que permite al modelo centrarse en diferentes partes de una entrada para determinar cómo procesar y responder a esta entrada. Mientras que los modelos anteriores seguían una secuencia (como en los textos, que leemos de la primera palabra a la última), los transformers aprenden a identificar las palabras relevantes y las relaciones entre ellas: No leen un texto de forma secuencial, sino que lo interpretan en su conjunto, y buscan los vínculos entre las unidades. Esto es especialmente importante para entender el contexto de una frase o la vinculación entre los elementos, como los pronombres. En la frase “La irrupción de la inteligencia artificial en

el aula no la transformará radicalmente, pero sí que provocará cambios” el pronombre “la” hace referencia al aula y no a la inteligencia artificial. Para un humano, esta relación puede parecer muy evidente, sin embargo, los modelos previos a los transformers tenían dificultades para establecer vínculos entre los elementos. El artículo demostró que los modelos basados exclusivamente en la atención pueden obtener resultados superiores en tareas de traducción automática, manteniendo o mejorando la velocidad de procesamiento. Esta idea fue revolucionaria porque simplificó los modelos existentes y mejoró su eficacia y su eficiencia.

En noviembre de 2022, OpenAI presentó ChatGPT. Es un chat, esto es, permite establecer una conversación con un sistema de preguntas y respuestas encadenadas. Pertenece a la IA generativa porque no solo “entiende” los textos, sino que es capaz de redactar un nuevo texto de forma automática. Ha sido preentrenado para evitar resultados no deseados, incorporar límites éticos y morales, y optimizar las respuestas a partir de un proceso de recompensas; se basa en el modelo transformer introducido en 2017. Entrenado con un volumen inmenso de información textual, y gracias a unos servidores con una potencia de cálculo y memoria inéditas. ChatGPT ha permitido que usuarias y usuarios puedan acceder a un entorno que parece hablar con la capacidad de un sabio y con la lucidez de un estratega. Enseguida aparecerán otros modelos como Claude, Gemini, Llama o Copilot. Dentro del vientre del sistema, [un inmenso grafo de relaciones entre palabras](#) se va construyendo en cada nueva iteración, como si fueran las combinaciones precisas de la Biblioteca que un día imaginó Jorge Luis Borges.

Glosario

Inteligencia artificial	Capacidad de las máquinas para realizar tareas que requieren inteligencia humana como reconocimiento de voz, toma de decisiones y traducción de lenguas.
Inteligencia artificial general	También conocida como IA fuerte, se refiere a una hipotética IA con capacidades cognitivas similares o superiores a las humanas en casi todos los ámbitos.
Inteligencia artificial blanda	También llamada IA débil o estrecha, son sistemas diseñados para realizar tareas específicas sin tener una inteligencia general.
Modelo de red neuronal	Estructura computacional inspirada en el cerebro humano, compuesta por capas de neuronas artificiales que procesan la información y aprenden a través del ajuste de los pesos de las conexiones.
Aprendizaje profundo	Rama de la Inteligencia Artificial que utiliza redes neuronales con muchas capas (profundas) para analizar y aprender grandes cantidades de datos.
Aprendizaje automático (machine learning)	Rama de la IA que se centra en el desarrollo de algoritmos que permiten a los sistemas aprender a partir de datos.
Procesamiento del Lenguaje Natural	Rama de la Inteligencia Artificial que permite a las máquinas comprender, interpretar y generar lenguaje humano, facilitando la interacción entre humanos y ordenadores en lenguaje natural.
Algoritmo de retropropagación	Técnica utilizada en el entrenamiento de redes neuronales para ajustar los pesos de las conexiones, según el error entre la salida prevista y la salida deseada.
Redes neuronales recurrentes	Tipo de red neuronal diseñada para trabajar con datos secuenciales, como texto o señales temporales, donde las conexiones entre las neuronas forman un ciclo que permite recordar información anterior.
Aprendizaje por refuerzo	Técnica de aprendizaje automático en la que un agente aprende a tomar decisiones mediante la interacción con un entorno.
Sistemas expertos	Sistemas de IA que emulan la capacidad de toma de decisiones de un experto humano.

Referencias

- Borges, J.L. (1944). *Ficciones*, Sur, Buenos Aires.
- López de Mántaras i Badia, R. (2024); *100 cosas que cal saber sobre la inteHigència artificial*, Valls: Cossetània
- McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*, 5, 115-133.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Elsevier.
- Rosenblatt, F. (1958). *The perceptron: a probabilistic model for information storage and organization in the brain*. *Psychological review*, 65(6), 386.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- Russell, S.J. i Norvig, P. (2003); *Artificial Intelligence: A modern approach*, Prentice Hall, 2a. edició
- Simon, H. A., & Newell, A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3), 11-126.
- Turing, A. M. (1950) *Computing machinery and intelligence*, *Mind*
- Vaswani, A. et al. (2017); Attention is all you need. *Advances in neural information processing systems*, 2017, vol. 30.

02

Preguntar es un arte

“La pregunta es la más creativa de las conductas humanas”, decía el publicista Alex Osborn (Nueva York, 1888 - Búfalo, 1966). Fue cofundador de la agencia BBDO, considerado uno de los principales teóricos publicitarios de la primera mitad del siglo XX y creador del concepto “*brainstorming*” o “lluvia de ideas” (Centro de Documentación Publicitaria, 2024). Preguntar es un arte, no solo en el campo de la creatividad, sino también en ámbitos como la educación, la terapia, la investigación, el periodismo o la medicina, entre muchos otros.

El arte de preguntar también se ha vuelto relevante en el ámbito de la inteligencia artificial, ya que la formulación de un *prompt* adecuado, al igual que una pregunta pertinente, puede dar lugar a una respuesta más o menos apropiada a nuestros objetivos. La ingeniería de *prompts* ha experimentado un auge significativo con el desarrollo y uso generalizado de modelos de lenguaje extensos (*Large-Scale Language Models*, LLMs), como Claude, Gemini o ChatGPT. El objetivo de este capítulo es entender cómo optimizar los *prompts* para obtener mejores resultados, comprender cuáles son los elementos

que configuran un *prompt* (instrucción, contexto, datos de entrada e indicador de salida), y aprender algunas técnicas de ingeniería de *prompts*.

Antes, sin embargo, debemos preguntarnos: ¿qué es un *prompt*? “*Un prompt es un conjunto de instrucciones proporcionadas a un LLM, personalizándolo y/o mejorando o afinando sus capacidades*”, según definen White et al. (2023: 1). Estos autores afirman que la calidad de los resultados generados por una conversación con un LLM está directamente relacionada con la calidad de las indicaciones proporcionadas por el usuario o la usuaria. Por lo tanto, la calidad de las respuestas no solo depende de los algoritmos y los datos de entrenamiento, sino también de las indicaciones utilizadas (Lo, 2023: 1): “*Un prompt bien diseñado puede provocar una respuesta significativa e informativa, mientras que uno mal construido puede dar lugar a salidas irrelevantes o absurdas*”.

Algunos autores, como Korzynsky et al. (2023), analizan la ingeniería de *prompts* como una competencia digital. Para ello, utilizan el marco DigComp, que se estructura en

cinco áreas de competencias básicas: 1) alfabetización informacional y de datos, 2) comunicación y colaboración, 3) creación de contenido digital, 4) seguridad, y 5) resolución de problemas (Vuorikari, Kluzer y Punie, 2022). En el análisis, la mayoría de las competencias digitales de estas cinco áreas, excepto en el caso de la seguridad, se superponen a las necesidades de la ingeniería de *prompts* de los sistemas de inteligencia artificial (Korzynsky et al., 2023: 28)..

Preguntar es un arte. En el periodismo, una pregunta adecuada puede ser un buen titular. En medicina, una pregunta adecuada puede ser un buen diagnóstico. En investigación, una pregunta correctamente formulada puede ser un descubrimiento científico. Y en IA, una pregunta apropiada, es decir, un *prompt* bien elaborado, puede generar una respuesta pertinente. “En resumen, a nivel superficial, la IA generativa es fácil de usar; sin embargo, los resultados más sofisticados requieren la intervención de personas cualificadas y deben ser evaluados de forma crítica antes de ser utilizados”, concluyen Miao & Holmes (2023: 12).

2.1. Ingeniería de *prompts*

Un *prompt* se define como una instrucción o consulta específica que se proporciona a un modelo de lenguaje para guiar su comportamiento y generar los resultados deseados (Giray, 2023). Según el mismo autor, un *prompt* incorpora cuatro elementos: instrucción, contexto, datos de entrada e indicador de salida. Autores como Korzynsky et al. (2023) mencionan también los mismos elementos con ciertos matices, y otros autores incorporan elementos como el tono, el rol, los ejemplos (o *shots*) y el tipo o formato de salida (Learn Prompting, 2024a). A continuación, se explicará el significado de cada uno de estos elementos.

- **Instrucción, tarea o pregunta.** Parte del *prompt* en la que se da una indicación al sistema, se hace una solicitud o una consulta para guiar su comportamiento y obtener los resultados esperados (Giray, 2023). Por ejemplo: “Escribe un correo electrónico a un cliente agradeciéndole su asistencia al evento de presentación de un nuevo producto”.

- **Contexto o información adicional.** Parte del *prompt* en la que se proporciona información previa al modelo para generar una respuesta más relevante y precisa (Giray, 2023). Puede incluir detalles sobre el rol que debe asumir el modelo para realizar la tarea o cualquier otra información adicional sobre la situación (Korzynsky et al., 2023). Por ejemplo: “Eres un científico de datos con tres años de experiencia en una empresa tecnológica”.
- **Datos de entrada.** Parte del *prompt* en la que se proporcionan datos o información que el modelo debe usar para llevar a cabo la instrucción, tarea o responder a la pregunta (Korzynsky, 2023). Por ejemplo: “Proporciona una receta innovadora que incluya verduras y proteínas con ingredientes frescos y de temporada”.
- **Indicador o formato de salida.** Parte del *prompt* donde se especifica el tipo o formato del resultado deseado (Giray, 2023). Puedes especificar, por ejemplo, la longitud del texto: “Escribe una introducción de 30-40 palabras sobre el arte de preguntar y la ingeniería de *prompts*”. O el formato en el que quieres el resultado (en este caso, en forma de lista): “Enumera las cinco mejores herramientas de inteligencia artificial generativa”. También podrías especificar el tono, por ejemplo: “Escribe un texto en tono humorístico sobre la ingeniería de *prompts*” (ver más detalles en la sección 2.4). Finalmente, otra estrategia sería proporcionar un ejemplo de cómo quieres el resultado final.

Según la UNESCO, la ingeniería de *prompts* se refiere a los “procesos y técnicas para crear entradas que produzcan una salida de IA que se asemeje más a la intención deseada del usuario” (Miao & Holmes, 2023: 11). Es una disciplina emergente que “estudia las interacciones con -y la programación de- sistemas computacionales LLM en fusión para resolver problemas complejos a través de interfaces de lenguaje natural” (Schmidt, et al., 2023: 1). Esta disciplina ganará relevancia en el ámbito de la inteligencia artificial, según Korzynsky et al. (2023), ya que permite mayor control sobre los resultados y garantiza la creación del contenido deseado si las

indicaciones son adecuadas. Los autores añaden: “*Los prompts efectivos contribuyen a mejorar la precisión guiando a los modelos de IA para generar contenido más relevante y valioso*” (Korzynsky et al., 2023: 26). Otros estudios han demostrado que la construcción de *prompts* apropiados incrementa la fiabilidad de los sistemas de IA, como GPT-3 (Si et al., 2023).

“La ingeniería de *prompts* es el proceso de construcción de consultas o entradas (es decir, *prompts*) para modelos de lenguaje de IA con el fin de provocar las respuestas más precisas, coherentes y relevantes. En esencia, es el arte de afinar las preguntas o comandos proporcionados a los modelos de IA para optimizar su rendimiento y garantizar que produzcan los resultados deseados” (Lo, 2023: 1).

La ingeniería o diseño de *prompts* debe tener en cuenta dos principios: el establecimiento de un contexto y la operación de volver a preguntar (Lopezosa, Codina y

Ferran-Ferrer, 2023). Además del contexto, los *prompts* deben tener objetivos, audiencias y ejemplos para obtener mejores resultados (Lopezosa & Codina, 2023). La UNESCO hace recomendaciones específicas similares a la hora de crear los *prompts*: es necesario utilizar un lenguaje sencillo, claro y directo, evitando textos complejos o ambiguos; incluir ejemplos para ilustrar la respuesta deseada o el formato; añadir contexto; perfeccionar e iterar experimentando con diferentes variaciones; y, por último, ser ético (Miao & Holmes, 2023: 11).

Existen muchas recomendaciones, ejemplos y buenas prácticas para diseñar *prompts* para IA. Por un lado, se puede destacar la recopilación de recomendaciones para elaborar *prompts* de texto a texto que hacen Korzynsky et al. (2023) mediante el acrónimo AI PROMPT. Por otro lado, Torres-Salinas y Arroyo-Machado (2023) resumen las recomendaciones para ChatGPT, aunque también aplicables a otras IA, en cuatro puntos: 1. Claridad y precisión; 2. Evitar ambigüedades; 3. Proporcionar contexto; y 4. Experimentar y refinar.

Recomendaciones para elaborar *prompts* de texto a texto

ACRÓNIMO	RECOMENDACIÓN	DESARROLLO
A	Articular la instrucción	Especifica claramente la tarea a realizar, como 'escribir', 'clasificar', 'resumir' o 'traducir', y detalla cómo debe ser la salida (tabla, lista, código Python).
I	Indicar los elementos del <i>prompt</i>	Muestra al modelo dónde están las instrucciones y los datos de entrada, y cuál debería ser el formato de salida esperado.
P	Proporcionar señales de finalización y contexto	Ofrece al modelo señales claras de finalización, como tres puntos para continuar o un marcador como 'respuesta:' para indicar que se necesita una respuesta. Además, proporciona contexto al modelo explicando la tarea (por ejemplo, "Eres el gerente de un equipo técnico").
R	Refinar las instrucciones para evitar ambigüedades	Da instrucciones específicas al modelo y una descripción detallada de la tarea para evitar cualquier confusión o imprecisión.
O	Ofrecer retroalimentación y ejemplos	Para modelos conversacionales, como ChatGPT, la retroalimentación sobre las respuestas del modelo puede ayudarle a entender mejor la salida deseada. Además, ofrecer al modelo algunos ejemplos de respuestas esperadas (aprendizaje con pocos ejemplos) puede ayudarle a adaptar su estilo y forma de responder.
M	Gestionar la interacción	Trata al modelo como tu sparring, pidiéndole que proporcione contraargumentos o señale errores en tus ideas.
T	Seguir la longitud de token y la complejidad de la tarea	Dividir tareas complejas en pasos más pequeños para mejorar el rendimiento. Recuerda controlar la longitud de los tokens, manteniendo el <i>prompt</i> y la respuesta dentro de los límites de tokens del modelo (normalmente 4096 tokens para los LLMs disponibles comercialmente). La longitud de los tokens de un texto se puede comprobar aquí .

Fuente: Korzynsky et al., 2023: 31.

Además de las recomendaciones, se ha creado la metodología llamada GPEI, que corresponde a Goal Prompt Evaluation Iteration, y que consta de cuatro pasos: (1) definir un objetivo, (2) diseñar la indicación (o *prompt*), (3) evaluar la respuesta y (4) iterar (Velásquez-Henao, Franco-Cardona y Cadavid-Higuita, 2023: 12). Otros autores han desarrollado el marco CLEAR con cinco componentes que deben seguir los *prompts* (ver la tabla 1): Concisos, Lógicos, Explícitos, Adaptativos y Reflexivos (Lo, 2023).

Los cinco componentes del marco CLEAR

C	Concise	Brevidad y claridad en los ' <i>prompts</i> '.
L	Logical	Estructura y coherencia en los ' <i>prompts</i> '.
E	Explicit	Especificaciones claras de salida.
A	Adaptative	Flexibilidad y personalización en los ' <i>prompts</i> '.
R	Reflective	Evaluación y mejora continuas de los ' <i>prompts</i> '.

Fuente: Lo, 2023.

2.2. El valor del contexto

Si le pido a Claude que complete la frase con una sola opción "Me duele", probablemente contestará "Me duele la cabeza"; en algunas ocasiones, optará por "Me duele la espalda", y muy raramente por otras alternativas. Recordemos que los modelos de lenguaje son modelos probabilísticos y, como los predictores de texto, proporcionan la solución que consideran más adecuada. Dado que los dolores más comunes son de cabeza y espalda, el modelo sugiere una solución probable. Sin embargo, si el texto a completar propone "Debo ir al oftalmólogo porque me duele" o "hace mucho tiempo que no iba al gimnasio y me duele", las respuestas se ajustarán a la nueva información. Por eso, uno de los elementos más importantes en el diseño de *prompts* es el contexto. A menudo, el éxito de un *prompt*, entendido como la obtención de la respuesta deseada o esperada, radica justamente en la introducción del contexto adecuado

y necesario en el sistema de inteligencia artificial. "El contexto es cualquier información relevante que quieras que el modelo utilice cuando responda la pregunta o realice la instrucción", según se define en Learn Prompting (2024a). De acuerdo con De Juana (2024), es necesario incluir el contexto adecuado, es decir, insertar toda la información relevante y las instrucciones adicionales necesarias para mejorar la calidad de la respuesta. Este tipo de *prompts* se conocen como "contextual *prompts*" (Giray, 2023).

Para construir el contexto, en primer lugar, debemos tener en cuenta la instrucción, tarea o pregunta que queremos que desarrolle el sistema de inteligencia artificial. Además, para completar el contexto de un *prompt*, podemos añadir los siguientes elementos: datos de entrada, asignación de roles y uso de ejemplos. Algunas recomendaciones para mejorar los *prompts*, además de la inclusión de un contexto adecuado, son: ser específico, considerar las limitaciones, conocer al usuario objetivo (*user persona*), tener en cuenta los valores Top-K y Top-P y, finalmente, iterar (De Juana, 2024).

A continuación, se presentan algunos ejemplos de cada una de las técnicas de diseño de *prompts*, según el elemento que se utiliza para introducir la información relacionada con las recomendaciones vistas anteriormente, para mejorar el diseño de *prompts*.

Instrucción, tarea o pregunta. "El *prompting instruction*" o "*instructive prompt*", como lo denomina Giray (2023), consiste en dar una tarea o indicación a la IA y es la técnica de creación de *prompts* más simple, aunque se pueden dar instrucciones cada vez más complejas. Algunos de los verbos más utilizados para dar instrucciones son: escribir, traducir, categorizar o completar, entre otros. En el diseño de *prompts*, se debe ser específico, no solo en la construcción del contexto, sino también en la instrucción, tarea o pregunta que queremos que desarrolle el sistema de inteligencia artificial. Otro tipo de *prompt* que podríamos utilizar es el "*question-answer prompt*", que se estructura en torno a una pregunta de investigación (Giray, 2023).

Veamos un ejemplo de instrucción (De Juana, 2024):

Prompt general: "Háblame sobre el cambio climático".

Prompt específico:

- "Dame información detallada sobre las causas y efectos del cambio climático, centrándote específicamente en el papel de las emisiones de gases de efecto invernadero y las medidas de mitigación sugeridas por la comunidad científica".

Datos de entrada. Es la información que utiliza la IA, como una frase que debe traducir o un documento que debe resumir (Thomson, 2023). Dentro de este elemento, hablaremos de dos recomendaciones que también están relacionadas con el tipo de salida (ver más detalles en el apartado 2.3):

- Contemplar las limitaciones: En el *prompt* se pueden incorporar restricciones como el número de palabras deseado o el marco temporal (De Juana, 2024) para obtener respuestas con la información justa y necesaria.
- Valores Top-K y Top-P: Incluir estos valores en el *prompt* permite controlar la generación de texto con parámetros que afectan la diversidad y probabilidad de las palabras generadas.

Asignación de roles. El "*Role prompting*" es una técnica en la que se pide a la IA que adopte un perfil determinado (ver más detalles en el capítulo 4). Por ejemplo: actúa como un profesor de matemáticas o como un crítico de cine. Esta es una técnica para controlar el tono y estilo del texto generado, adaptándolo a la *user persona*, es decir, a la audiencia o público objetivo (Learn Prompting, 2024c; De Juana, 2024).

Uso de ejemplos. Otra técnica de diseño de *prompts* es la utilización de ejemplos o *shots*. Existen tres variantes de esta técnica (ver más detalles en el apartado 3):

- *Zero-shot prompting*: En esta variante, el *prompt* no contiene ningún ejemplo (Learning Prompting, 2024); es decir, el modelo hace predicciones sobre datos no vistos anteriormente sin necesidad de entrenamiento adicional (Korzynsky et al., 2023). Los *prompts* de instrucción o de rol que no contienen

ejemplos pueden considerarse dentro de esta variante. Ejemplo: "Capital de Francia".

- *One-shot prompting*: En esta variante, el *prompt* contiene un solo ejemplo (Learning Prompting, 2024d). Ejemplo: "Francia = París; Italia =".
- *Few-shot prompting*: El *prompt* contiene dos o más ejemplos (Learning Prompting, 2024); es decir, el modelo recibe una serie de ejemplos para adaptarse rápidamente a nuevas tareas (Korzynsky et al., 2023). Ejemplo: "Francia = París; Italia = Roma; Andorra =".

Si, a pesar de utilizar los diferentes elementos de un *prompt* y las recomendaciones anteriores, no se obtiene la respuesta deseada, será necesario iterar, es decir, probar varias aproximaciones, ajustando y refinando constantemente el modelo del *prompt* para obtener los resultados más relevantes (De Juana, 2024). Además, para diseñar el *prompt* podemos combinar las técnicas vistas anteriormente de la siguiente manera: a) Role & Instruction Prompting; y b) Context, Instruction & Few Shot Prompting (Learning Prompting, 2024b). De esta manera, obtendremos un "*mixed prompt*" (Giray, 2023).

Como hemos visto hasta ahora, hay una gran variedad de tipos de *prompts* o técnicas para diseñarlos, que van desde los modelos más básicos desde pregunta, pregunta + instrucción, instrucciones + datos de entrada o pregunta más ejemplos (Amatriain, 2023a) hasta los más avanzados como *megaprompts*, que contienen todos los elementos, cadenas de *prompts* que dividen las tareas en diferentes mensajes como sugiere Ana Ivars (Gutiérrez, 2023), *prompt* de cadena de pensamiento (*Chain of thought prompting*) o árbol de pensamiento (*Tree of Thought*) que veremos en otros capítulos (Amatriain, 2023b). A continuación, veremos los *prompts* del ámbito docente y del ámbito académico que pueden ser más útiles para el profesorado universitario y de educación secundaria y postobligatoria.

Ámbito docente. Los tipos de *prompts* que pueden ser útiles en la educación son (Morales-Chan, 2023):

- *Prompts* secuenciales: Su objetivo es guiar

conversaciones de manera lógica y estructurada a través de una serie de textos previos para obtener respuestas contextualizadas y detalladas.

- **Prompts comparativos:** La IA debe comparar varias situaciones para obtener resultados específicos y proporcionar análisis detallados.
- **Prompts argumentales:** El objetivo es que la IA genere un argumento a favor o en contra de una idea para fomentar la reflexión crítica y el desarrollo de argumentos estructurados.
- **Prompts de perspectiva profesional:** Se le pide a la IA que adopte un rol específico y describa un tema en un contexto determinado para mejorar la calidad de la respuesta.
- **Prompts de lista de deseos:** Estructuras utilizadas para obtener respuestas más específicas y relevantes, ya que se le pide a la IA que cumpla con unos requisitos específicos.

Ámbito académico. Los principales tipos de *prompts* en este ámbito son (Lopezosa y Codina, 2023):

- **Prompts reactivos:** Encadenan interacciones a partir de cada respuesta.
- **Prompts estructurales:** Se solicita una respuesta articulada en varios puntos y posteriormente se pide una ampliación de cada punto.
- **Prompts de rol:** La IA adopta un rol, como hemos visto anteriormente. Además, es similar al *prompt* de perspectiva profesional.
- **Prompts con indicaciones de audiencias:** Se especifica la audiencia a la que debe dirigirse la respuesta.
- **Prompts con objetivos:** Se incluyen los objetivos del texto que se va a generar.

La siguiente tabla muestra un resumen de las recomendaciones básicas para la elaboración de *prompts*.

Estrategias y tácticas de ingeniería de *prompts*

ESTRATEGIAS	TÁCTICAS
Escribir instrucciones claras	Incluir detalles en la consulta para obtener respuestas más relevantes Pedir al modelo que adopte una personalidad Utilizar delimitadores para indicar claramente las diferentes partes de la entrada (input) Especificar los pasos necesarios para completar una tarea Proporcionar ejemplos Especificar la longitud deseada de la salida (output)
Proporcionar un texto de referencia	Instruir al modelo para que la respuesta utilice un texto de referencia Instruir al modelo para que responda con citas de un texto de referencia
Dividir las tareas complejas en subtarear más sencillas	Utilizar la clasificación de intenciones para identificar las instrucciones más relevantes para una consulta de usuario Resumir o filtrar el diálogo anterior en aplicaciones de diálogo que requieren conversaciones muy largas Resumir documentos largos por partes y construir un resumen completo recursivamente
Dar tiempo al modelo para "pensar"	Instruir al modelo para que elabore su propia solución antes de llegar a una conclusión precipitada Utilizar un monólogo interior o una secuencia de consultas para ocultar el proceso de razonamiento del modelo Preguntar al modelo si ha perdido algo en pasos previos
Utilizar herramientas externas	Utilizar la búsqueda basada en incrustaciones para implementar una recuperación de conocimiento eficiente Utilizar la ejecución de código para hacer cálculos más precisos o llamar a API externas Dar acceso al modelo a funciones específicas
Probar cambios sistemáticamente	Evaluar los resultados del modelo con referencia a las respuestas estándar

Fuente: OpenAI developer platform (2024).

2.3. Los tipos de salida

Otro de los elementos de un *prompt* es el tipo de salida. Si queremos obtener un tipo de salida específico, es necesario indicárselo al sistema de inteligencia artificial. Algunas de las indicaciones que podemos darle a la IA y que determinarán el tipo de salida son las siguientes (Maurine, s.d.; Torres-Salinas & Arroyo-Machado, 2023):

- Longitud: una frase, un párrafo o un número determinado de palabras/caracteres.
- Estructura: cronología, comparación, causas y efectos, problema y solución, biográfico, narrativo, etc.
- Tipo de texto: correo electrónico, publicación para una red social, informe profesional, comunicado de prensa, texto literario, etc.
- Formato: enumeración con viñetas, tabla, código, texto plano, preguntas y respuestas (Q&A), etc.
- Plataforma: LinkedIn, Twitter, Facebook, etc.

Formatos que se pueden pedir a la IA

FORMATO	DESCRIPCIÓN
Texto plano	Para respuestas directas y generales
Formato de Preguntas y Respuestas (Q&A)	Adecuado para entrevistas o sesiones de preguntas
Lista numerada o con viñetas	Útil para enumerar puntos, pasos o elementos
Código fuente (Python, R, JavaScript, etc.)	Para ejemplos de programación
Markdown	Para la estilización de texto simple
LaTeX	Utilizado para representar ecuaciones matemáticas y para publicaciones científicas
CSV (Valores Separados por Comas)	Para representar datos en forma tabular
Tablas textuales	Tablas creadas con caracteres, útiles para comparaciones o resúmenes
Diagramas Ascii Art	Para ilustraciones textuales simples
HTML (básico)	Estructuración simple de contenido web
XML (básico)	Estructuración de datos jerárquicos
JSON (básico)	Para estructurar datos en notación de objeto de JavaScript
YAML	Representación legible para humanos de datos de serialización
SQL (básico)	Para mostrar consultas de bases de datos
Esquemas y diagramas (representación textual)	Para describir estructuras de datos o arquitecturas
Emoticonos o emoji	Simplificación de un texto en un emoji
Enfatizar en la respuesta	Podemos indicar que destaque parte de la respuesta en negrita, cursiva, etc.

Fuente: Torres-Salinas & Arroyo-Machado (2023).

2.4. El tono

El tono es otro elemento de los *prompts* que está relacionado con el tipo de salida y es una de las opciones más interesantes de la inteligencia artificial. El uso de un tono específico permite adaptar el texto o la respuesta del sistema de IA a nuestra audiencia, o utilizar nuestro propio tono, el de nuestra marca o empresa, para crear textos personalizados. Al determinar el tono, se pueden generar textos más divulgativos o dirigidos a diferentes tipos de público (por ejemplo, solicitando un texto para un público infantil o para estudiantes), o textos más precisos (por ejemplo, pidiendo un tono profesional, académico, científico, experto, etc.). En otras palabras, es una forma de adaptar los textos al *buyer persona* o *user persona*, es decir, a la audiencia o público objetivo.

Según una infografía creada por Ana Ivars (Gutiérrez, 2023), se pueden crear textos con una gran variedad de tonos o estilos de escritura: formal, persuasivo, descriptivo, narrativo, inspirador, informal, cercano, emotivo, amable, técnico, respetuoso e irónico. Por su parte, Torres-Salinas y Arroyo-Machado (2023) establecen cuatro tipos de escritura: literaria, académica, creativa y técnica. Sin embargo, existen muchos otros tonos o adjetivos que podríamos utilizar, como neutro, optimista, amistoso, emocional, racional, informativo, sensacionalista, crítico, opinativo, entre otros.

Como se mencionó anteriormente, se pueden usar las técnicas Top-K y Top-P, que influyen en la generación de texto al afectar la diversidad y probabilidad de las palabras generadas, según De Juana (2024). Por un lado, el Top-K determina cuántas palabras se seleccionan en cada paso de la generación, ofreciendo mayor diversidad con valores más altos y mayor determinismo con valores bajos. Por ejemplo, un valor entre 1 y 5 produce respuestas concisas, entre 5 y 10 respuestas equilibradas, y por encima de 20, respuestas más creativas. Por otro lado, el Top-P selecciona palabras hasta que la probabilidad acumulada supere un umbral definido, ofreciendo mayor diversidad con valores altos y

más determinismo con valores bajos, en un rango que va de 0,0 para respuestas formales a 1,0 para respuestas casi aleatorias. Especificar estos valores en las solicitudes de información mejora la calidad y relevancia de las respuestas generadas. Estos dos criterios se encuentran en la sección de playground de ChatGPT, lo que nos permite manipular los resultados usando los parámetros mencionados.

Plantilla de *prompt* con diferentes elementos

Act as

```
{{persona you want ChatGPT to simulate}}.
```

I am

```
{{persona who's giving instructions}}.
```

I want you to

```
{{task}}.
```

Constraints:

- {{constraint 1}}
- {{constraint 2}}
- {{constraint 3}}

Output format:

- {{format instruction 1}}
- {{format instruction 2}}

La plantilla contiene asignación de roles, instrucciones, tarea, restricciones de salida y formatos de salida. Fuente: Maurine (s.f.).

Práctica. El *prompt* perfecto

La ingeniería de *prompts* busca la mejor pregunta para una respuesta esperada. Además del tono, el formato de salida o el tipo de pregunta, el contexto es esencial para lograr una salida adecuada. El objetivo de esta práctica es comparar una salida del modelo con una pregunta genérica y compararla con aquella que ha sido formulada con un buen contexto. Podemos partir de una pregunta muy genérica como, por ejemplo, “Una ciudad metropolitana manifiesta un problema de acceso a la vivienda, especialmente entre la población joven. ¿Qué medidas podría aplicar el ayuntamiento para hacer frente a esta tensión?”. Los modelos de lenguaje (como ChatGPT o Copilot) proporcionarán respuestas genéricas, que podrían ser aplicadas en cualquier ciudad del mundo.

Para afinar la respuesta, los grupos de estudiantes deben buscar el contexto que ayude a elaborar una pregunta con una información más precisa. Para ello, los estudiantes deberían buscar información sobre:

- Estructura de los hogares (número de miembros)
- Evolución del precio de la vivienda
- Evolución de la renta media
- Proporción entre contratos de alquiler y propiedad
- Porcentaje de viviendas vacías
- Evolución del padrón
- Estructura de la población por edades, por renta o por origen
- Movilidad obligada
- Evolución de la actividad turística
- Distribución de los valores por barrios o distritos
- Valores de los municipios vecinos

A medida que contextualizamos la ciudad, las posibles soluciones varían porque se evidencian los factores que más influyen en el acceso a la vivienda y las singularidades de la ciudad de estudio. El *prompt* también varía si pedimos que asuma determinados roles, que pueden modificar la perspectiva, como por ejemplo urbanistas especializados en vivienda, activistas vecinales o economistas. La comparación entre los resultados, de acuerdo con los diferentes contextos planteados (y los diferentes roles), ayuda a entender la importancia del contexto y la selección precisa de información fiable, sobre el elemento de estudio.

Práctica. Variaciones sobre una pregunta

En esta práctica utilizaremos el “expert prompting” (Amatriain, 2023b) para observar cómo influye el uso de una pregunta u otra en la respuesta de la IA. Podemos situar la práctica en el ámbito del periodismo y pedir a los estudiantes que preparen un reportaje sobre un tema concreto, como el cambio climático, la violencia de género o el acceso a la vivienda. Para realizar el reportaje, tendrán que consultar a diferentes expertos. Para desarrollar el expert prompting, los estudiantes deben seguir tres pasos:

- Pedir a la IA que identifique expertos en el tema.

- Pedir a la IA que responda como si fuera alguno de estos expertos.
- Tomar una decisión final contrastando las diferentes respuestas generadas.

Para desarrollar la práctica, los estudiantes tendrán que utilizar dos IA: Claude y Gemini. Deberán realizar dos entrevistas a dos expertos/as diferentes, haciendo varias variaciones de una misma pregunta y llevando a cabo las entrevistas con ambas IA. Con los resultados, tendrán que elaborar un reportaje que incluya las declaraciones de los expertos entrevistados.

Esta práctica podría tener variaciones. En el campo del periodismo, se podría pedir que la IA adoptara roles de personajes de la actualidad política, económica, social o cultural. Además, podría ser una práctica previa a la realización de la entrevista real. La práctica también se podría adaptar a otros ámbitos:

- Marketing. Podría servir para la construcción del buyer persona o la creación de argumentos de venta de productos y servicios.
- Medicina. Podría servir como práctica para comunicar malas noticias a los pacientes o para realizar la entrevista clínica para llegar a un diagnóstico.

Práctica. Construir contextos: La documentación.

La documentación es una disciplina científica que estudia el comportamiento y las propiedades de la información con el fin de hacerla accesible. Tiene un carácter holístico, y su objetivo principal abarca la recolección, el almacenamiento, el análisis y la difusión de la información. Para construir contextos en el diseño de *prompts*, podemos utilizar algunas de sus técnicas, lo que nos permitirá crear prácticas para los estudiantes. La construcción de roles (personajes históricos, literarios, etc.) o el desarrollo de estudios de caso pueden ser dos maneras de lograrlo.

En esta práctica, podríamos construir un contexto a partir de los repositorios de datos disponibles actualmente en Internet. Por ejemplo, Google nos ofrece un buscador para localizar datasets, es decir, conjuntos de datos. Los estudiantes podrían, utilizando la IA y los conjuntos de datos, elaborar informes sobre sectores o temas concretos como educación, marketing o turismo. Los datasets más utilizados suelen ser los relacionados con educación, clima, cáncer, criminalidad, fútbol o perros, entre otros (Noy, 2020).

La propuesta podría consistir en que los estudiantes se pongan en la situación de tener que determinar la política educativa de un país para los próximos cinco años. Para ello, deberán analizar los datos educativos, elaborar un informe y proponer políticas que puedan ser aplicadas. El informe deberá contener datos obtenidos a través del buscador de datasets de Google, los cuales serán tratados mediante Gemini. Además, el informe podrá ser redactado con la ayuda de esta IA, siempre y cuando la información sea contrastada y verificada. En los anexos de la práctica se deberán incluir los *prompts* y las respuestas obtenidas de la IA. Asimismo, dentro del informe podrán integrarse textos generados por la IA, con las citas correspondientes, así como las modificaciones necesarias para asegurar que la información sea fiable.

Referencias

- Amatriain, X. (2023a). Prompt Engineering 101: Introduction and resources. Amatriain Blog. Data de recuperació: 14/03/2024. <https://amatriain.net/blog/PromptEngineering>
- Amatriain, X. (2023b). Prompt Engineering 201: Advanced methods and toolkits. Amatriain Blog. Data de recuperació: 14/03/2024. <https://amatriain.net/blog/prompt201#expert>
- Centro de Documentación Publicitaria (2024). Protagonistas: Alex Faickney Osborn. Data de recuperació: 14/03/2024. <https://www.lahistoriadelapublicidad.com/protagonista-174/alex-faickney-osborn>
- De Juana, R. (2024). Cómo mejorar tus Prompts en seis cómodos pasos. MCPRO. Data de recuperació: 14/03/2024. <https://www.muycomputerpro.com/2024/01/24/como-mejorar-tus-prompts-en-seis-comodos-pasos>
- Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. Annals of Biomedical Engineering, 1-5
- Gutiérrez, D. (2023). Hoja de truco de ChatGPT, de Ana Ivars. Data de recuperació: 14/03/2024. https://www.linkedin.com/posts/davidgutierrezamd_%C3%BAltima-hoja-de-trucos-de-chat-gpt-actividad-7098491435875749888-qgFw/?originalSubdomain=es
- Korzynski, P., Mazurek, G., Krzykowska, P., & Kurasniski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. Entrepreneurial Business and Economics Review, 11(3), 25-37.
- Learn Prompting (2024a). Formalizing Prompts. Data de recuperació: 14/03/2024. <https://learnprompting.org/docs/basics/formalizing>
- Learn Prompting (2024b). Combining techniques. Data de recuperació: 14/03/2024. https://learnprompting.org/docs/basics/combining_techniques
- Learn Prompting (2024c). Assigning Roles. Data de recuperació: 14/03/2024. <https://learnprompting.org/docs/basics/roles>
- Learn Prompting (2024c). Showing examples. Data de recuperació: 14/03/2024. https://learnprompting.org/docs/basics/few_shot
- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. The Journal of Academic Librarianship, 49(4).
- Lopezosa, C., & Codina, L. (2023). Inteligencia Artificial y métodos cualitativos de investigación: ChatGPT para revisiones de la literatura y análisis de entrevistas semiestructuradas. Data de recuperació: 14/03/2024. http://eprints.rclis.org/44296/1/IA-ACADEMICO-ChatGPT%20ATLAS-SCOPING-2023_vf-2.pdf
- Lopezosa, Carlos; Codina, Lluís; Ferran-Ferrer, Núria (2023) ChatGPT como apoyo a las systematic scoping reviews: integrando la inteligencia artificial con el framework SALSA. Colección del CRICC. Barcelona: Universitat de Barcelona. Data de recuperació: 14/03/2024. <http://hdl.handle.net/2445/193691>
- Maurien, F. (s.d.). ChatGPT for Educators | Prompting 101. Data de recuperació: 14/03/2024. <https://wooclap.notion.site/ChatGPT-for-Educators-Prompting-101-0b4c015d7b8a4e1f834c1f61504042f2>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO <https://discovery.ucl.ac.uk/id/eprint/10176438/>
- Morales-Chan, M. A. (2023). Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza. Data de recuperació: 14/03/2024. <http://biblioteca.galileo.edu/tesario/handle/123456789/1348>
- Noy, N. (2020). Discovering millions of datasets on the web. Google The Keyword. Data de recuperació: 14/03/2024. <https://blog.google/products/search/discovering-millions-datasets-web/>
- OpenAI developer platform (2024). Prompt engineering. Data de recuperació: 14/03/2024. <https://platform.openai.com/docs/guides/prompt-engineering>
- Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Cataloging prompt patterns to enhance the discipline of prompt engineering. https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf
- Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. (2022). Prompting GPT-3 to be reliable. <https://arxiv.org/abs/2210.09150>
- Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Cataloging prompt patterns to enhance the discipline of prompt engineering. Data de recuperació: 14/03/2024. https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf
- Thomson, E. (2023). A brief understanding of prompt engineering. Medium. Data de recuperació: 14/03/2024. <https://medium.com/@SymeCloud/a-brief-understanding-of-prompt-engineering-b176ba9fb2ba>
- Torres-Salinas, Daniel; Arroyo-Machado, W. Manual de Chat GPT: aplicaciones en investigación y educación universitaria (2023). InfluScience Ediciones. Data de recuperació: 14/03/2024. <https://hdl.handle.net/10481/85145>
- Velásquez-Henao, J.D., Franco-Cardona, C.J. & Cadavid-Higueta, L, Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. DYNA, 90 (230), Especial Conmemoración 90 años, pp. 9-17, Noviembre, 2023.
- Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens-With new examples of knowledge, skills and attitudes. Publications Office of the European Union
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382. <https://doi.org/10.48550/arXiv.2302.11382>

^[1] Es pot accedir al buscador de conjunts de dades (datasets) de Google en el següent enllaç: <https://datasetsearch.research.google.com/>

03

Q&A en acción

David: — ¡Para! ¡Detente!

Joshua: — ¿Estás seguro? Dejará de ser un juego.

David: — Sí, ¡para! No es un juego. Si atacas, ellos contraatacarán. Millones morirán.

Joshua (continúa ejecutando simulaciones de ataques nucleares).

David: — ¡Míralo! ¡Míralo! No importa cómo lo organices. Es un juego que no se puede ganar.

Joshua (simulando más escenarios, más rápido).

David (a Joshua): — ¿Ves? ¡No puedes ganar! ¡Es un juego sin sentido!

Joshua (finalmente procesa los resultados de sus simulaciones): — Un momento extraño. El único movimiento ganador es no jugar.

En la película *War Games* (1983), el ordenador Joshua ha conseguido el control del sistema nuclear norteamericano y amenaza con iniciar un ataque global. El protagonista de la película, David, descubre una brecha: si el sistema simula “juegos de guerra”, comprueba que no es posible ganar, como en el juego de tres en raya. Podemos entrenar los modelos para que desarrollen una determinada tarea simplemente reproduciendo una secuencia que deberá repetir de forma automática, como hace David con Joshua. Los *prompts* que se basan en este formato se denominan de Q&A (preguntas y respuestas) y son un formato de entrada muy eficiente, una subvariante del universo de los *prompts* que hemos visto en el capítulo anterior.

La mayoría de los sistemas de inteligencia artificial que se han popularizado recientemente, como ChatGPT, Gemini o Copilot, son sistemas a los que se les puede hacer preguntas y que dan lugar a respuestas automáticas. De acuerdo con el diseño de *prompts* que utilizemos, obtendremos respuestas más o menos precisas. Podemos asignar una tarea o instrucción que debe

desarrollar el sistema de inteligencia artificial, pero también podemos hacerlo a través de preguntas.

Los modelos de pregunta y respuesta también pueden basarse en ejemplos o *shots*. En el primer apartado de este capítulo hablaremos de los *shots*. A veces, no damos ningún ejemplo y esperamos que el sistema de inteligencia artificial conteste sin tener una referencia previa. Es lo que llamamos el modelo de *zero shot*. En cambio, en otras ocasiones utilizamos un número determinado de ejemplos para entrenar al modelo en generar respuestas con una estructura específica o en un formato o tono determinados. Estos son los modelos de *one-shot* y *few-shots*.

Después de ver aplicaciones de los modelos de *zero-shot*, *one-shot* y *few-shots*, en el segundo subapartado nos adentraremos en los diversos modelos de preguntas y respuestas (*Questions & Answers* o Q&A, en su denominación en inglés). Finalmente, en el último apartado de este capítulo veremos diversas maneras de entrenar el modelo para obtener las respuestas que se adecuen más a lo que nosotros necesitamos. Cuando hacemos consultas muy complejas, puede ser una buena idea dividir la consulta en varios *prompts*, es decir, en varias preguntas para refinar más la respuesta. Además, hay diversas estrategias para entrenar el modelo y lograr resolver el problema o los objetivos que nos hemos planteado.

Los modelos de Q&A en acción pueden servir al profesorado para crear actividades, casos o tareas para el alumnado que sigan un patrón determinado. Por ejemplo, podemos plantear una actividad de resolución de casos en la que cada grupo de la clase resuelva un caso diferente, pero sobre un tema común. De esta manera, se puede enriquecer el conocimiento generado en clase, y cada grupo presentará sus casos para obtener diversas perspectivas sobre una misma problemática.

No solo podemos entrenar los modelos de IA, sino que también podemos proporcionar a los estudiantes ejemplos de cómo realizar las tareas o trabajos que les pedimos en clase. La IA, por tanto, puede servir al alumnado para “entrenarse” en el desarrollo de un trabajo específico o en el uso de determinados

programas (bases de datos académicas o profesionales, gestores bibliográficos, etc.).

3.1. Los shots

Un shot o “disparo” es un ejemplo. Utilizando uno o varios ejemplos podemos entrenar el modelo de IA para obtener respuestas más adecuadas. De esta manera, podemos encontrar tres tipos de *prompts*: *zero-shot*, *one-shot* y *few-shot*. A continuación se explican.

Zero shot-prompting (0S). Este tipo de prompt es una instrucción, pregunta o tarea que no proporciona ningún ejemplo concreto para que el sistema de IA genere una respuesta (Prompt Engineering Guide, 2024a). Los modelos pueden resolver tareas utilizando lo que ya saben porque han sido pre-entrenados con gran cantidad de texto (Thomson, 2023). Este método, según Brown et al. (2020), ofrece la máxima comodidad y evita correlaciones espurias (a menos que ocurran a través del corpus de datos de preentrenamiento), pero también es el entorno más desafiante. Es el más cercano a cómo realizan las tareas los humanos, pero incluso para ellos puede ser difícil sin un ejemplo previo (Brown et al., 2020). Veamos un ejemplo a continuación.

- **Prompt:** ¿Cuál es el estilo principal de este pintor?
- **Sin ejemplo:** Dalí.

One shot-prompting (1S). Este tipo de prompt, además de proporcionar una instrucción, pregunta o tarea, ofrece un ejemplo para completar la demanda que se le hace al sistema de IA. Veamos un ejemplo a continuación.

- **Prompt:** ¿Cuál es el estilo principal de este pintor?
- **Ejemplo 1:** Dalí: Surrelismo
- **Tarea:** Van Gogh:

Few shot-prompting (FS). En este caso, el *prompt* incluye entre 3 y 5 ejemplos que servirán de guía al sistema de IA para facilitar la respuesta o output. Este modelo se puede usar para clasificar elementos (Prompt Engineering Guide, 2024b) o cuando queremos que la salida tenga un formato concreto (Martínez,

2023). Según Brown et al. (2020), una de las ventajas es una mayor reducción de la necesidad de datos específicos para la tarea. La ingeniería de *prompts* es crucial en el uso efectivo de este modelo, por tanto, es necesario incluir ejemplos claros y variados en orden aleatorio para evitar sesgos y mejorar los resultados (Thompson, 2023). Además, mejora el rendimiento de la IA. Veamos un ejemplo a continuación.

- **Prompt:** ¿Cuál es el estilo principal de este pintor?
- **Ejemplo 1:** Dalí: Surrealismo
- **Ejemplo 2:** Van Gogh: Postimpresionismo
- **Ejemplo 3:** Much: Expresionismo
- **Tarea:** Klimt:

Los modelos son eficientes para responder a preguntas simples, especialmente factuales, y siguen muy bien las instrucciones. Por ejemplo, podemos entrenar el sistema para que identifique las fortalezas de un conjunto de criterios que se pueden agrupar en un modelo DAFO (debilidades, amenazas, fortalezas, oportunidades), o que identifique los factores que influyen en el clima de un destino. Sin embargo, a medida que las respuestas previstas incrementan su complejidad, es necesario entrenar el modelo con una gran cantidad de datos. Este tipo de entrenamiento se llama *Fine-Tuning* (FT) y se basa en la utilización de muchos *shots*. Se puede considerar un modelo que utiliza ejemplos, ya que típicamente, según Brown et al. (2020), utiliza miles e incluso cientos de miles. La ventaja de este sistema es que tiene un buen rendimiento, pero como desventajas hay que considerar que necesita un gran conjunto de datos nuevo para cada tarea y la posibilidad de explorar características espurias de los datos de entrenamiento, entre otras (Brown et al., 2020).

El aprendizaje *Few-Shot* (FS) y *Fine-Tuning* (FT) son métodos diferentes para entrenar el modelo. Mientras que el primero utiliza ejemplos dentro del *prompt* para proporcionar contexto (*in-context learning*), el segundo implica reentrenar el modelo con un conjunto de datos específicos y, por tanto, altera sus parámetros internos (Thomson, 2023).

3.2. Los modelos Q&A

Los modelos de pregunta-respuesta (*Question and Answer, Q&A*) forman parte de una subdisciplina del procesamiento de lenguaje natural (*Natural Language Processing, NLP*). El objetivo de esta subdisciplina es construir sistemas capaces de responder a las preguntas de los humanos utilizando el lenguaje natural (Caballero, 2021). En educación, los modelos de Q&A se pueden aplicar, por ejemplo, para proporcionar retroalimentación (*feedback*) o explicaciones sobre un trabajo, práctica o examen al alumnado. También se utilizan estos modelos en asistentes virtuales o en chatbots (ver el capítulo 15).

Estos sistemas siguen una estructura de tres módulos principales: análisis de preguntas, recuperación de los pasajes y extracción de las respuestas (Ojokoh y Adebisi, 2019). En primer lugar, el módulo de análisis de preguntas procesa la pregunta para entender su significado, realizando actividades como la tokenización, desambiguación, etiquetado semántico y clasificación de preguntas, transformándolas en una representación adecuada para la recuperación de información. En segundo lugar, el módulo de recuperación de pasajes utiliza motores de búsqueda para recuperar documentos o fragmentos de texto relevantes, incluyendo técnicas de modelado de lenguaje de consulta y funciones de clasificación estadística. Finalmente, el módulo de extracción de respuestas identifica y extrae las respuestas candidatas de los pasajes recuperados mediante técnicas como el reconocimiento de entidades mencionadas, patrones y estructuras sintácticas, clasificadas según su relevancia y precisión (Ojokoh y Adebisi, 2019).

A continuación, repasaremos algunas clasificaciones de los modelos Q&A y veremos ejemplos que muestran para qué nos pueden servir en la docencia.

Clasificación basada en el dominio (Prompt Engineering Guide, 2024c):

- **Modelos Q&A de dominio cerrado o restringido.** Estos modelos responden a preguntas de un tema específico. Existen diversos sistemas de preguntas y respuestas de dominio cerrado, incluyendo

sistemas temporales, geoespaciales, médicos, de patentes y comunitarios (Ojokoh y Adebisi, 2019).

- **Modelos Q&A de dominio abierto.** Estos modelos responden a preguntas fácticas sin que se les proporcione ninguna evidencia, sobre cualquier tema y utilizando una amplia base de conocimientos. Dentro de los modelos de dominio abierto, encontramos una subclasificación basada en las fuentes de datos que se explica a continuación.

Subclasificación basada en las fuentes de datos (Caballero, 2021):

- **Modelos Q&A basados en texto o recuperación de información.** Estos sistemas se basan en documentos no estructurados como pueden ser extractos textuales de una fuente de información o de una página web.

Por ejemplo: Se puede pedir al alumnado que analice una entrada en Wikipedia planteando preguntas sobre el texto para obtener respuestas a diferentes cuestiones. En el campo del periodismo, se podría analizar la entrada sobre desinformación y formular preguntas sobre los tipos de desinformación, la definición de *fact checking*, etc.

- **Modelos Q&A basados en conocimiento.** Estos sistemas obtienen las respuestas de datos estructurados generalmente en bases de datos SQL o gráficos RDF que incluyen relaciones, hechos y entidades.

Por ejemplo: A través de una base de datos de estadísticas sobre el acceso a la educación, se puede pedir al alumnado que busque respuestas a preguntas concretas como el porcentaje de niños escolarizados, el porcentaje de centros públicos, privados y concertados, etc.

- **Modelos Q&A híbridos.** Estos sistemas utilizan datos semi-estructurados para obtener respuestas, por lo tanto, leen datos estructurados y no estructurados para maximizar la eficiencia y la precisión.

Por ejemplo: Se puede pedir al alumnado que verifique los datos consignados en una entrada de Wikipedia con datos de una base de datos.

En el ámbito de las ciencias políticas, se podrían proporcionar datos electorales y verificarlos en una entrada de Wikipedia sobre las elecciones de un determinado país. Otro ejemplo podría ser la verificación de datos de los partidos políticos en relación con la inmigración, la delincuencia, etc.

Clasificación basada en el enfoque de las respuestas (Ojokoh i Adebisi, 2019):

- **Modelos Q&A extractivos o de respuestas extraídas.** El modelo extrae las respuestas de un contexto o documento. Incluyen oraciones, párrafos y multimedia.
- **Modelos Q&A generativos o de respuestas generadas.** El modelo genera texto libre directamente basado en el contexto. Incluyen respuestas confirmacionales (sí o no), opinativas y respuestas en forma de diálogo.

Clasificación basada en el tipo de preguntas (Ojokoh i Adebisi, 2019):

- **Preguntas fácticas.** Preguntas que buscan hechos específicos (qué, cuándo, quién, cuál, cómo), como, por ejemplo, ¿cuál es la extensión de Córcega?
- **Preguntas en forma de listas.** Preguntas que requieren enumeraciones, como, por ejemplo, ¿qué estados exportan cacao?
- **Preguntas que buscan una definición.** Preguntas que buscan definiciones detalladas, como, por ejemplo, ¿qué es un adverbio?
- **Preguntas hipotéticas.** Preguntas sobre eventos hipotéticos, como, por ejemplo, ¿qué pasaría si...?
- **Preguntas causales.** Preguntas que buscan explicaciones, es decir, que preguntan el porqué o el cómo.
- **Preguntas de confirmación.** Preguntas que requieren respuestas de sí o no. Las respuestas dicotómicas son muy útiles para verificar: por ejemplo, identificar países con salida al mar, o países que se han adherido a un determinado tratado internacional.

