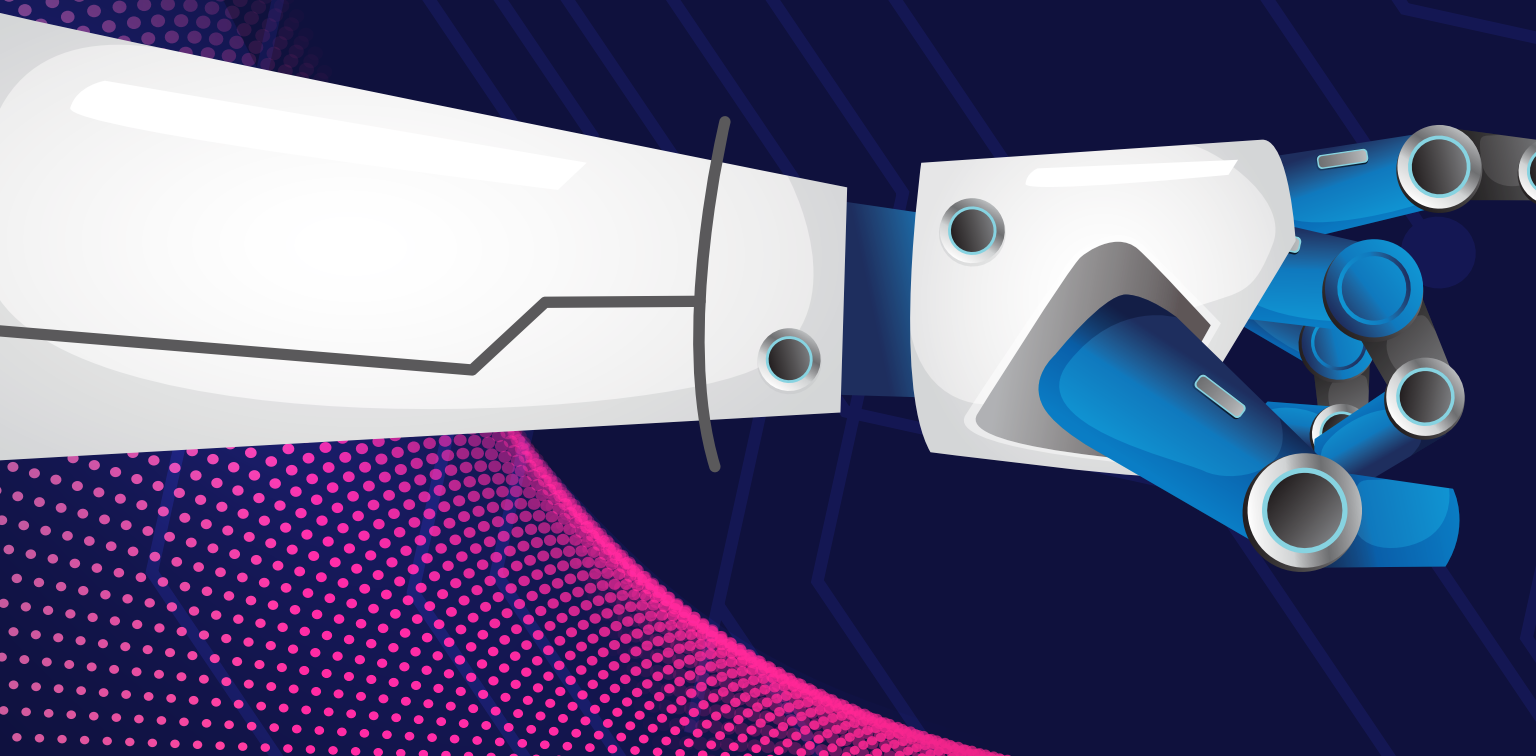


Preparándonos para confiar en los sistemas de inteligencia artificial de los equipos de combate

Mayor (capellán) Marlon W. Brown, Ejército de EUA

Estamos en el albor de una rápida integración de la inteligencia artificial (IA) en las operaciones militares. En la *Estrategia de Seguridad Nacional de Estados Unidos* se señala que los rápidos avances en el campo de la IA son de gran importancia para la seguridad nacional¹. El

Resumen de la Estrategia de Seguridad Nacional de Estados Unidos de 2018 se hace eco de esta declaración y añade que es necesario «invertir considerablemente en la aplicación militar de la autonomía, la inteligencia artificial y el aprendizaje automático, como también aplicar rápidamente los avances comerciales, para ganar ventajas competitivas militares» como



parte de los esfuerzos de modernización de capacidades clave para construir una fuerza más letal².

