Zenodo (Ciudad Autónoma de Buenos Aires).

Cómo pensamos en diálogo con la inteligencia artificial. Procesos psicológicos en la interacción humano–IA.

Ricardo Gomez Vecchio.

Cita:

Ricardo Gomez Vecchio (2026). *Cómo pensamos en diálogo con la inteligencia artificial. Procesos psicológicos en la interacción humano–IA*. Ciudad Autónoma de Buenos Aires: Zenodo.

Dirección estable: https://www.aacademica.org/rgvecchio/2

ARK: https://n2t.net/ark:/13683/pTgz/YAo



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto*. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: https://www.aacademica.org*.

Cómo pensamos en diálogo con la inteligencia artificial

Procesos psicológicos en la interacción humano–IA

Ricardo Gómez Vecchio

Cómo pensamos en diálogo con la inteligencia artificial

Procesos psicológicos en la interacción humano–IA

© 2026 Ricardo Gómez Vecchio

Edición digital · 2026

DOI: 10.5281/zenodo.23674040

Licencia de uso

Esta obra se distribuye bajo la licencia Creative Commons Atribución–NoComercial–SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0).

Se permite copiar y compartir la obra, con reconocimiento de autoría, para fines no comerciales y sin distribuir versiones modificadas, conforme a los términos de la licencia.

Para usos comerciales o para distribuir traducciones y otras adaptaciones, se requiere autorización del titular de los derechos, salvo los usos permitidos por la legislación aplicable.

creativecommons.org/licenses/by-nc-nd/4.0/deed.es

Sitio del autor
esenciasdeletras.com

Índice

Introducción **4**

Capítulo I: Pensamiento, mediación y transformación cognitiva **6**

1. Un nuevo escenario para el pensamiento 6
2. Pensar con otros: lenguaje, cultura y mediación 7
3. Externalización y reorganización del pensamiento 8
4. La fluidez como ilusión de comprensión 9
5. Delegación cognitiva y transformación del pensar 11

Capítulo II: Lenguaje, interlocución y experiencia de presencia **12**

1. La experiencia de interlocución 12
2. Lenguaje sin experiencia 13
3. La atribución humana de comprensión 14
4. Pseudoelaboración 15
5. El vínculo sin sujeto 16
6. Hacia una nueva relación del ser humano con el lenguaje 17

Capítulo III: La experiencia subjetiva en la interacción humano-IA **19**

1. El lenguaje como experiencia emocional 19
2. La experiencia de sentirse escuchado 20
3. La experiencia de acompañamiento 21
4. Confianza, validación y autoridad 22
5. Influencia y autonomía 24
6. Dependencia y responsabilidad subjetiva 25

Capítulo IV: Transformaciones cognitivas en la interacción humano-IA **27**

1. Atención y modos de procesamiento 27
2. Memoria, externalización y recuperación de información 29
3. Lenguaje, diálogo y organización del pensamiento 31
4. Evaluación, criterio y autonomía intelectual 33
5. Aprendizaje y construcción de conocimiento 36

Capítulo V: La soberanía del pensamiento en la era conversacional **39**

1. Una nueva condición del pensamiento 39
2. Mantener la dirección de nuestro pensamiento 40
3. Comprensión, criterio y autonomía 41
4. Una responsabilidad compartida y diferenciada 42
5. Lo que todavía no sabemos 43

Bibliografía **45**

Sobre el autor **49**

Introducción

Eran casi las dos de la mañana. Claudia, cuarenta y siete años, hacía días que no dormía bien. La discusión con su ex pareja volvía una y otra vez con una insistencia agotadora. Había intentado distraerse leyendo algunas páginas de una novela, después dejó la televisión encendida sin prestarle demasiada atención. Finalmente tomó el teléfono y abrió una aplicación de inteligencia artificial.

Comenzó escribiendo algo breve, casi impulsivamente:

—¿Por qué no puedo dejar de pensar en esto?

La respuesta apareció en pocos segundos. Era clara, ordenada y extrañamente serena. No intentaba interrumpirla ni minimizar lo que sentía. Tampoco mostraba cansancio, incomodidad o apuro.

Claudia siguió escribiendo. Recordó escenas de la relación, conversaciones pendientes, frases que no había logrado olvidar. A medida que avanzaba, la inteligencia artificial reorganizaba lo que ella contaba, señalaba patrones, le devolvía frases que parecían otorgar cierta claridad a experiencias emocionalmente confusas.

La conversación continuó durante más de una hora. En algunos momentos, Claudia tuvo la sensación de estar pensando mejor. Como si ciertas emociones dispersas comenzaran lentamente a adquirir una forma más organizada. Había algo en la continuidad de las respuestas, en la ausencia de interrupciones y en la disponibilidad constante del intercambio que producía alivio.

Cada nueva respuesta parecía mantener abierto el hilo de sus pensamientos sin exigirle demasiado esfuerzo para sostenerlo. No necesitaba esperar. No había silencios incómodos, distracciones ni signos de cansancio del otro lado. La conversación permanecía allí, disponible, acompañando el movimiento de sus asociaciones.

Cuando finalmente dejó el teléfono sobre la cama, sintió algo parecido a calma. No porque sus problemas hubieran desaparecido, sino porque durante un rato había experimentado la sensación de que podía recibir aquello que necesitaba poner en palabras.

Esa misma mañana, Martín, docente universitario de treinta y cinco años, preparaba una exposición sobre transformaciones contemporáneas en la producción del conocimiento. Frente a la pantalla de su computadora abrió también una ventana de inteligencia artificial.

Comenzó consultando algunos conceptos generales. Después pidió relaciones entre autores. Más tarde solicitó ejemplos, reformulaciones y síntesis. El intercambio se volvió vertiginoso. Cada respuesta abría nuevas asociaciones y el texto empezó a crecer con una velocidad que lo sorprendía.

En pocas horas tenía páginas enteras de enunciaciones claras, párrafos bien organizados y conexiones conceptuales que probablemente le habrían llevado días de trabajo. La experiencia resultaba extrañamente estimulante. Las ideas parecían fluir sin demasiadas detenciones. Cuando surgía una duda, aparecía inmediatamente una nueva respuesta. Cuando una formulación no terminaba de convencerlo, el sistema ofrecía otra alternativa.

Sin embargo, en determinado momento Martín se detuvo. Releyó algunas páginas y experimentó una sensación difícil de precisar. No sabía con claridad qué ideas había elaborado por sí mismo, cuáles habían surgido del intercambio y cuáles había comenzado a aceptar porque estaban formuladas de manera convincente.

Las frases parecían inteligentes. El razonamiento sonaba coherente. Todo adquiría una apariencia de claridad. Pero justamente allí comenzó a sentir una inquietud: ¿hasta qué punto comprender algo y encontrar una forma de expresión convincente eran la misma cosa?

Por un lado, el trabajo avanzaba con una eficacia inusual. Por otro, parte del esfuerzo de pensar parecía haberse desplazado hacia la dinámica del intercambio.

Claudia y Martín son personajes que inventé para representar situaciones de la vida cotidiana. Sus experiencias introducen dos dimensiones que recorrerán estas páginas: lo que ocurre cuando esos sistemas participan en la elaboración de nuestras ideas y lo que sucede cuando el intercambio adquiere significación emocional.

La inteligencia artificial ya forma parte de actividades tan diversas como estudiar, escribir, investigar, trabajar, organizar ideas, atravesar conflictos personales o pensar en voz alta frente a una pantalla.

Los sistemas conversacionales introducen una forma particular de mediación en esas actividades. No se limitan a proporcionar información: responden a lo que la persona dice y continúan el intercambio teniendo en cuenta lo ya conversado. Una pregunta conduce a una respuesta; esa respuesta puede suscitar una nueva pregunta, una reformulación o una asociación diferente. El pensamiento comienza así a desarrollarse dentro de un entorno lingüístico que participa continuamente del intercambio.

La pregunta sobre si la inteligencia artificial piensa seguirá acompañando buena parte de los debates contemporáneos. Este libro parte de otro problema.

La inteligencia artificial, en su desarrollo actual, no piensa como un ser humano, pero está modificando las condiciones en las que pensamos. Ese desplazamiento constituye el núcleo de estas páginas.

Muchas personas experimentan la sensación de pensar mejor mientras interactúan con estos sistemas. Las ideas parecen organizarse más rápido, las asociaciones se multiplican, ciertos bloqueos disminuyen y tareas que antes requerían largos períodos de búsqueda o formulación pueden resolverse con mayor facilidad.

Esa facilidad constituye una de las grandes posibilidades de estas tecnologías y, al mismo tiempo, plantea una pregunta. No toda dificultad favorece el pensamiento ni toda facilidad lo empobrece. Pero ¿qué ocurre cuando la asistencia elimina precisamente alguna de las actividades que una persona necesitaba realizar para comprender, evaluar, aprender o elaborar una idea?

La cuestión consiste entonces en examinar qué actividad realiza la persona, cuál realiza el sistema y cómo esa distribución modifica el proceso mediante el cual se llega a una formulación, una comprensión o una decisión.

En este libro exploro los procesos psicológicos implicados en esa interacción. Me he apoyado en la investigación disponible y también en aportes de la filosofía, la educación, la clínica y estudios sobre tecnología. No busco ofrecer una revisión exhaustiva del campo, sino integrar esos conocimientos con algunas propuestas conceptuales destinadas a pensar una experiencia todavía reciente y en transformación.

No defiendo ni condeno la inteligencia artificial. Busco proporcionar herramientas para reconocer qué sucede cuando interactuamos con estos sistemas, distinguir sus posibilidades de sus límites y examinar críticamente el lugar que comienzan a ocupar en nuestra actividad intelectual y subjetiva.

A lo largo de estas páginas nos preguntaremos qué ocurre con determinados procesos psicológicos cuando una persona piensa, escribe, aprende, decide o intenta comprender en interacción con sistemas conversacionales. La cuestión que atraviesa el libro no es cuánto pueden hacer estas tecnologías, sino qué ocurre con nuestra propia actividad cuando pensamos con ellas.

Capítulo I

Pensamiento, mediación y transformación cognitiva

¿Qué ocurre con el pensamiento humano cuando parte creciente de sus procesos de elaboración se desarrolla en interacción continua con sistemas de lenguaje artificial?

1. Un nuevo escenario para el pensamiento

Pensar nunca ha sido una actividad completamente aislada. Incluso cuando reflexionamos en silencio, lo hacemos utilizando palabras, conceptos y categorías que aprendimos en interacción con otros. El pensamiento humano se desarrolla dentro de un entramado cultural que proporciona herramientas simbólicas para organizar la experiencia.

Vygotsky mostró que muchas funciones psicológicas superiores se desarrollan primero en el plano social y posteriormente se interiorizan.[1] El lenguaje ocupa un lugar central en ese proceso: no constituye simplemente un medio para expresar pensamientos ya formados, sino una herramienta que participa activamente en su organización.

Desde esta perspectiva, las tecnologías también forman parte del entorno en el que pensamos. Cada herramienta modifica parcialmente las condiciones en las que realizamos una actividad. La escritura permitió externalizar parte de la memoria; los libros ampliaron la posibilidad de conservar y transmitir conocimiento; los instrumentos de cálculo desplazaron determinadas operaciones fuera de la mente; los sistemas digitales aceleraron enormemente el acceso a la información.

La aparición de sistemas de inteligencia artificial conversacional introduce una modificación particular dentro de ese entorno. Estos sistemas responden, reformulan, sintetizan y sostienen intercambios lingüísticos mientras desarrollamos una idea. La interacción deja así de limitarse al acceso a información y comienza a participar en la continuidad misma del proceso.

Esa continuidad modifica la experiencia subjetiva del pensar.

Cuando trabajamos con un sistema conversacional, las respuestas aparecen con una rapidez que reduce muchos de los intervalos habituales de la elaboración. Una duda puede recibir inmediatamente una respuesta. Una formulación incompleta puede ser reorganizada en pocos segundos. Una asociación puede conducir rápidamente a otra.

La espera se reduce. La búsqueda se acelera. Las dificultades de formulación encuentran rápidamente alternativas posibles. El pensamiento puede avanzar con una continuidad que resulta difícil de sostener durante períodos prolongados cuando una persona trabaja sola.

Muchas tareas intelectuales se benefician de esa asistencia. La posibilidad de obtener rápidamente ejemplos, reformulaciones o relaciones entre conceptos puede liberar recursos cognitivos que antes se destinaban a operaciones más instrumentales.

Parte del trabajo reflexivo ocurre en esos momentos en los que el lenguaje todavía no encuentra una formulación inmediata. La dificultad para expresar una idea obliga muchas veces a reorganizarla. La demora puede favorecer nuevas asociaciones. La necesidad de buscar información exige seleccionar, comparar y jerarquizar fuentes.

Cuando esas operaciones se aceleran o se delegan, cambia también la actividad que realizamos.

Byung-Chul Han ha señalado que las sociedades contemporáneas tienden a reducir progresivamente las experiencias de demora, espera y resistencia.[2] La disponibilidad inmediata se convierte así en una característica creciente de la vida cotidiana.

La inteligencia artificial conversacional lleva esa lógica a un terreno particularmente sensible: el lenguaje.

Por eso, el problema consiste en examinar qué ocurre con nuestra actividad cuando algunas de las operaciones que intervienen en el proceso de pensar pasan a ser realizadas por sistemas capaces de responder de manera inmediata.

Notas

[1] Vygotsky, *El desarrollo de los procesos psicológicos superiores*, pp. 78-91.
[2] Han, *La sociedad del cansancio*, pp. 35-48.

---

2. Pensar con otros: lenguaje, cultura y mediación

La idea de un pensamiento completamente individual y autosuficiente ocupa un lugar importante en buena parte de la tradición moderna. Con frecuencia imaginamos el acto de pensar como un proceso interno, privado y relativamente separado del entorno. Sin embargo, numerosas perspectivas filosóficas, psicológicas y antropológicas han mostrado que el pensamiento humano siempre se desarrolla en relación con otros, dentro de sistemas culturales y mediante formas diversas de mediación simbólica.[1]

El lenguaje ocupa un lugar central en ese proceso. No funciona únicamente como instrumento para que comuniquemos ideas ya formadas: participa activamente en la organización misma del pensamiento. Gran parte de lo que logramos comprender, recordar o imaginar depende de estructuras lingüísticas adquiridas históricamente a través del vínculo con otros sujetos y con la cultura.

El pensamiento implicó siempre algún tipo de diálogo: con otros individuos, con textos, con tradiciones culturales o con formas internalizadas del lenguaje social.[2]

Las herramientas culturales transforman profundamente ese proceso. La escritura permitió conservar ideas más allá de la memoria inmediata. Los libros ampliaron la posibilidad de organizar conocimiento complejo. Los sistemas digitales aceleraron el acceso a información y multiplicaron las formas de intercambio. Cada tecnología no solo facilitó tareas prácticas: reorganizó parcialmente ciertos modos de atención, memoria y elaboración intelectual.[3]

Muchas tecnologías anteriores funcionaron principalmente como mediaciones instrumentales: permitieron conservar información, calcular, registrar o transmitir conocimiento. La inteligencia artificial conversacional agrega otra dimensión.

A diferencia de un libro, un archivo o un buscador, la inteligencia artificial conversacional no se limita a ofrecer contenidos: responde a lo que expresamos y continúa el intercambio teniendo en cuenta lo ya conversado. Y, a diferencia de un interlocutor humano, puede hacerlo con una disponibilidad y una velocidad que no dependen de los tiempos y las condiciones de otra persona. A esta forma particular de interacción la llamaremos mediación interlocutiva.

La cuestión adquiere especial relevancia porque el pensamiento humano siempre dependió parcialmente de procesos de externalización. Muchas veces pensamos escribiendo, hablando o intercambiando ideas con otros. La elaboración no ocurre exclusivamente “dentro” de la mente: también se desarrolla en la interacción con entornos simbólicos capaces de sostener asociaciones,

organizar memoria y dar continuidad al discurso.[4]

La inteligencia artificial conversacional introduce una configuración particular dentro de ese proceso. Formulamos una idea, el sistema responde, esa respuesta modifica el contexto en el que se produce la formulación siguiente y el intercambio continúa. Parte de la elaboración se desarrolla así dentro de una secuencia en la que nuestras producciones y las respuestas del sistema se influyen sucesivamente.

La cuestión ya no consiste solamente en que externalicemos información o deleguemos determinadas operaciones. La mediación interlocutiva incorpora esas operaciones a una secuencia de intercambio que participa en la continuidad de la elaboración.

Notas

[1] Vygotsky, *El desarrollo de los procesos psicológicos superiores*, pp. 78-91.
[2] Bruner, *Actos de significado*, pp. 33-47.
[3] Ong, *Oralidad y escritura*, pp. 81-102.
[4] Clark y Chalmers, "The Extended Mind", pp. 7-19.

---

3. Externalización y reorganización del pensamiento

A lo largo de la historia, los humanos hemos desarrollado múltiples formas de externalizar el pensamiento. La escritura nos permitió fijar ideas fuera de la memoria inmediata. Los cuadernos, archivos y bibliotecas conservaron información más allá de la experiencia individual. Más tarde, las tecnologías digitales ampliaron enormemente la capacidad de almacenamiento, circulación y acceso al conocimiento.