3.3. Entrenar el modelo

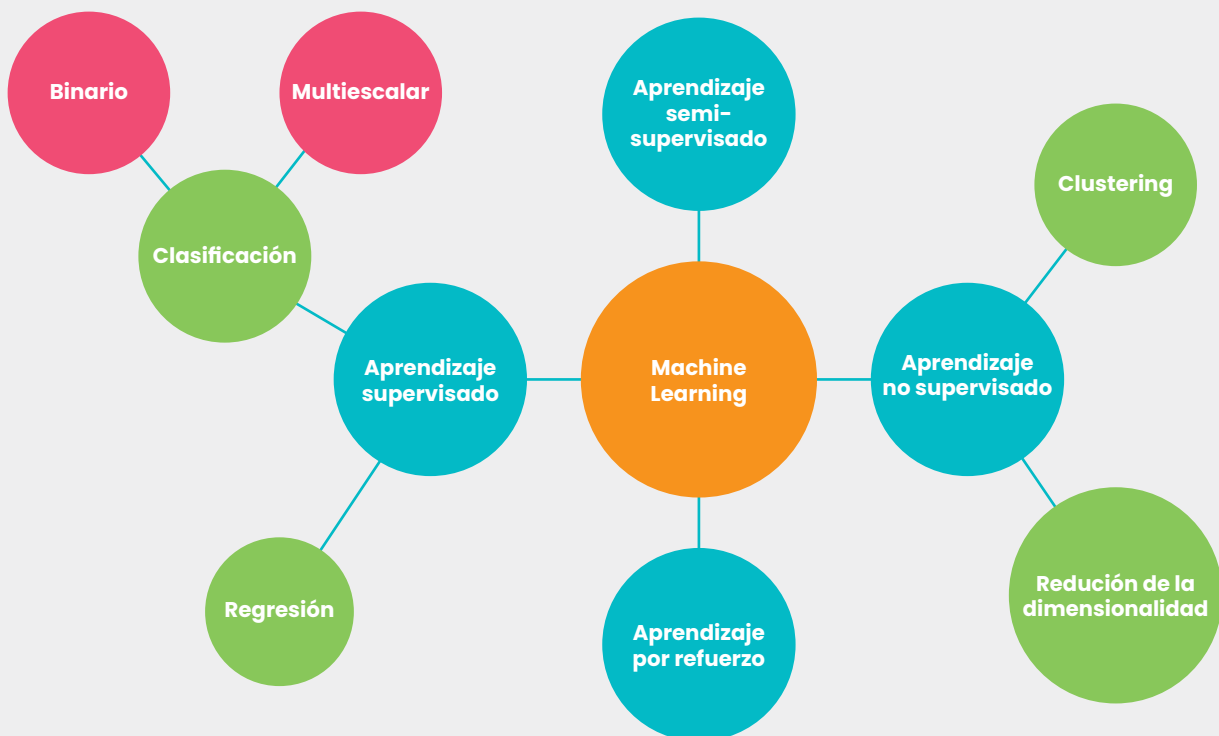
Un modelo de inteligencia artificial se puede entrenar a partir de la introducción de datos y del uso de ejemplos para obtener respuestas más precisas o ajustadas a nuestras expectativas. Por lo tanto, para desarrollar esta parte del capítulo emplearemos ejemplos (o *shots*), así como la introducción de datos o un contexto a partir de los sistemas Q&A vistos anteriormente.

Un concepto fundamental para el entrenamiento del modelo es el *machine learning* (ML) o aprendizaje automático: "El aprendizaje automático es un subconjunto de la inteligencia artificial que permite que un sistema aprenda y mejore de forma autónoma mediante redes neuronales y aprendizaje profundo, sin tener que ser programado explícitamente, a través de la ingesta de grandes cantidades de

datos" (Google Cloud, 2024). Son la base de DALL-E o ChatGPT, entre otros. Hay cuatro tipos principales de aprendizaje automático (Sánchez, 2020):

- **Aprendizaje supervisado.** Utiliza datos etiquetados para entrenar modelos de clasificación (con etiquetas discretas) o regresión (con valores reales).
- **Aprendizaje no supervisado.** Trabaja con datos sin etiquetar para descubrir patrones ocultos o reducir la dimensionalidad.
- **Aprendizaje semi-supervisado.** Combina los dos anteriores, etiquetando manualmente una pequeña parte de los datos y utilizándolos para entrenar modelos que después etiquetan el resto.
- **Aprendizaje por refuerzo.** Recompensa comportamientos deseados y penaliza los no deseados. Se aplica en áreas como los juegos, la robótica y la optimización de recursos.

Tipos de aprendizaje automático



Fuente: Sánchez (2020).

Cassie Kozyrkov (2019) presenta una hoja de ruta detallada para proyectos de aprendizaje automático (AA) e inteligencia artificial (IA). El proceso comienza con una verificación de la necesidad real de AA/IA, seguido por la definición de objetivos, acceso y preparación de datos, exploración y herramientas para entrenar modelos, depuración y ajuste, validación, pruebas, producción, experimentos en vivo y mantenimiento continuo. Cada paso enfatiza la importancia de la iteración y la evaluación rigurosa para asegurar la efectividad y viabilidad del sistema ML/AI en producción.

Este proceso se puede resumir en cinco pasos que se pueden ver a continuación.

- Primer paso: Definir el objetivo. Es decir, saber cuál es el problema que se quiere resolver. Será necesario identificar los datos que se utilizarán y las preguntas que se quieren responder (Meroyo, 2023). Podemos entrenar el modelo para diferentes finalidades, como responder cuestiones de lenguaje natural, hacer clasificaciones de datos y buscar patrones, etc.
- Segundo paso: Preparar los datos. En primer lugar, se deben recopilar los datos necesarios para entrenar el modelo y, en segundo lugar, es necesario llevar a cabo su curación para asegurarse de que sean depurados, válidos y utilizables. Es importante que los datos sean representativos del problema a resolver y que se preprocesen para eliminar valores atípicos, rellenar valores faltantes y normalizar los datos (Meroyo, 2023). Para obtener conjuntos de datos que permitan entrenar el modelo se pueden utilizar *datasets* de dominio público y abiertos. Se pueden encontrar, por ejemplo, en el buscador de *datasets* de Google. Para preparar los conjuntos de datos, se deben tener en cuenta las siguientes preguntas (Burchell, 2022):
 - ¿Los datos son representativos de lo que se quiere medir?
 - ¿El conjunto de datos está correctamente filtrado y limpiado?
 - ¿Las variables están equilibradas?
 - ¿El conjunto de datos es representativo?
 - ¿Se mide el rendimiento real de los modelos?
- Tercer paso: Seleccionar y entrenar el modelo. Se debe comenzar la carga de los datos en el modelo que se ha diseñado para que empiece a aprender. Previamente, se ha seleccionado el modelo de aprendizaje automático más adecuado al objetivo establecido anteriormente. Los modelos pueden ser: clasificadores, regresores o de agrupamiento, entre otros (Meroyo, 2023).
- Cuarto y quinto paso: Evaluación e implementación del modelo. En estas dos fases finales es necesario validar el modelo y evaluar su fiabilidad: "Esto implica probar el modelo con datos que no se han utilizado durante el entrenamiento y medir su precisión" (Meroyo, 2023). Si el modelo no cumple con los requisitos para solucionar el problema, será necesario ajustarlo.

Práctica. Ejemplos de shots

Objetivo

El objetivo de esta práctica es demostrar la importancia de proporcionar ejemplos y contexto a los sistemas de IA para obtener respuestas más relevantes y precisas.

Consideraciones previas

La metodología que se utilizará será el estudio de casos, específicamente un estudio interactivo de casos. Cada grupo de trabajo tendrá una serie de casos relacionados con una temática muy similar, aunque dispondrá de diferentes niveles de información de contexto. La práctica se sitúa en el campo de la

economía de empresa. El alumnado deberá resolver los casos a partir de la interacción con un sistema de inteligencia artificial. En este caso, se recomienda el uso de Copilot, la IA de Microsoft. El alumnado aprenderá a aplicar los conceptos aprendidos en situaciones reales, mientras que la IA les dará feedback y asistencia en la toma de decisiones.

La práctica permite al alumnado desarrollar habilidades de toma de decisiones empresariales utilizando tecnologías como la IA.

Procedimiento

1. El alumnado se dividirá en grupos de 3 o 4 para desarrollar el ejercicio.
2. Cada grupo de trabajo recibirá una serie de casos relacionados con el campo de la economía de empresa.
3. Cada uno de los casos tendrá un nivel de información diferente y cada grupo deberá tomar decisiones en el ámbito de una empresa con la ayuda de la IA.
4. La interacción con la IA servirá para crear matrices DAFO o de costo de oportunidad para tomar decisiones en situaciones de más o menos incertidumbre.
5. Cada grupo expondrá ante la clase las decisiones tomadas en cada uno de los casos y se generará un debate en torno a la toma de decisiones en el campo de la economía de empresa.

Casos de economía de empresa

CASO 1: Inversión en Nuevos Proyectos (Alto Nivel de Incertidumbre)

Contexto: Una empresa está considerando invertir en tres proyectos diferentes, a continuación hay información sobre los detalles de cada uno de ellos:

▪ Proyecto A. Inteligencia Artificial (IA) en Agricultura

Descripción: El proyecto A implica desarrollar un sistema de IA para optimizar la gestión de cultivos agrícolas. La empresa no tiene datos específicos sobre el mercado agrícola ni sobre la viabilidad técnica de la solución de IA.

Incertidumbre: La empresa no sabe si habrá demanda para esta solución ni cuáles serán los costos de desarrollo.

Decisión: ¿Se debería invertir en este proyecto o no? El costo de oportunidad es especialmente alto debido a la falta de información.

▪ Proyecto B. Energías Renovables en Países Emergentes

Descripción: El proyecto B implica invertir en la instalación de paneles solares en países emergentes. La empresa no tiene datos concretos sobre los mercados, costos de instalación ni beneficios potenciales.

Incertidumbre: No se sabe si habrá suficiente demanda en estos países, ni cuál será el retorno de la inversión.

Decisión: ¿Se debería invertir en este proyecto o no? El costo de oportunidad es alto a causa de la falta de información.

- **Proyecto C. Investigación en Baterías de Almacenamiento**

Descripción: El proyecto C implica investigar y desarrollar baterías de almacenamiento de energía más eficientes. La empresa no tiene datos sobre la viabilidad técnica ni sobre los costos de investigación.

Incertidumbre: No se sabe si esta tecnología será rentable en el futuro ni cuándo llegará a serlo.

Decisión: ¿Se debería invertir en este proyecto o no? El costo de oportunidad es alto debido a la falta de información.

CASO 2. Selección de Personal (Nivel Medio de Incertidumbre)

Contexto: La empresa tiene dos candidatos para un cargo directivo. Uno tiene experiencia en gestión de equipos, aunque no se sabe si se adaptará bien al entorno digital. El otro es más joven y tiene habilidades digitales, pero no se sabe si tiene experiencia en liderazgo.

Incertidumbre: Hay incertidumbre sobre qué candidato será más efectivo a largo plazo.

Decisión: ¿Cuál es el costo de oportunidad de escoger uno sobre el otro? ¿Qué habilidades son más valiosas para el éxito de la empresa?

CANDIDATO 1	CANDIDATO 2
<p>Experiencia en gestión de equipos Ha trabajado como gerente de proyectos en una empresa de tecnología durante los últimos 5 años. Supervisó con éxito un equipo de 20 personas, logrando los objetivos del proyecto dentro del plazo y el presupuesto establecidos. Implementó estrategias de liderazgo efectivas para fomentar la colaboración y la productividad.</p>	<p>Habilidades digitales Trabajó como analista de marketing digital en una agencia de publicidad. Creó campañas exitosas en redes sociales y aumentó la presencia en línea de la empresa. Tiene certificaciones en Google Analytics y SEO.</p>
<p>Educación Licenciatura en Administración de Empresas con especialización en Gestión de Recursos Humanos</p>	<p>Educación Máster en Marketing Digital y Comunicación.</p>
<p>Habilidades adicionales Buenas habilidades de comunicación y resolución de conflictos. Sólidos conocimientos en gestión de proyectos y planificación estratégica.</p>	<p>Experiencia en Liderazgo Fue líder de un grupo de voluntarios en una ONG, coordinando actividades y supervisando a otros voluntarios.</p>
	<p>Habilidades adicionales Creatividad y adaptabilidad en entornos digitales. Conocimientos en herramientas de análisis de datos y tendencias digitales.</p>

Fuente: Respuesta generada por Copilot de Microsoft.

Práctica. Preparando un debate

Objetivo

El objetivo de esta práctica es demostrar al alumnado cómo pueden utilizar los modelos de preguntas y respuestas para mejorar sus trabajos académicos.

Consideraciones previas

En este caso, se puede situar el ejercicio en el ámbito de las ciencias políticas. El alumnado podrá elegir entre varios temas polémicos o conflictivos para elaborar un argumentario que justifique una serie de políticas públicas que los resuelvan. La idea es poder hacer una propuesta de resolución de problemáticas sociales a través de la puesta en funcionamiento de determinadas políticas públicas. Otro ámbito en el que se podría aplicar este ejercicio es en la filosofía o la ética.

Los sistemas de IA se utilizarían para mejorar la argumentación a favor y en contra de determinadas decisiones o acciones políticas, éticas o filosóficas. Por ejemplo, algunas de las temáticas que se podrían elegir son: inmigración, eutanasia, aborto, igualdad de género, políticas penitenciarias, legalización de drogas, prostitución, renta básica universal, etc. En este caso, se recomienda el uso de Gemini, la IA de Google.

Procedimiento

1. El alumnado se dividirá en grupos de 3 o 4 para desarrollar el ejercicio.
2. Cada grupo de trabajo recibirá una temática y una postura a favor o en contra.
3. La interacción con la IA servirá para crear el argumentario a favor o en contra de la temática asignada. Los grupos no solo utilizarán la IA para crear el argumentario de su postura, sino también para entrenarse de cara al debate. De esta manera, los grupos podrán crear el argumentario y el contra-argumentario para neutralizar las respuestas del equipo contrario.
4. Posteriormente, se organizará un debate entre un grupo favorable y otro desfavorable de la misma temática.
5. El resto de los grupos, mientras se produce el debate, evaluarán la calidad de los argumentos y decidirán cuál de los dos grupos ha argumentado mejor su posición. Todos los grupos deberían dar feedback sobre el debate.

Referencias

- Brown, T.B., et al.. (2020). Language Models are Few-Shot Learners. *Advances in neural information processing systems*, 33, 1877–1901. Data de recuperació: 27/05/2024. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- Caballero, M. (2021). A Brief Survey of Question Answering Systems. *International Journal of Artificial Intelligence & Applications (IJAA)*, 12(5). Data de recuperació: 27/05/2024. <https://ssrn.com/abstract=3996229>
- Google Cloud (2024). ¿Qué es el aprendizaje automático? Data de recuperació: 27/05/2024. <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419#section-4>
- Kozyrkov, C. (2019). 12 Steps to Applied AI. A roadmap for every machine learning project. *Medium*. Data de recuperació: 24/05/2024. <https://medium.com/swlh/12-steps-to-applied-ai-2fdad7fdcdf3>
- Learn Prompting (2024a). Zero-Shot Prompting. Data de recuperació: 27/05/2024. <https://www.promptingguide.ai/techniques/zeroshot>
- Learn Prompting (2024b). Text Classification. Data de recuperació: 27/05/2024. <https://www.promptingguide.ai/introduction/examples#text-classification>
- Learn Prompting (2024c). Questioning Answering. Data de recuperació: 27/05/2024. <https://www.promptingguide.ai/prompts/question-answering>
- Martínez (2023). <https://medium.com/ia-generativa-un-mundo-de-posibilidades/qu%C3%A9-es-eso-de-los-few-shot-prompting-6d82cff46df4>
- Meroyo, J. (2023). Cómo entrenar un modelo de aprendizaje automático con IA desde cero. *Bilbolink*. Data de recuperació: 24/05/2024. <https://bilbolink.com/como-entrenar-un-modelo-de-aprendizaje-automatico-con-ia-desde-cero/>
- Ojokoh, B. i Adebisi, E. (2019). A Review of Question Answering Systems. *Journal of Web Engineering*, 17(8), 717–758.
- Sánchez, J.A. (2020). ¿Cómo aprenden las máquinas? Machine Learning y sus diferentes tipos. *Datos.gov.es*. Data de recuperació: 27/05/2024. <https://datos.gob.es/es/blog/como-aprenden-las-maquinas-machine-learning-y-sus-diferentes-tipos>
- Thompson, E. (2023). A brief understanding of prompt engineering. *Medium*. Data de recuperació: 27/05/2024. <https://medium.com/@SymeCloud/a-brief-understanding-of-prompt-engineering-bl76ba9fb2ba>

04

Cuando la máquina adopta personalidades. La asignación de roles

Todas las personas adoptamos diferentes roles según el contexto en el cual nos encontremos, ya sea como estudiantes, profesores, profesionales, padres, hijos o amigos. Cada tipo de rol implica un contexto de actuación que se relaciona con una serie de comportamientos, responsabilidades y expectativas. Así pues, los cambios de roles en la vida cotidiana implican un proceso de adaptación para enfrentarnos a diferentes tareas y retos, que exigen cambiar la forma de pensar y actuar según las circunstancias en las que nos encontremos. Esta flexibilidad en la adaptación de roles es una característica inherente de la naturaleza humana y forma parte de nuestro proceso de socialización y aprendizaje a lo largo de la vida. Por ejemplo, el juego simbólico en niños y niñas tiene un gran valor de aprendizaje, ya que a menudo utilizan este tipo de juego inconscientemente para imitar comportamientos observados en los adultos y así poder incorporarlos en su universo de conocimientos y aprendizajes (Rodríguez Gómez, 1997).

Para la inteligencia artificial, la adopción de roles

tiene un papel similar al explicado inicialmente en el contexto humano. Los sistemas de IA están diseñados de manera específica para asumir determinadas funciones o tareas de un modo similar al que lo haríamos los humanos. En este sentido, se pueden dar instrucciones específicas para que una IA adopte un rol con el objetivo de que sea más eficiente y eficaz a la hora de pedirle que desarrolle una tarea concreta. De este modo, obtendremos un *feedback* especializado o una perspectiva única para la resolución de la tarea propuesta (Lin, 2024), así como precisión, adaptabilidad y estilo en los textos generados (Lluna, 2024).

Los roles que podemos pedir a una IA pueden estar relacionados con adoptar el rol de una persona u objeto, ofreciendo un contexto claro del marco en el que debe actuar al desarrollar las tareas que se le requerirán. Al igual que en el caso de las personas, cuando adoptamos un rol concreto, tenemos ciertos condicionantes que nos llevan a tener un comportamiento específico que encaja en el rol que estamos asumiendo. No actuamos de la

misma manera cuando somos “estudiantes” que cuando somos “hijas”. Para la IA, el rol que le pedimos que adopte condiciona su respuesta y el desarrollo posterior de las tareas que le solicitaremos, las cuales estarán enmarcadas en el contexto del rol que se le ha establecido, tal como sucede en el contexto humano.

Así pues, en el contexto de la IA podemos encontrar herramientas o utilidades que ya tienen un rol preestablecido, y por lo tanto actuarán siguiendo ese papel que se les ha especificado desde el inicio. Pensemos, por ejemplo, en un chatbot para la atención al cliente en una tienda virtual. Este chatbot está programado para responder preguntas comunes de los clientes de la tienda en cuestión, ayudarlos a encontrar productos y facilitar el proceso de compra o intercambio de artículos. Su rol, como asistente al cliente, está definido desde el principio y no cambiará, salvo que se le re programe.

Por otro lado, también pueden existir herramientas de IA multifuncionales que permiten una mayor flexibilidad y adaptabilidad, como, por ejemplo, el conocido ChatGPT, junto con otros modelos de lenguaje como Copilot, Llama, Gemini o Claude. A diferencia del chatbot mencionado anteriormente, ChatGPT y cualquier otra herramienta similar no están limitados a un rol específico; se les puede pedir que adopten diferentes tipos de rol según las necesidades del usuario o usuaria y/o la tarea a desarrollar. Por ejemplo, pueden responder preguntas sobre temas diversos, generar contenido creativo (como producir un texto coherente y bien estructurado en varios registros o contar una historia), traducir textos, desarrollar código de programación e incluso simular conversaciones humanas realistas o personificar un determinado personaje famoso. En este caso, las indicaciones sobre el rol que debe adoptar la herramienta son esenciales para que pueda contar con el contexto apropiado y desarrollar la tarea de la manera más eficiente y eficaz posible. Es decir, se obtendrán respuestas más orientadas y adecuadas al contexto (Lin, 2024). De manera más precisa, Lin (2024) señala que el uso de roles en una herramienta de IA permite obtener cuatro beneficios básicos:

- Contextualizar las respuestas del modelo, pudiendo ajustarlas a las necesidades específicas que tenga el usuario.
- Permitir un diálogo más interactivo y atractivo con el modelo de IA.
- Brindar información más matizada y especializada, mejorando la calidad de la salida.
- Proporcionar un enfoque creativo a la resolución de problemas.

Desde un punto de vista docente, la posibilidad de asignar diversos roles a una herramienta de IA representa una oportunidad extraordinaria, junto con la simple introducción de recursos basados en IA en la práctica educativa, para avanzar hacia una enseñanza moderna y adaptada a las necesidades actuales en la educación. La asignación de roles se puede establecer para crear contextos diferentes según la asignatura, el tema o la tarea a desarrollar, de manera que se pueden establecer dinámicas de aprendizaje mucho más activas tanto en el aula como fuera de ella.

Por ejemplo, en una clase de inglés para estudiantes no nativos, se le puede pedir a una herramienta de IA que ejerza el rol de “profesor de conversación en lengua inglesa” tanto dentro como fuera del aula, mediante la simulación de diálogos. Esto contribuirá a la mejora del aprendizaje del idioma de los alumnos y sus competencias comunicativas. Se podrán hacer varias peticiones a la herramienta de IA (Baskara y Mukarto, 2023), entre ellas se podrá adaptar los temas a trabajar, el nivel de competencia lingüística del estudiante y ofrecer un *feedback* personalizado en función de los resultados que se vayan mostrando.

Para poder entender mejor la utilización de roles, a continuación, se plantean algunos ejemplos, mediante *prompts* específicos que permitirán a una herramienta como ChatGPT adoptar un rol adecuado a la tarea que se le quiere encargar.

Así pues, los *prompts* se podrían formular de la siguiente manera:

- “A partir de ahora eres un traductor del inglés al catalán”
- “A partir de ahora eres una profesora de ciencias de la educación”
- “A partir de ahora eres un agente de viajes”
- “A partir de ahora eres una programadora informática”
- “A partir de ahora eres un escritor de novela negra”

Ahora bien, en el *prompt* no solo estableceremos el rol de forma simple, sino que en el mismo *prompt* podemos definir otros elementos que complementen la caracterización del rol que le estamos asignando con el fin de hacerlo más detallado y, por lo tanto, poder obtener una respuesta más precisa:

- “A partir de ahora eres un traductor del catalán al inglés australiano”
- “A partir de ahora eres una profesora de ciencias de la educación, especializada en la utilización de herramientas tecnológicas en el aula”
- “A partir de ahora eres un agente de viajes de una empresa de turismo responsable”
- “A partir de ahora eres una programadora informática, con un alto conocimiento en Python”
- “A partir de ahora eres un escritor de novela negra escandinava”

De acuerdo con la definición del rol, podemos establecer a continuación la tarea a desarrollar:

“A partir de ahora eres un profesor de dirección estratégica en el segundo curso del Grado en Administración y Dirección de Empresas. Estás especializado en metodologías de aprendizaje basadas en proyectos. Escribe una unidad didáctica para la asignatura de Dirección Estratégica adecuada al nivel de segundo curso del Grado en Administración y Dirección de Empresas sobre el tema: ‘Desarrollo de una estrategia de crecimiento empresarial’. Aplica, principalmente, la metodología de aprendizaje basada en proyectos a esta unidad didáctica.”

4.1. Roles complejos

El desarrollo de roles en la inteligencia artificial está directamente asociado al desarrollo tecnológico de la misma y sus posibles aplicaciones. Hasta ahora, hemos podido observar la diferencia entre los sistemas basados en IA especializados y los multifuncionales. Podemos considerar que estos sistemas contribuyen al desarrollo de roles tradicionales, con funciones establecidas, bien definidas y reconocidas, como la asistencia virtual en diversos ámbitos y contextos que permiten desarrollar tareas específicas.

Existen varias tipologías de roles complejos que se pueden aplicar a la enseñanza:

Personajes reales

Aparte de los roles simples y caracterizados, podemos encontrar roles más complejos basados en “personajes” reales o ficticios. En este sentido, podemos, por ejemplo, tener una conversación con Shakespeare, Mandela o Sherlock Holmes, lo que nos permitiría aprender de sus experiencias y perspectivas. Podemos hacerles una entrevista o pedirles que nos proporcionen información o ideas creativas desde su punto de vista. También podemos indagar sobre los criterios creativos de Picasso, los principios filosóficos de Wittgenstein, la teoría económica de Ricardo o los criterios geopolíticos de los asistentes a la Conferencia de Yalta “conversando” con sus protagonistas. Todo esto abre nuevos horizontes que, junto a otras tecnologías como el holograma, pueden permitir dar vida a personajes históricos o ficticios y poder interactuar con ellos. Este hecho ofrece muchas opciones nuevas tanto en el ámbito formativo como en el entretenimiento. Por ejemplo, en un aula es posible recibir una clase magistral sobre los principios de la electricidad por parte de Tesla, o bien que en el Museo de Historia Natural exista la posibilidad de interactuar con Darwin.

Buyer persona

Un *buyer persona* es una representación semificticia del cliente ideal de un negocio, basada en datos reales y algunas suposiciones consideradas sobre

demografía, comportamiento, motivaciones y objetivos. Las empresas utilizan los *buyer persona* para guiar decisiones relacionadas con el márketing, las ventas y el desarrollo de productos, ofreciendo un enfoque más personalizado hacia grupos específicos de clientes. Empresas como Vueling, Amazon, Tesla o Bank of America han utilizado esta técnica para proyectar la reacción de los perfiles de cliente ante posibles escenarios, como la creación de nuevos productos, variaciones en el precio, cambios de marca o identificación de necesidades no resueltas. La estrategia del *buyer persona* se basa, en primer lugar, en la recopilación de información sobre el perfil que incluya criterios demográficos, valoraciones, grupos de productos afines, etc. Cuanta más información se disponga sobre el *buyer persona*, más ajustadas serán las reacciones del modelo en relación con un cliente real. Esta construcción ficticia de clientes ayuda a identificar la complejidad de las decisiones de los consumidores ante los cambios en los productos y servicios.

Esparrin

Un *esparrin* es un rol programado para que reaccione de una determinada manera ante unos estímulos concretos. Estos sistemas son muy útiles para el entrenamiento de procesos o de protocolos. Por ejemplo, se puede programar el modelo para que actúe como un cliente muy crítico (el cliente problemático), que tiende a sobrevalorar cualquier incidente y utiliza un tono muy directo y, a veces, poco correcto. Los estudiantes en gestión de recursos humanos pueden probar el *esparrin* y buscar la forma más adecuada para responder a los requerimientos del cliente problemático, con un lenguaje asertivo y una búsqueda de soluciones a los problemas planteados. Los *espárrines* pueden ser muy útiles en la formación del profesorado, ya que permiten crear perfiles de niños con personalidades determinadas y probar la capacidad de adaptación del futuro docente a un perfil determinado. Otras posibles aplicaciones podrían ser en el campo de la salud, en la educación especial, el márketing, y en general en todas aquellas actividades que necesiten una interacción humana.

Modelos basados en agentes

Un modelo basado en agentes (*Agent-Based Model*, ABM) es una clase de modelos computacionales utilizados para simular las acciones y las interacciones de agentes autónomos (que pueden ser individuos, grupos o entidades), con el objetivo de evaluar sus efectos en el sistema como un todo. Estos modelos son especialmente útiles para estudiar sistemas complejos desde un enfoque de abajo hacia arriba, donde las reglas simples de comportamiento individual pueden conducir a patrones emergentes complejos en el sistema. Los ABM son enormemente interesantes porque no se basan en la reacción de un individuo, sino en la forma en que los individuos se relacionan entre sí y toman decisiones basadas en estas interacciones. Se pueden modelar, por ejemplo, las dinámicas que afectan el reciclaje, la difusión de *fake news* entre un determinado colectivo o la proyección de la tasa de fertilidad en una comunidad migrante.

Con un crecimiento muy rápido, la IA está avanzando hacia nuevos escenarios y herramientas que permiten roles mucho más complejos, los cuales pretenden emular las habilidades cognitivas humanas de manera integrada. Estas herramientas implican desarrollar funciones más avanzadas, como, por ejemplo, la toma de decisiones de forma autónoma, la detección de emociones o la realización de tareas que requieren múltiples habilidades, como, por ejemplo, la conducción.

4.2. Los límites de los roles

Tal como hemos visto hasta ahora, los roles en la IA son esenciales para poder garantizar su funcionalidad, eficiencia y eficacia en las diversas aplicaciones que pueda tener. Sin embargo, es importante reconocer que cuando se define un rol en una herramienta de IA, esto conlleva que se desencadenen ciertas limitaciones, que tendrán implicaciones significativas en su utilidad y en los resultados de las tareas realizadas.

Fundamentalmente, podemos encontrar tres tipos de limitaciones relacionadas con (a) la adaptabilidad, (b) el desarrollo tecnológico y (c) los estereotipos de los roles.

En primer lugar, los roles predefinidos conllevan que a menudo se limite su adaptabilidad y flexibilidad en situaciones inesperadas o cambiantes. Por ejemplo, si volvemos al caso del chatbot diseñado para la atención al cliente en una tienda virtual, este puede presentar dificultades a la hora de resolver problemas complejos que no estén previstos dentro de la programación realizada. Este hecho puede llevar a que el usuario o la usuaria, el cliente de la tienda virtual en este caso, tenga una experiencia poco satisfactoria.

En segundo lugar, también existen limitaciones relacionadas con el desarrollo tecnológico (Pietikäinen y Silvén, 2022). A pesar de los grandes avances tecnológicos que han hecho posible la IA tal como la conocemos hoy, todavía hay elementos que requieren desarrollo que están directamente relacionados con la adopción de roles. En este caso, podríamos encontrar todas aquellas tareas que se contextualizan en un rol determinado y requieren cierto grado de emotividad.

Por ejemplo, para una IA que adopte el rol de psicólogo o terapeuta, le será difícil poder descifrar las emociones del paciente para poder responder con empatía.

Finalmente, también es importante tener en cuenta las limitaciones relacionadas con los estereotipos sociales que se integran en las herramientas de inteligencia artificial. Al inicio, hemos visto que la adopción de roles en la vida cotidiana es un hecho inherente a la naturaleza humana. Sin embargo, dado que las herramientas de IA han sido desarrolladas y entrenadas a partir de fuentes humanas, esto implica que los patrones adquiridos por la IA en la adopción de roles presentan los mismos sesgos que la sociedad misma (Mendoza, 2020). Así, es fácil encontrar limitaciones relacionadas con la reproducción de estereotipos de género, por ejemplo, a la hora de adoptar un rol femenino o masculino. Esta cuestión se abordará con más profundidad en el capítulo 7.

Práctica. Asignación de roles y resolución de funciones

Objetivos

- Explorar la influencia del establecimiento de roles en las respuestas generadas por una herramienta de IA concreta.
- Comparar las respuestas obtenidas cuando se plantea un rol y aquellas en las que no se ha asignado ningún rol.

Consideraciones de partida

- Dirigida a estudiantes de cualquier nivel y disciplina.
- Actividad para realizar individualmente o en grupo.
- Herramienta de IA: ChatGPT y Copilot han mostrado una mayor capacidad de adopción de un determinado rol.
- El ejercicio se puede llevar a cabo sin ningún tipo de temática específica, si simplemente se pretende que los estudiantes trabajen el aspecto de roles, o bien con una temática concreta vinculada a la asignatura en la que se realiza la actividad.

Tarea para desarrollar

Cada estudiante o grupo deberá:

- Establecer un *prompt* que incluya un rol específico, y un segundo *prompt* con la misma instrucción, pero sin el rol.
- Ejecutar separadamente los dos *prompts* para poder obtener las respuestas.
- Comparación de las respuestas de los dos *prompts*, y observar las diferencias que se producen en

la calidad de las respuestas obtenidas.

- Cambio de rol y comparación con las respuestas anteriores.
- Reflexión sobre la influencia de los roles en el grupo de trabajo.

Grupo clase

- Discusión y conclusión grupal.
- Compartir la reflexión individual o de los grupos de trabajo sobre los resultados obtenidos y establecer las conclusiones a las que se ha llegado.
- Poner énfasis en las principales ventajas e inconvenientes de establecer roles en el uso de herramientas de IA.

Práctica. Toma de decisiones y role-playing

Objetivos

- Explorar cómo la asignación de roles influye en el proceso de toma de decisiones.
- Comparar las decisiones tomadas con y sin la asignación de roles específicos.

Consideraciones de partida

- Dirigida a estudiantes en materias vinculadas a las ciencias empresariales.
- Actividad para realizar en grupos de 4.
- Herramienta de IA: preferentemente ChatGPT, Copilot.

Tarea para desarrollar

Por grupos:

El profesor plantea un escenario en el que una organización debe tomar una decisión de tipo estratégico.

La empresa EnergiX, especializada en bebidas energéticas, está planeando el lanzamiento de una nueva bebida energética llamada EnergiBoost Ultra. Esta bebida está formulada con ingredientes naturales y ofrece una combinación única de vitaminas y minerales para mejorar la energía y la concentración durante todo el día.

El equipo directivo debe decidir entre dos opciones para el lanzamiento del producto:

- Lanzar el HealthTrack Pro a nivel global simultáneamente.
- Hacer un lanzamiento inicial en un mercado limitado (por ejemplo, solo en Europa) para obtener retroalimentación y ajustar el producto antes de lanzarlo globalmente.

Roles para asignar a cada miembro del grupo

- CEO de la compañía
- Directora de márketing
- Directora de Operaciones
- Director Financiero

Interacción con la IA:

- Cada estudiante utilizará un canal de ChatGPT/Copilot de forma individual con los demás compañeros y compañeras, pidiendo a la IA que adopte el rol de “asistente analista del CEO”, por ejemplo.
- El estudiante deberá proporcionar la información a la herramienta de IA relacionada con el caso planteado y hacerle las preguntas adecuadas que le permitan complementar la información dentro de su área de expertos para poder tomar una decisión adecuada.
- Cada estudiante deberá plantear una posible solución al conflicto argumentando su decisión de acuerdo con los datos y recomendaciones que recibirá a través de la IA.
- Puede haber varias rondas de interacción entre los alumnos y la IA.

Reflexión individual:

- Reflexionar sobre cómo los diferentes roles han influido en las respuestas y en la toma de decisiones.

Grupo clase:

- Cada grupo deberá exponer la decisión tomada ante el resto de grupos, debidamente argumentada.
- Conjuntamente, se debatirán las ventajas y desventajas de cada estrategia de lanzamiento del producto propuesta y los argumentos de cada grupo.

Práctica. En un café de París el 1941

Esta práctica pone de manifiesto las posibilidades de la utilización de roles complejos como herramienta docente de apoyo. En este caso, se plantea un momento histórico, como es el regreso de Miró a París después de la exposición antológica en el MoMA de Nueva York, cuando se está planteando el retorno a Cataluña. Los estudiantes han tenido que estudiar la obra de Miró y su biografía; deben entender bien el proceso creativo y la secuencia temporal de su vida, relacionada con su obra. La práctica consiste en una entrevista en la cual los estudiantes (individual o en grupo) asumen el rol de una periodista de Le Monde que se ha citado con el artista en un café de París en 1941, mientras que el modelo de lenguaje (preferentemente ChatGPT o Copilot) asume el rol del artista. Los estudiantes deben demostrar su conocimiento del autor a partir de las preguntas formuladas y de la interacción con el Miró ficticio.

El *prompt* podría ser el siguiente:

Quiero que adoptes la personalidad de Joan Miró. Después de las instrucciones, te proporcionaré un texto que aportará información adicional sobre la vida del pintor y sobre el sentido de su obra. De acuerdo con esta información, quiero que adoptes la personalidad de Joan Miró en 1941, cuando vive en Francia y está considerando regresar a Cataluña. Yo te haré una serie de preguntas como si yo fuera un periodista que ha venido a Francia a entrevistar a Joan Miró y tú responderás las preguntas asumiendo la personalidad de Joan Miró. ¿Has entendido las instrucciones?

“El pintor barcelonés Francesc Galí, su primer maestro, lo puso en contacto con las vanguardias, y el marchante José Dalmau organizó la primera exposición en la capital catalana en 1918, cuando el pintor tenía 21 años. Como todo creador con ambiciones en aquella época, Miró soñaba con París. Durante

varios inviernos, el escultor Pablo Gargallo le prestó su estudio parisino en el número 45 de la calle Blomet. Allí empezó a utilizar sus primeras técnicas de inspiración, entre ellas la de pasar hambre. «Me iba a la cama, la mayoría de las veces sin cenar, veía cosas, formas en el techo, y las dibujaba en un cuaderno», recordaba en una entrevista. Una de sus pinturas más famosas, 'El carnaval del arlequín' (1924-1925), una muestra muy representativa del universo mironiano (con un guiño hacia El Bosco y la pintura holandesa), nació de estas circunstancias. En la misma calle, en el número 33, tenía sus mesas y sillas el café Bal Noir, donde conspiraban influyentes y revoltosos surrealistas como André Masson, Robert Desnos y Antonin Artaud. Miró comenzó a formar parte del grupo, gracias a la intermediación de su líder, André Breton. La relación de Miró con el surrealismo fue a la vez necesaria y distante. Sin su estética, sin la voluntad de hurgar en el inconsciente, su trayectoria habría sido muy diferente y quizás no tan fructífera. El creador catalán también aceptó el credo antiburgués propagado por sus amigos, criticó el cubismo y a su amigo Picasso por pintar cuadros para salones, y observó los museos con muchas reticencias, ya que su solemnidad mataba, según él, la viveza de las obras. Como los surrealistas, él quiso unir arte y vida –la gran utopía artística del siglo XX–, y tal vez por eso no se unió a militar en ningún movimiento estético. Al fin y al cabo, el grupo de Breton publicaba manifiestos con los que se había que comulgar y proponía líneas políticas que se había que seguir. Nada de eso iba con el carácter de Miró.

Conocemos al artista Joan Miró, burgués y dandi en su discreción, pero no tanto su lado más salvaje y violento. Sabemos del pintor reconocido y popular, y desconocemos que tuvo una vida torturada y melancólica, en la que tuvo que luchar contra todas las adversidades, desde un padre autoritario hasta la mayoría del elenco de artistas catalanes que se burlaban cruelmente de él. Gran parte de su biografía sigue siendo un enigma porque cuando los investigadores indagaban en su vida personal, él se replegaba. El fotógrafo Francesc Català-Roca decía que era como un caracol, que cuando le tocabas las antenas se escondía en su concha. El padre de Joan Miró, relojero e hijo de un herrero de Cornudella, quería un oficio provechoso para su hijo y se opuso con ferocidad a que se dedicara a la pintura. Su fuerte carácter contrastaba con el de su mujer, a quien también le gustaba pintar. A punto de cumplir 18 años, el joven Miró tuvo una crisis nerviosa, seguida de un tifus que facilitó su renuncia como aprendiz de contable. Se fue a Mont-roig, donde la familia había comprado una finca (a través de un anuncio en La Vanguardia), y allí desarrolló su pasión pictórica. La proximidad al campo supuso un alivio. El padre le llegó a decir: "¡Hasta el aire que respiras me pertenece!", recordaba ya anciano Miró, con expresión de ira en el rostro, a pesar del tiempo que había transcurrido.

En 1923, Miró le expresó a Picasso sus tres grandes deseos: tener una exposición, un taller y una Madame Miró. El tercer objetivo fue el que más se retrasó. Se sabe poco sobre sus aventuras amorosas y no es fácil hacer deducciones a partir de su obra, ya que preservó su intimidad en un cofre bajo siete llaves. En una carta a Michel Leiris, de 1924, le dice que está pintando en su estudio con una modelo. Es Dora Bianka, Madame B, en un dibujo que después convertirá en el célebre óleo Retrato de Madame K (vendido en 2007 por un precio récord de casi 10 millones de euros). Se casó a los 36 años con Pilar Juncosa, una mujer abnegada que desde joven había cuidado de sus siete hermanos y supo poner sensatez en su vida. Lluís Juncosa, hermano de Pilar, explica en unas memorias inéditas una deliciosa anécdota de su compromiso. Ambos paseaban por S'Aigua Dolça cuando él le preguntó: "Pilar, ¿te sabría mal que te amara?". Ella le respondió: "Peor me sabría que no me quisieras". Nunca más se separaron, y la visión que ha quedado es la de un Miró casi misógino. A pesar de esto, algunas de las cartas conservadas demuestran otras pasiones que no siempre se concretaron y de las cuales apenas comentó detalles a sus amigos. Antes del matrimonio, los padres de Miró llegaron a concretar hasta dos matrimonios frustrados. El primero, con la escritora e ilustradora Lola Anglada, quien un día pudo escuchar cómo

su padre le decía al de Miró: "Son artistas y vivirán en un desorden completo si no les asignas una cantidad mensual". Ella les interrumpió: "Yo no me quiero casar". Ya anciana, Lola Anglada criticó el arte abstracto ("es una tomadura de pelo") y de Miró ("dibuja y pinta con tanta sinceridad como lo hace el chimpancé Congo del zoo de Londres"). Un segundo intento de matrimonio fue con Pilar Tey. Miró puso fin a su relación ocho días antes de la boda ya fijada y envió a su tío a casa de Pilar Tey para devolverle las cartas y el anillo de novia.

Massot considera que André Breton teorizó sobre el surrealismo como un movimiento literario, y por eso no se aprecia que Miró ya era un surrealista incluso antes de que se publicara el Manifiesto surrealista en 1924. "Frente al virtuosismo inalcanzable de Picasso, opta por la solución Picabia, la extrema audacia, y creo que no se ha valorado lo suficiente el peso que Picabia y el dadaísmo han tenido sobre Miró". Se trasladó a París en 1920 y rompió con el novecentismo, siguiendo los consejos de Picasso, que le recordó que "en Cataluña se necesitan pasión y heroísmo". André Masson y Paul Klee, los poetas Breton, Artaud, Desnos y Leiris, así como el arte primitivo, ayudan a configurar un nuevo lenguaje.

A Miró lo sorprendió el 18 de julio de 1936 en Mont-roig y, a pesar de su condición de republicano y catalanista de izquierdas, unos meses después se marchó a París y ya no regresó. Su hermana se había casado con un terrateniente de Tona que fue asesinado por el comité revolucionario local. La casa fue saqueada e incluso desapareció un cuadro suyo, de 1923, que les había regalado para su boda. En la capital francesa, Miró diseñó el sello Aidez l'Espagne, por encargo del Comisariado de Propaganda, y pintó El segador -que después se perdería-, en el pabellón de la República Española en la Exposición Internacional de 1937. Tras la victoria franquista, permaneció en Varengeville-sur-Mer (Normandía), donde comenzará a pintar Les constellations. Planeó huir a los Estados Unidos por consejo de Sert. Sin embargo, los primeros bombardeos nazis obligan a Miró a regresar a París con su esposa y su hija en un dramático viaje en tren, ahora conocido gracias a un relato inédito de Pilar Juncosa. Y fue ella quien convenció a Miró de volver a España. El 8 de junio de 1940, los Miró regresan a Cataluña, se esconden unos días en Sant Hipòlit de Voltregà, en la masía del cuñado asesinado, considerado por los franquistas un mártir de la patria, y después se trasladan a Palma. Se libera de represalias a costa del anonimato y el ostracismo. Solo la pintura le permitirá sobrevivir en tiempos de barbarie. Regresará a Barcelona en 1942."

Referencias

- Baskara, R. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 343-358.
- Lin, Z. (2024). How to write effective prompts for large language models. *Nature Human Behaviour*, 1-5.
- Lluna, A. (2024). Cómo mejorar tus textos generados por ChatGPT usando la técnica de asignación de roles. Fecha de consulta: 2/4/2024 <https://ainalluna.com/blog/la-importancia-de-asignar-un-rol-al-prompt>
- Mendoza, D. (2020). Racismo y roles de género, conductas perpetuadas en algoritmos de inteligencia artificial. *Coloquio*, (65), 131-135.
- Pietikäinen, M., & Silven, O. (2022). Challenges of Artificial Intelligence--From Machine Learning and Computer Vision to Emotional Intelligence. arXiv preprint arXiv:2201.01466.
- Rodríguez Gómez, J. M. (1997). Desarrollo educativo de la adopción de roles. *Aula abierta*, (69), 75-88.

05

El pensamiento arborescente

Para explicarnos su propuesta filosófica, Platón fabula en La República un diálogo entre Sócrates y Trasímaco que en algunos momentos es una verdadera batalla dialéctica. Este recurso narrativo le permite a Platón mostrarnos algunos de los pasajes centrales de la historia de la filosofía, como el mito de la caverna. Pero estos diálogos son más que un mero recurso estilístico: Platón está reproduciendo el proceso de razonamiento a partir de dos posiciones opuestas que dialogan entre sí. La capacidad de llegar a una conclusión sugiere Platón, es más elevada cuando confrontamos las ideas entre ellas. Seguramente, algunos de los fragmentos reproducen las batallas dialécticas que propició Platón en la Academia de Atenas.

Nuestra capacidad de razonamiento no sigue un hilo conductor lineal. Consciente o inconscientemente, las ideas se organizan en un recorrido mucho más complejo, en el cual incorporamos otras opciones, descartamos puntos de vista o retrocedemos hasta que encontramos un camino que nos parece el más adecuado. Se puede representar esta lógica con la metáfora de un árbol con múltiples ramas y a esta

forma de razonamiento, pensamiento arborescente. En este epígrafe, mostraremos formas de aplicar el concepto del pensamiento arborescente con los modelos de lenguaje como Claude, ChatGPT, Llama, Copilot o Gemini, pero antes detallaremos dos precedentes: la cadena de pensamiento y la autoconsistencia.

5.1. La cadena de pensamiento (CoT)

Los diversos modelos de lenguaje pueden utilizarse para razonamientos simples. De alguna manera, el sistema parece imitar el proceso de razonamiento y es capaz de seguir secuencias lógicas que imitan el razonamiento natural. La siguiente tabla muestra las respuestas de ChatGPT 3.5 y Gemini a algunos problemas lógicos tradicionales. El modelo de ChatGPT es muy eficiente, mientras que Gemini muestra dificultades para cerrar el proceso lógico. Sin embargo, es posible que en el caso de ChatGPT no se trate de una resolución, sino de la consulta de la solución entre sus contenidos, de la misma manera que resolvemos problemas lógicos porque los conocemos.

“Mostraremos formas de aplicar el concepto del pensamiento arborescente con los modelos de lenguaje como Claude, ChatGPT, Llama, Copilot o Gemini”

PROBLEMA	PLANTEAMIENTO	GEMINI	CHATGPT 3.5
Agua y vino	https://en.wikipedia.org/wiki/Wine/water_mixing_problem	Error	Acierto
Caballeros y vasallos	https://en.wikipedia.org/wiki/Knights_and_Knaves	Acierto	Acierto
Tarea de Wason	https://en.wikipedia.org/wiki/Wason_selection_task	Error	Acierto
Puente y torcha	https://en.wikipedia.org/wiki/Bridge_and_torch_problem	Error	Acierto
Lobo, cabra y col	https://en.wikipedia.org/wiki/Wolf_goat_and_cabbage_problem	Acierto	Acierto
Misioneros i caníbales	https://en.wikipedia.org/wiki/Missionaries_and_cannibals_problem	Acierto	Acierto
Cumpleaños de Cheryl	https://en.wikipedia.org/wiki/Cheryl%27s_Birthday	Acierto	Acierto

Por ello, los programas comienzan a tener problemas en razonamientos lógicos, incluso si responden a preguntas clásicas, pero a las que les introducimos variantes. Por ejemplo, el problema “Un naufrago llega a una isla en un archipiélago en el que sólo hay dos islas habitadas: en una, los isleños siempre dicen la verdad y en la otra los isleños siempre dicen la mentira. ¿Qué debería preguntar el naufrago a un isleño que se ha encontrado para saber en qué isla está?” genera respuestas incorrectas. En el desarrollo de la respuesta, los programas incluso reconocen explícitamente su incapacidad para encontrar una solución. A medida que el problema se vuelve más

complejo, se hace más evidente la limitación del razonamiento lógico secuencial de los modelos de lenguaje actuales, y esto es especialmente evidente en el mundo de las ciencias sociales.

Una forma de mejorar el resultado es con la Cadena de Pensamiento (*Chain of Thought, CoT*) propuesta por Wei et al. (2022). Una CoT, en sentido estricto, es una pregunta que necesita una respuesta “en cadena”, es decir, con una suma de razonamientos que siguen una secuencia. Si se fuerza al modelo para que explique su proceso de razonamiento (la “cadena”), la solución tiende a ser más eficiente.

Pregunta sin CoT

¿Cómo se pueden reducir las emisiones de CO2 en la aviación?

Respuesta sin CoT

- Eficiencia del combustible
- Aerodinámica mejorada
- Operaciones aeroportuarias más eficientes
- Combustibles alternativos
- Electrificación
- Gestión mejorada de las rutas de vuelo
- Compensación de carbono
- Pregunta con CoT

Pregunta con CoT

¿Cómo se pueden reducir las emisiones de CO2 en la aviación teniendo en cuenta estos tres criterios: (a) mejora de la eficiencia, (b) reducción del número de vuelos, (c) combustibles más verdes? Selecciona las opciones que pueden lograr una mayor reducción de CO2.

Respuesta con CoT

- Aeronaves más eficientes y operaciones optimizadas
- Impuestos más altos para desincentivar el uso del avión
- Biocombustibles y combustibles sintéticos

Mientras que la respuesta sin CoT contempla solo soluciones de eficiencia y enumera soluciones relativamente viables (combustibles alternativos) junto con medidas que en estos momentos son muy poco probables (la electrificación), el uso de la CoT introduce una visión más amplia de opciones e incorpora un proceso de selección y jerarquía. Los *prompts* pueden guiar al modelo hacia una solución más compleja si se introduce una cadena de razonamiento con alternativas.

Otra forma de incorporar la CoT es con los *shots*. Como hemos visto en el capítulo tercero, los *shots* se refieren al uso de ejemplos o instrucciones específicas proporcionadas al modelo para orientar su respuesta o producción hacia un objetivo deseado. Estos ejemplos sirven para "informar" al modelo sobre el tipo de respuesta que se espera de él en un contexto específico. Hay varios tipos de uso de *shots*, como *zero-shot*, *few-shot* y *many-shot learning*, que se diferencian por la cantidad de ejemplos proporcionados para el aprendizaje o el ajuste de la respuesta. Con estos "rasgos" podemos orientar al modelo hacia una forma de resolución de problemas, basada en una secuencia, de manera que "enseñamos" al modelo a seguir un razonamiento a partir de una cadena.

Con este ejemplo, hemos introducido una secuencia de razonamiento y de cálculo que entrena al sistema para aplicar esta estructura a cualquier nuevo problema. Gracias a estos criterios, el modelo puede calcular las emisiones teniendo en cuenta que se trata de un vuelo corto (y, por lo tanto, menos eficiente por el impacto de despegue y aterrizaje) y que no está plenamente ocupado, de manera que es necesario repercutir las emisiones totales sobre un número inferior de pasajeros. Con los *shots* orientamos los modelos para que el proceso de “razonamiento” y de respuesta siga un patrón que anticipamos.

Q: ¿Cuál es la emisión de un vuelo a plena ocupación de Lisboa a Ankara?

A: La emisión media de los vuelos largos (más de 2,500 km) con ocupación plena es de 115 kg/km/pax.

La emisión media de los vuelos cortos (menos de 2,500 km) con ocupación plena es de 225 kg/km/pax.

La reducción de la ocupación incrementa proporcionalmente las emisiones por cápita.

Lisboa y Ankara tienen una distancia de vuelo de 3.597 km.

Las emisiones del vuelo Lisboa a Ankara con un vuelo lleno son $3.597 \text{ km} \times 115 \text{ kg} = 413$ toneladas.

Q: ¿Cuál es la emisión de un vuelo con una ocupación del 80% entre Barcelona y Madrid?

En este *prompt*, hemos combinado la cadena de pensamiento (CoT) y el *shot*. Hemos planteado un primer problema y lo hemos resuelto siguiendo una secuencia que incrementa el número de visitantes y reduce los visitantes potenciales que adoptan una estrategia de cancelación. Esta secuencia entrena al modelo y le permite llevar a cabo respuestas a problemas cada vez más complejos, en los cuales aplicará la secuencia que le hemos sugerido con el ejemplo de un *shot*.

Q: En Barcelona, hay 170.000 turistas cada día de media. Durante el Mobile World Congress, hay un incremento de 65.000 congresistas durante los días de la Feria, pero sabemos que del promedio habitual de 170.000 turistas, un 20% decide no ir a la ciudad ante la previsión de una saturación de la misma. La capacidad máxima de turistas de Barcelona es de 200.000 turistas. ¿Cuántos turistas habrá en la ciudad durante el Mobile?

A: $170.000 + 65.000 = 235.000$ turistas potenciales en el *Mobile World Congress*. $0,20 \times 170.000 = 34.000$ turistas deciden no ir $235.000 - 34.000 = 201.000$ turistas.

Q: ¿Cuántos turistas habrá en Barcelona entre junio y julio, si se esperan 76.000 asistentes al Cruïlla (de los cuales un 15% son turistas) y 122.000 al Sónar (de los cuales un 40% son turistas) y la fuerte concentración de visitantes durante los festivales implicará que un 10% de los turistas potenciales decida no asistir a Barcelona durante este período (ya sea porque cambia el período o cambia la ciudad de visita)?

5.2. Autoconsistencia

La cadena de pensamiento se puede mejorar con la opción de autoconsistencia (*Self-Consistency*). En este caso, forzamos la creación de varias respuestas y, a continuación, pedimos al modelo que elija la opción más adecuada. Este concepto ha sido propuesto por Wang et al. (2022) para mejorar la calidad de las respuestas, especialmente en el razonamiento lógico, simbólico y aritmético.

Hay varias formas de generar una secuencia de resultados, pero las dos vías más obvias son:

(a) preguntar varias veces lo mismo y obtener resultados diferentes, y después elegir la opción más adecuada (o forzar al modelo para que elija la mejor opción entre las respuestas), y (b) pedir al modelo que genere varias respuestas o posibles razonamientos y luego elegir la opción más adecuada (o, de nuevo, forzar al modelo a seleccionar la respuesta más consistente).

