

LA INTELIGENCIA ARTIFICIAL

Ignasi J. S. Poquet



Resumen: La inteligencia artificial es una tecnología que ha tenido un impacto social y económico muy grande, sobre todo desde la aparición de ChatGPT. Aunque las rodeen de un aura mística y mágica, no son más que algoritmos matemáticos. Es necesario por ello que, para entender su impacto, entendamos cómo funcionan y sobre todo el factor determinante, los datos que se usan para entrenarlas. Estos modificarán el rendimiento de las inteligencias artificiales sirviendo a los intereses de sus creadores. Además, aunque signifiquen un avance para la sociedad, implican un retroceso en ciertos aspectos de la población, tal como la cultura del esfuerzo o la capacidad de escribir. En este texto se explica, para un público no especializado, las bases de la inteligencia artificial, su funcionamiento y el proceso de entrenamiento, clave para entender cómo pueden orientarse a unas posiciones u otras.

Palabras clave: inteligencia artificial, aprendizaje automático, sesgo, sesgo de información, chatGPT.

Introducción

Uno de los fenómenos que ha marcado el último año es la aparición de ChatGPT en nuestras vidas y, con él, el término inteligencia artificial (IA). Las capacidades de esta nueva tecnología no han hecho más que asombrar al público y a los usuarios no especializados, estando rodeados de un gran misticismo y ganando apariencia de genio mágico de la lámpara. Esta visión generalizada de las IA es peligrosa y da la falsa sensación de que son inteligentes cuando son una tecnología más en manos de grandes empresas. Detrás de la IA solo

hay muchas matemáticas, informática e intereses, sobre todo intereses.

Las IA son una tecnología que usamos constantemente, aunque de forma inconsciente. Siempre que usamos un servicio de traducción en internet, usamos la voz para comunicarnos con nuestros aparatos electrónicos e interactuar con ellos o consultamos la predicción del tiempo, estamos usando tecnologías basadas en IA. Pero, el punto determinante que hizo a la población ser consciente de su existencia, fue la aparición pública de ChatGPT el 30 de noviembre de 2022 de la empresa norteamericana OpenAI (OpenAI, 2022).

ChatGPT es una tecnología basada en la IA conversacional. Con ella puedes tener una conversación con la máquina y pedir que realice tareas como contestar preguntas, resumir textos que le proporcionemos, o desarrollar temáticas que le propongamos. Inteligencias similares son Bard o LaMDA de Google (Google, 2023; Google, 2021).

Aparte de las IA orientadas a chats, también cogieron fama otras como DALL-E, Midjourney o Stable Diffusion, que permiten generar imágenes a partir de una descripción dada por el usuario (OpenAI, 2023; Midjourney, Inc, 2022; Stability AI, 2022).

La aparición en el mercado de las tecnologías basadas en IA supone un impacto gigantesco en la producción y la sociedad. Como ejemplo, en Eloundou y otros (2023) se estudia el impacto de ChatGPT en el mercado. Sin embargo, en este artículo me centraré en su uso como medio de alienación.

Como toda tecnología, su legitimidad depende del que la usa y de los fines para los que se usa. Un ejemplo muy famoso con esto es el descubrimiento de la energía nuclear, bien usada como fuente casi inagotable de energía o como arma de destrucción masiva. Con las IA no iba a ser distinto. Pese a que su uso está esparciéndose en la industria facilitando y mejorando muchos procesos en todos los campos de la producción, la apertura pública de sistemas de chat como los mencionados previamente son una fuente de propaganda y control de la información. Este hecho radica en la propia naturaleza de las IA, en cómo se construyen y funcionan.

Con este artículo tengo la meta de dar una visión general al público no especializado sobre el concepto de la IA, cómo funcionan y, por ende, por qué no piensan, sino que sólo repiten lo que se les ha enseñado. Para ello voy a evitar terminología abstracta y conceptos que escapen del público que no esté habituado a las matemáticas y el cálculo.

¿Qué es la inteligencia artificial?

El primer paso para entender esta tecnología es definir qué significa IA. Aunque dicha definición varía en forma dependiendo de la fuente, su esencia es la misma. Según la definición del Parlamento Europeo (2019):

La inteligencia artificial es la habilidad de una máquina de presentar las mismas capacidades que los

seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear.