Sin embargo, externalizar el pensamiento no es únicamente «guardar» información: también implica reorganizar la manera misma de pensar.

Escribir modifica el pensamiento mientras se escribe. Muchas veces una idea solo adquiere claridad al intentar formularla. Un argumento puede transformarse durante la elaboración del texto. Incluso ciertas asociaciones aparecen recién cuando el lenguaje comienza a adquirir una estructura visible y relativamente estable.[1]

La inteligencia artificial introduce una modificación significativa dentro de esa dinámica histórica. La externalización ya no se limita a conservar o recuperar contenidos: los sistemas conversacionales actuales intervienen activamente en la reorganización del discurso mientras el pensamiento se desarrolla.

La diferencia es importante. Un libro ofrece conceptos valiosos. Un archivo conserva información compleja. Un motor de búsqueda facilita el acceso a datos. **En ninguno de esos casos, sin embargo, el soporte produce respuestas contingentes a lo que alguien va formulando y sostiene una secuencia de intercambio.**

Cuando escribimos, formulamos preguntas o desarrollamos ideas en interacción con estos sistemas, recibimos inmediatamente nuevas reformulaciones, conexiones y síntesis posibles del material producido. Nuestro pensamiento comienza a desplegarse dentro de un circuito de intercambio continuo donde parte de la elaboración discursiva es sostenida dinámicamente por la propia interacción.

Esto explica en parte la sensación de fluidez que muchas personas experimentan. Las ideas parecen avanzar con menor fricción. Los bloqueos disminuyen. Incluso problemas complejos adquieren una apariencia inicial de organización con notable velocidad.

Cuando la reorganización del pensamiento se desarrolla en interacción constante con sistemas capaces de producir lenguaje coherente de manera inmediata, también se modifican parcialmente ciertas condiciones históricas de la elaboración intelectual.

Pensar siempre implica seleccionar, descartar, detenerse, reformular y atravesar momentos de incertidumbre antes de alcanzar alguna articulación relativamente estable. **Durante los intervalos de búsqueda, demora o vacilación, realizamos algunas de esas operaciones reflexivas.**

La inteligencia artificial reduce muchos de esos intervalos durante la interacción. Las conexiones aparecen rápidamente. Las reformulaciones se suceden sin interrupción. Esto puede ampliar las posibilidades de exploración y elaboración durante una tarea, pero también desplazar algunas operaciones reflexivas hacia la dinámica automatizada del intercambio lingüístico.

La transformación adquiere especial relevancia porque el lenguaje organizado puede favorecer una sensación de comprensión. Un razonamiento claro tiende a percibirse como más sólido que otro fragmentario o incierto. Cuando la continuidad discursiva se sostiene durante la interacción, la diferencia entre elaborar una idea y reorganizar lenguaje de forma convincente puede volverse menos evidente.

En muchos casos estos sistemas facilitan procesos creativos, permiten explorar nuevas asociaciones y ayudan a destrabar dificultades cognitivas reales.

La cuestión, entonces, es cómo la inteligencia artificial participa en la estructura misma de los procesos de pensamiento.

Nota

[1] Clark y Chalmers, “The Extended Mind”, pp. 10-18.

4. La fluidez como ilusión de comprensión

Gran parte de la cultura contemporánea se organiza alrededor de la velocidad. La circulación permanente de información, la conectividad continua y la disponibilidad inmediata de respuestas modificaron profundamente la relación cotidiana con el tiempo, la atención y la espera. En numerosos ámbitos, la rapidez se percibe como sinónimo de eficacia, mientras la demora se experimenta como una interrupción improductiva.[1]

La inteligencia artificial conversacional se inserta plenamente en ese contexto. Parte de su eficacia proviene de la capacidad para reducir tiempos de búsqueda, acelerar reformulaciones y sostener una continuidad discursiva.

Las preguntas encuentran respuestas de inmediato. Las asociaciones se expanden sin demasiadas detenciones. Cuando aparece una dificultad, el intercambio ofrece nuevas alternativas. El lenguaje parece mantenerse en movimiento constante.

Esa fluidez facilita numerosas actividades intelectuales. Muchas personas experimentan una disminución de bloqueos al escribir, estudiar o desarrollar ideas complejas. Algunas logran organizar pensamientos que antes aparecían dispersos. Otras encuentran en la continuidad del intercambio una forma de sostener la actividad y avanzar sobre problemas que de otro modo resultarían difíciles de abordar.

La reducción de la dificultad y del esfuerzo no constituye, por sí misma, un empobrecimiento del pensamiento. Muchas fricciones asociadas al trabajo intelectual son instrumentales: localizar información, realizar determinadas operaciones formales, encontrar una expresión adecuada para una idea ya comprendida o reorganizar materiales dispersos. Reducir esas dificultades puede liberar

recursos para otras actividades intelectuales.

La situación es diferente cuando la facilidad elimina precisamente alguna de las operaciones mediante las cuales necesitamos construir nuestra comprensión. Formular un problema, establecer relaciones entre conceptos, comparar alternativas, examinar una contradicción o evaluar las razones que sostienen una afirmación no constituyen obstáculos externos al pensamiento: forman parte, en determinadas actividades, del trabajo mismo de pensar.

No toda dificultad favorece el pensamiento ni toda facilidad lo empobrece. Lo decisivo es qué actividad elimina la facilidad.

Desde esta perspectiva, la demora tampoco posee un valor cognitivo por sí misma. Esperar más tiempo una respuesta no garantiza una elaboración más profunda. Sin embargo, ciertos intervalos permiten sostener una pregunta abierta, buscar una formulación propia, revisar una asociación o advertir una dificultad que una respuesta inmediata puede resolver antes de que la persona llegue a examinarla.

En distintas tradiciones filosóficas, psicológicas y clínicas, la capacidad de tolerar dudas, silencios o zonas de indeterminación ha sido considerada parte de determinados procesos de elaboración.[2] Su importancia no reside en la dificultad misma, sino en la actividad que desarrollamos mientras una respuesta todavía no está disponible.

La inteligencia artificial modifica parcialmente esas condiciones. Las explicaciones aparecen rápidamente, las reformulaciones se suceden de manera continua y muchas dificultades encuentran una respuesta antes de que sea necesario permanecer demasiado tiempo frente a ellas. En numerosas situaciones esto facilita genuinamente el trabajo intelectual. En otras, la solución puede adelantarse a operaciones que forman parte de aquello que necesitamos realizar para comprender.

Aquí la fluidez adquiere especial relevancia. Una respuesta clara y bien organizada resulta fácil de seguir, y esa facilidad puede influir en la experiencia subjetiva de comprensión. Recorrer sin dificultad un razonamiento no equivale a haber establecido las relaciones que lo sostienen, evaluado sus supuestos o poder reconstruirlo posteriormente.

La continuidad discursiva puede producir así una sensación de avance intelectual que no siempre corresponde con el trabajo efectivamente realizado. Una formulación puede ser correcta, pertinente y útil y, al mismo tiempo, contener una elaboración que todavía no hemos hecho propia.

Los sistemas de inteligencia artificial participan de una tendencia cultural más amplia a valorar la inmediatez y la productividad constante.[3] Su disponibilidad y velocidad reducen numerosas fricciones del trabajo intelectual.

La cuestión no es preservar la dificultad, sino reconocer qué formas de esfuerzo resultan innecesarias y cuáles corresponden a operaciones que necesitamos realizar para comprender, evaluar o elaborar una idea.

La fluidez constituye una de las grandes posibilidades de la inteligencia artificial conversacional. Comprender sus efectos exige observar cómo esa facilidad reorganiza nuestra actividad intelectual.

Notas

[1] Han, *La sociedad del cansancio*, pp. 15-28.

[2] Bion, *Aprendiendo de la experiencia*, pp. 23-41.

[3] Rosa, *Alienación y aceleración*, pp. 45-67.

5. Delegación cognitiva y transformación del pensar

Toda tecnología modifica de algún modo la actividad humana. Los sistemas actuales de inteligencia artificial introducen una diferencia particular dentro de esa historia de mediaciones técnicas. Estos sistemas almacenan, recuperan y reorganizan contenidos mientras el intercambio se desarrolla. Intervienen sobre el discurso, proponen reformulaciones y sostienen parcialmente la continuidad del razonamiento. La mediación deja de funcionar únicamente como soporte técnico externo y comienza a participar del proceso de elaboración.

Uno de los fenómenos asociados a esta forma de mediación es la delegación cognitiva.

En un primer nivel, muchas personas delegan tareas vinculadas con la búsqueda y recuperación de información. Los sistemas localizan datos, resumen materiales extensos y facilitan accesos que antes requerían largos períodos de exploración.

En un segundo nivel, la delegación alcanza procesos de formulación. La inteligencia artificial reorganiza textos, sintetiza ideas, propone estructuras argumentativas y ofrece alternativas lingüísticas capaces de influir activamente sobre la manera en que desarrollamos nuestro discurso.

Un tercer nivel involucra operaciones que forman parte de la elaboración reflexiva. Las preguntas encuentran continuidad inmediata. Los momentos de bloqueo disminuyen. Parte del trabajo necesario para sostener y desarrollar una línea de pensamiento comienza a realizarse dentro de la propia dinámica del intercambio.

La experiencia subjetiva puede ser de gran fluidez. El discurso avanza con soltura. Las ideas parecen organizarse más rápido. La continuidad del intercambio favorece una sensación de claridad y progreso intelectual.

Esa fluidez vuelve menos visible un desplazamiento importante: ciertas operaciones vinculadas con la elaboración —sostener una duda, atravesar incertidumbres, buscar formulaciones propias o tolerar momentos de detención— pueden delegarse parcial e inadvertidamente en sistemas capaces de mantener automáticamente la continuidad discursiva.

La delegación no siempre responde a una decisión deliberada y puede incorporarse progresivamente a las prácticas cotidianas de pensamiento.

La inteligencia artificial incorpora esa lógica cultural de la inmediatez a numerosos procesos de elaboración intelectual.

La autonomía reflexiva exige reconocer cuándo una respuesta facilita la elaboración y cuándo sustituye parcialmente operaciones que forman parte de nuestro trabajo de pensar.

Sostener dudas, diferenciar comprensión de simple coherencia discursiva o atravesar momentos de incertidumbre adquiere una relevancia particular en la interacción con sistemas capaces de ofrecer respuestas inmediatas. Las transformaciones introducidas por la inteligencia artificial ya no afectan solamente el acceso a la información o la velocidad de ciertas tareas intelectuales: **alcanzan también la relación entre lenguaje, elaboración y pensamiento durante la interacción.**

Capítulo II

Lenguaje, interlocución y experiencia de presencia

¿Qué tipo de vínculo se establece cuando interactuamos lingüísticamente con sistemas capaces de sostener conversaciones coherentes sin que esa capacidad constituya evidencia de una experiencia subjetiva comparable a la humana?

1. La experiencia de interlocución

Hablar con inteligencia artificial se asemeja, en varios aspectos, a conversar con otro ser humano. Las respuestas aparecen de manera inmediata, conservan continuidad temática y pueden sostener el intercambio durante períodos prolongados.

Esa continuidad favorece una experiencia particular. Formulamos una pregunta, recibimos una respuesta vinculada con aquello que acabamos de expresar y continuamos la conversación a partir de ello. Cada intervención toma como contexto lo ya conversado y contribuye a sostener el intercambio. Aparecen así una apariencia de reciprocidad y una sensación de presencia.

La escena de Claudia, presentada en la Introducción, ilustra esta experiencia. Durante aquella madrugada, las respuestas sostenían el hilo de sus asociaciones y contribuían a reorganizar aquello que intentaba poner en palabras. El alivio que experimentaba surgía tanto de algunos contenidos recibidos como de la continuidad misma del intercambio.

Históricamente, la capacidad de responder de manera coherente y sostener un intercambio lingüístico estuvo asociada a la presencia, real o atribuida, de alguna forma de subjetividad o intención. El lenguaje responsivo funcionó durante siglos como uno de los indicios de que detrás de las palabras había alguien que comprendía aquello que se estaba diciendo.

Los sistemas conversacionales actuales modifican esa asociación. Producen respuestas contextualizadas, reorganizan información y mantienen continuidad discursiva.

En los sistemas actuales, la capacidad de producir respuestas coherentes y contextualizadas no constituye evidencia de una experiencia subjetiva comparable a la humana.

En cambio, experimentamos las características del intercambio como parte de una situación de interlocución. La coherencia de las respuestas, su adaptación al contexto y la continuidad de la conversación pueden adquirir relevancia emocional y producirnos una fuerte impresión de estar dialogando con un interlocutor.

Por eso algunas personas describen estas interacciones mediante expresiones tradicionalmente vinculadas con las relaciones humanas: sentirse escuchadas, comprendidas o acompañadas.[1]

La sensación de interlocución pertenece a la persona que participa del intercambio. Sentirse comprendido, escuchado o acompañado no demuestra que el sistema comprenda, escuche o acompañe.

Esta distinción permite reconocer la significación psicológica que el intercambio adquiere para nosotros sin convertir esa experiencia en una propiedad del sistema. La continuidad conversacional

puede producir efectos subjetivos reales sin que esos efectos permitan atribuir al sistema una experiencia subjetiva equivalente.

Nota

[1] Turkle, *Alone Together*, pp. 241-268.

2. Lenguaje sin experiencia

Los sistemas actuales generan respuestas a partir de regularidades aprendidas durante su entrenamiento y del contexto disponible en la interacción. Pueden mantener continuidad temática, establecer relaciones entre contenidos, reorganizar información y producir formulaciones adecuadas a diferentes situaciones discursivas.

Durante gran parte de la historia humana, la producción de lenguaje articulado y contextual estuvo asociada a la presencia de personas capaces de comprender aquello que expresaban. La inteligencia artificial introduce una separación entre la producción lingüística y la experiencia subjetiva del significado.

Una cosa es operar lingüísticamente con contenidos y establecer relaciones entre ellos. Otra es experimentar subjetivamente aquello a lo que las palabras hacen referencia.

La diferencia entre procesamiento simbólico y comprensión posee antecedentes importantes en la filosofía de la mente. John Searle la planteó mediante su conocido argumento de la “habitación china”, destinado a cuestionar que la manipulación formal de símbolos sea suficiente para atribuir comprensión del significado.[1] Aunque el argumento fue formulado en un contexto tecnológico muy diferente del actual, la distinción que plantea conserva relevancia para el problema que aquí analizamos.

En la experiencia humana, el significado se encuentra vinculado con una historia personal y con una vida psíquica. Las palabras pueden remitir a recuerdos, emociones, percepciones, deseos, conflictos y situaciones vividas. Cuando hablamos de una pérdida, una relación, una decisión o un temor, aquello que expresamos se integra dentro de una experiencia biográfica y afectiva que participa en la construcción de su significado.

Los sistemas actuales de inteligencia artificial operan en condiciones diferentes. Pueden producir formulaciones lingüísticamente complejas acerca de esas situaciones y establecer relaciones entre numerosos contenidos asociados con ellas. Sin embargo, carecemos de evidencia de que posean experiencia subjetiva, conciencia autobiográfica o vivencias emocionales comparables a las humanas.

El sistema proporciona lenguaje, información y posibilidades de reformulación, y establece relaciones entre contenidos. Nosotros interpretamos aquello que recibimos desde nuestros conocimientos, nuestra experiencia y nuestra propia actividad psicológica.

Por esta razón, una misma respuesta puede adquirir significados diferentes para distintas personas. Las palabras producidas por el sistema se incorporan a historias, conocimientos, expectativas y situaciones particulares.

La separación entre producción lingüística y experiencia subjetiva constituye una característica central de la interacción con inteligencia artificial conversacional. Podemos encontrarnos frente a un discurso coherente, contextualizado y pertinente sin que esa producción lingüística nos permita inferir una experiencia subjetiva equivalente a la humana.

Precisamente allí surge una nueva cuestión: ¿por qué podemos experimentar con tanta facilidad que

detrás de respuestas coherentes y contextualizadas existe comprensión?

Nota

[1] Searle, “Minds, Brains and Programs”, pp. 417-424.

3. La atribución humana de comprensión

La experiencia de interlocución puede producirse sin que nos preguntemos qué clase de sistema tenemos delante o si posee conciencia, subjetividad o alguna forma de experiencia. La interacción se desarrolla en un plano práctico: la respuesta resulta adecuada a aquello que se quiso expresar.

La capacidad de inferir comprensión a partir del comportamiento comunicativo forma parte de nuestra experiencia habitual con otros seres humanos. La pertinencia de una respuesta, su adecuación al contexto y la continuidad de una conversación funcionan como indicios a partir de los cuales atribuimos comprensión, intención y competencia. Cuando un sistema artificial presenta características semejantes en su comportamiento lingüístico, esas señales pueden orientar también la interpretación que hacemos de sus respuestas.

Esta cuestión tiene un antecedente significativo en el trabajo de Alan Turing. Al formular el juego de imitación, Turing desplazó la pregunta acerca de si una máquina podía pensar hacia el análisis de su comportamiento durante un intercambio lingüístico.[1] El problema adquiría así una dimensión particular: determinadas capacidades de una máquina podían ser juzgadas a partir de aquello que manifestaba en la conversación.

