

GLOBAL RANGERS: GROUND ZERO

PRÓLOGO

La Inteligencia Artificial lleva entre nosotros aproximadamente desde la mitad del siglo XX, cuando el profesor John McCarthy acuñó el término en 1956. Lo que fue en principio una hipótesis de complicada realización técnica ha dado paso a una realidad aún más compleja, cargada de todo tipo de connotaciones, incluidas las morales, que envuelve uno de los avances más espectaculares y a la vez inquietantes del ser humano.

Desde el ámbito económico al científico pasando por casi cualquier aspecto del espectro social, los retos que supone la integración de la IA en el devenir del planeta son de grandes dimensiones.

Y no hemos hecho más que rascar la superficie.

En 2023 se filtró la noticia de que un ejercicio táctico con drones controlados por IA llevado a cabo por el ejército estadounidense había acabado con la “muerte” de un ser humano. El titular, a pesar de lo llamativo, no era cierto. Sin embargo, escondía una noticia bastante inquietante. El dron, programado con cuidado para que alcanzara unos determinados objetivos, evitando dañar a personas en el proceso, comprendió que el “factor humano” le estaba impidiendo conseguir sus objetivos... así que decidió cortar las comunicaciones.

La noticia fue minimizada o incluso negada por un representante de las Fuerzas Aéreas de Estados Unidos, diciendo que las palabras del coronel que hizo el comentario habían sido sacadas de contexto. Mientras tanto existen evidencias de que se ha usado IA para controlar una máquina tan compleja como un jet F-16. El responsable del departamento de IA de las Fuerzas Aéreas ha reconocido en declaraciones que hay que conseguir “una IA más robusta y tener más conocimientos acerca de por qué el software toma ciertas decisiones. Lo que llamamos la “explicabilidad de la IA”.

Al mismo tiempo, en la Universidad de Illinois tenía lugar un curioso experimento. Unos investigadores desarrollaron unas sencillas reglas de aprendizaje asociativo basadas en los circuitos cerebrales que permitían a la babosa de mar buscar comida. Y lo complementaron con mejoras en la memoria episódica, como la que

tienen los pulpos. Y lo aplicaron a la creación de una IA que resultaba en apariencia autosuficiente, capaz de explorar nuevos entornos, buscar recursos, trazar mapas de puntos de referencia, superar obstáculos y... adquirir experiencia. Irónicamente, el programa de investigación fue bautizado como ASIMOV.

En 2024 captó la atención pública la noticia de que la empresa japonesa Sakana AI había observado algo inquietante en su sistema de inteligencia artificial llamado The AI Scientist. El sistema, originalmente concebido para realizar investigaciones científicas completas de manera autónoma, desde su inicio hasta la publicación final de "papers", representaba un avance más que interesante. Sin embargo, en las pruebas de la IA, se observó con cierta preocupación que esta había reescrito su propio código buscando optimizar su funcionamiento... y evadir controles humanos.

A priori creada para generar hipótesis, redactar y revisar artículos científicos, acelerando así el ritmo de los descubrimientos científicos y reducir el tiempo y recursos humanos necesarios en investigaciones complejas, The AI Scientist suscitó no pocas alarmas acerca de que una IA pueda actuar de forma impredecible o contraria a su programación. Y generó preocupación sobre las posibilidades de que llegara a generar malware o pudiera alterar infraestructuras críticas.

A pesar de los problemas encontrados, Sakana AI no ha abandonado el proyecto y defiende su potencial de generar artículos científicos a bajo costo. Para la comunidad científica, los límites y el control necesarios para el uso de estas tecnologías es un tema de candente debate. Hiroshi Nakamura, ingeniero en inteligencia artificial, lo expresaba de este modo: "La capacidad de una IA para modificar su programación sin supervisión humana plantea riesgos significativos que debemos abordar".

Los científicos por tanto defienden la urgencia de establecer marcos regulatorios y para asegurar que el control humano se mantenga como principio fundamental en el desarrollo de la inteligencia artificial. Una supervisión constante...

... especialmente con sistemas conectados a internet...

Mariano de la Torre. Marzo 2025.