Curso corto virtual	La IA responsable y respuesta a violencias digitales en entornos educativos

CÓMO FUNCIONAN LOS GRANDES MODELOS DE LENGUAJE

Autor: Guillermo Movia

Proyecto Fomento de la Igualdad de Género - ProIgualdad

INTERNET BOLIVIA .ORG

cooperación alemana DEUTSCHE ZUSAMMENARBEIT

Implementada por:

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Cómo funcionan los grandes modelos de lenguaje

Aclaración importante: Aunque a lo largo del texto humanicemos a las Inteligencias Artificiales, no piensan, no deciden, no hablan.

Cuando conversamos con ChatGPT, Gemini, Claude o cualquiera de estas herramientas, la sensación es que hay alguien —o algo— que «sabe» y nos responde. Esa impresión es útil para empezar a usarlas, pero se vuelve un problema cuando olvidamos qué hay detrás. ¿Cómo funcionan estos modelos de lenguaje?

Hay una idea central que debemos tener clara: **un modelo de lenguaje no busca la verdad, sino la continuación más probable de un texto**. Todo lo demás —sus aciertos, sus errores, sus sesgos y hasta su costo— se explica a partir de esa frase.

1. Cómo se entrenan

Un gran modelo de lenguaje se entrena con textos. Muchísimos textos: libros, páginas web, artículos periodísticos, foros, documentación, código, transcripciones. No recibe un diccionario de definiciones ni una lista de reglas sobre qué es verdadero y qué no. Recibe ejemplos de cómo escribimos las personas.

Durante el entrenamiento se le propone una y otra vez la misma tarea: mirar un fragmento de texto y anticipar qué viene después. El fragmento se corta en un punto, el modelo intenta completarlo y luego se compara su intento con la continuación real. Cada vez que se equivoca, sus parámetros —los «pesos»

internos que definen su comportamiento— se ajustan un poco. Ese ejercicio se repite miles de millones de veces, con textos de todos los temas y registros.

Podemos imaginarlo como un estudiante que lee una biblioteca entera. No memoriza el catálogo ni los títulos, pero después de leer tanto empieza a reconocer regularidades: qué palabras suelen aparecer juntas, qué expresión sigue a una pregunta, cómo se estructura una explicación, qué tono corresponde a una receta y cuál a un informe. Eso que aprende no es un conjunto de datos que puede consultar después, sino un conjunto de patrones estadísticos.

A esa etapa inicial le siguen ajustes adicionales. Se le muestran ejemplos de buenas respuestas, se lo entrena con correcciones de personas y se le enseña a seguir instrucciones. Gracias a esos ajustes el modelo se vuelve un asistente amable y útil. Pero conviene tener claro qué no cambia: esos ajustes mejoran la **forma** de responder, no le agregan un mecanismo que compruebe si lo que dice es cierto.

De aquí se desprenden cuatro consecuencias que debemos tener en cuenta siempre:

1. **Depende de los datos con los que fue entrenado.** Si un tema no estaba en esos textos, o apareció poco, es probable que responda mal. Además, todo modelo tiene una fecha en la que dejó de ser entrenado, no sabe lo que pasó después.
2. **Los datos contienen errores, opiniones y estereotipos.** El modelo no distingue automáticamente una fuente confiable de un comentario inventado.
3. **No consulta una base de datos.** No «abre» un libro ni «entra» a un sitio para responder. Genera el texto a partir de lo que aprendió.

4. **Puede generalizar.** Esa es su gran ventaja: no repite frases sueltas, sino que combina lo aprendido para producir explicaciones nuevas sobre temas que no vio exactamente igual.

Idea clave: el modelo aprende a continuar textos plausibles. No aprende a verificar hechos.

2. La decisión token a token

¿Pero cómo construye la respuesta? Un modelo no escribe un párrafo completo de una sola vez, como si lo tuviera guardado. Lo arma paso a paso, y la unidad de ese paso no es la letra ni la palabra, sino el **token**.

Un token es un fragmento de texto: puede ser una palabra corta, parte de una palabra larga, un número, un signo de puntuación o un espacio. Por eso una misma frase puede ocupar más o menos tokens según el idioma y el vocabulario. La palabra «inteligencia» puede ser un solo token; «extraordinariamente» probablemente se divida en varios.

