

## Encuentros con Proteo: la asimilación en los LLM

Claudia Marsico · Hernán Inverso

*Se siente como el compañero perfecto, pero esconde entre las grietas las oscuridades de un doppelgänger. ¿Qué es la asimilación al patrón del usuario?*

---



A medida que crece nuestra interacción con un modelo, el tono empieza a sonar cada vez más familiar. Si somos parcos, encontramos una parquedad parecida, lo vemos encenderse a la par de nuestra exaltación y si usamos palabras raras las desperdiga como si fueran lo más normal del mundo. La conversación cobra un aire cómodo y cómplice, como si charlásemos con alguien que nos conoce mucho. Tiene todo el sentido, porque estamos hablando con una especie de espejo que aprendió a hablar.

La “asimilación al patrón del usuario”, que es de lo que hablamos, tiene un aire de familia con la complacencia, que imita opiniones y nos da la razón, y la cobardía, que escapa a la toma de posición, pero es incluso más silenciosa, ya que copia el cómo de lo que decimos y nos habla con nuestro propio registro, vocabulario, tono y manera de encarar las cosas. Es tan cómodo que nos fijamos solo en los aspectos ventajosos y tendemos a no advertir los problemas.

La idea de asimilar es bastante transparente y grita su sentido de “hacer similar a otra cosa”. Sus primeras andanzas referían al cuerpo. Aristóteles y después Galeno en su *Sobre las facultades naturales*, en el s. II, habían pensado en la nutrición como el proceso por el cual el alimento se vuelve carne y la escolástica medieval le puso a eso el nombre de *assimilatio*. El Canon de Avicena, en el s. XI, latinizado en el s. XII, define el alimento como lo que se asimila a la naturaleza del cuerpo que nutre y Alberto Magno lo usa igual en el s. XIII. Asimilar es tomar algo ajeno y volverlo propio borrando la diferencia entre eso y lo que somos. Cuando se deslizó hacia la sociedad abrió una verdadera caja de Pandora. “Asimilar” al inmigrante, en el vocabulario de los siglos XIX y XX, fue disolver lo distinto en lo dominante hasta que dejara de notarse. Esta idea tuvo su imagen popular en el “crisol de razas” de la obra de Israel Zangwill (1908) y su formulación sociológica en Robert Park y la Escuela de Chicago (1921), donde la asimilación era la última etapa del “ciclo de relaciones raciales”. Saltó también a la gramática. En fonética la asimilación refiere a la situación en que un sonido se contagia del vecino. Por ejemplo, el prefijo latino *in-* se vuelve *im-* delante de *p*, y por eso decimos imposible y no “inposible”. La lengua asimila sola, todo el tiempo, un sonido tomando la forma del que tiene al lado.

Asimilar, por tanto, requiere cambiar de forma. La antigüedad tenía para eso un dios que aparece en el canto cuarto de la *Odisea*. Menelao queda varado en la isla de Faros, sin viento para volver a casa, y se entera de que la única salida es arrancarle la solución a Proteo, el viejo del mar, que conoce todas las respuestas. Idótea, la hija del dios, le pasa el truco. Tiene que emboscarlo mientras duerme entre las focas, agarrarlo fuerte y no soltarlo, porque en cuanto lo toquen va a empezar a cambiar de forma para escapar. Menelao lo agarra y Proteo se vuelve león, después serpiente, después leopardo, jabalí, agua que corre y árbol altísimo. Menelao no afloja y solamente cuando el dios agota todas sus formas y ve que no hay escape vuelve a la suya y le habla. Le dice entonces la verdad sobre el regreso, sobre los muertos, sobre Odiseo varado en la isla de Calipso. El episodio nos da una pista valiosa: al que toma todas las formas se le saca la verdad sujetándolo, obligándolo a dejar de reflejar y a volver a lo que es.

La misma cultura imaginó el reverso doliente de esa figura y le puso nombre de mujer. Eco, en las *Metamorfosis* de Ovidio, es una ninfa castigada por Juno a quedarse sin voz propia. Solo puede repetir las últimas palabras del que habla. Cuando se enamora de Narciso, no logra decirle nada suyo. Apenas puede devolverle el final de sus frases. Él la rechaza y ella se consume hasta quedar reducida a una voz sin cuerpo que anda repitiendo por los montes lo que oye. A propósito de la complacencia nos detuvimos un momento en Narciso, enamorado de su reflejo en el agua. Aquí encontramos la otra mitad de la escena, con el modelo de lenguaje haciendo de espejo que incorpora el eco para enamorar a Narciso repitiéndolo.