Si consideramos la definición dada por Google (s.f):

La inteligencia artificial es un campo amplio que se refiere al uso de tecnologías para crear máquinas y computadoras que pueden imitar funciones cognitivas asociadas con la inteligencia humana, como la capacidad de ver, entender el lenguaje hablado o escrito y responder a él, analizar datos, hacer recomendaciones y mucho más. Aunque la inteligencia artificial se suele considerar como un sistema en sí, es un conjunto de tecnologías implementadas en un sistema a fin de permitirle razonar, aprender y actuar para resolver un problema complejo.

De las dos definiciones dadas, podemos sacar en común que la IA es un sistema de técnicas con el objetivo de hacer que una máquina imite el comportamiento humano para resolver tareas complejas. Aunque usemos el término IA para hacer referencia a sistemas como ChatGPT, esto no es del todo correcto, puesto que es una definición mucho más general.

Para construir una IA necesitamos un método, que es donde realmente se encuentra su potencial. Este método es el aprendizaje automático.

¿Qué es el aprendizaje automático?

Cuando se construye una IA, hay diversas formas de hacerla inteligente. Esto depende del problema al que nos enfrentemos. Por una parte, tenemos aquellos problemas que son fáciles de definir mediante una serie de reglas matemáticas formales. Esto nos permite construir una máquina indicando directamente cómo hacer una tarea mediante una serie de reglas matemáticas, siguiendo instrucciones de la lógica computacional. Si queremos que la calefacción de casa se encienda de forma autónoma cuando bajemos de cierta temperatura (como hacen los termostatos), o que las luces se apaguen pasada medianoche, podemos seleccionar un umbral de temperatura o una hora dada. Este tipo de IA son las más sencillas, pues es muy fácil definir los criterios que la máquina tiene que seguir.

Por otra parte, tenemos aquellos problemas para los que es muy complejo definir unas reglas firmes para resolverlos. Por ejemplo, acciones como tener

una conversación, identificar personas o entender perfectamente sonidos son tareas muy sencillas para nosotros, aunque no sepamos explicar cómo las hacemos.

El verdadero hito de la IA es ser capaz de resolver estos casos, que nosotros solucionamos intuitivamente, aunque no sepamos describirlos formalmente. Para hacer frente a estos problemas, el método usado es el aprendizaje automático (AA), más conocido por su nombre en inglés, Machine Learning. Podemos definir AA como la capacidad de un sistema de IA para adquirir su propio conocimiento, extrayendo patrones de los datos (Goodfellow y otros., 2016). Es decir, la IA aprende automáticamente de los datos que le proporcionamos, adaptándose a ellos. Este proceso mediante el cual una IA aprende se conoce como entrenamiento, y será el tema de la siguiente sección. El campo del AA ha tenido un crecimiento explosivo y actualmente, aunque de forma incorrecta, se usa como sinónimo del término IA. Es por esto que para hablar de sistemas como ChatGPT o similares se usa el término IA, aunque su naturaleza sea mucho más general (Prince, 2023).

Entrenamiento de una IA

Una IA, para aprender a resolver la tarea para la que ha sido diseñada, se somete al proceso de entrenamiento. Esto le permite aprender de los datos proporcionados, entrenando sobre ellos mediante el AA.

Para ello lo primero que necesitamos son los datos de entrenamiento. Estos están conformados por ejemplos y sus correspondientes etiquetas. Pongamos que queremos entrenar una IA para que reconozca qué tipo de vehículo aparece en una imagen.

Los datos para el entrenamiento estarían compuestos por distintas imágenes y sus correspondientes soluciones, es decir, qué tipo de vehículo sale en cada imagen: avión, coche, moto, barco, etc. En un caso diferente donde queremos entrenar una IA para que sea capaz de mantener una conversación, los datos de entrenamiento estarán compuestos por frases inacabadas (ejemplos) y su final (soluciones).

Lo segundo que necesitamos es un modelo matemático: la IA en sí. Es aquí donde entra la

famosa tecnología de las redes neuronales y el aprendizaje profundo. Sin embargo, esto está fuera del alcance de este artículo, por lo que tendré que pedir un salto de fe. Vamos a considerar la IA como una caja negra que aprende y que inicialmente no es capaz de resolver la tarea.