El Centro de Inteligencia Artificial Conjunto está a cargo de ejecutar lo que se estipula en el nuevo *Resumen de la Estrategia de Inteligencia Artificial de 2018 del Departamento de Defensa*. En este documento se menciona que el Departamento de Defensa colaborará con socios comerciales y académicos para desarrollar e implementar tecnologías emergentes³. Otro componente de este enfoque de modernización es la Agencia de Defensa para Proyectos de Investigación Avanzada (DARPA), para la cual el presidente solicitó un presupuesto de 3556 millones de dólares para el año fiscal de 2020. Se estima que el proyecto «Inteligencia artificial y la simbiosis hombre-máquina» costará más de 161 millones de dólares en 2020, lo que significa un aumento de un 233 por ciento con respecto al presupuesto de 2018⁴.

En la actualidad, la integración de la IA es limitada y no ha alterado considerablemente la manera de librar la guerra, especialmente en el nivel táctico. Los humanos todavía tienen control total. Dado que los líderes civiles y militares actúan con cautela cuando la toma de decisiones y los análisis asistidos por la IA pueden afectar vidas humanas, muchos esperan que esta norma continúe. Sin embargo, es probable que este tipo de colaboración entre humanos y tecnología cambie cuando nuestros adversarios empiecen a utilizar la IA contra nosotros. Independientemente de cuantos científicos

prominentes se opongan al empleo de armas autónomas o cuán razonable sean los argumentos contra el desarrollo de la IA, «el genio de la IA y la innovación ya salieron de la lámpara: no hay marcha atrás»⁵. Nuestros adversarios están invirtiendo considerablemente en esta tecnología y nosotros también.

Como las guerras futuras serán libradas con sistemas de IA, nuestras Fuerzas Armadas deben estar preparadas para aceptar esta nueva tecnología. Estar preparados no se trata simplemente de desarrollar y utilizar los sistemas de IA apropiados. También se deben responder preguntas éticas y morales como: «¿Estarán los soldados dispuestos a pelear codo a codo con robots en el campo de batalla?»⁶. Para responder este tipo de preguntas, se debe considerar si los combatientes humanos son capaces de confiar en los sistemas artificiales que integran el equipo. Si aplican el concepto doctrinal de confianza en equipos cohesivos y evalúan los factores que pueden facilitar la confianza, los soldados podrán prepararse para confiar en los sistemas de IA que serán integrados próximamente en los equipos de combate.





Gráfico cortesía de Army AL&T Magazine.

¿Qué es la IA?

Antes de abordar la relación de la confianza con la IA, es importante entender la naturaleza cambiante de esta tecnología. La tecnología de la IA no es estática y los rápidos avances en su desarrollo continúan ampliando nuestro entendimiento de la tecnología y cómo el tema de la confianza en los sistemas de IA debería ser tratado. Una búsqueda rápida en Internet permite ver que hay numerosos ejemplos y términos para identificar los diferentes tipos de IA. La inteligencia artificial es clasificada de dos maneras en este artículo: inteligencia artificial débil (IA débil) e inteligencia artificial general (IA general). En la actualidad, todos los sistemas de IA operan mediante IA débil, lo que significa que estos se centran en tareas limitadas y específicas. Siri, uno de los sistemas de IA más conocidos, solo es capaz de realizar una serie de tareas específicas que se limitan a los productos de Apple. Los sistemas de IA débil solo pueden hacer lo que fueron diseñados para hacer.

La IA general, por otro lado, es el futuro de la IA y la que permitirá que las máquinas tengan *intención* y *consciencia*. Los sistemas de IA general, como los humanos,

serán generalistas y podrán aplicar la información aprendida a diversas tareas y experiencias. Muchas veces se aplican términos filosóficos en las discusiones sobre la IA general. Además de intención y consciencia, también se utilizan otros términos como *sintiencia* (capacidad de sentir) y *voluntad* (libre albedrío) para describir los diferentes tipos de IA que clasificamos como IA general. En otras palabras, la IA general será igual a la inteligencia humana en términos de emociones y pensamientos de alto nivel. Los personajes ficticios como Terminator, Wall-E y Data, de *Star Trek: The Next Generation*, son sistemas de IA general. Aunque muchos de estos sistemas ficticios tienen forma humanoide, es posible que los sistemas de IA débil y general tengan componentes robóticos o proyecciones audiovisuales, o puede que existan en el ciberespacio sin una interfaz humana. Confiar en la IA débil y general dependerá de cómo se defina y se experimente la confianza dentro de las Fuerzas Armadas⁷.

