Tu jefe acaba de instalar un agente de IA en su ordenador: ¿qué puede leer ahora?

Computer use, ChatGPT Atlas, Gemini Deep Research: por qué los agentes autónomos están entrando en las empresas sin pasar por el CISO, y qué deberías estar preguntando antes de que aterricen en tu equipo.

Imagina la escena. Es lunes, son las nueve y media, y tu jefe entra a la oficina con cara de haber descubierto algo. Te dice, casi con orgullo, que ha instalado "una IA que va a ahorrarnos horas a todos". La activa en su ordenador. Le dicta una tarea. Y la pantalla empieza a moverse sola: abre el navegador, lee correos, hace clic en un documento compartido, cambia de pestaña.

Lo que tu jefe no te ha dicho —porque probablemente no lo sabe— es que ese agente de IA también puede ver el Slack abierto en su pantalla. El hilo de correos en el que estás copiado. El dashboard de clientes al que accede por cortesía de su rol. Y, en algunos casos, el documento que estás co-editando con él en tiempo real.

Bienvenido a 2026, el año en que la ciberseguridad dejó de ser un problema de fronteras para convertirse en un problema de gobernanza interna.

El concepto que nadie te ha explicado: qué es realmente "computer use"

Los modelos de IA de frontera —Claude con *computer use*, ChatGPT Atlas, Gemini Deep Research, entre otros— han incorporado en los últimos meses una capacidad que cambia las reglas del juego: la habilidad de **operar un ordenador como lo haría una persona**. No hablamos de un chatbot al que pides un texto. Hablamos de un agente que mira la pantalla, mueve el cursor, hace clic, rellena formularios, salta entre aplicaciones, lee documentos y, sí, envía correos en tu nombre.

La analogía más útil es esta: imagina que contratas a un becario rapidísimo y absolutamente obediente. Lo sientas delante del ordenador del jefe. Le das, sin pensarlo, las llaves de todo lo que tiene abierto. Y le pides que "se ocupe de las tareas pendientes". Ese becario, en realidad, es un sistema capaz de tomar miles de decisiones por minuto, que no se cansa, no pregunta dos veces y no recuerda haber sido contratado.

Lo importante no es lo que hace bien. Es lo que puede ver mientras lo hace.

Lo que ve un agente cuando entra en tu ordenador

La gente subestima sistemáticamente cuánta información sensible tiene abierta en pantalla en cualquier momento del día. Hagamos un inventario rápido y honesto: el correo corporativo con hilos confidenciales, el Slack con conversaciones internas, un Excel con datos de clientes, un Google Doc con la estrategia del próximo trimestre, el navegador con quince pestañas abiertas, el gestor de contraseñas desbloqueado, la VPN conectada al servidor central.

Cuando se activa un agente con *computer use*, todo eso pasa a estar **dentro de su campo de visión operativo**. No porque sea malicioso, sino porque así funciona la tecnología: para poder hacer clic en el botón correcto, primero necesita ver la pantalla completa. La cámara está siempre encendida; lo que cambia es a quién pertenece.

Y aquí aparece el primer problema que casi nadie está discutiendo en serio: los permisos del agente son los del usuario que lo activó. Si tu jefe es director financiero, ese agente hereda, al menos temporalmente, los privilegios del director financiero. Si él tiene acceso a la nómina, el agente lo tiene. Si él tiene acceso a los contratos firmados, el agente también. La IA no necesita escalar privilegios: los recibe en bandeja por delegación.

Del atacante externo al empleado digital invitado

Durante veinte años, la ciberseguridad ha enseñado a las empresas a defenderse de un atacante que estaba **fuera del perímetro**: el hacker que intenta entrar, el correo de phishing que llega desde un dominio falso, el malware que viaja en un USB. Toda la infraestructura defensiva —cortafuegos, antivirus, segmentación de red, controles de acceso— está construida sobre esa lógica.

Los agentes autónomos rompen esa lógica desde dentro. No "entran": son invitados. No vulneran credenciales: las usan legítimamente. No tienen que sortear el cortafuegos: están al otro lado del cortafuegos desde el segundo cero.

Esto significa que la pregunta relevante en 2026 ya no es solo *¿quién está intentando entrar?*, sino, sobre todo, **¿qué está mirando la IA autorizada que vive dentro de mis sistemas?** Es una pregunta distinta, y exige herramientas y políticas distintas. La mayoría de empresas todavía no las tiene, y muchas ni siquiera saben que deberían tenerlas.

Por qué los agentes están entrando sin pasar por el CISO

Aquí está la parte incómoda. En la teoría, cualquier herramienta que acceda a sistemas corporativos debería pasar por el responsable de seguridad: revisión, evaluación de riesgos, configuración de permisos, formación al equipo. En la práctica, los agentes de IA están entrando por tres puertas que sortean ese proceso.

La primera es la **puerta del directivo**. Un alto cargo prueba un agente en su ordenador personal, ve que le funciona, y lo extiende informalmente a su equipo. Como es él quien lo "aprueba", nadie lo cuestiona. Quien manda firma; quien firma instala.

La segunda es la **puerta del proveedor**. Una herramienta que la empresa ya usa —un CRM, un gestor de proyectos, un cliente de correo— incorpora capacidades agénticas en una actualización rutinaria. Nadie la activa explícitamente: simplemente, un martes por la mañana, está ahí. La cláusula que lo permite estaba en los términos del servicio que ningún jurista ha vuelto a leer desde 2023.