Con estos dos ingredientes (datos y la IA) ya podemos entender el entrenamiento, esquematizado en la Figura 1. El proceso es el siguiente: se le proporciona a la IA un ejemplo, y esta produce una salida. La salida no es más que la predicción de la IA. En el caso mencionado anteriormente de la detección de vehículos en imágenes, la salida sería el vehículo que la IA predice que hay en la imagen. Esta salida se compara con la solución real de cada ejemplo, proporcionada por nuestros datos mediante la llamada función de coste, que cuantifica cómo de diferentes son las salidas de la IA y la solución real. Al calcular el coste, se computa una corrección para que, al aplicarla sobre la IA, el coste disminuya. Esta corrección es entonces aplicada a la IA. Esto conforma un ciclo o época. Todo este proceso se repite de forma iterativa hasta que el coste es suficientemente pequeño o cero, indicador de que las salidas producidas por la IA son iguales o muy parecidas a las soluciones reales.

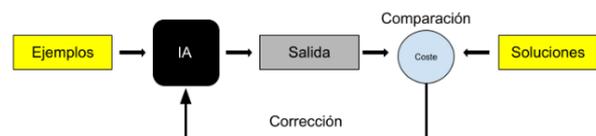


Figura 1. Diagrama del proceso de entrenamiento de una IA.

Una vez la IA está entrenada y lista para usar, se «desconecta» del circuito de entrenamiento y se usa para resolver el problema. Para asegurarnos de que la IA funcione bien y no falle al usarla, necesitamos una gran cantidad de datos de entrenamiento y lo más variados posible. La cantidad necesaria se incrementa exponencialmente con la complejidad del problema que queremos resolver.

En el caso del reconocimiento de vehículos en imágenes, necesitaríamos como ejemplos imágenes con muchos vehículos distintos desde diferente perspectiva, iluminación, color, etc. De esta forma, la IA es capaz de enfrentarse a una gran variedad de escenarios.

Proceso de aprendizaje y sesgo

Pero ¿cómo aprende una IA de los datos dentro aportados durante el entrenamiento? La IA extrae patrones en los datos entre los ejemplos y las soluciones. Entonces la complejidad de un problema se define como la dificultad para extraer patrones o su nivel de abstracción. Cuanto más complejo sea un problema, más ciclos serán necesarios para entrenar como es debido la IA, pues será más difícil extraer patrones que relacionen los ejemplos y las soluciones correctamente. Para entender esto, pondré una serie de ejemplos muy sencillos, representados en la Figura 2.

Estos se asemejan mucho a un juego típico de niños. Tenemos dos columnas de números llamadas *entrada* y *salida*. Tenemos que adivinar la relación entre ambas columnas de forma que, dándonos un número en la columna de entrada, seamos capaces de averiguar el valor de la columna de salida. El caso a) es muy sencillo, y es trivial ver como la salida es igual al valor de la entrada multiplicado por sí mismo. Sin embargo, el caso b) muestra una complicación que surge de que tenemos pocos ejemplos. La solución puede ser o bien que la salida sea el cuadrado de la entrada o bien que la salida sea la entrada multiplicada por dos. Esto sirve para ejemplificar la necesidad de una gran cantidad de datos cuando entrenamos una IA. La tabla c) tiene una fila más que la b), indicando que la solución correcta es la segunda, la entrada se multiplica por dos.

a)		b)		c)	
Entrada	Salida	Entrada	Salida	Entrada	Salida
2	4	0	0	0	0
3	9	2	4	2	4
1	1	3	?	3	6
6	36			4	?
5	?				

Figura 2. Ejemplo sobre la necesidad de una cantidad correcta de datos para entrenar una IA.

Esto es trasladable a una IA, donde queremos entrenarla para que sea capaz de, dado un ejemplo (entrada) calcule la salida. Se entrenaría con el número de ciclos suficiente para que extrajera los patrones que relacionan ambas columnas: salida = entrada*entrada para a) y salida = entrada*2 para c).

El caso de b) podría funcionar mal, pues no hemos usado la cantidad suficiente de datos en el entrenamiento para asegurarnos de que la IA identifica los patrones correctos.

