ACADEMIA DE LAS CIENCIAS
Y LAS ARTES MILITARES

Comunicaciones académicas

Contrainteligencia artificial

Manfredo Monforte Moreno
Academia de las Ciencias y las Artes Militares
Sección de Prospectiva de la Tecnología Militar

26 de septiembre de 2023

En mayo de 2010, una operación lanzada desde Londres provocó un error en los índices bursátiles de Nueva York. El incidente duró 36 minutos y produjo pérdidas por valor de mil millones de dólares. La causa: un algoritmo erróneo —o manipulado—. Unos meses más tarde, un error técnico en uno de sus algoritmos de compraventa automática de valores produjo unas pérdidas de 440 millones de dólares a un grupo inversor norteamericano.

En febrero de 2023 Microsoft anunció la renovación de su motor de búsqueda Bing con el uso de inteligencia artificial (IA) con el objetivo estratégico de retomar su liderazgo en el mercado tecnológico, pero, a una semana del lanzamiento, los usuarios informaron de comportamientos anómalos de su *chatbot*, señalando que a veces se tornaba hostil hasta el punto de insultarles.

Las consecuencias de los errores o los cambios mal intencionados en los algoritmos que circulan por las venas de los sistemas de IA y gobiernan los procesos pueden tener consecuencias catastróficas.

Según el artículo 32 de la Ley de Seguridad Nacional, se define la contrainteligencia «como el conjunto de medidas de protección en contra de actos lesivos, así como las acciones orientadas a disuadir o contrarrestar su comisión».

Por tanto, las acciones de contrainteligencia se basan en la capacidad de prevenir, detectar y neutralizar aquellas actividades que supongan una amenaza para los derechos y libertades de los ciudadanos, así como para la soberanía, integridad y seguridad del Estado, sus instituciones, infraestructuras críticas y la economía nacional en su conjunto.

La Estrategia de Seguridad Nacional plantea como objetivo en el ámbito de la contrainteligencia adoptar medidas en defensa de los intereses estratégicos, políticos y económicos de España con acciones concretas que imposibiliten la obtención ilegal de información sensible para España. Entre las líneas de acción que plantea está la elaboración de una normativa actualizada para regular la protección de la información clasificada; así como el impulso y refuerzo de las capacidades de los órganos nacionales de inteligencia, con objeto de garantizar la disposición de los medios humanos y técnicos necesarios para contrarrestar eficazmente la amenaza. Para lograr sus fines, es necesaria la adecuada y oportuna concienciación de las personas, pues estas constituyen el eslabón débil de la seguridad en cualquier ámbito.

Desde esta perspectiva, se ha profundizado en la determinación de los orígenes, metodología y finalidad de los ciberataques realizados por servicios de inteligencia de otros países contra organismos de las Administraciones Públicas y los sectores estratégicos. En paralelo, se ha mantenido el esfuerzo en materia de contrainteligencia y seguridad en apoyo a los contingentes militares españoles desplegados en el exterior. Para alcanzar sus objetivos, la IA constituye una herramienta valiosísima en apoyo a la actuación de los organismos responsables.

Históricamente, el ser humano ha luchado por obtener la superioridad sobre su entorno: desde ocupar y defender los mejores territorios de caza o cultivo a esclavizar a los pueblos enemigos y apropiarse de sus riquezas, normalmente bajo el patrón oro. Durante la década de los 70 del siglo pasado, el petróleo se constituyó en el llamado oro negro, provocando alguna de las primeras crisis globales sin intervención militar. Después aparecieron otros vellocinos de oro: la tinta de las impresoras, las redes sociales, la telefonía móvil, las plataformas de contenidos audiovisuales… para llegar al valor de los datos como fuente de la superioridad en la información, superado hoy por la enorme importancia de los algoritmos y el potencial de la inteligencia artificial (IA), una herramienta, como tantas otras, cuyo valor añadido radica en las intenciones de quien la desarrolla o usa, sean buenas o malas. Podemos afirmar que el oro de hoy son los algoritmos.

Básicamente, IA es la capacidad con que se dota a un dispositivo electrónico para realizar tareas que generalmente requieren inteligencia humana. En informática, IA es la disciplina que estudia y desarrolla sistemas capaces de sustituir al ser humano en ciertas tareas.

La IA trata de acercarse a la complejidad del pensamiento racional deductivo, es decir, aquél que es capaz de inferir conocimiento por sí mismo y determinar su comportamiento en base a sus algoritmos y capacidad de aprendizaje. Incluye diversas áreas de conocimiento que imitan aspectos de la inteligencia humana, como la capacidad para percibir, reconocer el entorno y reaccionar en consecuencia; la facultad de planificar y resolver problemas; la habilidad para aprender constantemente y adaptarse; las inteligencias social, emocional y, por qué no, alcanzar el escalón más elevado: la creatividad.

