Las 5 preguntas que más se repiten cuando hablo de IA y ciberseguridad (y lo que nadie te cuenta sobre ellas)

Subtítulo: Inteligencia artificial, amenazas digitales y decisiones humanas: guía para entender el nuevo escenario sin caer en el pánico ni en la ingenuidad.

Introducción: la pregunta detrás de todas las preguntas

Cuando hablo de ciberseguridad —en formaciones, en conversaciones con clientes o en redes sociales— noto que las dudas se repiten. No importa si estoy hablando con el director de una pyme, una madre preocupada por las apps que usa su hijo o un profesional que empieza a oír hablar de deepfakes. La inquietud de fondo es siempre la misma: *¿hasta qué punto estoy expuesto y qué puedo hacer sin ser un experto?*

La irrupción de la inteligencia artificial ha amplificado esa inquietud. Y con razón. La IA no solo ha cambiado lo que los atacantes pueden hacer; ha cambiado la velocidad a la que pueden hacerlo y el nivel de sofisticación que necesitan para lograrlo. Pero también —y esto se dice menos— ha abierto posibilidades enormes para defender mejor.

A continuación, respondo las cinco preguntas que más escucho. No para alarmar, sino para que tomes decisiones informadas.

1. "¿Puede la IA hacerme un ataque personalizado a mí, que no soy nadie famoso?"

Sí. Y eso es exactamente lo que la hace peligrosa para el usuario medio.

Durante años, los ataques masivos funcionaban como redes de pesca: correos genéricos, mensajes torpes llenos de errores ortográficos, URLs sospechosas. Era fácil entrenar el ojo para detectarlos. Hoy, un modelo de lenguaje puede generar en segundos un mensaje de phishing perfectamente redactado, adaptado a tu sector profesional, con tu nombre, mencionando tu empresa y simulando el estilo de tu banco o de tu proveedor habitual.

¿De dónde saca esa información? De lo que tú mismo has publicado. Tus redes sociales, tu perfil de LinkedIn, las reseñas que has dejado en Google, los comentarios en foros. Todo eso es OSINT —inteligencia de fuentes abiertas— y los atacantes llevan años usando esta técnica de forma manual. La IA simplemente automatiza y escala ese proceso a un coste casi cero.

El escenario típico: recibes un correo que parece ser de tu gestoría, con tu nombre, referenciando una factura real de un proveedor que nombras en tu web. El asunto es urgente. El enlace, malicioso. Sin entrenamiento previo, es muy difícil no caer.

La conclusión no es vivir en paranoia. Es entender que la fricción —ese segundo de duda antes de hacer clic— se ha vuelto más valiosa que nunca.

2. "Mi empresa es pequeña. ¿De verdad somos un objetivo?"

Esta es probablemente la creencia más peligrosa en el ecosistema de pymes español.

Los atacantes no eligen objetivos por tamaño; los eligen por rentabilidad y accesibilidad. Una empresa de 12 empleados puede tener acceso a datos bancarios, información de clientes, credenciales de proveedores y sistemas sin parchear. Y a diferencia de las grandes corporaciones, raramente cuenta con un equipo de seguridad dedicado, políticas de acceso claras o copias de seguridad verificadas.

Lo que ha cambiado con la IA es que el coste de lanzar un ataque sofisticado ha caído dramáticamente. Ya no hace falta un grupo de ciberdelincuentes especializados trabajando durante semanas. Hay herramientas automatizadas —algunas incluso disponibles como servicio en mercados ilegales— que pueden escanear vulnerabilidades, generar correos de ingeniería social adaptados y gestionar el proceso de extorsión de forma casi autónoma.

El ransomware —ese tipo de ataque que cifra todos tus archivos y pide un rescate para recuperarlos— afecta proporcionalmente más a pymes que a grandes empresas, precisamente porque la probabilidad de que paguen es mayor y la capacidad de resistencia es menor. Un negocio que no puede acceder a sus datos durante cinco días tiene un problema de supervivencia.

La buena noticia es que la mayoría de ataques exitosos explotan fallos básicos: contraseñas débiles, software desactualizado, ausencia de doble factor de autenticación, empleados sin formación. No hace falta una inversión desorbitada para cerrar esas puertas.

3. "¿Los deepfakes son reales o están exagerados?"

Son reales, están en uso y evolucionan más rápido de lo que la mayoría imagina.

Un deepfake es una simulación audiovisual generada por IA que imita la voz, el rostro o ambos de una persona real. Hace tres años, producirlos requería hardware potente y conocimientos técnicos avanzados. Hoy existen aplicaciones que clonan una voz con menos de un minuto de audio de referencia y generan en tiempo real una llamada que suena idéntica a la persona original.