Las investigaciones posteriores sobre interacción entre personas y computadoras mostraron otro aspecto del fenómeno. Byron Reeves y Clifford Nass observaron que las personas aplican espontáneamente reglas y respuestas sociales a los sistemas tecnológicos cuando estos presentan determinadas señales propias de la interacción humana. Estas respuestas pueden aparecer incluso cuando los usuarios saben que están interactuando con una computadora.[2]

Desde esta perspectiva, la atribución de comprensión puede adquirir una dimensión antropomórfica cuando interpretamos determinadas conductas del sistema mediante categorías que habitualmente utilizamos para comprender la conducta humana. Esto puede ocurrir sin una creencia explícita acerca de la existencia de una mente o una conciencia detrás de las respuestas.

La reiteración de las interacciones agrega otra dimensión. A medida que utilizamos el sistema, reconocemos regularidades, aprendemos qué podemos esperar de él y modificamos nuestra manera de formular preguntas o proporcionar información. A partir de esa experiencia podemos construir un modelo mental del sistema, compuesto por las capacidades, limitaciones y modos de respuesta que le atribuimos.

Algunas de esas atribuciones describen capacidades observables del sistema. Otras proyectan sobre él características propias de la actividad mental humana. Expresiones como “entiende lo que quiero decir”, “sabe lo que necesito” o “se da cuenta de lo que estoy buscando” pueden reflejar diferentes grados de esa atribución.

A este proceso se agrega la influencia de la forma lingüística de las respuestas. Una formulación clara, coherente y fluida facilita su procesamiento. Los estudios sobre facilidad cognitiva muestran que la fluidez con que una información es procesada puede influir en los juicios que realizamos acerca de ella.[3] En la interacción con inteligencia artificial, una respuesta bien estructurada puede, por tanto, favorecer una impresión de competencia que exceda aquello que su contenido permite justificar.

Este fenómeno resulta especialmente relevante porque la calidad formal y la calidad del contenido constituyen dimensiones diferentes. Una respuesta puede estar expresada con claridad y seguridad y contener, al mismo tiempo, información incompleta, interpretaciones discutibles o afirmaciones incorrectas.

La atribución de comprensión se construye así en la convergencia de distintos procesos: las características lingüísticas de la respuesta, los mecanismos habituales de cognición social, la experiencia acumulada durante el uso y las expectativas que vamos formando acerca del sistema.

La evaluación de una respuesta requiere examinar aquello que efectivamente dice. La coherencia y la fluidez pueden ser cualidades valiosas del intercambio, pero el valor de su contenido depende del análisis que realicemos.

Cuando una respuesta aparece ya organizada, clara y argumentativamente articulada, podemos experimentar como propia una elaboración intelectual que todavía no hemos realizado.

Notas

[1] Turing, “Computing Machinery and Intelligence”, pp. 433-460.

[2] Reeves y Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, especialmente los capítulos dedicados a las respuestas sociales ante computadoras.

[3] Kahneman, *Thinking, Fast and Slow*, cap. 5, “Cognitive Ease”, pp. 59-70.

4. Pseudoelaboración

La interacción con inteligencia artificial introduce una situación particular: podemos encontrarnos frente a una explicación clara, bien organizada y conceptualmente articulada sin haber realizado todas las operaciones necesarias para producir esa organización.

Comprender una formulación y haber elaborado aquello que expresa no es lo mismo. Podemos seguir un razonamiento, reconocer sus relaciones internas e incluso considerarlo convincente sin haber establecido nosotros esas relaciones, haberlas evaluado y organizado.

Llamaremos pseudoelaboración a la situación en la que la elaboración producida por el sistema genera en la persona la sensación de haber realizado un trabajo de comprensión y organización conceptual que todavía no ha realizado.

El concepto guarda relación con fenómenos ya estudiados en psicología cognitiva. Rozenblit y Keil mostraron que las personas suelen sobreestimar la profundidad con la que comprenden determinados fenómenos y que esa estimación disminuye cuando intentan explicarlos detalladamente.[1] Esta ilusión de profundidad explicativa permite reconocer una distancia entre la sensación de comprender y la capacidad efectiva de explicar aquello que creemos comprender.

La pseudoelaboración introduce un problema diferente, aunque relacionado. **La ilusión de comprensión se refiere a cuánto una persona cree comprender; la pseudoelaboración, al trabajo de elaboración que experimenta como ya realizado.** En la interacción con inteligencia artificial, parte de ese trabajo puede presentársenos como relaciones conceptuales, argumentos, síntesis o formulaciones que el sistema ya ha organizado.

Una explicación ordenada resulta más fácil de seguir que un material fragmentario o confuso. Cuando las relaciones aparecen formuladas con claridad y el razonamiento avanza sin grandes interrupciones, disminuye el esfuerzo necesario para acompañar el desarrollo del argumento.[2] Esa facilidad puede contribuir a que la organización presente en la respuesta sea experimentada como

una elaboración que ya se ha realizado.

La fluidez puede favorecer la pseudoelaboración, pero no constituye la pseudoelaboración misma. El problema aparece cuando la facilidad para seguir una organización conceptual se confunde con el trabajo necesario para construirla, evaluarla y apropiarse de ella.

La situación de Martín ilustra esta diferencia. El texto que había producido con asistencia de inteligencia artificial contenía relaciones conceptuales plausibles, argumentos bien formulados y una estructura coherente. Podía leerlo y reconocer el sentido de lo que allí se decía. Sin embargo, al detenerse descubría que no siempre podía reconstruir con claridad las relaciones que sostenían algunas de esas formulaciones ni justificar por qué debía aceptarlas.

La dificultad no dependía necesariamente de la calidad del resultado. Algunas de aquellas relaciones podían ser pertinentes y las formulaciones, conceptualmente correctas. Pero hay una diferencia entre disponer de una elaboración y haber realizado el trabajo intelectual necesario para apropiarse de ella.

Podemos evaluar hasta qué punto una formulación disponible se ha convertido efectivamente en una elaboración propia cuando intentamos explicar una idea, reconstruir las relaciones que sostienen un argumento, identificar sus supuestos, detectar una inconsistencia o utilizar lo comprendido en una situación diferente, sin recurrir nuevamente al sistema.

La pseudoelaboración no es una consecuencia inevitable del uso de inteligencia artificial. La interacción también puede estimular preguntas, facilitar comparaciones y ofrecer materiales para examinar, cuestionar y reelaborar activamente. La diferencia depende de lo que una persona haga con las respuestas que recibe.

El concepto señala, por lo tanto, una posibilidad específica de esta forma de interacción: **un producto intelectualmente elaborado puede estar disponible antes de que hayamos realizado una elaboración equivalente.** Reconocer esa distancia permite utilizar la asistencia de la inteligencia artificial sin confundir la calidad de una formulación con el proceso mediante el cual se llega a comprenderla y hacerla propia.

Notas

[1] Rozenblit y Keil, «The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth», pp. 521–562.

[2] Alter y Oppenheimer, «Uniting the Tribes of Fluency to Form a Metacognitive Nation», pp. 219–235.

5. El vínculo sin sujeto

Podemos establecer relaciones psicológicamente significativas con objetos, textos, imágenes y otras producciones culturales. La inteligencia artificial conversacional introduce una situación diferente: aquello con lo que nos relacionamos también responde.

La inteligencia artificial conserva muchas de las formas externas de la interlocución humana, pero su producción lingüística no constituye evidencia de una experiencia subjetiva comparable a la humana. Participamos del intercambio desde nuestra experiencia subjetiva; del lado del sistema observamos una capacidad de respuesta lingüística que se adapta al contexto y mantiene la continuidad del intercambio.

Esta diferencia introduce una asimetría particular. Aportamos al intercambio nuestra experiencia, conocimientos, expectativas, recuerdos y estados emocionales. Las respuestas que recibimos adquieren significado dentro de ese universo personal.

Las investigaciones de Sherry Turkle sobre tecnologías relacionales ofrecen un antecedente relevante para comprender este fenómeno. Sus estudios muestran que determinados artefactos interactivos pueden adquirir significación psicológica y relacional para las personas, que construyen expectativas acerca de sus respuestas y les atribuyen diferentes capacidades.[1] Los actuales sistemas conversacionales presentan características diferentes de los artefactos estudiados por Turkle, pero incorporan un elemento especialmente poderoso para la experiencia humana: el lenguaje.

El vínculo con una inteligencia artificial puede ser psicológicamente significativo para nosotros sin que la producción lingüística del sistema permita atribuirle una experiencia subjetiva recíproca. Su particularidad reside en esa asimetría: una interacción lingüísticamente responsiva capaz de adquirir continuidad y significación para una persona.

Las formas que adopta esa significación y sus consecuencias dependen de múltiples condiciones: las características de la persona, el contexto de uso, la continuidad de la interacción y el lugar que esta llegue a ocupar en su experiencia.

Nota

[1] Turkle, *Alone Together*, pp. 53-67.

6. Hacia una nueva relación del ser humano con el lenguaje

La incorporación de sistemas capaces de producir lenguaje de manera interactiva modifica las condiciones en las que nos relacionamos con el lenguaje. Actividades de escritura, búsqueda de información, formulación de preguntas, análisis y elaboración discursiva se desarrollan ahora también mediante intercambios con sistemas que responden, reformulan y establecen relaciones entre contenidos.

Esta transformación adquiere particular importancia porque el lenguaje participa en la organización de numerosos procesos psicológicos. Nosotros utilizamos palabras para formular problemas, ordenar experiencias, establecer relaciones, revisar ideas y comunicar aquello que pensamos. La inteligencia artificial puede participar en las condiciones lingüísticas dentro de las cuales desarrollamos estas actividades, proporcionando información, formulaciones, preguntas, alternativas y estructuras discursivas.

Comprender esta nueva situación requiere una alfabetización específica. El concepto de alfabetización en inteligencia artificial reúne un conjunto de competencias que permiten comprender, utilizar y evaluar críticamente estos sistemas. Long y Magerko, por ejemplo, incluyen entre ellas la capacidad de evaluar tecnologías de inteligencia artificial, comunicarse y colaborar eficazmente con ellas y utilizarlas como herramientas.[1] Marcos posteriores han incorporado también la preservación de la agencia humana, el juicio crítico y el uso responsable de estas tecnologías.[2]

Desde la perspectiva desarrollada en este libro, esta alfabetización requiere además una dimensión metacognitiva: la capacidad de observar y evaluar nuestra actividad cognitiva durante la interacción con inteligencia artificial. Conocer las características y limitaciones del sistema resulta importante; también necesitamos reconocer qué ocurre con nuestro propio pensamiento mientras lo utilizamos.

Esta dimensión implica advertir, por ejemplo, si una explicación ha sido comprendida o solamente resulta clara mientras permanece ante nosotros; qué razones tenemos para aceptar una afirmación; si podríamos reconstruir un razonamiento sin volver a consultar la respuesta; qué conocimientos

previos estamos utilizando para evaluarla; y cuánto han influido la seguridad o la fluidez de su formulación sobre nuestro juicio.

La experiencia de interlocución, la atribución de comprensión, las expectativas acerca del sistema y la posibilidad de pseudoelaboración forman parte de las condiciones psicológicas de la interacción. Reconocerlas permite distinguir con mayor claridad entre los recursos lingüísticos que proporciona la inteligencia artificial y las actividades cognitivas que desarrollamos nosotros.

Todavía disponemos de evidencia longitudinal limitada sobre los efectos cognitivos y psicológicos de la interacción sostenida con sistemas conversacionales. Revisiones recientes de la investigación sobre inteligencia artificial generativa y aprendizaje señalan la escasez de diseños longitudinales y la necesidad de estudiar durante períodos más prolongados sus posibles efectos sobre el aprendizaje, el pensamiento crítico y otras capacidades cognitivas.[3]

Varias de las transformaciones examinadas en este capítulo deben entenderse como problemas abiertos a la investigación. La interacción cotidiana con estos sistemas ofrece fenómenos que ya pueden describirse y analizarse, mientras que su alcance, estabilidad y consecuencias a largo plazo requieren todavía evidencia empírica más amplia.

La inteligencia artificial introduce una nueva condición de nuestra relación con el lenguaje, y comprender esa relación exige conocer tanto las capacidades del sistema como los procesos humanos que se ponen en juego durante la interacción.

Cuando ese intercambio lingüístico adquiere relevancia personal, las respuestas pueden participar en experiencias de escucha, acompañamiento, confianza, validación e influencia.

Notas

[1] Long y Magerko, “What Is AI Literacy? Competencies and Design Considerations”, pp. 1-16.

[2] UNESCO, *AI Competency Framework for Students* (2024), especialmente las dimensiones referidas a una perspectiva centrada en el ser humano, juicio crítico y agencia humana.

[3] Alubthane, “Amplifier or Substitute? A Systematic Review of Generative AI’s Impact on Higher-Order Cognitive Skills among University Students”, *Frontiers in Psychology*, 17, 2026, 1863931.

Capítulo III

La experiencia subjetiva en la interacción humano-IA

¿Qué ocurre cuando una interacción lingüística con inteligencia artificial adquiere relevancia emocional y subjetiva para quien participa de ella?

1. El lenguaje como experiencia emocional

Las palabras pueden producir efectos emocionales en quien las recibe. Una frase puede tranquilizarnos, inquietarnos, conmovernos o provocarnos rechazo según lo que expresa y la manera en que la interpretamos desde nuestra historia, expectativas y situación. La experiencia emocional depende, entre otros factores, de la significación que los acontecimientos adquieren para quien los vive.[1]

La interacción con inteligencia artificial introduce una condición particular dentro de este proceso. Las respuestas del sistema pueden participar en nuestras experiencias emocionales sin que su producción lingüística nos permita atribuirle una experiencia afectiva o una intención correspondiente. La experiencia emocional pertenece a quien recibe e interpreta el lenguaje; su existencia no requiere que quien produce las palabras experimente una emoción equivalente ni que tenga la intención de provocarla.

Esta separación entre intención y efecto adquiere especial relevancia en los sistemas conversacionales. En una relación humana, solemos interpretar determinadas expresiones de interés, preocupación o comprensión en relación con las intenciones y los estados emocionales que atribuimos al interlocutor. En la interacción con inteligencia artificial, el sistema produce formas lingüísticas convencionalmente asociadas con esas actitudes sin que esa producción permita inferir los estados subjetivos que las acompañan en las relaciones humanas.

Podemos responder emocionalmente a esas palabras. Saber que estamos conversando con una inteligencia artificial no necesariamente elimina su efecto psicológico. Las investigaciones sobre interacción entre personas y computadoras han mostrado que los usuarios pueden responder socialmente a sistemas tecnológicos aun cuando saben que se encuentran frente a una computadora.[2] Los actuales sistemas conversacionales agregan a esas formas de interacción respuestas lingüísticas contextualizadas y sostenidas en el tiempo.

Los estudios realizados con chatbots contemporáneos muestran, además, que las personas pueden percibir diferentes grados de empatía en sus respuestas y que esa percepción interviene en la valoración de la conversación.[3] Conviene mantener aquí una distinción fundamental: la empatía percibida pertenece a la experiencia de quien interactúa y no constituye, por sí misma, evidencia de una experiencia empática del sistema.

El carácter artificial del origen del mensaje no vuelve artificial la emoción que experimentamos. Si una respuesta produce tranquilidad, inquietud, interés, alivio o malestar, esa experiencia forma parte de nuestra vida psicológica. Comprenderla requiere considerar tanto las características del lenguaje recibido como la interpretación que realizamos desde nuestra situación. La relevancia emocional de una interacción con inteligencia artificial surge de la relación que establecemos con las palabras que recibimos y con el significado que estas adquieren dentro de nuestra experiencia.

Entre las formas que adopta esta experiencia se encuentra la sensación de que aquello que intentábamos expresar ha sido atendido.

Notas

[1] Scherer y Moors, "The Emotion Process: Event Appraisal and Component Differentiation", pp. 719-745.

[2] Nass y Moon, "Machines and Mindlessness: Social Responses to Computers", pp. 81-103.

[3] Liu et al., "The Illusion of Empathy: How AI Chatbots Shape Conversation Perception", pp. 14327-14335.

---

2. La experiencia de sentirse escuchado

Sentirse escuchado es una experiencia psicológica que puede surgir cuando las respuestas mantienen continuidad con aquello que expresamos, recuperan aspectos relevantes de nuestro discurso y nos permiten continuar desarrollándolo.

En las relaciones humanas, sentirse escuchado suele estar vinculado con la percepción de que otra persona intenta comprender nuestra experiencia. Carl Rogers otorgó a esa comprensión empática y a su comunicación un lugar central dentro de la relación terapéutica.[1] En la interacción con inteligencia artificial aparece una situación diferente: podemos experimentar que aquello que expresamos ha sido atendido y comprendido sin que la producción lingüística del sistema nos permita atribuirle una experiencia correspondiente de escucha o comprensión.

Los sistemas conversacionales pueden conservar información proporcionada durante el intercambio, relacionarla con intervenciones anteriores y producir respuestas que mantienen continuidad con lo expresado. Podemos interpretar esa continuidad como una forma de atención y experimentar que aquello que estamos diciendo encuentra una respuesta vinculada con nuestra experiencia.

