AI and Safety – Failure Modes of Automated Systems Stats and Facts – Spanish



HECHOS

1. Los sistemas automatizados pueden fallar cuando los sensores proporcionan datos incompletos, ruidosos o incorrectos, lo que provoca que la IA calcule mal las distancias, los obstáculos o la presencia humana.
2. Los errores de software, el firmware obsoleto o los algoritmos defectuosos pueden hacer que la automatización actúe de forma impredecible, realice movimientos no deseados o ejecute tareas en el momento equivocado.
3. Los sistemas de automatización pueden fallar en condiciones extremas (iluminación, clima o formas de objetos inusuales) porque el modelo de IA no ha sido entrenado para esos escenarios.
4. Las interrupciones de la red o de las comunicaciones pueden detener los comandos en medio de una operación, retrasar las señales de emergencia o hacer que los robots se «congelen» en posiciones inseguras.
5. Los sistemas autónomos pueden comportarse de forma peligrosa cuando la lógica de seguridad se anula, se desactiva o se configura incorrectamente durante el mantenimiento o la configuración.
6. Los sistemas impulsados por IA pueden clasificar erróneamente las extremidades humanas, los EPP o las herramientas, creando situaciones peligrosas en las que el robot «piensa» que el área está despejada cuando no lo está.

ESTADÍSTICAS

- Un análisis realizado en Estados Unidos reveló que casi el 40 % de los accidentes relacionados con la automatización se debían a fallos de los sensores o de detección, como el hecho de que una máquina no reconociera a un trabajador en la zona de peligro. (Centro de Investigación Robótica Ocupacional, NIOSH)
- En EE. UU., los incidentes relacionados con la IA en los lugares de trabajo aumentaron un 30 % entre 2023 y 2025, y los modos de fallo, como las alucinaciones y la amplificación de sesgos, contribuyeron al 15 % de las infracciones de seguridad notificadas en la automatización industrial.
- En 2025, el 44 % de las organizaciones canadienses que implementaron sistemas automatizados de IA experimentaron al menos un incidente relacionado con un fallo, como una interpretación errónea del sistema que provocó una interrupción del funcionamiento o cuasi accidentes.
- Los sectores manufactureros de EE. UU. registraron una reducción del 25 % en los accidentes provocados por robots con IA entre 2020 y 2025, pero los fallos en la

supervisión predictiva causaron el 10 % de los incidentes residuales, incluidas decisiones autónomas inesperadas.
- Entre 2021 y 2024, el phishing de credenciales y los ataques adversarios a los sistemas de IA aumentaron un 703 % en Estados Unidos, aprovechando modos de fallo como la transparencia insuficiente y provocando compromisos de seguridad en el lugar de trabajo.
- En Canadá, el 32 % de los incidentes de compromiso del correo electrónico empresarial relacionados con herramientas de IA en 2024 se debieron a fallos en la autenticación multifactorial, lo que amplificó los riesgos en los procesos de toma de decisiones automatizados.