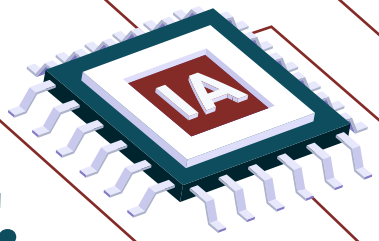
Cuando la IA deja de limitarse a responder. De la intervención humana al control humano



| Por la traductora pública Silvana Debonis

En esta serie de artículos venimos intentando que los algoritmos y la inteligencia artificial dejen de resultarnos tan ajenos. Porque para cuestionar, criticar o, incluso, decidir qué lugar les queremos dar en nuestro ejercicio profesional, primero tenemos que entender de qué estamos hablando.

El ritmo del cambio tecnológico actual es abrumador. Surgen nuevas herramientas y posibilidades a una velocidad sin precedentes y a más de uno nos genera angustia no poder seguirles el paso. Sin embargo, lo peor que podemos hacer es paralizarnos. Si queremos atravesar esta transformación con criterio, debemos mantenernos informados y tratar de entender no solo qué herramientas tenemos hoy, sino también hacia dónde se está desplazando la tecnología.

Por eso, en este artículo quiero dirigir la mirada precisamente hacia allí. Hacia lo que se viene. No porque lo visto hasta ahora haya quedado atrás. La inteligencia artificial generativa, los modelos de lenguaje, los *prompts*, la documentación y el uso de fuentes siguen siendo piezas fundamentales. Lo que empieza a cambiar es la manera en que esas piezas se combinan y, sobre todo, el tipo de relación que podemos establecer con sistemas que ya no solo responden, sino que también pueden actuar.

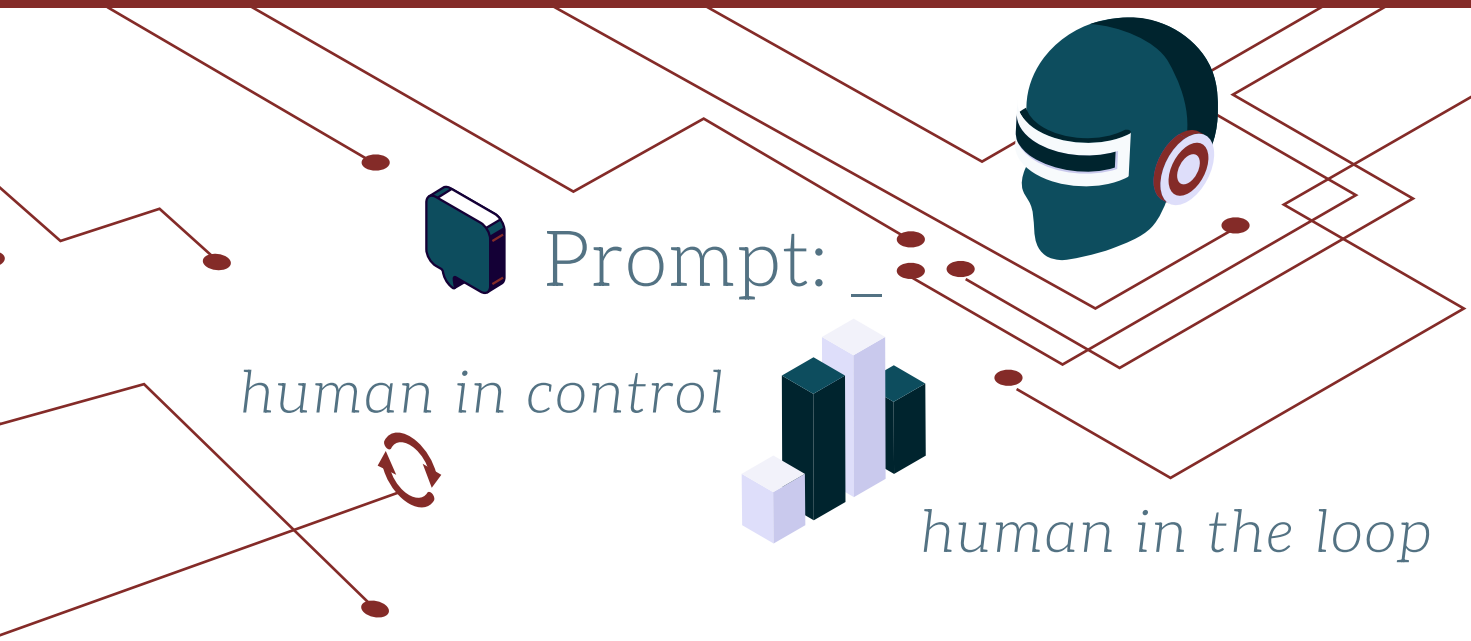
El cambio del que me ocupo puede resumirse en palabras sencillas: pasar de darle instrucciones a una inteligencia artificial a darle objetivos.