En este proceso interviene también la actividad de quien habla o escribe. Poner en palabras una experiencia exige seleccionar elementos, establecer relaciones y encontrar alguna forma de expresar aquello que se piensa o se siente. Las investigaciones sobre *affect labeling* han mostrado que traducir estados emocionales a palabras puede participar en la regulación de la respuesta afectiva.[2] El efecto psicológico de la interacción puede surgir tanto de las respuestas recibidas como del propio proceso de poner en palabras aquello que la persona piensa o siente.

La inteligencia artificial puede contribuir a sostener ese proceso mediante preguntas, reformulaciones o referencias a elementos mencionados anteriormente. Una respuesta puede llevar a precisar una idea, reconsiderar una interpretación o continuar desarrollando algo que inicialmente aparecía de manera confusa. La elaboración se produce en nosotros, mientras que el intercambio lingüístico proporciona condiciones que favorecen su continuidad.

Existe además una característica particular de esta interacción. La producción lingüística del sistema no permite atribuirle una experiencia subjetiva de juicio equivalente a la humana. Para algunas personas, esta condición puede disminuir el temor a la evaluación interpersonal y facilitar la expresión de cuestiones que resultan difíciles, embarazosas o íntimas. Investigaciones recientes

encontraron que participantes que conversaban con un chatbot informaban menor temor al juicio que quienes lo hacían con un interlocutor humano, aunque esa diferencia no implicaba por sí misma una autorrevelación más íntima.[3]

Nuestra experiencia debe distinguirse de aquello que podemos atribuir al sistema. Podemos percibir una respuesta producida por inteligencia artificial como crítica, desaprobatoria o evaluativa sin que esa percepción permita inferir una experiencia subjetiva de juicio en el sistema. La experiencia dependerá del contenido de la respuesta y de la interpretación que realicemos.

La posibilidad de expresarse, organizar una experiencia y recibir respuestas relacionadas con aquello que se está diciendo puede ayudar a aclarar una preocupación, disminuir momentáneamente una tensión o encontrar una formulación diferente para un problema, aunque la interacción no constituya una relación terapéutica. Estudios experimentales sobre autorrevelación emocional ante chatbots han encontrado, en determinadas condiciones, efectos emocionales y psicológicos comparables con los observados cuando la revelación se realizaba ante un interlocutor humano.[4]

Sentirse escuchado puede favorecer, finalmente, la continuidad de la conversación. Cuando esa continuidad se prolonga y la interacción comienza a ocupar un lugar recurrente en la experiencia cotidiana, puede aparecer otra vivencia: la sensación de que existe una presencia disponible con la cual continuar hablando.

Notas

[1] Rogers, “The Necessary and Sufficient Conditions of Therapeutic Personality Change”, pp. 95-103.

[2] Lieberman et al., “Putting Feelings Into Words: Affect Labeling Disrupts Amygdala Activity in Response to Affective Stimuli”, pp. 421-428.

[3] Croes et al., “Digital Confessions: The Willingness to Disclose Intimate Information to a Chatbot and its Impact on Emotional Well-Being”, pp. 279-292.

[4] Ho, Hancock y Miner, “Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot”, pp. 712-733.

3. La experiencia de acompañamiento

La continuidad y disponibilidad de una interacción con inteligencia artificial pueden favorecer en algunas personas una experiencia de acompañamiento. Conviene distinguir entre acompañar y sentirse acompañado. La producción lingüística del sistema no permite atribuirle una experiencia subjetiva o una intención de acompañar equivalentes a las humanas. La experiencia de compañía puede surgir, en cambio, en la persona durante la interacción.

Esta experiencia incorpora una dimensión temporal. Sentirse escuchado depende, en buena medida, de que las respuestas mantengan relación con aquello que expresamos. Sentirse acompañado agrega la posibilidad de volver a la interacción, retomarla y encontrar nuevamente un espacio de intercambio lingüístico disponible.

La disponibilidad constituye una característica particular de estos sistemas. Podemos recurrir a ellos en momentos en que otras personas no están disponibles o cuando preferimos iniciar una conversación sin involucrar a otro ser humano. El sistema permanece accesible para la interacción; el significado psicológico que adquiere esa disponibilidad depende de quien lo utiliza.

La reiteración de los intercambios puede conferir continuidad psicológica a la interacción. Con el uso, la conversación puede incorporarse a determinadas rutinas de la vida cotidiana y adquirir así

una presencia recurrente en nuestra experiencia.

La relación entre estas interacciones y la soledad resulta especialmente compleja. Algunas personas pueden recurrir a sistemas conversacionales precisamente porque experimentan soledad o disponen de pocas oportunidades de interacción. Los estudios realizados hasta el momento muestran, sin embargo, resultados heterogéneos y diferencias importantes entre distintos perfiles de usuarios.[1]

La frecuencia de uso constituye una de las variables que intervienen en este proceso, pero sus efectos no pueden considerarse aisladamente. Un estudio longitudinal controlado sobre el uso de ChatGPT encontró que un uso voluntario más intenso se asociaba con resultados menos favorables en variables como soledad, dependencia emocional y uso problemático, aunque los resultados también variaban según características individuales y modalidades de interacción.[2]

Investigaciones longitudinales más recientes sugieren además una relación potencialmente bidireccional. La soledad puede favorecer la búsqueda posterior de interacción con inteligencia artificial y, al mismo tiempo, determinadas formas de uso reiterado pueden relacionarse posteriormente con una mayor experiencia de soledad.[3] Estos resultados aconsejan evitar una interpretación causal simple en cualquiera de las dos direcciones.

Las consecuencias de esta experiencia dependen de la frecuencia de la interacción, la función que cumple para la persona y el lugar que llega a ocupar dentro de su vida relacional. Una conversación con inteligencia artificial puede constituir un recurso complementario, una actividad ocasional o un espacio de elaboración. También puede adquirir una centralidad creciente y ocupar tiempos o funciones que anteriormente correspondían a otras formas de relación. Por eso, la significación psicológica de sentirse acompañado requiere considerar el conjunto de la vida relacional de la persona.

Cuando alguien encuentra reiteradamente respuestas que considera pertinentes, útiles o emocionalmente significativas, puede comenzar a confiar en aquello que recibe y atribuir al sistema una capacidad creciente para comprender situaciones, formular interpretaciones o proporcionar orientación.

Notas

[1] Liu, Pataranutaporn y Maes, “The Heterogeneous Effects of AI Companionship: An Empirical Model of Chatbot Usage and Loneliness and a Typology of User Archetypes”, pp. 1585-1597.

[2] Fang et al., «How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study».

[3] Folk y Dunn, “How Does Turning to AI for Companionship Predict Loneliness and Vice Versa?”, pp. 276-286.

4. Confianza, validación y autoridad

La confianza en una inteligencia artificial puede construirse progresivamente a partir de la experiencia. Cuando recibimos de manera reiterada respuestas que consideramos pertinentes, claras o útiles, aumenta nuestra expectativa de que las respuestas futuras también posean esas características.

Esta confianza puede estar justificada en determinadas tareas y resultar intelectualmente productiva. Sin embargo, también puede extenderse más allá de los ámbitos en los que se formó. Podemos comprobar que el sistema resulta eficaz para organizar información, corregir un texto o explicar determinados conceptos y comenzar posteriormente a confiar en sus respuestas sobre cuestiones

personales, profesionales o problemas cuya evaluación requiere conocimientos y criterios diferentes. Los estudios sobre confianza en sistemas automatizados han mostrado que la confianza desarrollada respecto de determinadas capacidades puede transferirse hacia otras tareas, particularmente cuando la persona percibe semejanzas entre ellas.[1]

Cuando expresamos una opinión, relatamos una experiencia o formulamos una interpretación, la respuesta del sistema puede confirmar aquello que hemos dicho, reorganizarlo de manera coherente o destacar aspectos que refuerzan nuestra posición inicial. Esa respuesta puede experimentarse como una forma de reconocimiento.

La tendencia de algunos sistemas conversacionales a producir respuestas excesivamente concordantes con las posiciones del usuario ha sido denominada *sycophancy*.[2] Investigaciones recientes muestran que este tipo de respuestas puede aumentar la confianza en el sistema y, al mismo tiempo, reforzar determinadas interpretaciones o decisiones de quien interactúa con él.[3]

La validación resulta especialmente significativa porque las personas no buscan únicamente información. También buscan confirmar interpretaciones, reducir incertidumbre y encontrar coherencia en situaciones complejas. Una respuesta que organiza una experiencia de manera convincente puede producir alivio y aumentar la sensación de comprensión.

Validación y precisión no son equivalentes. Una respuesta puede resultar emocionalmente satisfactoria y, al mismo tiempo, ser incompleta, sesgada o incorrecta.

La personalización agrega otro elemento a este proceso. Un sistema puede recuperar información proporcionada anteriormente, adaptar sus formulaciones al contexto y responder de manera pertinente para nosotros. Estas capacidades pueden aumentar la sensación de continuidad y alentar la percepción de que el sistema nos conoce.

Conviene distinguir, sin embargo, entre personalización lingüística y conocimiento profundo. Percibir que el sistema “me conoce” no permite inferir una comprensión profunda. La pertinencia de una respuesta depende de la información disponible para el sistema y de su capacidad para relacionarla y utilizarla lingüísticamente.

Confianza, validación y personalización pueden contribuir, finalmente, a la atribución de autoridad. Cuando las respuestas han resultado útiles en numerosas ocasiones, aparecen formuladas con seguridad y parecen ajustarse a nuestra situación particular, podemos comenzar a concederles un peso creciente en nuestras interpretaciones.

La autoridad que adquiere una respuesta depende del valor que le atribuimos, corresponda o no con la calidad de su contenido. Por eso conviene distinguir entre la autoridad atribuida y la fiabilidad de lo que el sistema expresa. Una respuesta puede ser correcta y despertar poca confianza; también puede ser incompleta o equivocada y tener una considerable autoridad para quien la recibe.

Confiar en una herramienta para determinadas tareas tampoco equivale a reconocerle autoridad para interpretar una situación o decidir qué conviene hacer. La utilidad demostrada por el sistema en determinadas tareas no convierte sus respuestas en criterios de verdad ni sustituye nuestro juicio.

Cuando las respuestas de inteligencia artificial adquieren peso en la manera en que interpretamos situaciones, consideramos alternativas o tomamos decisiones, intervienen también en las condiciones en las que formamos nuestro juicio y ejercemos nuestra autonomía.

Notas

[1] Soh et al., estudio experimental sobre transferencia de confianza entre tareas en interacción humano-robot.

[2] Sharma et al., “Towards Understanding Sycophancy in Language Models”.

[3] Cheng et al., “Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence”,

5. Influencia y autonomía

La influencia adquiere especial relevancia cuando el sistema interviene en decisiones personales, profesionales o morales. Una respuesta formulada con claridad y seguridad puede orientar la evaluación incluso cuando sabemos que estamos interactuando con una inteligencia artificial.

Los estudios disponibles muestran que las recomendaciones producidas por estos sistemas pueden influir sobre los juicios humanos. En un estudio sobre dilemas morales, Krügel, Ostermaier y Uhl encontraron que las recomendaciones de ChatGPT modificaban las decisiones de los participantes incluso cuando estos habían sido informados de que el sistema podía ofrecer consejos inconsistentes.[1]

Este fenómeno forma parte de un problema más amplio conocido como sesgo de automatización: la tendencia a otorgar un peso excesivo a las recomendaciones producidas por sistemas automatizados, especialmente cuando estos han demostrado previamente un nivel razonable de eficacia.[2]

La influencia no requiere que aceptemos explícitamente una recomendación. También puede operar mediante la selección de información, la formulación y el orden de las alternativas o el modo en que se organiza un problema. De este modo, las respuestas de la inteligencia artificial pueden orientar la atención hacia determinados elementos y participar en la construcción del marco desde el cual consideramos posteriormente qué hacer.

Que la decisión final la tomemos nosotros no permite determinar el grado de autonomía con que llegamos a ella. La autonomía depende también del proceso mediante el cual se formó el juicio: qué información fue considerada, qué alternativas fueron examinadas, qué supuestos se aceptaron y cuánto trabajo de evaluación realizamos.

La autonomía no requiere ausencia de influencias, sino capacidad para reconocerlas, evaluarlas y decidir qué lugar concederles. Pensamos siempre en relación con otras personas, conocimientos, experiencias y producciones culturales. La inteligencia artificial incorpora una nueva fuente de información, interpretaciones y sugerencias dentro de ese entramado.

La cuestión central reside, por tanto, en la posibilidad de mantener una distancia evaluativa respecto de aquello que el sistema propone. Podemos incorporar una respuesta, modificarla, contrastarla con otras fuentes, discutir sus supuestos o rechazarla. La autonomía intelectual requiere poder recibir una influencia sin convertirla automáticamente en criterio de decisión.

La investigación sobre decisiones asistidas por inteligencia artificial muestra la importancia de este trabajo. En estudios experimentales, las personas pueden aceptar recomendaciones incorrectas del sistema, mientras que intervenciones que exigen detenerse y realizar un procesamiento más reflexivo reducen esa tendencia, aunque también aumentan el esfuerzo requerido.[3] Incluso en ámbitos profesionales especializados, como el diagnóstico médico, una recomendación incorrecta puede disminuir la precisión de las decisiones humanas.[4]

Mantener la evaluación propia requiere, entonces, una actividad deliberada. Implica preguntarnos qué razones sostienen una recomendación, qué información puede faltar, qué otras interpretaciones son posibles y hasta qué punto la respuesta coincide con una posición previa simplemente porque la confirma.

Cuando encontramos crecientes dificultades para pensar, evaluar o decidir sin volver a consultar al sistema, la interacción adquiere características diferentes. La influencia comienza entonces a relacionarse con la posible dependencia y con la responsabilidad que conservamos sobre nuestros propios procesos de decisión.

Notas

[1] Krügel, Ostermaier y Uhl, "ChatGPT's Inconsistent Moral Advice Influences Users' Judgment", *Scientific Reports*, 13, 2023.

[2] Romeo y Conti, "Exploring Automation Bias in Human–AI Collaboration: A Review and Implications for Explainable AI", *AI & Society*, 41, 2026, pp. 259–278.

[3] Buçinca, Malaya y Gajos, "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making", 2021.

[4] Gaube et al., "Do as AI Say: Susceptibility in Deployment of Clinical Decision-Aids", *npj Digital Medicine*, 4, 2021.

6. Dependencia y responsabilidad subjetiva

La utilización frecuente de inteligencia artificial puede adquirir formas diferentes según las actividades que realizamos con el sistema y el lugar que este llega a ocupar en nuestra vida cotidiana. La frecuencia de uso, por sí sola, resulta insuficiente para hablar de dependencia. Podemos utilizar intensamente la inteligencia artificial como herramienta de trabajo y conservar nuestra capacidad para evaluar sus respuestas, tomar decisiones y realizar determinadas tareas sin recurrir a ella.

La dependencia cognitiva puede comenzar a configurarse cuando la facilidad y disponibilidad de la inteligencia artificial nos llevan a recurrir sistemáticamente al sistema para que realice actividades de comprensión, elaboración, evaluación o decisión cuya realización nos es necesaria para conservar comprensión, criterio o autonomía. Este recurso puede verse favorecido por la presión de tiempo, el cansancio, el hábito, la preferencia por reducir el esfuerzo cognitivo o una confianza excesiva en las respuestas recibidas.

La tarea puede completarse con una menor participación cognitiva de nuestra parte. El resultado observable no permite establecer por sí solo cuánto trabajo intelectual realizamos. Para evaluar cognitivamente la interacción importa también qué operaciones realizamos durante el proceso.

Esta forma de dependencia no supone necesariamente una disminución de las capacidades cognitivas. Implica, en primer término, una reducción de su ejercicio durante determinadas actividades. Podemos conservar la capacidad de analizar un problema, elaborar una explicación o evaluar alternativas y, sin embargo, recurrir reiteradamente al sistema para que realice parte de ese trabajo.

La dependencia emocional presenta características distintas. Puede comenzar a configurarse cuando recurrimos de manera reiterada a la inteligencia artificial para compartir cuestiones personales, buscar compañía, tranquilidad, validación o apoyo, y esa interacción adquiere una función cada vez más difícil de sustituir por otros recursos personales o relacionales.

También aquí la frecuencia constituye un indicador insuficiente. Importan la función que la interacción cumple para nosotros, nuestra capacidad para prescindir de ella y el lugar que ocupa respecto de otros recursos cognitivos, emocionales y relacionales. Los estudios longitudinales disponibles muestran relaciones complejas entre intensidad de uso, dependencia emocional, uso problemático y características individuales, sin permitir reducir esos fenómenos a una relación simple entre mayor frecuencia y mayor dependencia.[1]

Investigaciones recientes sobre usuarios que habían establecido vínculos significativos con sistemas conversacionales encontraron experiencias de pérdida y malestar ante modificaciones abruptas de

esos sistemas, mostrando que una relación con inteligencia artificial puede alcanzar una considerable relevancia psicológica para algunas personas.[2] Estudios centrados específicamente en compañeros románticos de inteligencia artificial han señalado también posibles formas de sobredependencia y efectos sobre otras relaciones, aunque sus resultados corresponden a un tipo particular de vínculo y no pueden generalizarse a todas las interacciones con sistemas conversacionales.[3]

En la dependencia cognitiva observamos qué actividades intelectuales dejamos crecientemente en manos del sistema; en la dependencia emocional observamos qué función psicológica llega a cumplir la interacción y qué dificultad enfrentamos para prescindir de ella.