La tercera es la **puerta del empleado**. Alguien instala una extensión de navegador o una app de productividad personal en su equipo de trabajo, sin saber que esa herramienta lee la pantalla y envía datos a un modelo externo. No hay mala fe; hay desconocimiento. Y en ciberseguridad, el desconocimiento bien intencionado hace tanto daño como la mala fe.

El resultado es el mismo en los tres casos: agentes con acceso operativo a sistemas sensibles, sin auditoría, sin trazabilidad clara y sin que el equipo de seguridad sepa siquiera que existen.

Riesgos concretos que ya están ocurriendo

No hablamos de escenarios futuros. Hablamos de cosas que ya están pasando.

Un agente activado para "responder correos rutinarios" lee, en el proceso, un hilo confidencial sobre una negociación de fusión. Ese contenido pasa por la API del proveedor del modelo. Aunque no se almacene de forma persistente, ha existido en un sistema externo durante segundos. Para un regulador europeo, eso ya puede constituir una transferencia de datos que exige justificación.

Un agente con permisos para "agendar reuniones" puede ver, mientras lo hace, los nombres y horarios de todos los contactos del calendario. En un despacho de abogados, en una consulta médica o en una redacción de periodismo de investigación, eso no es

metadato neutro: es información sensible cuyo destino debería estar perfectamente trazado.

Un agente comprometido —porque su prompt ha sido manipulado, porque alguien ha explotado una vulnerabilidad del modelo, porque un sitio web malicioso le ha inyectado instrucciones ocultas a través de texto invisible— puede ejecutar miles de acciones por minuto antes de que un humano se dé cuenta. Transferencias, envíos masivos de correos, modificaciones de documentos compartidos. La velocidad, que era una virtud, se convierte en un acelerador del daño.

Y existe un riesgo más sutil, que es el que más me preocupa profesionalmente: la **normalización del acceso indiscriminado**. Cuando los empleados se acostumbran a que una IA "vea su pantalla" para ayudarles, el umbral psicológico de la privacidad digital se desplaza. Lo que hace tres años habría sido inconcebible —que un sistema externo lea en directo todo lo que tienes abierto— hoy se acepta con un clic. Y eso, una vez normalizado, es muy difícil de revertir.

Las preguntas que tienes derecho a hacer antes de aceptar uno

Si en tu empresa están desplegando agentes con *computer use*, o si tu jefe acaba de instalar uno "para probar", estas son las preguntas que tienes derecho —y diría que el deber profesional— de plantear, en voz alta y por escrito:

¿Quién ha autorizado formalmente este despliegue, y existe un documento que lo recoja? ¿Qué datos concretos puede ver el agente, y durante cuánto tiempo permanecen en sistemas del proveedor? ¿Existe un registro auditable de todas las acciones que ejecuta, accesible a posteriori? ¿Qué ocurre si el agente ejecuta una acción incorrecta: hay un mecanismo de reversión y de notificación al afectado? ¿Se ha informado a clientes, pacientes o proveedores cuyos datos puedan ser leídos por el agente? ¿Hay una política clara sobre qué pantallas, aplicaciones o conversaciones quedan **fuera** de su alcance, y cómo se hace cumplir esa política técnicamente?

No son preguntas hostiles. Son las preguntas básicas que cualquier empresa madura debería poder responder antes de poner un agente en producción. Si las respuestas no existen, no es un problema tuyo: es una señal nítida de que el despliegue se ha hecho sin la diligencia mínima.

La ciberseguridad ya no se defiende: se gobierna

Llevo tiempo diciéndolo en formaciones y conversaciones con responsables de seguridad: el problema de fondo en 2026 no es técnico, es **cultural y organizativo**. Las herramientas existen. Las políticas también. Lo que falta es la conciencia colectiva de que un agente autónomo no es un software más, sino una entidad operativa con privilegios delegados que requiere supervisión continua, igual que cualquier nuevo empleado con acceso a información crítica.

Mientras lo sigamos viendo como "una app que ayuda", seguiremos cometiendo el error de instalarlo sin pensar. Cuando lo veamos como **un empleado digital con acceso privilegiado**, empezaremos a hacer las preguntas correctas antes de contratarlo.

Esta transición no la van a liderar los equipos técnicos solos. La tienen que liderar empleados informados, managers con visión y consejos de administración que entiendan que la gobernanza de la IA es ya, de hecho, una parte central de la gobernanza corporativa. No hay departamento que pueda hacerlo en solitario; y, sin embargo, muchas empresas todavía actúan como si la decisión correspondiera solo a "los de informática".

Tu próximo paso

Hoy, antes de cerrar el ordenador, haz tres cosas. Revisa qué herramientas con capacidades agénticas están activas en tu equipo o en tu navegador, aunque sea por aproximación —si una extensión "lee la página" o "te ayuda con tu pantalla", probablemente lo es—. Pregunta abiertamente en tu empresa si existe una política sobre uso de agentes de IA con acceso a sistemas corporativos: si la hay, léela; si no la hay, esa es la conversación que toca abrir. Y comparte este artículo con alguien que esté a punto de instalar uno "porque le ahorra tiempo".

Si te ha resultado útil, compártelo. La conversación sobre agentes autónomos en empresas necesita más voces informadas y menos titulares alarmistas. Visita el blog para acceder a más recursos gratuitos sobre ciberseguridad aplicada, ingeniería social y gobernanza digital.

Isaac Ruiz Romero.