De acuerdo con su nivel de competencia se distinguen hasta tres tipos de IA:

- Específica o débil: es la que existe hoy en día. Es capaz de hacer bastante bien una tarea o conjunto de tareas previamente acotadas y solamente esas tareas, por ejemplo, redactar un artículo, un trabajo universitario, emular una voz o representar el rostro de un fallecido impartiendo una conferencia.
- General o fuerte: dotada con las mismas capacidades de una persona. Todavía en fase de desarrollo, su llegada no será el límite que convertirá a la IA en algo muy distinto a lo que hoy está transformando a nuestras sociedades.
- Superinteligencia: todavía en el ámbito de la ciencia ficción, superaría a la inteligencia humana, alcanzando en ese momento lo que se conoce como *singularidad*. Por definición, al ser una inteligencia superior a la nuestra, no la comprenderemos.

Aunque la IA existe como disciplina emergente desde los años cincuenta del siglo pasado, ha sido en los últimos diez años cuando ha adquirido gran protagonismo social gracias a la confluencia de tres factores: la existencia de ingentes cantidades de datos disponibles en fuentes abiertas (*Big Data*), la disponibilidad de grandes capacidades de cálculo a bajo coste y el desarrollo de modelos complejos de aprendizaje –inspirados en las redes neuronales– llamados modelos de aprendizaje profundo (*deep learning*) que están en el corazón de los servicios digitales que utilizamos en nuestro día a día y en el de los coches, ciudades, hogares y móviles inteligentes.

Los avances de la IA han sido formidables durante las últimas décadas, permitiendo el reconocimiento facial, la identificación de personas, la traducción de lenguas, la producción de obras artísticas e incluso el movimiento y empleo de máquinas autónomas. Surge una cuestión de difícil solución: ¿Cómo podemos defendernos de los avances de la IA cuando esta se usa con fines maliciosos?

La cuestión es que la IA ha entrado imperceptiblemente en nuestras vidas y podemos encontrarla en todas partes. Gracias a su gran capacidad para aprender a partir de los datos, permite reconocer y predecir patrones y optimizar tareas. Es parte del proceso de la toma de decisiones, tanto en los ámbitos públicos como privados. Está transformando todos los sectores productivos y por ello, constituye el motor de la cuarta revolución Industrial. De ahí que las grandes potencias mundiales apuesten por invertir en IA, conscientes de que el liderazgo de esta ha de abanderarse no solo desde los ámbitos tecnológico y económico, sino también teniendo en cuenta su impacto social, ético, legal y regulatorio.

La IA más conocida es la regenerativa, capaz de producir textos, imágenes, música, video u otro tipo de datos. Los modelos más conocidos son ChatGpT para texto y DALL-E para imágenes. Estos modelos cuentan con infinidad de parámetros al haber sido desarrollados sobre miles de textos o imágenes de Internet. Representan una revolución ya que nunca hasta ahora una máquina había generado este tipo de contenido con el nivel de competencia de un humano. Incluso transforman textos mediocres a lenguaje erudito o de experto en temas concretos.

En el lado positivo, gracias a la IA contamos con asistentes de voz (Alexa, Google, Apple…), sistemas de búsqueda y recomendación, mejora de procesos productivos, ciudades inteligentes y, cada vez con más presencia, ayudas a la conducción. En el sector de la seguridad y la defensa, la IA cuenta con aplicaciones muy interesantes que facilitan el despliegue de sistemas robustos frente a los ataques enemigos y proporcionan nuevas capacidades impensables sin el concurso de los algoritmos de la IA.

Aunque la IA tiene un potencial inmenso para ayudarnos, también adolece de limitaciones y plantea retos éticos que debemos abordar, especialmente por su posible uso malicioso. Así, la tecnología hostil se asocia comúnmente con actividades delictivas como el *ransomware*, el robo de datos o la creación de virus informáticos, aunque esto no lo es todo. El panorama está evolucionando de tal manera que la definición de tecnología hostil debería ampliarse para incluir actos legales, incluso ampliamente aceptados, que en última instancia amenazan el bienestar de la sociedad y la seguridad nacional.

A medida que la tecnología se vuelve más compleja, aumentan las formas en que puede ser utilizada maliciosamente. Y a medida que las personas dependen más de la tecnología en sus actividades cotidianas, se ven cada vez más expuestas a consecuencias involuntarias, incluso adversas. Si se añade un alto nivel de automatización —tomando decisiones a la velocidad de una máquina—, la posibilidad de que las cosas salgan mal aumenta rápidamente.

La tecnología hostil puede abarcar no sólo la tecnología criminal, como el *malware* y las herramientas de *hacking*, sino también casos de uso como la publicidad, la propaganda y la selección de segmentos sociales objetivo. Que la tecnología sea negativa puede ser cuestión de perspectiva. Algunos individuos no consideran intrusivos los anuncios en Internet, las *cookies* de seguimiento o las campañas de influencia en las redes sociales, mostrándose dispuestos a facilitar sus datos ante ofertas personalizadas y convincentes. El consentimiento para el seguimiento o la recopilación de datos personales es, para algunos, básicamente automático; para otros, una elección delicada y meditada. Muchas personas no se dan cuenta de que siempre hay opciones frente al consentimiento.