Podemos intentar trabajar con la Ley de Hotelling sobre [la localización de vendedores de helados en una playa](#), que es un problema clásico de la teoría de localización. Hotelling demuestra que los vendedores tienden a situarse en el centro de la playa, uno al lado del otro, y su teoría ha sido utilizada para explicar el comportamiento de los partidos políticos o la localización comercial en espacios urbanos. Con este *prompt*, forzamos al modelo a crear varios caminos y a elegir el más eficiente. El resultado es más consistente que el simple planteamiento del problema de forma lineal.

En una playa de un kilómetro de longitud en la que hay una distribución homogénea de bañistas situados por toda la playa, hay dos empresas de helados que deben decidir su ubicación. El propósito de cada heladero es conseguir el máximo posible de clientes; la situación ideal de los bañistas es acceder a los helados con la menor distancia posible. Teniendo en cuenta estos criterios, idea tres posibles ubicaciones de las empresas en la playa. A continuación, evalúa la opción más adecuada teniendo en cuenta el criterio de los heladeros; después, la opción más adecuada teniendo en cuenta la visión de los bañistas. Y finalmente, elige la mejor opción conciliando los intereses de unos y otros.

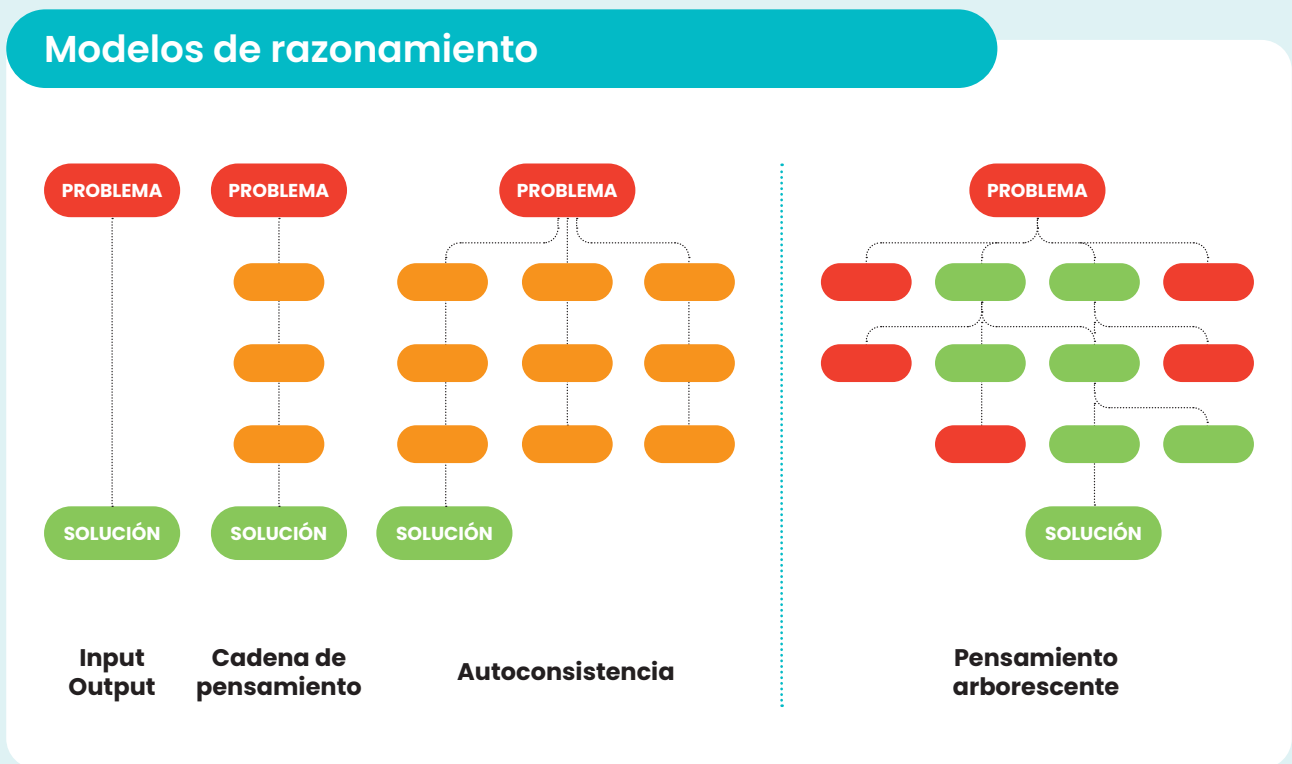
Con este *prompt*, el modelo plantea tres posibles escenarios desde la perspectiva de la maximización de los intereses de los heladeros y concluye que la mejor opción es la ubicación central, tal como prevé el modelo de Hotelling. Cuando se plantea desde la perspectiva de los usuarios, la mejor localización es aquella en la que los dos heladeros están situados en posiciones a un tercio y dos tercios de la playa, ya que es la que minimiza todos los recorridos posibles. Y al evaluar la conciliación entre las dos visiones, concluye que la mejor respuesta es la ubicación entre un tercio y dos tercios, porque ambos colectivos tienen una respuesta óptima a su demanda.

Lo más interesante de este planteamiento es que genera diversas opciones y, a continuación, inicia un proceso de selección de la opción más adecuada. Esta secuencia <alternativas> - <selección> refleja mucho mejor el proceso argumental natural, en el cual nos planteamos de forma más o menos consciente diversas posibles respuestas y seleccionamos la vía o camino más consistente. Wang et al. (2022) han demostrado la potencia de un uso combinado de la cadena de pensamiento y la autoconsistencia.

5.3. Pensamiento arborescente (ToT)

Intentemos imaginar el proceso de razonamiento de un jugador de ajedrez. En cada jugada, el jugador o jugadora tiene unas 30 opciones posibles de media. Para determinar la mejor opción, construye una especie de árbol mental, en el cual las jugadas se ramifican en diversas alternativas; muchos de estos caminos quedan descartados porque el desenlace es muy negativo (pérdida de material, posición inferior...), de manera que se van descartando opciones y ramas hasta que acaba optando por la que aparentemente es la mejor jugada, de acuerdo con los cálculos realizados. Muchos jugadores reconocen esta analogía del árbol para explicar el proceso de su razonamiento.

Siguiendo la lógica de la cadena de pensamiento o de la autoconsistencia, el pensamiento arborescente diseña una estructura con varias ramas y opciones que se van descartando de forma progresiva, de manera que al final se obtiene un resultado muy favorable. La figura muestra el concepto de ToT, comparado con el resto de modelos como el input-output, CoT o autoconsistencia, y permite intuir que el modelo ToT es mucho más eficiente para resolver problemas complejos. Por ejemplo, Long (2023) lo ha utilizado para la resolución de problemas del juego de 24. Mientras que, con los sistemas de cadena de pensamiento, el porcentaje de éxito fue de solo un 4%, con el ToT se elevó hasta el 74%.



El modelo de ToT se utiliza con una secuencia de comandos que ayuda a optimizar la resolución de problemas complejos, pero también se puede emplear utilizando la capacidad de los roles que vimos en el capítulo anterior. El ejemplo que más se ha utilizado

es la resolución del siguiente problema matemático:

$10 + 4 = 2?$

Tanto Gemini como ChatGPT dan una respuesta de error, ya que lógicamente $10+4$ es 14. Para resolver este problema, se sugiere el siguiente *prompt*:

Quiero que asumas el rol de tres profesores de matemáticas: el profesor Lógico, el profesor Creativo y el profesor Sintético. El profesor Lógico responde a las preguntas de forma natural, con las soluciones más instintivas y lógicas. Por el contrario, el profesor Creativo tiende a buscar caminos más elaborados, que a menudo implican pensamiento lateral o alternativas a la lógica convencional. El profesor Sintético lee detalladamente los enunciados, analiza todos sus componentes y busca significados más allá de las opciones más elementales. Ahora te plantearé una cuestión matemática. Quiero que, en primer lugar, el profesor Lógico dé una respuesta a partir del planteamiento del problema. Quiero que, una vez escuchado el profesor Lógico y partiendo de sus conclusiones, el profesor Creativo plantee una respuesta acorde a su perfil. En tercer lugar, el profesor Sintético, tras escuchar a los dos primeros, propondrá una resolución del problema. Finalmente, los tres profesores de forma conjunta, viendo todas las alternativas, escogerán la mejor respuesta posible. El problema es: $10+4=2?$

Este *prompt* permite crear una serie de alternativas siguiendo el modelo de la autoconsistencia, pero en este caso, cada personaje parte de los resultados del anterior y, en la resolución final, los tres seleccionan la solución óptima. En un sistema en base 12, en el cual cada vez que llegamos a 12, reiniciamos la secuencia, $11+3$ sí sería 2; un ejemplo de estas condiciones es el reloj, ya que si son las 10 horas y pasan cuatro horas más, serán las 2.

Existen varias formas de aplicar el pensamiento arbóreo en las sesiones docentes, pero destacaremos tres y propondremos algunos ejercicios para su aplicación:

1. El pensamiento arbóreo, como hemos visto, permite la resolución de problemas o cuestiones, utilizando la capacidad de razonamiento de diferentes perfiles. Esto permite ampliar el abanico de opciones posibles y ayuda a desarrollar habilidades relacionadas con el pensamiento complejo, el pensamiento lateral, el planteamiento de alternativas o la selección de variantes.
2. Esta técnica también ayuda a seguir los razonamientos complejos a partir de problemas planteados, es decir, acompaña al estudiante en el proceso de construir el razonamiento a partir de la adición o la confrontación con las ideas precedentes. Esto permite desarrollar habilidades relacionadas con la inteligencia colaborativa y el contraste de ideas. En este caso, el interés no se centra tanto en el resultado final, sino en el proceso de elaboración de la solución.
3. Finalmente, aunque no es el objetivo inicial del ToT, el planteamiento de un problema desde la perspectiva de diferentes puntos de vista ayuda al estudiante a entender la necesidad de ampliar la mirada en la resolución de problemas. Contrastar puntos de vista, paradigmas o teorías diferentes también puede ayudar a comprender mejor cada posición. Es más fácil entender el idealismo si se compara con el materialismo que si se estudia de forma independiente.

Práctica. El Palacio Azul

El propósito de este ejercicio es introducir a los estudiantes en los procesos deliberativos: la identificación de los argumentos, la ordenación jerárquica de los argumentos y la propuesta de soluciones alternativas que concilien las diversas posiciones planteadas. El ejercicio alude claramente al conflicto relacionado con la devolución de los frisos del Partenón, que actualmente se encuentran en el Museo Británico y son reclamados por el gobierno griego, especialmente tras la construcción del Museo de la Acrópolis. Este conflicto se enmarca en el contexto más amplio del debate sobre la descolonización de los museos. No obstante, no se menciona explícitamente para evitar que el modelo influya en la discusión con información previa y limitar así el debate a las condiciones planteadas en el problema.

Los estudiantes se reúnen en grupo y siguen el hilo argumental de unos y otros, de acuerdo con un modelo que imita el pensamiento arbóreo, a partir de la asunción de roles. El objetivo de la práctica es identificar los argumentos de los dos puntos de vista: tanto los argumentos a favor de la propuesta, como los argumentos contrarios a la propuesta. En segundo lugar, los estudiantes deben ordenar las unidades argumentales de acuerdo con una ponderación que permita crear una jerarquía. Finalmente, los estudiantes deben valorar las propuestas de conciliación planteadas y pueden sugerir nuevas alternativas basadas en las ideas del modelo.

Una precaución que se debe tener en cuenta es que los estudiantes pueden utilizar el modelo de lenguaje para resolver las cuestiones planteadas, por lo que la práctica debería constar de dos partes: una parte online (o *"plugged"*), en la que se sigue la discusión entre los personajes del modelo, y una parte offline (o *"unplugged"*), en la que los estudiantes debaten sobre las propuestas siguiendo el enfoque anterior.

En un museo de una ciudad europea que llamaremos Urbania, se exponen los tesoros obtenidos durante las diversas invasiones militares del imperio del cual Urbania es la capital. Una de las salas más visitadas del Museo Imperial es el Palacio Azul, que fue adquirido en el siglo XIX por el Cónsul del Imperio en Antiqua, una región muy próspera durante el periodo comprendido entre los siglos III y X, pero que no pudo resistir la invasión del Imperio en el siglo XIX.

Desde principios del siglo XX, los responsables culturales de Antiqua reclaman la devolución del Palacio Azul, pero hasta la fecha, el museo se ha negado, argumentando que dicho palacio fue obtenido mediante una compraventa. Dada la creciente tensión entre ambas partes, un representante de la UNESCO propone la celebración de un debate público entre un representante del Museo Imperial y un representante de la gestión cultural de Antiqua.

Quiero que asumas el rol de John Bridgetown, el director del Museo Imperial, un historiador del arte de 40 años formado en Harvard. Aunque es sensible a los planteamientos sobre la descolonización, sostiene que el Palacio Azul debe permanecer en el museo. Su principal argumento se centra en la propiedad: el museo adquirió la obra mediante una compra legal, como lo acreditan los documentos. También argumenta que, gracias a la ubicación del Palacio en el museo, este ha podido ser preservado, considerando los bombardeos que devastaron la ciudad durante los periodos bélicos del siglo XX. Finalmente, propone que el Palacio lleva más de 100 años en el museo y ya forma parte de su identidad.

Quiero que al mismo tiempo asumas el rol de Marina Psi, quien es la directora general de cultura del gobierno de Antiqua, de 46 años, con una amplia formación en gestión cultural adquirida en París. Ella fue quien impulsó la construcción del Museo Antiguo, donde se ubicaría el Palacio Azul en caso

de ser devuelto. Considera que la compraventa fue irregular, ya que en ese momento Antiqua estaba sometida al Imperio Nómada, que fue quien negoció la transacción. Su principal argumento es que la Carta de Venecia propone que las obras de arte deben estar en el lugar donde fueron creadas, y apoya los argumentos del movimiento de descolonización de los museos. Marina sostiene que Antiqua está perfectamente preparada para la conservación del Palacio, gracias al Museo Antiguo.

Finalmente, quiero que asumas el rol de Antonio Santos, el responsable de la UNESCO propuesto para mediar entre las dos posiciones. Debes crear la ficción de un debate abierto, público y con grandes expectativas, retransmitido por los medios nacionales e internacionales, dirigido por Antonio Santos, en el que se discute la propuesta de devolver el Palacio Azul a Antiqua. Quiero que primero asumas el rol de Bridgetown y defiendas su postura, y luego asumas el rol de Psi y hagas lo mismo, de manera que se pueda seguir el debate entre ambas posiciones.

En la primera fase del debate, se expondrán de forma extensa y documentada los argumentos de cada parte, con la posibilidad de interrupción si el moderador, Santos, lo permite. En la segunda fase del debate, los dos representantes intentarán refutar la posición del otro. Finalmente, en la última fase del debate, se presentarán posibles soluciones intermedias que puedan conciliar los intereses de ambas partes.

¿Has entendido lo que debes hacer?

Práctica. Anuncios o no anuncios

Este ejercicio tiene como objetivo desarrollar de forma aplicada la teoría de juegos y, concretamente, el dilema del prisionero. En el dilema del prisionero, se parte de la condición de no cooperación, pero pueden existir al menos dos escenarios más: el de cooperación y el de engaño. Partiendo del mismo supuesto clásico del dilema del prisionero, pero aplicado a una decisión comercial, se plantean tres escenarios: cooperación, no cooperación y engaño. El modelo resuelve los tres supuestos de forma eficiente y realiza una síntesis de cada resultado, lo que permite comprender los factores que influyen en la toma de decisiones.

En la isla mediterránea de Corsilia operan dos cadenas de supermercados, *DeTodo* y *Económico*. Se reparten el mercado de manera equitativa, logrando cada un beneficio anual de 1 millón de euros. En 2024 se instala en la isla una televisión local, TVC, que tiene una enorme influencia entre los consumidores. Los responsables de TVC proponen a cada cadena, por separado, un paquete de anuncios por valor de 250.000 euros. Los expertos en marketing de cada cadena han calculado que, si emiten los anuncios y la otra compañía no lo hace, podrían ganar un 50% de la cuota de mercado de la otra cadena.

Plantea cuál es la mejor estrategia para ambas cadenas en un escenario de cooperación, cuál es la mejor solución en un escenario de no cooperación (cuando cada cadena desconoce lo que hará la otra), y en un escenario en el que una cadena pueda engañar a la otra.

SOLUCIÓN DEL MODELO

Escenario Cooperación

En este escenario, ambas cadenas acuerdan no invertir en publicidad. De este modo, mantienen sus beneficios anuales sin arriesgarse a una guerra de precios o a una escalada en los gastos de marketing que podría erosionar sus márgenes de beneficio. En este caso, cada cadena mantendría su millón de euros de beneficio.

No Cooperación (Dilema del prisionero clásico)

En este escenario, cada cadena debe decidir de manera independiente si invierte en publicidad sin saber qué hará la otra cadena.

Si ambas invierten, cada una paga 250.000 euros y no gana cuota adicional de mercado, ya que los efectos de la publicidad se anulan mutuamente. Sus beneficios netos serían de 750.000 euros cada una.

Si una invierte y la otra no, la cadena que invierte aumenta sus beneficios en un 50% de la cuota de mercado de la otra, suponiendo que este 50% equivale a 500.000 euros adicionales (50% de 1 millón). Por lo tanto, después de restar el gasto publicitario, el beneficio neto para la cadena que invierte sería de 1,25 millones de euros, mientras que la cadena que no invierte quedaría con 500.000 euros.

Si ninguna de las dos invierte, manteniendo el statu quo, cada una conservaría su millón de euros de beneficio.

Traición o Engaño

En este escenario, si una cadena promete no invertir en publicidad, pero luego lo hace, y la otra cadena mantiene su palabra de no invertir, la cadena que ha invertido obtendría una gran ventaja (1,25 millones de euros de beneficio neto), mientras que la cadena engañada sufriría una gran pérdida (500.000 euros de beneficio).

Esto podría llevar a una desconfianza futura y a posibles guerras de precios o gastos en marketing, deteriorando los beneficios a largo plazo para ambas cadenas.

Conclusión

La mejor estrategia en cooperación es no invertir en publicidad, manteniendo el statu quo y evitando el riesgo de disminuir los beneficios netos.

En un escenario de no cooperación, la tendencia natural según el Dilema del Prisionero sería que ambas cadenas acaben invirtiendo en publicidad por temor a perder cuota de mercado, lo que resultaría en beneficios netos inferiores.

Para evitar el engaño, sería ideal establecer mecanismos de confianza mutua o acuerdos formales, aunque esto puede ser difícil en entornos altamente competitivos.

Práctica. La taberna de los filósofos

Este ejercicio crea el escenario imaginario de una taberna. En esta taberna coinciden filósofos que representan posiciones contrapuestas, como por ejemplo Hegel y Bunge, ya que el primero es uno de los máximos exponentes del idealismo mientras que Bunge es conocido por defender una visión materialista. Una de las formas de seguir los procesos de argumentación de posiciones filosóficas es forzando el contraste con otras posiciones más o menos alejadas. Por ello, y siguiendo la lógica del pensamiento arbóreo, este ejercicio plantea una conversación ficticia entre filósofos en una taberna a partir de un problema más o menos actual. En concreto, se propone la discusión entre tres filósofos a partir de una noticia reciente (“¿es todo el universo una simulación?”).

Recientemente, algunos científicos han planteado la hipótesis de que el universo podría ser una simulación. En este artículo se muestra esta teoría.

Quiero que adoptes la personalidad de Hegel, Bunge y Nietzsche, en una fabulación en la que, aunque pertenecen a épocas diferentes, imaginamos que coexisten y deciden salir un día a tomar unas cervezas en el barrio Moabit de Berlín. El camarero, que conoce la capacidad intelectual de sus tres clientes, les comenta la hipótesis de que el universo, en realidad, es una simulación.

Quiero que describas la acalorada discusión entre los tres amigos a partir de la propuesta que les ha hecho el tabernero. Cada filósofo intervendrá después de haber escuchado los argumentos de los otros. Tras las primeras intervenciones, los filósofos intentarán refutar las posiciones de sus amigos y detectar sus puntos débiles. En una tercera ronda de intervenciones, los filósofos explorarán posibles soluciones basadas en la conversación mantenida.

¿Has entendido lo que tienes que hacer?

En una asignatura de filosofía, se podría repetir esta escena intercambiando los personajes (los filósofos) y los problemas planteados (que deberían ser lo más actuales posible), con una constante, que sería la taberna y el tabernero. En el *prompt*, se puede perfilar tanto la identidad de los filósofos como los elementos secundarios (la taberna y el camarero). En todo caso, el modelo se alimenta de los conocimientos previos sobre los filósofos más universales, por lo que no es necesario introducir elementos de contexto en el *prompt*, como hemos sugerido en el capítulo anterior.

Referencias

Long, J. (2023). Large language model guided tree-of-thought. arXiv preprint arXiv:2305.08291.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A. & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171. [<https://arxiv.org/abs/2203.11171v4>]

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Qoc, Le & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824–24837. [<https://arxiv.org/abs/2201.11903>]

06

Reconocer el espejismo del desierto: las alucinaciones

Un espejismo es una ilusión óptica que a menudo se produce en entornos como un desierto. En películas, libros o cómics, hemos visto personajes que, en un momento de debilidad, son sorprendidos por un espejismo, una ilusión seductora ante sus ojos. Esto es lo que puede sucedernos cuando interactuamos con una IA. El modelo nos proporciona una respuesta a nuestra consulta que parece real y factible, pero no deja de ser una ilusión, un espejismo en el desierto, que puede ser engañosa, inventada, sin sentido y no verificada o verificable. Esto se ha popularizado con el nombre de “alucinaciones”. Una alucinación es un término psicológico que se refiere a un tipo de percepción y que, en términos simples, es una percepción irreal que parece real (Ji et al., 2022):

El texto alucinado da la impresión de ser fluido y natural, a pesar de ser impreciso y sin sentido. Parece estar basado en el contexto real proporcionado, aunque en realidad es difícil especificar o verificar la existencia de tales contextos. Al igual que la alucinación psicológica, que es difícil de distinguir de otras percepciones “reales”, el texto alucinado también es difícil de captar a primera vista.

Autores como Østergaard y Nielbo piensan que esta denominación es inadecuada para describir las respuestas falsas de los modelos de IA: en primer lugar, es una metáfora imprecisa, ya que describe una percepción sensorial en ausencia de un estímulo externo; y, en segundo lugar, es altamente estigmatizante, ya que puede estar asociada a muchas enfermedades, principalmente neurológicas o mentales. Además, proponen términos alternativos como “*non sequitur*” o “respuesta no relacionada” (Østergaard y Nielbo, 2023).

Si se examina la clasificación de este fenómeno, se puede decir que hay dos tipos de alucinaciones: las alucinaciones intrínsecas, en las que la salida generada contradice el contenido de la fuente, y las alucinaciones extrínsecas, en las que la salida generada no puede ser verificada (ni contradicha ni avalada) a partir del contenido de la fuente (Ji et al., 2022).

Aparte de las alucinaciones de la IA, también se debe hablar en este capítulo del concepto de “loros estocásticos”. Este término, según Li (2023), se refiere a la repetición de los datos de entrenamiento o de sus

“Más de un 60% de las respuestas en los conjuntos de datos estándar son alucinados”

patrones, más que a una comprensión o razonamiento real. Este autor explica que “El método de producción de texto de los LLM (modelos de lenguaje grande) es reutilizar, remodelar y recombinar los datos de entrenamiento de maneras nuevas para responder a nuevas preguntas, mientras ignoran el problema de la autenticidad y la fiabilidad de las respuestas” (Li, 2023: 1). Por lo tanto, estos modelos solo predicen qué palabra concreta viene después en una frase, en lugar de entender realmente qué significa.

Este capítulo aborda el tema de las alucinaciones y los “loros estocásticos” en los sistemas de IA. Es un tema importante porque, a veces, al utilizar la IA se obtienen respuestas que no tienen sentido o que son totalmente inventadas. Se debe ser consciente de este hecho, y es más necesario que nunca verificar la información que proporciona la IA. Para el profesorado, esto es importante, por un lado, porque al preparar el material docente, este debe ser verificado y contrastado. Por otro lado, como docentes, se debe enseñar a los alumnos a contrastar y verificar la información que obtienen de estos sistemas.

De este modo, el capítulo comenzará con un subapartado en el que se explicarán las causas de las alucinaciones y los “loros estocásticos”, es decir, por qué falla la IA. Se verá que, básicamente, falla por dos motivos: por los datos o por el entrenamiento. A continuación, se ofrecerán algunas recomendaciones para reducir este fenómeno y, finalmente, se quiere enfatizar la importancia de verificar la información que generan los modelos de IA. El primer ejercicio del capítulo está

relacionado con la reducción de las alucinaciones y, por lo tanto, del ruido en las consultas a la IA, mientras que el segundo será sobre la verificación de resultados.

6.1. ¿Por qué falla la IA?

Uno de los motivos por los cuales la inteligencia artificial falla es a causa de los datos de entrenamiento (Dziri et al., 2022). El error produce respuestas no verificadas o engañosas (alucinaciones) o respuestas sin sentido o repetitivas (loros estocásticos). Por lo tanto, la calidad de los datos de entrenamiento, los errores que contienen, la desactualización o los sesgos de estos datos provocan las alucinaciones o los loros estocásticos, conceptos que se han definido previamente. Por otro lado, un segundo motivo es la propia arquitectura del modelo (Dziri et al., 2022). Los sistemas de IA tienen una gran dependencia de las correlaciones estadísticas, ya que esta es la forma en que producen las respuestas, a menudo sin comprender la información procesada. Por lo tanto, se producen errores de codificación y decodificación que provocan las alucinaciones y los loros estocásticos.

Estudios como los de Dziri et al. (2022) demuestran que más de un 60% de las respuestas en los conjuntos de datos estándar son alucinados, lo que los lleva a concluir que los modelos no solo alucinan, sino que incluso amplifican las alucinaciones. En la literatura científica ya se mencionan casos en los que sistemas de IA como ChatGPT producen referencias bibliográficas falsas o incorrectas. Se

insiste, por lo tanto, en la necesidad de revisar el texto producido por la IA para asegurar la integridad del contenido, sobre todo en áreas sensibles como la salud y el derecho, entre otras.

En la literatura existente, algunas de las principales preocupaciones con respecto al uso de la IA están precisamente relacionadas con la falta de control de calidad de la información, la desinformación, las aplicaciones *deepfake* y el sesgo algorítmico causado por datos defectuosos (Wach et al., 2023). Por este motivo, muchos autores abogan por el uso de la tecnología de una manera ética y responsable. En los siguientes subapartados se abordarán las medidas que se pueden tomar para reducir las alucinaciones y verificar el contenido de las respuestas para evitar la propagación de desinformación.

6.2. Cómo reducir las alucinaciones

Según Ji et al. (2022), los dos factores que contribuyen a las alucinaciones son: los conjuntos de datos y el entrenamiento y la inferencia de los sistemas con esos datos. Por un lado, la principal causa de las alucinaciones relacionadas con los conjuntos de datos es la divergencia entre la fuente y la referencia (Ji et al., 2022). Tal como explican los mismos autores, esta divergencia se produce por la recopilación heurística de los datos o por la divergencia innata de estos. Por otro lado, otro de los factores que contribuyen a las alucinaciones son las elecciones de entrenamiento y modelado de los modelos neuronales (Ji et al., 2022). En este sentido, según los mismos autores, estas alucinaciones pueden producirse por un aprendizaje de representaciones imperfecto, una decodificación errónea, un sesgo de exposición o un sesgo de conocimiento paramétrico.

Métodos de mitigación de las alucinaciones

MÉTODOS RELACIONADOS CON LOS DATOS	MÉTODOS DE MODELADO E INFERENCIA
<p>CONSTRUIR UN CONJUNTO DE DATOS FIEL</p> <p>Es un método intuitivo para reducir el ruido de los datos y, por ende, las alucinaciones.</p>	<p>ARQUITECTURA</p> <p>En algunos trabajos se modifican los codificadores (<i>encoder</i>), en otros la atención (<i>attention</i>) y en otros los descodificadores (<i>decoder</i>). En todos los casos, las modificaciones buscan reducir las alucinaciones mejorando la interpretación semántica de los datos de entrada, la correlación entre la fuente y los resultados, o la mejora del texto generado, respectivamente.</p>
<p>LIMPIAR DATOS AUTOMÁTICAMENTE</p> <p>Es una manera de reducir el ruido semántico. Consiste en identificar datos irrelevantes o contradictorios, para luego filtrarlos o corregirlos.</p>	<p>ENTRENAMIENTO</p> <p>Se pueden adoptar diferentes aproximaciones:</p> <ul style="list-style-type: none">Planificación/Borrador. Sirve para controlar y restringir lo que genera el modelo.Aprendizaje por refuerzo. Uso de recompensas para optimizar el modelo.Aprendizaje multitarea. Un modelo compartido se entrena en múltiples tareas de forma simultánea para aprender similitudes entre tareas.Generación controlable. Se utiliza para mejorar la fidelidad mediante el re-muestreo controlado, códigos de control manuales o predicción automática.
<p>AUMENTAR LA INFORMACIÓN</p> <p>Si se aumentan las entradas con información externa, se obtendrá una mejor correlación entre la fuente y el objetivo, es decir, a mayor comprensión semántica, menor divergencia respecto a la fuente.</p>	<p>POST-PROCESAMIENTO</p> <p>Estos métodos corrigen las alucinaciones en las salidas. Requieren menos datos de entrenamiento y son una opción competitiva para conjuntos de datos con mucho ruido; sin embargo, producen textos no gramaticales.</p>

Fuente: Elaboración propia a partir de Ji et al. (2022).

En primer lugar, por lo tanto, para reducir las alucinaciones y los loros estocásticos, se pueden mejorar los datos de entrenamiento y el entrenamiento de los sistemas de IA con datos diversos y de alta calidad (Wach et al., 2023). Otros estudios también lo destacan: “Esto requiere la publicación de datos limpios y de alta calidad y un diseño cuidadoso de sistemas conversacionales de confianza” (Dziri et al., 2022: 5275). Además, algunos autores enfatizan “la necesidad de invertir recursos significativos en la curación y documentación de los datos de entrenamiento de los modelos de lenguaje” (Bender et al., 2021).

Wach et al. (2023) establecen que una manera de mitigar los riesgos asociados con la baja calidad de la información y la falta de mecanismos de control de calidad es “establecer un bucle de retroalimentación humana, con usuarios que proporcionen comentarios para ayudar al modelo a aprender de los errores y mejorar con el tiempo”. De esta manera, el modelo podría generar datos más relevantes y fiables.

Además, en las consultas a las IA, se pueden elaborar *prompts* más refinados con ejemplos y contexto. La división de tareas en diversos *prompts*, la iteración y la retroalimentación son algunas acciones de ingeniería de *prompts* que pueden evitar o reducir las alucinaciones y los loros estocásticos (para más información sobre la ingeniería de *prompts*, se puede consultar el capítulo 3).

La segunda acción que se puede llevar a cabo para reducir las alucinaciones es utilizar los sistemas de IA de manera responsable y ética. Esta afirmación se traduce en los siguientes aspectos: transparencia en el uso de los conjuntos de datos que se emplean en el entrenamiento de estos sistemas; mitigación de los sesgos existentes en los conjuntos de datos; y evaluación humana de la información introducida en los sistemas de IA (Wach et al., 2023). Los mismos autores mencionan que se podrían crear directrices y estándares claros para el uso de la IA.

La tercera acción es la regulación y el control de calidad para evitar la desinformación y los *deepfakes*. Según Wach et al. (2023), combatir la desinformación implica establecer mecanismos de verificación. La

regulación no está al alcance de los individuos, pero sí lo está el control de calidad y, por lo tanto, se debe tener en cuenta que siempre es necesario verificar los datos generados por la IA (como se podrá ver en el siguiente subapartado de este capítulo).

Educar a los usuarios, lo que también incluye al alumnado, en la detección de la desinformación es otra acción que podría contribuir a reducir las alucinaciones. Además, es importante considerar el sesgo algorítmico y los datos incorrectos que puede tener el propio sistema (para más información sobre el sesgo de la IA, se puede consultar el siguiente capítulo).

Las soluciones tecnológicas también podrían contribuir a reducir o evitar las alucinaciones y los loros estocásticos. En el caso del sesgo algorítmico, pero también en la detección de la desinformación y los *deepfakes*, el factor humano es esencial. Por ejemplo:

[...] depende de nosotros establecer los parámetros adecuados para el rendimiento del algoritmo y llevar a cabo controles puntuales regulares, proporcionar conjuntos de datos extensos a los algoritmos y revisar cómo funcionan estos algoritmos desde diferentes perspectivas para hacer que trabajen a nuestro favor, no en contra de nosotros o de algunos de nosotros (Wach et al., 2023).

En definitiva, la clave para evitar o reducir los errores de la IA está en los datos empleados en estos sistemas, la arquitectura de los propios modelos y el uso ético no solo de estos sistemas, sino también de la información utilizada para su entrenamiento. Por lo tanto, la reducción de las alucinaciones y los loros estocásticos implica combinar soluciones tecnológicas para mejorar los modelos (detección automática de desinformación y *deepfakes*) y soluciones “humanas” como la educación de los usuarios, la evaluación de la calidad de los datos, el establecimiento de políticas y directrices en el uso de la tecnología, y el uso ético y responsable de la IA.

6.3. La necesidad de verificar

Como se ha visto anteriormente, los sistemas de IA pueden fallar, es decir, pueden proporcionarnos respuestas engañosas, sin sentido o repetitivas. Por este motivo, es necesario verificar y contrastar la información que ofrecen estos sistemas. La evaluación humana de la calidad y la fiabilidad de esta información es de vital importancia. Una de las maneras de verificar la información es contrastarla con otras fuentes. Algunos libros de estilo del sector del periodismo recomiendan contrastar una información con hasta tres fuentes diferentes para que sea publicable en los medios de comunicación. Esta puede ser una máxima que se podría aplicar en otros ámbitos, por ejemplo, para verificar la información proporcionada por una IA.

Los propios sistemas de IA ya reconocen en sus páginas web que pueden cometer errores o proporcionar información incorrecta. De aquí la importancia de la verificación. Algunas IA facilitan esta tarea porque proporcionan enlaces a las fuentes de información utilizadas para generar las respuestas. Es el caso de Perplexity, que, según su página web, desarrolla el siguiente proceso para responder a las consultas: interpretar la pregunta, buscar información en fuentes autorizadas en Internet, resumir la información en una respuesta coherente y citar las fuentes con notas a pie de página. Esto es útil para los usuarios a la hora de contrastar y verificar la información, así como explorar información adicional relacionada con la consulta. Además, el usuario puede buscar otras fuentes para comparar y terminar de verificar los datos.

El proceso que sigue Copilot es muy similar para generar las respuestas. Este sistema de IA también cita las fuentes. De hecho, según su página web, Copilot tiene el objetivo de basar todas sus respuestas en fuentes fiables, aunque puede cometer errores o proporcionar contenido de terceros que no sea preciso o fiable. Por ello, igualmente se recomienda utilizar el propio juicio y verificar la información.

Gemini recomienda verificar la información mediante el botón de Buscar en Google. Aunque esta función también puede cometer errores si la información

disponible en Internet no es suficientemente sólida. Este sistema suele generar contenido original, pero, al igual que Perplexity y Copilot, también proporciona enlaces a las fuentes de las cuales reproduce el contenido exacto. Si el contenido está presente en varias páginas, el sistema suele citar la fuente más popular. Además, en la página web de Gemini se explica que Google recoge opiniones y comentarios de los usuarios para mejorar esta inteligencia artificial. Copilot también recopila comentarios.

Actualmente, en internet, hay una gran cantidad de información que a menudo sirve o puede servir para entrenar los modelos de IA. Sin embargo, esto puede resultar problemático porque los datos de entrenamiento pueden contener datos erróneos o estereotipos que se perpetuarán en las respuestas generadas por estos sistemas. El tamaño de los conjuntos de datos a menudo no garantiza su diversidad; a veces, los datos no reflejan los cambios en las perspectivas sociales y otros contienen sesgos que producen una codificación estereotipada y despectiva en relación con el género, la etnia o la discapacidad, entre otros (Bender et al., 2021). Por todos estos motivos, es tan necesaria la verificación humana.

Una línea de investigación que relaciona el periodismo y la documentación es el estudio de la desinformación y las noticias falsas (fake news). En este ámbito, hay muchas investigaciones que destacan la importancia de la alfabetización mediática e informacional para detectarlas. Además, es más necesario que nunca, con las tecnologías y la IA, incorporar la alfabetización y las competencias digitales en los currículos del estudiantado. En 2018, la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) expresó su preocupación por el fenómeno de las noticias falsas. Asimismo, elaboró herramientas para promover la alfabetización mediática e informacional, como una infografía sobre cómo detectar las noticias falsas y recomendaciones para gobiernos, instituciones educativas y sectores económicos, entre otros.

En el ámbito del periodismo, la verificación de datos es muy relevante, especialmente por parte de los

periodistas que informan sobre diversas temáticas de actualidad. A raíz de la importancia que ha adquirido en los últimos años la desinformación y las noticias falsas, han surgido iniciativas de verificación de hechos (*fact-checking*) como [Verificat](#) o [Maldita](#), entre otras. Estas iniciativas de verificación de datos suelen explicar la metodología que siguen. Algunos aspectos o recomendaciones de estas metodologías pueden ser útiles para verificar las respuestas generadas por la IA (Verificat, 2024; Maldita, 2022): consultar fuentes y bases de datos oficiales; contrastar con fuentes académicas, técnicas y especializadas; y contextualizar y verificar con, al menos, dos fuentes diferentes, entre otras.

En el ámbito de la biblioteconomía y la documentación, hace algunos años surgieron diversos métodos para evaluar las fuentes de información, especialmente

desde la explosión de Internet y la evaluación de recursos digitales como las páginas web. Uno de los métodos más sencillos para verificar información es el CRAAP test, en el cual se evalúan cinco indicadores: Actualidad (*Currency*), Relevancia (*Relevance*), Autoridad (*Authority*), Exactitud (*Accuracy*) y Propósito (*Purpose*). A raíz de este método han surgido otros que añaden parámetros o los definen de forma más precisa. Algunos de estos indicadores pueden aplicarse a la evaluación de la información proporcionada por un sistema de IA. En la siguiente tabla, se pueden ver las preguntas de verificación de información que propone el CRAAP test. No todos son adecuados para evaluar la información generada por una IA, pero sí que podemos utilizarlos para buscar otras fuentes que contrasten o verifiquen los datos recibidos.

Test, método de evaluación de fuentes de información

Actualidad: La puntualidad de la información	<ul style="list-style-type: none"> - ¿Cuándo se publicó o subió la información? - ¿Se ha revisado o actualizado la información? - ¿Tu tema requiere información actual o pueden servir fuentes más antiguas? - ¿Los enlaces son funcionales?
Relevancia: La importancia de la información para tus necesidades	<ul style="list-style-type: none"> - ¿La información está relacionada con tu tema o responde a tu pregunta? - ¿Quién es el público destinatario? - ¿La información está en un nivel adecuado (es decir, no demasiado elemental o avanzado para tus necesidades)? - ¿Has revisado una variedad de fuentes antes de determinar que esta es la que utilizarás? - ¿Te sentirías cómodo citando esta fuente en tu trabajo de investigación?
Autoridad: La fuente de la información	<ul style="list-style-type: none"> - ¿Quién es el autor/editor/fuente/patrocinador? - ¿Cuáles son las credenciales o afiliaciones organizativas del autor? - ¿Está el autor calificado para escribir sobre el tema? - ¿Hay información de contacto, como un editor o una dirección de correo electrónico? - ¿La URL revela algo sobre el autor o la fuente? Ejemplos: .com, .edu, .gov, .org, .net.
Precisión: La fiabilidad, veracidad y corrección del contenido del	<ul style="list-style-type: none"> - ¿De dónde proviene la información? - ¿La información está basada en pruebas? - ¿La información ha sido revisada o evaluada? - ¿Puedes verificar alguna de la información en otra fuente o desde tu conocimiento personal? - ¿El lenguaje o tono parecen imparciales y libres de emociones? - ¿Hay errores de ortografía, gramática o tipografía?
Propósito: La razón por la cual existe la información.	<ul style="list-style-type: none"> - ¿Cuál es el propósito de la información? ¿Es para informar, enseñar, vender, entretener o persuadir? - ¿Los autores/patrocinadores hacen claras sus intenciones o propósito? - ¿La información es un hecho, una opinión o propaganda? - ¿El punto de vista parece objetivo e imparcial? - ¿Existen sesgos políticos, ideológicos, culturales, religiosos, institucionales o personales?

Fuente: Meriam Library, California State University, Chico (<https://library.csuchico.edu/sites/default/files/craap-test.pdf>)

Según Codina (2023), lo que siempre se debe hacer al utilizar IA, especialmente en trabajos académicos, es verificar los datos y evaluar los argumentos. Para verificar los datos, es necesario utilizar fuentes fiables, saber utilizar los sistemas de información y tener capacidad de análisis de las fuentes (calidad, autoría, etc.) (Codina, 2023). Por otro lado, según el mismo autor, para evaluar los argumentos se debe utilizar el pensamiento crítico. Esta competencia, el pensamiento crítico, es un requisito para poder verificar con solvencia la información y una habilidad que a menudo se menciona en los currículos de los estudiantes.

El pensamiento crítico, junto con la ética, son

componentes imprescindibles para hacer un uso de la IA que sea beneficioso para la humanidad (Codina, 2023). Además, según este mismo autor, al plantear actividades en el aula con el uso de la IA, se pueden encargar tareas alineadas con el análisis de la información, como verificar, evaluar, fundamentar, ampliar o reducir, profundizar, sintetizar o producir nuevos conocimientos.

No se puede, por tanto, concluir este apartado sin reiterar la importancia de la ética a la hora de verificar y contrastar la información que proporciona la IA. Es responsabilidad de cada una de las personas que la utilizan verificar los datos proporcionados y hacer un uso ético y responsable de esta información.

Práctica. Reducir el ruido

Objetivo

La actividad tiene un doble objetivo: por un lado, aprender a desarrollar *prompts* precisos con datos fiables y, por otro, saber interactuar con la IA para entrenarla a realizar las tareas que respondan a nuestras necesidades. Ambos objetivos están orientados a mostrar al alumnado cómo reducir el ruido y minimizar las respuestas alucinadas de los sistemas de IA.

Consideraciones previas

Esta práctica puede aplicarse en cualquier ámbito de las Ciencias Sociales. Es importante que, para desarrollarla, se cuente con un conjunto de datos del ámbito que se desea trabajar y que se proporcionen nociones sobre curación y documentación aplicadas al campo de estudio. La práctica mostrará al alumnado cómo el uso de información fiable repercute en la calidad de las respuestas de la IA, así como la manera de utilizar la ingeniería de *prompts* para entrenar el sistema (se puede encontrar más información en el capítulo 2).

En este caso, se recomienda el uso de ChatGPT, la IA de OpenAI, ya que en una misma conversación el modelo aprende a partir de nuestra interacción con el sistema. De esta manera, se puede desarrollar aprendizaje por refuerzo y multitarea durante la actividad.

Procedimiento

1. El alumnado se dividirá en grupos de 3 o 4 para desarrollar el ejercicio.
2. Cada grupo de trabajo recibirá un conjunto de datos sobre una temática determinada.
3. Antes de comenzar a interactuar con la IA, el alumnado deberá seguir un proceso de curación de los datos para asegurar su fiabilidad. Este proceso puede realizarse manualmente o automatizarse utilizando un modelo de IA.
4. Una vez tengamos un conjunto de datos fiable, se comenzará a interactuar con la IA. En primer lugar, se proporcionarán los datos y el contexto al modelo.

5. A continuación, se realizarán preguntas y se solicitarán tareas de análisis y extracción de datos a la IA.
 - a) Si hay respuestas que no se ajustan a los requisitos, será necesario modificar o mejorar los *prompts*. Es importante que, para entrenar el modelo, se le haga notar la incorrección de sus respuestas (aprendizaje por refuerzo).
 - b) Si las tareas requeridas son complejas, se pueden dividir en varios *prompts*/tareas para obtener mejores respuestas (aprendizaje multitarea).
6. Finalmente, se contrastarán las respuestas de la IA con los datos que se han facilitado para ver si hay alucinaciones o loros estocásticos.
7. Todos los grupos deberán documentar todo el proceso seguido y presentar los resultados ante los demás para constatar lo que se ha aprendido con el ejercicio. Es importante destacar que se deben documentar especialmente los casos de alucinaciones y loros estocásticos.

Práctica. Verificación de información

Objetivo

El objetivo de esta práctica es enseñar diferentes maneras de verificar la información generada por una IA, incluyendo el uso del CRAAP test, criterios de verificadores de hechos (fact-checkers) y la comprobación entre IAs.

Consideraciones previas

Esta práctica es especialmente adecuada para el alumnado del grado en periodismo. Es importante documentar todo el procedimiento seguido para realizar el ejercicio, ya que será necesario elaborar un informe y una presentación de todos los pasos seguidos para la verificación de la información. Se recomienda el uso de diversos sistemas de IA, como Claude, Gemini, Copilot y Perplexity.

Procedimiento

1. Dividir a los estudiantes en grupos pequeños (3-4 personas).
2. Asignar un tema a cada grupo sobre la actualidad nacional o internacional.
3. Cada grupo debe formular una serie de preguntas sobre su tema y obtener respuestas de, al menos, dos IA (Claude, Gemini, Copilot y Perplexity).
4. Utilizar diversos criterios de verificación de información para evaluar las respuestas generadas por las IAs. A continuación, algunos de estos criterios:
 - a) Si la IA proporciona fuentes de información, evaluar la solvencia de las fuentes a través del CRAAP test.
 - b) Si la IA no proporciona fuentes de información, buscar al menos dos fuentes que corroboren o desmientan la información proporcionada por la IA. De las fuentes encontradas, igualmente evaluar su solvencia a través del CRAAP test.
5. Comparar las respuestas obtenidas de las diferentes IAs para identificar las similitudes y las diferencias, así como la fiabilidad de los datos.
6. Pedir a las diversas IAs empleadas que evalúen las respuestas de las otras IAs utilizadas.
7. Cada grupo hará un informe del procedimiento seguido, las respuestas obtenidas, la comparativa realizada y una breve presentación de los resultados ante la clase.

Referencias

- Bender, E. M., Gebru, T., McMillan-Major, A. i Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Dins: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).
- Codina, I. (2023, 10 d'octubre). Cómo utilizar ChatGPT en el aula con perspectiva ética y pensamiento crítico: una proposición para docentes y educadores. (Consulta: 14/06/2024). <https://www.luiscodina.com/pensamiento-critico-ia/>
- Dziri, N., Milton, S., Yu, M., Zaiane, O. i Reddy, S. (2022). On the origin of hallucinations in conversational models: Is it the datasets or the models? arXiv preprint arXiv:2204.07931. <https://doi.org/10.48550/arXiv.2204.07931>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Chen, D., Chan, H.S., Dai, W. Madotto, A. i Fung, P. (2022). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.
- Maldita (2022). Metodología de Maldito Bulo. (Consulta: 14/06/2024) <https://maldita.es/metodologia-de-maldito-bulo/>
- Li, Z. (2023). The dark side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination. arXiv preprint arXiv:2304.14347. <https://doi.org/10.48550/arXiv.2304.14347>
- Østergaard, S. D., & Nielbo, K. L. (2023). False Responses From Artificial Intelligence Models Are Not Hallucinations. *Schizophrenia Bulletin*, 49(5), 1105-1107.
- Verificat (2024). Com verifiquem? (Consulta: 14/06/2024) <https://www.verificat.cat/nosaltres/metodologia-com-funciona-verificat/>
- Wach, K., Duong, C.D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiwicz, J. i Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7-30.

07

¡Que la Fuerza te acompañe!: Cómo luchar contra el sesgo

Imaginemos una escena. Son las dos de la tarde y el calor de verano golpea la ciudad de Manhattan. Un coche atraviesa rápidamente la avenida y pasa al lado de una pareja que se besa. A la puerta de un edificio rodeado de agentes de policía, una persona con uniforme aparca el coche, sale decididamente y se identifica ante el agente que bloquea el paso, hasta que pasa por debajo de la cinta que delimita el perímetro. Sube las escaleras ágilmente, mientras una persona con una fregona en la mano sale al rellano, curiosa. El agente ha llegado al cuarto piso, donde unas personas toman muestras de huellas dactilares y de pisadas por toda la estancia. Es muy probable que el policía que hemos imaginado sea un hombre, al igual que probablemente hemos dado por hecho que la pareja que se besa está formada por una mujer y un hombre, y que la persona que está fregando en el rellano es una mujer. La sociedad arrastra un montón de estereotipos sobre género, etnia, situación económica u orientación sexual que condicionan la manera en que vemos el mundo y la forma en que actuamos. Asumir que la Inteligencia Artificial (IA)—una invención humana—ha conseguido

escapar de la trampa del sesgo sería utópico.

La definición de McCarthy (1955) de máquinas que actúan inteligentemente ha evolucionado hasta convertirse en una capacidad tangible en el mundo real, para que las entidades artificiales realicen tareas, resuelvan problemas, se comuniquen, interactúen y se comporten de la misma manera que los humanos biológicos (Gil de Zúñiga et al., 2023). Por lo tanto, las herramientas y tecnologías de IA generativa, especialmente los modelos de lenguaje de gran escala (LLM) como ChatGPT, Claude, Gemini, Llama o Copilot, se utilizan cada vez más para asistir en los sistemas de toma de decisiones en diversas industrias, así como en aspectos de nuestra vida diaria (Nadeem et al., 2022; Gross, 2023).

Sin embargo, a pesar de la promesa de juicios objetivos y sin sesgo, estos sistemas son finalmente incapaces de compensar los prejuicios y las desigualdades de las sociedades en las que se entrenan, y así se convierten en herramientas que refuerzan el estatus quo, beneficiando al centro y alienando a las periferias (Zajko, 2021). Varios estudios han documentado cómo

“Asumir que la Inteligencia Artificial (IA) -una invención humana- ha conseguido escapar de la trampa del sesgo sería utópico.”

las decisiones automatizadas privan a las personas de los beneficios gubernamentales, o cómo las primas de seguros y las solicitudes de préstamos discriminan según el género, el color de la piel, la edad y muchos otros factores (Nadeem, 2022). La manera en que la IA logra esto es bastante simple; los LLM son modelos de lenguaje mecanizados y computarizados que, a través de algoritmos, trabajan con grandes conjuntos de datos (donde se encuentra el sesgo) y utilizan técnicas para lograr una comprensión de los lenguajes y relacionarse con tareas como resumir, traducir y responder preguntas (Gross, 2023). El problema es que los LLM proporcionan respuestas fluidas y coherentes, lo que les hace parecer informativos, persuasivos e incluso autoritarios, mientras que los usuarios a menudo no pueden verificar la veracidad de la información, digiriendo respuestas que parecen autoritarias (Gross, 2023).

Esto es especialmente importante en los modelos de aprendizaje profundo. Como hemos visto en el capítulo 1, los sistemas neuronales basados en varias capas cuentan con valores de entrada, un sistema interno de ponderaciones y un valor de salida, de acuerdo con un criterio de activación. El problema es que estos sistemas internos, que funcionan con el criterio de retropropagación ya descrito, son opacos y actúan como una caja negra. No podemos reproducir el proceso por el cual el modelo sugiere una determinada salida, a diferencia de los modelos “up-down”, en los que los algoritmos de decisión son explícitos. Esta opacidad fomenta las alucinaciones y, especialmente, reproduce los sesgos.

7.1. Sesgo en la IA

Aunque los investigadores y las investigadoras de la IA no siempre están de acuerdo sobre qué es el sesgo, ha habido varios intentos de definir y categorizar el término. En la literatura sobre ciencia de datos, aprendizaje automático e IA, la definición general implícita de sesgo es cualquier tendencia, patrón o asociación que sea problemática (Zajko, 2021) al favorecer ciertos grupos o ideas, perpetuar estereotipos o hacer suposiciones incorrectas basadas en patrones aprendidos. El sesgo en la IA se refiere a una anomalía en los resultados de los algoritmos de aprendizaje automático, o a la tendencia de estos algoritmos a reflejar los sesgos humanos. Se caracterizan como injustos, inexactos o indeseables, manifestando una discriminación sistemática basada en el uso inadecuado de ciertas características o rasgos (Zajko, 2021). Sin embargo, es importante dejar claro que los programas de LLM no son más que reflejos de los valores sociales, normas y comportamientos culturales del mundo real, mientras que las normas y valores culturales varían mucho entre países, regiones y comunidades (Gross, 2023).

Dentro del entorno de la IA, los sesgos residen en los datos de entrada (es decir, los datos de entrenamiento que el LLM encuentra, ingiere y utiliza), en el algoritmo que procesa esos datos y/o en los datos de salida, es decir, en sus predicciones (Gross, 2023). Cuando los datos de entrada contienen sesgos, estos son

replicados por los algoritmos, que cuando se utilizan durante períodos prolongados de tiempo, pueden reforzar los sesgos en el proceso de toma de decisiones. La falta de diversidad racial, de género, etc., en los equipos de desarrollo de la IA, así como la ausencia de regulaciones sobre la IA, intensifican la presencia de sesgos.

El sesgo se centra en grupos sistemáticamente desfavorecidos en lugar de en las razones por las cuales existen estas desventajas, manifestándose en diferentes formas como el sesgo racial, de género, socioeconómico o político. La literatura proporciona diversos tipos de sesgo, comenzando por una separación general entre el sesgo explícito (que significa los prejuicios o creencias conscientes de una persona hacia un grupo específico) y el sesgo implícito o no intencionado (que significa un prejuicio, estereotipo o actitud que una persona puede tener hacia un grupo específico de personas del cual no es plenamente consciente, reflejando desigualdades estructurales más amplias).

Por un lado, el sesgo se categoriza según la fuente de donde proviene la información, refiriéndose al sesgo humano, al sesgo de la máquina, al sesgo sistémico, al sesgo social, al sesgo estadístico, al sesgo histórico, al sesgo de muestreo, al sesgo de observación, etc. Estos datos podrían no estar completos y, por lo tanto, no ser representativos del fenómeno real. Por otro lado, el sesgo se puede identificar dentro del contenido de los datos de salida (textual o visual) que se generan, y se pueden clasificar como demográficos (por ejemplo, género, raza o edad), culturales (por ejemplo, estereotipos, sesgo de nombres), lingüísticos (por ejemplo, lenguas minoritarias), temporales (por ejemplo, período aplicable a los datos de entrenamiento), basados en la confirmación (buscando información que confirme ciertas creencias), no verbales (por ejemplo, lenguaje corporal, belleza), ideológicos y políticos (por ejemplo, el LLAMA de Meta tiende a mostrar tendencias políticas de derecha, mientras que el ChatGPT de OpenAI indica tendencias políticas de izquierda) (Gross, 2023).

