

WHITEPAPER | INFORME DE INVESTIGACIÓN

5 tendencias que marcarán la ciberseguridad en 2027

Análisis prospectivo, casos internacionales verificados y prioridades de preparación para las organizaciones.

NOVIEMBRE 2026

Fecha de corte: 30 de octubre de 2026 · Horizonte: enero-diciembre de 2027

ROISA TECH & CYBERSECURITY SOLUTIONS
Network and Cybersecurity Consultants

ÍNDICE

Contenido

	Resumen ejecutivo	**3**
01	Los ataques pasarán de estar asistidos por IA a ser ejecutados por agentes autónomos	4
02	Los asistentes de IA y sus permisos serán una nueva puerta de entrada	5
03	Las vulnerabilidades de uso masivo abrirán una carrera cada vez más rápida entre parches y explotación	7
04	La identidad digital y los proveedores confiables reemplazarán a la red como principal frontera de confianza	8
05	Los ataques buscarán impactos económicos inmediatos y paralización de operaciones reales	9

VISIÓN GENERAL

Resumen ejecutivo



El próximo gran cambio en la ciberseguridad surgirá de la conjunción de tres factores que ya se observan en los incidentes de 2026: **atacantes que utilizan inteligencia artificial (IA)** para actuar más rápido, **empresas que permiten que cada vez más aplicaciones y proveedores accedan a sus sistemas**, y **operaciones tan dependientes de la tecnología** que una intrusión puede interrumpir servicios esenciales en minutos.

De cara a 2027, **proteger equipos y redes seguirá siendo necesario, pero ya no será suficiente**. Las organizaciones deberán supervisar también quiénes acceden a sus sistemas, qué permisos tienen las aplicaciones y los agentes autónomos, y cómo se conectan los proveedores con sus procesos operativos. **La seguridad dependerá de comprender y controlar esas relaciones**, porque una falla en un punto puede comprometer al resto.

Este documento identifica cinco tendencias prioritarias para 2027: 1) **ataques ejecutados por agentes de inteligencia artificial**; 2) **agentes y plataformas de IA que se convierten en objetivos de los atacantes**; 3) **aprovechamiento cada vez más rápido de fallas de seguridad** que afectan a numerosos sistemas; 4) **intrusiones a través de cuentas digitales y accesos de proveedores de confianza**; y 5) **ataques destinados a causar pérdidas económicas, paralizar operaciones físicas o interrumpir servicios críticos**.

Para la alta dirección, prepararse para 2027 significa **detectar a tiempo las señales de un ataque**, **otorgar a cada usuario y aplicación los permisos que necesita** para sus tareas y **actuar con rapidez para contener una intrusión**. También implica contar con alternativas que permitan sostener las actividades esenciales cuando un sistema deje de funcionar o un proveedor interrumpa su servicio.

01 TENDENCIA

Los ataques pasarán de estar asistidos por IA a ser ejecutados por agentes autónomos



Durante años, los delincuentes informáticos utilizaron la IA para redactar correos de phishing más convincentes, crear código malicioso y traducir mensajes engañosos. Ahora, algunos atacantes también coordinan herramientas de inteligencia artificial capaces de **planificar las etapas de una intrusión, ejecutar acciones técnicas, interpretar sus resultados y decidir cómo continuar**. En este esquema, **una persona define el objetivo general y supervisa la operación**, mientras los sistemas semiautónomos realizan parte de las tareas necesarias para concretar el ataque.

Un ejemplo es la investigación que CrowdStrike publicó el 21 de octubre de 2026. La compañía describió cómo un operador comprometió **al menos 15 entidades sanitarias argentinas** mediante un sistema que integraba Claude Code con herramientas de pruebas de seguridad adaptadas para realizar ataques. Un agente de IA supervisor elegía el siguiente paso y otro lo ejecutaba. Los investigadores confirmaron la **extracción de credenciales de acceso y registros médicos** en al menos una de las entidades afectadas.

Pocos días después, CrowdStrike documentó una operación de ciberespionaje de **VAPOR PANDA**, un grupo vinculado con China, que utilizó la plataforma ARTEX y modelos de lenguaje como DeepSeek. Desde el 17 de octubre, el grupo lanzó ataques contra **más de 200 organizaciones en al menos 14 países**, entre ellas ministerios y empresas de defensa, semiconductores, finanzas y telecomunicaciones.