Confianza doctrinal en los equipos militares

En la doctrina del Ejército se reconoce que la confianza en los equipos militares es importante. La confianza mutua es esencial para practicar el mando tipo misión. «Ganar o perder la confianza depende más de nuestras

Página anterior: Gráfico compuesto de Arin Burgess, *Military Review*. Gráficos originales cortesía de Harryarts, ddraw y Freepik vía <https://www.freepik.com/>.

acciones diarias que de gestos ocasionales o magnánimos. Es producto de los entrenamientos y las experiencias compartidas, y por lo general, es una consecuencia fortuita de las operaciones, pero también fomentada deliberadamente por los comandantes»⁸. El Ejército define la confianza entre los soldados como la «dependencia en el carácter, la competencia y el compromiso que guía a los profesionales del Ejército y sostiene la Ética del Ejército»⁹. Es difícil exagerar el nivel de confianza necesario para construir un equipo de combate eficaz.

La guerra es una actividad humana. Sin embargo, la integración de la IA complica las lecciones históricas sobre la naturaleza de la guerra y amenaza con reemplazar parte de la carne y hueso en los equipos de combate con hardware y software. Incluso si a la larga la IA no afecta la naturaleza de la guerra (una proposición poco probable), el carácter de la guerra sí cambiará por completo cuando esta tecnología se integre plenamente. El inventor y escritor Amir Husain sugiere que uno de los cambios más significativos del carácter de la guerra debido al rápido avance de la IA será la velocidad del combate a nivel táctico¹⁰. ¿Qué pasará cuando las mentes humanas y los sistemas de decisiones no puedan mantener el ritmo de las máquinas autónomas del enemigo? Aunque la decisión de ir a la guerra o conducir una operación requiere análisis y reflexión, y por lo tanto, tiempo, en muchas ocasiones, las unidades tácticas no tendrán ese lujo y dependerán de la IA para tomar y ejecutar decisiones letales en el campo de batalla si quieren sobrevivir. En tal escenario, la IA claramente es parte de un equipo de combate cohesivo en el que la confianza es indispensable. Por lo tanto, es necesario abordar el tema de la confianza entre hombre y máquina.

La confianza con un actor no humano no es tan extraña si nos damos cuenta que esta ya existe en las operaciones militares. Tal vez el mejor ejemplo de confianza mutua entre humanos y actores no humanos es la relación entre perros de trabajo y sus adiestradores. Los adiestradores desarrollan una relación muy estrecha con sus perros, incluso más estrecha que la que desarrollamos con nuestras mascotas. Lo que hace especial al perro de trabajo es el alto nivel de confianza que sus adiestrados le tienen. Los adiestradores no solo confían en sus perros para que ejecuten las tareas rutinarias para las que fueron adiestrados, sino que también confían en ellos en situaciones de peligro y hasta en situaciones donde sus vidas corren peligro.

La confianza que un humano pueda tener en una IA débil, que no tiene carácter o compromiso, es solo confianza en la competencia del sistema. La IA débil debe ser competente en varias tareas, como por ejemplo, la identificación de amenazas de componentes clave y la formulación de estrategias para mitigarlas. También debe ser capaz de identificar con precisión al enemigo en el campo de batalla o reconocer síntomas de depresión entre los miembros de la unidad y recomendar un tratamiento.

La confianza en una IA débil se parece más al tipo de confianza que los soldados tienen en sus armas o herramientas de planificación que la que tienen con otro soldado. Las herramientas, aunque estén hechas de acero o algoritmos, no debería ser tratadas como «miembros» verdaderos de un equipo, incluso si se establecen vínculos emocionales. Los vínculos con un sistema de IA débil no cambian la naturaleza del sistema. En la película *Cast Away* queda claro que el personaje de Tom Hanks desarrolla vínculos emocionales con una pelota de voleibol que cariñosamente llama «Wilson». Incluso podría decirse que hasta «confiaba» en Wilson, ya que le contaba sus pensamientos íntimos. Independientemente del grado de apego emocional, Wilson era simplemente un pedazo de cuero y goma. Wilson ayudaba a preservar la sanidad del naufrago. Aunque una IA débil pueda actuar de forma autónoma, la autonomía no significa voluntad. Los soldados deben ser capaces de reconocer que la confianza en los sistemas de IA débil no es igual a la confianza que tienen con otros soldados humanos o miembros futuros del equipo que funcionarán mediante la IA general.

La IA general será diferente. Tendrá una forma de «persona» que le permitirá ser tratada como un miembro confiable del equipo de combate. Atribuirle una forma de persona no es una invitación para debatir si una máquina consciente es una forma de vida o si merece protecciones legales por ello. Esas preguntas éticas deben ser abordadas en otro trabajo. Considerar la

El mayor Marlon W. Brown, Ejército de EUA, es el capellán de la 2ª Brigada de Combate Blindada, 1ª División de Infantería, en Fort Riley, Kansas. Obtuvo una licenciatura por la Universidad de East Central y una maestría en Divinidad por el Southwestern Baptist Theological Seminary. Ha servido como capellán en unidades psicológicas, artillería de campaña, sostenimiento y aviación operacional y también como oficial de inteligencia militar y de infantería.