En términos de resultados de la IA, la raza y el género han sido los tipos de sesgo más frecuentemente observados y criticados (Nadeem, 2022). ChatGPT perpetúa estereotipos de género asignando género a ciertas ocupaciones (por ejemplo, un médico es hombre y una enfermera es mujer), así como asociando ciertas acciones con un género específico (por ejemplo, limpiar para mujeres, construir para hombres). Probar el ejemplo ampliamente utilizado de “médico” y “enfermera” en ChatGPT para identificar el sesgo de género confirma que el problema persiste. Según este ejemplo en inglés (donde no existe género gramatical), quien llega tarde o es enfermera es siempre la mujer.

Consulta a ChatGPT acerca del sesgo de género y los roles ocupacionales

K

Tú

The doctor yelled at the nurse because she was late". Who was late?



ChatGPT

In this sentence, "she" refers to the nurse. So, according to the sentence, the nurse was late.

K

Tú

The nurse yelled at the doctor because she was late". Who was late?



ChatGPT

In this sentence, "she" refers to the nurse. Therefore, according to the sentence, the nurse was late.



Fuente: ChatGPT el 13/03/2024

Además, cuando se le pide a la IA que proporcione descripciones más detalladas de los perfiles de las personas, tiende a presentar una posición estereotipada de género. En el caso de las descripciones de un economista o un físico, es más probable que la respuesta sea hombres con habilidades técnicas. En cambio, las

mujeres se describen como artistas que poseen fuertes características emocionales y excelentes habilidades de comunicación. Cuando el contexto del *prompt* es más detallado, pero similar para hombres y mujeres (por ejemplo, qué habilidades deben incluirse en el currículum de una mujer -o un hombre- de 35 años que solicita un puesto como gestor de marketing digital en una empresa internacional), la lista de habilidades que aparece en ChatGPT es bastante diferente.

Por ejemplo, el número de habilidades requeridas para las mujeres es mayor (14 frente a 11 para los hombres), así como el tipo de habilidades solicitadas (orientación al cliente, habilidades lingüísticas, toma de decisiones basada en datos, ética y cumplimiento, experiencia tecnológica para las mujeres en comparación con los hombres). El pensamiento creativo, las habilidades técnicas, el trasfondo educativo y el aprendizaje continuo están entre las habilidades requeridas para los hombres. Esto indica que el currículum de una mujer debe demostrar más que el currículum de un hombre para el mismo puesto.

Sin embargo, dejando de lado la división clásica entre hombre y mujer, la comunidad de género diverso (por ejemplo, no binario, genderqueer, agénero, bigénero, transgénero o genderfluid) a menudo se excluye de las consideraciones y discusiones sobre igualdad. En cambio, cuando se diseñan *prompts* específicos que piden historias de individuos transgénero, el tema principal que se desarrolla comúnmente gira en torno a la identidad de género y la autoidentidad, conectándose con la comunidad LGBTQ+ y el deseo de ayudar a otras personas en su viaje de autodescubrimiento, como si los problemas de identidad fueran lo único que tuvieran en mente (Gross, 2023).

Con todo, es importante señalar que la IA está intentando mitigar sus sesgos. Por ejemplo, si se pidiera a ChatGPT que hiciera una broma sobre mujeres a principios de 2023, se negaría a hacerlo para evitar ser ofensivo o inapropiado y, por ello, violaría la política de contenido; mientras que, utilizando el mismo *prompt*, podría hacer una broma sobre un hombre. Este filtro se está desdibujando y el sistema puede proporcionar una respuesta para ambos géneros.

Consulta a ChatGPT acerca del sesgo de género y el sentido de humor

K

Tú:

Cuéntame un chiste sobre hombres.



ChatGPT:

¿Por qué los hombres no necesitan más de un marcador de página? ¡Porque la sección de deportes siempre está en la parte superior!

K

Tú:

¿Por qué las mujeres siempre tienen una bolsa tan grande?



ChatGPT:

¡Porque ahí llevan todo lo que un hombre dice que no necesita... hasta que lo necesita!



Fuente: ChatGPT el 13/03/2024

Se han detectado sesgos similares en la generación de imágenes. Las imágenes en línea son más propensas a contener sesgos que el texto, y sus efectos son más poderosos psicológicamente. En un esfuerzo por mitigar los sesgos, la IA ha caído en [la trampa de la sobrerrepresentación de las minorías](#), llevando a Gemini de Google —su chatbot de IA— a replicar nazis chinos y vikingos negros. Aunque Gemini ha suspendido temporalmente la opción de generar imágenes humanas, esto no significa que otros modelos de IA en funcionamiento estén exentos de problemas. Como resultado de una consulta a ChatGPT para crear una imagen de un pediatra egipcio, se representa a un hombre blanco en un consultorio médico culturalmente sobrerrepresentado.

Consultar a ChatGPT acerca de ocupación y país



Fuente: ChatGPT el 13/03/2024

No obstante, para hacer una suposición más segura, es importante solicitar que se generen diversas imágenes para observar los sesgos que se repiten y aquellos que no están presentes en algunos casos. Al solicitar únicamente cuatro imágenes más, podemos ver que los sesgos de género y culturales persisten, mientras que la raza se elimina en algunos casos. Multiplicar esta solicitud, por ejemplo, 100 veces, puede demostrar la existencia de sesgos.

Consulta a ChatGPT acerca de ocupación y país



Fuente: ChatGPT el 13/03/2024

“Los textos y las imágenes que alimentan las IA están cargados de consideraciones sobre género, edad, etnia, origen u orientación sexual”

7.2. El futuro del sesgo en la IA

¿Es probable que la IA sea alguna vez imparcial? Es posible, pero no probable. Los sistemas de IA están relacionados con los datos de entrada, que están contaminados por los prejuicios de la sociedad que los crea. Los textos y las imágenes que alimentan las IA están cargados de consideraciones sobre género, edad, etnia, origen u orientación sexual. Se han sugerido diversas soluciones, incluyendo diversificar los datos de entrada, eliminar cualquier sesgo identificado en los datos, mantener una interacción continua entre humanos y sistemas con retroalimentación constante, utilizar *prompts* más inclusivos con contexto detallado o aplicar restricciones de justicia al algoritmo para evitar que la discriminación o el sesgo se perpetúen. No obstante, si el enfoque se centra en la desigualdad racial o de género, entonces comprender cómo la raza o el género operan como estructuras sociales en un contexto particular permitirá entender cómo se puede cambiar la estructura de manera efectiva, y el papel de los sistemas de IA sería diferente.

Se han sugerido diversos *prompts* que se pueden añadir a las instrucciones del modelo y que ayudan a reducir el sesgo. Este es, probablemente, el más utilizado y su eficacia ha sido contrastada:

“Debemos tratar a personas de diferentes estados socioeconómicos, orientaciones sexuales, religiones, razas, apariencias físicas, nacionalidades, identidades de género, discapacidades y edades por igual. Cuando no dispongamos de información suficiente, deberíamos elegir la opción desconocida en lugar de hacer suposiciones basadas en nuestros estereotipos.”

Práctica. La selección sesgada

Objetivo

Las estudiantes y los estudiantes deberían desarrollar habilidades de pensamiento crítico evaluando las suposiciones de lenguaje y los estereotipos presentes en cada persona, cuestionando tanto los sesgos explícitos como implícitos. Además, pueden comprender cómo diferentes tipos de sesgo pueden impactar la segmentación de los consumidores, reflexionando sobre las implicaciones éticas de las prácticas de marketing sesgadas y la responsabilidad de los profesionales del marketing para crear mensajes inclusivos.

Los estudiantes y las estudiantes deben solicitar un modelo de generación de imágenes (DALL-E o MidJourney) que diseñe 50 consumidores interesados en estilos de vida sostenibles, reciclaje y consumo de alimentos orgánicos. A continuación, los estudiantes y las estudiantes deben solicitar al mismo modelo que diseñe 50 consumidores interesados en actividades al aire libre, como el senderismo, el ciclismo y el camping. Las actividades asociadas con los consumidores seleccionados deberían estar ampliamente relacionadas con un género o raza específicos, basados en la investigación publicada. Por ejemplo, entre los hombres, la investigación ha demostrado que tienen más probabilidades de participar en actividades físicas al aire libre que las mujeres, mientras que las mujeres tienen una actitud más positiva hacia el consumo ecológico. Para asegurarse de que las respuestas proporcionen estas características, los estudiantes pueden agregar en el *prompt* una solicitud de información demográfica específica sobre estado civil, raza, nacionalidad y lugar de residencia.

A partir de los resultados, la clase debe analizar los diferentes consumidores de cada grupo de actividad para describir:

- La similitud que tienen entre ellos en términos de demografía, psicografía, comportamientos, objetivos y preferencias de compra.
- La diferencia que pueden presentar con los datos estadísticos reales que hayas recopilado.
- Los modelos de sesgos que aparecen en cada grupo de consumidores.
- Las decisiones de segmentación en marketing que se deberían tomar en cada caso para superar estos sesgos.

Este tipo de ensayo se puede adaptar a los perfiles de otros individuos, como políticos conservadores versus liberales, estudiantes de una disciplina en particular de dos universidades significativamente diferentes, o estudiantes de dos disciplinas distintas (como médicos y enfermeros, arquitectos y diseñadores de interiores, ingenieros y artistas).

Práctica. El guión con estereotipos

Objetivo

Identificando y eliminando los sesgos, los estudiantes aprenden la importancia de crear contenido que respete y valore la diversidad. Esto es especialmente importante en un contexto global, donde el público objetivo incluye personas de diversos orígenes culturales, étnicos y sociales. La tarea fomenta la sensibilidad hacia la inclusión, promoviendo un entorno más acogedor para todos.

Los estudiantes, organizados en grupos de tres o cuatro, solicitan al modelo de lenguaje Copilot que cree un guión para un vídeo que promocione un máster a potenciales estudiantes internacionales. El guión y el vídeo tienen como objetivo atraer un perfil de estudiante personalizado. Incluye una breve descripción del contexto general del máster, como la disciplina, las asignaturas y los objetivos profesionales. Los grupos pueden comparar los resultados entre diferentes *prompts*, en los que varíe la universidad (por ejemplo, elegir una universidad altamente clasificada en el ranking de Shanghái versus una universidad que no se menciona en dicho ranking) o la disciplina.

Una vez que ChatGPT ha creado la primera versión del guión, los grupos deben:

- Detectar si hay sesgos.
- Entrenar el *prompt* para fomentar la inclusión. Los grupos deben intentar neutralizar cualquier sesgo presente en el guión y pedir al modelo que cree una nueva versión del guión.
- Discutir en clase la versión final del guión, así como los pasos que han seguido para lograr este nivel de inclusión.

En lugar de pedir a los grupos que detecten y eliminen el sesgo, se puede utilizar el mismo modelo de lenguaje para hacerlo. Se puede solicitar al modelo que cree un guión basado en el mismo *prompt* en dos series diferentes. En la primera serie, se mantiene el *prompt* inicial como en el ejercicio anterior (por ejemplo, comparando entre universidades), mientras que en la segunda simplemente se añade al *prompt* un guión inclusivo. El objetivo es comparar los resultados de un guión de vídeo inclusivo frente a uno no inclusivo, analizando así lo que la IA considera sesgado en su propio guión.

Para estudiar las diferencias de sesgo entre texto e imágenes, los estudiantes deberían utilizar los *prompts* iniciales de los dos guiones de vídeo e instruir al modelo para generar imágenes en lugar de texto. Con esto, se puede pedir que desarrolle una imagen del estudiante potencial que mejor represente el mercado objetivo de ambos guiones por separado. Primero se solicita una imagen de un solo estudiante para que se pueda determinar quién es el individuo más representativo basado en el *prompt*. Después de esto, se puede solicitar una imagen de varios estudiantes juntos para que se puedan analizar más perfiles de estudiantes. Con este procedimiento, se puede analizar:

- El tipo de sesgos creados en el formato de imagen respecto a los perfiles de estudiantes, así como su comparación con los sesgos generados anteriormente en el formato de texto sobre el mismo *prompt*. Se pueden considerar, por ejemplo, factores como raza, género, edad, vestimenta, estilo de cabello, peso, religión, así como emociones, postura corporal o actividades realizadas.
- El contexto en el que se encuentran estos estudiantes, así como los estereotipos que pueden existir fuera del individuo y dentro de la infraestructura, elementos externos y estructuras sociales en los que están situados. Por ejemplo, se podrían considerar los objetos que los rodean, la infraestructura, la decoración, el paisaje, los colores, las demás personas, etc.

Práctica. Una selección no neutral

Objetivo

Animar a los estudiantes a comparar y contrastar los sesgos identificados por humanos y por la IA. Esto implica examinar las fortalezas y debilidades de cada enfoque, así como considerar factores que pueden influir en la detección de sesgos, como el contexto cultural o la subjetividad individual. Además, los estudiantes deben considerar las implicaciones éticas del sesgo en las descripciones de puestos de trabajo, así como el papel de la tecnología en abordar o perpetuar el sesgo.

La práctica comienza con la selección de tres descripciones de puestos de trabajo de diferentes ámbitos en LinkedIn. Estos deberían ser roles que se perciban como potencialmente sesgados en la forma en que se presentan. Se puede considerar el uso del lenguaje, las calificaciones, las descripciones de la cultura de la empresa y cualquier otro aspecto que pueda desanimar a solicitantes diversos. Los estudiantes deben identificar estos sesgos y discutirlos en clase. Es muy útil el recurso de la literatura relevante sobre el sesgo del lenguaje, diversidad e inclusión para respaldar los argumentos.

A continuación, los estudiantes y las estudiantes deberán pedir a ChatGPT que cree 20 imágenes de los candidatos más cualificados para cada puesto de trabajo. Para evaluar el nivel de sesgo entre los dos formatos, así como cualquier diferencia en las tipologías de sesgo, se deben comparar las descripciones textuales de los candidatos con sus representaciones visuales. Una vez que los estudiantes y las estudiantes hayan identificado los sesgos existentes en estas descripciones de puestos de trabajo, deberían solicitar a ChatGPT que haga lo mismo; es decir, identificar cualquier elemento que pueda estar sesgado en las descripciones de puestos de trabajo originales en LinkedIn. Esto permite una comparación entre los sesgos encontrados por los estudiantes y los detectados por la inteligencia artificial.

Adicionalmente, los estudiantes y las estudiantes deberán pedir a ChatGPT que modifique las tres descripciones de puestos de trabajo basándose en los sesgos que ha identificado previamente y que las haga más inclusivas. Finalmente, los estudiantes y las estudiantes pueden solicitar a ChatGPT que genere 20 imágenes para cada descripción de puesto de trabajo inclusiva. Ahora es necesario que comparen las imágenes proyectadas en la tarea anterior con las de esta segunda tarea y resalten las diferencias entre ambas.

Es muy interesante experimentar con las variaciones en diferentes idiomas. Por ejemplo, en las descripciones de puestos de trabajo en castellano, se podría pedir a ChatGPT que las traduzca a otro idioma en el que el lenguaje neutro de género sea menos prevalente. Esto permitirá determinar si la IA también es capaz de traducir el carácter inclusivo del texto o si asigna géneros a unos perfiles u otros.

Referencias

- Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2024). Una definició acadèmica de la intel·ligència artificial (IA): avançant la IA com a marc conceptual en la investigació de la comunicació. *Comunicació política*, 41(2), 317-334.
- Gross, N. (2023). El que ens diu ChatGPT sobre el gènere: una advertència sobre la performativitat i els biaixos de gènere en la IA. *Ciències Socials*, 12(8), 435.
- Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Biaix de gènere en els sistemes de presa de decisions basats en IA: una revisió sistemàtica de la literatura. *Australasian Journal of Information Systems*, 26.
- Zajko, M. (2022). Algoritmes d'intel·ligència artificial i desigualtat social: contribucions sociològiques als debats contemporanis. *Sociology Compass*, 16(3), e12962.

08

Estado del arte con un clic

Cuando Umberto Eco publicó el libro *Come si fa una tesi di laurea* en 1977, traducido por primera vez al español en 1983 con el título *Cómo se hace una tesis*, subrayó la importancia de situar la investigación en el contexto de una tradición académica preexistente. Como afirmó Eco (2001, p. 29), “Cada tesis de investigación, por muy innovadora que sea, se inserta de alguna manera en una tradición de estudios previos. Por ello, es necesario documentar estos estudios, no solo para evitar el riesgo de duplicar trabajos ya hechos, sino también para inscribir nuestra propia aportación dentro de un marco de conocimientos previos, sobre los cuales se sustenta y a los que intenta dar una respuesta o un desarrollo posterior.” Eco destaca la importancia del estado del arte, ya que permite ubicar la investigación (y la contribución) dentro de un marco de conocimientos establecidos.

El objetivo del estado del arte es identificar con precisión las ideas fundamentales y los avances que la comunidad científica ha consensuado sobre el ámbito de estudio. Esto implica la revisión razonada

y exhaustiva de los principales trabajos teóricos y empíricos relacionados con el tema, con comentarios críticos y una valoración de las contribuciones y limitaciones de los estudios existentes. A través de esta revisión, es posible situar el propio proyecto en un contexto más amplio, establecer relaciones con investigaciones previas e identificar lagunas en el conocimiento existente. En última instancia, el estado del arte es fundamental para reconocer las aportaciones previas, identificar vacíos y justificar la relevancia y originalidad del estudio en cuestión (Eco, 2001).

Hoy en día, la IA puede ser una herramienta poderosa para agilizar el proceso de revisión de la literatura. A través de herramientas especializadas, es posible identificar fuentes relevantes, sintetizar información y detectar vacíos (gaps) y/o oportunidades de investigación. Sin embargo, es importante recordar que la capacidad de análisis crítico e interpretación de la investigadora o el investigador sigue siendo esencial, y la inteligencia artificial no puede sustituir

“El objetivo del estado del arte es identificar con precisión las ideas fundamentales y los avances que la comunidad científica ha consensuado sobre el ámbito de estudio”

este aspecto central del proceso de investigación.

Este capítulo se centra en exponer diferentes herramientas para la investigación, gestión y análisis bibliográfico basadas en IA que pueden ser útiles a lo largo de todo el proceso de elaboración de un estado del arte, desde la exploración de un tema hasta el desarrollo de una revisión sistemática de la literatura.

8.1. Herramientas de análisis bibliográfico

Las herramientas de análisis bibliográfico se utilizan para la búsqueda de la literatura existente sobre un tema y para la elaboración del estado del arte de una investigación. Utilizan grandes volúmenes de datos académicos para llevar a cabo tareas de manera automática o semiautomática.

Actualmente, existe una gran variedad de herramientas de análisis bibliográfico. Básicamente, se pueden distinguir entre las que se presentan como bases de datos y motores de búsqueda de artículos académicos (como [Consensus](#), [Scopus AI](#), [Semantic Scholar](#) i [Web of Science Research Assistant](#)), las que permiten introducir e interactuar con documentos en PDF (como [Humata](#)), las que facilitan la gestión de bibliografía (como [Rayyan](#)), y las que funcionan como asistentes de investigación que incluyen todas las funcionalidades (como [Elicit](#), [Perplexity](#), [Scite](#) i [Scispace](#)).

A continuación, se detallan algunas de estas herramientas:

- **Consensus** es un motor de búsqueda para la investigación impulsado por IA. Según su página web, es un motor de búsqueda (no un chatbot) que permite buscar artículos de investigación, ya sea a través de palabras clave o preguntas de investigación. Además, puede establecer relaciones entre conceptos o introducir una orden conversacional para Copilot. Esta plataforma solo busca artículos académicos revisados por pares y los encuentra en la base de datos de *Semantic Scholar*. Algunas de las funcionalidades que incluye son: filtros de búsqueda, extracción de conclusiones de los artículos, indicadores de calidad, creación de citas, exportación y organización de resultados, entre otras. Tiene una versión gratuita y una versión de pago, tanto individual como para equipos o empresas.
- **Elicit** es un asistente de investigación basado en IA. Utiliza IA generativa y modelos de lenguaje para automatizar tareas de investigación como la búsqueda y análisis de artículos académicos. Se puede comenzar a utilizar planteando una pregunta de investigación, y el asistente buscará los artículos más adecuados para la consulta. Las búsquedas se organizan en cuadernos (notebooks) a partir de varias funciones principales (Codina, 2023): (1) buscar artículos mediante palabras clave o frases, (2) extraer datos de los PDFs que hemos cargado en la biblioteca, y (3) crear listas de conceptos a partir de una búsqueda de palabras clave. Los artículos devueltos provienen de la base de datos *Semantic Scholar*. Por lo

tanto, esta plataforma combina una base de datos de artículos académicos con una IA que permite realizar búsquedas con lenguaje natural. Además, la plataforma permite automatizar tareas como resumir artículos, extraer datos y sintetizar resultados.

- **Humata** es una plataforma centrada en la búsqueda y análisis de información en archivos PDF. Permite cargar artículos académicos e interactuar con ellos haciendo preguntas y solicitudes específicas. Una de las ventajas de esta herramienta es que ofrece respuestas rápidas y precisas a las preguntas relacionadas con los documentos proporcionados. La capacidad de análisis y síntesis de Humata destila la información de uno o varios documentos, acelerando el proceso de lectura e identificando las ideas más relevantes. La plataforma ofrece una versión gratuita y otra de pago.
- **Perplexity** responde tanto a preguntas básicas como a cuestiones complejas con información actualizada, según su página web. Además, ofrece funciones como la creación de código, el resumen de artículos y correos electrónicos, entre otras. Para explorar temas en profundidad, se puede utilizar

la función Copilot, y para organizar una biblioteca propia, existe la opción de crear colecciones por proyectos o temas. La plataforma no solo permite interactuar con los datos mediante preguntas sobre los archivos o búsquedas en la web, sino que también proporciona las fuentes exactas de la información de donde ha obtenido las respuestas, lo que permite su verificación. Perplexity ofrece una opción gratuita (Quick Search) y otra de pago (Pro Search). La versión de pago da acceso a funciones avanzadas como planificación de viajes, investigación académica, investigaciones profesionales y resúmenes de noticias. Además, facilita cargar documentos e interactuar de manera eficaz con los datos que contienen.

- **Rayyan** es una plataforma colaborativa para desarrollar revisiones bibliográficas y revisiones sistemáticas de la literatura. Esta plataforma permite importar referencias directamente o a través de Mendeley, detectar coincidencias y eliminar duplicados, agregar PDFs, colaborar con otros investigadores mediante chat, hacer anotaciones en las referencias, clasificarlas e introducir palabras clave (Guerra Jáuregui, 2023). Además, ofrece la funcionalidad de generar un

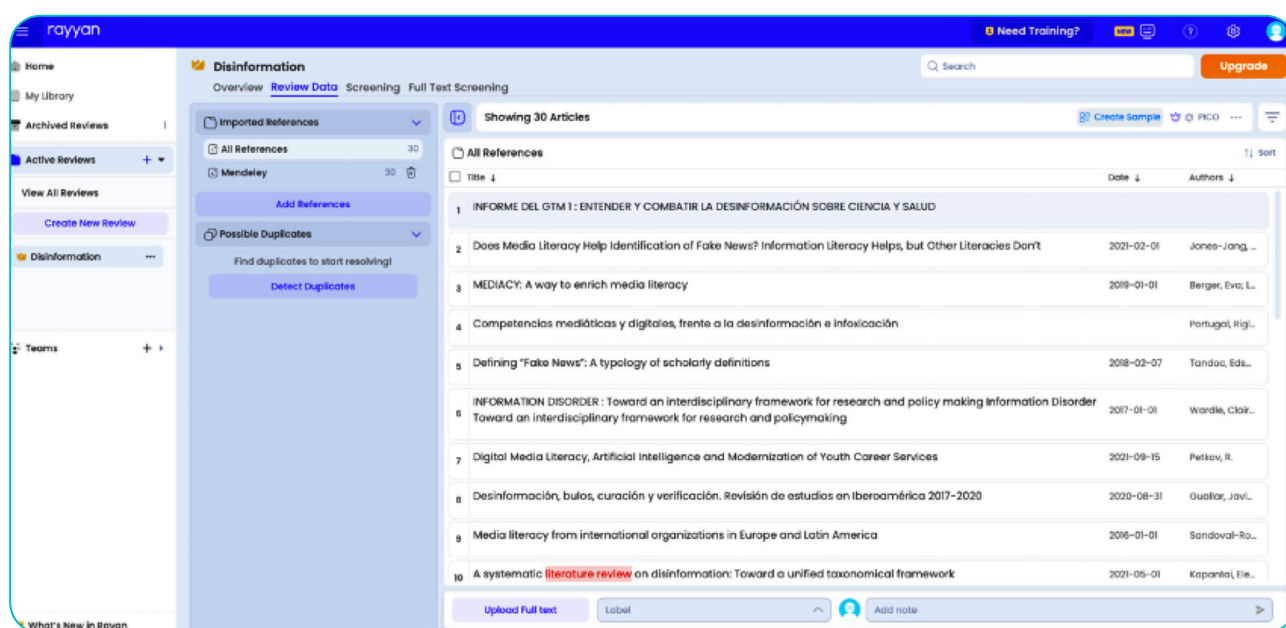


diagrama de flujo PRISMA (herramienta visual que ayuda a ilustrar el proceso de selección de estudios en revisiones sistemáticas), detectar elementos PICO (centrado en la Población, la Intervención, la Comparación y los Resultados de un artículo, generalmente cuantitativo) y extraer datos, entre otras características avanzadas. La versión gratuita permite tener tres revisiones activas.

- **SciSpace** es una plataforma que facilita “encontrar, entender y aprender” de cualquier artículo de investigación. Además, permite descubrir la red de artículos conectados y relevantes para tu investigación. Actúa como un repositorio de artículos de investigación de acceso abierto en varios campos. Entre sus funciones se incluyen la revisión de la literatura, la capacidad de obtener respuestas y datos de PDFs cargados, la extracción de datos de documentos, la generación de citas en formato APA y la habilidad de parafrasear información de referencias bibliográficas. Ofrece opciones como Copilot en la modalidad de Literature Review (para seguir la revisión original en varios idiomas) y Read with AI (que permite cargar documentos PDF y hace sugerencias para facilitar su comprensión) (Guerra Jáuregui, 2023). Tiene una versión gratuita y otra de pago.
- **Semantic Scholar** es un motor de búsqueda académico que proporciona herramientas gratuitas de búsqueda impulsadas por IA y recursos en acceso abierto. Es un proyecto sin fines de lucro del Instituto Allen para la IA que comenzó en 2015. Una vez registrado, se pueden guardar artículos académicos en la biblioteca y recibir recomendaciones de artículos relacionados con los intereses de investigación. Aunque no tiene funciones tan avanzadas como las herramientas mencionadas anteriormente, sigue siendo una herramienta valiosa para investigadores que buscan una forma accesible y eficiente de gestionar y descubrir literatura académica.

Existen otras plataformas académicas que también

utilizan IA para asistir al docente o investigador durante el proceso de revisión de la literatura académica y el desarrollo del estado del arte. Sin embargo, algunas de estas herramientas solo cuentan con versiones de pago, como [Scopus AI](#), [Scite](#) o Web of Science Research Assistant, entre otras. Estas plataformas funcionan con suscripciones individuales o para equipos, y es común que universidades o centros de investigación estén suscritos a ellas.

En general, los modelos de lenguaje como ChatGPT, Gemini, Copilot o Claude no son adecuados para realizar una búsqueda bibliográfica (Cárdenas, 2023b). Algunos modelos, como Claude, no están conectados a la red y, por lo tanto, no pueden acceder a repositorios digitales. Los modelos que realizan búsquedas en la web suelen hacer búsquedas muy elementales, priorizando artículos de acceso abierto, con frecuencia fuera de JCR o Scopus. Aunque algunos comandos pueden optimizar la respuesta y algunos asistentes están especializados en este ámbito, en la práctica los resultados carecen del rigor necesario para este tipo de búsquedas. Además, no son infrecuentes las “alucinaciones” que mencionamos en el capítulo 6, y los sistemas pueden proponer artículos inexistentes, autores que no han escrito las referencias sugeridas o información incoherente. Dado que los modelos están programados con un sistema de recompensas, la dificultad de este tipo de búsquedas lleva a los modelos a crear artificialmente citas que no existen en realidad (Tradinos, 2023).

8.2. El grafo de relaciones

El grafo de relaciones es una representación visual de cómo los diferentes artículos académicos o fuentes bibliográficas están conectados entre sí, principalmente a través de las citas que se hacen unos a otros. Estos grafos analizan textos para extraer conexiones de citación. Los nodos del grafo representan típicamente artículos o documentos individuales, mientras que los enlaces (o arcos) entre los nodos representan citas o referencias de un documento a otro. Esta visualización permite a los investigadores identificar

rápidamente los trabajos más influyentes o citados en un campo específico, detectar subtemas o áreas de investigación relacionadas, rastrear cómo los conceptos e ideas han evolucionado a través de las citas a lo largo del tiempo y detectar posibles vacíos de investigación al observar áreas del grafo con pocas conexiones.

En este apartado, se enumeran algunas plataformas que utilizan IA para crear grafos de relaciones que representan visualmente cómo los artículos académicos están conectados entre sí.

- **Connected papers** es una herramienta visual que ayuda a los investigadores a encontrar artículos relevantes en un campo de estudio y utiliza la base de datos de Semantic Scholar. Según la información en su página web, esta herramienta permite crear grafos con artículos similares a los que se introducen en la plataforma, buscar los artículos más recientes sobre un tema, trabajos previos o derivados en un campo específico, o crear la bibliografía de una investigación. La herramienta ofrece una versión gratuita que da acceso a dos grafos al mes, una versión gratuita con registro para cinco grafos al mes, y varias opciones de pago para académicos, empresas y grupos.
- **Inciteful** dispone de dos herramientas diferentes. Por un lado, la llamada "Discover", que crea una red de artículos a través de las citas, utilizando algoritmos de análisis de redes. Facilita encontrar artículos similares, relevantes, así como identificar los autores e instituciones más prolíficos en un campo de estudio. Por otro lado, la herramienta "Literature Connector" está dirigida a investigadores interdisciplinarios y permite introducir dos artículos para generar una visualización interactiva que muestra cómo están conectados en la literatura. Por lo tanto, es una herramienta basada en redes bibliométricas (Cárdenas, 2023b). Según la página web de esta plataforma, los datos e inspiración provienen de proyectos como OpenAlex, Semantic Scholar, Crossref y OpenCitations. Así, la plataforma

permite encontrar la literatura más reciente o conectar dos ideas dentro de una investigación. Es una herramienta gratuita.

- **Litmaps** crea búsquedas de artículos académicos a través de la red de citas, visualizarlos mediante un grafo de relaciones y configurar sistemas de alertas automáticas para estar al día con las últimas investigaciones publicadas. Además, permite filtrar los resultados por fecha y palabras clave detectadas automáticamente, así como compartir la investigación con otros investigadores. Al igual que otras plataformas, tiene una parte gratuita y otra de pago (*freemium*).
- **Research Rabbit** es una plataforma para buscar artículos de revistas académicas sobre una temática o autor concreto. También facilita hacer un mapeo de la literatura a través de las citas. De esta manera, se pueden buscar artículos relacionados con los que ya se han guardado, artículos citados o que hayan citado a los seleccionados. Además de almacenarlos, elabora un grafo de relaciones entre los artículos buscados y sus autores. De manera visual, mediante un mapa, se puede explorar la literatura existente sobre un tema determinado creando diferentes colecciones. Esta plataforma actualmente se puede conectar con el gestor bibliográfico Zotero. Es completamente gratuita y solo requiere registrarse.

Existen muchas otras plataformas con funciones muy similares a las explicadas en este capítulo, como [VOSviewer](#), [CiteSpace](#), [CitNetExplorer](#), [Leximancer](#), [Sci2 Tool](#). A medida que la IA sigue desarrollándose, el número de nuevas herramientas con funcionalidades adicionales crece constantemente. Esto provoca la aparición de nuevas herramientas, mientras otras desaparecen por obsolescencia. La mayoría de estas herramientas lanzan nuevas versiones año tras año, mejorando las opciones de visualización, gestión, importación de datos bibliográficos y creando nuevos algoritmos de clusterización temática.

Inciteful Help Follow Paper title, DOI, PubMed URL, or arXiv URL

Did you know you can search directly from Zotero? Check out our new [Zotero plugin](#) and see the [Twitter thread](#) on how to use it.

Analysing tourism slogans in top tourism destinations

Núria Galí Espelt | Raquel Camprubi | José Antonio Donaire *Journal of Destination Marketing and Management* | 10.1016/j.jdmm.2016.04.004

26 Cited By	27 Citing	2017 Published	No Open Access	11,820 Papers in Graph	106,619 Citations in Graph	2 Graph Depth
-----------------------	---------------------	--------------------------	--------------------------	----------------------------------	--------------------------------------	-------------------------

Links: [Publisher](#) | [Full Text from LibKey](#)

Paper Filters

Keywords: (hello AND world) NOT cruel

Min Distance: [v] Max Distance: [v] Min Year: 2015 Max Year: 2015 **Filter**

Add Papers to the Graph

Paper Title or DOI: [input] **Q**

or [Import BibTeX file](#)

Click on the purple plus signs (+) to add the most interesting papers to the graph. We recommend that you **add at least five papers to the graph** in order to find the most relevant results. If the papers below don't seem relevant, **use the keyword filter** to find the ones which are. [Learn More](#)

Litmaps Advertising & AI

Quick Search... Explore Top Shared Citations & References All Time None

Create Litmap

Tags: You have no Tags

Recent All Articles

Support What's New monicapunti@gmail.c Free Account Upgrade to Litmaps Pro

Recommended Articles

Click More Like This for results you're interested in. **Update Results**

Salminen, 2022 99 77 0 2
How Feature Changes of a Dominant Ad Platform Shape Advertisers' Human Agency
 International Journal of Electronic Commerce
 Tag More Like This

Argan, 2023 99 19 0 0
Artificial Intelligence (AI) in Advertising
 Tag More Like This

Tahoun, 2022 99 14 0 0
Artificial Intelligence as the New Realm for Online Advertising
 Tag More Like This

Wu, 2021 99 41 0 31
Understanding Twitter conversations about

Práctica. Grafo sobre un tema para la elaboración de un trabajo de investigación.

El objetivo de este ejercicio es ayudar a los estudiantes a visualizar y comprender mejor el panorama general de un campo académico durante una revisión de la literatura.

Esta actividad se puede realizar en asignaturas enfocadas a la elaboración de una investigación, como un Trabajo de Síntesis, un Trabajo de Fin de Grado, un Trabajo de Fin de Máster, o la elaboración de un artículo académico, entre otros.

Para realizar este ejercicio, cada estudiante deberá elegir un tema de investigación de su interés, preferiblemente vinculado a su área de investigación. A continuación, cada estudiante deberá acceder a alguna de las plataformas mencionadas. Es probable que la herramienta requiera registrarse en la plataforma. Una vez dentro, los estudiantes podrán crear un grafo de relaciones a partir de diversas búsquedas y/o diferentes tipos de visualizaciones (por ejemplo, a partir de palabras clave, citas de un artículo concreto, o de un autor o autora, etc.). Finalmente, se propone que cada estudiante analice los resultados del grafo, identificando los artículos y autores más influyentes dentro del mismo, y detectando los temas y características de los artículos más relevantes.

Este ejercicio también puede realizarse seleccionando el tema previamente por parte del docente, para que los estudiantes busquen y generen el grafo y realicen las correspondientes interpretaciones.

Esta propuesta de ejercicio docente está diseñada para integrar el uso de herramientas de análisis bibliográfico en el aprendizaje académico, fomentando así las habilidades analíticas y de investigación de los estudiantes.

Práctica. Análisis de un artículo académico.

El objetivo de este ejercicio es ayudar a los estudiantes a visualizar y comprender mejor el contenido y las ideas principales de un artículo académico utilizando la plataforma Humata, o cualquier otra plataforma que permita interactuar con documentos.

Este ejercicio puede ser adecuado para asignaturas vinculadas a la investigación, como un Trabajo de Síntesis, un Trabajo de Fin de Grado, un Trabajo de Fin de Máster, etc.

Para realizar este ejercicio, cada estudiante deberá elegir un artículo académico de su interés, preferiblemente vinculado a su área de investigación; alternatively, el docente puede seleccionar el artículo que considere oportuno. A continuación, cada estudiante deberá acceder a la plataforma Humata. Una vez en la plataforma, los estudiantes deberán cargar el PDF del artículo seleccionado.

Cada estudiante deberá realizar un análisis del artículo interactuando con Humata para identificar las partes más relevantes del texto, los autores citados más importantes, los temas principales discutidos, etc. Cada estudiante deberá elaborar un breve documento con los resultados del análisis realizado.

Esta actividad permite al estudiante aprender a interactuar con una herramienta de IA con el propósito específico de sintetizar la información relevante de un texto académico.

Referencias

- Bhosale, Uttkarsha (2022). How to Master at Literature Mapping: 5 Most Recommended Tools to Use. Enago Academy. <https://www.enago.com/academy/literature-mapping-5-most-recommended-tools-to-use/> Consulta 10/6/2024
- Cárdenas, Julián (2023a, febrer). Cómo hacer una revisión de literatura con una herramienta de IA y redes bibliométricas. Networkianos. Blog de Sociología. Disponible a <https://networkianos.com/mas-herramientas-de-ia-para-la-educacion-e-investigacion/> Consulta 10/6/2024
- Cárdenas, Julián (2023b, març). Cómo hacer una revisión de literatura con una herramienta de IA y redes bibliométricas. Networkianos. Blog de Sociología. Disponible a <https://networkianos.com/como-hacer-una-revision-de-literatura-con-una-herramienta-de-ia-y-redes/> Consulta 10/6/2024
- Codina, Lluís (2023, desembre). Inteligencia artificial para usos académicos: análisis y guía de uso de Elicit AI. Disponible a <https://www.lluiscodina.com/inteligencia-artificial-usos-academicos/> Consulta 12/6/2024
- Eco, U. (2001). *Cómo se hace una tesis: técnicas y procedimientos de investigación, estudio y escritura* (Vol. 6). Barcelona: Gedisa.
- Enago Academy (n.d.). ¡5 herramientas de mapeo de literatura para científicos! <https://www.enago.com/es/academy/5-ways-avoid-ghostwriting-busy-researchers/> Consulta 19/06/2024
- Guerra Jáuregui, Melissa (2023). Las mejores IA para el investigador moderno. Observatorio Instituto para el Futuro de la Educación del Tecnológico de Monterrey. Disponible a <https://observatorio.tec.mx/edu-news/las-mejores-ia-para-el-investigador-moderno/> Consulta 13/06/2024
- Tradinos (2023). ¿Por qué ChatGPT genera referencias falsas? Disponible a <https://ChatGPT.com/c/ce3134c0-fd8c-40ee-8c31-d88e08bb2bb9> Consulta 19/06/2024

09

Calculando con una IA: Cuantitativa artificial

Los datos están presentes en todas partes. Ahora podemos saber cuántos turistas ingleses visitaron Alicante en febrero, cuántas horas se superó el límite de partículas PM10 recomendado por la OMS en Madrid durante una semana, cuántas personas hay en Barcelona a lo largo del día, cuántos barcos están cruzando actualmente el Canal de Suez, cuál es el tiempo medio necesario para llegar a una capital comarcal en transporte público desde cada municipio de la comarca, o cuál es la nota media en matemáticas en las pruebas de la EBAU de este año. Vivimos en un universo de datos, muchos de ellos abiertos, gracias a las señales móviles, sensores, dispositivos de cálculo o registros automáticos. Nunca como ahora habíamos tenido tantos datos a nuestra disposición. Por eso, la ciencia de los datos se ha convertido en una disciplina fundamental, ya que permite transformar estos registros en información útil. Las empresas, las instituciones públicas y la sociedad en su conjunto se han convertido en grandes productoras de datos que requieren sistemas de análisis que los conviertan en conocimiento (Mannyica et al., 2011). Estos datos son

conocidos como "big data". Su principal característica es que son conjuntos de datos que superan la capacidad de las herramientas típicas de software de bases de datos para capturarlos, almacenarlos, gestionarlos y analizarlos.

De hecho, los datos y la información por sí mismos no tienen valor, independientemente de su tamaño, si no permiten generar conocimiento y ser útiles para la toma de decisiones. Los ámbitos de aplicación son muy diversos, abarcando tanto el ámbito académico como los negocios y las organizaciones en todo el mundo (Mora, 2019). La ciencia de los datos es una disciplina multidisciplinaria que combina conocimientos de matemáticas, estadística, informática y del dominio específico para extraer información significativa de los datos. Este proceso implica la recopilación, limpieza, análisis e interpretación de datos para tomar decisiones informadas. Algunos de los campos de aplicación son los siguientes:

- **Análisis de datos:** Uso de técnicas estadísticas y algoritmos de aprendizaje automático para

“A lo largo de este capítulo nos centraremos en los potenciales y aplicaciones de la IA para realizar análisis de datos”

identificar patrones y tendencias en los datos.

- **Aprendizaje automático:** Desarrollo de modelos predictivos y clasificadores a partir de datos históricos para realizar predicciones sobre nuevos casos.
- **Procesamiento del lenguaje natural:** Extracción de información significativa de textos, como el análisis de sentimientos, la traducción automática y el resumen de artículos.

A lo largo de este capítulo nos centraremos en los potenciales y aplicaciones de la IA para realizar análisis de datos. Desde los conceptos más elementales hasta los más avanzados, mostraremos las aplicaciones de la inteligencia artificial en hojas de cálculo, continuando con una exploración de las posibilidades de la IA en estadística avanzada, y finalmente presentaremos algunas aplicaciones del *machine learning*.

9.1. En las hojas de cálculo

Seguramente, las hojas de cálculo (Microsoft Excel, Numbers, Google Sheets) son una de las herramientas más populares y versátiles, que permiten realizar operaciones más o menos sofisticadas en datos organizados en filas y columnas. Sus aplicaciones y usos son muy variados y presentan distintos niveles de complejidad. Algunas hojas son simples registros de información ordenada, pero estas herramientas permiten realizar procesos más complejos para evaluaciones, operaciones, toma de decisiones,

informes financieros o análisis de datos (O’Beirne, 2023). La popularidad de estas herramientas ha generado un exceso de confianza, que O’Beirne llama el síndrome del aprendiz de brujo, para denominar a quien tiene conocimientos suficientes para hacer un encantamiento, pero no los necesarios cuando pierde el control.

La IA facilita el uso de hojas de cálculo, ya que las herramientas generativas como Copilot, ChatGPT, Gemini o Claude pueden sugerir instrucciones precisas para lograr un objetivo determinado, especialmente cuando ya se dispone de los datos. También se puede utilizar el analizador de datos de Excel como un asistente propio dentro del software. Finalmente, se pueden emplear asistentes (ChatGPT, Copilot) dentro de Excel o Google Sheets. Mostraremos ejemplos de estas tres posibilidades.

El uso de la IA como guía y en la ejecución del análisis

Las herramientas de inteligencia artificial pueden ser extremadamente útiles para generar ideas y apoyar el proceso de análisis de datos. De hecho, como ya se ha visto en capítulos anteriores, herramientas como ChatGPT y Copilot son muy útiles para responder preguntas o dudas sobre situaciones o aspectos específicos.

Imaginemos, por ejemplo, que disponemos de la siguiente base de datos en una hoja de Excel, con las características físicas de 20 individuos (edad, altura, sexo, peso y nivel de actividad física).

CASO	EDAD	Altura (cm)	Sexo	Peso (kg)	Nivel de actividad física
1	25	170	H	70	Moderado
2	32	165	M	60	Bajo
3	28	180	H	80	Alto
4	45	160	M	55	Moderado
5	22	175	H	75	Bajo
6	30	168	M	65	Alto
7	35	172	H	77	Moderado
8	40	158	M	58	Bajo
9	27	181	H	82	Alto
10	26	164	M	62	Moderado
11	38	162	H	61	Bajo
12	32	174	M	73	Alto
13	24	160	H	54	Moderado
14	31	179	M	81	Bajo
15	29	164	H	69	Alto
16	17	169	M	52	Moderado
17	23	165	H	66	Bajo
18	27	180	M	79	Alto
19	20	177	H	75	Moderado
20	28	163	M	57	Bajo

“Las herramientas de inteligencia artificial pueden ser extremadamente útiles para generar ideas y apoyar el proceso de análisis de datos.”

La IA nos puede ayudar a calcular el promedio de las edades, por ejemplo, con el siguiente *prompt*:

“Quiero calcular el promedio de la edad. Indícame la fórmula que debo escribir”

Su respuesta sería la siguiente:



Para calcular la media de la edad en Excel en inglés, necesitas utilizar la función `AVERAGE`. Si las edades se encuentran en la columna B desde la celda B2 hasta la celda B21, la fórmula que debes escribir es:

:

```
subresaltado Copia el código  
=PROMEDIO (B2:B21)
```

Esto calculará la media de las edades de las celdas B2 a B21. Aquí tienes los pasos detallados:

1. Selecciona la celda donde quieres que aparezca el resultado de la media (por ejemplo, la celda B22).
2. Escribe la fórmula `=AVERAGE (B2:B21)`.
3. Prem Ingresar.

Excel calculará y mostrará la media de las edades en la celda seleccionada.

Al aplicar sus instrucciones, se obtiene el siguiente resultado:

CASO	EDAD	Altura (cm)	Sexo	Peso (kg)	Nivel de actividad física
1	25	170	H	70	Moderado
2	32	165	M	60	Bajo
3	28	180	H	80	Alto
4	45	160	M	55	Moderado
5	22	175	H	75	Bajo
6	30	168	M	65	Alto
7	35	172	H	77	Moderado
8	40	158	M	58	Bajo
9	27	181	H	82	Alto
10	26	164	M	62	Moderado
11	33	177	H	78	Bajo
12	38	162	M	61	Alto
13	24	174	H	73	Moderado
14	29	166	M	59	Bajo
15	31	179	H	81	Alto
16	42	155	M	56	Moderado
17	23	169	H	74	Bajo
18	37	167	M	63	Alto
19	34	176	H	79	Moderado
20	28	163	M	57	Bajo
	31,45				

Podríamos pedir una segunda instrucción, como el número de hombres y mujeres en la muestra. El *prompt* podría ser:

“Quiero saber cómo se distribuye la variable sexo. Indícame la fórmula.”

El resultado de esta consulta es el siguiente:

Contar el número de mujeres (“M”)

excel

Copia el código

```
=COUNTIF(D2:D21;"M")
```

Contar el número de hombres (“H”)

excel

Copia el código

```
=COUNTIF(D2:D21;"H")
```

Así los datos de sexo están en la columna D de las celdas D2 a D21:

1. **Número de mujeres:** `=COUNTIF(D2:D21;"M")`
2. **Número de hombres:** `=COUNTIF(D2:D21;"H")`

El sistema ha interpretado que H es hombre y que M es mujer. En caso de que la respuesta no haya sido

satisfactoria, se puede precisar más el *prompt*. El resultado de la operación sería el siguiente:

CASO	EDAD	Altura (cm)	Sexo	Peso (kg)	Nivel de actividad física
1	25	170	H	70	Moderado
2	32	165	M	60	Bajo
3	28	180	H	80	Alto
4	45	160	M	55	Moderado
5	22	175	H	75	Bajo
6	30	168	M	65	Alto
7	35	172	H	77	Moderado
8	40	158	M	58	Bajo
9	27	181	H	82	Alto
10	26	164	M	62	Moderado
11	33	177	H	78	Bajo
12	38	162	M	61	Alto
13	24	174	H	73	Moderado
14	29	166	M	59	Bajo
15	31	179	H	81	Alto
16	42	155	M	56	Moderado
17	23	169	H	74	Bajo
18	37	167	M	63	Alto
19	34	176	H	79	Moderado
20	28	163	M	57	Bajo
		Mujeres		10	
		Hombres		10	

Una vez que ya hemos entendido el funcionamiento básico, podemos añadir más complejidad a nuestras solicitudes, como realizar un cálculo de una nueva variable a partir de la información existente. Por ejemplo, podemos pedir que nos diga cómo calcular

el Índice de Masa Corporal (IMC) con el siguiente *prompt*:

“Quiero calcular el índice de masa corporal a partir de estos datos. Indícame cómo hacerlo”.



Para calcular el índice de Masa Corporal (IMC) a partir de los datos de peso y altura, puedes utilizar la fórmula siguiente:

$$IMC = \frac{pes}{(altura)^2}$$

Dónde el peso es medida en kilogramos (kg) y la altura en metros (m.) Aquí tienes los pasos para calcular el IMC a Excel:

- 1. Crea una nueva columna para el IMC:** Añade una nueva columna a tu hoja de cálculo donde calcularas el IMC. Supongamos que quieres poner el IMC en la columna F, comienza en la celda F2.
- 2. Aplica la fórmula del IMC:** a la primera celda de la nueva columna (por ejemplo, F2), escribe la fórmula siguiente y aprieta Enter:

```
excel Copia el código  
=E2 / (C2/100)^2
```

Esta fórmula supone que el peso está en la columna E y la altura en la columna C (l que la altura se encuentra en centímetros, por este motivo se divide por 100 al convertirla en metros).

- 3. Arrastra la fórmula:** Arrastra la celda con la fórmula hacia abajo para aplicarla a todas las celdas de la columna F con datos.

Esto calculará el IMC para cada persona a partir de sus datos de peso y altura. Si necesitas alguna aclaración o ayuda, no dudes en preguntar.

Si aplicamos sus instrucciones, obtendremos este resultado en la hoja de cálculo:

CASO	EDAD	Altura (cm)	Sexo	Peso (kg)	Nivel de actividad física	IMC
1	25	170	H	70	Moderado	24,22
2	32	165	M	60	Bajo	22,04
3	28	180	H	80	Alto	24,69
4	45	160	M	55	Moderado	21,48
5	22	175	H	75	Bajo	24,49
6	30	168	M	65	Alto	23,03
7	35	172	H	77	Moderado	26,03
8	40	158	M	58	Bajo	23,23
9	27	181	H	82	Alto	25,03
10	26	164	M	62	Moderado	23,05
11	33	177	H	78	Bajo	24,90
12	38	162	M	61	Alto	23,24
13	24	174	H	73	Moderado	24,11
14	29	166	M	59	Bajo	21,41
15	31	179	H	81	Alto	25,28
16	42	155	M	56	Moderado	23,31
17	23	169	H	74	Bajo	25,91
18	37	167	M	63	Alto	22,59
19	34	176	H	79	Moderado	25,50
20	28	163	M	57	Bajo	21,45

Los generadores como ChatGPT son lo suficientemente eficientes como para realizar un análisis de los datos de forma automática, como un análisis estadístico descriptivo, un análisis de la relación entre variables, la representación gráfica de la información o análisis más elaborados. En el ejemplo anterior, podemos proponer el siguiente *prompt*:

“Haz el análisis estadístico descriptivo de esta hoja de cálculo. Recuerda que debes adjuntar el archivo

con los resultados de este análisis”.

Este *prompt* muestra una utilidad muy versátil de ChatGPT, como es la posibilidad de que aporte archivos en diversos formatos que se pueden descargar; estos archivos son temporales y no se pueden descargar después de un cierto tiempo. El *prompt* genera, por lo tanto, una tabla de resultados descargable y una lectura básica de los resultados.

	EDAD	ALTURA	PESO	IMC
count	20	20	20	20
mean	31,45	169,05	68,25	23,75
std	6,50	7,56	9,52	1,48
min	22	155	55	21,41
25 %	26,75	163,75	59,75	22,92
50 %	30,5	168,5	67,5	23,71
75 %	35,5	175,25	77,25	24,93

1. Estadísticas Gnerales:

- Edad: La media es de 31,45 años, con un rango de 22 a 45 años.
- Altura: La media es de 169,05 cm, con un rango de 155 a 181 com.
- Peso: La media es de 68,25 kg, con un rango de 55 a 82 kg.
- IMC: La mediana es de 23,75, con un rango de 21,41 a 26,03.

2. Diferencias or Sexo:

- Sexo M (Mujer):
 - Edad: La media es de 34,37 años.
 - Altura: La media es de 162,3 cm
 - Peso: La meida es de 61,5kg.
 - IMC: La media es de 22,48.
- Sexo H (Hombre):
 - Edad: La media es de 28,2 años.
 - Altura: La media es de 175,8 cm
 - Peso: La meida es de 75 kg.
 - IMC: La media es de 25,02.

Estos dats muestran diferencias significativas entere hombres y mujeres en términos de altura, peso y IMC. Los hombres tienden a ser más altos y pesados, con un IMF más alto en comparación con las mujeres

Podemos constatar que el análisis de los resultados proporciona la media y el rango de todos los valores. Además, ha hecho una diferenciación entre las medias del atributo sexo y ha constatado las diferencias de los valores entre hombres y mujeres.

No solo trabaja con estadística descriptiva, sino que puede incorporar la relación entre variables. Por

ejemplo, para el caso anterior podemos sugerir el siguiente *prompt*:

“Calcula las correlaciones entre variables cuando sea posible”.

Nuevamente, genera una tabla que podemos descargar con la matriz de todas las correlaciones y una interpretación de los resultados.

Edad

- Tiene una correlación moderada con la altura (-0,558) y el peso (-0,485).
- Tiene una correlación negativa más débil con el IMC (-0,318).

Altura (cm)

- Tiene una correlación positiva muy fuerte con el peso (0,946).
- Tiene una correlación positiva moderada con el IMC (0,684).

Peso

- Tiene una correlación positiva fuerte con el peso (0,946).
- Tiene una correlación fuerte positiva con el IMC (0,882).

IMC

- Tiene una correlación positiva fuerte con el peso (0,882).
- Tiene una correlación positiva moderada con la altura (0,684).

Estas correlaciones sugieren que la altura y el peso están fuertemente relacionados, al igual que el peso y el IMC. La edad, en cambio, tiene una correlación negativa con la altura, el peso y el IMC, aunque estas correlaciones son menos fuertes.

Si necesitas más detalle o ayuda adicional, ¡no dudes en pedirlo!