Hasta ahora, buena parte de nuestra relación cotidiana con la IA generativa se construyó en torno al *prompting*. Le pedimos que resuma un documento, compare dos versiones, identifique términos, reformule un párrafo o sugiera alternativas. La herramienta responde y nosotros decidimos qué hacer con esa respuesta. Incluso con tareas más complejas, también le marcamos un recorrido con el diseño de cadenas de *prompts*. Podemos pedirle primero que analice un texto, después que extraiga la terminología especializada, que la compare con un glosario y que, finalmente, señale las incongruencias. Puede haber varias acciones y varios intercambios, pero seguimos siendo nosotros quienes decidimos el siguiente paso.

Sin embargo, la novedad es que ahora, en lugar de indicarle cada una de las tareas, podemos definir un objetivo o resultado. La diferencia no radica en escribir un *prompt* más largo o más detallado. Es una diferencia conceptual. Una cosa es diseñar el recorrido para que el sistema ejecute las instrucciones; otra muy distinta es definir el resultado que buscamos y dejar parte de la planificación y ejecución en manos del sistema. Aquí entramos en el terreno de la llamada IA agéntica y, más específicamente, de los agentes de IA.

Para muchos, el término «agente» puede sonar un tanto perturbador. Sin embargo, no estamos hablando de máquinas con intenciones propias ni de sistemas que deciden como decidimos las personas. Hablamos de sistemas basados en IA que pueden recurrir a distintas fuentes de información y herramientas, y encadenar una serie de acciones para alcanzar un objetivo.

Para quienes trabajamos con traducción, varias de esas piezas no son nuevas. Estamos acostumbrados a trabajar con glosarios, memorias, corpus, documentos anteriores y material de referencia. En el artículo anterior hablábamos justamente de la posibilidad de darle a la IA documentación concreta para que no dependa únicamente del conocimiento incorporado durante su entrenamiento.



Prompt: _

human in control

human in the loop

Con los agentes aparece un paso adicional. Esa documentación ya no sirve solamente para mejorar la respuesta que produce el modelo; también puede orientar lo que el sistema hace después.

Pensemos, por ejemplo, en una revisión terminológica. En lugar de indicarle cada paso, podríamos pedirle que revise la coherencia de un proyecto a partir del glosario y de los antecedentes del cliente. El sistema podría hacer las comprobaciones necesarias y dejarnos para revisión aquellos casos en los que la información disponible no alcanza para tomar una decisión.

No hace falta imaginar escenarios mucho más complejos para entender que estamos frente a un cambio fundamental en el que se da a la IA cierto margen de maniobra para actuar en pos de un objetivo.

Una vez más, los avances tecnológicos nos ponen delante de una pregunta casi existencial. Hasta ahora estábamos obsesionados con saber qué puede hacer la IA. Ahora tenemos que empezar a preguntarnos qué estamos dispuestos a delegar.

No todas las tareas tienen el mismo peso ni las mismas consecuencias. Podemos sentirnos perfectamente cómodos dejando que una herramienta detecte incongruencias, compare documentos o prepare una lista preliminar de problemas. Pero otra cosa es permitir que esa misma herramienta pueda modificar automáticamente una traducción, incorporar una equivalencia a una base terminológica o aplicar una decisión a cientos de segmentos sin una nueva intervención nuestra.

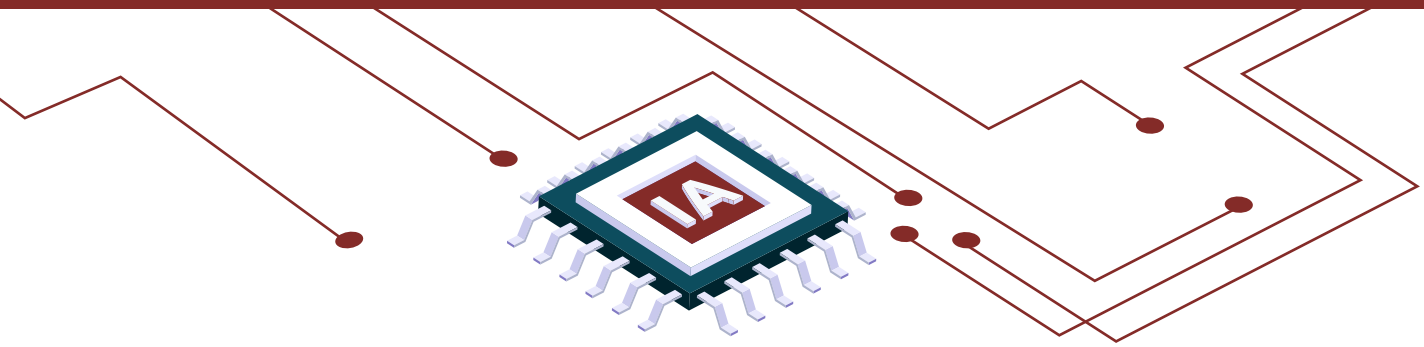
En este punto creo importante aclarar que la autonomía no es todo o nada. Entre ejecutar manualmente cada paso y permitir que un sistema actúe sin restricciones existe una enorme cantidad de posibilidades intermedias.

Ahí entra en juego una expresión que aparece constantemente cuando se habla de inteligencia artificial: *human in the loop*.