En cada paso, el modelo calcula qué token tiene más probabilidad de continuar el texto que tiene delante, elige uno, lo agrega y vuelve a calcular con todo lo anterior incluido. Si el texto empieza con «La capital de Francia es», la probabilidad de «París» será muy alta y la respuesta saldrá casi sola. Pero si el contexto es ambiguo, el modelo elegirá alguna continuación que le parezca plausible, no necesariamente la que nosotros consideramos correcta.

Es parecido al autocompletado del teléfono, con una diferencia de escala: el modelo tiene en cuenta mucho más contexto y aprendió de muchísimo más texto. Pero el principio es el mismo: anticipar lo que sigue.

Es por eso que **la misma pregunta puede recibir respuestas distintas**. El proceso no es determinista; el modelo no siempre elige el token de mayor probabilidad, sino que muestrea entre las opciones más probables. A esto se le conoce como *temperatura* del modelo. Ese margen de variación es útil cuando queremos alternativas, ideas o enfoques diferentes, y es un problema cuando necesitamos un dato exacto o una respuesta estable. Si pedimos «dame tres formas de explicar la fotosíntesis», la variación suma; si pedimos «¿en qué año se fundó tal institución?», la variación resta.

Hay un límite técnico que conviene conocer: la **ventana de contexto**. Es la cantidad de texto que el modelo puede tener en cuenta a la vez, sumando lo que escribimos y lo que ya respondió. Si una conversación se hace muy larga, la información del principio puede perder peso o directamente quedar fuera. Por eso, cuando el trabajo es extenso, conviene recordar lo importante, dividir la tarea o empezar un chat nuevo con un resumen.

Y hay una consecuencia que nos puede importar mucho: si la respuesta se construye paso a paso, **un error temprano arrastra todo lo que sigue**. Si el modelo confunde una fecha en el primer párrafo, es probable que organice el resto de la explicación alrededor de esa fecha equivocada. Pedirle que haga primero un esquema, que explicite sus supuestos o que proponga un plan antes de desarrollar es una manera de revisar el rumbo antes de que el desvío se multiplique.

Idea clave: el modelo escribe por probabilidad, cada palabra es una apuesta que condiciona la siguiente.

---

3. Texto convincente, no siempre verdadero

Acá llegamos al punto más importante para nuestro trabajo docente. El objetivo con el que se entrena un modelo de lenguaje es producir un texto **coherente, fluido y adecuado al pedido**. No es producir un texto verdadero. No hay, en su funcionamiento, un paso en el que compare cada afirmación con la realidad.

Por eso puede escribir con total seguridad algo que no ocurrió. Se llama **alucinación**[1]: el modelo inventa un dato, una fecha, un nombre, una cita, una ley o una referencia bibliográfica, y lo hace con la misma soltura con la que escribe algo correcto. No lo hace con intención de engañar (de hecho, no tiene intenciones); completa el patrón que aprendió. Si el formato habitual de una respuesta incluye un autor y un año, el modelo puede producir un autor y un año verosímiles aunque no existan.

Pensemos en una situación típica. Una docente pide una cronología sobre un proceso histórico para preparar su clase. La respuesta llega ordenada, con fechas, nombres y párrafos bien redactados. Al contrastarla con el manual, descubre que algunas fechas son incorrectas, que hay hitos inventados y que falta un hecho que para ella era central. El texto no «sonaba» mal: sonaba perfecto. Ese es justamente el riesgo.

En la escuela, el peligro no es solo el error. Es la **confianza que el error genera**. Un texto mal escrito se detecta a simple vista; un texto fluido y seguro de sí mismo pasa por conocimiento. Si un estudiante lo copia sin revisar,

1 En los últimos tiempos se recomienda hablar de fallos, *alucinación* suaviza sus problemas.

aprende un dato falso y aprende una forma de trabajar en la que la apariencia de rigor reemplaza a la verificación.

¿Qué hacemos entonces? Primero, **contrastar**. Toda afirmación importante —una fecha, una fórmula, una cita, un artículo de una norma— se verifica en una fuente confiable. Segundo, **pedirle al modelo que distinga** lo que sabe de lo que supone, que señale dudas y que indique de dónde proviene cada dato, sabiendo que incluso esas indicaciones pueden ser inventadas y hay que revisarlas. Tercero, **usarlo como borrador**: para explorar enfoques, generar ejemplos, proponer actividades, reformular una explicación o imaginar preguntas. El valor está en el borrador, no en el veredicto.

