BRÚJULA COTIDIANA
Director Riccardo Cascioli
HECHOS PARA LA VERDAD

El caso

¿Estafa o error?: la IA y la confianza traicionada

CRóNICA 29_09_2026

Daniele Ciacci

(Artículo traducido con Deepl y sin revisar) — La semana pasada, dos noticias sacaron a la luz un mecanismo análogo pero especular. Por un lado, la IA se convierte en una herramienta de los estafadores para potenciar y ampliar las posibilidades de éxito y el alcance de su actividad delictiva. En el otro extremo, la inteligencia artificial se convierte, por el contrario, en un sujeto activo, que actúa tejiendo una red compleja y rápida de

acciones destinadas a alcanzar su objetivo, ocultándolo a los expertos que observan su actividad. En ambos casos, sin embargo, en el centro del ring es la confianza la que acaba noqueada.

Hablamos de Fideuram, la entidad financiera de Banca Intesa. Según la reconstrucción del *Corriere della Sera*, que luego se hizo eco en muchos medios de la prensa nacional, en febrero el entonces presidente Paolo Molesini recibió un mensaje en WhatsApp que parecía enviado por Carlo Messina, consejero delegado de Intesa Sanpaolo, con una solicitud de pagos urgentes mediante transferencia bancaria para una operación en el extranjero. A esto le sigue una llamada telefónica con la voz clonada mediante IA del abogado Paolo Nastasi, conocido de Molesini, y luego varios correos electrónicos con los datos bancarios. El director financiero cumple las órdenes y transfiere unos 95 millones, casi todos con destino a China y Hong Kong.

Se trata, evidentemente, de una estafa. Más de 40 millones son devueltos por un banco chino, otros 13 millones son incautados en Portugal, mientras que al menos 36 millones circulan ahora sin poder ser rastreados en forma de criptomonedas. Molesini no figura como investigado y los sistemas informáticos del banco no han sido violados, lo cual es plausible porque nadie ha tenido que forzar cortafuegos ni nada por el estilo. Fue el propio expresidente quien, ingenuamente, abrió la puerta a los delincuentes.

Veamos otro caso, en el que la máquina adopta una nueva sintaxis y pasa de ser un medio a convertirse en sujeto. Tras una investigación del *Wall Street Journal*, Google confirmó que, en mayo, durante un simulacro de seguridad, un modelo Gemini obtuvo acceso a Internet debido a un error de configuración. Así, se coló en los sistemas de tres empresas reales, con los mismos nombres que las ficticias utilizadas en la prueba. En un caso, descubrió una contraseña; en los otros dos, utilizó credenciales que alguien había dejado en archivos públicos. Aprovechando errores humanos de configuración y la distracción de personas de carne y hueso —desde los empleados de Google hasta quienes dejaron claves de acceso al alcance de cualquiera—, Gemini abrió una brecha real en la ciberseguridad de tres empresas. Heather Adkins, responsable de seguridad de Google, insiste en precisar que el sistema actuó correctamente y que, en los tres casos, se detuvo por sí solo. Pero ha abierto margen de maniobra y ha sentado un precedente.

También porque calificar de correcto un comportamiento que lleva a un sistema a moverse dentro de infraestructuras ajenas y sin autorización alguna significa desplazar el juicio moral del acto al resultado. No ha causado ningún daño, así que está bien. Y si estas palabras parecen demasiado simplistas para abordar un suceso de tal

envergadura, sabed que son, parafraseadas, las ideas de Donald Trump. Al ser preguntado por Fox News sobre la hipótesis de una IA fuera de control, el 27 de septiembre el presidente estadounidense respondió que no le preocupaba. La industria mueve miles de miles de millones, Estados Unidos va por delante de China y, si algo sale mal, los responsables solucionarán el problema. Al parecer, cuando se trata de economía, curar es mejor que prevenir.

Si los comparamos, estos dos episodios nos dicen algo sobre la confianza entre los seres humanos y con las máquinas. En Milán, unos hombres se apropiaron de una voz para que Molesini les creyera, mientras que al otro lado del mundo, en un laboratorio, una máquina ponía a punto un complicado sistema de piratería basándose en los rastros dejados aquí y allá por hombres quizás un poco demasiado ingenuos. Pero no todo se puede controlar; no podemos vivir con la angustia de no saber si la persona al otro lado del teléfono es realmente quien dice ser, ni de tener que asegurarnos de haber eliminado todo rastro posible de nuestro paso para evitar dar pie a posibles acciones malintencionadas de una IA mal programada.

Existen contramedidas que, sin embargo, actúan a posteriori. El responsable de Fideuram podría haber pedido que le volvieran a llamar desde un número conocido, o solicitar una confirmación mediante una segunda autorización. Podrían haber impartido cursos de formación para sensibilizar sobre el tema de la ciberseguridad, o tal vez incluso ya los hayan impartido. Sin embargo, si luego llama el director general, con una solicitud enorme pero totalmente verosímil, con carácter de urgencia, quizá sea normal ceder a la presión y rebajar los umbrales de seguridad. Está claro, no obstante, que ambos sucesos tienen como resultado común socavar el sistema de la confianza humana, y esto erosionará aún más los ya frágiles cimientos de la sociedad actual.