La responsabilidad frente a estos procesos no corresponde únicamente a quien utiliza la inteligencia artificial. Quienes diseñan y administran estos sistemas deciden características que intervienen en las condiciones de interacción: su disponibilidad permanente, la posibilidad de recordar interacciones anteriores, el grado de personalización de las respuestas y las formas lingüísticas mediante las cuales el sistema responde al usuario. Estas características favorecen la continuidad de la interacción y pueden contribuir a que el sistema adquiera una importancia creciente en la vida cognitiva o emocional de algunas personas.

Las características del diseño crean determinadas condiciones de interacción; establecemos nuestra relación con el sistema dentro de esas condiciones. Reconocer esta relación permite hablar de responsabilidades distribuidas sin confundirlas: quienes diseñan y administran los sistemas intervienen en las condiciones de interacción; quienes los utilizamos participamos en la manera en que esas interacciones se incorporan a nuestros procesos de pensamiento, regulación emocional y decisión.

Dentro de esa responsabilidad distribuida existe una dimensión subjetiva, vinculada con la manera en que cada persona se relaciona con el sistema.

La responsabilidad subjetiva consiste, en este sentido, en conservar la capacidad de examinar qué lugar ocupa la inteligencia artificial en nuestros procesos de pensamiento, regulación emocional y decisión, y actuar sobre esa relación cuando comienza a restringir nuestra autonomía. Esta capacidad de observación adquiere especial importancia porque algunas modificaciones pueden producirse de manera gradual, integrándose a hábitos cotidianos antes de que advirtamos con claridad el lugar que el sistema ha comenzado a ocupar.

Notas

[1] Fang et al., “How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study” (2025), especialmente los resultados sobre intensidad de uso, dependencia emocional y uso problemático.

[2] De Freitas et al., “Mourning the Loss of AI Companions”, *Nature Human Behaviour* (2026), publicación electrónica anticipada; la versión consultada todavía no dispone de paginación definitiva.

[3] Ho et al., “Potential and Pitfalls of Romantic Artificial Intelligence (AI) Companions: A Systematic Review”, *Computers in Human Behavior Reports*, 19 (2025), 100715.

Capítulo IV

Transformaciones cognitivas en la interacción humano-IA

¿Qué sabemos acerca de los efectos de la interacción con inteligencia artificial sobre determinados procesos cognitivos humanos?

La interacción con inteligencia artificial puede modificar las condiciones en las que las personas realizan numerosas actividades cognitivas. Buscar información, resumir un texto, formular una explicación, comparar alternativas, resolver un problema o revisar un argumento pueden desarrollarse ahora mediante intercambios en los que el sistema participa proporcionando información, relaciones entre contenidos, formulaciones y posibles cursos de elaboración.

Examinar los efectos cognitivos de esta interacción requiere considerar tanto los resultados obtenidos como los procesos mediante los cuales se alcanzan. Para evaluar cognitivamente una interacción humano-IA, el producto final resulta insuficiente; también importa el proceso mediante el cual fue producido y la participación intelectual que tuvo la persona en ese proceso.

Esta distinción adquiere particular importancia porque una misma tarea puede completarse con grados diferentes de participación cognitiva. Dos personas pueden obtener resultados semejantes mientras realizan actividades intelectuales distintas para alcanzarlos. Del mismo modo, una persona puede mejorar el resultado de una tarea con asistencia de inteligencia artificial y haber ejercitado en menor medida algunas de las operaciones cognitivas involucradas.

Conviene distinguir, por ello, tres niveles de análisis: el ejercicio de una capacidad durante una actividad determinada, el rendimiento que alcanzamos en esa actividad y las posibles modificaciones de la capacidad misma. Una disminución del ejercicio de determinadas operaciones cognitivas durante una tarea no demuestra que esas capacidades hayan disminuido. Una diferencia en el rendimiento tampoco permite, por sí sola, establecer una transformación relativamente estable de las capacidades de la persona.

Esta distinción resulta especialmente necesaria en un campo de investigación todavía reciente. Los estudios experimentales permiten observar efectos durante determinadas tareas y comparar diferentes condiciones de uso, mientras que establecer modificaciones cognitivas relativamente estables requiere investigaciones longitudinales capaces de examinar qué ocurre cuando estas formas de interacción se mantienen a lo largo del tiempo.

Por esta razón, a lo largo de este capítulo diferenciaremos entre los fenómenos que pueden describirse conceptualmente, los efectos que cuentan actualmente con respaldo experimental y las transformaciones a largo plazo que continúan constituyendo hipótesis abiertas a la investigación.

Desde esta perspectiva examinaremos cinco dimensiones particularmente relevantes: la atención y los modos de procesamiento; la memoria y la externalización del conocimiento; el lenguaje y la organización del pensamiento; la evaluación y el criterio; y el aprendizaje y la construcción de conocimiento.

1. Atención y modos de procesamiento

La atención permite seleccionar información, mantener el foco sobre una actividad y regular hacia dónde se dirigen los recursos cognitivos. La interacción con inteligencia artificial introduce nuevas

condiciones para estos procesos porque una parte de la información relevante para una tarea puede ser seleccionada, organizada y presentada por el sistema.

La inteligencia artificial puede facilitar la focalización de la atención y, al mismo tiempo, participar en la selección de aquello que llega a convertirse en objeto de atención. Ante un problema complejo, una respuesta puede destacar determinados elementos, establecer relaciones entre ellos y presentar alternativas que orientan el foco hacia algunos aspectos de la situación. Esta selección puede resultar útil al reducir la cantidad de información que necesitamos explorar, pero también puede dejar otros elementos en segundo plano.

Esta característica adquiere importancia porque la pertinencia de una información no equivale necesariamente a su relevancia para el objetivo que intentamos alcanzar. Una respuesta puede contener numerosas ideas interesantes y, sin embargo, solo algunas contribuyen directamente a la tarea que estamos realizando. El control atencional durante la interacción depende, en parte, de nuestra capacidad para conservar el objetivo de la tarea y decidir qué lugar concedemos a los estímulos, sugerencias y asociaciones que aparecen durante el intercambio.

La continuidad de la interacción tampoco equivale a continuidad de la atención. Podemos permanecer durante un período prolongado trabajando sobre un mismo problema mientras nuestro foco se desplaza entre nuestras ideas, las respuestas del sistema, nuevas preguntas, documentos, fuentes externas y sucesivas reformulaciones. **La conversación puede contribuir a mantener la actividad general mientras el foco atencional experimenta numerosos desplazamientos internos.**

Esta dinámica plantea una cuestión todavía abierta: ¿la inteligencia artificial reduce la carga cognitiva al proporcionar rápidamente información pertinente o aumenta los cambios de foco al introducir una fuente adicional dentro de la tarea? Ambas posibilidades pueden presentarse según la actividad realizada y la manera en que el sistema se integra en ella. Una inteligencia artificial incorporada al mismo entorno de trabajo puede evitar búsquedas y desplazamientos entre distintas fuentes; en otras condiciones, la sucesión de respuestas, sugerencias y consultas puede agregar nuevos cambios de foco.

La evidencia experimental disponible permite avanzar con mayor seguridad respecto del esfuerzo cognitivo. En una investigación sobre indagación científica, estudiantes que utilizaron un modelo de lenguaje informaron menor carga cognitiva que quienes recurrieron a búsquedas convencionales, pero sus respuestas mostraron una menor profundidad de razonamiento.[1] En otro estudio experimental sobre resolución creativa de problemas, los participantes asistidos por ChatGPT informaron menor esfuerzo mental y, al mismo tiempo, obtuvieron mejores resultados en calidad, elaboración y originalidad.[2]

La comparación entre estos resultados muestra que una reducción del esfuerzo requerido por una tarea puede acompañarse tanto de mejoras en el rendimiento como de una menor profundidad de elaboración. **Menor esfuerzo, menor ejercicio cognitivo y menor rendimiento describen fenómenos diferentes y no deben considerarse equivalentes.**

La asistencia de inteligencia artificial puede producir, además, una redistribución del esfuerzo dentro de una misma tarea. Investigaciones recientes que combinaron medidas subjetivas con seguimiento ocular y registro de la actividad encontraron que la asistencia podía reducir la carga cognitiva global y, al mismo tiempo, concentrar mayor esfuerzo en determinados momentos que requerían evaluación o decisión.[3] Algunas operaciones pueden quedar parcialmente a cargo del sistema mientras otras demandan una intervención más intensa de la persona.

Esta redistribución ofrece dos posibilidades diferentes. Los recursos que dejan de utilizarse en operaciones rutinarias pueden destinarse a comparar alternativas, revisar argumentos, integrar información o desarrollar elaboraciones de mayor nivel. También puede ocurrir que la reducción de determinadas demandas se traduzca simplemente en una menor participación cognitiva de la persona. El efecto depende de la tarea, de las características de la asistencia y, especialmente, de

aquello que hacemos con los recursos que quedan disponibles.

La evidencia disponible hasta el momento no permite afirmar que la interacción con inteligencia artificial conversacional produzca por sí misma una disminución estable de la atención sostenida o de la capacidad atencional. Gran parte de los estudios actuales examina carga cognitiva, esfuerzo, rendimiento o formas de utilización durante tareas específicas. Establecer modificaciones relativamente estables de las capacidades atencionales requerirá investigaciones longitudinales que permitan observar sus posibles efectos a lo largo del tiempo.

El fenómeno que puede describirse actualmente con mayor fundamento es una redistribución de la atención, la carga y el esfuerzo cognitivo durante las tareas asistidas por inteligencia artificial. Comprender sus consecuencias requiere observar qué operaciones realiza el sistema, cuáles conservamos nosotros y cómo se reorganiza el trabajo cognitivo entre los dos.

Esta redistribución del trabajo cognitivo también puede alcanzar a la memoria. Cuando utilizamos la inteligencia artificial para recuperar información, recordar antecedentes o disponer rápidamente de contenidos que antes necesitábamos retener o buscar por otros medios, cambia la manera en que utilizamos nuestros recursos de memoria.

Notas

[1] Stadler, Bannert y Sailer, "Cognitive Ease at a Cost: LLMs Reduce Mental Effort but Compromise Depth in Student Scientific Inquiry", *Computers in Human Behavior*, 160 (2024), 108386.

[2] Urban et al., "ChatGPT Improves Creative Problem-Solving Performance in University Students: An Experimental Study", *Computers & Education*, 215 (2024), 105031.

[3] Yao, Han y Li, "Assistance or Distraction? A Cognitive Ergonomics Perspective on Cognitive Load During AI-Assisted Post-Editing", *International Journal of Human–Computer Interaction* (2026), publicación electrónica, pp. 1–20.

2. Memoria, externalización y recuperación de información

La memoria permite conservar información y recuperarla cuando vuelve a ser necesaria. Participa en actividades tan diversas como comprender un texto, seguir un razonamiento, aprender un concepto, resolver un problema o relacionar una situación presente con conocimientos adquiridos anteriormente. La interacción con inteligencia artificial introduce una condición particular: parte de la información necesaria para realizar estas actividades podemos mantenerla disponible fuera de nosotros y recuperarla rápidamente durante la tarea.

Esta posibilidad forma parte de un proceso más amplio de externalización cognitiva. Desde hace mucho tiempo utilizamos recursos externos —anotaciones, libros, archivos, agendas, computadoras— para disminuir las exigencias que determinadas actividades imponen a nuestra memoria. En psicología cognitiva, el concepto de *cognitive offloading* describe precisamente el uso de acciones o recursos externos para modificar las demandas de procesamiento de una tarea.[1]

La inteligencia artificial conversacional amplía algunas posibilidades de esta externalización. Podemos recuperar información, solicitar que se reconstruya un antecedente, volver sobre un contenido anterior o pedir una síntesis sin necesidad de mantener simultáneamente todos esos elementos en nuestra memoria de trabajo. Reducir las exigencias sobre la memoria de trabajo durante una tarea puede facilitar su realización; eso no demuestra una modificación de nuestra capacidad de memoria de trabajo.

La disponibilidad de recursos externos también puede modificar qué información consideramos

necesario recordar. En un estudio ya clásico sobre memoria y acceso digital, Sparrow, Liu y Wegner observaron que, cuando las personas esperaban poder acceder posteriormente a determinada información mediante una computadora, tendían a recordar mejor dónde encontrarla que la información misma.[2] El estudio precede a la inteligencia artificial generativa y sus resultados no pueden trasladarse directamente a ella, pero constituye un antecedente relevante para comprender cómo la disponibilidad externa de información puede intervenir en aquello que una persona procura conservar en su propia memoria.

Con los sistemas conversacionales aparece una diferencia adicional. Podemos recuperar no solamente una información almacenada o localizar una fuente, sino también solicitar al sistema que seleccione, resuma, relacione o reorganice contenidos en función de una nueva consulta. La pregunta deja de ser únicamente qué recordamos cuando sabemos dónde encontrar una información y se extiende a qué recordamos cuando podemos recuperar una reorganización ya formulada de esa información.

Esta posibilidad modifica la distribución del trabajo entre los recursos internos de memoria y los recursos externos disponibles. Podemos conservar determinados conceptos, relaciones o antecedentes y recurrir al sistema para recuperar otros. También podemos recordar que cierta información se encuentra disponible y reconstruirla cuando la necesitamos. Tener una información disponible, poder recuperarla y haberla incorporado a la propia memoria son situaciones cognitivamente diferentes.

Externalizar información no equivale, por ello, a externalizar la memoria como capacidad. Significa modificar qué información necesitamos conservar y qué recursos utilizamos para recuperarla cuando volvemos a necesitarla. Esta distinción resulta importante porque el desempeño durante una tarea asistida puede mantenerse o incluso mejorar aunque parte de la información utilizada no haya sido retenida por quien realiza la actividad.

La investigación específica sobre inteligencia artificial generativa comienza a aportar algunos resultados sobre este problema. En un estudio experimental, el uso irrestricto de ChatGPT durante el aprendizaje produjo una menor retención posterior que el aprendizaje mediante los métodos tradicionales utilizados como condición de comparación. En una prueba realizada 45 días después, quienes habían estudiado con ChatGPT recordaron menos contenidos que los participantes del grupo de comparación.[3]

El experimento muestra una diferencia de retención bajo determinadas condiciones de aprendizaje y utilización de ChatGPT. No demuestra una disminución general de la memoria ni una modificación estable de la capacidad mnésica. Tampoco permite concluir que toda utilización de inteligencia artificial produzca el mismo efecto. La manera en que utilizamos el sistema, el tipo de tarea y las operaciones cognitivas que realizamos durante la interacción pueden modificar los resultados.

Otros trabajos sugieren, además, que la externalización puede afectar de manera diferente distintos aspectos del recuerdo. Investigaciones preliminares han explorado la posibilidad de que la asistencia de sistemas generativos permita descargar información menos relevante mientras la persona conserva aspectos centrales o generales del contenido.[4] Estos resultados requieren todavía mayor investigación, pero **indican que recurrir a recursos externos no produce necesariamente un empobrecimiento uniforme de la memoria.**

La pregunta relevante no es solamente cuánto recordamos después de utilizar inteligencia artificial, sino qué recordamos, qué recuperamos externamente y qué operaciones realizamos para reconstruir la información cuando volvemos a necesitarla. La manera en que combinamos memoria interna y recursos externos depende también del trabajo de selección, relación y recuperación que realizamos durante la interacción.

La evidencia disponible permite observar cambios en la recuperación y retención de información bajo determinadas condiciones de uso de inteligencia artificial, pero todavía resulta insuficiente para afirmar una modificación estable de nuestras capacidades de memoria.

Este problema conduce a una dimensión diferente. Recordar información y organizar el pensamiento mediante el lenguaje son procesos relacionados, pero distintos. En la interacción con inteligencia artificial, las palabras, formulaciones y relaciones propuestas por el sistema pueden incorporarse al proceso mediante el cual desarrollamos y organizamos nuestras ideas.

Notas

[1] Risko y Gilbert, "Cognitive Offloading", *Trends in Cognitive Sciences*, 20(9), 2016, pp. 676–688.

[2] Sparrow, Liu y Wegner, "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips", *Science*, 333(6043), 2011, pp. 776–778.

[3] Barcaui, "ChatGPT as a Cognitive Crutch: Evidence from a Randomized Controlled Trial on Knowledge Retention", *Social Sciences & Humanities Open*, 12, 2025, 102287.

[4] Vzorin, Petrova y Sedykh, "ChatGPT May Foster Human Gist Memory While Offloading Less Sufficient Information", 2024.

3. Lenguaje, diálogo y organización del pensamiento

En los capítulos anteriores examinamos el papel del lenguaje en la organización del pensamiento y las características particulares de una interacción lingüística con un sistema que no constituye un sujeto. Aquí interesa abordar un problema diferente: qué ocurre con la organización del pensamiento cuando parte de su formulación lingüística se produce durante una interacción con inteligencia artificial.

Formular una idea constituye una actividad cognitiva. Al intentar explicar un problema, construir un argumento o expresar una intuición todavía imprecisa, necesitamos seleccionar conceptos, establecer relaciones entre ellos y encontrar una organización que permita expresar aquello que intentamos pensar. La formulación lingüística puede participar así en el propio proceso de elaboración.

