



SALTO BASE

# “Vamos a ser tontos, pero muy tontos...”

Máquinas contra máquinas, equipos de IAs como asistentes de los CISOs, IAs parando ataques e infiltrándose en las redes criminales para rescatar información empresarial, SOC de IAs -jefas de departamentos de IAs-, en jornadas laborales de 24x7, recorriendo las redes y comunicaciones en busca de indicios de incidentes, de posibles *insiders*, incluso identificando a los que lo van a ser en el más puro estilo de ‘Minority Report’ (2002). La imaginación permite vislumbrar un mundo inquietante plagado de IAs, no siempre al servicio de las personas. Y donde la seguridad de la información, la ciberseguridad, se va a convertir en la ‘última frontera’ frente a los desmanes con datos, identidades,



**El simbólico ‘Reloj del Apocalipsis’ ha pasado de estar a siete minutos de la medianoche en 1947... a 89 segundos en 2025. Y la culpa, en**

**parte, es de la IA... ¿Se imaginan a los científicos del proyecto Manhattan si, además de la energía nuclear, hubieran tenido las capacidades de los aviones supersónicos y las armas láser? Para evitar que llegue esa ‘temida hora’ habrá que desarrollar una nueva ciberseguridad que ni imaginamos.**

operaciones de negocio, despliegue de tecnología a cargo de sistemas autónomos que seguramente precizarán, incluso, de un ‘botón rojo’ para apagarlos en caso de riesgo crítico –la propia **Google** reconoció que había patentado algo parecido–.

¿Ciencia o ficción? Recientemente, Sam Altman, el CEO de OpenAI, dijo que “no hay búnker que te libre de una IA fuera de control”. Lógicamente, luego matizó sus palabras. Porque da que pensar que el responsable de una de las IAs Generativas más usadas del mundo plantee el concepto de ‘IA fuera de control’ –¿Qué habrá visto en sus laboratorios de pruebas?–. Y no es el único: **Geoffrey Hinton**, el científico estadounidense que ganó el Premio Nobel de Física en 2024, junto a **John Hopfield**, por sus trabajos en este campo ha reconocido que ve “muy preocupantes” los avances tecnológicos en esa materia y advirtió de una catástrofe si no se controlan. Y ese control dará pie a una nueva familia de lo que denominamos ciberseguridad, pero aplicada al control de las ‘máquinas pensantes’, término con el que se denominaba esta tecnología hasta que, en 1956, el matemático e informático teórico, **John McCarthy** en la ‘Convención de Dartmouth’, enunció lo que entendemos por IA: “la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes”.

## Esperando la singularidad, cibersegura

Y es que no hay día que alguien no se plantee cuándo podría alcanzarse la temida –¿y esperada?– “singularidad tecnológica”. Dos palabras que marcan el momento en el que la IA creada por el hombre sea mucho más ‘inteligente’ que él, bautizada como Superinteligencia Artificial (ASI). Que sea capaz de hacer cosas que nunca llegaremos a comprender. **Elon Musk** sitúa este mismo año la IA general (AGI), la que nos iguala. Otros como **Ray Kurzweil** ven posible la ASI en 2030, muchos otros para 2040 –cuando consideran que conviviremos con robots–. Llegará. Máxime si se tiene en cuenta que son las IAs las que están trabajando en su propio código, además de muchos humanos.

Uno de los cofundadores de **OpenAI**, **Ilya Sutskever**, abandonó en 2024 la compañía y fundó **Safe Superintelligence Inc. (SSI)**, que en menos de un año ha recibido una inversión de más de 900 millones de euros para desarrollar una Superinteligencia que, además, asegura, sea cibersegura. Para ello, para ‘domarla’, para evitar que haga cosas fuera de nuestra comprensión tendrá que partir de la ciberseguridad actual pero desarrollar una nueva que pueda detectar y parar riesgos y límites de la ASI que no alcanzamos a imaginar. ¿Llegará la ‘SuperCiberseguridad’?

De momento, usar la IA sin control nos vuelve tontos –¿quizá como los ocupantes de la nave de la película **Wall-e?**–, “pero muy tontos”, como dice **José Mota**. Así lo destaca una investigación de **Microsoft Research** y la **Universidad Carnegie Mellon** que alerta de que la dependencia de herramientas de IA, sin cuestionar su validez, reduce el esfuerzo cognitivo aplicado al trabajo. Lo dicho: nos hace tontos.

En 1947 muchos de los científicos del ‘Proyecto Manhattan’ –que creó la primera bomba atómica– ‘pusieron en marcha’ un simbólico ‘Reloj del Apocalipsis’ (a siete minutos de la medianoche). Este año se adelantó un segundo, quedándose a 89. Muchos han defendido que la IA es un *hype*, una moda que pasará (como ya hizo en los años 50 y 80), que está cerca de su ‘invierno’... Pero cada vez más científicos consideran que vive su primera gran primavera y que lo que viene, sin ‘ciberseguriAdad’, no tiene pinta de terminar bien...



**José Manuel Vera**  
Redactor  
Revista SIC