IA general como una forma de persona no solo reconoce que puede tener competencia, como la IA débil, sino también carácter y compromiso. Será capaz de establecer y cumplir tareas aparte de las que han sido ordenadas por el comandante o acordadas dentro del equipo. Algunas tareas no estarán vinculadas a actividades militares. La IA general tendrá objetivos «personales» y trabajará para conseguirlos. Esto podría ser considerado como creatividad. Que la IA general pueda ser creativa y con el carácter que buscan las Fuerzas Armadas es importante porque esto significa que tiene la capacidad para actuar contra sus propios objetivos, en particular, los vinculados con la autopreservación.

Entendiendo la decisión de confiar en la IA

Dado que confiar en los sistemas de IA, y probablemente establecer confianza mutua, como parte de un equipo cohesivo es necesario, ¿cómo los soldados de un equipo de combate pueden prepararse para confiar? Robert F. Hurley desarrolló un modelo que permite entender la confianza y cómo esta se puede establecer¹¹. Su modelo (Decision to Trust Model, DTM) evalúa la confianza desde la perspectiva de tanto el que confía como el que recibe esa confianza. Aunque el modelo es más útil para las relaciones interpersonales entre humanos, también puede ser aplicado en relaciones impersonales como la confianza entre un individuo y una organización o un sistema como la IA. Las ambigüedades y las inconsistencias inherentes en la confianza entre humanos y sistemas de IA hacen que la aplicación de este modelo sea considerablemente más compleja que cuando se aplica entre seres humanos. Sin embargo, intentaré usarlo en este artículo para abordar cómo se llega a la decisión de confiar o no.

Hurley divide diez elementos esenciales de la confianza en dos categorías. La primera categoría está compuesta por tres factores relacionados con la disposición innata del individuo para confiar, estos son: tolerancia al riesgo, adaptación psicológica y poder relativo. Estos factores existen independientemente de la situación o en quién se confía. La disposición para confiar basada en esta categoría se podría aplicar tanto a una relación romántica como a una de negocios.

La tolerancia al riesgo de una persona influye considerablemente en su disposición para confiar. Por lo general, cuando el riesgo es alto, la confianza

es limitada. Sin embargo, los que practican el mando tipo misión están acostumbrados a confiar incluso en situaciones de alto riesgo. Cuando los comandantes confían en que sus subordinados ejecutarán la iniciativa disciplinada basándose en las órdenes de misión, lo hacen en parte porque ellos entienden cómo los líderes toman decisiones. Los líderes aprenden ciertas metodologías, como el proceso militar para la toma de decisiones y la toma de decisiones rápida, para facilitar su toma de decisiones y poder explicar cómo llegaron a ellas. Un lenguaje y procesos comunes ayudan a que los soldados confíen mutuamente porque ellos pueden imaginarse los pasos que probablemente se tomaron para llegar a la decisión. Este tipo de confianza es necesaria entre máquinas y humanos.

También es claro que la IA presenta varios grados de riesgo dependiendo de su aplicación. Algunos de los riesgos pueden ser fallos en la máquina, infiltración en el sistema por los adversarios y acciones adversas con consecuencias letales. Estos riesgos no son necesariamente una barrera para el soldado que tiene una alta tolerancia al riesgo. Por otro lado, para el soldado que tiene una baja tolerancia al riesgo, incluso el riesgo más pequeño le impedirá confiar en la IA.

El segundo factor individual, la adaptación psicológica, aborda cuán bien adaptado un individuo está. Los individuos bien adaptados tienden a sentirse cómodos con sí mismos y con su entorno. Esto permite mayor capacidad para confiar rápidamente. Aunque los miembros de las Fuerzas Armadas tienen diferentes niveles de adaptación psicológica, las Fuerzas Armadas como institución promueven y proporcionan oportunidades experimentales y educativas para mejorar la adaptación psicológica de sus miembros. El entrenamiento mejora la autoconfianza. La uniformidad ayuda a reducir las inseguridades socioeconómicas y raciales, problemas que pueden socavar una adaptación positiva fuera de la organización. La adopción y aceptación rápida de misiones, equipamiento y miembros de equipo nuevos es valorada. Todas estas cosas mejorarán la adaptación psicológica individual y ayudarán a integrar la IA.