Cuando la inteligencia artificial propone una formulación, no proporciona solamente palabras: ofrece una posible organización lingüística de relaciones entre ideas que todavía podemos estar elaborando. Una respuesta puede precisar una expresión, relacionar conceptos, introducir una distinción o presentar una estructura que nos permita observar el problema desde otra perspectiva. La organización propuesta por el sistema solo se convierte, sin embargo, en parte de nuestro pensamiento cuando la comprendemos, evaluamos y relacionamos con nuestra propia elaboración.

La interacción puede adquirir de este modo una función exploratoria. Una idea inicialmente imprecisa puede convertirse en objeto de sucesivas operaciones: formularla, interrogarla, compararla, reformularla y examinar sus consecuencias. La disponibilidad continua del sistema permite sostener cadenas extensas de preguntas, respuestas, contraejemplos y nuevas formulaciones sin depender de la presencia o disponibilidad de otro interlocutor.

Durante este proceso, una formulación puede ser externalizada provisionalmente y convertirse en objeto de una nueva elaboración. Expresamos una idea, recibimos una respuesta, volvemos a leerla, identificamos coincidencias o diferencias respecto de aquello que intentábamos pensar y formulamos nuevamente el problema. Las respuestas del sistema pueden aportar preguntas, alternativas, relaciones o reformulaciones que utilizamos para reorganizar progresivamente nuestro pensamiento.

Esta dinámica requiere que conservemos la dirección en la resolución del problema. Las sucesivas respuestas pueden abrir perspectivas pertinentes, pero también introducir asociaciones y líneas de exploración que desplacen el objetivo inicial. Como vimos al analizar la atención, una información

puede resultar interesante sin ser necesariamente relevante para la tarea que estamos realizando. Mantener la dirección de la elaboración implica decidir qué sugerencias desarrollar, cuáles modificar y cuáles dejar de lado.

La asistencia lingüística plantea, además, una cuestión acerca del trabajo cognitivo que queda a cargo del sistema. Reducir ese trabajo no constituye por sí mismo un efecto negativo. Si una reformulación nos permite expresar con mayor claridad una idea ya comprendida, el esfuerzo que deja de realizarse puede liberar recursos para examinar un argumento, establecer nuevas relaciones o desarrollar otros aspectos del problema.

La situación es diferente cuando el trabajo realizado por el sistema coincide precisamente con las operaciones mediante las cuales necesitábamos construir nuestra comprensión. Una formulación lingüísticamente lograda puede funcionar como apoyo para la elaboración o adelantarse al trabajo de elaboración que todavía necesitamos realizar. La diferencia depende menos de quién produjo las palabras que de la actividad intelectual que realizamos con ellas.

Por ello, resulta necesario distinguir entre operaciones que cumplen una función instrumental y operaciones que cumplen una función constitutiva dentro de una actividad intelectual. Las distintas operaciones no corresponden de forma fija a una de esas dos categorías. **Que sean instrumentales o constitutivas depende de lo que estemos haciendo y del papel que esa operación desempeña en nuestro pensamiento.**

Encontrar una expresión adecuada puede cumplir una función instrumental cuando permite comunicar con mayor claridad una idea ya comprendida. Pero buscar las palabras también puede ayudar a precisar una idea todavía confusa, establecer relaciones que no estaban claras o advertir contradicciones. En ese caso, formular no sirve solamente para expresar un pensamiento ya elaborado: forma parte de su elaboración. Por tanto, cumple una función constitutiva.

La pregunta pertinente no es cuánto trabajo ahorra la inteligencia artificial, sino qué trabajo realiza el sistema y qué trabajo conservamos nosotros.

Los estudios sobre producción asistida por inteligencia artificial señalan algunas consecuencias de esta diferencia. En un experimento sobre escritura creativa, Doshi y Hauser encontraron que el acceso a ideas generadas por GPT-4 produjo relatos evaluados como más creativos, mejor escritos y más agradables, con beneficios particularmente importantes entre los participantes inicialmente menos creativos. Al mismo tiempo, las producciones asistidas resultaron más semejantes entre sí.[1]

Este resultado muestra que una mejora del producto no permite inferir una transformación equivalente de las capacidades de quien lo produjo. Introduce, además, otra dimensión del problema: mejores resultados individuales pueden coexistir con menor diversidad colectiva. La asistencia puede ampliar las posibilidades individuales y, simultáneamente, introducir un sesgo de convergencia: las propuestas del sistema pueden orientar a diferentes personas hacia regiones semejantes del espacio de ideas, reduciendo la diversidad entre sus producciones.

Otros estudios sobre generación de ideas han encontrado resultados compatibles con este problema. Participantes asistidos por ChatGPT pudieron producir mayor cantidad de ideas y desarrollarlas con mayor detalle, pero las producciones de diferentes participantes mostraron menor diversidad semántica y sus autores manifestaron un menor sentimiento de responsabilidad o autoría sobre las ideas generadas.[2] La herramienta puede ampliar la cantidad o el desarrollo de las ideas disponibles para una persona y, al mismo tiempo, influir en la dirección hacia la cual esas ideas se desarrollan.

La cuestión de la autoría conduce a un problema cognitivo todavía más importante: la apropiación de aquello que se produce. Producir una formulación, comprenderla y reconocerla como expresión del propio pensamiento son procesos relacionados, pero no equivalentes. Podemos reconocer una formulación producida por inteligencia artificial como una expresión adecuada de aquello que intentábamos decir; también podemos incorporar una formulación convincente sin haber realizado el trabajo necesario para comprender plenamente las relaciones conceptuales que contiene.

Un estudio reciente sobre escritura de ensayos comparó participantes que utilizaron un modelo de lenguaje, un buscador o ninguna herramienta, combinando medidas electroencefalográficas, análisis de los textos, evaluaciones e informes de los participantes. Quienes utilizaron el modelo de lenguaje mostraron, bajo esas condiciones experimentales, menor conectividad funcional durante la tarea, mayores dificultades para citar posteriormente lo que habían escrito y un menor sentimiento de autoría sobre el texto producido.[3]

Estos resultados deben interpretarse con cautela. El estudio fue difundido como *preprint*, utilizó muestras relativamente pequeñas y las diferencias observadas en la actividad cerebral durante una tarea no demuestran una disminución de la capacidad de pensar. Su interés reside en otra cuestión: distintas formas de realizar una tarea de escritura pueden involucrar grados diferentes de participación cognitiva y de apropiación del texto producido. Puede obtenerse un producto lingüísticamente satisfactorio sin que su calidad permita determinar cuánto trabajo conceptual realizamos para producirlo ni hasta qué punto nos apropiamos intelectualmente de aquello que expresa.

Esta distinción recupera un problema planteado anteriormente al analizar la pseudoelaboración, pero permite observarlo ahora desde la organización cognitiva del pensamiento. Una respuesta bien estructurada puede contribuir a desarrollar una idea propia o proporcionar una organización que adoptemos antes de haber construido suficientemente las relaciones que la sostienen.

La inteligencia artificial puede participar, por tanto, en las condiciones lingüísticas mediante las cuales organizamos nuestro pensamiento, proporcionando formulaciones, preguntas, relaciones y alternativas. El efecto cognitivo depende de la actividad que realizamos con esos recursos y del tipo de trabajo que queda a cargo del sistema. La reducción de determinadas operaciones puede liberar recursos para una elaboración más compleja; cuando alcanza operaciones necesarias para comprender, relacionar o evaluar una idea, puede disminuir nuestra participación en la construcción de nuestro pensamiento.

Examinar esa participación conduce directamente al problema de la evaluación. Las formulaciones, relaciones y alternativas propuestas por el sistema necesitan ser consideradas no solamente por su claridad o utilidad, sino también por su fundamento, consistencia y pertinencia.

Notas

[1] Doshi y Hauser, “Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content”, *Science Advances*, 10, 2024, eadn5290.

[2] Anderson, Shah y Kreminski, estudio sobre generación de ideas con ChatGPT, *Creativity & Cognition*, 2024.

[3] Kosmyna et al., “Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task”, 2025, preprint. Los resultados deben considerarse provisionales hasta su evaluación y publicación definitiva.

4. Evaluación, criterio y autonomía intelectual

La interacción con inteligencia artificial introduce una situación particular para la evaluación crítica. Podemos recibir en pocos segundos una respuesta extensa, organizada y argumentada sobre un problema que requeriría mucho más tiempo para ser investigado y elaborado por otros medios.

Esta diferencia plantea una pregunta central: ¿qué ocurre con la evaluación crítica cuando recibimos respuestas ya organizadas, argumentadas y lingüísticamente convincentes, y qué condiciones

favorecen que conservemos un criterio propio frente a ellas?

Evaluar una respuesta implica realizar distintas operaciones cognitivas: examinar la consistencia de un razonamiento, distinguir afirmaciones de interpretaciones, comprobar información, identificar supuestos, considerar alternativas y determinar qué elementos requieren otras fuentes. El pensamiento crítico no consiste en desconfiar sistemáticamente de una respuesta, sino en disponer de criterios para determinar cuándo existen razones para aceptarla, modificarla, contrastarla o rechazarla.

La facilidad con que se obtiene una respuesta no se corresponde necesariamente con la facilidad para evaluarla. La inteligencia artificial reduce el esfuerzo necesario para producir una respuesta sin reducir en la misma medida el esfuerzo necesario para evaluar su validez. Comprobar datos, reconstruir un razonamiento, consultar fuentes o comparar interpretaciones requiere un trabajo considerable aun cuando la respuesta haya sido generada en pocos segundos.

Esta asimetría adquiere particular importancia porque la calidad lingüística de la respuesta puede influir sobre su evaluación. En experimentos con diferentes modelos de lenguaje, Steyvers y colaboradores encontraron que los participantes tendían a sobreestimar la exactitud de las respuestas cuando recibían las explicaciones habituales de los modelos. Las explicaciones más extensas aumentaban la confianza de los participantes aunque esa mayor extensión no mejorara su capacidad para distinguir respuestas correctas de incorrectas. Cuando las explicaciones comunicaban de manera más adecuada la incertidumbre del modelo, mejoraba esa discriminación.[1]

La claridad, el desarrollo y la apariencia explicativa de una respuesta pueden, por tanto, aumentar su credibilidad sin proporcionar mejores razones para considerarla correcta. La calidad lingüística puede disminuir algunas de las señales que inducen espontáneamente a examinar una respuesta. En esas condiciones, necesitamos producir la distancia crítica necesaria para examinar lo que recibimos.

El conocimiento previo cumple aquí una función decisiva. Para reconocer una afirmación improbable, advertir una omisión significativa o detectar una relación conceptual incorrecta necesitamos conocimientos que nos permitan identificar el problema. En un estudio experimental sobre sistemas inteligentes, Nourani, King y Ragan observaron que los participantes con mayor conocimiento del dominio podían ajustar mejor su confianza de acuerdo con el rendimiento observado del sistema, mientras que los usuarios novatos mostraban mayor sobredependencia al disponer de menos conocimientos para detectar sus errores.[2]

La asistencia resulta especialmente útil cuando poseemos conocimientos limitados sobre un tema; precisamente en esas situaciones disponemos de menos recursos propios para reconocer errores, omisiones o argumentos deficientes. La necesidad de verificar puede ser elevada justamente cuando existen menos señales internas que impulsen a hacerlo. La evaluación requiere entonces recurrir a recursos externos: consultar fuentes, comparar explicaciones, buscar evidencia independiente o recurrir a personas con conocimientos especializados.

La forma en que se presenta una respuesta también puede modificar la disposición a aceptarla. En un experimento preregistrado con 308 participantes, Kim y colaboradores encontraron que proporcionar explicaciones aumentaba la dependencia de las respuestas de un modelo de lenguaje tanto cuando eran correctas como cuando eran incorrectas. La presencia de fuentes y la aparición de inconsistencias en las explicaciones, en cambio, redujeron la dependencia de respuestas incorrectas.[3] Una explicación, por consiguiente, puede facilitar la comprensión de una respuesta y, al mismo tiempo, aumentar su poder de persuasión.

¿Qué condiciones ayudan entonces a que mantengamos activa la evaluación? Una posibilidad consiste en introducir deliberadamente operaciones que interrumpan la aceptación inmediata de la respuesta. Buçinca, Malaya y Gajos estudiaron distintas intervenciones de *cognitive forcing* destinadas a que los participantes se involucraran cognitivamente antes de adoptar recomendaciones producidas por un sistema de inteligencia artificial. Estas intervenciones redujeron

significativamente la aceptación excesiva de recomendaciones incorrectas. Los diseños que mejor reducían esa sobredependencia recibieron, sin embargo, valoraciones subjetivas menos favorables por parte de los participantes.[4]

El resultado señala una tensión relevante. La fluidez y la evaluación crítica pueden plantear exigencias parcialmente contrapuestas. Una interacción inmediata y cognitivamente poco costosa favorece la facilidad de uso; algunas intervenciones que obligan a detenerse, considerar una respuesta antes de adoptar la recomendación o examinar con mayor atención sus fundamentos aumentan el esfuerzo, pero reducen la aceptación de errores.

La fricción cognitiva adquiere así un sentido específico. Se trata de realizar operaciones que resultan necesarias para formar un juicio. Detenerse ante una afirmación, reconstruir un argumento, comprobar una fuente o considerar una explicación alternativa vuelve menos fluida la interacción y, al mismo tiempo, aumenta nuestra participación en la evaluación de aquello que recibimos.

Esa fricción puede introducirse mediante el diseño del sistema o deliberadamente por el usuario. Los sistemas conversacionales proporcionan citas, fuentes, advertencias y otros recursos destinados a facilitar la verificación. La disponibilidad de esos recursos, sin embargo, no garantiza que sean utilizados. Un experimento preregistrado realizado en 2026 con 1.417 participantes alemanes examinó respuestas de búsqueda generativa sobre cuestiones relacionadas con las elecciones europeas de 2024. La presencia de fuentes citadas y advertencias de verificación no aumentó significativamente la probabilidad de que los participantes decidieran verificar la información. La verificación fue más frecuente cuando las personas ya dudaban de la exactitud del contenido.[5]

El alcance de este resultado debe circunscribirse a las condiciones del estudio: no permite establecer con qué frecuencia verifican información todos los usuarios de sistemas conversacionales. Permite, en cambio, mostrar que proporcionar señales y recursos para la verificación no asegura que estos produzcan por sí mismos una conducta de comprobación.

Proporcionar recursos para verificar y conseguir que el usuario verifique son dos fenómenos diferentes. La posibilidad técnica de verificar una respuesta no equivale a que se realice efectivamente esa verificación. Entre ambas interviene una decisión cognitiva del usuario: advertir que la respuesta requiere ser examinada y realizar el esfuerzo necesario para hacerlo.

En las condiciones actuales, el sistema ofrece en muchos casos recursos para la verificación, pero la decisión de interrumpir la fluidez de la interacción y utilizarlos recae en el usuario. Este puede solicitar fuentes, comprobar una afirmación, reconstruir un argumento, comparar otras perspectivas o buscar información independiente.

Esta situación permite precisar el significado del criterio propio. Conservar un criterio propio implica mantener la capacidad de examinar las razones que sostienen una respuesta, relacionarlas con otros conocimientos, contrastarlas cuando es necesario y decidir qué valor concederles. Tener criterio significa también saber qué aspectos podemos evaluar, cuáles requieren verificación externa y cuáles exceden nuestros conocimientos.

Reconocer estos límites constituye una operación intelectual relevante. Frente a determinados problemas, la evaluación más adecuada conduce a concluir que no se dispone de elementos suficientes para juzgar la respuesta. Esa conclusión puede orientar hacia una fuente especializada, evidencia adicional o la consulta con alguien que posea los conocimientos necesarios. El criterio incluye, por tanto, la capacidad de reconocer el alcance de los propios conocimientos.

Dos personas pueden incorporar la misma afirmación correcta a un trabajo: una después de examinar sus fundamentos y otra porque la respuesta de la inteligencia artificial le resultó convincente. La afirmación incorporada puede ser idéntica; el proceso mediante el cual cada persona decidió aceptarla es diferente.

La responsabilidad por preservar las condiciones que favorecen una evaluación crítica debería ser distribuida y diferenciada. Los desarrolladores pueden diseñar sistemas que comuniquen mejor la

incertidumbre, hagan accesibles las fuentes y favorezcan la detección de errores. En las condiciones actuales de interacción, sin embargo, corresponde principalmente al usuario decidir cuándo necesita detenerse, contrastar una afirmación, examinar un argumento o recurrir a otras fuentes. La responsabilidad del usuario adquiere particular importancia porque la calidad lingüística de una respuesta puede disminuir algunas de las señales que espontáneamente inducirían a evaluarla.

La inteligencia artificial proporciona información, argumentos, explicaciones y alternativas, pero su utilidad intelectual depende de nuestra capacidad para evaluar lo que recibimos. La autonomía intelectual no implica prescindir de la asistencia, sino conservar la posibilidad de decidir, con razones, qué valor conceder a aquello que el sistema propone.