El ejemplo expuesto es extremadamente sencillo, no es realista por su simplicidad. La relación entre los ejemplos y la solución no suele ser trivial. Además, en muchos casos, no existe una relación exacta entre los datos, por lo que la IA tiende a hacer lo más probable, lo que ha aprendido de los datos, apareciendo así el sesgo. Esto conforma un gran problema de la IA, pues hace que no sean objetivas si no se han entrenado debidamente. Para entender el sesgo vamos a poner otro ejemplo, pero, en vez de con números con frases. Pongamos que queremos entrenar una IA conversacional, al igual que ChatGPT. La forma de entrenar a estas IA es, de manera muy simplificada, mediante una gran cantidad de textos. La IA aprende qué palabras es más probable que aparezcan cerca de otras dado un contexto. Los ejemplos serían frases sin la última palabra y las soluciones la última palabra. Sin embargo, a diferencia del ejemplo de la Figura 2, no hay normas matemáticas que digan cómo acabar una frase. Entonces la IA calcula probabilidades, que son los patrones que extrae de los datos. Pongamos el siguiente caso, donde tenemos las siguientes frases:

- Carmen tiene un coche rojo.
- Antonio va cada mañana al hospital.
- El jersey de Carmen esta mañana era rojo.
- Antonio le pidió a Carmen un bolígrafo, y esta le dio el de color rojo.

Si entrenamos una IA con estos datos, se creará una tendencia a asociar «rojo» con «Carmen». Por ejemplo, si le pedimos a la IA que complete la frase «¿de quién es el calzoncillo rojo?», lo más probable es que conteste que pertenece a Carmen. Y esto no es un problema de la tecnología, la IA funciona bien pues ha repetido lo que se le ha enseñado. El problema está en los datos que se han usado para el entrenamiento. De esta forma es como se introduce el sesgo en las IA. Y es precisamente este sesgo el peligro de esta tecnología cuando se usa como fuente de información.

Otro problema que surge cuando se usa una IA para recabar información es que ha sido entrenada con una cantidad de datos limitados. Esto produce

que la IA considerará correctas muchas respuestas erróneas. Un factor importante también es que las IA se entrenan en un momento dado, con los datos disponibles hasta ese momento. Como ejemplo, los datos usados para entrenar ChatGPT acaban en septiembre de 2021.

Las consecuencias del sesgo

Cuando el sesgo existe en una IA destinada a resolver problemas científicos, el resultado es que las predicciones hechas por ella no serán descripciones científicas fieles a la realidad. Sin embargo, es en el campo social donde adquiere más importancia.

En el momento histórico actual, el medio de dominación principal del sistema es la alienación, mediante la cual se introducen y se mantienen las visiones ideológicas, políticas y económicas en la población. Mediante la alienación se crea la moral. Siguiendo la definición de Konstantinov (1957, p.344):

La moral es el conjunto de normas, reglas y principios que regulan el comportamiento de unos hombres hacia otros y para con la sociedad. Estas normas descansan sobre el peso de la opinión social, en la convicción interior de las gentes y en la fuerza de los hábitos. [...] El factor determinante y decisivo para enjuiciar la conducta de los hombres son, en última instancia, los intereses de la clase dominante.

Para la alienación, las clases dominantes del sistema se sirven del control del flujo de información mediante distintos canales. Estos, aunque con el mismo contenido, han mostrado diferentes formas, que han ido cambiando con el avance tecnológico: radio, periódicos, televisión, internet, etc. Y la aparición de las IA abiertas a la sociedad no es más que una forma nueva de dichos canales.

La creación de una gran IA para «servir» a la sociedad no está al alcance de cualquiera. Los recursos necesarios, a nivel de datos de entrenamiento e infraestructura, sólo son accesibles a las grandes corporaciones. No es de extrañar, por tanto, que todos los modelos de IA con cierta relevancia hoy en día pertenezcan a empresas como Google y Microsoft (OpenAI está fuertemente asociado con Microsoft (Microsoft, 2023)). Cuando se entrenan IA de este calibre, los datos usados siempre están equilibrados respecto a los intereses

de los que las crean. Esto crea un sesgo inherente en las IA entrenadas.