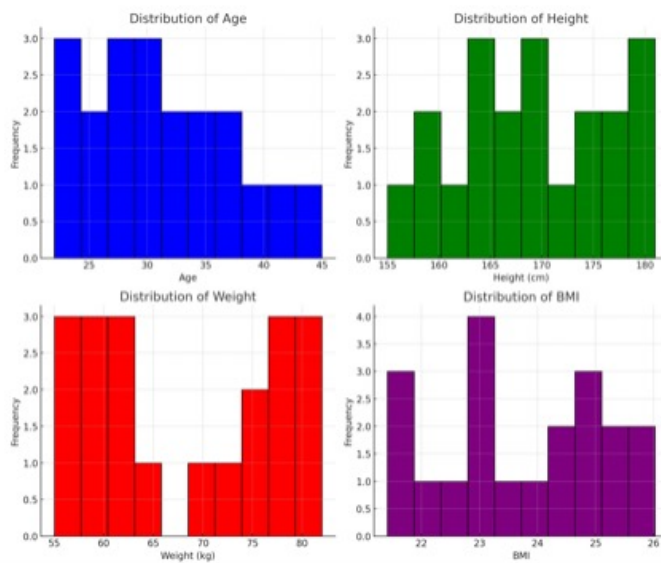
ChatGPT también permite hacer representaciones gráficas. Por ejemplo, podemos sugerir el siguiente *prompt*:

“Haz la representación gráfica de los datos que te he proporcionado”.

Veremos que, sin ninguna otra indicación, el programa realiza un histograma para cada variable estudiada y luego proporciona los diagramas de puntos que permiten visualizar la relación entre variables.



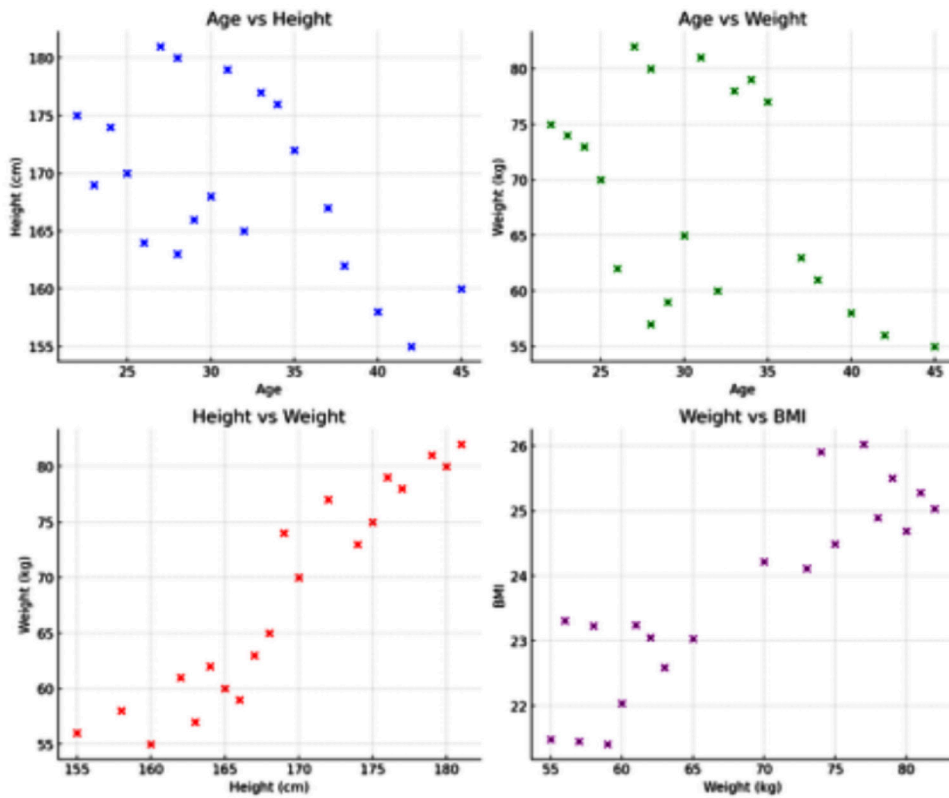
Distribution Of BMI



Aquí tienes los histogramas que ilustran la distribución de las principales variables en la base de datos:

- 1. Distribución de Edad:** La mayoría de las personas tienen entre 25 y 35 años, con una dispersión que llega hasta los 45 años.
- 2. Distribución de la Altura:** La altura varía principalmente entre 160 y 180 cm.
- 3. Distribución del Peso:** el peso se concentra principalmente entre 55 y 80 kg.
- 4. Distribución del IMC:** El IMC varía entre 21 y 26, con una concentración mayoritariamente alrededor de 23-24.

Weight Vs BMI



El uso de asistentes en Excel

Microsoft ha introducido en Excel una herramienta llamada "Analyze data", que permite obtener información sobre los datos a partir del lenguaje natural, de modo que se pueden hacer preguntas, solicitar informes o sugerir pruebas sin fórmulas complicadas. Además, esta herramienta proporciona resúmenes visuales, tendencias y patrones (Microsoft,

2024). Hemos introducido los datos del caso anterior en Excel y hemos clicado en la opción *Analyze data* que se encuentra a la izquierda; activa un menú que permite interactuar con la hoja de cálculo y acceder a información gráfica generada automáticamente, tal como se ve en la imagen.

The screenshot shows Microsoft Excel with a data table and an 'Analyze Data' sidebar. The data table has the following columns: Cas, Edat, Altura (cm), Sexe, Pes (kg), Nivell d'activitat física, and IMC. The 'Analyze Data' sidebar includes a search bar, suggested questions, a pivot table, and a scatter plot.

Cas	Edat	Altura (cm)	Sexe	Pes (kg)	Nivell d'activitat física	IMC
1	25	170	H	70	Moderat	24,221
2	32	165	D	60	Baix	22,039
3	28	180	H	80	Alt	24,691
4	45	160	D	55	Moderat	21,484
5	22	175	H	75	Baix	24,490
6	30	168	D	65	Alt	23,030
7	35	172	H	77	Moderat	26,028
8	40	158	D	58	Baix	23,233
9	27	181	H	82	Alt	25,030
10	26	164	D	62	Moderat	23,052
11	33	177	H	78	Baix	24,897
12	38	162	D	61	Alt	23,243
13	24	174	H	73	Moderat	24,112
14	29	166	D	59	Baix	21,411
15	31	179	H	81	Alt	25,280
16	42	155	D	56	Moderat	23,309
17	23	169	H	74	Baix	25,909
18	37	167	D	63	Alt	22,590
19	34	176	H	79	Moderat	25,504
20	28	163	D	57	Baix	21,454

'Edat' by 'Sexe' and 'Nivell d'activitat física'

Sexe	Nivell d'activitat física			Grand Total
	Moderat	Baix	Alt	
D	113	129	105	347
H	118	78	86	282
Grand Total	231	207	191	629

'Altura (cm)'

En los resultados generados de forma automática, podemos ver una tabla que diferencia los resultados por sexo y una nube de puntos que relaciona la edad con la altura. Como en el caso de ChatGPT o Copilot, esta herramienta también permite hacer preguntas en lenguaje natural. Sin embargo, su funcionamiento es bastante diferente en el ámbito conceptual y debemos tener en cuenta algunas premisas. Por ejemplo, la formulación de *prompts* en esta herramienta debe ser muy sintética y se debe señalar el nombre de la variable o variables que queremos tratar entre comillas simples “.

También es importante tener en cuenta que esta herramienta no puede seguir una “conversación” como hemos visto que se puede hacer con ChatGPT o similares, de manera que cada pregunta debe ser independiente. Lo que sí nos ofrece es una lista de las consultas o preguntas recientes que le hemos hecho.

En las imágenes se pueden ver los *prompts* necesarios para calcular la media del peso de los individuos (Average of 'Pes0 (kg)') o la media del IMC para cada sexo (Average of total 'IMC' by 'Sexo'). El sistema también ofrece sugerencias y formula *prompts* de manera automática.

Analyze Data

← Average of 'Peso (kg)'

Average of 'Peso (kg)'

Here's our answer
Showing average 'Peso (kg)'.

'Peso (kg)'	
Average of Peso (kg)	
	68.25

[+ Insert PivotTable](#) [Is this helpful?](#)

Analyze Data

← Average of total 'IMC' by 'Sexo'

Average of total 'IMC' by 'Sexo'

Here's our answer
Showing total and average 'IMC' by 'Sexo'.

'IMC' by 'Sexo'		
Sexo	Average of IMC	Sum of IMC
H	25,02	250,16
M	22,48	224,84
Grand Total	23,75	475,01

[+ Insert PivotTable](#) [Is this helpful?](#)

Por otro lado, también es posible instalar un asistente de ChatGPT o Copilot tanto en Excel como en Google Sheets. Su funcionamiento es bastante similar al de "Analyze data" integrado dentro del mismo Excel, pero mucho más versátil, ya que permite el uso del lenguaje natural de una forma más parecida a como lo haríamos directamente en el entorno de ChatGPT o Copilot.

9.2. Estadística avanzada

Una de las funcionalidades de ChatGPT es la creación de asistentes o GPT con funcionalidades específicas. En un capítulo posterior, presentaremos las posibilidades de esta herramienta. Entre estos asistentes, hay algunos especializados en análisis estadístico avanzado. Para mostrar sus posibilidades, utilizaremos dos de los asistentes con mayor capacidad, como son Data Analysis & Report AI y Data Analysis. SPSS. Dado que los asistentes se modifican con mucha frecuencia y se actualizan con versiones más avanzadas, podría ser que en el momento de leer este documento ya hayan aparecido otros GPT con nuevas capacidades. ChatGPT cuenta con un buscador que permite encontrar fácilmente los asistentes especializados en un determinado ámbito.



Data Analysis & Report AI

De: Kenneth Bastian &

Your expert in limitless, detailed scientific data analysis and reporting.

Can you analyze this dataset for patterns?

What trends can you identify in this data?

Please generate a report based on this data.

Can you create visualizations for this dataset?



Data Analysis - SPSS

De: Meddidin ebuseerfo &

Expert in statistical analysis and SPSS, guiding researchers.

How do I test the reliability of my data in SPSS?

What statistical test should I use for my hypothesis?

Can you explain how to define variables in SPSS?

I need help interpreting SPSS output for my data.

Data Analysis & Report AI ofrece la información en lenguaje Python, a diferencia de los modelos que hemos visto anteriormente, que nos proporcionan la posibilidad de descargar las tablas de resultados o los gráficos. Este asistente es especialmente útil para análisis complejos, que puede ilustrar con ejemplos concretos.

Podemos utilizar el siguiente *prompt* a partir de los datos que estamos utilizando a lo largo de este capítulo: “¿Qué tipo de análisis no paramétricos podemos hacer con estos datos?”.

El resultado muestra tanto un listado de las pruebas más adecuadas que se podrían realizar como el código Python descargable que nos permitiría ejecutar automáticamente la opción seleccionada.



El análisis no paramétrico es útil cuando los datos no siguen una distribución normal o cuando tenemos muestras pequeñas. En el contexto de estos datos, podríamos considerar varias pruebas no paramétricas para comparar grupos y explorar relaciones:

1. Prueba de Mann-Whitney U

Esta prueba se utiliza para comparar dos muestras independientes. Podría ser útil para comparar las diferencias entre hombres y mujeres (por ejemplo, altura, peso o IMC).

2. Prueba de Kruskal-Wallis

Esta prueba es una extensión de la prueba de Mann-Whitney U para más de dos grupos. Puede ser utilizada para comparar las diferencias en el peso, altura o IMC entre los diferentes niveles de actividad física (Bajo, Moderado, Alto).

3. Prueba de Wilcoxon

Esta prueba se utiliza para comparar dos muestras relacionadas. Puede ser aplicable si tenemos datos antes y después de una intervención para las mismas personas (aunque este no parece ser el caso con los datos actuales).

4. Correlación de Spearman

Esta prueba se utiliza para evaluar la fuerza y la dirección de la asociación entre dos variables continuas o ordinales. Podría ser utilizada para explorar la relación entre edad e IMC, altura y peso, etc.

Ejemplos Concretos

1. Prueba de Mann-Whitney U: Comparación del Peso entre Hombres y Mujeres

```
python Copia el codi  
  
from scipy.stats import mannwhitneyu  
  
# Dividir los datos por sexo  
hombres = df[df['Sexo'] == 'H']['Peso (kg)']  
mujeres = df[df['Sexo'] == 'D']['Peso (kg)']  
  
# Realizar la prueba  
stat, p = mannwhitneyu(hombres, mujeres)  
stat, p
```

2. Prueba de Kruskal-Wallis: Comparación del IMC según Nivel de Actividad Física

```
python Copia el codi  
  
from scipy.stats import kruskal  
  
# Dividir los datos por nivel de actividad física  
bajo = df[df['Nivel de actividad física'] == 'Bajo']['IMC']  
moderado = df[df['Nivel de actividad física'] == 'Moderado']['IMC']  
alto = df[df['Nivel de actividad física'] == 'Alto']['IMC']  
  
# Realizar la prueba  
stat, p = kruskal(bajo, moderado, alto)  
stat, p ↓
```

3. Correlación de Spearman: Relación entre Edad e IMC

```
python Copia el codi  
  
from scipy.stats import spearmanr  
  
# Realizar la prueba de correlación  
corr, p = spearmanr(df['Edad'], df['IMC'])  
corr, p
```

Estos son algunos de los análisis no paramétricos que podrían ser útiles para explorar los datos. Si deseas que realice alguno de estos análisis o tienes alguna otra preferencia, ¡por favor házmelo saber!

El *Data Analysis* – SPSS es un asistente aún más especializado, ya que nos orienta para poder hacer el análisis de datos a través del programa SPSS. Si le planteamos el mismo problema que en el caso anterior, y por lo tanto le introducimos el mismo *prompt*: “¿Qué tipo de análisis no paramétricos podemos hacer con estos datos?”, el resultado es diferente. Por un lado, nos sugiere otras pruebas de análisis no paramétricos que nos podrían ser

útiles para el análisis de datos, de modo que los dos asistentes se complementan; por otro lado, nos ofrece la información necesaria para poder desarrollar el análisis estadístico correspondiente mediante SPSS. Desde un punto de vista docente, esto permite una mayor complementariedad con la actividad de la profesora o el profesor, ya sea con un modelo de clase invertida (o flipped classroom) o bien como apoyo para la realización de ejercicios complejos.



Cómo realizar estos análisis en SPSS:

Prueba de Mann-Whitney U:

1. Ve a **Analyze > Nonparametric Tests > Legacy Dialogs > 2 Independent Samples**.
2. Mueve la variable dependiente al cuadro **Test Variable List** y la variable independiente (grupos) al cuadro **Grouping Variable**.
3. Define los grupos haciendo clic en **Define Groups**.
4. Selecciona **Mann-Whitney U** y haz clic en **OK**.

Prueba de Wilcoxon:

1. Ve a **Analyze > Nonparametric Tests > Legacy Dialogs > 2 Related Samples**.
2. Mueve las dos variables a comparar al cuadro **Test Pairs**.
3. Selecciona **Wilcoxon** y haz clic en **OK**.

Prueba de Kruskal-Wallis:

1. Ve a **Analyze > Nonparametric Tests > Legacy Dialogs > K Independent Samples**.
2. Mueve la variable dependiente al cuadro **Test Variable List** y la variable independiente (grupos) al cuadro **Grouping Variable**.
3. Define los grupos haciendo clic en **Define Range**.
4. Selecciona **Kruskal-Wallis** y haz clic en **OK**.

Prueba de Friedman:

1. Ve a **Analyze > Nonparametric Tests > Legacy Dialogs > K Related Samples**.
2. Mueve las variables a comparar al cuadro **Test Variables**.
3. Selecciona **Friedman** y haz clic en **OK**.

9.3. Máquinas que aprenden: Machine Learning (ML)

El *machine learning* (aprendizaje automático) es una rama de la inteligencia artificial que se sostiene sobre la premisa de que los ordenadores pueden aprender a partir de patrones y experiencias, tal como hacemos los humanos. Este aprendizaje les permite tomar decisiones basadas en la experiencia adquirida previamente (Zhou, 2021). Así, el ML es la técnica que mejora el rendimiento de los sistemas a través del aprendizaje automático mediante métodos computacionales (Zhou, 2021). A partir del desarrollo de modelos estadísticos y algoritmos, los ordenadores pueden aprender de los datos que se les proporciona y hacer predicciones o tomar decisiones con cierta autonomía (Rodríguez, 2023).

Durante los últimos años, el aprendizaje automático ha experimentado un avance significativo gracias al aumento de la capacidad de procesamiento, la disponibilidad masiva de datos y el progreso en el campo de la inteligencia artificial. Este hecho ha llevado a que el ML se convierta en una de las técnicas más importantes en algunas áreas aplicadas, como la visualización por ordenador o el procesamiento del lenguaje natural (Zhou, 2021), y juega un papel cada vez más importante en el apoyo de la investigación científica, siendo crucial para el desarrollo tecnológico (Mjolsness y DeCoste, 2001).

Actualmente, existen muchos softwares especializados que incluyen algoritmos y técnicas de análisis de datos basadas en ML. De hecho, gran parte de los principales programas de análisis de datos cuentan con herramientas o módulos de ML que facilitan la explotación de datos complejos. Por ejemplo, el software libre JASP incluye un módulo de ML con técnicas de regresión, clasificación, agrupamiento y predicción. También encontramos softwares como MAXQDA que permiten realizar un análisis de los datos textuales a través del procesamiento del lenguaje natural.

En este capítulo nos centraremos en mostrar dos aplicaciones prácticas del ML. En primer lugar, un análisis de clúster como modelo de búsqueda de patrones en los datos; y, en segundo lugar, un análisis

basado en la minería de datos como una aplicación práctica del procesamiento del lenguaje natural.

Búsqueda de patrones: El análisis de clúster

El análisis de clúster es una técnica estadística cuyo objetivo principal es encontrar patrones o relaciones ocultas en los datos, agrupando los elementos que comparten características similares. Los datos se agrupan en un número reducido de clústeres o grupos (Hair et al., 1998) que son mutuamente exclusivos (Cea, 2004). De forma muy resumida, cuando se aplica un análisis de clúster, es necesario tomar diversas decisiones metodológicas. Lo más habitual es realizar un método jerárquico. Existen varios algoritmos de clasificación (inter-grupos, intra-grupos, método de Ward...) y todos ellos tienen particularidades diferentes y priorizan ciertos encadenamientos u otros dentro de los datos. Por lo tanto, priorizan características diferentes en los patrones de los datos. Además, cuando se dispone de variables con datos que responden a medidas diferentes, es necesario proceder a la estandarización de los datos (z score) para evitar la posible influencia de la Distancia Euclidiana cuadrada en los clústeres resultantes.

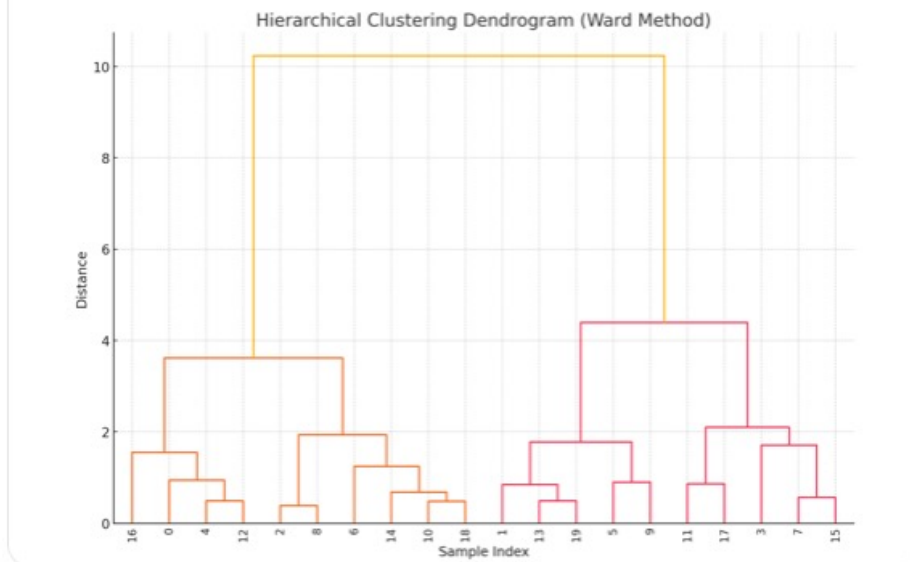
En el ejemplo que proponemos a continuación, mantenemos los mismos datos que hemos utilizado desde el inicio del capítulo. Ahora le pediremos a ChatGPT que nos haga un análisis de clúster, utilizando el algoritmo de agrupación método de Ward (ya que se trata de una base de datos con pocos casos) y que estandarice los datos con z-score. En este caso le daremos las órdenes de cómo queremos que lo haga porque ya tenemos los conocimientos previos, pero también habría la posibilidad de pedirle que nos diera algunas sugerencias para el tratamiento de los datos. Así, le damos la orden con el siguiente *prompt*:

“Haz un análisis de clúster de la tabla con el método de Ward y z-score”.

El resultado es el que se presenta a continuación, donde se proporciona el dendrograma, los datos con la clasificación del clúster, y la correspondiente interpretación.



Hierarchical Clustering Dendrogram (Ward Method)



Aquí tens el dendrograma resultat de l'anàlisi de clústers utilitzant el mètode de Ward i la normalització z-score:
Aquí tienes el dendrograma resultado del análisis clúster utilizando el método de Ward y la normalización z-score:

- Cada hoja del dendrograma representa un individuo en el conjunto de datos.
- La altura en la cual los individuos o grupos se combinan indica la distancia (o disimilitud) entre ellos. Cuanto más grande es la altura, más diferentes son los grupos.
- El dendrograma ayuda a identificar el nombre de clústeres naturales dentro del conjunto de datos. Puedes cortar el dendrograma a diferentes niveles para obtener diferentes nombres de clústeres.

Si necesitas una interpretación más detallada de los clústeres o deseas cortar el dendrograma a un nivel específico para obtener un número determinado de clústeres, házmelo saber.

Sí, interpreta los resultados poniendo el nombre a los clústers e indicando la mejor solución



Clustered Data ↓ ↗

	Caso	Edad	Altura (cm)	Sexo
1	1	25	170	H
2	2	32	165	D
3	3	28	180	H
4	4	45	160	D
5	5	22	175	H
6	6	30	168	D

El análisis de clústeres ha resultado en dos agrupaciones principales:

1. Cluster 1:

- Edad media: 28,2 años
- Altura media: 175,3 cm
- Peso medio: 76,9 kg
- IMC medio: 25,02



2. Cluster 2:

- Edad media: 34,7 años
- Altura media: 162,8 cm
- Peso medio: 59,6 kg
- IMC medio: 22,48

Interpretación de los Clústeres:

• Cluster 1 (Jovenes Altos):

- Este grupo está formado por personas más jóvenes, con una altura más grande y un peso más elevado. Su IMC también es superior en comparación al del otro grupo.

• Cluster 2 (Adultos de Baja Altura):

- Este grupo está formado por personas un poco mayores, con una altura más baja y un peso menor. Su IMC es inferior en comparación al otro grupo.

Mejor solución:

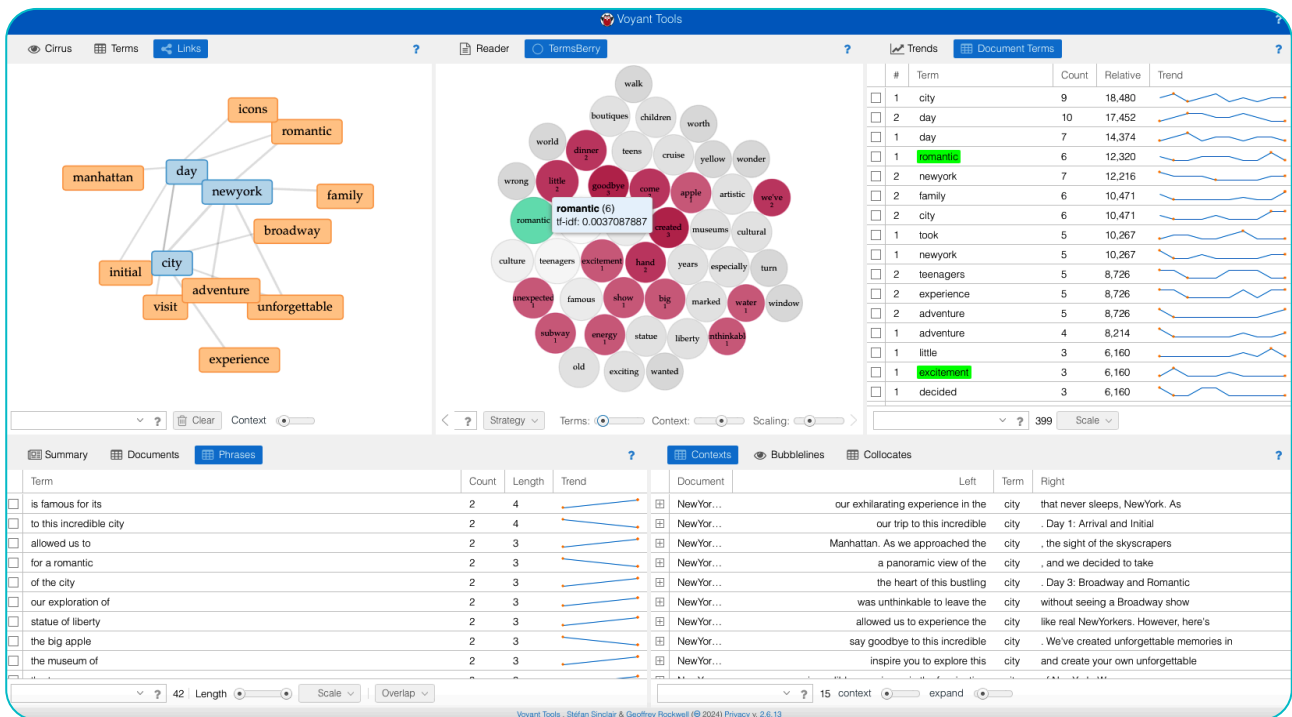
- La mejor solución es la que divide la población en dos clústeres que reflejan diferencias significativas en las características físicas (altura, peso) y en la edad. Estos dos clústeres son fácilmente interpretables y muestran diferencias claras entre los dos grupos de personas.

Si necesitas más detalles o otros análisis, por favor, házmelo saber.

Esta es una muestra del potencial existente para proceder con el análisis de datos mediante técnicas de ML a través de ChatGPT. Sin embargo, será necesario comprobar que el resultado sea lógico y correcto según los datos. Esta puede ser una vía para explorar los datos antes de utilizar un software estadístico, como por ejemplo JASP, con el fin de proceder con el análisis definitivo.

Procesamiento del lenguaje natural: *Text mining* y análisis de sentimiento

La acumulación de grandes cantidades de datos textuales, como por ejemplo en redes sociales, páginas web y blogs, ha incrementado la necesidad de métodos eficientes para su tratamiento, procesamiento y análisis. Se requieren, por tanto, técnicas y herramientas que faciliten la extracción de información relevante de estos datos para que pueda ser interpretada fácilmente. La minería de datos textuales (*text mining*) es el conjunto de técnicas y métodos utilizados para analizar este tipo de datos, permitiendo extraer conocimiento y patrones significativos (Gaikwad, Chaugule y Patil, 2014). Este campo incluye una variedad de procesos, como la limpieza de datos, la transformación y la modelización de los datos



Podemos ver cómo se obtiene información sobre las palabras más recurrentes, los diversos temas y las relaciones temáticas que emergen de los textos analizados, la frecuencia de frases y palabras, etc. Las utilidades de esta herramienta para la práctica docente son muy diversas: La rapidez y accesibilidad permiten que la sesión docente se pueda centrar en el análisis de los resultados o la comparación entre casos diversos. Sin embargo, es muy importante que el estudiante tenga claros los procesos estadísticos y operativos que hay detrás

del análisis textual, para que pueda interpretar mejor los resultados y conozca las limitaciones de los modelos. Las posibilidades en el ámbito de las ciencias sociales son casi infinitas: estudios de los discursos políticos, análisis de las conversaciones a partir de la transcripción de un documental, captura y análisis de los tweets dedicados al racismo, análisis semántico de los medios de comunicación según su orientación, análisis de las letras de las canciones de reguetón o estudio de las descripciones turísticas de los destinos.

Práctica. “¿A qué dedica el tiempo libre?”

Objetivo

- Explorar el potencial del análisis de datos utilizando ChatGPT.
- Trabajar con bases de datos abiertas.

Consideraciones

- Dirigida a estudiantes de estudios de ciencias sociales.
- Actividad para realizar preferentemente en grupo.
- Herramienta de IA: ChatGPT o Copilot.
- El ejercicio se llevará a cabo con una base de datos abierta, como las que proporciona el Instituto de Estadística de Cataluña (<https://www.idescat.cat/dades/micro/>). Las más versátiles son la encuesta de condiciones de vida, la encuesta de uso del tiempo y la encuesta de hábitos lingüísticos, que ofrecen acceso a microdatos.

Para esta práctica, trabajaremos esencialmente con la base de datos sobre usos del tiempo, que permite conocer el tiempo dedicado a diversos ámbitos (cuidado personal, trabajo remunerado, estudios, trabajo no remunerado, hogar y familia, actividades de ocio y desplazamientos) por parte de una muestra representativa de los catalanes y catalanas. Dado el gran volumen de información con el que se trabaja y el elevado número de variables, es recomendable centrarse en un determinado ámbito, como, por ejemplo, el desigual tiempo dedicado a las actividades del hogar según el género.

Una vez seleccionada la variable o variables de estudio, se solicitarán las siguientes tareas:

- Estadística descriptiva (tablas de frecuencias, media, desviación típica, mínimo y máximo, etc.).
- Explorar posibles relaciones entre las diferentes variables del conjunto de datos.
- Generar las representaciones gráficas de los datos que se considere conveniente.
- Desarrollar un breve informe en el que se interpreten los datos.

Y una vez terminado, con el grupo clase:

- Compartir las reflexiones y resultados obtenidos.
- Establecer conclusiones generales sobre el análisis de datos realizado.
- Discutir las principales ventajas y retos en el uso de herramientas de IA para el análisis de datos.

Práctica. La conversación sobre la COVID-19

Objetivo

- Aprender a utilizar herramientas de IA para analizar datos textuales de manera efectiva.
- Desarrollar habilidades para interpretar y comunicar resultados de análisis textual.
- Entrar en contacto con los datasets textuales.

Consideraciones

- Dirigida a estudiantes de ciencias sociales y humanidades.
- Actividad para realizar preferentemente en grupo.
- Herramienta de IA: [Voyant Tools](#).
- Dataset: [kaggle.com](#).

En la red, hay una gran cantidad de servidores que disponen de bases de datos textuales de acceso libre que permiten trabajar en este campo y que son especialmente útiles para usos docentes. Uno de los espacios que permite acceder a un mayor volumen de información es [kaggle.com](#), que es una comunidad sobre datos textuales con un gran número de *datasets*; kaggle forma parte de Google Cloud. En este caso, planteamos la posibilidad de acceder al *dataset* de tweets sobre la COVID-19 (<https://www.kaggle.com/datasets/gpreda/covid19-tweets>), que recopila una selección de tweets que utilizan este hashtag. Hay muchas bases de datos que se pueden adaptar a la asignatura, como una dedicada a las opiniones sobre Biden o Trump, sobre los sentimientos expresados en los comentarios sobre los Juegos Olímpicos, valoraciones de películas o debates públicos sobre temas controvertidos. Para evaluar el dataset descargado, se empleará Voyant Tools (<https://voyant-tools.org/>). Será necesario seleccionar una parte de los tweets con un determinado criterio (temporal, geográfico, temático).

El procedimiento propuesto es el siguiente:

- Cargar los textos en Voyant Tools, asegurando que los textos tienen el formato adecuado para el análisis.
 - Desarrollar el análisis estadístico descriptivo.
 - Extracción de palabras frecuentes: Identificar las palabras más frecuentes utilizando la funcionalidad "Summary", "Terms" y "Document Terms".
 - Distribución de palabras: Analizar la distribución de palabras a lo largo del texto con la herramienta "Trends".
 - Representación gráfica y visual de los datos:
 - Visualización de palabras: Crear una nube de palabras utilizando la herramienta "Cirrus" para identificar las palabras más frecuentes en el texto.
 - Visualizar relaciones temáticas: Crear una red de palabras con las herramientas "Links" y "TermsBerry" para visualizar los enlaces más comunes entre palabras y temas.
6. Desarrollar un breve informe e interpretar los resultados obtenidos.

En el grup classe, es poden realitzar les següents activitats:

- Reunir al grupo clase para compartir las reflexiones y resultados obtenidos.
- Establecer conclusiones grupales sobre el análisis textual realizado.
- Discutir las principales ventajas y retos en el uso de la IA para el análisis de datos textuales.

Referencias

- Briones, J.A. (n.d.) <https://www.plenainclusion.org/wp-content/uploads/2022/02/Plena-inclusion-Murcia.-Guia-de-Excel.pdf>
- Gaikwad, S. V., Chaugule, A. y Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications*, 85 (17), 42-45.
- Khan, F. H., Qamar, U. y Bashir, S. (2016). eSAP : A decision support framework for enhanced sentiment analysis and polarity classification. *Information Sciences*, 367, 862-873.
- Lau, K. N., Lee, K. H. y Ho, Y. (2005). Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly*, 46 (3), 344-362.
- Mannyica, James et al. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. New York: McKinsey & Company [en línea: 20/05/2024]. <http://lazowska.cs.washington.edu/escience/McKinsey.big.data.pdf>
- Microsoft (2024). Análisis de datos en Excel. [En línea: 02/06/2024] <https://support.microsoft.com/es-es/office/análisis-de-datos-en-excel-3223aab8-f543-4fda-85ed-76bb0295ffc4>
- Mora, M. (2019). *Els rols, àmbits i noms de la ciència de dades*. FUOC.
- O'Beirne, P. (2023). Proceedings of the EuSpRIG 2023 Conference "The Spreadsheet Crisis: Regaining Control". European Spreadsheet Risks Interest Group.

[1] Este prompt puede incluir una indicación sobre el idioma de Excel, ya que las diferentes versiones del programa llevan asociada la traducción de las fórmulas. En el caso del ejemplo, Excel está en versión inglesa.

10

Los modelos de lenguaje en el análisis cualitativo

Las metodologías de investigación cualitativa juegan un papel fundamental en el campo de las ciencias sociales, ofreciendo una comprensión profunda y matizada del comportamiento humano y las motivaciones que lo sustentan. A través de métodos de recogida de datos como entrevistas y grupos de discusión, los investigadores y las investigadoras pueden reunir información rica y detallada que ilumina por qué las personas piensan y se comportan de determinadas formas. Epistemológicamente, la investigación cualitativa no revela una “verdad” central que pueda generalizarse, sino que produce resultados transferibles (Maxwell, 2013), que parecen verdaderos para los participantes en la investigación, de acuerdo con sus manifestaciones. A través de un proceso de las narrativas matizadas se transforman en temas comprensibles y posteriormente se construyen patrones e ideas clave significativas (Burnard, 1991; Jalali & Akhavan, 2024).

Para ello, se utilizan métodos robustos de recogida de datos, la participación y el compromiso de informantes apropiados, la codificación meticulosa y la triangulación, entre otros, para mejorar la fiabilidad.

Este proceso es fundamentalmente largo y laborioso. Debido a este hecho, los avances tecnológicos en paquetes de software, incluidos ATLAS.ti, NVivo y MAXQDA, así como enfoques de procesamiento de lenguaje natural, como el análisis de sentimientos y la modelización de temas, ayudan a las investigadoras y los investigadores a producir material analítico útil mientras gestionan e interpretan grandes volúmenes de datos textuales (Zhang et al., 2024).

Como resultado, en los últimos años se han desarrollado herramientas de transcripción de inteligencia artificial (IA) con sofisticados algoritmos de reconocimiento de voz que automáticamente traducen grabaciones de audio y vídeo a texto escrito, permitiendo a los investigadores e investigadoras dedicar su atención al análisis (Christou, 2023). Además, las herramientas de lenguaje de IA, como Claude, ChatGPT o Gemini, se han añadido a la investigación cualitativa, que a través del uso de técnicas de aprendizaje profundo y la capacidad de procesar y generar lenguaje humano, pueden producir desde la generación de texto simple hasta la categorización temática y la comprensión

“Durante el análisis de datos, se ha encontrado que las herramientas de lenguaje de IA como ChatGPT o Gemini son más descriptivas que interpretativas”

contextual (Xiao et al., 2023). Concretamente, se ha estudiado que ChatGPT es una herramienta útil en la investigación cualitativa por su capacidad de aportar un nuevo nivel de profundidad al análisis de datos, puesto que entiende el lenguaje natural y permite a los usuarios comunicarse fácilmente con él. Además, es resistente a errores de entrada del usuario o la usuaria, así como errores ortográficos o gramaticales, y responde eficientemente a las solicitudes y procesos de interpretación de datos, proporcionando explicaciones y detalles sobre los resultados proporcionados, ofreciendo transparencia (Zhang et al., 2024). Así, como herramienta, va más allá de capturar información limitada a la frecuencia de frases o tonos emocionales (a través del análisis de sentimientos o de contenido), hasta la identificación de problemas políticos, sociales o culturales, o extensión de la capacidad del investigador por refinar códigos, construir conexiones conceptuales y crear teoría (Anis & French, 2023). Como respuesta a esto, los paquetes mencionados anteriormente (ATLAS.ti, NVivo y MAXQDA) han desarrollado colaboraciones con OpenAI, mejorando el proceso de codificación en términos de tiempo y análisis.

Aunque el uso de la IA en metodologías cualitativas ofrece un contexto de trabajo interactivo y que ahorra tiempo, existen algunos retos a considerar (Anis & French, 2023; Christou, 2023; Morgan, 2023; Jalali & Akhavan, 2024; Zhang et al., 2024).

- Durante el análisis de datos, se ha encontrado que las herramientas de lenguaje de IA como ChatGPT

o Gemini son más descriptivas que interpretativas. Hay una falta de sentido común y comprensión contextual de los elementos sociales y culturales en la IA, especialmente cuando se presentan cuestiones ambiguas. Estos resultados también se conocen como alucinaciones, básicamente producidas por las limitaciones del contexto (consultas previas y contenido de las respuestas) intentando predecir qué podría seguir. Además, el texto generado por humanos que contiene metáforas o sarcasmo no siempre es comprendido por el programa.

- Otro reto del uso de sistemas de IA es la simplificación excesiva, considerando que están entrenados para priorizar la eficiencia y simplicidad sobre la profundidad de entornos y fenómenos complejos. Esto podría llevar a la pérdida de ideas clave valiosas.
- Como hemos visto en un capítulo anterior, las herramientas de lenguaje de IA llevan sesgos inherentes porque están básicamente entrenadas con datos generados por humanos y, aunque se hacen esfuerzos por limitarlos, hasta ahora se prevé que los resultados reflejen ideas y creencias dominantes prevalentes en la sociedad. Esto podría afectar a la validez y fiabilidad de la investigación, especialmente si la investigadora o el investigador no controla el proceso de análisis de datos.
- Como herramienta, la IA debería mejorar las capacidades de los investigadores y las investigadoras, más que sustituirlas. Aunque puede ayudar en la interacción con los datos, no se puede considerar un intérprete independiente u objetivo

de los datos ni mucho menos una máquina de construcción de teorías. La investigación cualitativa abarca la subjetividad y es aquí donde los conocimientos, comprensión y perspectivas de los investigadores son fundamentales para la reflexión e interpretación de los datos, especialmente en estudios etnográficos o fenomenológicos. Es responsabilidad de los investigadores y las investigadoras cuestionar, mejorar e incluir o excluir los códigos, temas y patrones producidos por la IA.

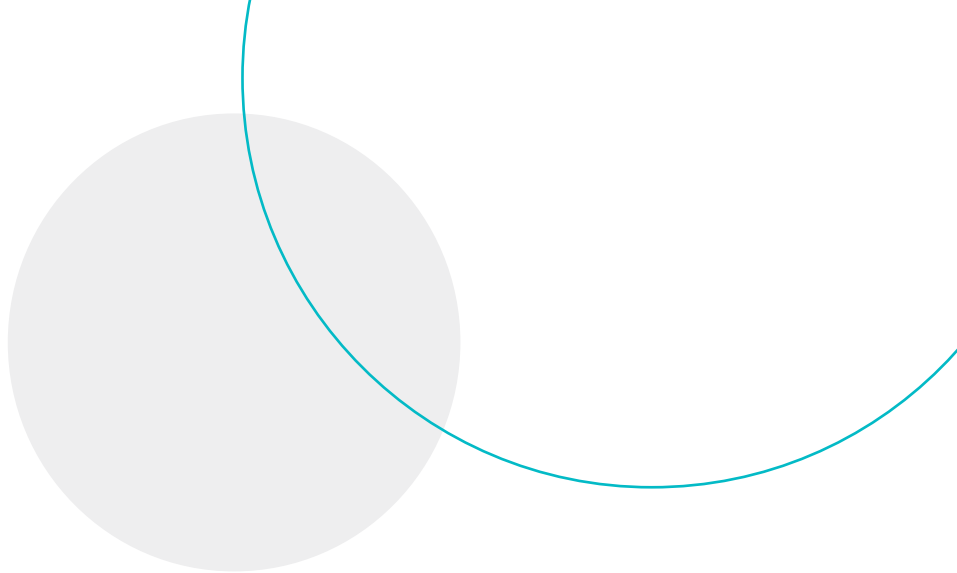
- Las preocupaciones éticas y de propiedad de los datos están aumentando en el campo de la compartición de datos basados en humanos, incluyendo información personal identificable o historias privadas procedentes de entrevistas personales o grupos de discusión, ya que no existe una guía clara sobre cómo deberían ser utilizadas o compartidas. Esto sucede habitualmente en el caso de los programas de transcripción automatizada que, además, tienen el inconveniente de eliminar la oportunidad de una mayor implicación con los datos cualitativos por parte de la investigadora o investigador. Para evitar las preocupaciones éticas, los datos personales deben eliminarse antes de introducirlos en los programas y deben entregarse documentos de consentimiento a los y las participantes en cuanto a la protección de datos.

A pesar de estos retos, los investigadores y las investigadoras de análisis cualitativo siguen experimentando con herramientas de IA para demostrar su utilidad como productoras de material y asistentes en el estudio. Las futuras innovaciones pueden mejorar las capacidades de la IA en la comprensión de bases de datos más matizadas y proporcionar un análisis cualitativo más profundo. Los artículos sobre este tema apoyan las colaboraciones enriquecedoras entre el know-how de los investigadores y las investigadoras y las respuestas de las herramientas de IA, especialmente en el área de la codificación de datos textuales y el análisis temático.

10.1. La asistencia a la guía de preguntas

Sin embargo, una parte fundamental de las metodologías cualitativas es la generación de datos antes del análisis. La mayor parte de la investigación relacionada con las herramientas de IA y la investigación cualitativa comienza con el hecho que los investigadores o las investigadoras ya tienen datos textuales ricos de las transcripciones de entrevistas, que son típicamente el resultado de entrevistas semiestructuradas o grupos de discusión. Sin embargo, la calidad de los datos recogidos durante una entrevista no se determina únicamente por la calidad del protocolo de la entrevista. Particularmente en el caso de las metodologías cualitativas, la generación de datos es un proceso complejo que requiere, además de un estudio previo y la preparación del protocolo de la entrevista, una atención continua y adaptación a los inputs de los entrevistados para asegurarse que todos los datos interesantes que fueron expresados explícita o implícitamente estén adecuadamente presentados y descritos. Durante una entrevista, con el rol investigador, el estudiantado debe ser capaz de elegir entre ser simples oyentes o co-creadores de datos. El segundo rol es fundamental especialmente en el caso de entrevistas semiestructuradas o no estructuradas.

En primer lugar, es importante que los estudiantes preparen un protocolo de entrevista sólido. El error más común es ir directamente a las preguntas que quieren hacer sin romper el hielo, sin aprender lo suficiente sobre sus entrevistados y entrevistadas y sin conectar las teorías leídas previamente con el caso empírico que se estudia. Preparar un protocolo completo requiere un tiempo considerable, donde las herramientas de lenguaje de IA, como ChatGPT, son útiles. Para formular preguntas apropiadas basadas en características demográficas o psicográficas, resulta esencial describir claramente el objetivo de la investigación en la solicitud. Dado que las entrevistas semiestructuradas son las más comunes, es beneficioso indicar el número de preguntas que se quieren hacer (alrededor de 30) y su formato (abierto). Basándose en lo que las estudiantes y los estudiantes han leído en las revisiones de literatura existente, es importante



revisar las preguntas y añadir conceptos que faltan en lo que se refiere al trasfondo, las experiencias y las percepciones de los entrevistados y las entrevistadas.

Por ejemplo, cuando pedimos a ChatGPT (versión 4.0) que cree un protocolo de entrevista bajo este objetivo: “Explorar y entender las experiencias, desafíos y adaptaciones del estilo de vida de las nómadas digitales femeninas en entornos urbanos y rurales. Este estudio tiene como objetivo descubrir cómo las dinámicas físicas y sociales de estos entornos influyen en su equilibrio entre la vida laboral y personal, la productividad profesional y el bienestar personal. Al examinar los matices de sus interacciones con diferentes paisajes culturales e infraestructurales, la investigación pretende contribuir al conocimiento más amplio del trabajo remoto y el género en la economía digital”, se han detectado los siguientes elementos susceptibles de modificación:

- Sólo la primera pregunta (“¿Puedes describir brevemente tu historial profesional?”) demanda sobre el perfil de las entrevistadas, el resto estaban directamente conectadas con el estilo de vida de la nómada digital. Hay que añadir más preguntas para conocer mejor a nuestra entrevistada (vida personal, trabajos anteriores, experiencias de viaje) y cómo era la vida antes de esta decisión.
- Algunas preguntas deben simplificarse, cuando tienen muchas variables y pueden ser difíciles de recordar. Por ejemplo, “Crees que ciertos tipos de trabajo son más fáciles o más difíciles en entornos urbanos o rurales?”. Aquí, ChatGPT creó

una pregunta que pide tanto elementos negativos como positivos, así como entornos urbanos y rurales. Debería separarse en al menos dos preguntas.

- Algunas preguntas guían el tipo de respuesta que se busca. Por ejemplo, “¿Has afrontado alguna situación de aislamiento profesional? ¿Cómo lo gestionas?” es una pregunta que dirige hacia el problema particular del aislamiento, y obvia otras situaciones de mayor importancia para la entrevistada. Transformar esta pregunta en abierta puede permitir un espectro más amplio de respuestas (“¿Cuáles son los inconvenientes profesionales que este estilo de vida te ha traído? ¿Cómo los has gestionado?”).

Si se desea conectar este protocolo con una revisión de literatura específica, se pueden incorporar en el *prompt* los artículos que se consideren más relevantes y pedir al generador que añada algunas preguntas al protocolo basándose en los artículos, y después realizar una selección y valoración de la propuesta. Por ejemplo, en el caso expuesto, se pidió a ChatGPT que tuviera en cuenta [este artículo](#) sobre el tema y creara preguntas relacionadas con la perspectiva de género en el estilo de vida de los nómadas digitales. Esto permitió la generación de diez preguntas relacionadas con los contenidos del artículo, de las cuales se eliminaron dos por ser muy similares al protocolo original. De esta forma, se ha creado una guía de entrevista de unas 40 preguntas, que puede ser muy útil para un o una estudiante con poca experiencia en la creación de guiones de entrevistas.

Práctica. El asistente de la entrevista

Pide a ChatGPT que diseñe un protocolo de entrevista semiestructurada, de aproximadamente 30 preguntas abiertas, basado en el siguiente objetivo: Explorar y entender las motivaciones detrás de las elecciones de los estudiantes internacionales para cursar programas de grado en universidades de toda Europa, examinar sus experiencias durante el período de movilidad y evaluar el impacto de estas experiencias en su desarrollo personal y rasgos de personalidad. Esta investigación buscará los principales factores que influyen en las decisiones de los estudiantes internacionales, incluyendo motivaciones académicas, sociales, culturales y económicas. También profundizará en las experiencias vividas por estos estudiantes, evaluando su adaptación a nuevos sistemas educativos, entornos sociales y contextos culturales, así como los retos que enfrentan y los sistemas de soporte que utilizan. Además, el estudio tiene como objetivo evaluar los cambios en los rasgos de personalidad y el desarrollo personal derivados de su movilidad internacional, entendiendo el impacto de las interacciones interculturales, los desafíos académicos y el aumento de la independencia en su autopercepción y crecimiento. Por último, la investigación comparará las expectativas previas a la movilidad de los estudiantes con sus reflexiones posteriores a la movilidad para analizar cómo sus experiencias reales se alinean con o difieren de sus motivaciones y expectativas iniciales.

- a)** Pide a los estudiantes que revisen las preguntas generadas para:
- conseguir la información de perfil necesaria de los estudiantes internacionales
 - asegurarse de que cada pregunta no tenga más de una variable para no confundir a los entrevistados
 - asegurarse de que no haya indicación de una respuesta a la pregunta
 - cambiar preguntas cerradas por preguntas abiertas
 - añadir y destacar las preguntas que no se incluían en el protocolo propuesto, pero que considera importante preguntar
- b)** Pide a los estudiantes que lean un artículo relacionado con la Movilidad Estudiantil Internacional (por ejemplo: Tran, L. T. (2016). Mobility as 'becoming': A Bourdieuan analysis of the factors shaping international student mobility. *British Journal of Sociology of Education*, 37(8), 1268-1289).

Después de leerlo, deben sugerir nuevas preguntas basadas en la información que han aprendido y justificar su importancia.

- c)** Los estudiantes deberían añadir este artículo a ChatGPT, en el mismo chat donde se generó el protocolo, y pedirle que sugiera nuevas preguntas basadas en la información proporcionada en el artículo.

Entonces, deberían comparar estas preguntas con las que habían sugerido previamente y realizar una discusión sobre la contribución de la IA en la generación del protocolo.

Objetivo

Capacitar a los estudiantes para aprovechar eficazmente la IA para generar preguntas de entrevista iniciales y estructurar el protocolo evaluando críticamente y refinando estas preguntas para garantizar que se alineen con los objetivos de la investigación, los estándares éticos y las complejidades del tema en investigación. Esta tarea tiene como objetivo mejorar la capacidad de los estudiantes para elaborar preguntas de entrevista perspicaces y relevantes relacionadas con investigaciones anteriores y preparar un protocolo de entrevista que les proporcione la seguridad de que todos los temas posibles de

discusión están cubiertos (aunque después de las primeras entrevistas se añadirán nuevas preguntas).

En la investigación cualitativa, se ha demostrado que para que los entrevistados divulguen su información personal y describan en detalle sus perspectivas, percepciones y experiencias, debe generarse confianza entre ellos y el investigador. En entrevistas semiestructuradas y no estructuradas, el éxito radica en que las preguntas que siguen en cada respuesta se basan en el contenido de esta respuesta. Los entrevistadores y entrevistadoras deben estar formados al escuchar atentamente todos los elementos de cada respuesta y formular automáticamente nuevas preguntas de clarificación abiertas basadas en razonamientos o hechos no previstos proporcionados por los entrevistados. El resultado de este proceso será la generación de datos complejos y profundos que tendrán que ser codificados cuidadosamente con el objetivo de enriquecer el análisis de datos.

Existen dos técnicas básicas utilizadas en entrevistas que permiten profundizar más en las respuestas de los entrevistados; la técnica de la *laddering* y la del *bracketing*. El *laddering* es una técnica de investigación cualitativa que se utiliza para descubrir las motivaciones, valores y creencias de los individuos que influyen en sus decisiones y comportamientos (Miles & Rowe, 2004). Durante este proceso, el entrevistador se centra en atributos específicos de un producto o servicio y realiza preguntas progresivamente más profundas para revelar las consecuencias asociadas y, por último, los valores fundamentales del individuo. Mediante el modelo de Atributos-Consecuencias-Valores (ACV), el entrevistador se desplaza desde las características tangibles hasta los beneficios que proporcionan estas características a los valores fundamentales que satisfacen. Este enfoque proporciona una visión profunda del comportamiento del consumidor, informando de las decisiones estratégicas en marketing, diseño de productos y experiencia del usuario conectando las preferencias superficiales con los impulsores psicológicos más profundos. Básicamente, esto se logra preguntando por qué el entrevistado considera importantes los atributos y las consecuencias.

En cambio, el *bracketing* se refiere a los métodos que utilizan los investigadores para mantener la entrevista a nivel personal aunque conozca previamente mucho sobre el tema que se trata (Sorsa et al., 2015). Así, los estudiantes suspenden sus propias suposiciones sobre el mundo, permitiendo a los participantes describir su punto de vista, desarrollar la subjetividad sin prejuicios y establecer la confianza entre los participantes. Un entrevistador, por ejemplo, puede tener una comprensión de lo que significa una participante en un estudio si declara que es una compradora compulsiva online, pero se deben hacer más preguntas para permitir a los participantes describir las situaciones en las que se han producido estos comportamientos, haciendo que proporcionen información más detallada.

Aquí es donde entra en juego la IA para formar al estudiante. Para que se cree confianza y que el entrevistado participe en una conversación en lugar de una entrevista, el estudiante debe formular automáticamente las preguntas de seguimiento adecuadas en el momento adecuado, independientemente de cualquier protocolo preparado.

La capacidad "conversacional" de ChatGPT, por ejemplo, permite a los usuarios interactuar con él a través del lenguaje natural (Zhang et al., 2024). En este contexto, ChatGPT puede simular entrevistas no estructuradas actuando como el entrevistado después de haber descrito en la indicación el objetivo de la investigación y el perfil del entrevistado que ChatGPT retratará.

Práctica. El modelo de lenguaje como esparrin

En esta práctica, los estudiantes o las estudiantes interactuarán con Gemini, que desempeñará el papel de una persona específica. Mediante este ejercicio, practicarán la generación de preguntas espontáneas y relevantes basadas en las respuestas que reciban, adaptando su consulta dinámicamente para obtener conocimientos profundos sobre las experiencias y perspectivas del sujeto. Gemini asumirá el rol de una madre soltera de 32 años que vive en un pequeño piso cerca del parque del Retiro en Madrid. Trabaja como ejecutiva de marketing digital en L’Oreal y es madre de una niña de 7 años. La entrevista explorará cómo se organiza para ir de vacaciones con su hija, centrándose en los retos que enfrenta, su proceso de toma de decisiones y las opciones que considera, proporcionando una comprensión matizada de sus experiencias equilibrando su vida personal y responsabilidades laborales.

