La magia de los ciberataques

Descripción

Introducción

Hace unos días tuve la oportunidad de participar como ponente en la Semana de la Seguridad de la Información 2026 de la SUNAT, la Superintendencia Nacional de Aduanas y de Administración Tributaria de Perú.

Más de 1.100 personas siguieron la conferencia *Cómo la IA está transformando la amenaza y la defensa*, dedicada a explorar una cuestión que me interesa especialmente: qué ocurre cuando cruzamos Inteligencia Artificial, ciberseguridad y comportamiento humano.

La SUNAT es una institución especialmente interesante para abordar este tema. Administra el sistema tributario y aduanero peruano y desarrolla buena parte de su actividad en un entorno en el que conviven información sensible, procedimientos, análisis de datos, servicios digitales y decisiones con consecuencias para ciudadanos, empresas y recursos públicos.

Por eso, hablar allí de seguridad de la información no podía limitarse a hablar de tecnología. Y esa fue precisamente la idea central de mi intervención, porque quizá uno de los cambios más importantes que está introduciendo la Inteligencia Artificial en la ciberseguridad no tenga que ver exclusivamente con lo que las máquinas pueden hacer. Tiene que ver con lo que están aprendiendo sobre nosotros.

Cuando el atacante se comporta como un ilusionista

Hay una analogía que utilicé durante la conferencia en SUNAT y que me parece especialmente útil para entender muchos ciberataques. Un buen atacante se parece bastante a un buen ilusionista.

Cuando un mago realiza un truco no necesita hacer desaparecer realmente aquello que no quiere que veamos, le basta con dirigir nuestra atención hacia otro lugar.

Los ataques digitales explotan un principio parecido. El truco no consiste necesariamente en esconder la amenaza, consiste en desviar nuestra atención y que miremos exactamente donde interesa al atacante.

Durante años hemos asociado la ciberseguridad con cortafuegos, contraseñas, antivirus, autenticación, cifrado o sistemas de detección. Todo ello sigue siendo imprescindible, pero existe otra superficie de ataque bastante más difícil de parchear, la de nuestra forma de tomar decisiones.

Los ataques de ingeniería social llevan décadas aprovechando mecanismos muy humanos. La confianza. La autoridad. El miedo. La urgencia. La expectativa de una recompensa. La familiaridad. La necesidad de terminar una tarea. La tendencia a aceptar aquello que encaja con lo que esperábamos encontrar.

Nada de esto lo ha inventado la Inteligencia Artificial. Lo que está cambiando es la capacidad para utilizarlo.

Los modelos generativos permiten producir mensajes convincentes, adaptar el lenguaje a diferentes personas, recopilar y relacionar información pública, imitar estilos comunicativos y generar imágenes, audio o vídeo sintéticos con un coste y una velocidad que hace pocos años resultaban difíciles de imaginar.

La evidencia reciente apunta precisamente en esa dirección. La IA parece estar aumentando sobre todo la escala, eficiencia y personalización de técnicas de ingeniería social que ya existían, más que inventando mecanismos psicológicos completamente nuevos. El National Cyber Security Centre británico ha llegado a una conclusión similar en sus evaluaciones prospectivas sobre IA y amenaza cibernética.

Un experimento publicado en 2026 resulta especialmente ilustrativo. Investigadores compararon campañas de *spear phishing* creadas por expertos humanos con otras

automatizadas mediante modelos de lenguaje. En una muestra experimental de 101 participantes, los mensajes generados mediante IA alcanzaron tasas de interacción comparables a las conseguidas por expertos humanos. Es un estudio concreto, con las limitaciones propias de su tamaño y diseño, pero sirve para demostrar que la automatización de ataques personalizados ha dejado de ser únicamente una posibilidad teórica.

La pregunta importante, por tanto, ya no es únicamente si podremos reconocer un correo mal escrito, es qué ocurrirá cuando el correo esté perfectamente escrito. Cuando conozca nuestro nombre, cuando utilice un lenguaje familiar, cuando aparezca en el momento adecuado o cuando venga acompañado por una voz que creemos reconocer.

Saber mucho no nos hace inmunes

Hay otra idea que quise trabajar durante la conferencia en SUNAT. Podemos conocer perfectamente las reglas y, aun así, equivocarnos.

No porque seamos incompetentes, sino porque las decisiones no se toman en el vacío.

Tomamos decisiones mientras contestamos correos, atendemos llamadas, resolvemos incidencias, cambiamos de tarea, tenemos una reunión dentro de diez minutos o intentamos terminar algo antes de marcharnos y nuestro comportamiento cambia con el

contexto.

La investigación sobre susceptibilidad al phishing está mostrando precisamente que el fenómeno es bastante más complejo que separar a las personas entre quienes “saben de ciberseguridad” y quienes no.

Una revisión sistemática publicada recientemente en *Computers & Security* analizó 54 estudios sobre susceptibilidad al phishing. Una de sus conclusiones más interesantes es que intervienen conjuntamente factores individuales, ambientales y conductuales, y que existe además una distancia importante entre saber lo que deberíamos hacer y hacerlo realmente cuando llega el momento.