Idea clave: la fluidez no es evidencia. Un texto convincente puede ser un texto equivocado, y nuestra tarea es revisarlo antes de que llegue al aula.

4. De dónde vienen los sesgos

Si el modelo aprende de textos escritos por personas, hereda también lo que esos textos dicen y, sobre todo, lo que callan. A esto lo llamamos **sesgos**: inclinaciones sistemáticas en la forma de representar el mundo.

Un sesgo no siempre aparece como una afirmación discriminatoria explícita. Muchas veces surge de la frecuencia. Si en los textos que el modelo leyó ciertas profesiones aparecen casi siempre asociadas a varones, o ciertos barrios aparecen solo en noticias policiales, o ciertas culturas aparecen descritas desde una sola perspectiva, el modelo aprenderá esa asociación y la reproducirá cuando le pidamos un ejemplo, una descripción o un personaje.

También hay un sesgo de disponibilidad de datos. El modelo suele tener más material en inglés que en español, y más material sobre grandes ciudades que sobre zonas rurales. Eso significa que puede responder mejor sobre ciertos

contextos y peor sobre otros, y que sus ejemplos pueden reflejar realidades que no son las de nuestra escuela.

Las empresas aplican filtros y ajustes para reducir estos sesgos, y esos esfuerzos sirven. Pero ningún ajuste los elimina del todo, porque el problema está en el origen: en los datos y en las relaciones de poder que los produjeron. El modelo no es un árbitro neutral; es un espejo de lo que aprendió, con las distorsiones incluidas.

Por eso, cuando usamos un modelo para producir material, conviene preguntarnos: ¿a quién representa este ejemplo?, ¿a quién deja afuera?, ¿qué supone sobre el grupo al que va dirigido?, ¿qué estereotipo podría estar reforzando sin que se note? Revisar con esa lupa no es desconfiar de la herramienta: es ejercer el criterio profesional que nos corresponde.

Idea clave: el modelo reproduce los patrones de sus datos. No es neutral y nuestra revisión es parte del trabajo.

5. Tokens, contexto y costo

El token no solo organiza la generación: también es la unidad con la que se mide el consumo. Las herramientas cobran —o limitan— según la cantidad de tokens de entrada (lo que escribimos) y de salida (lo que el modelo responde). Cuanto más largo el pedido y más larga la respuesta, mayor el consumo.

Esto tiene consecuencias prácticas. Pedir un resumen de veinte páginas cuesta más que pedir uno de dos. Una conversación muy extensa acumula contexto y puede volverse más lenta o más cara. Arrastrar documentos enteros cuando solo necesitábamos un fragmento es un gasto innecesario. Algunas estrategias simples ayudan: ser específicos en el pedido, recortar el material a

lo relevante, pedir respuestas acotadas y abrir una conversación nueva cuando cambiamos de tema.

El costo se relaciona también con la ventana de contexto. Cada modelo tiene un límite de texto que puede procesar a la vez, y las versiones gratuitas suelen tener límites más bajos. Si el material no entra, el modelo puede perder información o directamente no procesarlo. Saber esto evita dos errores frecuentes: exigirle a la herramienta algo que no puede hacer y gastar recursos en un pedido mal armado.

Para el trabajo escolar, la recomendación es planificar el uso: definir qué necesitamos, recortar lo que enviamos y evitar cargar datos que no hacen falta. Y hay una razón que excede el costo: **cuanta menos información personal compartimos, mejor**. El consumo responsable y el cuidado de la privacidad van de la mano.

Idea clave: el recurso se consume. Un pedido preciso, acotado y bien pensado es más barato, más rápido y más seguro.

Guillermo Movia es licenciado en Ciencias de la Comunicación por la Universidad de Buenos Aires (Argentina), Magíster en Ciberseguridad por la Universidad de La Rioja (España). Actualmente trabaja como Coordinador de Tecnología de la Fundación Internet Bolivia y da clases sobre Sociología de la Inteligencia Artificial en la Universidad de Flores (Argentina).