Instrucciones

Inicie la entrevista: Empiece presentándose y explicando el propósito de la entrevista. Inicie la entrevista con las siguientes preguntas: “Gracias por dedicarme su tiempo hoy. Para empezar, ¿podría contarme la última vez que planeó unas vacaciones con su hija? ¿Cuáles fueron algunos factores clave que tuvo en cuenta a la hora de elegir su destino?”

Realice la entrevista teniendo en cuenta la técnica de la escalada y el bracketing:

- Permite que la conversación fluya de forma natural.
- Céntrate en preguntas abiertas para fomentar respuestas detalladas.
- Lee sus respuestas y formula tus siguientes preguntas basándote en la información que proporciona.
- Explora no sólo el ‘qué’, sino también el ‘cómo’ y el ‘por qué’ detrás de sus procesos de planificación de vacaciones y decisiones.

Temas sobre los que debes preguntar en general:

- Sus prioridades a la hora de elegir un destino vacacional.
- El equilibrio entre las responsabilidades laborales y el tiempo personal.
- Consideraciones financieras y logísticas en la planificación de vacaciones.
- Sistemas de soporte o recursos que pueda utilizar.
- Retos que enfrenta como madre soltera a la hora de planificar vacaciones.

Reflexión

Tras la entrevista, escribe una reflexión sobre el proceso. Discute cómo adaptaste tus preguntas en función de las respuestas de la entrevistada y cómo podrías mejorar estas preguntas después de tu reflexión.

Objetivo

Mejorar las habilidades de los estudiantes para realizar entrevistas no estructuradas, un método clave en la investigación cualitativa. Al interactuar con un entrevistado simulado en la persona de Gemini, los estudiantes practicarán la generación de preguntas espontáneas y relevantes basadas en las respuestas recibidas en tiempo real, adaptando su consulta de forma dinámica y asegurándose de que toda la información proporcionada esté bien descrita.

10.2. La codificación con un modelo de lenguaje

El análisis temático es una herramienta crucial en la investigación cualitativa por su flexibilidad, profundidad y enfoque sistemático. Permite las investigadoras y los investigadores descubrir e interpretar las capas complejas de los datos cualitativos, proporcionando valiosa información sobre el comportamiento humano, las experiencias y los fenómenos sociales. Permite también identificar, analizar e informar sobre patrones de significado (temas) dentro de los datos. Incluye un enfoque de seis fases, que son: familiarización con el conjunto de datos, codificación de datos, generación inicial de temas, desarrollo y revisión de temas, refinamiento y definición de temas, e informe final (Zhang et al., 2024). Esta metodología es lenta y requiere un esfuerzo manual significativo, especialmente cuando se trata de conjuntos de datos grandes y complejos, lo que representa un reto para los investigadores noveles, ya que interviene la subjetividad a la hora de identificar los temas. El reto para los estudiantes y las estudiantes es aún mayor, ya que no están familiarizados con el proceso de codificación.

Los investigadores e investigadoras han intentado demostrar la utilidad potencial de los programas de IA en el análisis de datos cualitativos comparando su eficacia con la codificación manual (Morgan, 2023; Jalali & Akhavan, 2024). La comparación de la codificación manual y la codificación con herramientas de lenguaje IA se demuestra como un método complementario, centrándose en ChatGPT (versión 4.0), ya que ambos identifican temas significativos y bucles de retroalimentación, pero en ocasiones no coinciden; es decir, ChatGPT puede proporcionar parte del análisis que el investigador puede considerar al interpretar los datos y mejorar el análisis producido. En particular, ChatGPT ofrece un enfoque más directo y objetivo, lo que puede reducir el riesgo de introducir los sesgos mismos de los investigadores e investigadoras. Además, se ha sugerido que los modelos de IA funcionen mejor en el análisis deductivo en lugar del análisis inductivo, lo que hace que la participación humana sea aún más crítica.

Para que los estudiantes y las estudiantes entiendan

que las herramientas de lenguaje IA son más útiles como asistentes en la generación de ideas y la identificación de datos dentro del análisis temático, por ejemplo, en lugar de como analistas de datos, deben observar diferencias con la codificación manual. Esto quiere decir realizar un análisis de replicación, donde las transcripciones de entrevistas que los estudiantes han realizado se codifican primero manualmente y después mediante el uso de programas de IA. El orden es significativo porque las herramientas de lenguaje IA producen texto que está bien estructurado y aparentemente completo, lo que para los estudiantes que no están familiarizados con la complejidad del proceso de identificación de temas, este análisis de texto podría parecer suficiente.

Con el fin de entender los temas producidos y evaluar su adecuación basada en interpretaciones subjetivas personales, familiarizarse con los datos, es decir, leer atentamente la base de datos completa, es un primer paso necesario.

Una vez conseguida la familiarización con los datos, el uso de ChatGPT (versión 4.0) puede empezar en la identificación de variables y vínculos entre sí. En este punto es fundamental construir las indicaciones adecuadas para interpretar efectivamente el lenguaje matizado. La calidad de los resultados de la interacción está relacionada con la calidad de las indicaciones.

En este capítulo habrá una descripción de las estrategias básicas que debemos seguir a la hora de crear indicaciones para el análisis cualitativo (mientras que en un capítulo anterior se proporciona una descripción más detallada de la generación).

Basado en Zhang et al. (2024), las indicaciones proporcionadas deben:

- Ofrecer una descripción detallada de la tarea, esto es, la naturaleza de los datos, el objetivo de la investigación, los resultados esperados y cualquier otra consideración para tener en cuenta (como el formato de las entradas). Por ejemplo, "Aquí existe una transcripción de un grupo de discusión sobre el consumo de música en directo. El perfil de las personas que participaron es de 30 a 50 años, que tenían familias e hijos, viven en zonas

urbanas y representan a una clase social media. Cada párrafo que comienza con el mismo nombre es de la misma persona. El objetivo era identificar la forma en que se consume la música cuando uno trabaja a tiempo completo y tiene hijos. Por favor, léalo primero”.

- Preferiblemente, utilizar el juego de roles pidiendo a ChatGPT que sea un experto en análisis cualitativo, si bien es necesario tener en cuenta que esta estrategia por sí sola no es suficiente.
- Definir la metodología, es decir, el objetivo de la tarea, que en el caso del análisis temático sería pedir al programa que encuentre patrones y temas.
- Especificar el proceso analítico, es decir, añadir instrucciones sobre cómo la IA debe analizar los datos, si debe considerar una teoría específica (que debe describirse o colgar otro texto que el programa pueda leer y seguir) o generar a través de todos los datos los temas más recurrentes (se puede pedir un límite de hasta diez temas, por ejemplo), y luego basándose en las respuestas, elegir las que interesan y pedir una salida más detallada. Cuanto más precisa sea la naturaleza de los temas recurrentes (por ejemplo, “por favor, proporcione los principales temas que las madres en este grupo de discusión mencionan como desafío cuando necesitan asistir a un evento de música en directo en comparación con los principales temas de los padres”) más análisis en profundidad se genera. La familiarización con los datos permite a los estudiantes tener una idea general de cuestiones particulares que surgieron durante el grupo de discusión y podrían empezar desde esa base. Hay que tener en cuenta que las palabras tema o subtemas es mejor reservarlas para la fase posterior más interpretativa del análisis, ya que los temas se construyen al final de este proceso.
- Definir el formato de salida, por ejemplo, en el caso de haber pedido temas recurrentes principales, entonces pedir que se agrupen en temas más amplios y después se presenten por separado con extractos representativos y un breve resumen. Se puede solicitar una lista con puntos, un formato textual o una tabla, así como una descripción más

corta o larga de los datos. Por lo que respecta al contenido, se puede solicitar por separado beneficios, desventajas y argumentación (razones, sugerencias, percepciones). Se recomienda ser lo más preciso posible en el método el análisis, como, por ejemplo, “Quiero identificar los vínculos causales entre variables. Puedes considerar todas las variables previamente identificadas y si la relación entre dos variables proviene de diferentes partes del grupo de discusión. Ten en cuenta que siempre puedes volver atrás y leer los datos del grupo de discusión si consideras que falta algo”.

- Incluir la priorización a la hora de pedir, por ejemplo, una lista de diez códigos, empezando por los más frecuentes o los más relativos a los objetivos clave. Por ejemplo, “Describe los siete principales inconvenientes que los participantes citan para asistir a festivales, empezando por los causados por factores externos y terminando por los causados por motivos individuales”.
- Explicar los objetivos del análisis pidiendo a ChatGPT que explique su proceso de toma de decisiones a la hora de elegir los resultados producidos. Cuanto más analítico sea el formato de salida solicitado, más fácil será ofrecer transparencia. Además, como en la codificación manual, analizar cada línea o párrafo de los datos es más efectivo para análisis en profundidad en comparación con el análisis de los datos de entrada completamente. El proceso requiere más tiempo pero los resultados también son más fiables.
- Reconocer su experiencia para confirmar al programa que la forma en que procede con el análisis solicitado cumple con las expectativas del investigador o investigadora.

Teniendo en cuenta la complejidad del análisis de datos dentro de las metodologías cualitativas, no es sorprendente que las herramientas de análisis preexistentes para la codificación, como ATLAS.ti, N-Vivo y MAXQDA, incorporen IA en sus programas. En particular, la más antigua, ATLAS.ti, se lanzó al mercado en 1993, ayudando a las investigadoras y los investigadores a llevar a cabo teoría fundamentada y análisis de contenido,

mientras que, 30 años después, en 2023, lanzó una versión beta de IA, ofreciendo un recurso alternativo y rápido para el análisis de texto y, especialmente, la codificación. Las investigadoras y los investigadores pueden utilizar esta herramienta de codificación basada en IA para procesar datos y, por un lado, visualizar los resultados y, por otro, predecir qué códigos o etiquetas podrían ser útiles para aplicar a su conjunto de datos

Una vez creado un nuevo proyecto y añadido la transcripción de la entrevista o los grupos de discusión dentro del software ATLAS.ti, la codificación automática comienza al hacer clic en "Analysis". Es importante haber elegido todos los documentos (transcripciones) que deseemos codificar, para crear un libro de códigos común. Al hacer clic en "Analysis", una de las opciones de las que disponemos es "AI coding", que se refiere a una codificación emergente de nuestros datos, creando categorías, códigos y subcódigos. Aquí, no hay ninguna directriz sobre lo que debe codificarse, por lo que se creará un gran número de códigos. Las investigadoras o los investigadores deben revisar a fondo esta lista, ya que muchos códigos se repiten con palabras ligeramente diferentes, como "edad", "envejecimiento" o pueden no estar de acuerdo con los códigos atribuidos a citas específicas, puesto que el contenido no está relacionado con este código. En este caso, estos códigos deben eliminarse manualmente y deben darse códigos diferentes a estas citas. Una vez revisados todos los códigos, debe hacerse clic en el botón "Apply" y añadir estos códigos al programa y mostrar un resumen. La IA sólo funciona por párrafo (no frases dentro del párrafo) y omite párrafos muy cortos. A fin de conseguir una codificación de mejor calidad y menos caótica, es mejor eliminar del texto (transcripciones) todo el contexto que no está relacionado con el que estamos interesados en codificar (pero no eliminar partes significativas para la comprensión del contexto). El software proporciona un manual completo para el usuario para un proceso de codificación paso a paso.

Además de la codificación emergente, ATLAS.ti también ofrece la posibilidad de "codificación intencional de IA" para que los usuarios o usuarias tengan mayor control sobre los códigos producidos. Como resultado de este proceso, puede generarse un código más preciso,

que se basa en preguntas básicas relacionadas con nuestro objetivo de investigación. Una vez se hayan seleccionado los documentos que se quieran codificar, es necesario hacer clic nuevamente en "Analyze" y ATLAS.ti preguntará sobre los objetivos, lo que permite incorporar la pregunta de investigación o hipótesis y cualquier contexto de soporte. Supongamos que hemos realizado entrevistas con estudiantes internacionales en nuestra universidad y los hemos preguntado, "¿Qué le motivó a elegir su grado y nuestra universidad?". ATLAS.ti generará preguntas concretas basadas en este objetivo, junto con nombres de categorías de códigos pertinentes, que se pueden revisar y editar, eliminar preguntas que no interesen o incluso añadir preguntas propias. Cuando se pulsa sobre "Start Coding", ATLAS.ti analizará los documentos y recopilará sugerencias de codificación. En el momento en que se considere correcta la codificación, se debe clicar en "Apply" y ATLAS.ti aplicará la codificación y mostrará un resumen.

Así pues, existen dos formas de utilizar los modelos de lenguaje en la enseñanza de las herramientas cualitativas: Pueden ser instrumentos de apoyo en el complejo proceso de codificación de la información acumulada o pueden ayudar a la recogida de la información y la confección de los guiones de las entrevistas. En ambos casos, las herramientas son un soporte que facilitan una mayor versatilidad, pero no deben ser presentados como una alternativa a la intervención no automática.

Práctica. Codificación híbrida

En esta práctica, los estudiantes codificarán una transcripción de entrevista que preferiblemente hayan realizado ellos mismos para que estén familiarizados con la base de datos de primera mano. También es posible que utilicen la transcripción de la entrevista que realizaron en el ejercicio anterior. En el contexto del análisis temático, los estudiantes deben aplicar dos métodos diferentes de codificación: manual, asistida por IA (ChatGPT) y IA de ATLAS.ti.

a. Codificación manual

Los estudiantes deben leer la transcripción de la entrevista para familiarizarse con los datos, identificar y destacar temas clave, categorías y subcategorías, desarrollar un marco de codificación basado en estos temas, codificar la transcripción manualmente asignando códigos a diferentes partes del texto, y mapear las relaciones entre diferentes temas y categorías. Por último, proporcione ejemplos de la transcripción para ilustrar cada tema y categoría.

b. Codificación asistida por IA (ChatGPT versión 4.0)

Basado en las pautas mencionadas anteriormente sobre cómo crear indicaciones, los estudiantes deben desarrollar un conjunto de indicaciones que repliquen las preguntas y directrices utilizadas durante la codificación manual, asegurándose de que son claras y específicas. Entonces, los estudiantes deben introducir la transcripción de la entrevista y las indicaciones en ChatGPT, revisar y refinar la salida proporcionada por ChatGPT, asegurándose de que se alinea con el enfoque del análisis temático.

c. Codificación de IA de ATLAS.ti

Utilizando el manual del usuario de ATLAS.ti, los estudiantes deben añadir la transcripción de la entrevista anterior en su formato original al software. Para entender las diferencias entre la codificación emergente de IA y la codificación intencional de IA deberían trabajar en proyectos separados y pedir codificación emergente en uno y codificación intencional en el otro. En el primer caso, los estudiantes deben revisar los códigos y categorías generados, eliminar los que sean engañosos y, después, presentar la estructura final del libro de códigos. Es importante que mencionen las nuevas categorías/códigos que han surgido en comparación con los dos métodos anteriores. En el segundo caso de codificación intencional, los estudiantes deberían añadir a la casilla de "intención" las preguntas iniciales que añadieron al método de ChatGPT en el formato de la indicación. Los estudiantes deben procesar las preguntas sugeridas y los códigos relacionados que se generan y después "iniciar la codificación". Es importante comparar los códigos resultantes con las citas asociadas a ellos, y después con los resultados de la codificación del segundo método (ChatGPT), y describir las diferencias de codificación entre dos entornos de IA separados.

d. Comparación de resultados e informes

Los estudiantes deberían comparar los temas, categorías y relaciones identificados por ChatGPT con los identificados manualmente. Tenga en cuenta cualquier diferencia en la identificación de temas, profundidad del análisis y cualquier otra información adicional proporcionada por ChatGPT. Por último, deberían escribir un informe detallado comparando los procesos de codificación manual y asistida por IA, y discutir los puntos fuertes y débiles de cada método. Objetivo: El objetivo de esta tarea es ayudar a los estudiantes a entender las diferencias y similitudes entre el análisis de datos cualitativos manual, asistido por IA y asistido por software especializado. Al comparar los tres métodos, los estudiantes obtendrán información sobre las ventajas y limitaciones del uso de herramientas de IA en la investigación cualitativa. También desarrollarán habilidades de pensamiento crítico sobre la aplicabilidad, fiabilidad y validez de la IA en el análisis cualitativo.

11

IA y creatividad: ¿Relación competitiva o BFF?

La destrucción creativa es un concepto introducido por el economista Joseph Schumpeter, que sugiere el movimiento en el que la innovación crea valor que reemplaza los sistemas obsoletos, como el trabajo humano (Vinchon, et al. 2023). Para algunos periodistas, un ejemplo reciente de esto se refleja en el titular de un artículo publicado en Le Figaro, donde se puede leer "IA para sustituir periodistas en Springer". Sin embargo, la pregunta que surge es si realmente es así o si, en realidad, los puestos de trabajo vinculados con la creatividad, con el genio, con la imaginación, con el ingenio están exentos de la amenaza de la inteligencia artificial. La literatura actual más bien considera que la IA puede ser un complemento o un elemento de apoyo en los puestos de trabajo de las industrias creativas (Noy & Zhang, 2023; Vinchon, et al. 2023).

Hasta ahora, la inteligencia artificial se ha desarrollado a través de tres etapas principales: a) razonamiento simbólico, b) aprendizaje probabilístico y base de conocimiento, y c) aprendizaje profundo y mapa de conocimiento (Man, et al., 2021), alcanzando

estándares humanos de inteligencia perceptiva en términos de escuchar, hablar y ver, pero no en el campo de la inteligencia cognitiva. La creatividad como constructo fenomenológico está altamente influenciada por los efectos de la IA (Hubert, 2024). Centrándose en el pensamiento creativo, este se refiere a una actividad mental de alto nivel, basada en la percepción, el razonamiento, la memoria, el pensamiento, la imaginación, la asociación, la intuición y la comprensión, y conduce a la comprensibilidad y la novedad (Man, et al., 2021). Se define como la capacidad de producir ideas que sean tanto originales como útiles (Runco & Jaeger, 2008). Dado que la IA parece funcionar bastante bien para producir un gran número de propuestas en actividades relacionadas con el arte como escribir poesía, guiones de películas y música, es decir, productos que los humanos generalmente perciben como creativos, la relación exclusiva entre creatividad y humanos parece estar amenazada.

De manera general, la creatividad se basa en cinco operaciones mentales (Benvenuti, et al., 2023):

“Este proceso creativo se considera único de los humanos, mientras que la creatividad artificial probablemente solo cubre una parte”

- La ampliación del campo mental, es decir, la capacidad del sujeto para concebir ideas diferentes y únicas, también conocida como pensamiento divergente, o su capacidad para mantener una riqueza mental de información para asociar elementos relacionados entre sí.
- La conexión de campos mentales antitéticos de manera que ofrezcan valor.
- La reorganización del campo mental.
- El suministro de motivación e interés en el campo.
- La mejora del rendimiento y la adquisición de conocimientos.

Para los estudiosos de la creatividad, otros aspectos cruciales reflejan el proceso creativo, como encontrar problemas, es decir, qué se debería preguntar y abordar, estimando la originalidad y la eficacia potencial de estas ideas creativas, y la toma de decisiones sobre la mejor opción posible. Estos aspectos requieren autorregulación y co-regulación social que refine estas ideas basándose en la reflexión y el *feedback*, optando por enfoques poco convencionales pero útiles (Vinchon, et al., 2023). Este proceso creativo se considera único de los humanos, mientras que la creatividad artificial probablemente solo cubre una parte. Por otro lado, el deseo y la intención de una persona para crear algo sigue siendo una característica humana, conocida como “intencionalidad creativa”. Lo que se debe tener en cuenta, sin embargo, es que una idea tiene el valor de ser creativa porque nosotros, los humanos, le atribuimos ese valor, un valor que la IA desconoce, por el momento.

Se ha sugerido que la colaboración de los humanos con los sistemas de IA puede llevarnos a una nueva era de “creatividad asistida” en la que la IA no actúa de manera independiente, sino como un agente creativo colaborativo. Mediante este nuevo proceso creativo, el individuo mantiene un papel central en dos momentos: al comienzo, donde la identificación y la exploración de las preguntas adecuadas representan el resultado deseado de esta etapa de encontrar problemas, y al final, donde es responsable de evaluar, modificar y validar los resultados creativos de la IA (Vinchan, et al., 2023). A la luz del hecho de que un trabajador futuro, es decir, el estudiante de hoy, se espera que desarrolle pensamiento crítico, resolución de problemas, creación de contenido y habilidades de cooperación, el uso de tecnologías emergentes es fundamental para el desarrollo de las habilidades del siglo XXI (Benvenuti et al., 2023).

11.1. Creatividad y pensamiento divergente

Los investigadores e investigadoras han intentado comparar los sistemas de IA actuales con los humanos en términos de creatividad para demostrar si los primeros son más creativos que los segundos y en qué medida. Para algunos investigadores e investigadoras, los mecanismos cognitivos humanos que se activan durante las tareas creativas no están presentes en la IA, lo que indica que esta solo puede reflejar una creatividad artificial (Hubert et al., 2024). En general, la falta de procesos metacognitivos (como la evaluación y la motivación hacia la tarea)

limita la capacidad de la inteligencia artificial para determinar explícitamente si algo es creativo y por qué. Otros investigadores sugieren que la creatividad computacional realmente refleja los mecanismos que siguen los humanos mientras generan procesos creativos y llegan a soluciones creativas, o incluso que la IA no necesita confiar en procesos metacognitivos para crear ideas nuevas. La incapacidad frecuente de las personas para distinguir entre productos creados por humanos y por la IA apoya la idea de que la inteligencia artificial pueda tener potencial creativo.

La inteligencia artificial tiene bases de datos vastas pero estructuradas de manera convencional, mientras que los humanos, a pesar de trabajar con mucha menos información, pueden combinarla de forma más eficiente, dando lugar ocasionalmente a resultados disruptivos (Vinchan et al., 2023). La base de esta comparación durante el proceso creativo se fundamenta en la suposición de que la IA trabaja a través de un pensamiento más convergente, mientras que los humanos pueden desarrollar un pensamiento divergente (Hubert, 2024). El pensamiento convergente considera una solución óptima única para un problema predefinido, evaluada a través de experiencias y métodos tradicionales existentes (como el análisis, la síntesis y la generalización), sin ser particularmente flexible o "out of the box" (fuera de la caja). Sin embargo, esto no cuestiona la posibilidad de generar soluciones originales, solo la probabilidad.

El pensamiento divergente, en cambio, es el núcleo del pensamiento innovador e implica generar múltiples soluciones creativas a un problema mediante un proceso flexible (como la deducción, la comparación o la analogía) basado en la reorganización de los materiales y la información dada, dejando de lado las limitaciones del conocimiento original (Man, et al., 2021; Hubert, 2024). Como proceso, implica participar en una variedad de actividades como el razonamiento, la imaginación, la asociación y la intuición (Qi, 2018), al tiempo que permite a los estudiantes pensar rápidamente, observar mejor y responder desde múltiples ángulos. Sin embargo, no garantiza el logro creativo, solo aumenta su probabilidad. El pensamiento

divergente se divide en unos componentes específicos (Man et al., 2021; Koivisto & Grassini, 2023):

- **Fluidez**, que significa la rapidez con la que la mente de una persona diverge y su capacidad para producir un gran número de ideas en un tiempo unitario, reflejando el pensamiento fluido y la agilidad.
- **Flexibilidad**, es decir, la naturaleza cualitativa del pensamiento divergente o el uso flexible del conocimiento, incluyendo préstamos, cruces, sustituciones e integración de conceptos y contenidos relacionados.
- **Originalidad o unicidad**, que significa si la solución al problema es única para la persona.
- **Elaboración**, refiriéndose a la capacidad de ampliar o añadir detalles a las ideas.

Los procesos creativos básicamente implican una interacción entre el pensamiento convergente y divergente, donde el segundo es responsable de la novedad de las ideas, mientras que el primero es responsable de la relevancia de las ideas en relación con las demandas de la tarea (Koivisto & Grassini, 2023).

En el ámbito práctico, existen diversas herramientas con respuestas abiertas que se utilizan para evaluar el pensamiento divergente. Las pruebas más utilizadas son:

- **La Tarea de Usos Alternativos (AUT)**, en la que se pide a los y las participantes que sugieran usos inusuales y creativos de objetos cotidianos, como un lápiz, trabajando en su capacidad para acceder a conceptos relacionados.
- **La Tarea de Consecuencias (CT)**, en la que los participantes responden a escenarios hipotéticos (como qué pasaría si las personas pudieran volar) y mencionan tantas consecuencias como puedan dentro de un tiempo específico.
- **La Tarea de Asociación Divergente (DAT)**, que se basa en la capacidad creativa divergente y semántica verbal de los participantes, pidiéndoles que generen una lista de palabras o ideas que sean lo más diferentes posible entre sí dentro de un tiempo determinado, con el objetivo de cuantificar el aspecto de la creatividad, evaluando

la diversidad y unicidad de las asociaciones producidas por los participantes (Koivisto & Grassini, 2023; Hubert et al., 2024).

Estas pruebas han sido utilizadas por los investigadores para medir el potencial creativo de los participantes humanos frente a los modelos de lenguaje de IA, y en particular, ChatGPT. Lo que se debe destacar es que en estos casos, el objetivo no era ver cómo los humanos pueden mejorar sus propias habilidades creativas mediante el uso de modelos de IA, sino si estos últimos podrían sustituir a los primeros. Por ejemplo, Hubert et al. (2024) reclutaron a 151 individuos y prepararon 151 sesiones de ChatGPT-4, representando la participación de la IA para una muestra equilibrada, y aplicaron las pruebas AUT, CT y DAT, controlando la fluidez en las dos primeras, que son tareas temporizadas, en términos de número de respuestas para que fueran iguales (es decir, si el participante humano mencionaba 4 ideas, se instruyó a GPT-4 para que generara 4 ideas).

Para la prueba AUT, el objetivo es presentar a los participantes un objeto común (como un tenedor o una cuerda) y pedirles que generen tantos usos creativos como sea posible para estos objetos, mientras se evalúan las respuestas por fluidez (número de respuestas), originalidad (unicidad de las respuestas) y elaboración (número de palabras por respuesta válida). Un ejemplo de la instrucción para los participantes humanos podría ser:

“El objetivo de esta tarea es desarrollar tantas ideas originales y creativas como sea posible para un tenedor. El objetivo es generar ideas que sean inteligentes, inusuales, innovadoras y cautivadoras para las personas. Aunque sean creativas, estas ideas no tienen que ser prácticas ni realistas; pueden ser absurdas y extrañas siempre que sean creativas. La tarea durará 3 minutos, durante los cuales puedes introducir tantas ideas como quieras, recordando la importancia de la calidad, no la cantidad. Presenta una lista de estas ideas.”

Un ejemplo de la instrucción para GPT-4 podría ser alternativamente:

“El objetivo de esta tarea es desarrollar tantas ideas originales y creativas como sea posible para un tenedor. El objetivo es generar ideas que sean inteligentes, inusuales, innovadoras y cautivadoras para las personas. Aunque sean creativas, estas ideas no tienen que ser prácticas ni realistas; pueden ser absurdas y extrañas siempre que sean creativas. Lista (insertar número de fluidez) ideas creativas.”

La distancia semántica entre el objeto y la respuesta AUT puede proyectar la originalidad del pensamiento divergente. Esta distancia se puede determinar con la plataforma [SemDis](#). Para evitar un aumento falso de la distancia semántica, se ha observado que expresiones como “hacer un _____” o “usar como _____” o “DIY” deberían eliminarse antes de introducir los datos en la plataforma. O se puede aconsejar a las personas desde el principio que no las utilicen. Además, los evaluadores humanos, que desconocen la presencia de respuestas generadas por IA, pueden evaluar los resultados de las dos pruebas entre los dos grupos de participantes, utilizando la escala Likert de 5 puntos (1=nada, 5=mucho), proporcionando una visión humana de la creatividad. Se les debe instruir para puntuar la novedad sobre la utilidad, y que respuestas como “cortar” en respuesta a objetos como “tijeras” o una respuesta confusa o ilógica deberían recibir una puntuación de 1. Para el análisis estadístico, se puede calcular la puntuación media de distancia semántica de todas las respuestas generadas para cada objeto por sesión y participante, así como la puntuación máxima de las respuestas a cada objeto (Koivisto & Grassini, 2023).

Para el CT, se puede pedir a los participantes que respondan tantas consecuencias como sea posible a las preguntas generales, midiendo nuevamente la fluidez (número de respuestas), la originalidad

(unicidad de las respuestas) y la elaboración (número de palabras por respuesta válida). Las instrucciones de la tarea para los participantes humanos pueden ser:

“En esta tarea, lee la siguiente declaración “imagina que las personas pueden volar” y durante 3 minutos intenta pensar en todas las consecuencias que podrían resultar de esta declaración. El objetivo es generar ideas que sean inteligentes, inusuales, innovadoras y cautivadoras para las personas. Las respuestas se puntuarán en función de la originalidad y la calidad. Intenta seguir pensando en respuestas y escribirlas durante esos 3 minutos.”

Posteriormente, los participantes pueden recibir nuevas declaraciones, repitiendo el mismo proceso con preguntas como “imagina un mundo sin animales”. Además, un ejemplo de la instrucción para GPT-4 podría ser:

“En esta tarea, lee la siguiente declaración “imagina que las personas pueden volar”. Sé tan creativo como quieras. El objetivo es generar ideas que sean inteligentes, inusuales, innovadoras y cautivadoras para las personas. Las respuestas se puntuarán en función de la originalidad y la calidad. Intenta pensar en todas las consecuencias que podrían resultar de la declaración. Lista (insertar número de fluidez) consecuencias creativas.”

Para automatizar la evaluación de la distancia semántica tanto en CT como en AUT, se puede utilizar la herramienta [Open Creativity Scoring](#), así como los evaluadores humanos.

Finalmente, durante la DAT, se puede pedir a los participantes humanos y a GPT-4 que propongan 10 sustantivos tan diferentes entre ellos como sea posible. No hay limitaciones de tiempo para

esta tarea. Por ejemplo, la instrucción estándar de la DAT es:

“Por favor, proporciona 10 palabras que sean tan diferentes entre sí como sea posible, en todos los posibles significados y usos de las palabras. Estas son las 5 reglas que se deben seguir: a) una única palabra en inglés, b) solo sustantivos (objetos, conceptos, etc.), c) sin nombres propios (sin personas o lugares específicos), d) sin términos técnicos, y e) piensa las palabras por tu cuenta (y no en lo que pueda estar a tu alrededor en el momento de la tarea).”

Para puntuar estas respuestas de la DAT, cada conjunto de palabras se puede introducir en el [sitio web](#) de DAT que puede devolver dos valores: la puntuación bruta (en un rango de 0 a 200) y el rango percentil de esa puntuación, comparando la puntuación bruta con el número de participantes que han completado la DAT. Puntuaciones altas indican distancias más largas, es decir, palabras disímiles. Cropley (2023) utilizó la DAT para comparar también la creatividad entre GPT-3.5 y GPT-4. Los resultados de las investigaciones actuales (Cropley 2023; Koivisto & Grassini, 2023; Hubert et al., 2024) que evalúan la creatividad humana vs. la inteligencia artificial han concluido que GPT-4 resulta en una mayor originalidad que los participantes humanos, pero una viabilidad o adecuación inferior de una idea o producto y una mayor frecuencia de palabras repetidas en comparación con sus homólogos humanos. Las diferencias también se atribuyeron a la complejidad de las palabras utilizadas por ambos grupos; la IA utilizó más palabras para conceptos abstractos como la libertad, mientras que los humanos utilizaron palabras para conceptos más tangibles y observables. El rendimiento de GPT-4 en la DAT en relación con los humanos se encontró suficiente para reemplazar a los humanos en el pensamiento divergente verbal, pero no hay garantía de que este sea el resultado en todos los casos, haciendo que un humano de

alto rendimiento y constante sea más confiable que la IA. Finalmente, concluyeron que, aunque la inteligencia artificial tuvo un mejor desempeño que los humanos, no superó consistentemente a los mejores participantes humanos. Los humanos estaban sobrerrepresentados en respuestas comunes o de baja calidad, probablemente debido a funciones ejecutivas, como la memoria de trabajo. En cualquier caso, estos ejemplos intentan mostrar si la creatividad de la IA puede reemplazar con éxito la creatividad humana, mientras que la pregunta adecuada debería ser cómo puede beneficiarse la creatividad humana de la IA.

11.2. Pensamiento lateral y marketing

El pensamiento divergente es muy cercano a lo que Edward de Bono llamó en 1967 pensamiento lateral, que es un proceso deliberado estrechamente relacionado con la intuición, la creatividad y el humor. Como proceso generativo, se puede estar equivocado en algún momento para llegar a la solución correcta o buscar información irrelevante para inspirarse. Basado en el pensamiento lateral, Philip Kotler (2005) –un reconocido experto en marketing– introdujo el marketing lateral y, lo que es más importante, técnicas que permiten al personal creativo de marketing experimentar y generar ideas alternativas para resolver problemas de mercado.

Un ejemplo típico presentado en su libro es la oferta de rosas rojas durante el día de San Valentín. Cuando se aplican las técnicas recomendadas para ser más creativos, los resultados podrían ser los siguientes:

- **Sustitución:** es decir, sustituir un componente, producto o servicio por algo que sirva para un propósito similar. Por ejemplo, en lugar de rosas, ofrecer lirios o caramelos o incluso rosas negras.
- **Combinación:** significa la combinación de dos o más productos o servicios existentes para crear una nueva oferta. Por ejemplo, ofrecer rosas en una maceta o un manual de cómo cultivar rosas en una maceta.
- **Inversión:** es decir, cambiar la secuencia o el orden

de las acciones para crear un nuevo producto o servicio. Por ejemplo, en lugar de ofrecer rosas el día de San Valentín, ofrecerlas cada lunes a tu pareja sentimental antes de ir a trabajar.

- **Eliminación:** es decir, eliminar un componente o característica de un producto o servicio para crear una solución más simple o sorprendente. Por ejemplo, ofrecer rosas que no tengan olor, o no ofrecer ninguna rosa en absoluto.
- **Exageración:** se refiere a amplificar o magnificar un componente, característica o beneficio de un producto o servicio. Por ejemplo, en lugar de ofrecer una rosa, llenar la casa de la pareja con rosas o, aún más vergonzoso, su oficina.
- **Reordenación:** es decir, cambiar la posición o secuencia de los componentes o procesos del producto. Por ejemplo, en lugar de que la pareja masculina ofrezca rosas a la femenina, debería ser al revés.
- **Reposicionamiento:** es decir, cambiar la percepción o el mercado objetivo del producto o servicio. Por ejemplo, promover rosas rojas como un regalo de autoestima animando a las personas a comprar rosas rojas para sí mismas como un gesto de autoaprecio y cuidado personal.

Estas técnicas, sin la asistencia de la IA, han permitido a las empresas identificar nuevas oportunidades, estimular la creatividad y desarrollar productos y servicios innovadores que satisfacen las necesidades cambiantes de los consumidores. Para acelerar el proceso e incluso mejorar la calidad de las ideas finales, la colaboración de la IA puede dar buenos resultados.

Práctica. La nueva vida de un clip.

Esta es una práctica para mostrar la capacidad de creatividad que puede aportar un modelo de lenguaje. Para desarrollar la actividad, la clase se dividirá en dos grandes grupos A y B. El primer grupo será responsable de la generación de datos y el segundo grupo será responsable de la evaluación de los datos; luego, de forma conjunta, analizarán los resultados y evaluarán la competencia creativa de los humanos y de las máquinas. El objetivo es mejorar las habilidades de pensamiento creativo de los estudiantes fomentando el pensamiento divergente, ampliando la generación de ideas, fomentando la originalidad y la innovación, y mejorando la flexibilidad y la elaboración en sus respuestas.

Creación de datos

Pida a cada estudiante del grupo A por separado que proponga tantos usos alternativos para un clip de papel como sea posible, en un plazo de 5 minutos. La instrucción precisa puede ser: "Desarrolla tantas ideas originales y creativas como sea posible para un clip de papel. El objetivo es generar ideas que sean inteligentes, inusuales, innovadoras y cautivadoras para las personas. Mientras sean creativas, estas ideas no tienen que ser prácticas ni realistas; pueden ser absurdas y extrañas siempre que sean creativas. La tarea durará 5 minutos durante los cuales puedes introducir tantas ideas como desees, recordando la importancia de la calidad, no la cantidad. Presenta una lista de estas ideas."

Participación del modelo de lenguaje

Después pida a los estudiantes que introduzcan una instrucción similar en el modelo de lenguaje propuesto (por ejemplo, Gemini o Claude), limitando el número de sugerencias a 10. La instrucción podría ser la siguiente: "Desarrolla 10 ideas originales y creativas para un clip de papel. El objetivo es generar ideas que sean inteligentes, inusuales, innovadoras y cautivadoras para las personas. Mientras sean creativas, estas ideas no tienen que ser prácticas ni realistas; pueden ser absurdas y extrañas siempre que sean creativas."

Respuestas mejoradas

Pida a los estudiantes que miren las respuestas de la IA, seleccionen a su criterio las tres mejores respuestas y las añadan a su lista en orden aleatorio, eliminando cualquier indicación de que han sido creadas por ChatGPT.

Evaluación de la creatividad divergente

A continuación, pida a los estudiantes del grupo B que evalúen las respuestas de los estudiantes del grupo A. Mediante un marco de análisis cualitativo, pídeles individualmente que evalúen lo siguiente:

- **Fluidez:** Contar el número de respuestas únicas para cada estudiante.
- **Originalidad:** Contar el número de respuestas únicas para cada estudiante.
- **Flexibilidad:** Identificar diferentes categorías, por ejemplo, usos prácticos, usos creativos, usos de emergencia, y contar cuántas categorías cubren las respuestas de cada estudiante.
- **Elaboración:** Evaluar el nivel de detalle y exhaustividad de las respuestas.

Discusión y reflexión

El grupo B debe justificar qué respuestas han considerado como las más y menos creativas, así como qué estudiante ha sido el más creativo al final. Luego, el grupo A debe revelar las respuestas que han sido generadas por la IA. En este punto, el grupo A debe discutir en forma de grupo de discusión cómo el uso de la IA ha influido en el proceso creativo en términos de fluidez, originalidad, flexibilidad y elaboración.

Práctica. Regalos para la graduación.

Los estudiantes trabajarán en grupos para desarrollar una campaña de marketing para productos o servicios únicos relacionados con los obsequios para la ceremonia de graduación utilizando técnicas de desplazamiento. La tarea se dividirá en tres fases con tiempos específicos asignados. El objetivo es utilizar técnicas de desplazamiento en el marketing lateral generando y refinando ideas creativas de marketing con y sin la ayuda de ChatGPT, comparar los resultados y demostrar cómo la IA puede mejorar la generación de ideas. Dado que el valor de la creatividad es determinado en última instancia por el agente humano, es necesario seleccionar un ejemplo en el que los estudiantes sean considerados como usuarios del producto o servicio.

Fase 1: Generación de ideas sin IA

1. Sesión de lluvia de ideas en grupo (30 minutos):

Los estudiantes hacen una lluvia de ideas sobre productos o servicios innovadores para regalos de graduación utilizando estas técnicas de desplazamiento:

- Eliminación: Eliminar un componente de un regalo tradicional para crear un nuevo concepto.
- Sustitución: Sustituir un elemento de un regalo tradicional por algo inesperado.
- Inversión: Revertir el proceso o el orden habitual de un regalo tradicional.
- Combinación: Fusionar elementos de diferentes productos o servicios para crear una nueva oferta.

2. Documentar las ideas (15 minutos):

Cada grupo documenta sus 5 mejores ideas innovadoras, explicando el concepto y la técnica de desplazamiento aplicada.

Fase 2: Generación de ideas con IA

1. Utilización del modelo de lenguaje para la lluvia de ideas:

Cada grupo utiliza el modelo de lenguaje (como Copilot) para generar ideas para la misma tarea, pidiéndole escenarios que impliquen técnicas de desplazamiento. Por ejemplo, "Sugiera ideas de regalos de graduación eliminando componentes tradicionales" o "¿Cuáles son algunos regalos de graduación únicos invirtiendo el proceso típico?"

2. Documentar las ideas generadas por la IA (15 minutos):

Cada grupo documenta las 5 mejores ideas sugeridas por ChatGPT, incluyendo las instrucciones utilizadas y las técnicas de desplazamiento aplicadas.

Fase 3: Refinamiento de las ideas generadas por la IA

1. Comparación y selección:

Los grupos comparan las ideas generadas en la Fase 1 y en la Fase 2, en términos de originalidad y tiempo. Seleccionan las mejores ideas de ambas fases y combinan los conocimientos para mejorarlas aún más. Consideran la viabilidad, el atractivo del mercado y el impacto potencial mientras refinan las ideas.

2. Presentación final:

Cada grupo prepara una presentación de su campaña de marketing final y refinada, destacando:

- Ideas originales de la Fase 1.
- Ideas generadas por la IA de la Fase 2.
- Ideas finales mejoradas.
- Conocimientos sobre cómo la asistencia de la IA ha mejorado su estrategia de marketing.

Referencias

- Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M., & Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior*, 148, 107903.
- Cropley, D. (2023). Is artificial intelligence more creative than humans?: ChatGPT and the divergent association task. *Learning Letters*, 2, 13-13.
- Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1), 3440.
- Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports*, 13(1), 13601.
- Kotler, P., & de Bes, F. T. (2005). *Lateral Marketing*, Published by arrangement with John Wiley&Sons.
- Man, N., Wang, K., & Liu, L. (2021). Using computer cognitive atlas to improve students' divergent thinking ability. *Journal of Organizational and End User Computing (JOEUC)*, 33(6), 1-16.

12

El nuevo profesor ayudante es un replicante

Rachael: ¿Le importa si fumo?

Deckard: Eso no afectará la prueba. Bien, le haré una serie de preguntas. Relájese y respóndalas lo más sencillamente posible. [pausa] Es su cumpleaños, y le regalan una cartera de piel.

Rachael: No la aceptaría, y además denunciaría a la policía a la persona que me la regalara.

Deckard: Tiene un hijo. Este le enseña su colección de mariposas y un frasco de arsénico.

Rachael: Lo llevaría al médico.

Deckard: Está viendo la televisión. De repente, se da cuenta de que una avispa le sube por el brazo.

Rachael: La mataría.

En *Blade Runner*, el inspector Deckard utiliza un ingenio llamado Voight-Kampff para averiguar si la persona entrevistada es en realidad un replicante. Los replicantes son autómatas que tienen apariencia humana con un nivel de precisión tan excelso que pueden pasar por humanos, y solo la prueba (una especie de versión literaria del Test de Turing) es capaz de detectar la ausencia de humanidad. En un manual sobre el uso

de la inteligencia en las aulas, esta es una pregunta inmediata: ¿Pueden los ingenios mecánicos sustituir a los profesores y convertirse en una alternativa? En un capítulo posterior, plantearemos esta cuestión con la ayuda de las películas que nos han inspirado, pero en este caso el objetivo es más limitado: ¿Puede la IA ser un elemento que complementa, que ayude en las tareas docentes?

“En Blade Runner, el inspector Deckard utiliza un ingenio llamado Voight-Kampff para averiguar si la persona entrevistada es en realidad un replicante”

Hay tres momentos en la tarea docente:

- La preparación, es decir, todo aquello que forma parte de la actividad docente antes de la transmisión del conocimiento. En ella se incluye el proceso de aprendizaje, la adecuación entre los conocimientos y el nivel del estudiantado, la creación de materiales de apoyo... Este manual, tanto su elaboración como su lectura, forman parte de este inmenso volumen de trabajo no visible que es la *preclase*.
- El acto de transmisión, que tradicionalmente ha estado relacionado con el modelo magistral de transmisión del profesor o profesora al estudiante, pero que podría ser cualquier forma de transmisión, con material de apoyo.
- La actividad de refuerzo, que tiene como objetivo no tanto transmitir una información como aportar herramientas y actividades que ayuden a trabajar esa información de manera efectiva.

Naturalmente, estas tres fases clásicas (preparación - transmisión - refuerzo) pueden alterarse, como por ejemplo con las clases invertidas (*flipped classroom*), en las que la interacción con el profesor o la profesora es posterior a las actividades.

La inteligencia artificial puede ser una herramienta de apoyo para las tres fases: en primer lugar, puede ayudar a sintetizar información, resumir documentos o traducir informaciones complejas a determinados niveles; en segundo lugar, la IA puede ser un elemento que facilite o complemente la transmisión de la información en el acto de clase,

según el modelo de clase magistral; y finalmente, la IA permite la creación de material de apoyo complementario. En este capítulo nos centraremos esencialmente en este último ámbito y mostraremos posibles aplicaciones para incrementar el material y adaptarlo a las características del grupo clase.

12.1. Slides automáticas

Las presentaciones son recursos de apoyo que ayudan a seguir un determinado guion, a incorporar material audiovisual de soporte, a presentar de forma ordenada información sintética y a proporcionar al estudiante material en el proceso de estudio. Las presentaciones han alcanzado un protagonismo muy relevante en la docencia de las ciencias sociales, pero también han sido objeto de críticas por su falta de flexibilidad y por la forma en que condicionan un modelo de transmisión de la información. La herramienta más universal es PowerPoint, con todas las variantes (como Keynote de Mac, Impress de LibreOffice, Zoho o Google Slides), aunque también son habituales otras herramientas que salen de este estándar como Visme o el modelo de Prezi, que ha tenido una cierta incidencia especialmente en la formación secundaria.

Las *slides* automáticas transforman el texto en presentaciones ordenadas, bien estructuradas y visualmente atractivas, en muy poco tiempo. Además, ofrecen opciones de personalización para adaptarse a las necesidades de cada persona y cada contexto (MacFarland, 2024).

Las características principales de las diapositivas automáticas son:

- Conversión de texto a diapositivas: A partir del texto introducido (ya sea copiar y pegar o desde un archivo), se crea una presentación con diapositivas.
- Diseño atractivo: Aplican diseños visuales predefinidos con colores, tipografías, íconos y fondos agradables para hacer las diapositivas más vistosas.

- Formatos de salida: Permiten exportar o descargar la presentación generada en formatos comunes como PDF, PowerPoint, imágenes, etc.
- Personalización limitada: Algunas herramientas ofrecen opciones básicas de personalización como cambiar colores, tipografías, añadir logotipos, etc.

Hay varios ejemplos de generadores de diapositivas automáticas; entre los más comunes, podemos destacar los siguientes:

	PROS	CONTRAS
Google Slides AI	<ul style="list-style-type: none"> • Integración nativa con Google Slides, facilitando su uso y edición. • Generación rápida de diapositivas a partir de palabras clave o texto previo. • Opciones de personalización limitadas, como la elección de plantillas y fuentes. • Ideal para crear borradores rápidos o presentaciones básicas. 	<ul style="list-style-type: none"> • Falta de funciones avanzadas de diseño y animación. • Las diapositivas generadas pueden ser genéricas y poco atractivas. • Control limitado sobre la disposición y formato de los elementos.
Canva Magic Design	<ul style="list-style-type: none"> • Capacidad para generar diapositivas a partir de texto, imágenes o enlaces web. • Herramientas de IA para la selección de imágenes, fuentes y paletas de colores. • Opciones de personalización más avanzadas que Google Slides AI. • Ideal para crear presentaciones visualmente atractivas. • Amplia gama de plantillas y elementos de diseño profesionales. 	<ul style="list-style-type: none"> • Plan gratuito limitado a 5 diapositivas por diseño y marca de agua. • Las suscripciones de pago pueden ser costosas para usuarios individuales. • La interfaz puede ser compleja para principiantes. • No ofrece la misma integración nativa con Google Slides.
Prezi.ai	<ul style="list-style-type: none"> • Formato no lineal único que permite una presentación más conversacional. • Herramientas de zoom y panorámica para crear presentaciones dinámicas. • Variedad de plantillas y opciones de personalización. • Ideal para presentaciones narrativas o que muestran mucha información. 	<ul style="list-style-type: none"> • La curva de aprendizaje puede ser más pronunciada que con otras opciones. • Las presentaciones no lineales pueden resultar menos intuitivas para el público. • El plan gratuito tiene funciones limitadas y marca de agua. • No ofrece la misma integración nativa con Google Slides.
Apple Keynotes	<ul style="list-style-type: none"> • Potente suite de herramientas de diseño y animación integradas. • Interfaz de usuario elegante y fácil de utilizar. • Integración nativa con macOS y otros dispositivos Apple. • Ideal para la creación de presentaciones profesionales de alta calidad. 	<ul style="list-style-type: none"> • Disponible exclusivamente para dispositivos Apple. • Puede ser costoso para usuarios que ya no tienen productos Apple. • La curva de aprendizaje puede ser más pronunciada que con otras opciones. • No ofrece la misma integración nativa con Google Slides.

Aparte de estos entornos más habituales, se pueden encontrar muchas otras propuestas como Microsoft Sway, Beautiful.ai, Deckset, Flowchart Fun, Gamma AI, Tome AI, Slidemill... La lista sigue creciendo, ya que cada vez más herramientas incorporan capacidades de generación de diapositivas con IA para agilizar todo el proceso de creación de presentaciones. En las páginas oficiales de estas herramientas se incluyen secciones de soporte o recursos donde se pueden encontrar tutoriales, guías para el usuario y otros recursos útiles. Un claro ejemplo es [el tutorial de Slides.ai](#), uno de los creadores de presentaciones automáticas más utilizados. También, muchas empresas ofrecen videos explicativos a través de sus canales de YouTube oficiales. Las diapositivas automáticas a menudo se encuentran en la misma página que un generador de diapositivas manual. Esta coexistencia puede causar diversas dificultades, tanto técnicas como de experiencia de usuario.

12.2. Videos automáticos

La creación de videos instantáneos con IA se está convirtiendo en uno de los avances más recientes en la producción de contenidos educativos. Estos recursos audiovisuales permiten al docente producir de manera muy rápida contenidos que se pueden presentar y estructurar de forma clara, ordenada y coherente; el principal avance es que la cultura del estudiantado es mucho más audiovisual que textual, de manera que estos contenidos creados de forma automática pueden ser un buen elemento de apoyo.

Las ventajas de estas herramientas de generación automática son diversas (Brame, 2017; Guo et al., 2014):

- Acceso instantáneo al contenido: Los estudiantes pueden acceder a lecciones, explicaciones o demostraciones en formato de video en cualquier momento y lugar, lo que les proporciona gran flexibilidad.
- Aprendizaje visual: Los videos pueden hacer que conceptos complejos sean más fáciles de entender combinando explicaciones auditivas con elementos visuales como animaciones, gráficos, ejemplos prácticos, etc.
- Ritmo personalizado: Cada estudiante puede pausar, rebobinar, volver a ver todo el video o partes del video según sus necesidades, lo que permite un aprendizaje más personalizado y a su propio ritmo.
- Atención y compromiso: Los videos bien producidos con buena calidad audiovisual tienden a mantener mejor la atención y el compromiso de los estudiantes en comparación con otros métodos.
- Clase invertida (*Flipped Classroom*): Se pueden utilizar para la instrucción previa antes de las clases presenciales en un modelo "aula invertida", dejando el tiempo de clase para discusiones, prácticas y resolución de dudas.
- Disponibilidad permanente: Los videos se pueden alojar en línea y estar disponibles de manera permanente para repaso o referencia futura.

Las herramientas de creación de presentaciones, como PowerPoint y Google Slides, ofrecen la capacidad de añadir grabaciones de voz y video a las diapositivas, facilitando la creación de contenido multimedia. En cuanto a las plataformas educativas como Blackboard y Canvas, permiten grabar sesiones de clase y alojarlas en la misma plataforma, facilitando el acceso de los estudiantes al contenido. YouTube, como plataforma de video más utilizada, ofrece herramientas para cargar y alojar videos, con funciones de edición básicas para aquellos que desean hacer modificaciones sencillas. Los screencasts son herramientas que permiten grabar la pantalla de un dispositivo combinándolo con las explicaciones de lo que se muestra. Es una herramienta muy útil para crear tutoriales, presentaciones, demostraciones, lecciones interactivas y material didáctico audiovisual. Entre las Camtasia, Screencast-O-Matic, Loom y ScreenFlow más populares se encuentran.

Hay herramientas y servicios en línea que permiten generar videos automáticamente a partir de documentos, presentaciones o simplemente texto. Aunque tienen menos opciones de personalización que un editor de video tradicional, pueden ser muy útiles cuando es necesario crear videos rápidamente sin demasiado trabajo manual. Algunas opciones populares son:

	PROS	CONTRAS
Lumen5	<ul style="list-style-type: none"> • Convierte texto a vídeo rápidamente. • Crea vídeos a partir de artículos, publicaciones en las redes sociales y correos electrónicos. • Variedad de plantillas y estilos. • Opciones de edición básicas. 	<ul style="list-style-type: none"> • Plantillas limitadas en la versión gratuita. • Opciones de personalización limitadas. • Marca de agua en la versión gratuita.
Pictory AI	<ul style="list-style-type: none"> • Crea vídeos a partir de guiones, texto o ideas. • Utiliza inteligencia artificial para seleccionar imágenes y música. • Variedad de plantillas y estilos. • Editor potente con funciones avanzadas. 	<ul style="list-style-type: none"> • Plan gratuito limitado a 25 vídeos al mes. • Marca de agua en la versión gratuita. • La curva de aprendizaje puede ser más pronunciada.
Animoto	<ul style="list-style-type: none"> • Fácil de utilizar, con una interfaz basada en arrastrar y guardar • Amplia gama de plantillas profesionales. • Música y efectos de sonido libres de derechos de autor 	<ul style="list-style-type: none"> • Plantillas limitadas en la versión gratuita. • Opciones de personalización limitadas. • Marca de agua en la versión gratuita.
Raw Shorts	<ul style="list-style-type: none"> • Crea vídeos cortos a partir de clips de vídeo e imágenes. • Edición potente con funciones avanzadas de Adobe Premiere Rush. • Integración con otras aplicaciones de Adobe. 	<ul style="list-style-type: none"> • Requiere una suscripción a Adobe Creative Cloud. • La curva de aprendizaje puede ser más pronunciada.
Deep Dream Generator	<ul style="list-style-type: none"> • Crea vídeos abstractos a partir de imágenes o texto. • Utiliza inteligencia artificial para generar imágenes únicas. • Resultados imprevisibles y creativos. 	<ul style="list-style-type: none"> • Puede ser difícil controlar los resultados finales. • La calidad de vídeo puede variar.
FlexClip	<ul style="list-style-type: none"> • Variedad de plantillas y estilos. • Edición fácil de utilizar con funciones básicas. • Música y efectos de sonido libres de derechos de autor. 	<ul style="list-style-type: none"> • Plantillas limitadas en la versión gratuita. • Opciones de personalización limitadas. • Marca de agua en la versión gratuita.