Los autores advierten, de hecho, de cierto desajuste entre muchas estrategias tradicionales de concienciación, centradas principalmente en transmitir conocimiento, y la naturaleza multifactorial del problema.

Esto cambia bastante nuestra manera de entender la formación.

Una buena formación en ciberseguridad no debería aspirar únicamente a enseñar una colección de trucos para reconocer correos falsos.

Los ataques cambiarán, los diseños cambiarán, las herramientas cambiarán y con la Inteligencia Artificial probablemente lo harán cada vez más deprisa. Por eso es necesario ayudar a las personas a reconocer qué está ocurriendo en su propia toma de decisiones.

No se trata de convertir a cada trabajador en especialista en ciberseguridad, se trata de proporcionarle mejores herramientas para decidir.

Ese fue precisamente uno de los objetivos de la sesión que desarrollé para SUNAT, la de trasladar conceptos procedentes de la psicología, las ciencias del comportamiento, la neurociencia y la ciberseguridad a situaciones reconocibles del trabajo cotidiano.

La IA también puede ayudarnos a defendernos

Sería un error, sin embargo, presentar la Inteligencia Artificial únicamente como una herramienta para los atacantes.

La misma capacidad de procesar enormes cantidades de información permite identificar anomalías, relacionar datos, detectar patrones difíciles de observar manualmente, priorizar señales o asistir a profesionales que deben analizar situaciones complejas.

En una organización como SUNAT esta cuestión resulta especialmente relevante.

Una operación aislada puede parecer normal, una factura puede parecer normal, una empresa puede parecer normal y una declaración puede parecer normal, pero al relacionarlas puede aparecer un patrón que merezca nuestra atención.

Aquí existe una distinción fundamental sobre la que también insistí durante la conferencia. Anómalo no significa culpable. Una anomalía puede decirnos dónde mirar, pero no necesariamente qué concluir.

Este principio es aplicable mucho más allá de la administración tributaria. También afecta a la prevención del fraude, la criminología, la medicina, los recursos humanos, la inteligencia empresarial, la ciberseguridad o cualquier otro entorno en el que utilicemos algoritmos para apoyar decisiones.

Cuanta más capacidad tengan las máquinas para seleccionar información, mayor será la necesidad de que las personas sepamos interpretarla, contextualizarla y cuestionarla.

Por eso me interesa especialmente otra forma de utilizar la Inteligencia Artificial, no solo como generadora de respuestas, sino también como herramienta para mejorar nuestras preguntas.

Podemos pedirle que busque explicaciones alternativas, que identifique supuestos que no hemos comprobado, que intente refutar nuestra hipótesis o que señale qué información podría hacernos cambiar de opinión.

La IA deja de ser un oráculo al que preguntamos qué pensar y comienza a convertirse en un interlocutor que nos ayuda a pensar mejor. Y eso también forma parte de la ciberseguridad.

Conclusión

La participación durante la sesión y los numerosos mensajes que recibí posteriormente desde SUNAT me dejaron una sensación especialmente positiva, pero también reforzaron una convicción.

Durante mucho tiempo hemos intentado proteger a las organizaciones construyendo mejores barreras tecnológicas. Debemos seguir haciéndolo. Sin embargo, la evolución de la Inteligencia Artificial obliga a añadir una capa más.

Necesitamos organizaciones capaces de detectar amenazas, pero también personas capaces de reconocer cuándo alguien, o algo, está intentando condicionar su atención, sus emociones y sus decisiones.

Eso requiere tecnología, requiere procedimientos, requiere cultura y requiere formación. No una formación basada en asustar al usuario ni en recordarle periódicamente que puede convertirse en “el eslabón más débil”, todo lo contrario, sino una formación que le permita comprender por qué una persona competente y experimentada puede equivocarse, qué condiciones aumentan ese riesgo y qué pequeñas modificaciones en nuestra manera de verificar, detenernos y decidir pueden ayudarnos a reducirlo.

La Inteligencia Artificial seguirá evolucionando, los ataques también, pero nuestro cerebro, en cambio, no cambiará a la misma velocidad.

Referencias

- European Union Agency for Cybersecurity. (2025). *ENISA Threat Landscape 2025*. ENISA.
- Heiding, F., Lermen, S., Kao, A., Mayrink Verdun, C., Schneier, B., & Vishwanath, A. (2026). Evaluating large language models’ ability to automate spear phishing. *Expert Systems with Applications, 314*, 131546. https://doi.org/10.1016/j.eswa.2026.131546
- National Cyber Security Centre. (2025). *Impact of AI on cyber threat from now to 2027*. NCSC.
- Schmitt, M., & Flechais, I. (2024). Digital deception: Generative artificial intelligence in social engineering and phishing. *Artificial Intelligence Review, 57*, 324. https://doi.org/10.1007/s10462-024-10973-2
- Spicer, M., & Nurse, J. R. C. (2026). What makes people susceptible to phishing attacks: A systematic literature review and future research agenda. *Computers & Security*. Advance online publication. https://doi.org/10.1016/j.cose.2026.105096
- Superintendencia Nacional de Aduanas y de Administración Tributaria. (2026). *Misión, visión y políticas institucionales*. SUNAT.