

Nueva métrica clave:

Por fin entendió cómo funciona la IA en el trabajo. Ahora viene la cuenta

WSJ

CONTENIDO LICENCIADO POR
THE WALL STREET JOURNAL

KATHERINE BINDLEY
The Wall Street Journal

El creador de la plataforma de automatización equipada con inteligencia artificial (IA), Zapier, tiene un nuevo tipo de tablero de mando para mantenerse informado sobre el uso de la IA por parte de sus trabajadores.

¿La nueva métrica clave? Cuántos *tokens* están gastando.

El rendimiento de la IA, que ha impulsado la productividad y promete cambiar la naturaleza del trabajo, podría parecer como si se materializara por arte de magia. Pero es realmente el trabajo de los centros de datos que mueven e interpretan instrucciones en un proceso complicado y caro, aunque invisible.

Cada vez que una persona da instrucciones a un bot o tiene un agente que escribe códigos, los recursos computacionales se miden en *tokens*. Para la IA basada en texto, es muy sencillo: generar 750 palabras requiere alrededor de mil *tokens*. Se complica más cuando escribe código, crea un video y audio o recluta a agentes para que realicen complejas tareas que pueden durar días, pero la idea sigue siendo la misma: mientras más trabajo hace, más *tokens* se utilizan.

“Tenemos este nuevo tipo de línea presupuestaria”, señala Brandon Sammut, director de transformación de IA de Zapier. La asistencia que proporciona la IA —ya sea que maneje una solicitud de soporte o cierre un trato— tiene un costo y las empresas tienen que considerar eso.

Aunque el precio de los *tokens* ha bajado, su costo puede ser más alto para algunos modelos más nuevos y con mayor demanda, y su uso por parte de las empresas en general está aumentando. A veces las empresas optan por planes de pago según consumo, mientras que otras compran planes empresariales que incluyen una cierta cantidad de uso por trabajador.

Una mayoría de empresas aún está tratando de lograr que sus empleados utilicen la IA.

Las empresas que ahora utilizan inteligencia artificial en forma habitual están empezando a hacer un seguimiento al uso que hacen sus trabajadores de los *tokens*, la unidad de medida de la IA.



Una mayoría de empresas aún está tratando de lograr que sus empleados utilicen la IA.

Pero aquellas que están más avanzadas ya están haciendo un seguimiento del uso de *tokens* y empezando a estimar los costos. Están explorando qué estrategias de IA se deberían amplificar después de generar un gran retorno, y qué despilfarro se debería eliminar.

Si los jefes de Zapier ven que el uso de *tokens* que hace una persona es cinco veces mayor que el de sus compañeros, tienen curiosidad por saber qué está sucediendo: la persona podría ser muy ineficiente o una superestrella, dependiendo de qué pasó con esos *tokens*.

“Empezamos a sacar conclusiones”, dice Sammut, “ya sea que se trate de un patrón exitoso que queramos multiplicar entre sus pares o de un antipatrón sobre el cual queramos prepararlos para que puedan salir”.

Brian Jabarian, investigador de la Escuela de Negocios Booth de la Universidad de Chicago que estudia cómo las nuevas tecnologías redefinen los lugares de trabajo, afirma que las empresas tienen que empezar a medir el uso de *tokens*.

“Todo el mundo pensaba que solo con el uso de *tokens* de IA tenía un aumento de la productividad, y ya está”, dice. “Pero la realidad es más complicada”.

Suponga que una empresa

ahorra dinero por adelantado mediante el uso de IA como reclutador. Si esta no hace una buena labor, la compañía tiene que pagar a un ser humano para que escoja personal, o gastar más *tokens* de IA para resolver el problema. Cuando una compañía entrega herramientas de IA a 500 mil empleados, agrega, “estos problemas de *tokens* se vuelven prioritarios”.

Algunos trabajadores que saben que deberían utilizar IA podrían pensar que si se abren paso rápidamente a través de los *tokens*, eso los hará merecedores de una medalla de honor. Depende de qué resultados demuestren.

Un alto ingeniero de Vercel, una nueva empresa de plataforma en la nube con IA, desplegó un equipo de agentes de IA para que analizaran un trabajo de investigación y crearan un nuevo servicio de infraestructura crítica basado en esta. Los bots generaron una valiosa base de código en un día; algo que a los humanos les habría llevado semanas, si no meses.

¿La cuenta por ese trabajo? Alrededor de US\$ 10 mil.

“Es un poco como dar a las

personas una manguera para incendios con combustible”, indica Guillermo Rauch, director ejecutivo de Vercel, quien otorga a sus empleados un presupuesto ilimitado de *tokens*.

En Kumo AI, un emprendimiento con cerca de 60 empleados, se empezó a monitorear el uso de *tokens* por ingeniero a principios de año. Hema Raghavan, cofundadora de la compañía, afirma que sus mejores ingenieros utilizan agentes de IA como si tuvieran un ejército de ayudantes.

A algunos ingenieros les gusta esquivar los fines de semana pero sus agentes siguen realizando las tareas.

Raghavan explica que el gasto de *tokens* forma parte de una estructura más amplia de costos de investigación y desarrollo que no se puede evaluar en forma aislada. A veces los altos costos en una parte de la compañía pueden significar ahorros en otra parte.

“Encontramos ejemplos en que los agentes nos ayudaron realmente a escribir códigos más óptimos, así es que eso de verdad ha disminuido nuestros costos de la nube”, asegura.

Mark Hull fundó Exceeds AI,

la que crea un *software* que mide los impactos de la IA en organizaciones de ingeniería. Gran parte de sus clientes están enfocados actualmente en lograr que todos sus empleados utilicen IA. Los más avanzados miden su rendimiento en general de las inversiones en IA, pero no con un nivel máximo de detalle.

Aunque Hull no cree que los empleados vayan a estar viendo un medidor de *token*, sí tendrán que ser cuidadosos con respecto a las formas más eficientes y eficaces de utilizar la IA. Prevé que las empresas van a desarrollar una gobernanza en torno al uso de *tokens*, como establecer límites sobre qué modelos se pueden utilizar para determinadas tareas. Incluso la IA puede ayudar en esto, automatizando esas elecciones.

Hace poco, Hull utilizó Claude Code de Anthropic para desarrollar tres herramientas diferentes de flujo de trabajo, lo que llegó a un total de 300 mil líneas de código aproximadamente. ¿El costo en *tokens*? Alrededor de US\$ 2 mil. Y decidió que todos en su compañía de 15 personas deberían utilizar la plataforma de código en el futuro.

“En un lapso de 48 horas nuestros costos se dispararon”, cuenta, y agrega que desde entonces ha establecido límites.

Rauch, el director ejecutivo de Vercel, afirma que, hasta ahora, sus empleados que más gastan *tokens* son también los de mejor desempeño. Por ahora, se siente tranquilo dejándolos funcionar. Estima que un gasto de US\$ 10 mil por el trabajo de un día probablemente le ha ahorrado millones.

Sin embargo, no será extraño en un futuro cercano que grandes y pequeñas empresas consideren apartar a un empleado, para una conversación en privado, cuando vean un aumento repentino en el uso de *tokens*.

“Podrían estar utilizándolos en proyectos adicionales. Podrían estar utilizándolos por su cuenta, en emprendimientos, en actividades secundarias, cualquier cosa”, comenta. “Va a haber mucho abuso”.

Artículo traducido del inglés por “El Mercurio”.