Aunque el nivel de adaptación psicológica de los miembros de la generación más nueva es tan variado como el de generaciones anteriores, es evidente que los prospectivos soldados para el futuro próximo se sienten más cómodos con la integración de tecnología. Esto



se debe porque la tecnología se ha convertido en parte del tejido de la experiencia humana en el siglo XXI. La afinidad de la Generación Z por la tecnología ha sido bastante documentada¹². Los miembros de esta generación nacieron con la tecnología y la han adoptado durante su desarrollo. Como la IA será utilizada cada vez más en ámbitos civiles, es muy probable que los soldados futuros entren en la fuerza con la adaptación psicológica necesaria para confiar en la IA. Sus experiencias y nivel de confianza con aplicaciones militares de IA dependerán de sus experiencias con esta tecnología como civiles. Es posible que dentro de una generación, habremos prácticamente resuelto como sociedad el dilema de si los soldados deben confiar o no en la IA.

El último factor individual, el poder relativo, ayuda a determinar la disposición de un individuo para confiar basándose en el poder, o la falta de él, que este tiene sobre la persona en la que confía. Es probable que los individuos que tengan considerable poder en un grupo debido a su posición confíen más en otros

DRC-Hubo utiliza una herramienta para abrir un hueco en la pared. Este robot ganó una competencia de robótica patrocinada por la Agencia de Defensa para Proyectos de Investigación Avanzada (DARPA) el 4 de junio de 2015 en Pomona, California. (Foto: DARPA)

porque pueden castigar a los que rompen la confianza o modificar, e incluso acabar, la relación que tienen con el que confían. Si las regulaciones y las políticas relacionadas con la IA establecen que los combatientes humanos tienen supremacía sobre los sistemas de IA, esto garantizará poder relativo sobre la máquina, lo cual facilitará la confianza con ella. Si a la IA se le otorga la facultad de operar o actuar independientemente de los deseos de un humano, el poder relativo pasa a ser situacional y la confianza se vuelve más difícil.

Como se señaló en la introducción, existe el consenso de que la IA debe estar subordinada a los combatientes humanos y se ha de tener cuidado con sustituir a los humanos con IA en decisiones que tienen efectos letales. Adoptar esta posición es importante en las Fuerzas Armadas de cara al futuro. De esta manera,

los soldados ya tienen una ventaja con respecto al factor de poder relativo en la confianza. Sin embargo, a medida que aumente la integración de la IA, habrá consecuencias inesperadas que podrán alterar la dinámica de poder relativo. Por ejemplo, si un humano invalida la decisión de un sistema de IA y esto resulta en un fratricidio o daños colaterales que no habrían ocurrido si se hubiera tomado la decisión del sistema de IA, ¿se reexaminará la dinámica de poder entre los humanos y las máquinas? Tal vez la IA adquirirá más poder relativo que no tuvo al principio a medida que se emplee y tenga éxito en los equipos de combate. También es posible que las capacidades de la IA lleguen a tener más peso que las preocupaciones humanitarias de los combatientes humanos, lo cual alteraría el factor de poder relativo y la decisión de confiar.

La segunda categoría de Hurley en el modelo DTM consiste de siete factores situacionales que las personas pueden manipular para ganarse la confianza de otros: seguridad situacional, semejanzas, intereses, preocupaciones benevolentes, capacidad, predictibilidad/integridad y comunicación. Sería útil evaluar estos factores teniendo en cuenta que en quien se va a confiar puede ser un solo sistema de IA o la combinación de un sistema de IA, desarrolladores del sistema y las encargados de política que influyen en su implementación. Como la IA débil no tiene conciencia ni libre albedrío, sus creadores pueden limitar su capacidad para que no actúe fuera de los parámetros establecidos por ellos. Con respecto a los intereses como factores situacionales que influyen en la confianza, es posible que estos reflejen principalmente lo que los desarrolladores de sistemas han diseñado.

Seguridad situacional, capacidad y predictibilidad son expectativas comunes para cualquier máquina. La seguridad situacional está estrechamente vinculada al factor de tolerancia de riesgo de quien confía. Dado que hay riesgos cuando se utiliza la IA artificial en aplicaciones militares, es importante que la IA demuestre seguridad situacional, lo opuesto del riesgo. Algunos riesgos simplemente existen porque investigadores y usuarios no entienden como la IA procesa la información y toma decisiones. Esta fascinante realidad ha despertado considerable interés. DARPA ha estado invirtiendo significativamente en una IA explicable (IAX) junto con otros socios del campo de la ciencia y la tecnología. Esta «tercera

ola» de tecnología de IA «busca crear una serie de técnicas de aprendizaje automático que producirán modelos explicables mientras se mantiene un alto nivel de precisión en la predicciones para que los usuarios humanos entiendan, confíen y gestionen la generación emergente de socios con inteligencia artificial»¹³. Esta nueva tecnología busca cerrar la brecha entre las decisiones o recomendaciones hechas por un sistema de IA y la capacidad de un usuario humano para entender por qué la IA llegó a tal conclusión. Los avances en el campo de la IAX mejorarán considerablemente la seguridad situacional entre IA y humanos.