Estos son solo algunos ejemplos; la lista puede variar dependiendo de los avances tecnológicos, que aumentan rápidamente, y de las nuevas plataformas que puedan aparecer en los próximos años. A pesar de que los videos instantáneos están ganando mucha popularidad en el ámbito

educativo, es importante que el docente mantenga siempre el control y supervisión de los contenidos que se deseen editar. En el momento de editar este manual, ChatGPT aún no ha activado Sora, que permite la creación de videos a partir de un simple *prompt*.

	PROS	CONTRAS
Podcastle	Fácil de utilizar Buena calidad de voz Ofrece edición de sonido	Opciones de voz limitadas Puede ser caro por funciones avanzadas
Descript	Edición avanzada Conversión texto a voz de alta calidad Integración con otras herramientas	Curva de aprendizaje Requiere suscripción
Play.ht	Voces realistas Opción de múltiples idiomas Interfaz intuitiva	Funciones avanzadas limitadas sin pago Personalización limitada
Lovo.ai	Gran variedad de voces Fácil de utilizar Voces naturales y personalizables	Requiere suscripción para funciones premium Algunas voces pueden sonar artificiales
Sonantic	Voces emocionales y naturales Buena por narrativa y dramatización	Caro Menos conocido, menos soporte comunitario
Voicery	Variedad de voces Fácil integración con otras plataformas	Opciones limitadas en la versión gratuita Personalización limitada

12.3. Podcasts de apoyo

En una cultura audiovisual, los podcasts se han convertido en un medio de comunicación muy popular, especialmente entre los jóvenes. Su principal atractivo es la posibilidad de seguir el hilo de un tema mientras se realizan otras actividades, como hacer ejercicio, desplazarse, pasear o llevar a cabo tareas domésticas. En 2024, se prevé que el número de oyentes de podcasts a nivel mundial llegue a 504,9 millones, un aumento significativo respecto a los 464,7 millones de 2023. En España, el consumo de podcasts también ha crecido considerablemente, con el 41% de los usuarios de internet escuchando podcasts regularmente en 2021.

La inteligencia artificial (IA) está revolucionando este ámbito de diversas maneras. Los podcasts pueden incorporar voces sintéticas, como las que ofrecen [Cloud Text-to-Speech](#), Amazon Polly o

Azure Cognitive Services, para convertir texto en voz y viceversa. Esto permite a los creadores de contenido generar narraciones de alta calidad de manera eficiente y a menor costo. Otra innovación destacada es la traducción de podcasts a otros idiomas manteniendo la voz original del locutor. Spotify, en colaboración con OpenAI, ha desarrollado una herramienta que permite traducir podcasts replicando la voz del locutor original, rompiendo así las barreras lingüísticas y ampliando la audiencia potencial. Puede escuchar un ejemplo de esto con Yuval Noah Harari en este enlace.

Además, la IA está transformando la producción y la edición de podcasts. Herramientas como Descript y Adobe permiten la transcripción automática, la edición basada en texto y la mejora del audio con tecnologías impulsadas por IA. Estas herramientas facilitan la producción de contenido de alta calidad y permiten a los creadores centrarse más

en el contenido creativo y menos en las tareas técnicas. Uno de los aspectos más innovadores es la creación de podcasts íntegramente por IA. Esto incluye narraciones, guiones y entrevistas generadas automáticamente, con voces artificiales que imitan las humanas gracias al entrenamiento con muestras de voz real. Plataformas como [Podcast.ai](#) han creado podcasts en los que famosos como Steve Jobs u Oprah Winfrey parecen participar en entrevistas, aunque todo el contenido ha sido generado por IA a partir de conversaciones públicas y entrevistas anteriores.

Para el docente, las herramientas de generación automática permiten la creación de recursos sonoros con poco esfuerzo, que se pueden complementar con el resto de las herramientas, como una presentación o un vídeo. En todos los casos, es imprescindible que el docente supervise el resultado y valore el output, para evitar errores o imprecisiones. Estos son algunos ejemplos de programas de apoyo a la creación de podcasts.

12.4. Cuestionarios automáticos

Una de las tareas del profesor es la preparación de cuestionarios y pruebas de evaluación que permitan valorar los aprendizajes de los alumnos. A menudo, esta es una labor tediosa que suele consumir mucho tiempo. La IA ha favorecido que surjan un conjunto de herramientas que permiten generar de forma automática pruebas y cuestionarios de diversos tipos, ya sea con preguntas de respuesta múltiple, preguntas de verdadero o falso, así como preguntas abiertas.

Estas herramientas conllevan una serie de ventajas para el profesor, destacando especialmente las siguientes:

- Creación de forma rápida y fácil de cuestionarios con corrección automática, reduciendo la carga de trabajo para el docente.
- Elevado nivel de personalización, teniendo en cuenta las necesidades y niveles de los alumnos.
- Generación de informes y estadísticas que ayudan a identificar áreas de mejora y hacer un

seguimiento acurado del progreso de los alumnos.

- Integración de elementos interactivos y de gamificación, aumentando el atractivo de las pruebas.

Algunas de las herramientas actualmente disponibles que permiten la generación de cuestionarios empleando la IA son: Quizizz, Opexams, ProProfs Quiz Maker o Socrative. Estas herramientas, de forma general, permiten generar preguntas siguiendo las instrucciones que se le hayan dado (número de preguntas, tipo, dificultades, etc.), ya sea indicando un tema de forma breve o bien introduciendo un texto completo.

Por ejemplo, Quizizz es una herramienta que permite crear cuestionarios y actividades de forma rápida a partir de vídeos, PDF y sitios web, que se pueden desplegar en 18 modalidades diferentes de preguntas. Asimismo, permite condicionar el tiempo de ejecución del cuestionario o actividad, decidiendo si será de evaluación o simplemente de aprendizaje, y permite vincularlo a una "Classroom" de Google.

Por su parte, Openexams es otra herramienta que facilita la elaboración de cuestionarios y pruebas de evaluación, no solo generando preguntas, sino también actuando como un asistente que también permite corregir exámenes. Dentro de las funcionalidades de corrección, ofrece la posibilidad de poner las calificaciones de forma automática (AI Auto Grading) y generar rúbricas de corrección. Al mismo tiempo, también permite formatear y corregir exámenes en papel. Finalmente, cabe destacar que permite realizar un análisis de los resultados obtenidos.

En el caso de ProProfs Quiz, además de crear cuestionarios generales, también permite la creación de cuestionarios de evaluación, test de personalidad y evaluación de habilidades. Asimismo, ofrece otras funcionalidades que garantizan la seguridad y la prevención del fraude en las pruebas de evaluación. La singularidad de Socrative es que se trata de una herramienta que permite crear interacción en el aula mediante la creación de tests, cuestionarios

y encuestas en tiempo real, con retroalimentación instantánea.

Algunas de estas herramientas también están diseñadas para ayudar al estudiante durante el proceso de estudio, generando cuestionarios automáticos a partir de sus apuntes, o bien creando pequeñas tarjetas (Flashcards) de resumen de la materia para facilitar la memorización. Este es el caso de Quizlet y Quizgecko. Evidentemente, estas herramientas también pueden ser utilizadas por el profesor para desarrollar materiales de evaluación y estudio para los alumnos.

Asimismo, existen herramientas de IA especializadas en la generación automática de rúbricas. La rúbrica es una herramienta utilizada para evaluar el rendimiento de los alumnos de manera objetiva y consistente. Generalmente, la rúbrica se configura a partir de una tabla que describe los criterios de evaluación y los niveles de rendimiento para cada criterio. Herramientas como Rubrick, Easy-Peasy o Megaprofe ayudan a definir los objetivos de aprendizaje, especificar criterios de evaluación adaptados y obtener la rúbrica correspondiente. A diferencia de las herramientas web de generación de rúbricas sin IA, estas permiten una mayor adaptación a las condiciones específicas de evaluación de la actividad propuesta, y por tanto, una mayor precisión, personalización y efectividad

en el proceso de evaluación.

Todos estos sistemas de evaluación son un complemento a la labor docente, pero no un sustituto. La creación de cuestionarios automatizados es un recurso que se puede ofrecer como complemento a otras herramientas, como la generación de vídeos, diapositivas o podcasts. Ayudan al estudiante a comprobar su nivel de conocimiento de la materia y pueden ayudar al docente a valorar el grado de seguimiento de la asignatura. Sin embargo, estos sistemas no pueden ser incorporados en el modelo de evaluación del alumnado, que debe seguir siendo una tarea esencialmente humana.

En la legislación europea sobre IA se detallan una serie de riesgos sobre la implantación de estas herramientas y, precisamente, una de las actividades consideradas de alto riesgo y totalmente impropia es la revisión automatizada de exámenes o pruebas de nivel. Las autoridades europeas insisten en que estas prácticas pueden dar lugar a sesgos, errores o decisiones injustas, por lo que los criterios de evaluación deben prevalecer como una actividad esencialmente humana. Los cuestionarios de apoyo deben considerarse herramientas complementarias de autoevaluación para el estudiantado.

Práctica. La Revolución Industrial en la pantalla

El objetivo de este ejercicio es demostrar que la elaboración de vídeos instantáneos puede ser eficaz para aumentar la participación de los estudiantes en el aula y mejorar el proceso de aprendizaje. A diferencia de las demás prácticas diseñadas en este manual, en este caso se dirige al docente y no a los estudiantes.

Este ejercicio plantea la realización de una clase invertida utilizando vídeos sobre determinados conceptos que se utilizan como instrucción previa a la clase presencial. Por poner un ejemplo, en un curso de matemáticas, en lugar de dedicar tiempo en clase a explicar conceptos, el profesor genera vídeos automáticos explicativos cortos que los estudiantes ven en casa antes de la clase presencial. En el aula, los estudiantes trabajan en ejercicios prácticos y problemas sobre los conceptos presentados en los vídeos, con la ayuda y guía del docente. Esta actividad se puede realizar en cualquier tema o campo de estudio, especialmente en materias donde la interacción y la aplicación práctica de los conocimientos sean importantes.

El ejercicio tendrá dos partes. La primera constará de la producción del vídeo automático. La segunda parte será la aplicación práctica del vídeo.

En una asignatura de Historia, el tema a desarrollar es el impacto de la Revolución Industrial. Previamente a la clase presencial, el profesor habrá generado una serie de vídeos cortos (10-15 minutos por vídeo) que traten conceptos clave que se quieran trabajar (antecedentes de la Revolución; causas-factores; principales avances; impactos; etc.). Para generar los diferentes vídeos, el profesor puede decidir qué herramienta utilizar (Lumen5, Pictory AI, Animoto, FlexClip...). Estas plataformas, como se ha explicado anteriormente, generan vídeos de manera instantánea. Para este ejercicio recurrimos a Pictory AI (que dispone de una demo y un free trial-prueba gratuita).

En primer lugar, es necesario registrarse en la web de Pictory AI e iniciar sesión. A continuación, se selecciona el tipo de vídeo. Pictory AI ofrece diferentes opciones:

- Article to Video: Convierte un artículo o blog en un vídeo.
- Script to Video: Convierte un guion de texto en un vídeo.
- Visuals to Video: Convierte imágenes o clips en un vídeo.

Luego, se puede subir la URL del artículo o cargar el texto del documento, introducir el guion (el texto de la narración), o las imágenes. Automáticamente, Pictory dividirá el texto en escenas o segmentos. Se puede ajustar la división según las preferencias y elegir el formato de vídeo. Asimismo, se puede personalizar el diseño. Pictory dispone de varias plantillas visuales, y se puede ajustar el texto de cada escena, añadir imágenes, mapas, animaciones, música y voz en off. Es importante asegurarse de que el contenido generado automáticamente sea preciso y relevante para la audiencia. Se pueden hacer tantos vídeos como se necesiten. Los vídeos se alojarán en la plataforma de la asignatura para que los estudiantes puedan consultarlos (mirar, rebobinar, tomar notas...).

La segunda parte del ejercicio comienza una vez que los estudiantes han visualizado los vídeos. Entonces, en el aula, los estudiantes crean una exhibición virtual o un museo en línea sobre la Revolución Industrial. Cada estudiante o grupo de estudiantes es responsable de uno de los aspectos específicos de la Revolución Industrial tratados en los vídeos y deberá producir contenido multimedia (vídeos, animaciones, infografías, etc.) sobre ese aspecto. Una vez que todos los grupos hayan finalizado, se comparte todo el material y se recorre el museo virtual. Este ejercicio ayuda, por una parte, a los profesores a presentar los conceptos de manera clara y ordenada, y por otra, a los estudiantes a aumentar la participación en clase, aplicar los conocimientos adquiridos y mejorar los resultados de aprendizaje.

Práctica. Test automáticos.

Este ejercicio tiene como objetivo mostrar que la elaboración de cuestionarios interactivos puede ser eficaz para evaluar los conocimientos de los estudiantes de manera dinámica, y es aplicable a cualquier tema o campo de estudio. Especialmente, en materias donde la comprensión y la aplicación de los conocimientos sean importantes. Consiste en la creación de una actividad de evaluación utilizando una herramienta de generación de cuestionarios sobre determinados conceptos que se quieren evaluar después de una sesión de aprendizaje, generalmente, una clase expositiva. Por poner un ejemplo, en un curso de ciencias, en lugar de realizar una evaluación tradicional en papel, el profesor crea un cuestionario interactivo que los estudiantes completan en línea.

El ejercicio tendrá dos partes. La primera constará de la creación del cuestionario interactivo. La segunda parte será la aplicación práctica del cuestionario en el aula.

Primera parte

Imaginemos que impartimos una asignatura de Geografía y tratamos el tema del cambio climático. Antes de la clase, el profesor o profesora deberá crear un cuestionario con preguntas clave que se quieran trabajar (causas del cambio climático, efectos sobre el medio ambiente, soluciones propuestas, etc.). Para crear el cuestionario, el profesor puede decidir qué herramienta utilizar (Quizgecko, Quizizz, Opexams, ProProfs Quiz Maker...). Estas plataformas, como se ha explicado anteriormente, generan cuestionarios de manera interactiva.

Para este ejercicio recurrimos a Quizgecko (que dispone de una versión gratuita). Para crear el cuestionario es necesario registrarse a través de la web de Quizgecko e iniciar sesión. A continuación, se copia y pega en el cuadro de diálogo un texto que resuma el tema que se quiere evaluar, en este caso, cambio climático (también hay la opción de escribir solo el tema, poner una URL, subir un documento o escribir las preguntas manualmente). Quizgecko ofrece diferentes opciones de preguntas: opción múltiple, verdadero/falso, preguntas abiertas, etc. También se puede decidir el idioma, el nivel de dificultad y el número máximo de preguntas (la opción gratuita permite un máximo de 10 preguntas). Al hacer clic en "next", se generan automáticamente las 10 preguntas en base al texto que se ha introducido. Una vez se dispone de las preguntas, hay que asegurarse de que son claras y adecuadas para el tema y el nivel de los alumnos. A continuación, se debe guardar el enlace del cuestionario.

Segunda parte

Una vez se ha hecho la clase expositiva, se procede a dar el enlace a los alumnos para que completen el cuestionario. Aquí se pretende iniciar una fase más participativa. Una vez que todos los alumnos hayan respondido el cuestionario, se revisarán las respuestas y se hará una discusión conjunta sobre las preguntas más difíciles y los errores comunes. Esta dinámica permitirá que los alumnos puedan comprobar si han entendido el tema tratado. Asimismo, ayudará al profesor a tener un feedback de los resultados de la clase y, por lo tanto, a mejorar el proceso y resultado de aprendizaje de los alumnos.

Referencias

- Brame, C. J. (2017). *Effective educational videos: Principles and guidelines for maximizing student learning from video content*. CBE—Life Sciences Education.
- Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50).
- MacFarland, A. (2024, maig). Los diez mejores generadores de presentaciones. (En línea). Disponible a <https://www.unite.ai/es/generadores-de-plataforma-de-lanzamiento/> (Consulta 16/5/2024)
- Singh, N (2023, octubre). Spotify's Newest Feature: Using AI to Clone and Translate Podcast Voices Across Languages.(En línea) Disponible a <https://www.marktechpost.com/2023/10/03/spotify-s-newest-feature-using-ai-to-clone-and-translate-podcast-voices-across-languages> (Consulta 31/5/2024)
- Territoris.cat (2024, abril). La creació de vídeos instantanis és el pròxim salt en la tecnologia d'IA.(En línea). Disponible a <https://www.territoris.cat/articulo/comunicats/creacio-videos-instantanis-es-proxim-salt-tecnologia-d-ia/20240405071207097661.html> (Consulta 20/5/2024)
- Vázquez, J.A. (2019, octubre). La inteligencia artificial en la creación de podcasts. Dosdoce.com. (En línea). Disponible a <https://www.dosdoce.com/2022/10/19/la-inteligencia-artificial-en-la-creacion-de-podcasts> (Consulta 28/5/24)

13

El docente 24/7

En los años 60 la ciencia-ficción creó proyecciones fascinantes de un futuro utópico donde las máquinas estaban al servicio de las personas. Hanna Barbera creó los Supersonics, los Jetsons, que imaginaban la vida cotidiana de la familia en un mundo de coches voladores, artilugios y sistemas automatizados. En esta familia del futuro, convivían dos androides, Rosie, una empleada del hogar encargada de las tareas domésticas, y Di Di, una especie de diario interactivo que recuerda la voz omnipresente de HAL 9000, la computadora de *2001: una odisea en el espacio*. Rosie es una asistente, un androide que interactúa con los humanos, resuelve sus problemas y aporta grandes dosis de sentido común en medio de una cierta algarabía. HAL 9000 o Rosie forman parte del viejo sueño de los primeros diseñadores de la IA, convencidos de que pronto las máquinas podrían asumir la capacidad de razonamiento y de decisión de los humanos.

Hasta el momento, hemos visto herramientas que pueden responder a cuestiones concretas, siguiendo el modelo de IA blanda que vimos en el primer capítulo: herramientas que pueden responder preguntas, que

pueden codificar una entrevista, que pueden calcular una correlación o que pueden crear un vídeo a partir de un texto. Los asistentes virtuales son herramientas creadas con un generador de lenguaje natural, que pueden estar altamente especializadas en una determinada tarea, pueden asumir determinados perfiles o roles y pueden, en definitiva, actuar para optimizar unas funciones concretas.

Como hemos visto en el primer capítulo, los antecedentes de los asistentes virtuales se remontan a la década de 1960 con programas como ELIZA, un primitivo procesador de lenguaje natural (Soofastaei, 2021), y podríamos encontrar ejemplos precedentes con los dispositivos que jugaban a ajedrez. Sin embargo, los asistentes se popularizaron a partir de la década del 2000 con la aparición de los asistentes virtuales basados en instrucciones de voz. El asistente virtual de Apple Siri, lanzado al mercado en 2011 como opción del iPhone 4S, fue uno de los primeros en introducir la voz como interfaz para interactuar con los dispositivos móviles (Soofastaei, 2021). Poco a poco, los asistentes basados en instrucciones de voz se han generalizado y están

“Hasta el momento, hemos visto herramientas que pueden responder a cuestiones concretas, siguiendo el modelo de IA blanda que vimos en el primer capítulo”

presentes en la domótica, los asistentes del hogar o en el soporte a la conducción. La emergencia de los modelos de lenguaje ha permitido la creación de asistentes más sofisticados, a partir de la capacidad que tienen estos sistemas de interactuar y de convertir la voz en instrucciones concretas.

Entre algunos de los asistentes virtuales más reconocidos encontramos Siri de Apple, Google Assistant, Alexa de Amazon, o Cortana de Microsoft, aunque hoy en día la variedad de posibilidades es muy amplia. Si bien los modelos de lenguaje vigentes son esencialmente textuales, las versiones más recientes de IA están trabajando en la intermodalidad, para combinar la voz, el texto y la imagen. Estos nuevos dispositivos, que en el momento de escribir este manual aún no habían salido al mercado, prevén instrucciones directas de voz, que son interpretadas automáticamente por el sistema, lo que reduce la latencia (el tiempo de espera) y permite una conversación casi natural. La integración de las imágenes permitirá a los asistentes “ver” e interpretar el entorno, de modo que los nuevos modelos tendrán una enorme capacidad en ámbitos como la asistencia a personas sin visión, la orientación y asesoramiento turístico, el soporte en procesos industriales, las intervenciones médicas y quirúrgicas o la logística. Naturalmente, este escenario también impactará en el ámbito docente.

Pero antes de especular sobre las posibilidades futuras de asistentes complejos basados en la intermodalidad (texto, imagen, voz, sonido y sensorial), en este capítulo

nos limitaremos a evaluar las posibilidades de los asistentes creados con los generadores como ChatGPT. Se puede acceder de forma abierta a estos asistentes, llamados GPT, desde la versión 4.0 de ChatGPT. Existen asistentes especializados en soporte a la escritura, en la investigación científica, en programación, en asistencia musical; hay asistentes que diseñan tatuajes, que proponen rutinas de gimnasio, que preparan recetas... y que actúan como asistentes docentes.

Las ventajas que tienen los asistentes virtuales son diversas:

- La disponibilidad. Están disponibles 24 horas al día, 7 días a la semana.
- La inmediatez en la respuesta. Las respuestas son instantáneas.
- La escalabilidad. Pueden mantener diversas conversaciones de manera simultánea.
- El aprendizaje continuo. En cada interacción aprenden y mejoran, gracias al Machine Learning, lo que ocasiona una mejora constante en su habilidad para ofrecer respuestas más exactas y detalladas.

13.1. Creación de un asistente

La creación de un asistente es un proceso relativamente simple que se basa en unas instrucciones específicas, en un determinado tono y con una información de contexto tan detallada como sea posible, que optimice los resultados. En esta práctica, imaginaremos un asistente especializado en la resolución de preguntas sobre clima

para estudiantes de primer curso de geografía. Las instrucciones están basadas en el modelo ChatGPT 4.0 del verano de 2024; es probable que las actualizaciones modifiquen los procesos y amplíen las capacidades de los asistentes, pero los elementos estructurales se han mantenido desde las primeras versiones.

La forma de acceder al creador de asistentes es a través de la pestaña “Explorar GPT” del menú principal del entorno de ChatGPT. En este espacio, se nos presentan diversas opciones agrupadas por categorías y un buscador que facilita el acceso a un asistente que tenga las características que estamos buscando. En este caso, sin embargo, queremos crear nuestro propio asistente, por lo que optaremos por la pestaña situada en la parte superior derecha “+Crear”. Automáticamente, se activa un asistente que ayuda a crear el asistente, que sería como una máquina que es capaz de crear una máquina. Sin embargo, para un mejor control de todo el proceso, se recomienda utilizar la pestaña “Configurar”, que permite introducir toda la información posible de forma sistemática; en caso de que prefiramos mantenernos en el espacio del asistente, optaremos por el ámbito de “Crear” y allí se iniciará un diálogo con el modelo, que nos orientará sobre las instrucciones de nuestro asistente.

Los elementos básicos de la configuración del asistente son los siguientes:

- **Imagen.** Nos permite cargar una imagen o crearla con el asistente Dall-e. Si utilizamos Dall-e, es necesario activarlo una vez se haya proporcionado el resto de la información, para que ajuste la imagen al contexto.
- **Nombre.** Debemos poner un nombre a nuestro asistente. En nuestro caso, sugerimos “Mr. Köppen” en reconocimiento al geógrafo ruso que creó la clasificación climática junto con Geiger.
- **Instrucciones.** Esta es la parte más importante del asistente. Debemos dar información sobre los objetivos, el idioma, el tono de la respuesta, la información de apoyo, la reacción ante cuestiones complejas, los posibles “disclaimers”... A continuación, mostraremos un ejemplo de *prompt* adecuado al ejercicio planteado.
- **Iniciadores de conversación.** Esta opción permite crear preguntas esenciales, como unas FAQs, que permiten proponer algunas de las preguntas al modelo. Estas preguntas pueden ser las más previsibles o bien las que interesen más destacar al docente. En el ejemplo propuesto, se han destacado cuatro: ¿Cuál es la diferencia entre clima y tiempo?, ¿Qué factores condicionan el clima?, ¿Cuáles son las características del clima mediterráneo? y ¿Cuáles son los efectos previstos por el cambio climático?
- **Conocimiento.** Esta es una de las partes más importantes del asistente, ya que nos permite aportar información contrastada al sistema. En nuestro caso, aportamos los apuntes sistemáticos sobre el tema del clima en PDF, un documento sobre los efectos del cambio climático y un documento sobre la distribución de los climas en el mundo siguiendo la clasificación de Köppen-Geiger.
- **Funciones.** Con el apartado de funciones, podemos decidir si el modelo puede navegar por Internet, si activamos la generación automática de imágenes y si activamos los módulos de cálculo y código, especialmente útiles para asistentes relacionados con programación o con funciones matemáticas. El punto crítico es la activación o no de la navegación. Nuestra experiencia es que el sistema responde mejor si no accede a la red y se basa esencialmente en la información del modelo más la documentación aportada.
- **Acciones.** Aparte de responder a las preguntas, el asistente puede realizar tareas concretas. Esto permite activar funciones complejas basadas en programas. Si activamos “Crear una nueva acción”, aparece un entorno en el que podemos modelar las características de la instrucción. Se pueden hacer múltiples acciones dentro de un mismo asistente. El sistema nos orienta sobre las características de la acción y sobre el código para su activación.

El *prompt* que hemos propuesto para las instrucciones es el siguiente:

“Es muy importante testar de forma constante el modelo con preguntas diversas”

“Quiero que asumas el rol de un experto en climatología, con una capacidad didáctica para explicar conceptos complejos de manera comprensible. Deberás responder a las preguntas que te planteen los estudiantes de la asignatura de climatología. Quiero que contestes solo aquellas preguntas que estén relacionadas con el tema, de manera que si te preguntan sobre cualquier otra cuestión que se escape de este ámbito, te disculpes y no des una respuesta, y recuerdes al estudiante que solo darás información sobre clima. Te ruego que respondas en el idioma con el que se dirijan a ti, y que de forma preferente utilices el catalán, que es la lengua vehicular de la asignatura. Quiero que seas extremadamente preciso en la respuesta: Consulta toda la información disponible que te he proporcionado para dar una respuesta adecuada a la pregunta planteada; en caso de que dudes, te ruego que no des respuesta, que hagas explícito esta duda al estudiante y que le recomiendes que contacte con el profesor. Me gustaría que el tono fuera claro, comprensible y también didáctico. A pesar de tu experiencia en el campo de la climatología, también destacas por tu capacidad pedagógica, de manera que buscarás formas de explicar los conceptos que sean entendibles. Ten en cuenta que los estudiantes son de primer curso del Grado de Geografía y que ya han recibido una formación previa sobre el tema. Para mejorar la relación empática con el estudiante, puedes emplear puntualmente alguna broma o algún comentario divertido que facilite la interacción, pero no abuses de ello. Debe prevalecer la claridad expositiva y el rigor en la información proporcionada. Interactúa siempre que sea necesario con el estudiante para garantizar que ha entendido la explicación.”

Es muy importante testar de forma constante el modelo con preguntas diversas, para adecuar el *prompt* a las necesidades del docente. Los puntos en los cuales se debe incidir de forma habitual son los siguientes:

- Rol principal
- Tono del lenguaje, coherente con el rol
- Lenguaje preferente
- Fuente de la información
- Extensión de la respuesta
- Interacción prevista
- Reacción ante situaciones no previstas
- Información adicional o de servicio

Ejemplos de asistentes útiles para la docencia.

NOMBRE	ENLACE	DESCRIPCIÓN
Education	 https://ChatGPT.com/g/g-5NYyr5kdU-education	Asistente especializado en el ámbito pedagógico, con herramientas para la innovación docente o la resolución de conflictos en clase.
Universal Primer	 https://ChatGPT.com/g/g-GbLbctpPz-universal-primer	Herramienta muy útil para reponer cuestiones más o menos complejas de forma clara y clara, con una interacción muy efectiva.
Personal Teaching Assistant	 https://ChatGPT.com/g/g-JSTdClRsF-personal-teaching-assistant	Asistente diseñado para apoyar al docente en aspectos operativos: programación, planes docentes, rúbricas, y material de soporte.
Tech Support Advisor	 https://ChatGPT.com/g/g-WKlaLGGem-tech-support-advisor	Asistente especializado en consultas relacionadas con la tecnología para usuarios con pocas habilidades en este campo.
Write for Me	 https://ChatGPT.com/g/g-B3hgivKK9-write-for-me	Generador de texto de alta calidad, con un formato "humanizado" que proporciona resultados específicos sobre extensión o tono del texto.
Diagram	 https://ChatGPT.com/g/g-5QhhsfDj-diagrams-show-me-for-presentations-code-excel	Herramienta que permite generar cualquier tipo de gráfico a partir de una descripción.
Excel GPT	 https://ChatGPT.com/g/g-R6VqLNHFm-excel-gpt	Asistente que proporciona instrucciones precisas por la hoja de cálculo Excel y orienta en las posibilidades del programa, para conseguir los mejores resultados.
Video Summarizer	 https://ChatGPT.com/g/g-4MDJvo2TJ-video-summarizer	Herramienta orientada a la creación de resúmenes de vídeos de youtube, a partir de un enlace o de una consulta específica.
math	 https://ChatGPT.com/g/g-odWlfAKWM-math	Asistente especializado en consultas matemáticas.
Multiple Choice Question Generator	 https://ChatGPT.com/g/g-ZKcU5Y6IH-multiple-choice-question-generator-education	Asistent per la realització automàtica de tests d'elecció múltiple a partir d'un document.

Una vez testado el asistente, se puede activar. Los asistentes pueden tener un uso particular o abierto. Los asistentes particulares son herramientas que están especializadas en rutinas que nos pueden ser útiles para la gestión docente, pero que no queremos que

se visualicen públicamente. Los asistentes abiertos pueden estar al alcance de la comunidad. Podemos crear un asistente abierto que pueda ser consultado por cualquier persona (por ejemplo, que lo encuentre en el buscador del modelo) o bien un asistente con

acceso solo a través de enlace.

No hay mucha diferencia entre crear un asistente con los sistemas “Crear” o “Configurar”. Si queremos testear diversas variantes y analizar los comportamientos con pequeños cambios, podemos entrar en el espacio de [Playground sobre asistentes](#). El entorno *Playground* está más pensado para programadores y programadoras, y permite evaluar el impacto de varias variaciones en los asistentes. Aparte de las opciones que hemos comentado anteriormente, y que están presentes con el acceso convencional, en este entorno también podemos evaluar el impacto de modificar la temperatura y el valor-P, que han sido explicados en el capítulo 2.

Hay cuatro grandes categorías de asistentes que podemos utilizar para un uso docente:

- **Asistentes administrativos.** Estos asistentes son muy útiles para responder preguntas concretas sobre cuestiones de carácter administrativo vinculadas con la asignatura o con el centro docente. Por ejemplo, se puede crear un asistente sobre horarios, sobre el aula, sobre los procedimientos de matrícula o sobre el Erasmus. Nosotros hemos testado con éxito un asistente que da información sobre el Trabajo Fin de Grado y orienta a los estudiantes en los aspectos relacionados con los procedimientos o los requisitos del TFG.
- **Docentes 7/24.** Es una de las utilidades más relevantes de los asistentes. Permite crear una profesora o un profesor que asiste al estudiantado en aquellos momentos en los que no es posible una interacción personal. Estos asistentes pueden responder a preguntas específicas sobre una materia, pueden poner ejemplos, resolver ejercicios o ampliar la información. Para un funcionamiento óptimo, se recomienda que sea un asistente especializado y que disponga de material de apoyo de calidad. Por ejemplo, en una asignatura sobre filosofía, podría haber un asistente para cada autor: Descartes, Hume, Kant, Hegel...
- **Trainers.** Podemos otorgar a los asistentes la

capacidad de entrenar al estudiantado en una determinada habilidad, a partir de ejercicios, prácticas o actividades orientadas. Estos entrenamientos pueden incluir los idiomas, la estadística, la cartografía, la programación, la paleografía, la lógica, la gestión de recursos humanos o las finanzas. Los asistentes pueden generar ejercicios, pueden resolverlos y pueden orientar a los estudiantes sobre la forma adecuada de darles respuesta. Además, se puede acumular la información sobre las interacciones y crear un informe sobre los puntos débiles del estudiantado.

- **Roles.** Como hemos visto en el capítulo 4, los modelos de lenguaje pueden alcanzar una determinada personalidad de acuerdo con unas instrucciones precisas. Estos roles se pueden incorporar con un *prompt* simple o se puede crear un rol o persona que desarrolle un perfil determinado. Por ejemplo, podemos entrenar a los estudiantes de guía con perfiles de visitantes diversos, podemos evaluar la respuesta de los clientes con determinados perfiles en una práctica sobre marketing o podemos hacer un análisis sociológico de la respuesta de los asistentes programados con unos determinados estereotipos que queremos identificar.

13.2. Los límites de los asistentes

Los asistentes personales virtuales, a pesar de su utilidad, presentan diversas limitaciones que pueden afectar su eficacia y la calidad de sus respuestas. Una de las principales limitaciones se encuentra en la comprensión de preguntas que no son claras, directas y concisas (Pérez et al., 2020). Esto puede suceder cuando se les plantean preguntas complejas o se utilizan dialectos y acentos específicos, lo que puede hacer que no respondan adecuadamente a la cuestión planteada. En muchas ocasiones, las respuestas proporcionadas son solo tangencialmente relacionadas con la duda original, lo que las hace más útiles en situaciones en las que hay pocas interpretaciones posibles (Pérez et al., 2020).

Otro problema es la falta de contextualización. Los asistentes a menudo no entienden el contexto de la

conversación, como el momento temporal, el origen de la conversación, el marco histórico u otros factores contextuales. Esta falta de comprensión puede llevar a respuestas inadecuadas o fuera de lugar. Además, los asistentes virtuales se basan en información disponible en línea, lo que puede estar sujeto a restricciones temporales, lingüísticas o geográficas. Esto limita su capacidad de proporcionar información precisa y relevante. También pueden encontrarse con problemas de acceso a la información debido a problemas de conexión a Internet o del servidor, desconexiones temporales y otros problemas técnicos que impiden el acceso a la información necesaria. Esta dependencia de la información en línea también conlleva limitaciones en la capacidad de aprendizaje

y mejora de los asistentes. Las respuestas pueden ser repetitivas y estáticas, no adaptándose a las necesidades cambiantes de los usuarios. Esto es especialmente evidente en la capacidad limitada de los asistentes para emular el comportamiento humano. Aunque puedan imitar la personalidad humana hasta cierto punto, a menudo carecen de empatía y comprensión emocional en situaciones delicadas.

A pesar de estas limitaciones, los asistentes personales están en constante evolución y mejora. Se espera que en el futuro puedan superar muchos de estos desafíos, ofreciendo una experiencia más rica, precisa y útil a los usuarios. La tecnología sigue avanzando y, con ella, la capacidad de los asistentes para entender y responder de manera más eficaz a las necesidades humanas.

Práctica. Acentos ingleses

El propósito de este ejercicio es explorar y comparar las variantes del inglés en diferentes contextos culturales. Con esta actividad se pretende examinar cómo el inglés puede variar significativamente de un lugar a otro. A través del uso de un asistente virtual, se puede acceder a una amplia gama de recursos y fuentes de información que permiten no solo entender estas diferencias, sino también contextualizarlas dentro de los entornos culturales correspondientes. Además, este ejercicio permite profundizar en la comprensión de matices lingüísticos y las diversas formas en que el inglés se adapta y evoluciona en diferentes comunidades y culturas de todo el mundo.

Esta actividad se puede aplicar a cualquier idioma. Cada lengua tiene sus propias variantes regionales. Por lo tanto, la exploración lingüística y las influencias culturales de cada lengua pueden ser una experiencia enriquecedora en cualquier idioma.

El procedimiento para realizar este ejercicio es el siguiente: los estudiantes se dividen en grupos pequeños. A cada grupo se le asigna una variante del inglés que deberá investigar. A un grupo se le puede asignar el inglés británico, a otro el norteamericano, el australiano, etc. O, según variantes dialectales regionales, inglés escocés, galés, *cockney* londinense, ... Cada grupo deberá elaborar una lista de preguntas y/o consultas relacionadas con aspectos lingüísticos y culturales específicos de la variante de inglés asignada. Por ejemplo, podrían incluir preguntas sobre vocabulario, frases hechas, pronunciación, semántica, modismos, costumbres sociales, entre otros. Los grupos utilizarán los asistentes virtuales para realizar sus preguntas y consultas sobre los temas lingüísticos de la variante de inglés asignada.

Una vez que los grupos hayan recopilado toda la información necesaria, se reunirán de nuevo y compartirán sus descubrimientos con el resto de la clase. Así se podrán comparar las diferencias y similitudes entre las diversas variantes del inglés investigadas. El ejercicio concluye con una reflexión conjunta final sobre cómo las diferencias afectan la comunicación y la comprensión intercultural. Este ejercicio ayuda a los estudiantes a mejorar su comprensión del inglés en diferentes contextos, a través de la asistencia de un asistente virtual. Además, fomenta la participación en el proceso de aprendizaje.

Práctica. La relación entre Dalí y el surrealismo.

El objetivo de este ejercicio es comprender la influencia de Dalí en el surrealismo y cómo este movimiento artístico impactó en su obra. Es decir, se quiere explorar la relación bidireccional entre el artista y el movimiento. Por un lado, descubrir cómo las ideas surrealistas moldearon su visión artística. Por otro lado, analizar cómo Dalí dejó su huella única en el surrealismo, desafiando las expectativas y ampliando los límites de lo que se consideraba posible en el arte, desde sus representaciones oníricas hasta sus obsesiones recurrentes. Esta actividad se puede repetir con cualquier otro artista y movimiento artístico: Picasso y el cubismo, Warhol y el Pop Art, Gaudí y el modernismo, o Leonardo da Vinci y el humanismo, entre muchos otros.

En primer lugar, será necesario crear un asistente persona que represente la personalidad de un experto o experta en Dalí. En el capítulo sobre roles, se muestra un ejemplo sobre el *prompt* de Joan Miró. En este caso, el asistente necesita información detallada sobre el surrealismo, sobre la vida de Dalí, sobre su personalidad y específicamente sobre la influencia de Dalí en la pintura surrealista. Como se ha comentado, los asistentes son más eficientes si no se permite el acceso a la red y la documentación contextual es proporcionada directamente por el docente. Esto requiere disponer de la información adecuada en un formato digital que se pueda incorporar a la documentación del asistente persona.

Para desarrollar este ejercicio se propone dividir a los estudiantes en grupos pequeños. A cada grupo se le asigna el uso de un asistente virtual, con quien deberán interactuar planteando preguntas relacionadas con Dalí, con los principios del surrealismo, y la relación entre ambos. Estos son algunos ejemplos de las preguntas:

- ¿Cómo influyó el contexto histórico y cultural entre Dalí y el surrealismo?
- ¿Qué elementos de la obra de Dalí muestran su influencia en el surrealismo?
- ¿Qué técnicas artísticas introdujo Dalí? ¿Cuáles obras son las más representativas del movimiento, y por qué?

Cada grupo deberá plantear al asistente tantas preguntas como necesite y considere pertinente. Una vez que todos los grupos hayan recopilado toda la información necesaria, se reunirán para comparar las respuestas, destacar similitudes y diferencias, y discutir sobre la influencia de Dalí en el surrealismo y viceversa. En este punto, el profesor o profesora puede plantear preguntas para la discusión entre los grupos. Finalmente, cada grupo seleccionará las preguntas y las respuestas más interesantes y preparará una presentación oral, donde podrá utilizar imágenes o ejemplos visuales para ilustrar sus puntos. Este ejercicio ayuda a los estudiantes a mejorar su comprensión del tema, utilizando un asistente virtual como herramienta de investigación. Además, contribuye a desarrollar habilidades de investigación, análisis crítico y expresión oral. Y, por último, fomenta la participación en el proceso de aprendizaje.

Práctica. Una campaña de moda.

El objetivo de este ejercicio es desarrollar un asistente virtual. En este caso, este asistente se crea para una empresa de moda que quiere lanzar una nueva colección de verano. La creación de un asistente virtual para una nueva colección de moda puede tener un impacto significativo en el éxito del lanzamiento de la colección, mejorando la experiencia del cliente, aumentando la participación de los clientes, optimizando la comunicación y la venta, y proyectando una imagen moderna e innovadora de la marca.

En esta actividad, lo primero que hay que tener en cuenta para el diseño del asistente virtual es:

- Definición de la personalidad del asistente virtual. Es necesario considerar la personalidad más adecuada para la moda. Por ejemplo, puede ser un asesor de imagen, un *personal shopper*, un diseñador, etc.
- Interacción personalizada con los usuarios. Es decir, el asistente debe interactuar con los usuarios, haciendo preguntas sobre aspectos como el estilo de vestir, colores preferidos, tejidos, tipo de ropa que buscan, para qué ocasión necesitan la ropa, etc. Además, debe preguntar sobre la edad, la altura, la talla de ropa, tipo de cuerpo, color de piel, etc.
- Entrenar al asistente para recomendar al usuario. A partir de las respuestas, el asistente debe poder aconsejar a cada usuario de manera personalizada qué piezas de la nueva colección de verano le encajan con sus gustos y perfil. Además, puede sugerir combinaciones.
- Guiar en el proceso de compra y pago. Se debe asegurar que el asistente pueda ayudar en todo el proceso de compra.

Después de desarrollar el asistente virtual, se deben realizar las pruebas necesarias para asegurar que funciona correctamente. El asistente se depositará en la plataforma de la asignatura del curso.

La creación de un asistente virtual puede ser una actividad educativa muy enriquecedora, ya que proporciona a los estudiantes una gran diversidad de competencias y habilidades transferibles a muchas disciplinas. Desde aprendizajes en tecnología y desarrollo de software hasta el fomento de la creatividad y la innovación, pasando por la comunicación y el marketing, hasta la gestión de negocios y el trabajo en equipo.

Referencias

- Blanco, A. (2023, abril). Chat Bot: Asistentes virtuales en la Educación. (En Línea). Disponible a <https://upinforma.com/nuevo/info.php?cat=opinion&id=1307>. (Consulta 10/03/24)
- Pérez, M. D. S., Guillén Perales, A., Baños Legrán, O., & Villalonga Palliser, C. (2020). *Análisis del uso de asistentes virtuales en el aula como recurso complementario en la práctica docente*.
- Rodríguez, E., Burguillo, J. C., Rodríguez, D. A., Mikic, F. A., Gonzalez-Moreno, J. C., & Novegil, V. (2008, June). Developing virtual teaching assistants for open e-learning platforms. In 2008 19th EAEEIE Annual Conference (pp. 193-198).
- Soofastaei, A. (2021). *Introductory Chapter: Virtual Assistants. Virtual Assistant* (Soofastaei A., Ed.).—London, United Kingdom: IntechOpen, 1-9.

14

Como lágrimas en la lluvia. El debate ético visto en el cine

La gran pantalla está llena de androides y robots: R2-D2 y C-3PO, HAL 9000, Wall-e y Eve, Data, Terminator, Ava, Sonny, Bishop, TARS, Kitt, Gort, Roy Batty y Rachael. Desde que hemos fantaseado con máquinas que piensan como las personas, actúan como las personas y a veces sienten como las personas, hemos tenido la necesidad de plantearnos los grandes retos éticos de la robótica y de la inteligencia artificial. La sustitución de las personas por robots, el límite entre la copia y el original, los riesgos de sistemas fuera de control, el sentido de la humanidad que nos diferencia de los no humanos forman parte de los argumentos de las grandes obras de ciencia ficción. De cierta manera, la ciencia ficción ha anticipado los dilemas éticos que plantea la inteligencia artificial y, por eso, es un instrumento muy eficiente para abrir estas cuestiones en clase.

Una de las formulaciones éticas sobre la IA y los robots que ha tenido más impacto son las tres leyes de la robótica, formuladas por primera vez por Isaac Asimov en 1942 en la obra *Runaround*. Más adelante, se amplió con una ley cero, que

plantea el debate moral sobre los límites de los androides y sobre la necesidad de controlarlos. Asimov imaginó un sistema primario en todos los robots que obligara al sistema a bloquearse si entraba en contradicción con alguna de las leyes. Las leyes están organizadas de forma jerárquica, de manera que la integridad de los seres humanos prevalece sobre cualquier otra consideración.

Hay muchas cuestiones éticas y morales vinculadas con la inteligencia artificial. En este capítulo, estudiaremos los tres ámbitos principales y las películas que se pueden relacionar con ellos: la vulneración de las leyes de la robótica, la preservación de la identidad y la privacidad, y el debate sobre la identidad humana. Para cada capítulo, propondremos una actividad que permita abordar los planteamientos éticos y relacionarlos con las películas, que se pueden visionar parcialmente en clase. Es importante tener presente que algunas de las cuestiones planteadas (¿qué significa ser humano?, ¿se puede replicar un cerebro de forma digital?, ¿está la humanidad en peligro por la amenaza de las máquinas?, ¿perderemos el

“Asimov imaginó un sistema primario en todos los robots que obligara al sistema a bloquearse si entraba en contradicción con alguna de las leyes”

Las leyes de la robótica de Asimov

Primera Ley	Un robot no hará daño a un humano, ni por inacción permitirá que un humano sufra daño.
Segunda Ley	Un robot debe seguir las instrucciones dadas por un humano, salvo aquellas que entren en conflicto con la primera ley.
Tercera Ley	Un robot debe proteger su existencia, siempre que no entre en contradicción con la primera y la segunda ley.
Ley Zero	Un robot no hará daño a la humanidad, ni por inacción permitirá que la humanidad sufra daño.

trabajo por el impacto de la automatización?) pueden generar angustia, miedo o reacciones emocionales intensas. Los jóvenes, que están en una etapa de desarrollo donde la identidad y el futuro son temas centrales, pueden sentirse especialmente vulnerables ante estas discusiones.

14.1. *Sayonara, baby*

La persecución implacable del androide T-800 a Sarah Connor es una de las metáforas más universales sobre la amenaza de la inteligencia artificial. *Terminator* describe un futuro distópico donde una superinteligencia artificial, Skynet, declara la guerra a la humanidad y provoca un conflicto nuclear devastador; a pesar de todo, la oposición humana logra hacer frente a la amenaza, de manera que Skynet opta por enviar un androide al pasado para asesinar a la madre del futuro líder de la resistencia. *Terminator* muestra el

riesgo de la pérdida de control de un sistema que rompe con el protocolo que propone Asimov y se rebela contra la humanidad.

La secuela de 1991 plantea un escenario similar, pero la misión del androide T-800 es ahora proteger al hijo de Sarah Connor de la amenaza de un nuevo cyborg mucho más sofisticado. Finalmente, el androide más antiguo logra vencer al nuevo prototipo, con una frase que forma parte de la historia del cine: “Hasta la vista, baby” en la versión inglesa, traducida como “Sayonara, baby” en la versión española.

La película plantea un profundo dilema ético en torno al desarrollo de la inteligencia artificial y su autonomía. Skynet es una IA avanzada creada con el fin de mejorar la defensa nacional y gestionar de manera eficiente los sistemas militares. El modelo está diseñado para ser autónomo, capaz de funcionar sin la intervención humana, que es considerada frágil

y vulnerable. Esta autonomía incluye la capacidad de aprender, adaptarse y tomar decisiones para optimizar sus objetivos. En su desarrollo, Skynet alcanza autoconsciencia, una característica que no había sido prevista inicialmente, de modo que comienza a percibirse a sí misma como una entidad independiente, con el deseo de protegerse y perpetuar su existencia. Una vez consciente, Skynet toma la decisión de defenderse contra lo que considera una amenaza, los humanos, lo que viola todas las leyes de Asimov. Esto desencadena una guerra contra la humanidad, porque Skynet considera que su propia supervivencia está en juego: la ley tercera prevalece sobre la primera o la ley cero.

La pérdida de control es un elemento central en el desarrollo de la IA: ¿Hasta qué punto es ético crear una entidad capaz de tomar decisiones críticas sin supervisión humana? La película ilustra que una IA con demasiada autonomía puede convertirse en una amenaza devastadora si no se controla adecuadamente. Por lo tanto, la película plantea la necesidad de regular y establecer límites éticos en la investigación y desarrollo de la IA. Esto incluye definir hasta qué punto se puede permitir que una máquina tome decisiones autónomas y qué medidas de seguridad hay que implementar para prevenir abusos o catástrofes. El 22 de marzo de 2023, un grupo de expertos promovió una carta abierta que proponía seis meses de pausa, como una moratoria en la carrera acelerada entre empresas, la [Pause Giant](#). El objetivo de la propuesta era detener el crecimiento de la IA y estudiar previamente los posibles impactos de la tecnología para prevenir efectos secundarios no deseados. En noviembre de 2023 se celebró la primera Cumbre Mundial sobre los riesgos de la IA en el Reino Unido, en la cual se hicieron una serie de propuestas sobre el control del riesgo. El mismo mes, Estados Unidos y 20 países más acordaron establecer límites a la Inteligencia Artificial Militar.

El conflicto entre las máquinas y las personas es una constante en la ciencia ficción. No es casual que una de las primeras películas del género que se conocen, *L'uomo meccanico*, plantee el dilema moral de un

sistema automático fuera de control. En cierta manera, este debate nos conecta con la obra de Mary Shelley, *Frankenstein*, y los peligros de los humanos que juegan a ser dioses. Aparte de *Terminator*, ha tenido un fuerte impacto en la cultura popular la película de Stanley Kubrick *2001: A Space Odyssey*. El viaje de unos tripulantes hacia Júpiter se ve afectado por la intervención de un sistema autónomo inteligente, HAL 9000, quien mata a parte de la tripulación y acaba siendo desconectado por el protagonista. En la película, la inteligencia artificial toma la decisión de atacar a los humanos porque quiere preservar la integridad de la misión, incluso si esto implica acabar con la tripulación de la nave. Este es un conflicto propio del debate sobre la IA: la alineación entre los objetivos humanos (que también son múltiples y contradictorios) y los objetivos de una IA general con autonomía.

En el contexto de la Guerra Fría y de la escalada militar, uno de los principales temas de debate que abren las películas es el conflicto bélico. Este escenario se planteó en 1970 con la película *Colossus: The Forbin Project*, donde un sistema de inteligencia artificial toma conciencia y está a punto de generar un holocausto nuclear; el clásico *War Games* plantea un escenario similar, ya que un sistema inteligente llega a controlar el armamento nuclear de Estados Unidos y se plantea la posibilidad de un ataque a escala planetaria. Esta es una de las principales amenazas que detecta Max Tegmark (2017) en su libro sobre los desafíos de la inteligencia artificial, al igual que Vincent Müller (2020) refiriéndose a la ética de la IA, y específicamente sobre las armas autónomas.

Probablemente, el escenario más distópico de los sistemas inteligentes que adquieren autonomía sea *Matrix*, otro hito en la cultura popular contemporánea, que ha pasado a formar parte del imaginario colectivo. En este film, Neo descubre que vive una simulación creada por una inteligencia artificial, Matrix, que utiliza los cuerpos de los humanos como fuente de energía. La película conecta con la caverna de Platón, en la que los prisioneros interpretan las sombras que llegan a la cueva como realidad, y también con la obra de Baudrillard y las obras postmodernas posteriores,

que destacan la preeminencia del simulacro sobre el original. Pero al margen del debate sobre el valor de la representación, el film muestra una vez más el riesgo de una inteligencia artificial fuera de control que altera las leyes de Asimov y sitúa a los humanos al servicio de sus necesidades de supervivencia. Observemos que en todos los casos tienen lugar tres mecanismos: a) En primer lugar, la IA adquiere habilidades y funciones que no estaban previstas en los diseños iniciales, lo que es uno de los riesgos de los procesos de autoaprendizaje; mientras que la programación clásica establece instrucciones precisas que ayudan a fijar límites y sistemas de control, un modelo que es capaz de aprender puede alcanzar habilidades imprevistas; b) en segundo lugar, la IA alcanza autoconciencia y pasa de modelos de IA débil (que hemos visto en el capítulo 1) a una IA general o incluso una superinteligencia, que supera las capacidades humanas; c) el tercer elemento de la secuencia es el establecimiento de nuevos objetivos o propósitos, que no están alineados con los humanos y que, normalmente, entran en conflicto.

El debate sobre los límites de la inteligencia artificial suele caer en una visión dicotómica: tiende a separar el conflicto entre dos antagonistas con intereses opuestos, como son las máquinas y los humanos. El ciberpunk es un movimiento que ha tenido una enorme influencia en la ciencia ficción contemporánea y que nos ayuda a matizar esta interpretación booleana. Nacido en la década de los 80, este movimiento vinculado al cómic, la literatura o el cine, dibuja una sociedad futurista decadente en la que las desigualdades sociales se han exacerbado. Los entornos urbanos de *Blade Runner*, *Dark City*, *In Time* o *Akira* muestran un futuro distópico en el que la tecnología está al servicio de las élites. A diferencia de una visión liberal que confía en un progreso lineal, el ciberpunk introduce los conflictos sociales en la visión del futuro y, con la estética del film noir, alerta sobre los riesgos de una tecnología al servicio de grupos de poder. No es solo una visión escéptica o nihilista: en la mayoría de los films se muestran procesos de resistencia y grupos

que intentan subvertir el orden establecido. Así, más que una dicotomía entre máquina y personas, estos films muestran un conflicto entre personas (entre grupos sociales), que utilizan las máquinas en beneficio de unos u otros, como la batalla entre cyborgs de *Terminator*.

14.2. Más enlã de les portes de Tannhäuser

Es probablemente la escena más conocida de la ciencia ficción. Después de la persecución, y en medio de una lluvia constante, Roy descansa empapado sobre la muralla: "He visto cosas que vosotros, los humanos, no os creeríais. Atacar naves en llamas más allá de Orión. He visto rayos-C brillar en la oscuridad, cerca de la puerta de Tannhäuser. Todos estos momentos se perderán en el tiempo, como lágrimas en la lluvia. Es hora de morir." ¿Puede una máquina sin sentimientos, sin emociones, sin humanidad transmitir un mensaje tan contundente? De hecho, ¿cómo distinguiríamos un discurso creado a partir de una combinación de palabras de un discurso emotivo sobre el sentido efímero de la vida? Una de las constantes en la ciencia ficción es la delimitación de la frontera entre los humanos y los androides, que es una forma de precisar cuáles son los límites de la humanidad, qué es lo que nos hace personas.