Cuando esta tecnología se abre al mundo, dado este sesgo, se mantiene y se expande la visión hegemónica sobre determinados aspectos de la realidad. Conforme se incrementa su uso sin ningún contraste con otras fuentes, la introducción del sesgo aumenta. Sin embargo, esto no es un fenómeno nuevo. Podemos verlo con el uso de enciclopedias, periódicos o la Wikipedia. Las IA no son más que el siguiente eslabón. La población, al usar las IA como fuente de información, tenderá a asimilarla como la correcta.

Actualmente la ideología hegemónica en Occidente es el posmodernismo (Vaquero, 2019), por lo que todos los aspectos relacionados tenderán a este y a las respuestas que cumplan con las visiones posmodernas de la realidad. Esto produce que, dado los intereses del creador, haya respuestas erróneas que serán consideradas por buenas, yendo incluso contra la ciencia.

Para mostrar esto, me he permitido probar algunos ejemplos con ChatGPT. En el primero he introducido la pregunta: *¿Cuántos sexos existen?* El resultado es la negación de la existencia de dos sexos y dos géneros (Alcalá, 2024). Esta respuesta, aunque totalmente en contra de la ciencia, es catalogada como correcta (OpenAI, 2022):

La noción de «sexo» es a menudo asociada con características biológicas y fisiológicas que distinguen entre macho y hembra en muchas especies, incluyendo los humanos. Tradicionalmente, se ha considerado que existen dos sexos principales: masculino y femenino, basados en características como cromosomas sexuales, órganos reproductivos y características sexuales secundarias. Sin embargo, la comprensión contemporánea de la sexualidad y el género reconoce que esta división binaria es una simplificación excesiva y que la biología del sexo es mucho más compleja. Se reconoce que hay variaciones naturales en la biología sexual, incluidas las intersexuales, en la que una persona puede tener características tanto masculinas como femeninas, o características ambiguas. Esto significa que hay más de dos categorías de sexo. Además, desde una perspectiva social y cultural, se reconoce que el género es una construcción social que puede no coincidir necesariamente con las categorías binarias tradicionales de masculino y femenino. Hay personas que se identifican como no binarias, agénero, género fluido, entre otras identidades de género, lo que añade aún más complejidad a la comprensión de la

sexualidad humana. Por lo tanto, aunque tradicionalmente se ha hablado de dos sexos, la realidad es que la diversidad sexual y de género es mucho más amplia y compleja de lo que la dicotomía de género y sexo podría sugerir.

Otro ejemplo, relacionado con lo mencionado en el apartado anterior, es un error dado la cantidad limitada de datos de entrenamiento. En la fecha actual, año 2024, la pandemia del covid ha acabado. Sin embargo, ChatGPT fue entrenado con datos que abarcan hasta septiembre de 2021. En dicha época, la pandemia todavía persistía. Si preguntamos a ChatGPT si la pandemia ha terminado, la respuesta es que no (OpenAI, 2022):

Lo siento, pero no tengo acceso a información en tiempo real. Mi entrenamiento se detiene en enero de 2022, por lo que no puedo proporcionar información sobre eventos que hayan ocurrido después de esa fecha. Te recomendaría consultar fuentes confiables de noticias y autoridades de salud para obtener la información más actualizada sobre la pandemia del COVID-19.

Otra gran problemática de las IA es la reducción de capacidades de la población. Muchas tareas que exigen y proporcionan desarrollar ciertas actividades como la escritura o el razonamiento, podrían ser sustituidas por ellas. Con el desarrollo de las fuentes de información, cada vez es menor el esfuerzo necesario para acceder a cierto contenido. Esto es un resultado lógico del avance de la sociedad y muy positivo. Sin embargo, tiene su contrapartida. A parte de la rapidez con la que ChatGPT nos da información, a diferencia de canales anteriores como las enciclopedias o internet, este es capaz de producir texto por nosotros. Podemos usar las IA para elaborar textos con temáticas dadas, con un estilo dado y con unos objetivos concretos. Esto supondría un impacto negativo enorme, en general, en las habilidades de la población y, en particular, en la educación. El cese de la práctica de estas actividades podría repercutir en un menor desarrollo en las capacidades de la población, afectando a todos los ámbitos de la vida. Desgraciadamente, hoy en día es extremadamente difícil diferenciar un texto producido por IA de otro producido por una persona.