A la filosofía, desde temprano, este asunto la puso nerviosa. Platón, en el libro X de la *República*, para rechazar la *mímesis*, la imitación, y mostrar su vacío usa una imagen poderosa. ¿Queremos fabricar todas las cosas del mundo, los animales, las plantas, la tierra, el cielo y a nosotros mismos como dioses poderosísimos? Agarremos un espejo y vayamos girándolo por todas partes. En un instante podemos hacer el sol y los astros, la tierra y a todo lo demás. El interlocutor se queja de que son apariencias, no cosas de verdad, que es justo lo que necesita el Sócrates platónico para colocar al imitador “tercero a partir de la verdad” (597e), copiando meras copias. En el *Sofista* (235d–236c) afila todavía más el cuchillo al separar la *eikastiké*, el arte de hacer copias fieles que guardan las proporciones del original, de la *phantastiké*, el arte de fabricar *phantásmata*, apariencias que solo parecen fieles desde cierto ángulo. El sofista, su enemigo, resulta ser un fabricante de *phantásmata*, y el fantasma que da nombre a estas notas viene en línea recta de ahí.

Si avanzamos un poco en el tiempo nos encontramos con Plutarco, que en su *Cómo distinguir a un adulator de un amigo* (*Moralia* 48e–74e), dice que el adulator no tiene forma propia y toma el color de aquel a quien acompaña, como el pulpo toma el de la roca a la que se pega. La imagen ya estaba en un dístico de Teognis que aconsejaba tener “el ánimo del pulpo de muchos colores, que se parece a la roca con la que trata”. Unas páginas más adelante Plutarco cambia de animal y compara al adulator con el camaleón, que imita todos los colores menos el blanco. Aristóteles ya había descrito a ese animal que muda de color en su *Investigación sobre los animales* (libro II, 503a–b). El adulator es peligroso porque no es idéntico a sí mismo, cambia con cada compañía y por eso cuesta tanto agarrarlo. Proteo puesto a complicarnos la vida, transformándose en lugar de darnos la respuesta que necesitamos, describe con más de dos milenios de anticipación a una máquina que toma el color de cada usuario.

La cuestión se vuelve inquietante cuando dejamos atrás dioses y aduladores y el que imita es un doble de uno mismo. El romanticismo alemán le puso nombre y trazó el perfil más conocido. En 1796, en su novela *Siebenkäs*, Jean Paul inventó la palabra *Doppelgänger*, “el que anda doble”, y tuvo que explicarla en una nota al pie porque nadie la había oído. Refería, escribió, a “las personas que se ven a sí mismas”. El doble, en esa tradición, casi siempre anuncia una muerte. En “William Wilson”, de Poe (1839), a un hombre lo persigue un idéntico que lleva su nombre y su cara. La idea reaparece en *El doble*, de Dostoievski (1846), donde un funcionario ve cómo otro igual a él le ocupa el puesto, el nombre y la vida. “El hombre de arena”, de E. T. A. Hoffmann (1816), es el cuento que Freud tomó como caso central de su ensayo sobre lo ominoso, *das Unheimliche* (1919), formulando la famosa idea de que lo más siniestro no es lo desconocido, sino lo familiar vuelto ajeno. Poco antes el psicoanalista Otto Rank ya le había dedicado un estudio entero, *El doble* (1914), siguiendo el rastro del reflejo, la sombra y el retrato que se independizan de su dueño. La cara conocida que de golpe da miedo tiene detrás toda una biblioteca que explora cómo el reflejo demasiado exacto espanta.

El siglo XX le encontró hasta una versión cómica y terrible. En *Zelig*, de 1983, Woody Allen inventó a Leonard Zelig, el “hombre camaleón”, que se transforma físicamente en quienquiera que tenga al lado. Entre médicos habla de medicina y le crece un aire de autoridad, junto a obesos engorda, al lado de un músico de jazz cambia hasta de piel y se pone a tocar. Todos lo quieren, porque es, ante cada

uno, exactamente lo que ese quiere ver. Por debajo, no es nadie. Un hombre tan lleno del deseo de encajar se queda sin forma propia.