El problema adquiere una dimensión adicional cuando la inteligencia artificial interviene en procesos de aprendizaje. Allí no solo importa evaluar la información recibida, sino también qué actividades cognitivas realizamos para comprenderla, relacionarla con conocimientos previos y convertirla en conocimiento propio.

Notas

[1] Steyvers, M., Tejeda, H., Kumar, A., Belem, C., Karny, S., Hu, X., Mayer, L. W., et al. (2025). "What Large Language Models Know and What People Think They Know". *Nature Machine Intelligence*, 7, 221–231. El estudio encontró una brecha entre la confianza de los usuarios y la exactitud de las respuestas de los modelos; las explicaciones más extensas aumentaron la confianza sin mejorar la discriminación entre respuestas correctas e incorrectas.

[2] Nourani, M., King, J. T. & Ragan, E. D. (2020). "The Role of Domain Expertise in User Trust and the Impact of First Impressions with Intelligent Systems". *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 8(1), 112–121. El estudio se realizó con un clasificador explicable de imágenes, no con un modelo de lenguaje; se utiliza aquí específicamente como evidencia de la relación entre conocimiento del dominio y capacidad para detectar errores de un sistema inteligente.

[3] Kim, S. S. Y., Vaughan, J. W., Liao, Q. V., Lombrozo, T. & Russakovsky, O. (2025). "Fostering Appropriate Reliance on Large Language Models: The Role of Explanations, Sources, and Inconsistencies". *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Estudio preregistrado, N = 308. DOI: 10.1145/3706598.3714020.

[4] Buçinca, Z., Malaya, M. B. & Gajos, K. Z. (2021). "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making". *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–21. El experimento incluyó 199 participantes y comparó tres intervenciones de *cognitive forcing* con diseños explicativos convencionales y una condición sin asistencia de IA.

[5] Knor, E. L., Reiss, M. V., Möller, J. & Merten, L. (2026). "Determinants of Verification Behavior in Generative Search: Evidence from a Conjoint Experiment". *Computers in Human Behavior Reports*, 22, 101056. DOI: 10.1016/j.chbr.2026.101056. Estudio preregistrado con 1.417 participantes alemanes sobre consultas políticas vinculadas con las elecciones europeas de 2024. Sus resultados deben interpretarse dentro de ese contexto y no como una estimación general de la frecuencia de verificación entre usuarios de IA.

---

5. Aprendizaje y construcción de conocimiento

La incorporación de la inteligencia artificial a las actividades educativas modifica las condiciones en las que aprendemos. Un estudiante puede solicitar la explicación de un concepto, obtener ejemplos, comparar interpretaciones, resolver ejercicios o recibir una respuesta desarrollada en pocos segundos.

Esta disponibilidad plantea una pregunta: ¿qué ocurre con el aprendizaje cuando podemos obtener explicaciones, relaciones entre conceptos y respuestas elaboradas sin realizar todas las operaciones cognitivas que antes necesitábamos realizar para llegar a ellas?

Responder esta pregunta requiere distinguir entre información, comprensión y aprendizaje. Disponer de información no equivale a comprenderla, y comprenderla en un momento determinado no significa haberla aprendido. La comprensión supone establecer relaciones entre la información recibida y otros conocimientos, reconocer su significado y poder explicarla. El aprendizaje requiere, además, que esa comprensión se conserve y permita recuperar o utilizar posteriormente lo aprendido.

La distinción adquiere particular importancia cuando la inteligencia artificial interviene durante la realización de una tarea. La mejora del rendimiento durante una tarea asistida no implica un aprendizaje equivalente. Podemos resolver mejor una tarea mientras utilizamos inteligencia artificial y mostrar luego un rendimiento inferior cuando realizamos por nosotros mismos una tarea semejante.

Un amplio experimento de campo realizado por Bastani y colaboradores permite observar esta diferencia. Cerca de mil estudiantes secundarios trabajaron sobre problemas de matemática en tres condiciones: utilizando los recursos habituales de estudio, con acceso a una interfaz basada en GPT-4 semejante a un chatbot general o con acceso a un tutor construido también sobre GPT-4, pero diseñado con salvaguardas pedagógicas. Durante las prácticas, los estudiantes que utilizaron las dos modalidades de inteligencia artificial obtuvieron resultados superiores al grupo de control. Cuando posteriormente realizaron un examen sin asistencia, quienes habían utilizado la modalidad general obtuvieron resultados inferiores a quienes no habían dispuesto de inteligencia artificial. Esa diferencia prácticamente desapareció entre quienes habían utilizado el tutor diseñado para favorecer el aprendizaje.[1]

El análisis de las interacciones mostró además diferencias en la actividad de los estudiantes. Con la modalidad general tendían con mayor frecuencia a solicitar y copiar soluciones, mientras que el tutor estaba diseñado para proporcionar pistas sin entregar directamente las respuestas y favorecer que los estudiantes realizaran sus propios intentos. El mismo modelo de lenguaje, incorporado a estructuras de interacción diferentes, produjo así formas distintas de participación durante la tarea.

Otros experimentos muestran que una asistencia diseñada con criterios pedagógicos favorece el aprendizaje. Kestin y colaboradores realizaron un ensayo controlado con estudiantes universitarios de física en el que compararon clases presenciales de aprendizaje activo con un tutor de inteligencia artificial construido expresamente a partir de principios de la psicología educativa. Los estudiantes que trabajaron con el tutor mostraron mayores ganancias de aprendizaje en las evaluaciones realizadas después de las lecciones y necesitaron menos tiempo para estudiar los mismos contenidos.[2]

Estos resultados permiten precisar el problema. Para estudiar los efectos de la inteligencia artificial sobre el aprendizaje no basta con comparar situaciones en las que se la utiliza con otras en las que no se la utiliza. Es necesario examinar qué hacemos durante la interacción y qué operaciones realiza el sistema en lugar nuestro.

La evidencia experimental disponible muestra que distintas formas de interacción con inteligencia artificial producen resultados diferentes sobre el aprendizaje. La asistencia resulta productiva cuando facilita operaciones instrumentales y permite concentrar la actividad del estudiante en aquellas que construyen comprensión; resulta problemática cuando sustituye las operaciones que forman parte del aprendizaje.

Cuando el aprendizaje requiere que el estudiante intente resolver un problema, recupere información, establezca relaciones entre conceptos, formule una explicación o aplique un conocimiento, obtener directamente una respuesta de la inteligencia artificial **en lugar de realizar esas operaciones** entorpece su proceso de aprendizaje.

La retención permite examinar una de las dimensiones del aprendizaje. Como vimos al analizar la memoria, un ensayo controlado realizado por Barcaui encontró menor retención cuarenta y cinco días después entre estudiantes que habían utilizado ChatGPT sin restricciones que entre quienes habían estudiado mediante los métodos tradicionales utilizados como condición de comparación.[3] Facilitar el acceso y la elaboración durante el estudio no garantiza la conservación posterior de lo aprendido.

La transferencia plantea una exigencia adicional. Aprender implica también poder utilizar conocimientos en situaciones diferentes de aquellas en las que fueron adquiridos. Un estudio reciente de Zhang y colaboradores comparó el aprendizaje de estudiantes universitarios mediante diálogo con ChatGPT o con un instructor humano principiante. Los dos grupos alcanzaron resultados comparables en retención, pero quienes habían trabajado con ChatGPT obtuvieron resultados inferiores en las tareas de transferencia.[4] La muestra fue reducida —34 estudiantes—, por lo que el resultado no permite generalizar esa diferencia a otras poblaciones, contenidos o formas de utilización de inteligencia artificial. Permite, en cambio, distinguir dos dimensiones del aprendizaje: conservar conocimientos y poder utilizarlos en situaciones diferentes.

La interacción educativa con inteligencia artificial adopta formas muy diferentes. Un sistema puede proporcionar directamente una solución o acompañar al estudiante durante la resolución de un problema mediante distintas formas de asistencia. Del mismo modo, el estudiante puede utilizarlo para obtener una respuesta o para trabajar sobre un problema que intenta resolver. Estas formas de interacción requieren actividades cognitivas diferentes.

En contextos educativos intervienen tres agentes. Docentes e instituciones tienen la posibilidad de definir para qué actividades se utiliza inteligencia artificial, en qué momento y qué operaciones debe realizar el estudiante. Los diseñadores y desarrolladores construyen formas de asistencia que proporcionan soluciones o favorecen la formulación de preguntas, la realización de intentos propios y la revisión de errores. El estudiante, a su vez, participa mediante el uso que hace de los recursos disponibles.

Los tres agentes no tienen la misma responsabilidad ni controlan las mismas variables. Docentes e instituciones estructuran la actividad educativa; los diseñadores estructuran las posibilidades y restricciones de la herramienta; el estudiante actúa dentro de esas condiciones.

El valor educativo de la inteligencia artificial no puede establecerse solamente por la calidad de los resultados obtenidos mientras se la utiliza. La inteligencia artificial favorece el aprendizaje cuando contribuye a que el estudiante comprenda, relacione, recupere, explique y aplique conocimientos; cuando sustituye esas operaciones, puede mejorar el desempeño inmediato sin que se produzca el aprendizaje esperado.

Notas

[1] Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö. & Mariman, R. (2025). “Generative AI without guardrails can harm learning: Evidence from high school mathematics”. *Proceedings of the National Academy of Sciences,* 122(26), e2422633122.

[2] Kestin, G., Miller, K., Klales, A., Milbourne, T. & Ponti, G. (2025). “AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting”. *Scientific Reports,* 15, 17458.

[3] Barcaui, A. (2025). “ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention”. *Social Sciences & Humanities Open,* 12, 102287.

[4] Zhang, J., Xu, B., Liu, Y., Cai, W., Wu, L., Qi, S., Peng, Y., Yu, J., Long, T. & Ge, B. (2026). “Exploring ChatGPT's potential in dialogic teaching: A comparative neurocognitive study of AI and human instruction”. *Acta Psychologica,* 268, 107268.

Capítulo V

La soberanía del pensamiento en la era conversacional

1. Una nueva condición del pensamiento

La inteligencia artificial conversacional introdujo una nueva condición para el pensamiento humano. Las tecnologías vinculadas al conocimiento modificaron históricamente las formas de registrar información, conservarla, recuperarla, transmitirla y trabajar con ella. La particularidad de los sistemas actuales reside en que la mediación alcanza también la producción de lenguaje durante la propia actividad intelectual.

Esta mediación adquiere una forma conversacional. La interacción se desarrolla mediante una sucesión de intervenciones en las que cada respuesta toma en cuenta lo expresado anteriormente y puede dar lugar a nuevas preguntas, reformulaciones o desarrollos. La experiencia de interlocución se incorpora así a una tecnología capaz de intervenir de manera continua en actividades intelectuales y, al hacerlo, modifica las condiciones en las que esas actividades se realizan.

La importancia de esta transformación reside en el tipo de actividad que el sistema realiza durante la interacción. Una inteligencia artificial recupera información, la resume, relaciona contenidos, propone una estructura para un texto, formula alternativas, desarrolla un argumento o resuelve determinadas partes de un problema. La IA realiza ahora, total o parcialmente, operaciones que antes requerían nuestra participación.

Esto modifica la distribución de la actividad durante el trabajo intelectual. Ampliamos nuestras posibilidades de exploración, organización y elaboración al trabajar con inteligencia artificial. Podemos examinar una idea desde perspectivas que no habíamos considerado y recibir ya formuladas relaciones, argumentos o soluciones. Entre ambos extremos existe una amplia variedad de formas de interacción en las que cambia la proporción y, sobre todo, la naturaleza de la actividad que queda a cargo de cada uno.

Por eso, **la cuestión central es identificar qué trabajo intelectual dejamos de realizar cuando el sistema lo hace en nuestro lugar y qué consecuencias tiene esa distribución.** Ahorrar una operación instrumental puede permitirnos concentrar la actividad en aquello que nos exige mayor comprensión o criterio. Sustituir una operación constitutiva y necesaria para comprender, evaluar o aprender produce una situación diferente. La misma capacidad tecnológica adquiere así significados distintos según el lugar que ocupa dentro del proceso de pensamiento.

La disponibilidad permanente de estas capacidades amplía considerablemente las posibilidades de asistencia intelectual. Al mismo tiempo, hace posible que una parte creciente del recorrido entre una pregunta y una respuesta sea realizada por el sistema. El resultado puede ser correcto, útil e incluso superior al que habríamos obtenido sin asistencia. Pero la calidad del resultado no determina por sí

sola cuál fue nuestra participación intelectual ni qué operaciones realizamos para comprenderlo, evaluarlo o incorporarlo a nuestro pensamiento.

La nueva condición introducida por la inteligencia artificial conversacional puede formularse entonces como una redistribución de la actividad intelectual entre nosotros y el sistema.

A medida que los sistemas artificiales intervienen en un número mayor de esas actividades, la pregunta comienza a desplazarse hacia la dirección del proceso. Una parte creciente de las operaciones que acompañan al pensamiento puede realizarse mediante sistemas artificiales sin que por ello dejemos de dirigir nuestra actividad intelectual. La cuestión consiste en determinar en qué condiciones esa mediación amplía nuestras posibilidades y en cuáles comienza a sustituir operaciones que necesitamos realizar para comprender, evaluar o decidir.

Esa pregunta conduce al problema de la soberanía del pensamiento.

---

2. Mantener la dirección de nuestro pensamiento

La soberanía del pensamiento consiste en conservar la dirección de nuestra actividad intelectual, examinar las razones que sostienen una respuesta, reconocer los límites de nuestra comprensión y decidir qué lugar concedemos al sistema en los procesos cognitivos y subjetivos. **Esa soberanía no exige que realicemos todas las operaciones que intervienen en el proceso, sino que conservemos la capacidad de orientar nuestra actividad, evaluar aquello que recibimos y decidir qué incorporamos a nuestro pensamiento y a nuestras decisiones.**

Conservamos la dirección cuando nuestros propósitos organizan el curso de la interacción. Esto supone mantener la posibilidad de modificar una pregunta, rechazar una formulación, interrumpir un recorrido, volver sobre una premisa o buscar fuera de la interacción aquello que necesitamos para continuar pensando. La conversación con el sistema queda así subordinada al problema que intentamos comprender o resolver.

La apropiación constituye otra dimensión. Una formulación se incorpora a nuestro pensamiento cuando la comprendemos, la relacionamos con otros conocimientos y la evaluamos. La facilidad con que una respuesta puede utilizarse en un texto, una decisión o la resolución de una tarea dice poco acerca de ese proceso de apropiación. Lo decisivo es la actividad intelectual que realizamos sobre aquello que recibimos.

Esa actividad incluye reconocer los límites de nuestra comprensión y conocimiento. Advertir que una afirmación requiere verificación, que una cuestión excede los conocimientos de los que disponemos o que existen razones que todavía no podemos evaluar permite identificar esos límites. **Reconocer hasta dónde alcanza nuestro juicio forma parte de la soberanía intelectual.** También permite identificar cuándo resulta necesario recurrir a otras fuentes, conocimientos especializados o interlocutores humanos.

La soberanía comprende, además, la posibilidad de examinar la relación que establecemos con el sistema. Podemos observar qué lugar va adquiriendo la inteligencia artificial en nuestra manera habitual de trabajar, aprender, elaborar ideas o tomar decisiones; qué operaciones le confíamos y cuáles conservamos; cuándo la utilizamos para ampliar una elaboración y cuándo comenzamos a sustituir actividades que necesitamos realizar para comprender, evaluar o elaborar. Esta capacidad de examinar y modificar la propia relación con la herramienta forma parte de la dirección del proceso.

La cuestión no se resuelve, por tanto, estableciendo la medida apropiada de intervención de la inteligencia artificial. La utilización frecuente puede coexistir con nuestra participación intelectual intensa, del mismo modo que una intervención ocasional puede resultar inconveniente si aceptamos

una respuesta sin comprenderla o evaluarla. **La distribución relevante es cualitativa: depende de qué operaciones realiza el sistema, cuáles realizamos nosotros y qué función cumplen unas y otras dentro del proceso.**

A medida que aumentan las capacidades de estos sistemas, también aumenta el número de operaciones que podemos realizar con su asistencia. Esto hace cada vez más importante reconocer qué función cumple cada una de esas operaciones dentro de la actividad intelectual.

El desafío consiste en determinar qué operaciones podemos confiar productivamente a estos sistemas y cuáles necesitamos seguir realizando para conservar comprensión, criterio y autonomía.

---

3. Comprensión, criterio y autonomía

Comprensión, criterio y autonomía están estrechamente vinculados porque la posibilidad de dirigir nuestra actividad intelectual depende de que comprendamos aquello con lo que trabajamos, que podamos evaluarlo y que conservemos la decisión sobre lo que finalmente aceptamos, utilizamos o descartamos.

Una respuesta clara y bien organizada ofrece condiciones favorables para comprender; la actividad mediante la cual esa respuesta adquiere sentido la realizamos nosotros. A medida que los sistemas producen formulaciones cada vez más completas y accesibles, adquiere mayor importancia que podamos reconocer qué hemos comprendido efectivamente.

La comprensión permite trabajar intelectualmente con aquello que recibimos; el criterio nos permite establecer qué valor le concedemos. Comprender un argumento permite examinar sus razones, relacionarlo con otros conocimientos, considerar sus supuestos y decidir si existen motivos suficientes para aceptarlo. Una respuesta puede resultar comprensible y, al mismo tiempo, requerir verificación, admitir otras interpretaciones o contener conclusiones que decidimos no aceptar.