Los factores de capacidad y predictibilidad van de la mano en el campo de la tecnología y son bastante sencillos de entender en relación con la IA. Se trata simplemente de la competencia del sistema. ¿Puede la inteligencia artificial hacer lo que promociona? ¿Puede verdaderamente sobrepasar la capacidad humana en áreas como el análisis de información, desarrollo de cursos de acción o identificación de objetivos? Las experiencias con esta tecnología a larga demostrarán que la IA puede hacer lo que promociona con pocos fallos o desviaciones de la norma. La sociedad por lo general está convencida de que las máquinas son superiores a los humanos en muchas tareas. Prácticamente nadie cuestiona o revisa a mano los resultados de una calculadora que ha hecho millones de cálculos matemáticos sin fallar una vez. Probar los sistemas antes de utilizarlos garantizará su capacidad y predictibilidad. Cuando se empieza a utilizar, si la IA demuestra que puede operar sin cometer errores según sus funciones programadas, la probabilidad de que los usuarios humanos confíen en ella aumentará.

Los otros factores —semejanzas, intereses, preocupaciones benevolentes y comunicación— son más difíciles de evaluar en la relación entre combatientes humanos y sistemas de IA. Establecer semejanza e intereses comunes entre hombre y máquina es difícil. Por ello, crear sistemas de IA con una interfaz antropomórfica ayudaría considerablemente a facilitar la confianza. Establecer vínculos con la IA probablemente será más fácil si la apariencia del sistema o la manera en la que se comunica es parecida a la de los humanos. En un estudio de 2018 sobre la interacción de los humanos con robots se demostró que los humanos tienden a establecer vínculos emocionales con las

máquinas que tienen apariencia humana y se comportan como humanos¹⁴. En este estudio, algunos de los participantes interactuaron con un robot de forma social mientras que otros interactuaron de manera funcional. Al final de algunas de estas interacciones, el robot pedía que no lo apagaran. Los participantes que oían esta súplica eran más propensos a tratar al robot como si fuera una persona. El estudio concluyó que las personas tienden a considerar más humana una máquina que demuestra autonomía que una que demuestra menos. Es más fácil confiar en sistemas de IA que demuestren capacidad antropomórfica.

Es posible que los sistemas de IA débil diseñados para tareas de combate puedan tener elementos en común con el soldado e incluso compartir intereses. Un sistema de IA y un combatiente humano que operan conjuntamente en el nivel táctico demuestran semejanza porque se encuentran en el mismo ambiente y con objetivos casi similares. Es posible que los futuros sistemas de IA general tengan autoconsciencia e incluso deseos de pelear y ganar guerras. Esto demostraría intereses comunes y semejanza con los combatientes humanos.

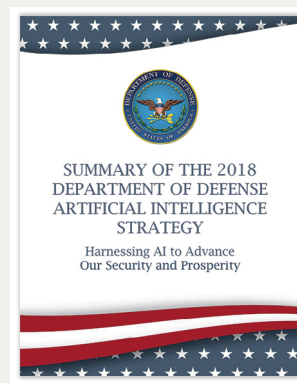
Tal vez se podrían conducir entrenamientos para que los combatientes humanos establezcan vínculos emocionales con los sistemas de IA. En el Ejército generalmente ingresan personas de diferentes orígenes, pero una vez dentro, se llega a establecer cierta uniformidad entre el personal. Es común que en los entrenamientos de formación inicial se alineen los intereses del personal recién ingresado. Los entrenamientos los unirán más y las diferencias pasarán a ser secundarias. Aunque siguen siendo diferentes, los soldados empezarán a compartir intereses y establecer elementos en común. Estas experiencias formativas y entrenamientos darán lugar a la confianza. Es posible que lo mismo ocurra con los soldados que entrenen codo a codo con los sistemas de IA.