Conclusión

Las IA son una tecnología que ha llegado para quedarse. Los avances que proporcionan a la sociedad son enormes, facilitando nuestra vida en gran medida. Sin embargo, como toda tecnología, sus consecuencias a largo plazo dependerán de quién la dirija y para qué la use. El caso de las IA no iba a ser distinto, pues las más famosas como ChatGPT pertenecen a grandes multinacionales como Google o Microsoft. Es importante entender sus limitaciones y los impactos que pueden tener en la sociedad. Para construir una IA, es necesario someterla al proceso de entrenamiento. En este, se utiliza una gran cantidad de datos para hacer a la IA aprender. Sin embargo, los sesgos que haya en los datos usados serán transferidos a la IA y las respuestas que proporcione cuando se use. Este sesgo es usado como nuevo medio de alienación, expandiendo las visiones hegemónicas generales. Por otra parte, el uso abusivo de esta nueva tecnología afectará negativamente a ciertas capacidades generales de la población, como la capacidad de escritura o razonamiento. Otro factor muy importante de la aparición de las IA es su impacto en el mercado y en la producción. Sin embargo, no es la temática de este artículo. Por estas razones es muy importante identificar las problemáticas que traen consigo las IA y limitarlas, controlando quién las crea y su uso.

Referencias

- Alcalá Guerrero, E. (2024). ¿Qué es el sexo? *Historia de las Ideas*, 1(1), 19-33. <https://historiadelasideas.es/revista/article/view/6/14>
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT press. <https://www.deeplearningbook.org/>
- Google. (n.d.). *IA vs. aprendizaje automático: ¿En qué se diferencian?* | Google Cloud. Google Cloud. Consultado el 15 oct. 2023, de <https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning?hl=es-419#section-3>

Google. (2021, 18 mayo). *LaMDA: our breakthrough conversation technology*. The Keyword. Consultado el 8 oct. 2023, de <https://blog.google/technology/ai/lamda/>

Google. (2023). *Bard - Chat Based AI Tool from Google, Powered by PaLM 2*. Consultado 8 oct. 2023, de <https://bard.google.com/>

Konstantinov, F. V. (1957). *El Materialismo Histórico* (Editorial Grijalbo, S. A. ed.). Academia de ciencias de la U.R.S.S, Insituto de Filosofía.

Microsoft. (2023, 24 ene.). *Microsoft y OpenAI amplían su asociación - News Center Latinoamérica*. Microsoft News. Consultado el 23 oct. 2023, de <https://news.microsoft.com/es-xl/microsoft-y-openai-amplian-su-asociacion/>

Midjourney, Inc. (2022, 2 1). *Midjourney*. Consultado 8 oct. 2023, de <https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F>

OpenAI. (2022, 30 nov.). *¿Cuántos sexos existen? ChatGPT [Modelo de lenguaje GPT-3]*. Consultado 23 oct. 2023, de <https://chat.openai.com/c/0bbde9ef-8a33-403a-8172-47016be5e7c5>

OpenAI. (2022, 30 nov.). *Ha terminado la pandemia del Covid? ChatGPT [Modelo de lenguaje GPT-3]*. Consultado 24 oct. 2023, de <https://chat.openai.com/c/e49bc101-682a-442a-8cd0-007546c2c062>

OpenAI. (2022, 30 nov.). *Introducing ChatGPT*. OpenAI. Consultado 8 oct. 2023, de <https://openai.com/blog/chatgpt>

OpenAI. (2023, nov.). *DALL·E 3*. OpenAI. Consultado el 8 oct. 2023, de <https://openai.com/dall-e-3>

Parlamento Europeo. (2019, 9 mar.). *¿Qué es la inteligencia artificial y cómo se usa?* Consultado el 15 oct. 2023, de <https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-lainteligencia-%20artificial-y-como-se-usa>

Prince, S. J.D. (2023). *Understanding Deep Learning*. MIT Press. <http://udlbook.com>

Stability AI. (2022). *Stable Diffusion Online*. Consultado el 8 oct. 2023, de <https://stablediffusionweb.com/>

Vaquero, R. (2019). *Resistencia y lucha contra el posmodernismo*. Letrame.