El escenario más utilizado en entornos empresariales se conoce como fraude del CEO: alguien llama al departamento financiero simulando ser el director general, con urgencia, pidiendo una transferencia a una cuenta nueva. La persona al otro lado escucha la voz de su jefe. No sospecha. Transfiere.

En el entorno familiar, el vector más preocupante son los menores. Sus voces, imágenes y hábitos son mucho más accesibles —a través de redes sociales, grupos de WhatsApp o plataformas de juego— y son el público menos equipado para evaluar críticamente lo que reciben.

La defensa aquí no es tecnológica en primer lugar; es cultural. Establecer protocolos de verificación —una palabra clave acordada en familia, un proceso de doble confirmación en empresa para transferencias— vale más que cualquier software de detección de deepfakes.

4. "¿La IA que usamos en el trabajo puede filtrar información de nuestra empresa?"

Depende de cómo la uses, y la respuesta honesta es que muchas empresas no lo están usando bien.

Cuando un empleado introduce información confidencial en un modelo de lenguaje público —un contrato, datos de clientes, una propuesta estratégica— esa información sale del entorno controlado de la empresa y entra en sistemas externos cuyos términos de uso pocas personas leen. En algunos casos, esos datos pueden usarse para entrenar modelos futuros. En otros, simplemente están expuestos a brechas de seguridad en terceros.

Esto no significa que no deba usarse IA en el trabajo. Significa que hay que hacerlo con políticas claras: qué información puede introducirse, en qué herramientas, bajo qué condiciones. Las empresas que están gestionando bien este tránsito son las que han tratado la adopción de IA no como un fenómeno espontáneo, sino como una decisión organizativa con implicaciones de seguridad y cumplimiento normativo.

El RGPD europeo ya regula aspectos relacionados con el tratamiento de datos personales mediante IA. No es un detalle menor: las sanciones por incumplimiento pueden ser significativas, y la responsabilidad recae sobre la organización, no sobre el empleado que usó la herramienta.

5. "¿Qué puedo hacer yo, sin ser técnico, para protegerme?"

Más de lo que crees. Y la mayor parte no requiere ningún conocimiento técnico.

La ciberseguridad tiene una dimensión humana que ningún software cubre completamente. La ingeniería social —el arte de manipular a las personas para que cedan información o realicen acciones— es el vector de entrada en la mayoría de ataques exitosos, no las vulnerabilidades técnicas. Lo que eso significa en la práctica es que entrenar el juicio humano es tan importante como instalar un antivirus.

Tres hábitos que marcan una diferencia real: usar contraseñas únicas para cada servicio (gestionadas con un gestor de contraseñas), activar la autenticación en dos pasos siempre que sea posible, y desarrollar el hábito de verificar antes de actuar cuando algo genera urgencia —ya sea un correo, un mensaje o una llamada. La urgencia artificial es la herramienta favorita de la ingeniería social porque desconecta el pensamiento crítico.

En el ámbito familiar, la conversación más importante no es sobre qué aplicaciones están prohibidas, sino sobre cómo funciona la manipulación digital: por qué ciertos mensajes están diseñados para generar miedo, ilusión o prisa, y cómo reconocer esas señales. Un menor que entiende la mecánica del engaño digital es mucho más resistente que uno que simplemente tiene una lista de apps vetadas.

Reflexión final: el problema no es la IA, es la asimetría

Hay una frase que repito mucho en mis formaciones: *la IA no ha creado nuevos problemas de seguridad, ha acelerado y democratizado los que ya existían.* El phishing existía antes de ChatGPT. La ingeniería social es tan antigua como la comunicación humana. Los deepfakes son una evolución de técnicas de falsificación que llevan décadas existiendo.

Lo que ha cambiado es la asimetría. Los atacantes tienen acceso a herramientas muy potentes, de bajo coste y fáciles de usar. El ciudadano medio, la familia, la pyme, siguen funcionando con los mismos patrones de comportamiento digital de hace diez años. Cerrar esa brecha no es un problema tecnológico; es un problema de cultura, de educación y de decisión estratégica.

Y esa es exactamente la conversación que necesitamos tener.

¿Quieres profundizar?

Si este artículo te ha generado preguntas concretas sobre tu situación —como empresa, como familia o como profesional—, **escríbeme directamente**. En el blog encontrarás más recursos gratuitos sobre ingeniería social, privacidad digital y gestión del riesgo humano.

Comparte este artículo si conoces a alguien que crea que "esto no va con él". Porque esa creencia es, en sí misma, el primer vector de riesgo.

Isaac Ruiz Romero.