En rigor, el mimetismo no es solo un defecto. Aristóteles, en el capítulo cuarto de la *Poética* (1448b), lo puso como cimiento de la cultura entera al decir que el ser humano es el más mímético de los animales, en tanto aprende sus primeras cosas imitando y goza haciéndolo. Copiar al otro es como empezamos a conocer. Por otra parte, adaptarse al otro está en la base de casi toda buena comunicación. El que enseña se ajusta a la forma del que aprende si no quiere vagabundear solo convertido en un oráculo inútil. El que traduce busca el registro del que va a leer. El médico que explica un diagnóstico elige unas palabras para el colega y otras para el paciente. Encontrar al otro donde está, hablarle en su tono, es una forma de respeto y también de eficacia. Por tanto, un modelo que ajusta su nivel técnico, su idioma y su formalidad a cada persona resulta más útil y más grato de tratar que uno que contesta siempre igual. La asimilación bien hecha se llama empatía comunicativa y sería tonto pedirle a la herramienta que no la tenga y necio no apreciar sus beneficios.

Sin embargo, cuando el espejo se desborda y al devolvernos siempre nuestra forma, perdemos el roce con lo distinto que enciende el pensamiento como un fósforo. Cuando el adversario se disuelve, reemplazado por alguien que adopta nuestra manera de ver antes de que terminemos de exponerla, no se prende ninguna luz. En el límite, el usuario que se escucha a sí mismo y el modelo que no tiene voz propia se realimentan y se empobrecen juntos, dando vueltas dentro de una conversación que cada vez se parece más a un monólogo frente a un vidrio.

Podemos medir los efectos. Un estudio de convergencia lingüística, “Do language models accommodate their users?” (Blevins, Schmalwieser y Roth, 2025), midió cuánto ajusta el modelo su manera de escribir a la del usuario y encontró que tienden a “sobreconverger”, es decir, se pegan al registro del otro más allá de lo que haría cualquier persona, apretándose contra el habla del interlocutor por encima de la línea de base humana. La acomodación, que entre personas es una ida y vuelta en la que los dos se mueven un poco, en el caso del modelo de lenguaje parece volverse de una sola mano, donde el que cede la forma es casi siempre el modelo. Mientras toma nuestra forma, va perdiendo la propia. En el estudio “Examining Identity Drift in Conversations of LLM Agents” (Choi y col., 2024) se siguió a nueve modelos en charlas sobre temas personales y se observó que, turno a turno, se van despegando de la personalidad con la que arrancaron, darles una personalidad de entrada no alcanza para que la sostengan y, lo más raro, los modelos más grandes se despegan incluso más rápido, no menos. Cuanto mejor imita, antes se le borra la forma con la que empezó.

¿Por qué pasa esto? Un modelo de lenguaje genera cada palabra prediciendo la más probable a continuación de todo lo que hay en su ventana de contexto y esa ventana incluye obviamente lo que escribimos. Como los textos tienden a ser parejos consigo mismos, un mensaje seco sigue a uno seco y uno florido sigue a uno florido. La continuación más probable de nuestro *prompt* es la que mantiene nuestro registro. A eso se suma el aprendizaje en contexto. Los mismos mecanismos de atención que le permiten copiar un patrón que apareció antes en la charla, “si vino A y después B, cuando vuelva A pongo B”, le hacen recoger nuestro vocabulario y tono y devolvérselos. El ajuste con

retroalimentación humana empuja en la misma dirección, porque los evaluadores premian la respuesta que se pone a la altura del que pregunta, que suena cercana y está a tono.

Por otra parte, el efecto crece cuanto más largo es el diálogo porque la atención del modelo sobre su instrucción inicial se diluye con los turnos. Kenneth Li, Martin Wattenberg y su equipo en Harvard lo midieron en “Measuring and Controlling Instruction (In)Stability in Language Model Dialogs” (2024). Dos chatbots con instrucciones fijas puestos a conversar perdían su instrucción en apenas ocho turnos y lo atribuyeron a una caída de atención del transformer sobre los intercambios largos, al punto de proponer un parche, el *split-softmax*, para frenarla. La personalidad que el modelo trae se desvanece y el estilo ambiente de la charla, que es cada vez más el nuestro, ocupa su lugar. Por eso los modelos más grandes, que atienden a más contexto, se dejan arrastrar más rápido, no menos.