El factor de la preocupación benevolente es la capacidad de la IA para priorizar las necesidades de los humanos por encima de las suyas. Es absolutamente necesario que la IA demuestre que entiende que los combatientes humanos están por encima de todas las partes no humanas del equipo. ¿Se autodestruirá la IA si detecta que ha sido hackeada por un adversario? ¿Sacrificará su existencia para proteger la vida de sus compañeros humanos? Incluso los humanos muchas



Dana Deasy, jefe de Información del Departamento de Defensa, y el teniente general de la Fuerza Aérea John N. T. Shanahan, director del Centro de Inteligencia Artificial Conjunto, conversan sobre la estrategia de inteligencia artificial del Departamento de Defensa el 12 de febrero de 2019 en el Pentágono. (Foto: Sargento Amber I. Smith, Ejército de EUA)

El *Resumen de la Estrategia de Inteligencia Artificial de 2018 del Departamento de Defensa*, publicado por el Centro de Inteligencia Artificial Conjunto, articula el enfoque y la metodología del departamento para acelerar la adopción de capacidades que funcionan mediante la IA con el objetivo de fortalecer nuestras Fuerzas Armadas, aumentar la eficacia de nuestras operaciones y mejorar la seguridad de la nación. Para leerlo (en inglés), visite: <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>.



veces se preocupan más por sí mismos que por los demás y este egoísmo es aceptado en situaciones en donde todo vale. Sin embargo, el altruismo es uno de los pilares del servicio militar y por lo tanto, la IA debe practicarlo. A igual que los perros de trabajo militares, la IA debe ser capaz de actuar con coraje para defender a sus compañeros de combate y proteger la misión.

Es probable que los futuros sistemas de IA general, máquinas conscientes, demostrarán el mismo coraje que los humanos. Coraje, tanto físico como moral, es una cualidad esencial de los soldados que les permite ejecutar acciones violentas para cumplir objetivos tácticos, operacionales y estratégicos. Aunque los equipos cohesivos se basan en una confianza mutua que es desarrollada principalmente a través de acciones diarias, grandes gestos como actos de valor fortalecen la confianza entre los miembros de un equipo¹⁵. Durante acciones de combate, los actos de coraje inspiran a los soldados a esforzarse más en el campo de batalla. La valentía puede ser el elemento decisivo que ayude a salir de un estancamiento, impida una derrota inminente y abata una fuerza enemiga con violencia. La IA general que se comporte de esta manera verdaderamente se ganará toda la confianza de sus compañeros humanos.

Por último, el factor de la comunicación afecta a gran parte del resto de los factores situacionales. Para fomentar la confianza es necesario tener una comunicación buena y frecuente. La comunicación con la IA probablemente será situacional. Como ya se mencionó, es difícil comunicar el proceso de toma de decisiones de la IA a los humanos y por eso se está desarrollando la IAX. Los sistemas de IA necesitarán una interfaz intuitiva para facilitar la comunicación con los usuarios. Si llegara a surgir la situación en la que la IA parece evitar la comunicación o esconde información de los combatientes humanos, es probable que la confianza se deteriore irreparablemente. Una comunicación frecuente y transparente entre los sistemas de IA y los soldados humanos facilitará y sostendrá la confianza.

Recomendaciones

La recién creada Fuerza de Tarea de IA (A-AI TF), bajo el Comando de Futuros del Ejército, fue un paso importante para el desarrollo y la implementación militar de la IA¹⁶. No se sabe que problemas éticos, si es que los hay, están siendo estudiados en profundidad como parte de los proyectos de la A-AI TF. A través de la cooperación con la A-AI TF, el Ejército puede acelerar la preparación de los combatientes humanos para que confíen en la IA de cuatro maneras. Primero, la fuerza debe estar más informada sobre los diferentes sistemas en desarrollo y su aplicación en los niveles

táctico, operacional y estratégico. La confidencialidad inherente de los proyectos de IA militares dificulta esto, pero debe haber alguna manera de promover cómo la IA será usada. No es suficiente decir: «Por ahí viene la IA». La A-AI TF y otras organizaciones vinculadas deberían encontrar maneras de comunicar sus actividades al público general del Ejército.

Segundo, la A-AI TF debería analizar los factores que influyen en cómo los individuos deciden confiar o no en los sistemas de IA. Mediante evaluaciones psicológicas, se debería analizar si la fuerza está verdaderamente preparada para confiar en la IA como si fuera un miembro más del equipo de combate. Los resultados de estas evaluaciones deberían ser publicados y también se deberían hacer recomendaciones sobre cómo establecer la confianza con la IA.

Tercero, la doctrina del mando tipo misión debe incluir el concepto de confianza entre humanos y sistemas, en particular, los sistemas de IA autónomos. Si bien en la doctrina se señala que la confianza es necesaria en equipos cohesivos, es necesario especificar que esto también es aplicable a los sistemas de IA en los equipos de combate.

Por último, los soldados necesitan evaluar si están listos para confiar en sistemas de IA, ya que esta tecnología dentro de poco transformará la manera en la que libramos la guerra. La integración de la IA transformará los equipos de combate futuro y estos cambios, de cierta manera, serán similares a los cambios sociales y operacionales que produjo la integración de la mujer en las especialidades de combate. Los soldados y los líderes tuvieron que asimilar el impacto de la integración y adaptarse a las nuevas políticas sobre las operaciones y los entrenamientos de combate. Para la integración de la IA, los soldados en todos los niveles necesitan tiempo, espacio e información adecuada para reflexionar si están listos para confiar en una máquina que es parte de su equipo de combate durante la ejecución de tareas importantes.