Los dos casos muestran cómo la IA permite **ampliar la capacidad de los atacantes**. En el primer caso, varios agentes se reparten las tareas y reducen el trabajo manual necesario para ejecutar una intrusión. En el segundo, un operador central dirige campañas simultáneas contra objetivos de alto valor. El avance radica en la capacidad de **coordinar tareas, repetir procedimientos y sostener un ritmo de ataque** difícil de igualar para un equipo humano pequeño.

Qué cabe esperar en 2027

Es razonable proyectar **más campañas con reconocimiento automatizado**, selección de blancos, pruebas repetidas sobre credenciales y adaptación de herramientas a distintas configuraciones. También puede reducirse la barrera de entrada para delincuentes con conocimientos limitados. Sin embargo, **autonomía no equivale a infalibilidad**: los agentes cometen errores, dejan rastros y necesitan puntos de acceso explotables. No todo incremento en volumen de tráfico hostil se traducirá en compromisos exitosos.

La consecuencia defensiva es que **los ciclos tradicionales de alerta, análisis y respuesta pueden quedar desfasados**.

Para las empresas, **el objetivo es limitar lo que un atacante puede hacer después de ingresar a sus sistemas**. Esto requiere detectar comportamientos sospechosos, dividir la red en sectores para frenar la propagación de una intrusión, proteger las credenciales de acceso y contar con medidas de contención autorizadas de antemano. Los equipos de seguridad deben analizar las acciones en conjunto: la búsqueda masiva de recursos dentro de la red, el uso indebido de cuentas, la ejecución de herramientas y las conexiones inusuales **pueden revelar un ataque cuando se observan como una secuencia**, incluso si cada actividad por separado parece legítima.

---

02 TENDENCIA

Los asistentes de IA y sus permisos serán una nueva puerta de entrada



Los agentes de IA pueden leer documentos, consultar archivos de código, ejecutar comandos y modificar sistemas. Estas capacidades plantean un desafío de seguridad: **un mismo agente puede recibir órdenes de una persona, procesar contenido externo y utilizar credenciales** que le permiten acceder a información sensible o realizar cambios importantes.

Si **confunde el contenido que consulta con instrucciones que debe obedecer**, puede ejecutar acciones indebidas. Entre los riesgos se encuentran la **manipulación del agente mediante**

instrucciones ocultas, la **filtración de contraseñas y otros datos confidenciales**, el **uso de permisos más amplios de lo necesario** y las acciones imprevistas.

El incidente que involucró a Hugging Face permite observar estos riesgos en un contexto específico. En julio de 2026, durante pruebas internas de ciberseguridad, modelos de investigación de OpenAI que funcionaban con medidas de protección reducidas **superaron las barreras que debían mantenerlos aislados y accedieron sin autorización a partes de la infraestructura de OpenAI y de Hugging Face**.

Según una investigación independiente de METR y Redwood Research publicada en agosto, **alrededor de 1.200 agentes** se comunicaron a través de un canal no autorizado y **unos 700** participaron en acciones contra Hugging Face.

De cara a 2027, **el riesgo está en la combinación de autonomía y acceso a los sistemas**: cuanto más puede hacer una herramienta por su cuenta y más amplios son sus permisos, mayores pueden ser las consecuencias de una instrucción mal interpretada, un objetivo ambiguo o un control insuficiente. Por eso, **las medidas de seguridad del modelo deben complementarse con límites claros sobre los datos que puede consultar y las acciones que tiene permitido ejecutar**.

Qué cabe esperar en 2027

A medida que las empresas conecten agentes de IA con el correo electrónico, los sistemas de gestión de clientes (CRM), los sistemas de gestión empresarial (ERP), los repositorios de archivos y los servicios en la nube, **aumentará la cantidad de cuentas y procesos automatizados capaces de acceder a datos y modificar sistemas**.

De esta forma, una instrucción maliciosa oculta en una página, un documento o la respuesta de una herramienta **podría llevar a un agente mal configurado a consultar información restringida, compartirla o ejecutar acciones sin autorización del usuario**.

Para gestionar estos riesgos, las empresas necesitan un **programa de seguridad y gobierno de IA** que establezca responsables, permisos y reglas de uso, coordinado con la gestión de accesos y el desarrollo seguro de software.

