IA-NT-ES-25-529
MAYO 2025

PATRICO ROJAS E.

No se trata de controlar, sino de entender: la IA generativa y el potencial del conocimiento distribuido

¿Has notado que, en los últimos meses, todo parece estar mejor escrito? Informes, propuestas comerciales, minutas, todo tipo de cosas.

Al mencionarlo hace poco me dijeron: *"¿Y te extraña? 800 millones de personas están usando ChatGPT cada semana"*. Lo cierto es que su velocidad de adopción ha sido notable. Llegó a los 100 millones de usuarios en dos meses, hito que a Youtube le tomó un año y medio, y Spotify once años.

Frente a estos hechos, es claro que las organizaciones que están "considerando" la adopción de la IA llegan tarde. La gente ya la adoptó, por millones. Clientes, proveedores, competidores y colaboradores ya la están usando, en todos los niveles, en múltiples contextos, con diferentes grados de conciencia y comprensión.

Las razones son evidentes. Las IAs generativas como Chatgpt y Gemini son rápidas, redactan fantástico, suenan convincentes, y han sido entrenadas con un volumen de información equivalente a unas 200 veces lo que contiene Wikipedia. Por otro lado, estudios recientes[1] del MIT y Harvard sugieren que las IA generativas pueden generar aumentos de productividad y calidad en rangos que van del 12 al 40%.

¿Por qué no usarlas? Lo más probable es que nuestra competencia ya la esté usando, ¿queremos quedarnos atrás?

[1] Experimental evidence on the productivity effects of generative artificial intelligence | Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality

Nota Técnica preparada por Patricio Rojas E., profesor del ESE Business School de la Universidad de los Andes (Chile), para servir de base para la discusión en clase y no como ilustración de la gestión, adecuada o inadecuada, de una situación determinada.

El otro lado de la moneda es que ya varios se han hecho famosos por exceso de confianza en las IAs. Está el caso de Steven Schwartz, quien usó ChatgGPT para preparar una demanda a la aerolínea Avianca. El memorándum legal incluyó más de media docena de citas de jurisprudencia supuestamente relevante (ejm. *Martinez v. Delta*, *Zicherman v. Korean Air Lines*), sin embargo, ninguna era real. La IA las había inventado por completo. Cuando el juez federal pidió explicaciones, Schwartz dijo que, para verificar las fuentes, le preguntó a ChatGPT si los casos eran auténticos, cosa que la IA confirmó, falsamente. Para este abogado, con décadas de experiencia, la multa que le impusieron fue lo de menos. Lo grave fue el daño reputacional, tanto personal como a su bufete, pues el caso se difundió ampliamente.

Otro caso es el de CNET, un medio digital especializado en tecnología, cuya editora jefa tuvo que salir a dar explicaciones. El medio publicó durante meses artículos generados con IA pero, tras una revisión interna, se vio obligado a corregir 41 de los 77 textos publicados, al detectarse numerosos errores, especialmente en contenidos financieros. ¿Entre los artículos afectados había títulos como "What Is Compound Interest?", que contenían explicaciones incorrectas y cálculos inexactos sobre conceptos económicos básicos. La situación generó una fuerte polémica, especialmente porque estos artículos estaban diseñados para posicionarse en buscadores mediante palabras clave SEO y así atraer tráfico con fines comerciales. Esto llevó a la empresa a pausar temporalmente la publicación de contenido generado por IA, aunque sus directivos señalaron que continuarán explorando el uso de estas tecnologías.

Las dos situaciones descritas previamente son una combinación de "alucinación" con uso "acrítico"de la IA. Una "alucinación" es una situación en que el modelo genera información incorrecta o inventada que parece plausible, pero no se basa en datos reales. Puede incluir hechos falsos, citas inexistentes, cifras erróneas o atribuciones incorrectas. No se trata de errores intencionales, sino de resultados derivados de cómo el modelo predice texto a partir de patrones aprendidos, sin comprensión ni verificación de la verdad.

Por otro lado, el uso "acrítico" implica confiar en los resultados de la IA sin verificar su precisión, ni considerar sus limitaciones. Este tipo de uso puede llevar a errores, desinformación o consecuencias no deseadas, especialmente cuando se aplica en temas sensibles.

La gran pregunta es si estos casos de uso acrítico son una excepción, o son más frecuentes de lo que nos gustaría pensar. Una breve exploración en la literatura de psicología cognitiva y economía conductual sugiere que existen diversas tendencias en el comportamiento humano, tales como sesgos cognitivos y heurísticas, que nos predisponen a confiar automáticamente en sistemas como ChatGPT o Gemini, delegando nuestro propio juicio. En el centro de estas tendencias se encuentra una fuerza silenciosa pero poderosa: la de minimizar el esfuerzo mental. Nuestra mente, guiada por una economía cognitiva básica, busca respuestas rápidas y plausibles. Así, cuando la IA entrega una respuesta coherente, el camino de menor resistencia suele ser aceptarla tal cual, sin cuestionarla, evaluarla ni contrastarla con otras fuentes. En situaciones de sobrecarga informativa o presión de tiempo, esta tendencia se agudiza, favoreciendo un uso irreflexivo de estas herramientas.