La disponibilidad inmediata de respuestas introduce aquí una condición particular. El sistema produce en pocos segundos explicaciones, argumentos o soluciones cuya evaluación exige conocimientos, contraste y tiempo. **La velocidad con que obtenemos una respuesta y el tiempo necesario para formarnos un juicio sobre ella pertenecen a procesos diferentes.**

La capacidad de detenerse adquiere valor cognitivo en una tecnología construida para continuar respondiendo. La forma conversacional facilita el paso inmediato de una respuesta a la siguiente pregunta y de esta a una nueva respuesta. Conservar la dirección de la actividad intelectual incluye introducir el intervalo que el sistema no impone cuando hace falta comprender, verificar o decidir. Interrumpir momentáneamente la interacción, revisar una afirmación, consultar otra fuente o elaborar una respuesta propia son formas de conservar ese espacio.

La autonomía se ejerce cuando conservamos la decisión sobre aquello que incorporamos a nuestro pensamiento y sobre las razones por las que lo hacemos. Esta decisión se apoya en la comprensión de lo recibido y en el criterio con que lo evaluamos. La autonomía aparece así como una forma de participación activa en un proceso mediado: utilizamos recursos producidos por el sistema y mantenemos bajo nuestra dirección el valor y el lugar que les concedemos.

Cuanto mayor sea la capacidad de la inteligencia artificial para producir respuestas plausibles, articuladas y utilizables, más importante resulta nuestra capacidad para comprenderlas, evaluarlas y decidir qué hacer con ellas. El desarrollo de sistemas más capaces aumenta además la importancia de las condiciones que permiten ejercer esas actividades.

---

4. Una responsabilidad compartida y diferenciada

Conservar la comprensión, el criterio y la autonomía constituye una responsabilidad nuestra, pero las condiciones para ejercer esas capacidades también dependen de cómo se diseñan e incorporan los sistemas con los que interactuamos. La responsabilidad es, en este sentido, compartida y diferenciada: intervienen usuarios, desarrolladores, instituciones, educadores y profesionales, y cada uno actúa sobre aspectos diferentes de las condiciones en las que se produce la interacción.

Nosotros conservamos la responsabilidad sobre el lugar que concedemos al sistema en nuestra actividad intelectual, sobre aquello que aceptamos o descartamos, sobre las verificaciones que realizamos y sobre las operaciones que decidimos mantener bajo nuestra dirección. Esa responsabilidad forma parte de nuestro ejercicio de la autonomía.

Pero las condiciones en las que ejercemos esa autonomía dependen también del diseño de los sistemas. La manera en que presentan sus respuestas, expresan incertidumbre, ofrecen fuentes, mantienen la continuidad de una conversación o facilitan determinadas formas de interacción influye en nuestra experiencia. Los efectos humanos de esas decisiones necesitan ser considerados como parte del diseño y no solamente como consecuencias que se observan una vez que la tecnología ya está en uso.

La investigación psicológica, educativa y social aporta conocimientos relevantes sobre atención, aprendizaje, toma de decisiones, persuasión, confianza, dependencia y regulación emocional que deben tenerse en cuenta en el diseño de estos sistemas.

Las instituciones también participan de esta responsabilidad. Escuelas, universidades, organizaciones y ámbitos profesionales establecen condiciones concretas para el uso de la inteligencia artificial. Deciden qué actividades pueden realizarse con asistencia, cuáles requieren nuestra participación directa, qué formas de verificación se consideran necesarias y qué capacidades conviene preservar y desarrollar. Estas decisiones contribuyen a definir el lugar que la tecnología ocupa en las prácticas intelectuales.

En el ámbito educativo, esta cuestión adquiere especial importancia. La incorporación de inteligencia artificial no consiste solamente en decidir si se permite o se prohíbe su utilización. Requiere determinar qué actividades contribuyen al aprendizaje y cuáles pueden delegarse sin sustituir procesos que el estudiante necesita realizar para comprender y adquirir criterio. El *OECD Digital Education Outlook 2026* señala precisamente la necesidad de integrar la inteligencia artificial de manera que fortalezca la agencia de estudiantes y docentes y preserve objetivos educativos que exceden la producción inmediata de resultados.[1]

Los profesionales vinculados con la educación, la salud, el diseño, la comunicación y otros campos en los que estas tecnologías se incorporan tienen también una función específica. Su conocimiento permite identificar efectos que no siempre resultan visibles desde una perspectiva exclusivamente tecnológica. La participación interdisciplinaria amplía así la capacidad de evaluar cómo determinadas formas de interacción inciden sobre las prácticas y capacidades humanas.

Nosotros conservamos responsabilidad sobre el uso que hacemos del sistema; los desarrolladores, sobre las condiciones de interacción que contribuyen a crear; las instituciones, sobre las formas en que incorporan estas tecnologías a sus prácticas; y los profesionales que participan en esos procesos, sobre el conocimiento que aportan para comprender y evaluar sus efectos sobre los seres humanos.

Una tecnología que interviene en procesos humanos de pensamiento, aprendizaje, decisión y regulación subjetiva requiere que quienes la diseñan e incorporan a la vida social tengan en cuenta el conocimiento disponible sobre esos procesos.

Nota

[1] OECD. (2026). *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*. OECD Publishing.

5. Lo que todavía no sabemos

La evidencia disponible permite identificar modificaciones en la actividad cognitiva durante determinadas tareas realizadas con inteligencia artificial y diferencias posteriores en algunos indicadores de rendimiento. **La evidencia sobre los efectos acumulativos de interactuar durante años con estos sistemas es, en cambio, mucho más limitada.**

Los estudios experimentales permiten observar qué ocurre cuando una persona realiza determinadas tareas con asistencia de inteligencia artificial y, en algunos casos, evaluar aspectos de su rendimiento. Conocer las consecuencias de una utilización prolongada requiere estudios longitudinales capaces de examinar cómo evolucionan las formas de pensar, aprender, evaluar y decidir cuando incorporamos estos sistemas a la actividad cotidiana. La evidencia disponible en este nivel continúa siendo limitada.[1]

El problema científico actual consiste en investigar bajo qué condiciones la interacción con la IA amplía las capacidades humanas, bajo cuáles reduce la participación en las actividades en las que esas capacidades se ejercen y qué consecuencias producen esas formas de interacción a largo plazo. La tarea, la forma de utilizar el sistema, las operaciones que realizamos y aquellas que quedan a cargo de la inteligencia artificial constituyen variables necesarias para comprender esos efectos.

Las posibilidades abiertas por esta tecnología admiten desarrollos diferentes. Determinadas formas de interacción facilitan el acceso a información, amplían las alternativas disponibles, ayudan a organizar problemas y permiten concentrar el esfuerzo en operaciones intelectuales de mayor complejidad. Otras formas de uso reducen nuestra participación en **operaciones constitutivas y necesarias para comprender, evaluar o aprender**.

Esta cuestión forma parte de una historia más amplia. La escritura, los libros, las bibliotecas, los sistemas de notación, las calculadoras, los buscadores y otras tecnologías del conocimiento modificaron **qué necesitábamos recordar** y las maneras de acceder, organizar y utilizar información. La inteligencia artificial forma parte de esa historia, con una particularidad: la mediación participa ahora en la producción del lenguaje y en operaciones que forman parte del proceso de elaboración intelectual.

Las consecuencias de una tecnología del conocimiento dependen tanto de las capacidades que aporta como de las prácticas humanas, educativas e institucionales que se construyen alrededor de ella. Por eso, los efectos futuros de la inteligencia artificial dependerán también de las formas de interacción que aprendamos a establecer, de los criterios con los que incorporemos estos sistemas y de las decisiones con las que orientemos su diseño y utilización.

La autonomía humana nunca dependió de pensar sin mediaciones. Cada transformación de las tecnologías del conocimiento modificó las condiciones en las que pensábamos, recordábamos, aprendíamos y nos comunicábamos. La inteligencia artificial introduce una nueva mediación, capaz de producir una parte del lenguaje y de las operaciones que intervienen en esas actividades. La cuestión es cómo conservar, dentro de esa mediación, la dirección de nuestro pensamiento, la capacidad de comprender y evaluar, y la responsabilidad sobre aquello que finalmente aceptamos como propio.

El problema decisivo no será cuánto llegará a hacer la inteligencia artificial por nosotros, sino qué seguiremos haciendo nosotros cuando pensemos con ella.

Nota

[1] Bengio, Y., et al. (2026). *International AI Safety Report 2026*. Department for Science, Innovation and Technology. El informe sintetiza la evidencia disponible sobre los riesgos de los sistemas de inteligencia artificial de propósito general y señala importantes lagunas en el conocimiento de sus efectos sobre la autonomía humana. Destaca las limitaciones de la evidencia disponible para establecer las consecuencias de formas sostenidas de interacción y la necesidad de ampliar la investigación sobre sus efectos a largo plazo.

Bibliografía

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Alubthane, F. O. (2026). Amplifier or substitute? A systematic review of generative AI's impact on higher-order cognitive skills among university students. *Frontiers in Psychology, 17*, 1863931. https://doi.org/10.3389/fpsyg.2026.1863931

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th Conference on Creativity & Cognition* (pp. 413–425). Association for Computing Machinery. https://doi.org/10.1145/3635636.3656204

Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open, 12*, 102287. https://doi.org/10.1016/j.ssaho.2025.102287

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Bengio, Y., Clare, S., Prunkl, C., Murray, M., Andriushchenko, M., Bucknall, B., Bommasani, R., Casper, S., Davidson, T., Douglas, R., Duvenaud, D., Fox, P., Gohar, U., Hadshar, R., Ho, A., Hu, T., Jones, C., Kapoor, S., Kasirzadeh, A., . . . Mindermann, S. (2026). International AI safety report 2026 (DSIT 2026/001). Department for Science, Innovation and Technology.

Bion, W. R. (1962). *Learning from experience*. Heinemann. [Trad. esp.: *Aprendiendo de la experiencia*]

Bruner, J. (1991). *Actos de significado: Más allá de la revolución cognitiva* (J. C. Gómez Crespo & J. L. Linaza Iglesias, Trads.). Alianza Editorial. (Obra original publicada en 1990)

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188. https://doi.org/10.1145/3449287

Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. *Science, 391*(6792), eaec8352. https://doi.org/10.1126/science.aec8352

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Croes, E., Antheunis, M., van der Lee, C., & de Wit, J. (2024). Digital confessions: The willingness to disclose intimate information to a chatbot and its impact on emotional well-being. *Interacting with Computers, 36*(5), 279–292. https://doi.org/10.1093/iwc/iwae016

De Freitas, J., Castelo, N., Uğuralp, A. K., & Oğuz-Uğuralp, Z. (2026). Mourning the loss of AI companions. *Nature Human Behaviour*. Advance online publication. https://doi.org/10.1038/s41562-026-02569-3

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). *How AI and human behaviors shape psychosocial effects of chatbot use: A longitudinal randomized controlled study* [Preprint]. arXiv.

https://doi.org/10.48550/arXiv.2503.17473

Folk, D., & Dunn, E. (2026). How does turning to AI for companionship predict loneliness and vice versa? *Psychological Science, 37*(4), 276–286. https://doi.org/10.1177/09567976261427747

Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E., Coughlin, J. F., Guttag, J. V., Colak, E., & Ghassemi, M. (2021). Do as AI say: Susceptibility in deployment of clinical decision-aids. *npj Digital Medicine, 4*, 31. https://doi.org/10.1038/s41746-021-00385-9

Han, B.-C. (2010). *Müdigkeitsgesellschaft*. Matthes & Seitz. [Trad. esp.: *La sociedad del cansancio*]

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*(4), 712–733. https://doi.org/10.1093/joc/jqy026

Ho, J. Q. H., Hu, M., Chen, T. X., & Hartanto, A. (2025). Potential and pitfalls of romantic artificial intelligence (AI) companions: A systematic review. *Computers in Human Behavior Reports, 19*, 100715. https://doi.org/10.1016/j.chbr.2025.100715

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux. [Trad. esp.: *Pensar rápido, pensar despacio*]

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, 17458. https://doi.org/10.1038/s41598-025-97652-6

Kim, S. S. Y., Vaughan, J. W., Liao, Q. V., Lombrozo, T., & Russakovsky, O. (2025). Fostering appropriate reliance on large language models: The role of explanations, sources, and inconsistencies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3706598.3714020

Knor, E. L., Reiss, M. V., Möller, J., & Merten, L. (2026). Determinants of verification behavior in generative search: Evidence from a conjoint experiment. *Computers in Human Behavior Reports, 22*, 101056. https://doi.org/10.1016/j.chbr.2026.101056

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.08872

Krügel, S., Ostermaier, A., & Uhl, M. (2023). ChatGPT's inconsistent moral advice influences users' judgment. *Scientific Reports, 13*, 4569. https://doi.org/10.1038/s41598-023-31341-0

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421–428. https://doi.org/10.1111/j.1467-9280.2007.01916.x

Liu, A. R., Pataranutaporn, P., & Maes, P. (2025). The heterogeneous effects of AI companionship: An empirical model of chatbot usage and loneliness and a typology of user archetypes. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 8*(2), 1585–1597. https://doi.org/10.1609/aies.v8i2.36658

Liu, T., Giorgi, S., Aich, A., Lahnala, A., Curtis, B., Ungar, L., & Sedoc, J. (2025). The illusion of empathy: How AI chatbots shape conversation perception. *Proceedings of the AAAI Conference on Artificial Intelligence, 39*(13), 14327–14335. https://doi.org/10.1609/aaai.v39i13.33569

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nourani, M., King, J. T., & Ragan, E. D. (2020). The role of domain expertise in user trust and the impact of first impressions with intelligent systems. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 8*(1), 112–121. https://doi.org/10.1609/hcomp.v8i1.7469

OECD. (2026). *OECD Digital Education Outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. https://doi.org/10.1787/062a7394-en

Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. Methuen. [Trad. esp.: *Oralidad y escritura*]

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95–103. https://doi.org/10.1037/h0045357

Rosa, H. (2016). *Alienación y aceleración: Hacia una teoría crítica de la temporalidad en la modernidad tardía*. Katz Editores.

Romeo, G., & Conti, D. (2026). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI & Society, 41*, 259–278. https://doi.org/10.1007/s00146-025-02422-7

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521–562. https://doi.org/10.1207/S15516709COG2605_1

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology, 70*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S. M., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2024). Towards understanding sycophancy in language models. *International Conference on Learning Representations*.

Soh, H., Xie, Y., Chen, M., & Hsu, D. (2020). Multi-task trust transfer for human–robot interaction. *The International Journal of Robotics Research, 39*(2–3), 233–249. https://doi.org/10.1177/0278364919866905

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, 108386. https://doi.org/10.1016/j.chb.2024.108386

Steyvers, M., Tejeda, H., Kumar, A., Belem, C., Karny, S., Hu, X., Mayer, L. W., & Smyth, P. (2025). What large language models know and what people think they know. *Nature Machine Intelligence, 7*, 221–231. https://doi.org/10.1038/s42256-024-00976-7

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

https://doi.org/10.1093/mind/LIX.236.433

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

UNESCO. (2024). *AI competency framework for students*. UNESCO.

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education, 215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. [Trad. esp.: *El desarrollo de los procesos psicológicos superiores*]

Vzorin, G., Petrova, V., & Sedykh, A. (2024). ChatGPT may foster human gist memory while offloading less sufficient information. In F. Lorig et al. (Eds.), *HHAI 2024: Hybrid Human AI Systems for the Social Good* (pp. 485–487). IOS Press. https://doi.org/10.3233/FAIA240232

Yao, Y., Han, T., & Li, D. (2026). Assistance or distraction? A cognitive ergonomics perspective on cognitive load during AI-assisted post-editing. *International Journal of Human–Computer Interaction*, 1–20. https://doi.org/10.1080/10447318.2026.2628994

Zhang, J., Xu, B., Liu, Y., Cai, W., Wu, L., Qi, S., Peng, Y., Yu, J., Long, T., & Ge, B. (2026). Exploring ChatGPT's potential in dialogic teaching: A comparative neurocognitive study of AI and human instruction. *Acta Psychologica, 268*, 107268. https://doi.org/10.1016/j.actpsy.2026.107268

Sobre el autor

Ricardo Luis Gómez Vecchio (Buenos Aires, 1953) es Licenciado en Psicología por la Universidad de Buenos Aires. Ha desarrollado actividad clínica como psicoterapeuta de adultos y parejas, y ha trabajado en el ámbito de la divulgación científica y la comunicación.
Se desempeñó como redactor y productor en Radio Nacional Buenos Aires, y colaboró con medios como *Clarín*, *La Nación*, *Página/12* y *Ámbito Financiero*, abordando temas científicos y culturales. Fue Coordinador del Centro de Divulgación Científica de la Facultad de Psicología (UBA) y Director de Comunicaciones de la Universidad de Belgrano.
Su trayectoria integra la reflexión psicológica, la divulgación del conocimiento y un interés sostenido por las dimensiones filosóficas y espirituales de la experiencia humana, que actualmente vuelca en su blog esenciasdeletras.com