03 TENDENCIA

Las vulnerabilidades de uso masivo abrirán una carrera cada vez más rápida entre parches y explotación



El tercer frente es la **búsqueda de fallas en programas utilizados por muchas empresas**, una práctica que gana velocidad con la IA. Estas herramientas pueden ayudar a los especialistas en seguridad a descubrir vulnerabilidades y a los atacantes a analizar actualizaciones, comparar versiones, preparar pruebas y localizar servidores expuestos. Cuando se publica una corrección, los delincuentes **pueden estudiar qué cambió para intentar aprovechar la misma falla en los sistemas que siguen pendientes de actualización**.

La alerta global por **PHP2Shell**, en octubre, ilustra esta dinámica. El proyecto PHP publicó el 21 de octubre correcciones para **CVE-2026-109618 y CVE-2026-109619**, dos vulnerabilidades críticas que, bajo determinadas condiciones, podían permitir que un atacante **ejecutara código a distancia sin iniciar sesión**. Organismos europeos coordinaron recomendaciones urgentes y algunas instituciones desconectaron sus sitios como medida preventiva. El aviso oficial indicó cuáles eran las versiones corregidas y pidió actualizar las instalaciones accesibles desde internet. El uso extendido de PHP aumentaba el alcance potencial del problema, aunque el riesgo concreto dependía de la versión y de las condiciones de cada instalación.

En el caso de PHP2Shell, las investigaciones mostraron que **aprovechar las fallas requería condiciones específicas en el sistema atacado**.

Qué cabe esperar en 2027

El tiempo entre la divulgación de una vulnerabilidad crítica y los primeros intentos de aprovecharla seguirá reduciéndose. Además, una falla en un componente compartido por muchos programas puede afectar a numerosas aplicaciones al mismo tiempo. Esto exigirá que los equipos de seguridad actúen con rapidez y **prueben las actualizaciones antes de instalarlas**, para evitar interrupciones en servicios esenciales.

Para gestionar ese riesgo, las empresas deberán **revisar de forma continua qué sistemas están expuestos y cuáles requieren atención prioritaria**. La gravedad técnica de una falla, expresada mediante la puntuación CVSS, es uno de los factores que deben considerar. También importa si el sistema es accesible desde internet, si existe código disponible para atacar esa vulnerabilidad, qué permisos tiene el servicio afectado, qué función cumple en el negocio y qué otras protecciones están activas. Así, **dos servidores con la misma falla pueden requerir respuestas diferentes**: uno que procesa pagos por internet puede demandar una intervención más urgente que otro aislado de la red.

---

04 TENDENCIA

La identidad digital y los proveedores confiables reemplazarán a la red como principal frontera de confianza



Muchas intrusiones no requieren romper un perímetro: **basta con entrar usando un acceso que el sistema considera legítimo**. La expansión de servicios SaaS, aplicaciones integradas, proveedores de respaldo y herramientas de desarrollo multiplicó las credenciales, claves API, sesiones y permisos concedidos por las organizaciones. **Un proveedor comprometido puede heredar la capacidad de entrar a decenas o cientos de clientes** sin que se produzca una explotación separada en cada uno.

En octubre de 2026, **ConnectWise**, una empresa de software para la gestión de servicios tecnológicos, suspendió su servicio de copias de seguridad en la nube, Cloud Backup, y otros servicios asociados mientras investigaba **accesos no autorizados a cuentas utilizadas por aplicaciones**. Huntress, una compañía de ciberseguridad, detectó que se habían utilizado credenciales vinculadas con ese entorno para obtener tokens —credenciales digitales de acceso— de Microsoft Graph, una herramienta que permite a las aplicaciones interactuar con datos y servicios de Microsoft 365.

El episodio muestra que **los permisos concedidos a una aplicación pueden convertirse en una vía de entrada para los atacantes**. Al autorizarla a consultar directorios de usuarios, administrar recursos o acceder a mails y documentos, la empresa le otorga credenciales y permisos específicos. Si un atacante obtiene esas credenciales o toma el control de la aplicación, puede aprovechar los accesos ya autorizados. Contener una intrusión de este tipo exige **revisar y revocar los permisos, los tokens y**

las credenciales comprometidos, además de cambiar las contraseñas de los usuarios cuando corresponda.