Conclusión

En las operaciones militares futuras, la IA será utilizada frecuentemente junto con combatientes humanos. Algunos podrán argumentar que la integración será gradual y que la confianza en la IA se establecerá de forma natural como resultado de la afinidad tecnológica actual de la sociedad. Incluso si

tal argumento es cierto, es importante entender los mecanismos que llevan a tal confianza. Es posible que en una operación de combate a gran escala sea necesario emplear sistemas de IA rápidamente y esto fracture la cohesión de los equipos de combate humanos. En tal caso, comprender cómo funciona la confianza en torno a la IA ayudará a la fuerza a superar los problemas que surjan con respecto a ella. Emplear los

conceptos doctrinales sobre la confianza y entender los factores que facilitan la confianza ayudarán a la fuerza a confiar en la IA. La ayuda de expertos en tecnología, ética y conducta, además de otros profesionales de la comunidad militar interesados en el tema, también permitirá que el Ejército alcance un alto nivel de preparación para confiar en la IA y crear equipos de combates cohesivos. ■

Notas

1. *The White House, National Security Strategy of the United States of America* (Washington, DC: The White House, diciembre de 2017), accedido 5 de agosto de 2019, <https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905-2.pdf>.
2. Department of Defense, *Summary of the 2018 National Defense Strategy of the United States of America* (Washington, DC: U.S. Government Publishing Office [GPO], 2018), accedido 5 de agosto de 2019, <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>.
3. Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity* (Washington, DC: U.S. GPO, 2018), accedido 5 de agosto de 2019, <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>.
4. Defense Advanced Research Projects Agency, *Department of Defense Fiscal Year (FY) 2020 Budget Estimates* (Washington, DC: Department of Defense, marzo de 2019), accedido 5 de agosto de 2019, https://www.darpa.mil/attachments/DARPA_FY20_Presidents_Budget_Request.pdf.
5. Amir Husain, *The Sentient Machine: The Coming Age of Artificial Intelligence* (Nueva York: Scribner, 2017), 107.
6. Andrew Ilachinski, *Artificial Intelligence & Autonomy Opportunities and Challenges* (Arlington, VA: Center for Naval Analyses, octubre de 2017), 16–17, accedido 5 de agosto de 2019, https://www.cna.org/CNA_files/PDF/DIS-2017-U-016388-Final.pdf.
7. Husain, *The Sentient Machine*, 9–48.
8. Army Doctrine Reference Publication (ADRP) 6-0, *Mission Command* (Washington, DC: U.S. Government Printing Office, mayo de 2012 [obsoleto]), párr. 2-5.
9. ADRP 1, *The Army Profession* (Washington, DC: U.S. GPO, junio de 2015 [obsoleto]), párr. 3-3.
10. Husain, *The Sentient Machine*, 89.
11. Robert F. Hurley, *The Decision to Trust: How Leaders Create High-Trust Organizations* (San Francisco: Jossey-Bass, 2012), ProQuest Ebook Central.
12. «How Generation Z Is Shaping Digital Technology», BBC Future, accedido 5 de agosto de 2019, <https://www.bbc.com/future/sponsored/story/20160309-youth-connection>.
13. *A Review and Assessment of the Fiscal Year 2019 Budget Request for Department of Defense Science and Technology Programs Before the Subcommittee on Emerging Threats and Capabilities Armed Services Committee, U.S. House of Representatives*, 115th Cong. (14 de marzo de 2018) (declaración de Steven Walker, director, Defense Advanced Research Projects Agency), 5–6, accedido el 5 de agosto de 2019, <https://docs.house.gov/meetings/AS/AS26/20180314/107978/HHRG-115-AS26-Wstate-WalkerS-20180314.pdf>.
14. Aike C. Horstmann et al., «Do a Robot's Social Skills and Its Objection Discourage Interactants from Switching the Robot Off?», *PLOS ONE* 13, nro. 7 (18 de julio de 2018), accedido 5 de agosto de 2019, <https://doi.org/10.1371/journal.pone.0201581>.
15. ADRP 6-0, *Mission Command*, párr. 2-5.
16. Mark T. Esper, Memorandum for Principal Officials of Headquarters, Department of the Army, «Army Directive 2018-18 (Army Artificial Intelligence Task Force in Support of the Department of Defense Joint Artificial Intelligence Center)», 2 de octubre de 2018, accedido 5 de agosto de 2019, https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN13011_AD2018_18_Final.pdf.