En términos generales, la expresión transmite la idea de que debe haber una persona dentro del proceso que intervenga en determinados momentos para supervisar, validar, corregir o autorizar lo que hace el sistema. Es un principio importante y absolutamente necesario. Sin embargo, cuando empezamos a pensar en sistemas capaces de planificar y encadenar acciones, considero que se necesita más que eso.

Cada vez me inclino más por la expresión *human in control*. No lo planteo como una disputa terminológica ni como la necesidad de reemplazar una expresión por otra. El punto es que describen preocupaciones diferentes. *Human in the loop* nos lleva a preguntarnos en qué momento interviene la persona. *Human in control* nos plantea una cuestión más de fondo: quién determina las condiciones dentro de las cuales puede actuar el sistema.

Que haya una persona en algún punto del proceso no garantiza, por sí solo, que esa persona conserve el control sobre él. Imaginemos un sistema que realiza veinte acciones y solicita aprobación humana en la número diecinueve. Técnicamente habría una persona en el proceso, pero muchas de las decisiones que dieron forma al resultado ya se habrían tomado. También puede ocurrir que el sistema nos pida confirmar una acción sin mostrarnos con suficiente claridad qué fuentes consultó, qué acciones realizó o qué otras modificaciones introdujo para llegar hasta allí.



Por eso, cuando hablamos de control humano, la cuestión no debe reducirse a insertar una aprobación en algún punto de la cadena. Importa quién fija el objetivo, qué herramientas puede utilizar el sistema, qué información tiene permitido consultar, qué fuentes deben prevalecer cuando existen contradicciones, qué acciones puede ejecutar directamente y cuáles requieren autorización. También es fundamental definir qué hacer cuando la información no alcanza, cuando dos instrucciones se contradicen o cuando el sistema se encuentra con una situación que no estaba prevista.

Por ejemplo, supongamos que durante una revisión el sistema detecta que un término aparece traducido de tres maneras y que una de ellas coincide con el glosario aprobado. Podríamos autorizarlo a reemplazar automáticamente las otras dos variantes. A primera vista parece una decisión perfectamente razonable. Si existe un glosario, ¿no debería respetarse?

Quienes trabajamos con traducción sabemos que no siempre es tan sencillo. Las variantes pueden corresponder a contextos diferentes. El glosario puede haber quedado desactualizado o puede existir una instrucción posterior del cliente que nunca llegó a incorporarse formalmente. Más aún, la equivalencia registrada podría estar en el glosario simplemente por error.

Si nos topamos con un caso de equivalencia equivocada, podemos corregirlo, actualizar el glosario o simplemente cambiar el término en la traducción. Si un agente se cruza con este término, podría aplicarlo automáticamente al resto del proyecto y, si tiene autorización para actualizar los recursos de referencia, utilizarlo después en otros documentos. Aquí estamos frente a un verdadero problema. La misma capacidad que permite automatizar una decisión correcta puede multiplicar una decisión equivocada.

Con esto no estoy diciendo que debamos oponernos de plano a la automatización. Forma parte de una transformación que ya está en marcha en todas las profesiones. Además, la posibilidad de automatizar secuencias de tareas puede ahorrarnos trabajo mecánico y permitir que nos concentremos en decisiones que realmente requieren conocimiento profesional.

Pero para aprovechar esa capacidad necesitamos pensar con más precisión dónde queremos que esté el control. Y control no significa necesariamente revisar manualmente cada acción, porque eso implicaría convertir al profesional en una especie de supervisor que pasa el día haciendo clic en «aceptar».

Para empezar, considero que el control debe estar en el diseño previo del proceso. Debemos poder establecer qué fuentes son válidas y qué jerarquía tienen entre sí, determinar qué acciones son reversibles, fijar los casos que requieren consulta y decidir de antemano qué cosas el sistema no puede hacer. También puede implicar que queden registradas las fuentes consultadas, las acciones realizadas, los cambios introducidos y las autorizaciones requeridas, de modo que podamos reconstruir posteriormente qué ocurrió durante el proceso.

A medida que estas herramientas se incorporen a nuestro trabajo, seguramente tendremos que revisar muchas de nuestras prácticas, aprender nuevas formas de supervisión y, probablemente, redefinir algunas tareas que hasta hace poco considerábamos propias del ejercicio profesional. Pero el desafío no pasa solamente por aprender a utilizar la tecnología. También tendremos que decidir qué lugar queremos ocupar nosotros en esa nueva distribución del trabajo.

Quizá sea en ese punto en el que se juegue una parte importante de nuestro ejercicio profesional. Si pasamos de darle instrucciones a la IA a darle objetivos, nuestra tarea no puede reducirse a aprobar o rechazar lo que hace al final del recorrido. También tendremos que decidir de antemano con qué información puede trabajar, qué acciones puede realizar, hasta dónde puede avanzar y cuándo debe detenerse. Delegar pasos no equivale a delegar criterio. El control no está solamente en intervenir durante el proceso, sino también en haber definido antes las condiciones dentro de las cuales la IA puede actuar. Quizá de eso se trate, en definitiva, el control humano.■