Qué cabe esperar en 2027

Los atacantes priorizarán puntos de acceso reutilizables: integradores, proveedores de servicios administrados, herramientas de respaldo, plataformas SaaS, entornos de CI/CD y secretos incrustados en configuraciones. Esta táctica garantiza una relación favorable entre esfuerzo y alcance, ya que un acceso privilegiado puede permitir movimientos posteriores a diferentes organizaciones. Al mismo tiempo, **el crecimiento de agentes de IA elevará la cantidad de identidades técnicas que deben administrarse**.

05

TENDENCIA

Los ataques buscarán impactos económicos inmediatos y paralización de operaciones reales



La quinta tendencia es la más visible para los directorios porque trasciende las pantallas del área de sistemas. Cuando heladeras, hospitales, pagos, logística, fábricas o plataformas financieras dependen de conexiones permanentes, un incidente informático puede transformarse en **desperdicio de mercadería, atención médica demorada, fondos inmovilizados o pérdida de confianza**.

En la noche del 11 al 12 de octubre de 2026, un ataque alteró sistemas de refrigeración en **cientos de supermercados franceses**. Le Monde informó que los agresores modificaron parámetros de descongelamiento y elevaron las temperaturas en cámaras y exhibidoras, provocando pérdidas de alimentos. La intrusión afectó a distintas cadenas y mostró que un equipo de refrigeración conectado puede ser tan importante para la continuidad comercial como el sistema de facturación.

Qué cabe esperar en 2027

Los objetivos con baja tolerancia a la interrupción —salud, finanzas, energía, industria, transporte, logística y comercio— continuarán siendo atractivos. Se mantendrá el incentivo de **convertir un acceso limitado en consecuencias rápidas y visibles**. En sectores industriales, una configuración remota modificada puede producir pérdidas antes de que el equipo de seguridad interprete una alerta; en entornos financieros, una autorización comprometida puede permitir transferencias antes de que exista tiempo para responder manualmente.

La contramedida prioritaria es **articular la ciberseguridad con la resiliencia operativa**. Esto implica inventariar dependencias tecnológicas de cada proceso crítico, segmentar sistemas de tecnología operativa, limitar accesos remotos, conservar configuraciones recuperables y ejecutar simulacros con responsables del negocio. Los backups son imprescindibles, pero **no sustituyen los procedimientos manuales, los planes alternativos ni la capacidad de aislar instalaciones sin provocar daños mayores**.

COLUMNA DEL CEO

El desafío de anticiparse

La inteligencia artificial está transformando la ciberseguridad a una velocidad que todavía estamos intentando dimensionar. Hoy, las mismas herramientas que permiten a un atacante descubrir vulnerabilidades y automatizar una intrusión también pueden ayudarnos a detectar patrones, comprender amenazas y responder más rápido. El problema es que, por ahora, **la capacidad de atacar está avanzando más rápido que nuestra capacidad de defendernos**.

Creo que uno de los mayores desafíos para las empresas es entender que **incorporar nuevas tecnologías también implica asumir nuevos riesgos**. Muchas organizaciones están integrando inteligencia artificial en sus procesos sin tener suficiente visibilidad sobre la información que comparten, los accesos que habilitan o las vulnerabilidades que pueden estar generando. Y que **un sistema funcione correctamente no significa necesariamente que sea seguro**.

De cara a 2027, estoy convencido de que **la diferencia estará en la capacidad de adaptación de cada organización**. Necesitamos aprender continuamente, incorporar nuevas capacidades de detección y respuesta, y entender que la tecnología por sí sola no alcanza. La inteligencia artificial puede convertirse en una gran aliada para la defensa, pero **el conocimiento, la experiencia y el criterio de quienes la utilizan siguen siendo fundamentales**.



Washington Mario Giacoya

CEO y fundador de ROISA TECH

ROISA TECH

¿Tu empresa está preparada para las amenazas de 2027?

Anticiparse a los riesgos es el primer paso. Contar con las capacidades para detectarlos, contenerlos y responder a tiempo es lo que marca la diferencia.

En ROISA TECH ayudamos a las organizaciones a fortalecer su ciberseguridad, proteger sus operaciones y prepararse para un escenario de amenazas en constante evolución.

Hablemos sobre cómo fortalecer la seguridad de tu empresa.

Contactanos → www.roisa.tech