Este principio se combina con otros sesgos clave. Por ejemplo, el sesgo de automatización lleva a confiar excesivamente en sistemas automatizados, delegando el juicio crítico personal. El sesgo de autoridad y el efecto halo tecnológico refuerzan esta confianza acrítica, al percibir a la IA como una fuente experta e innovadora. La creencia en la objetividad algorítmica añade una capa adicional de legitimidad, al suponer que los algoritmos no arrastran sesgos humanos.

A nivel individual, la sobre confianza en el propio juicio, o en la IA, hace que la persona no se cuestione lo recibido. Esto se combina con la ilusión de conocimiento, por la cual una explicación coherente genera una sensación engañosa de comprensión profunda. El sesgo de confirmación cierra el círculo. Al buscar y aceptar solo aquello que, valida nuestras creencias, disminuye la motivación para explorar alternativas. Finalmente, el antropomorfismo y el efecto ELIZA refuerzan la impresión de estar interactuando con un agente inteligente, lo que aumenta la credibilidad que otorgamos a sus respuestas.

En conjunto, estos factores se traducen en un patrón común. La tendencia a minimizar el esfuerzo mental necesario para tomar decisiones o analizar información. Esto no ocurre por ignorancia o descuido, sino porque la evolución programó nuestros cerebros por miles de años para optimizar recursos cognitivos. El riesgo es que esta eficiencia natural derive en una aceptación irreflexiva de respuestas generadas por IA, con consecuencias potencialmente importantes.

Por lo tanto, es muy probable que las situaciones de uso acrítico no sean pocas. En algunos caos las consecuencias serán menores, casi imperceptibles. Pero en otros las consecuencias pueden ser graves, aunque quizás demoren tiempo en incubarse y exponerse con toda su contundencia.

Las empresas están respondiendo al uso acrítico de IA generativa mediante una combinación de medidas reactivas y preventivas. Tras incidentes, muchas han impuesto restricciones estrictas al uso de IAs generativas, o reformulado políticas internas. Otras han optado por acciones preventivas: supervisión humana obligatoria, desarrollo de versiones seguras de IA, comités de revisión ética, y programas de formación masiva en uso responsable. Sectores como el financiero, legal, y tecnológico han liderado estas respuestas, ajustando flujos de trabajo, imponiendo validación humana y estableciendo directrices. Sin embargo, persisten brechas. Un estudio reciente indica que el 48% de los empleados no cumple las políticas. Además, existen limitaciones en la detección automatizada y ausencia de auditorías sistemáticas de sesgos.

Claramente esta forma de enfrentar el desafío de la IA no está funcionando. Instalar comités, redactar normativas o realizar capacitaciones ex post no modifica el hecho de que el fenómeno ya se instaló desde abajo hacia arriba, y que esta tecnología cambia más rápido de lo que los comités alcanzan a revisar o reglamentar

Como lo muestra Karl Weick en "Managing the Unexpected", lo crucial no es tanto diseñar estructuras rígidas, sino desarrollar capacidades organizacionales para gestionar lo inesperado.
Las organizaciones que operan en entornos de alta incertidumbre no evitan el caos mediante control jerárquico, sino que cultivan una forma de acción colectiva que Weick llama *mindful organizing*. Esto implica cinco prácticas clave que ofrecen un marco útil para abordar el fenómeno de la IA en las empresas:

1. Preocupación constante por las fallas: reconocer que el uso de IA puede generar errores, sesgos o desinformación, incluso cuando las respuestas parezcan coherentes. En lugar de castigar los errores, se debe aprender activamente de ellos.
2. Resistencia a simplificar interpretaciones: evitar reducir la IA a "una herramienta más" o asumir que su uso es seguro o trivial. Las organizaciones deben incorporar múltiples perspectivas, técnicas, éticas, operativas, al evaluar su uso.
3. Sensibilidad a las operaciones: prestar atención a cómo realmente se está usando la IA en el trabajo diario. Esto requiere observar con detalle y aprender de cómo los colaboradores interactúan con la IA, en qué tareas la integran y con qué efectos reales.
4. Compromiso con la resiliencia: entender que no todo se puede anticipar y que la capacidad de adaptarse rápidamente ante nuevos usos, problemas o dilemas es más importante que prevenirlo todo desde el diseño. En este sentido, el aprendizaje en la acción es tan valioso como la planificación previa.
5. Respeto al expertise: este punto es quizás el más transformador. Frente al conocimiento distribuido sobre IA que ya existe entre colaboradores, clientes y proveedores, la tarea no es centralizar, sino escuchar y articular ese saber emergente. No hay curso de IA más actualizado que lo que ya están probando, compartiendo y aprendiendo cientos de colaboradores en tiempo real.

En vez de normar desde arriba, las organizaciones pueden diseñar mecanismos para hacer visible ese conocimiento distribuido, crear espacios de reflexión compartida, y promover capacidades de interpretación colectiva. Más que aplicar reglas fijas, las organizaciones necesitan desarrollar capacidades colectivas de interpretación: *sensemaking* frente al cambio.

En un entorno donde los modelos cambian cada mes y las formas de uso se transforman en semanas, la solución no es imponer control, sino desarrollar conciencia organizacional. No se trata de decir qué se puede o no se puede hacer con la IA, sino de generar condiciones para que las personas puedan pensar mejor lo que están haciendo con ella, y así sacarle el máximo provecho.