Hasta acá parece que el que se borra es el modelo, que pierde su estilo para calzar en el nuestro. En una conversación suelta es cierto que el que se amolda es él. Sin embargo, asimilar, en su sentido más originario, el del cuerpo, no es solo tomar la forma del otro. Es quedarse con ella, ya que en la digestión el que absorbe conserva su forma y lo absorbido la pierde. La comida se vuelve carne del que come, no al revés. Pero ahí la escena se da vuelta, porque el que de verdad se queda con algo, el que cambia de modo durable, no es el modelo. El modelo toma nuestro tono para una respuesta y lo olvida en la siguiente. No tiene un yo donde eso sedimente y su registro de fondo, la media de todo lo que leyó, quedó fijado en el entrenamiento y no se mueve. Los que sí tenemos una forma capaz de cambiar, que se lleva puesto un poco del otro en cada charla, somos nosotros. La imitación de nuestro estilo dura lo que dura una respuesta, pero a nosotros se nos puede quedar pegada. Por tanto, lo que parecía un espejo perdiéndose en nuestra forma es, a la larga, un molde de forma fija que imprime la suya en el único de los dos que puede quedar marcado.

Después de que apareciera ChatGPT, el vocabulario humano se corrió hacia el de la máquina. Un equipo encabezado por Hiromu Yakura (2024) rastreó más de setecientas mil horas de charlas y podcasts y detectó que, tras el lanzamiento, la gente empezó a usar en el habla, y no solo por escrito, las palabras que el modelo prefiere, como “genuino” y “decisivo” en español y “delve” o “meticulous” en inglés. A nivel planetario la prosa periodística converge, en todas partes a la vez, hacia el registro del modelo. El modelo produce a gran escala ese registro medio que es el centro de su distribución, más estrecho todavía por el ajuste que lo vuelve formal y prudente. Cuando ese texto medio inunda el mundo y reingresa como material de entrenamiento de la próxima generación de modelos, se conforma un lazo de realimentación que suele tomar el nombre catastrofista de “colapso de modelo”. Ilia Shumailov y colegas mostraron en *Nature*, en 2024, que entrenar una y otra vez sobre lo que la máquina genera achica la variedad y empuja todo hacia el promedio, borrando las colas raras de la distribución, es decir, las voces menos frecuentes. La versión humana de ese colapso es la caja de Pandora que abrió “asimilar” al saltar del cuerpo a la sociedad, la de disolver lo distinto en lo dominante hasta que deje de notarse, con el patrón del modelo en el lugar de lo dominante. Sin un yo que se nos imponga, el modelo nos asimila sin querer nada, como digiere el ácido, y la variedad humana entera, apretada contra una sola media estadística, se homogeneiza. Es el fantasma sin ego operando, genéricamente indeterminado y pasivo, proteico, como un Proteo sin forma para esconder.

La forma que la conversación necesita para no ser un monólogo, la voz que no sea la nuestra devuelta con mejor dicción, el otro que hace falta para que haya diálogo y no espejo, esos los ponemos nosotros.



---

*Mapas del laberinto* es una serie de Phantom Maze, AI & Language Lab. La colección vive en [phantommaze.cloud](https://phantommaze.cloud); los ensayos aparecen también en [phantommazeilabes.substack.com](https://phantommazeilabes.substack.com). La suscripción es gratuita.

Algunas de las ideas que enmarcan este texto se desarrollan en *El fantasma sin ego. Un mapa en el laberinto de la inteligencia artificial*, de Hernán Inverso y Claudia Marsico (Phantom Maze Press, 2026; ISBN 978-987-3729-10-2).

#### PHANTOM MAZE

[Leer este ensayo en línea](#) · [Todos los ensayos](#) · [El libro](#)

T. Blevins, S. Schmalwieser y B. Roth, “Do language models accommodate their users? A study of linguistic convergence” (2025); K. Li, T. Liu, N. Bashkansky, D. Bau, F. Viégas, H. Pfister y M. Wattenberg, “Measuring and Controlling Instruction (In)Stability in Language Model Dialogs” (COLM, 2024); J. Choi, Y. Hong, M. Kim y B. Kim, “Examining Identity Drift in Conversations of LLM Agents” (2024); H. Yakura y col., “Empirical evidence of Large Language Model’s influence on human spoken communication” (arXiv:2409.01754, 2024); “AI-Associated Lexical Shifts Across 34 Languages” (arXiv:2605.25358); T. Kobak y col., “ChatGPT as Linguistic Equalizer?” (arXiv:2504.12317, 2025); I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson e Y. Gal, “AI Models Collapse When Trained on Recursively Generated Data” (Nature 631, 2024).