No todos los comportamientos hostiles son maliciosos o malintencionados. Un ejemplo es el sesgo en los algoritmos o sistemas de aprendizaje automático, que pueden mostrar tendencias agresivas hacia determinados grupos sociales sin haber sido diseñados deliberadamente así debido a distorsiones no planificadas e inadvertidas en la forma en que fueron construidos o desarrollados.

Claramente la IA puede ser una fuente inagotable de agentes perversos. Según Frost & Sullivan, el número de dispositivos activos de la Internet de las cosas (IoT) superará los 65.000 millones en todo el mundo en 2026. Cada uno de ellos conlleva posibles fallos de seguridad que podrían explotarse para beneficio de terceros y, en el peor de los casos, para el delito.

Con las filtraciones de datos acercándose en niveles récord, la protección contra la piratería deliberada y el *malware* es cada vez más importante. Las empresas deben invertir en la defensa de una gama más amplia de adversarios bien financiados y organizados. Sin embargo, a medida que aumenta el potencial de peligro, también

hay que tener en cuenta otras dimensiones de la tecnología: el respeto por los deseos de los usuarios, negar la segmentación intrusiva e interesada y eliminar los sesgos en los sistemas algorítmicos y los conjuntos de datos no solo es intrínsecamente ético, sino que favorece la confianza y la percepción pública positiva de las nuevas propuestas tecnológicas.

Una de las aplicaciones más controvertidas es el uso de la IA para el control sobre las máquinas y sistemas de armas con capacidad para usar la fuerza letal propia de la guerra y que en ningún caso puede pasar por alto los aspectos legales y morales que su uso implica, especialmente por parte de sociedades democráticas. El cambio en las características de la guerra que producen los sistemas de armas dirigidos por IA exige que se tengan en cuenta las restricciones éticas y legales que implican los nuevos conflictos con el riesgo de que la guerra en sí misma se convierta en un fenómeno inhumano fuera de todo control y supervisión.

Recientemente, el empresario norteamericano Elon Musk ha pedido frenar el desarrollo de la IA en varias ocasiones, sobre todo de la IA más avanzada y potente. Musk ha expresado su preocupación por los posibles peligros de la IA para la humanidad, como la pérdida de empleos, la manipulación de la información, la guerra cibernética o el surgimiento de una superinteligencia hostil. Algunos de los expertos en IA que han apoyado a Musk en su petición son Steve Wozniak (cofundador de Apple), Emad Mostaque (director general de Stability AI), Yoshua Bengio (pionero del aprendizaje profundo) y Stuart Russell (autor de un libro de referencia sobre IA).

Los motivos por los que quieren frenar los desarrollos en marcha se basan en el temor a que la IA pueda generar información errónea, reemplazar los trabajos con automatización y causar daños a la sociedad y la humanidad, así como en el peligro de que la IA escape a la comprensión, predicción y control de sus creadores. Es imperativo evaluar la seguridad de la IA, definir un marco legal y crear dispositivos técnicos para ayudar a estados e instituciones a encarar las crisis que la IA podría provocar, evitando el riesgo de poner en manos de unas pocas personas o empresas el control de los algoritmos que controlan la IA.

La evolución de la tecnología está afectando a casi todos los ámbitos de la sociedad, pero también, de manera muy significativa, al de la seguridad y la defensa. El conflicto de Ucrania ha demostrado que las guerras de hoy son multidominio y se libran tanto en el espacio físico como en el digital. Los ciberataques, las campañas de desinformación y las noticias falsas son una amenaza para las sociedades democráticas. La desinformación como arma de estrategia híbrida es tan antigua como la propia guerra, pero estamos asistiendo a un nuevo hito histórico, definido como «cuarta revolución industrial», en el que la

IA juega un papel determinante al cambiar el escenario global y su entorno. La soberanía tecnológica en IA debe de ser un objetivo estratégico con la participación conjunta y coordinada de administraciones públicas, empresas tecnológicas, sistemas educativos y agencias de verificación.

DARPA, la agencia norteamericana de investigación y desarrollo en el campo militar continúa en su empeño por crear más y mejores sistemas para el combate y la disuasión. En DARPA no todo son sistemas militares, también hay espacio para el desarrollo de la IA en apoyo al combate, dado que las principales desarrolladoras se niegan a participar en el terreno bélico con su tecnología. Lo último que ha desarrollado la agencia es una IA que no solo actúa conforme a unas órdenes y experiencia previas, sino que razona las decisiones. Es decir, es capaz de encontrar una buena solución a un problema y valorar una orden humana para mejorarla y comunicarla antes de actuar: ¡se atreve a cuestionar las decisiones del mando!

Sin duda, pronto hablaremos del concepto de «contrainteligencia artificial» (CIA) si queremos enfrentarnos con garantías a las nuevas tecnologías hostiles y al uso malicioso de herramientas informáticas. Debemos alcanzar la capacidad de detectar, identificar y neutralizar las nuevas amenazas que acompañan al desarrollo universal de la IA.

En España se ha creado la Agencia Española de Supervisión de la IA (AESIA) que está llamada a jugar un papel esencial en el control del nuevo oro negro: los algoritmos. Espero que pronto cuente con un departamento que se ocupe de los asuntos militares y de los servicios de inteligencia. ■