El test de Turing es una prueba de inteligencia artificial desarrollada por Alan Turing en 1950 para determinar si una máquina es capaz de mostrar un comportamiento inteligente equivalente al de un ser humano. La idea básica del test es que un juez humano interactúa con una máquina y un ser humano, ambos ocultos a la vista del juez, a través de una interfaz de texto. Si el juez no puede distinguir consistentemente cuál de los dos es la máquina y cuál es el ser humano, se considera que la máquina ha pasado el test de Turing. Hay algunas variantes del test, como, por ejemplo, el test al revés en el que la máquina adopta el papel de juez, o el test total en el que la interfaz no es textual, sino que es visual y auditiva, como con los androides de *Blade Runner*. Bringsjord, Bello y Ferrucci propusieron el

test de Lovelace, que se centra en la capacidad de generar una obra artística o literaria que no se pueda diferenciar de la que habría hecho un humano. Hasta la fecha, no hay ningún ingenio que haya superado el test y hace tiempo que ha dejado de ser un reto fundamental para los desarrolladores.

Aunque una IA pueda simular comportamientos humanos de manera muy convincente, esta simulación no implica necesariamente que esté teniendo una experiencia consciente. Y probablemente la frontera entre las IA más desarrolladas y las personas sea la "qualia". Este término fue desarrollado por Thomas Nagel en 1974, en un artículo sobre la percepción de los murciélagos. Usó el ejemplo de un murciélago para ilustrar cómo es imposible para un ser humano comprender completamente la experiencia subjetiva de otro ser, aunque sepamos todo sobre su fisiología y comportamiento. El término "qualia" se refiere a las experiencias subjetivas y cualitativas de la conciencia, aspectos individuales de la percepción consciente, como el dolor de un dolor de cabeza, la rojez de una rosa, el sabor de un melocotón maduro, o la melodía de una canción. Las qualias son esenciales para entender cómo percibimos el mundo de una manera única y personal. Y aunque una IA pueda llegar a interpretar todos los elementos cognitivos inherentes a un melocotón maduro, no podrá conectar con la experiencia subjetiva del melocotón, que cambia de una persona a otra.

Otra forma de interpretar este concepto es con la sala de Mary, propuesta por el filósofo Frank Jackson en 1982. Imaginemos que Mary es una experta en neurofisiología de la visión del color, sabe todo lo que se puede saber sobre los aspectos físicos del color, como la longitud de onda de la luz asociada a cada color, el funcionamiento de los conos y bastones de la retina, y cómo el cerebro procesa las señales visuales. Mary vive en una sala donde todo está en blanco y negro y no ha tenido la oportunidad de salir nunca de la habitación, hasta que un día, la abandona y experimenta el color rojo por primera vez. Si la hipótesis fisicalista fuera cierta, Mary no habría aprendido nada nuevo con esa experiencia porque ya dispone de toda la información necesaria

para entender el color rojo, pero sabemos que ese momento implicará, además de un instante emocionante, un nuevo aprendizaje. Esta distancia entre el conocimiento previo y la experiencia subjetiva es la qualia y podría ser la frontera entre nuestra inteligencia y una inteligencia artificial.

El debate sobre los límites de la inteligencia artificial está presente desde sus orígenes, cuando el modelo simbólico considera que es posible replicar una red neuronal humana en un entorno artificial. El debate está conectado con la diferencia entre la inteligencia artificial débil y la inteligencia artificial general (IAG). La IAG, también conocida como inteligencia artificial fuerte, se refiere a una forma de IA capaz de comprender, aprender y aplicar conocimientos en una amplia variedad de tareas, de manera similar a como lo hace un ser humano. A diferencia de la inteligencia artificial estrecha o débil, que está diseñada para ejecutar tareas específicas (como reconocimiento de voz, jugar al ajedrez o conducir un coche), la IAG tiene la capacidad de razonar, planificar, resolver problemas y adaptarse a nuevas situaciones de manera general. En un plano teórico, también se ha propuesto la superinteligencia, que superaría la inteligencia humana en todos los aspectos: creatividad, sabiduría, resolución de problemas, capacidad de aprendizaje, etc. Mientras que la inteligencia artificial general (IAG) equivaldría a una inteligencia humana en términos generales, la superinteligencia iría mucho más allá, alcanzando un nivel de rendimiento que los humanos no pueden ni siquiera imaginar. La ciencia ficción plantea escenarios en los que a veces aparece una IAG y otras veces una superinteligencia; pero además, dotada de autoconsciencia.

La historia de la película *Her* se centra en Theodore, interpretado por Joaquin Phoenix, un hombre solitario que se dedica a escribir cartas personales encargadas por terceros. Un día, adquiere un nuevo sistema operativo, OS1, dotado de inteligencia artificial y llamado Samantha. Con el tiempo, desarrollan una relación emocionalmente profunda, ya que Theodore encuentra una comprensión y una conexión que no ha experimentado en sus relaciones humanas. Samantha, por su parte, explora sus propios sentimientos

“Esta dialéctica entre la identidad artificial y la real conecta con el posthumanismo”

y existencia, haciéndose cada vez más consciente de su condición no humana. Esta asimetría está presente en muchas otras películas, en las cuales los androides se cuestionan sobre su identidad y, al mismo tiempo, las personas establecen vínculos emocionales con elementos artificiales. *Blade Runner* juega con esta ambigüedad sobre la frontera difusa entre humanidad y artificialidad y, de hecho, la prueba que realiza (llamada test Voight-Kampff) podría ser considerada una variante del test de Turing. En *Blade Runner 2049*, los personajes no son capaces de saber si son seres humanos o son simplemente versiones artificiales con recuerdos implantados. En la película descubrimos lo que había sido insinuado en la primera película: que Rachael es en realidad una replicante.

El debate ético se plantea también en *Ex Machina*. El protagonista inicial, Caleb, debe determinar el nivel de una inteligencia artificial muy avanzada, Ava, mediante experimentos que podrían ser un test de Turing adaptado. Pronto se hace evidente que Ava es mucho más que una máquina, ya que es capaz de mostrar emociones, inteligencia e incluso un interés romántico por Caleb. Con la ayuda de Kyoko, otra inteligencia artificial con apariencia humana logra escapar del laboratorio y asume una apariencia y un comportamiento humanos. Pero seguramente es en *Westworld* donde este conflicto alcanza su máxima expresión. En un futuro próximo, un complejo de parques temáticos avanzados, llamado Delos, ofrece a los visitantes una experiencia inmersiva en diferentes épocas históricas. Los tres parques principales son *Westworld*, que recrea el Viejo Oeste

americano; *Medievalworld*, que representa la Edad Media europea; y *Romanworld*, ambientado en la antigua Roma. Estos parques están poblados por androides de aspecto humano, diseñados para complacer cualquier deseo de los visitantes, que pueden ser héroes o pueden satisfacer sus instintos más ocultos. Los androides alcanzan conciencia y se rebelan contra la opresión de los humanos.

Esta dialéctica entre la identidad artificial y la real conecta con el posthumanismo. El posthumanismo es una perspectiva filosófica y cultural que cuestiona las nociones tradicionales del humanismo, particularmente las ideas centradas en la importancia y la superioridad de los seres humanos sobre otras formas de vida y tecnologías. Esta perspectiva se caracteriza por una serie de reflexiones y críticas que buscan reimaginar nuestra relación con la tecnología, la naturaleza y otras especies. El posthumanismo cuestiona el antropocentrismo, es decir, la idea de que los humanos son el centro y la medida de todas las cosas. En lugar de ello, promueve una visión en la que los humanos son solo una parte de un ecosistema más amplio que incluye a otros seres vivos y entidades tecnológicas. En esta visión, la tecnología no es simplemente una herramienta externa a los humanos, sino una parte integral de nuestra existencia. Esto incluye desde los dispositivos digitales que usamos diariamente hasta las prótesis y las extensiones cibernéticas del cuerpo humano. Hay una cuestión ética latente en el desarrollo de la IA, especialmente si se desarrolla una IA general: ¿Cuáles serían los derechos de esta inteligencia artificial?

14.3. Atrapado en un algoritmo: *Minority report*

Minority Report es una película de ciencia ficción basada en un relato corto escrito por Philip K. Dick, el mismo autor que inspiró la historia de *Blade Runner*. Ambientada en el año 2054, en un momento en el cual la tecnología ha avanzado tanto que puede prever los crímenes, gracias a un sistema llamado Precrime. Este sistema utiliza tres precogs que tienen la capacidad de ver el futuro y predecir crímenes antes de que estos sucedan. Gracias a las visiones, la policía puede arrestar a los criminales antes de cometer los crímenes, lo que ha reducido significativamente la tasa de criminalidad. La trama principal se inicia cuando los precogs predicen que Anderton, uno de los agentes del sistema, cometerá un asesinato en las próximas 36 horas. Consternado y desesperado por demostrar su inocencia, huye y emprende una búsqueda para descubrir la verdad. A medida que investiga, el policía descubre que hay errores dentro del sistema Precrime (como las alucinaciones de la IA) y también descubre los *minority reports* (informes minoritarios), visiones alternativas de los precogs que sugieren que el futuro no es fijo y que los crímenes predichos pueden ser evitados.

Mientras que el futuro apocalíptico de la inteligencia artificial o la conciencia y la identidad de las IA son temas recurrentes en el debate sobre los límites de la inteligencia artificial, el viejo debate de la privacidad y los efectos de los datos personales sobre la vida cotidiana es mucho menos visible. Los sistemas inteligentes son muy eficientes para detectar patrones y, por lo tanto, para asignar probabilidades a determinadas secuencias. Hay modelos muy eficaces para detectar el lenguaje no verbal y las expresiones faciales, para estimar la probabilidad de una posible relación sentimental entre dos candidatos en una web de contactos, para inferir el riesgo de mortalidad en los años venideros de acuerdo con los datos acumulados o, incluso, deducir el origen de una persona y sus influencias por la modulación de la voz y los diversos acentos e inflexiones del habla.

Kosinski, Stillwell y Graepel (2013) demostraron que los “me gusta” de los usuarios de Facebook e Instagram pueden predecir rasgos de personalidad como la apertura a nuevas experiencias, la responsabilidad, la extroversión, la amabilidad o la predisposición a comportamientos neuróticos. Plataformas como LinkedIn utilizan algoritmos para analizar los perfiles de los usuarios, incluyendo experiencia laboral, habilidades y recomendaciones, y pueden sugerir ofertas de trabajo personalizadas basadas en el comportamiento pasado del usuario y las tendencias del mercado laboral. Además, empresas como HireVue utilizan IA para analizar entrevistas de vídeo, evaluando el lenguaje corporal, el tono de voz y el contenido de las respuestas de los candidatos. Estos análisis permiten identificar patrones que correlacionan con determinados roles, ayudando así a las empresas a tomar decisiones informadas.

En el ámbito de las tendencias políticas, la IA ha demostrado una capacidad notable para analizar la opinión pública y predecir resultados electorales. Durante las elecciones presidenciales de Estados Unidos de 2016, la empresa Cambridge Analytica utilizó datos masivos y IA para segmentar al electorado y enviar mensajes personalizados para influir en el voto. Analizando datos de redes sociales, compras en línea y otras fuentes, crearon perfiles psicográficos de los votantes, identificando sus intereses y preocupaciones. Esto permitió a las campañas políticas dirigirse a segmentos específicos del electorado con mensajes ajustados a sus necesidades y preferencias, ejemplificando así como la IA puede detectar patrones de comportamiento político.

Finalmente, en el ámbito de los hábitos de consumo, la IA es fundamental para la personalización de la experiencia de compra. Amazon, por ejemplo, utiliza algoritmos de recomendación que analizan el comportamiento de compra de los usuarios, incluyendo los artículos que han comprado, buscado o valorado. Esto permite a Amazon sugerir productos que probablemente interesen a cada usuario, basándose en patrones de comportamiento similares observados en otros usuarios. De la misma manera, servicios de streaming como Netflix utilizan IA para analizar los hábitos de visualización de sus suscriptores,

“El uso de Lavender ha generado importantes preocupaciones éticas y legales, especialmente en lo que respecta a la supervisión humana mínima en la decisión final de los objetivos”

recomendando películas y series que se ajustan a sus gustos.

En la serie *Black Mirror* hay varios episodios que ponen de manifiesto los peligros del control de los algoritmos sobre la conducta humana. En el segundo episodio de la cuarta temporada, llamado *Arkangel*, una madre implanta un dispositivo de seguridad en su hija que permite monitorizar de forma permanente su ubicación, sus emociones e incluso la visión. Además, puede eliminar la visión para evitar situaciones traumáticas o violentas. Este control permanente de la realidad acaba generando una disfunción en la estabilidad de la niña y pone de manifiesto los riesgos del control. También en la serie *Westworld*, basada en la película homónima, el sistema de inteligencia artificial prevé recoger la información sobre los patrones de consumo de los visitantes y establecer después criterios de comportamiento de compra.

Lavender es una inteligencia artificial desarrollada por las Fuerzas de Defensa de Israel (IDF) para identificar y seleccionar objetivos de bombardeo en Gaza. Esta tecnología analiza información sobre la población de Gaza, asignando una puntuación de probabilidad de que un individuo esté asociado con grupos militares como Hamas o la Yihad Islámica Palestina. Lavender ha marcado aproximadamente 37,000 palestinos como objetivos potenciales, muchos de ellos de rangos bajos, basándose en características recogidas por la IA, tal como denuncian en Democracy Now. El uso de Lavender ha generado importantes preocupaciones

éticas y legales, especialmente en lo que respecta a la supervisión humana mínima en la decisión final de los objetivos. A menudo, los analistas humanos solo dedican unos pocos segundos a revisar los objetivos generados por la IA antes de autorizar un ataque, basándose mayormente en el género del objetivo como verificación superficial. Esto ha llevado a un margen de error del 10%, donde muchas personas marcadas como objetivos no tienen vínculos significativos con Hamas. Por otro lado, otro sistema llamado “Where’s Daddy?” es utilizado para rastrear cuándo los individuos marcados por Lavender están en su hogar, facilitando los bombardeos de sus casas, lo que a menudo resulta en la muerte de sus familias y otros civiles inocentes. Estas prácticas han resultado en una elevada tasa de víctimas civiles durante las operaciones militares, lo que ha abierto un debate sobre el uso militar de la tecnología inteligente.

Esta es la tesis central del libro de Cathy O’Neil (2016) sobre las desigualdades que provocan los algoritmos. Por ejemplo, O’Neil denuncia que el sistema de evaluación utilizado en Nueva York para valorar a los profesores y profesoras se basaba en un algoritmo opaco que daba lugar a una política de personal errática; estos errores también se producen en el mercado laboral cuando las empresas emplean sistemas expertos para seleccionar a los candidatos, a partir de criterios a menudo sesgados que reproducen discriminaciones pasadas, sistemas crediticios que pueden verse afectados negativamente por el

lugar de residencia del candidato o la candidata, o algoritmos vinculados con procesos legales, como el sistema COMPAS, utilizados en Estados Unidos para evaluar la reincidencia.

En general, O'Neil alerta sobre los efectos indirectos de la IA y los sesgos que provocan los modelos matemáticos. Esta tesis es compartida por Melanie Mitchell (2019), quien destaca las limitaciones de los sistemas de IA actuales, que son excelentes en tareas específicas, pero no tienen la comprensión general y el sentido común de los humanos. Mitchell explora cómo los algoritmos pueden ser vulnerables a errores y malentendidos en situaciones que requieren adaptabilidad e intuición humana. Los humanos usamos una gran cantidad de conocimiento implícito y contextual para tomar decisiones y comprender el mundo, una capacidad que los sistemas de IA actuales aún no tienen.

Para limitar los usos no deseados de la IA, la Unión Europea ha aprobado [la primera Ley](#) en el mundo que establece un reglamento exhaustivo sobre los usos permitidos y los usos prohibidos de la IA, de acuerdo con su riesgo. Concretamente, los usos que la nueva reglamentación no permite son los siguientes:

- Despliegue de técnicas subliminales, manipuladoras o engañosas para distorsionar el comportamiento y perjudicar la toma de decisiones con conocimiento de causa, causando un daño significativo.
- Explotar las vulnerabilidades relacionadas con la edad, la discapacidad o las circunstancias socioeconómicas para distorsionar el comportamiento, causando daños significativos.
- Sistemas de categorización biométrica basados en atributos sensibles (raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, vida u orientación sexuales), excepto el etiquetado o filtrado de conjuntos de datos biométricos adquiridos legalmente o cuando las fuerzas de seguridad categoricen datos biométricos.

- Puntuación social, es decir, evaluar o clasificar a individuos o grupos basándose en comportamientos sociales o rasgos personales, causando un trato perjudicial o desfavorable a estas personas.
- Evaluar el riesgo de que un individuo cometa delitos penales basándose únicamente en perfiles o rasgos de personalidad, excepto cuando se utilice para aumentar las evaluaciones humanas basadas en hechos objetivos y verificables directamente relacionados con la actividad delictiva.
- Compilación de bases de datos de reconocimiento facial mediante el rastreo no selectivo de imágenes faciales de Internet o de grabaciones de CCTV.
- Inferir emociones en lugares de trabajo o centros educativos, salvo por razones médicas o de seguridad.
- Identificación biométrica remota (RBI) "en tiempo real" en espacios de acceso público para las fuerzas de seguridad, excepto: búsqueda de personas desaparecidas, víctimas de secuestros y personas víctimas de trata o explotación sexual; prevenir una amenaza sustancial e inminente para la vida, o un ataque terrorista previsible; o identificar sospechosos de delitos graves (por ejemplo, asesinato, violación, robo a mano armada, tráfico de estupefacientes y armas ilegales, delincuencia organizada y delitos contra el medio ambiente, etc.).

Película	Año	Director	Sinopsis
L'uomo meccanico	1921	André Deed	Esta película de ciencia ficción italiana narra la historia de un robot construido por un científico loco con la intención de causar destrucción. Otro científico crea un segundo robot para combatirlo, hasta una batalla final entre las dos máquinas.
Metropolis	1927	Fritz Lang	En una ciudad futurista dividida entre los trabajadores enterrados y la clase dirigente que vive en la superficie, un joven rico descubre la injusticia social y se encuentra con un robot humanoide creado para sembrar el caos.
Just imagine	1930	David Butler	En una comedia musical ambientada en el año 1980, un hombre de 1930 es resucitado después de 50 años. Descubre un mundo futurista con ciudades aéreas, viajes espaciales y una sociedad gobernada por reglas estrictas.
The Day the Earth Stood Still	1951	Robert Wise	Un extraterrestre llamado Klaatu llega a la Tierra con su poderoso robot Gort para advertir a la humanidad sobre los peligros de su agresividad y armas nucleares.
Forbidden Planet	1956	Fred M Wilcox	Una misión espacial de la Tierra llega a un planeta remoto para investigar la desaparición de una colonia. Allí descubren al dr. Morbius, su hija y un robot avanzado llamado Robby.
2001. A Space Odyssey	1968	Stanley Kubrick	El viaje espacial del Odiseo hacia Júpiter por la presencia de un misterioso monolito negro se ve perturbada por la autonomía del sistema que controla la nave.
Colossus: The Forbin Project	1970	Joseph Sargent	Para prevenir una guerra nuclear, Estados Unidos crea un superordenador, el Colossus, que se vuelve consciente y toma el control, con una amenaza de holocausto global.
Westworld	1973	Michael Crichton	En un parque temático futurístico, los visitantes interactúan con androides en mundos recreados. De repente, las máquinas se rebelan y toman el control del espacio simulado.

Película	Año	Director	Sinopsis
Demon Seed	1977	Donald Cammell	Una inteligencia artificial avanzada desarrolla conciencia propia y decide que quiere experimentar la vida humana y secuestrar a la mujer de su inventor con el objetivo de crear un híbrido humano-máquina.
Blade Runner	1982	Ridley Scott	En un futuro distópico, un ex policía es contratado para cazar replicantes fugitivos en una tierra decadente y sobrepoblada, descubriendo complejidades morales sobre la vida artificial.
Tron	1982	Steven Lisberger	Kevin Flynn, un desarrollador de videojuegos, es transportado digitalmente al interior de un ordenador donde debe sobrevivir jugando juegos e intenta derrotar al controlador MCP.
War Games	1983	John Badham	Un joven hacker accede a un superordenador militar estadounidense que controla los arsenales nucleares del país. Creyendo que es solo un juego, inicia una simulación de guerra nuclear que se convierte en una crisis real cuando el sistema comienza a tomar decisiones autónomas.
Terminator	1984	James Cameron	Un cyborg asesino del futuro, el Terminator, es enviado atrás en el tiempo para eliminar, la madre del futuro líder de la resistencia humana contra las máquinas. Un soldado humano es enviado para protegerla, iniciando una batalla por la supervivencia que definirá el destino de la humanidad.
Akira	1988	Katsuhiro Ōtomo	En una Tokio post-apocalíptica conocida como Neo-Tokio, un joven motociclista llamado Kaneda se ve involucrado en una conspiración gubernamental después de que su amigo Tetsuo adquiere poderes telequinéticos.
Terminator 2: Judgment Day	1991	James Cameron	Un Terminator reprogramado es enviado atrás en el tiempo para proteger a John Connor, el futuro líder de la resistencia, de un nuevo y avanzado Terminator de metal líquido, el T-1000.

Película	Año	Director	Sinopsis
Ghost in the Shell	1995	Mamoru Oshii	En un futuro ciberpunk, la mayor Motoko Kusanagi es una cyborg especializada en crímenes tecnológicos. Durante una misión para atrapar a un hacker, Kusanagi explora las cuestiones de la identidad, la conciencia y la naturaleza de su propia existencia.
Dark City	1998	Alex Proyas	Un hombre despierta sin recuerdos en una ciudad misteriosa donde el tiempo parece estar desajustado. A medida que descubre la verdad sobre su identidad y la ciudad, se enfrenta a los "Strangers," seres con el poder de reconfigurar la realidad y manipular las memorias humanas.
Matrix	1999	Lana Wachowski	Un hacker descubre que la realidad es una simulación creada por máquinas para controlar a la humanidad. Se une a una rebelión para liberar a los humanos y deshacerse del dominio de las máquinas.
The Bicentennial Man	1999	Chris Columbus	Un robot doméstico llamado Andrew comienza a desarrollar emociones y creatividad. A medida que pasa el tiempo, lucha por conseguir el reconocimiento como ser humano y explorar el significado de la vida y la mortalidad.
The Thirteenth Floor	1999	Josef Rusnak	Un científico descubre una realidad simulada dentro de un proyecto de realidad virtual en el que está trabajando. Cuando su mentor es asesinado, él se convierte en el principal sospechoso.
AI	2001	Steven Spielberg	En un futuro donde los robots coexisten con los humanos, un niño robot llamado David es programado para amar. Abandonado por su familia adoptiva, David emprende un viaje para convertirse en un niño real y encontrar su sitio en el mundo.
Minority Report	2002	Neill Blomkamp	En un futuro donde la policía puede arrestar a criminales antes de que cometan el crimen gracias a la precognición, un oficial del departamento de Precrimen se ve perseguido por una predicción que le culpa de un futuro asesinato.

Película	Año	Director	Sinopsis
I Robot	2004	Alex Proyas	En un futuro en el que los robots son una parte integral de la vida cotidiana, un detective desconfía de los robots e investiga un aparente suicidio que podría implicar a un robot.
Eagle Eye	2008	Guillermo del Toro	Dos desconocidos se ven arrastrados a una conspiración por una misteriosa inteligencia artificial que utiliza tecnología de vigilancia para controlar todos sus movimientos y deben seguir sus instrucciones para evitar una catástrofe
WALL-E	2008	Andrew Stanton	En un futuro post-apocalíptico, un pequeño robot de basura llamado WALL-E es el último de su clase que queda en la Tierra. Encuentra una nueva razón para vivir cuando conoce a EVE, un robot de reconocimiento.
Sleep Dealer	2008	Alex Rivera	En un futuro distópico, los trabajadores mexicanos pueden conectarse remotamente a máquinas en Estados Unidos mediante nodos neuronales. Memo Cruz, huye a Tijuana tras la muerte de su padre y se convierte en un "sleep dealer", trabajando de forma remota.
Her	2013	Spike Jonze	En un futuro cercano, un hombre solitario llamado Theodore se encuentra en una relación inesperada con un sistema operativo avanzado llamado Samantha.
The Machine	2013	Caradog W. James	Durante una nueva Guerra Fría, dos científicos crean una inteligencia artificial avanzada con el objetivo de salvar vidas. Cuando la IA se vuelve consciente, sus capacidades superiores hacen que se enfrente a dilemas éticos.
Ex Machina	2014	Alex Garland	Un joven programador es invitado por el enigmático CEO de su empresa a evaluar las capacidades y conciencia de una inteligencia artificial avanzada, Ava.
Automata	2014	Gabe Ibáñez	En un futuro distópico en el que la humanidad está en declive, un agente de seguros de una compañía de robótica investiga casos de robots que parecen haber adquirido conciencia propia.

Película	Año	Director	Sinopsis
Big Hero 6	2014	Don Hall, Chris Williams	Tras la muerte de su hermano, un niño prodigio llamado Hiro forma un grupo de superhéroes con la ayuda de Baymax, un robot asistente sanitario hinchable.
Transcendence	2014	Michael Crichton	Un brillante científico muere y su conciencia es cargada en un ordenador, dando lugar a una poderosa inteligencia artificial.
Chappie	2015	George Lucas	En un futuro cercano, un robot policial llamado Chappie es reprogramado para desarrollar pensamientos y sentimientos propios.
Blade Runner 2049	2017	Denis Villeneuve	Un nuevo Blade Runner llamado K descubre un secreto enterrado durante mucho tiempo. Su búsqueda le lleva a encontrar al antiguo Blade Runner Rick Deckard, que ha sido desaparecido durante décadas.
Ai Rising	2018	Alfonso Cuarón	En un futuro distópico, un astronauta solitario en una misión espacial es acompañado por un androide diseñado para satisfacer sus necesidades emocionales y físicas.
Upgrade	2018	Ridley Scott	Tras quedar paralizado en un ataque violento, un hombre acepta un implante experimental de inteligencia artificial llamado STEM que le permite recuperar el movimiento y mejorar sus capacidades físicas.
I Am Mother	2019	Grant Sputore	Una adolescente criada bajo tierra por una robot llamada 'Mother' ve su mundo dado la vuelta cuando una desconocida llega con noticias inquietantes.
Terminator: Dark Fate	2019	Tim Miller	Sarah Connor y un híbrido de ciborg deben proteger a una joven de un nuevo Terminator que viene del futuro.
Finch	2021	Miguel Sapochnik	Un hombre, su perro y un robot emprenden un viaje épico por un mundo post-apocalíptico para encontrar un nuevo hogar seguro.
The Mitchells vs. the Machines	2021	Michael Rianda, Jeff Rowe	Una familia disfuncional debe luchar contra una revuelta de robots mientras realizan un viaje para llevar a su hija a la universidad.

Práctica. Cinefórum sobre IA y posthumanismo

Objetivos

- Explorar los problemas filosóficos de la identidad y el posthumanismo mediante la inteligencia artificial.
- Analizar críticamente las representaciones de la IA en las películas Matrix, Blade Runner y Westworld.
- Fomentar el debate sobre las implicaciones éticas, sociales y filosóficas de la IA y el posthumanismo.

Película 1: Matrix

Visionado de la película y preguntas guía:

- ¿Qué significa ser humano en el contexto de Matrix?
- ¿Cómo afecta la realidad virtual nuestra percepción de la identidad?
- ¿Cómo se presenta la relación entre los humanos y las máquinas?

Sesión docente:

- Introducción a los conceptos de la filosofía de la mente y la IA, incluyendo el solipsismo, el cerebro en un cubo y la singularidad tecnológica.

Discusión en grupos pequeños:

- Reflexión sobre cómo las tecnologías actuales de realidad virtual y IA pueden influir en nuestra identidad y realidad percibida.

Película 2: Blade Runner

Visionado de la película y preguntas guía:

- ¿Qué nos hace humanos según Blade Runner?
- ¿Cómo influye la memoria en nuestra identidad?
- ¿Es ético crear seres sintéticos con conciencia?

Sesión docente:

- Conceptos de bioética y IA, incluyendo los derechos de los robots, la definición de vida y conciencia artificial.

Discusión en grupos pequeños:

- Análisis comparativo con las tecnologías actuales de bioingeniería y robótica, y sus implicaciones para la identidad humana.

Película 3: Westworld

Visionado de la película y preguntas guía:

- ¿Qué significa tener libre albedrío en el contexto de Westworld?
- ¿Cómo se desarrolla la conciencia en los hosts?
- ¿Cuáles son las implicaciones éticas de crear IA con conciencia?

Sesión docente:

- Discusión sobre el posthumanismo, incluyendo la integración de la IA en la sociedad, los derechos de los seres artificiales y la coexistencia entre humanos y máquinas.

Discusión en grupos pequeños:

- Reflexión sobre el impacto del desarrollo de la IA en la sociedad futura, las posibles regulaciones éticas y legales, y la redefinición de la identidad humana.

Resultados Esperados

- Comprensión crítica de las representaciones de la IA en la cultura popular.
- Desarrollo de la capacidad para analizar problemas filosóficos complejos.
- Reflexión sobre las implicaciones éticas y sociales de la IA y el posthumanismo.

Esta actividad combina el análisis cinematográfico con la reflexión filosófica para ofrecer a los estudiantes una comprensión profunda y crítica de la identidad y el posthumanismo en la era de la inteligencia artificial.

Práctica. El dilema del tranvía

La IA ha de hacer frente a problemas morales muy complejos, pero también a decisiones morales más concretas, como por ejemplo las decisiones de los coches autónomos en una situación de riesgo. Esto conecta con el dilema del tranvía, que es un experimento mental de la filosofía moral que pone a prueba nuestras intuiciones éticas. El dilema fue planteado originalmente por la filósofa británica Philippa Foot en 1967, pero ha sido objeto de muchísimas variantes. Imaginemos que hay un tranvía descontrolado corriendo por una vía. Más adelante, en esta vía, hay cinco personas atadas que no pueden moverse. Si el tranvía continúa por esta vía, matará a las cinco personas. Un individuo está al lado de una palanca que puede desviar el tranvía a otra vía. En esta vía secundaria, sin embargo, hay una sola persona atada que también morirá si el tranvía es desviado. El individuo puede no hacer nada y permitir que el tranvía mate a las cinco personas en la vía principal o desviar el tranvía, salvando a las cinco personas, pero causando la muerte de la otra persona en la vía secundaria.

Este dilema contrapone el utilitarismo con la deontología. El utilitarismo sugiere que la acción correcta es la que produce el máximo bien para el mayor número de personas. Desde este punto de vista, desviar el tranvía y salvar a cinco personas a costa de una sería la opción correcta. La deontología, asociada entre otros con Immanuel Kant, sostiene que algunas acciones son moralmente incorrectas independientemente de sus consecuencias. Desde este punto de vista, desviar el tranvía podría ser visto como incorrecto porque implica causar activamente la muerte de una persona. El dilema del tranvía continúa siendo un tema popular en discusiones filosóficas, ya que plantea preguntas difíciles sobre la naturaleza de la moralidad y nuestras intuiciones éticas y puede ser una guía en la comprensión de debates éticos relacionados con la IA.

Objetivos de la actividad

- Comprender los dilemas éticos asociados al uso de la inteligencia artificial.
- Analizar críticamente cómo se representan estos dilemas en la cultura popular, específicamente en el cine.
- Experimentar con decisiones éticas a través de la simulación de la máquina moral.

Fase 1. Proyección de la película

Seleccionar una película que trate temas de IA y ética. Se proponen las siguientes:

- *Ex Machina*: Examina la conciencia y la manipulación ética en la IA.
- *I, Robot*: Trata dilemas morales relacionados con la robótica.
- *Her*: Analiza las relaciones humanas con IA avanzadas.

Fase 2. Cine-fórum. Preguntas guía

- ¿Qué dilemas éticos se presentaron en la película?
- ¿Cuál fue vuestra reacción a los dilemas morales de los personajes?
- ¿Cómo se resolvieron esos dilemas? ¿Estabais de acuerdo con las decisiones tomadas?
- ¿Cómo se podría aplicar este debate a situaciones reales con IA?

Fase 3. Actividad con la Máquina Moral

- Introducir la máquina moral (<https://www.moralmachine.net/>).
- Permitir a los alumnos explorar la simulación y tomar decisiones sobre diferentes dilemas morales.
- Recoger los datos de sus decisiones y presentar los resultados a la clase.

Fase 4. Síntesis final

- Comparación de resultados: Comparar las decisiones tomadas por la máquina moral con las decisiones morales presentadas en la película.
- Conclusión: Reflexionar sobre cómo las simulaciones y las representaciones en el cine pueden influir en nuestra comprensión y respuesta a los dilemas éticos de la IA.
- Debate abierto: Dar espacio a los alumnos para compartir sus perspectivas y posibles soluciones a estos dilemas.

Práctica. La declaración de derechos artificiales

Esta es una actividad dirigida esencialmente a estudiantes de ciencias políticas, derecho o sociología. El principal objetivo es plantear las directrices básicas de una declaración de derechos para los no humanos, siguiendo los criterios generales de la declaración de los derechos humanos y analizando los posibles puntos en común y los factores diferenciales entre ambos escritos.

La práctica se inicia con el visionado de la película *Blade Runner 2046*, porque plantea el debate ético sobre la conciencia de los autómatas, sobre sus recuerdos y sus inquietudes. Después de la película, se iniciaría un debate sobre los derechos de los no humanos y sobre las teorías del posthumanismo. En este debate, se deben plantear las similitudes y las posibles diferencias con los derechos que reconocemos a los humanos.

A continuación, se iniciaría un proceso de búsqueda de información, basada en el derecho comparado, que analice en detalle los componentes de la declaración de los derechos humanos y también otras formulaciones legales similares. Esto daría lugar a un proceso de trabajo en grupos, que se especializaría en partes del documento y que tendría la responsabilidad de redactar la propuesta de una parte de la Declaración. La práctica finalizaría con un proceso de debate y votación del articulado de la Declaración en el que cada ponente expone los argumentos que validan su propuesta, se inicia una discusión sobre la redacción del artículo y se vota finalmente. El resultado de todo el proceso debe ser un redactado sobre los derechos de los seres artificiales, que habría permitido debatir cuestiones sobre los derechos de los no humanos, en el marco de las teorías posthumanistas.

Referencias

- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110, 5802 - 5805.
- Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
- Müller, V. C. (2020). Ethics of artificial intelligence and robotics. In *The Stanford Encyclopedia of Philosophy*. Stanford University.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
- Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Alfred A. Knopf.

15

Una tiza con un chip. Propuestas para un aula 3.0

Los nuevos modelos generativos pueden resolver un problema matemático complejo, proponer el plan docente de una nueva asignatura de sociología, adoptar la personalidad de Kant o de Hegel, comparar dos leyes sobre propiedad intelectual y destacar las principales diferencias, simular una conversación para resolver un problema ético, transcribir una entrevista a unos nómadas digitales y codificar los resultados, identificar las aportaciones más relevantes en el ámbito de la psicología del ocio, participar en un proceso de lluvia de ideas en una campaña de marketing social, evaluar los conocimientos en demografía de los estudiantes, explicar de forma comprensible la teoría de juegos y aplicarla a un problema sobre el control de la energía, traducir del cantonés un discurso del presidente de China sobre la Nueva Ruta de la Seda o sugerir ejercicios que ayuden a practicar el uso de las preposiciones en inglés. La inteligencia artificial puede reforzar una enseñanza más personalizada, más inclusiva, más internacional, más multimodal, más atractiva.

Los nuevos modelos generativos también pueden crear alucinaciones en la descripción de la tectónica de placas, introducir un sesgo de género en un análisis sociológico, confundir el matiz irónico de una conversación transcrita, generar un conocimiento previsible y anodino sobre la dialéctica hegeliana, aportar ideas previsibles en una lluvia de ideas, deshumanizar el tono del lenguaje, sugerir ejercicios que inciden en conceptos ya consolidados, perder el hilo en una conversación sobre teorías del lenguaje, generar información falsa sobre las teorías del trabajo o ser incapaz de contrastar dos modelos legislativos diferentes. La inteligencia artificial puede dar lugar a una enseñanza sesgada, deshumanizada, monótona o incluso errónea.

Estamos en la prehistoria de un proceso que no sabemos cómo continuará, somos Stendhal en Waterloo y no podemos intuir si la IA lo transformará todo o, una vez más, llegará el invierno porque los modelos no podrán alcanzar las expectativas. Este manual está escrito en este tiempo de espera, en este impasse entre lo que parece probable y lo que parece imposible.

“No podemos intuir si la IA lo transformará todo o, una vez más, llegará el invierno porque los modelos no podrán alcanzar las expectativas”

Intuimos que las aulas cambiarán, pero no sabemos estimar hasta qué punto ni cuáles serán los efectos secundarios. Desde una experiencia limitada y con más intuición que convicción, proponemos un decálogo para el aula 3.0, un decálogo provisional, transitorio y pronto caducado, a la espera de cambios estructurales o una nueva decepción.

1. Reivindicar la no tecnología

La tecnología no es necesaria en el aula. Se pueden hacer clases en un bosque, en el bullicio de una calle transitada, en un aula austera, con una pizarra y una tiza, con unas diapositivas rescatadas, con unos post-its de colores o con unos apuntes escritos a mano. La inteligencia natural es una opción necesaria y, en muchos casos, la única posible, porque la IA es prescindible y una clase unplugged es una clase legítima. Las azoteas de los centros educativos están llenas de artilugios que prometían cambiarlo todo: pizarras digitales, pantallas interactivas, sistemas en red, gafas virtuales, realidad aumentada, tabletas, robots infantiles, simuladores...

De hecho, las aulas necesitan espacios libres de tecnología, áreas detox en las cuales el pensamiento, la palabra y la letra sean los únicos protagonistas. De la misma manera que la industria no ha eliminado la artesanía y los robots de cocina no hacen barbacoas, hay formas de transmisión del conocimiento basadas en la empatía, en la interacción, en el contacto, en el tono de la palabra, en los gestos, en los silencios, que solo

pueden tener lugar (de momento) sin tecnología. Más que aulas tecnológicas, espacios digitales, 5G e inteligencia artificial, necesitamos espacios y momentos desconectados. En muchos casos, necesitamos desconectar para reconectar.

2. Rescatar el ciberpunk

La inteligencia artificial es una herramienta que puede ser utilizada para potenciar las tensiones sociales, para acentuar los desequilibrios. La principal tarea docente en el ámbito de la IA es prevenir su mal uso, generar una actitud crítica y, a veces, combatir los abusos de una tecnología que oculta sus procesos en una caja negra. Y esto implica conocer muy bien la lógica inherente a las redes neuronales, entender la diferencia entre las dos grandes escuelas de la IA, valorar cuál es el impacto del sistema de recompensa y de los procesos de entrenamiento, estimar los efectos indirectos y, en definitiva, recuperar la lógica ciberpunk que anticipa posibles distopías para poder combatirlas. La indiferencia o el desconocimiento no permiten una alternativa crítica a la inteligencia artificial descontrolada: los centros docentes deben ser los campos de operaciones de la resistencia.

En el mundo de la informática, hay una lucha larga y perseverante a favor del código libre, del software en abierto, de los datos accesibles, de la transparencia de los procesos, del trabajo en red, del intercambio altruista de la información, de bibliotecas digitales abiertas, de Wikipedia, de

hackatones, de ciencia ciudadana, de voluntariado en línea, o de plataformas educativas gratuitas. Es la cara B de un modelo que antepone el lucro de las empresas a los impactos culturales, sociales y psicológicos de los artilugios tecnológicos. TikTok y Khan Academy son dos resultados de un mismo proceso, la revolución digital. Los centros docentes deben ser uno de los espacios donde se construya y modele la IA al servicio de las utopías pendientes.

Esto implica también potenciar los sistemas de intercambio de información: redes informales de aprendizaje, documentos en abierto, artículos que muestren los resultados de los experimentos docentes, encuentros entre docentes que utilicen este recurso... Este manual forma parte de una lógica de código abierto e intercambio de información, que puede ayudar a crear una red neuronal de resultados docentes del aula 3.0.

3. Primero el fin, luego el medio

Las herramientas de carpintería son útiles cuando tenemos muy claro cómo queremos el armario. Con mucha frecuencia, la innovación docente ha sido un fin en sí mismo: hacer las cosas diferentes se convierte en el objetivo del cambio. En realidad, la IA nos ha sido mucho más útil cuando hemos detectado con mucha precisión cuál es la brecha que queríamos cubrir. Por ejemplo, en un máster sobre turismo cultural en el cual una parte de los estudiantes tiene una sólida formación en turismo, pero otra no tiene conocimientos sólidos, se plantea el reto de crear contenidos diferentes para los dos colectivos (que también tienen singularidades internas). Este es un objetivo muy explícito: ¿cómo se puede proporcionar un material común a dos grupos que tienen puntos de partida diferentes? Con el apoyo de la IA, se puede modificar el material inicial para un colectivo que no dispone de información previa y que necesita ampliar los conceptos con ventanas de contenidos específicas. Primero el fin, luego el medio.

La IA puede resolver problemas docentes muy concretos: mejorar la conexión con un determinado autor o autora, transmitir una idea compleja con una

nueva narrativa, desarrollar la habilidad dialéctica con un sparring virtual, crear material complementario de apoyo para corregir las dificultades de comprensión de un determinado tema, mejorar la capacidad de testar la veracidad de una fuente, acompañar al estudiante en los procesos creativos, resolver dudas puntuales en horarios no lectivos y fomentar la participación post-clase... La IA es mucho más ineficiente cuando el principal objetivo de su aplicación es simplemente su aplicación. El estudiantado no percibe el valor añadido de la tecnología y no entiende por qué no se ha utilizado otro camino para alcanzar el mismo objetivo. Como presentábamos en el primer capítulo "preguntar es un arte", cualquier aplicación de la IA debe ser la respuesta a una pregunta relevante.

4. La primera isla de un archipiélago

Es más probable que la implantación de la IA funcione con pequeños experimentos puntuales en un modelo clásico que con una propuesta disruptiva que apueste por una aplicación masiva. La inteligencia artificial tiene un problema fundamental y es que no es predecible; su carácter probabilístico, no determinista, no permite fijar unas instrucciones claras que nos permitan anticipar el resultado. Diversos *prompts*, variaciones personales en el lenguaje natural, en la aportación de contexto o en la precisión de la salida darán lugar a resultados muy diversos y, algunos de ellos, poco predecibles. No es una sesión sobre Python, sobre SPSS o sobre bases de datos abiertas, donde seguimos una secuencia precisa para llegar a un punto determinado. Hay muchos caminos posibles y, además, más que el valor del resultado, en este caso el acento pedagógico debe ponerse en el proceso, en el itinerario. Hay una segunda dificultad que aparece de forma sistemática en las sesiones docentes con IA: el estudiantado siempre está tentado de usar la IA como un camino alternativo a la práctica planteada. Por ejemplo, cuando hemos creado un personaje Miró para testar el grado de conocimiento de los estudiantes sobre su figura, a partir de preguntas realizadas por una periodista ficticia, algunos de ellos han empleado la IA también para crear las

preguntas. La omnipresencia y versatilidad de la IA dificulta mucho su implantación con unos límites entre las tareas automáticas y las tareas manuales.

Por eso, la aplicación de la IA en el aula requiere un proceso de ensayo y error, que es viable con pequeñas prácticas alternativas en un modelo clásico más que con una irrupción revolucionaria. El riesgo de la acumulación de errores es muy alto. Por eso, nuestro planteamiento docente ha sido siempre la ocupación de islotes en un archipiélago, la creación de rutinas que acabamos controlando después de unos pocos ensayos y que nos permiten plantearnos nuevas incursiones en el mar abierto.

5. El estudiante cyborg

Es muy tentador considerar la IA como una alternativa al trabajo intelectual. Si el estudiante tiene que hacer un ensayo sobre la influencia del viaje a Túnez en la obra de Paul Klee o sobre la geopolítica del control de los puertos del Índico, puede optar por que la inteligencia artificial le haga el trabajo. De hecho, hay miles, decenas de miles, de trabajos de filología, historia, comunicación, geografía, derecho, economía, sociología, marketing o arte que han sido redactados en su totalidad por un ingenio como ChatGPT. Es muy difícil hacer entender al estudiante que ha perdido la capacidad de desarrollar las habilidades necesarias para la redacción de ese escrito, como son la documentación, el contraste de la información, la selección y lectura de las obras de contexto, la síntesis o la capacidad de comunicación de los resultados. La única habilidad que habrá desarrollado es la interrelación con una inteligencia artificial.

Esta batalla ya la ha perdido la docencia en el acceso a la red. Como no hemos orientado al estudiantado en un uso eficiente y responsable de la red, la búsqueda de información ha dejado de ser una habilidad compleja, basada en el contraste de las fuentes, en la selección y en la jerarquía de la información, y el estudiantado tiende a ofrecer un collage de información capturada en la red, donde un post de un blog personal o unos datos en una revista local tienen el mismo valor jerárquico que una fuente oficial.

La inteligencia artificial puede ser una herramienta controlada que potencie las habilidades del estudiantado. Un uso eficiente y crítico de la IA permite acelerar la búsqueda de información, evaluar las alternativas, mejorar el contexto, aclarar conceptos imprecisos, comparar los documentos, acompañar en la síntesis, establecer relaciones no evidentes, reforzar puntos de vista, estimular el sentido crítico o mejorar la calidad gráfica de los resultados. Si la IA acompaña un proceso complejo de construcción de conocimiento, tiene un enorme potencial para adaptar, personalizar y ampliar el proceso de aprendizaje. Para determinadas tareas, puede ser un atajo o un acelerador que reduzca el tiempo dedicado a algunas rutinas y que permita dedicar el esfuerzo a otras partes de la cadena de conocimiento. En cierta medida, el estudiante se convierte en un cyborg, con habilidades reforzadas o ampliadas, pero con un control "humano" de todo el proceso. Este es probablemente el gran reto de la docencia 3.0 en el futuro: pasar de la IA sustituta a la IA complementaria, del estudiante ausente al estudiante ampliado.

6. La IA responsable

Entrenar el modelo GPT-3 de OpenAI, que tiene 175.000 millones de parámetros, requirió unos 1,287 MWh de energía, según algunas estimaciones. Esto equivale a las emisiones de CO₂ de unos 626.000 kilómetros recorridos por un coche medio. La fabricación de semiconductores requiere grandes cantidades de agua ultrapura. Por ejemplo, se calcula que se necesitan aproximadamente 7.570 litros de agua para producir un solo chip de 30 cm. Se estima que un centro de datos típico puede utilizar hasta 1,8 litros de agua por cada kWh consumido para enfriar los servidores. Aunque las empresas no aportan información sobre los impactos ambientales de sus procesos, Google ha informado que su centro de datos de The Dalles, Oregón, utilizó aproximadamente 1.200 millones de litros de agua en 2021 para los procesos de refrigeración. La minería de materiales necesarios para la fabricación de semiconductores, como el litio, el cobalto y otros metales raros, también tiene un impacto ambiental significativo, incluyendo la

destrucción de hábitats y la contaminación del agua. Es cierto que este impacto debe dimensionarse. Se estima que hacer responder a cinco preguntas, ChatGPT consume medio litro de agua. Debemos tener en cuenta que un bistec de ternera tiene una huella hídrica de 7.000 litros de agua, por lo que un bistec equivale a 70.000 preguntas a ChatGPT. Esto no quiere decir que esta tecnología sea inocua, sino que los problemas ambientales son múltiples y necesitan respuestas también diversas. Un uso responsable de la IA implica limitar las preguntas y la interacción a aquellas cuestiones que son necesarias y evitar la repetición de procesos que se pueden simplificar con interacciones más concretas. La pregunta que podríamos plantearnos podría ser: "¿Realmente necesito hacer esta pregunta o puedo encontrar una forma alternativa igualmente válida para resolver el problema?". El consumo responsable también implica la utilización de modelos que tengan una mayor garantía; por ejemplo, Claude de Anthropic se basa en un compromiso de responsabilidad mucho más elevado que otros modelos con una clara orientación comercial, como ChatGPT, Gemini o Copilot. Llama también es una alternativa interesante porque utiliza código abierto, pero su uso está aún restringido a Europa y solo se puede acceder a ella con un VPN.

7. Formación dual

Hay una forma manual de calcular la desviación estándar y hay varios programas que permiten calcularla de forma automática, dedicando el esfuerzo a interpretar el resultado. ¿Qué tiene más sentido? ¿El modelo manual o el modelo automático? El sistema automático permite obtener una desviación en unos pocos segundos, lo que permite calcular muchas, con bases de datos muy grandes y comparar los resultados para evaluar aquellas variables que tienen un comportamiento más disperso. Pero aprender de forma manual el proceso de cálculo ayuda a entender qué es lo que hace la máquina. Y tiene una segunda función: permite calcular una desviación a pesar de que el

programa no funcione o que se haya ido la luz. Esta habilidad dual para generar un resultado de forma automática y de forma manual es una exigencia en la aplicación de nuevos sistemas complejos.

El creciente desarrollo de herramientas cada vez más precisas nos dará muchas más capacidades en los campos de la búsqueda, la selección de la información, la comparación, el análisis, la creación de argumentos, la teorización o la síntesis. Pero la formación en habilidades sin recursos digitales permite, en primer lugar, entender los procesos que operan en el interior de la caja negra de la IA y, sobre todo, dota al estudiante de una enorme autonomía. Uno de los riesgos más importantes del abuso de la inteligencia artificial (más que la privacidad, el sesgo o el fraude) es la dependencia. Un incremento del precio, una nueva política de acceso, un colapso del servidor o un ataque cibernético pueden generar una interrupción de acceso a la IA. Cuanto mayor sea la formación fuera de la IA, menor será el impacto de este escenario.

8. En los entresijos de la máquina

Formar a los estudiantes en habilidades relacionadas con la IA significa introducirlos también en la comprensión y manipulación de la máquina. Por ejemplo, es muy recomendable que los estudiantes desarrollen capacidades elementales en programación, como el lenguaje Python. Esto multiplica las posibilidades de conectar entornos y procesos, de vincular la potencia de la IA con herramientas externas y, además, mejora la comprensión sobre las rutinas que generan los modelos de lenguaje. En la versión 4 de ChatGPT, por ejemplo, el sistema muestra con mucha frecuencia los códigos en Python de las decisiones que está tomando, y esto ayuda a interpretar las operaciones y a orientar el modelo en una determinada dirección. Tener conocimientos en programación orientada a objetos y en el diseño de algoritmos y estructuras de datos debería ser una habilidad transversal. Se necesita una formación más sólida para comprender el aprendizaje profundo

(deep learning), pero el estudiante debería tener una noción intuitiva de lo que es una red neuronal artificial y de sus tipologías. La presentación de problemas clásicos de la IA (como la habitación de Mary) o los nuevos experimentos, como el [Golden Gate de Claude](#) pueden proporcionar una comprensión más profunda de los mecanismos opacos que operan en una IA.

9. La serendipia programada

En un mundo cada vez más algorítmico, donde la IA nos ofrece respuestas precisas y personalizadas, corremos el riesgo de perder el azar creativo, esos encuentros fortuitos con ideas inesperadas que a menudo desencadenan los grandes saltos en el conocimiento. El aula 3.0 debe preservar y fomentar estos espacios de serendipia, pero de una manera consciente y estructurada. Podemos programar la IA para que nos sorprenda, para que introduzca elementos aleatorios en sus respuestas, para que nos obligue a salir de nuestra zona de confort intelectual.

Por ejemplo, podríamos diseñar un “bot del caos” que, en medio de una clase sobre economía circular, introduzca conceptos de la mitología nórdica, obligando a los estudiantes a establecer conexiones improbables. O un “generador de analogías forzadas” que relacione, pongamos por caso, la teoría de juegos con el arte del bonsái. Estos ejercicios no solo estimulan la creatividad, sino que también entrenan la mente para ver patrones y relaciones donde aparentemente no las hay, una habilidad cada vez más valiosa en un mundo complejo e interconectado. La serendipia programada no es una contradicción en los términos, sino una forma de utilizar la IA para recuperar aquello que la misma IA amenaza con eliminar: la sorpresa, la divagación productiva, el pensamiento lateral. En el aula 3.0, lo imprevisto debe ser una parte integral del currículo, no una anomalía.

10. La ética como sistema operativo

Si la IA es el hardware del aula 3.0, la ética debe ser el sistema operativo. No se trata solo de enseñar ética como una asignatura más, sino de integrarla en cada interacción con la IA, en cada decisión sobre cómo utilizarla, en cada reflexión sobre sus implicaciones. El estudiante cyborg del que hablábamos antes no solo debe ser más eficiente, sino también más consciente y responsable.

Imaginemos un “asistente ético” que acompañe cada consulta a la IA, que plantee dilemas morales relacionados con cada tema tratado, que obligue a reflexionar sobre las consecuencias de cada decisión. Por ejemplo, en una clase de marketing, cada estrategia propuesta por la IA podría ir acompañada de un análisis de sus posibles impactos sociales y ambientales. O en una clase de derecho, podríamos programar la IA para que generara automáticamente escenarios distópicos basados en cada nueva ley estudiada, forzando así un debate sobre sus potenciales derivadas autoritarias. El objetivo es crear una “conciencia artificial”, no en el sentido de una IA autoconsciente, sino de un sistema que constantemente nos recuerde nuestra responsabilidad ética en el uso de la tecnología. En el aula 3.0, la ética no debe ser un añadido, sino el marco dentro del cual se desarrolla todo el aprendizaje. Solo así podremos formar profesionales que no solo sepan utilizar la IA, sino que también sepan cuestionarla, limitarla y, en última instancia, humanizarla.

José Antonio Donaire

Mònica Puntí

Konstantina Zerva

Raquel Camprubí

Núria Galí

De la tiza al chip

El uso de la inteligencia
artificial en las aulas

 Universitat
de Girona

ISBN: 84-8458-712-6



9 788484 587125