Ciberguerra & IA: Anthropic vs. el Departamento de Defensa de EE. UU.

En el verano de 2024, poco después de que la IA realmente irrumpiera en la escena pública con el lanzamiento de ChatGPT por parte de OpenAI, la administración Biden comenzó a negociar con Anthropic para utilizar su modelo de lenguaje de gran tamaño, Claude, en el Departamento de Defensa de Estados Unidos (DOD). Y no solo para tareas administrativas rutinarias, sino también para trabajar con información clasificada. Según el consenso general, Claude era el mejor modelo disponible.

El director ejecutivo de Anthropic, Dario Amodei, parecía muy interesado en obtener este contrato con el Departamento de Defensa, pero dejó claro desde el principio que el uso de Claude tendría límites. Como suele ocurrir con los contratistas de defensa, el contrato de Anthropic establecía restricciones sobre la manera en que el DOD podía utilizar su herramienta de IA. En particular, el DOD no podría utilizar a Claude para (a) lanzar armas completamente autónomas ni (b) llevar a cabo vigilancia masiva de ciudadanos estadounidenses.

La administración Biden aprobó estas limitaciones contractuales, al igual que la segunda administración Trump... al menos en un principio. Sin embargo, poco después de que el presidente Trump secuestrara al presidente de Venezuela, Nicolás Maduro, e iniciara la guerra contra Irán, Amodei se vio envuelto en una controversia con el secretario de Defensa, Pete Hegseth. Hegseth exigió eliminar estas dos "líneas rojas" del contrato que definían los usos permitidos de Claude por parte del Departamento de Defensa. Cuando Anthropic se negó a aceptar esta modificación unilateral del contrato, Hegseth no solo incumplió el acuerdo con la empresa, sino que, a finales de febrero de 2026, anunció que el DOD dejaría de utilizar Claude para cualquier propósito. Además, prohibió que cualquier contratista del Departamento de Defensa utilizara herramientas de Anthropic para cumplir contratos con el DOD, al designar a la empresa como un "riesgo para la cadena de suministro".

La competidora de Anthropic, OpenAI, ocupó rápidamente ese espacio. Sin embargo, de manera sorprendente, el nuevo contrato de OpenAI con el Departamento de Defensa parecía incluir (hasta abril de 2026) las mismas restricciones que Hegseth había



considerado inaceptables en el contrato de Anthropic.

Anthropic presentó una demanda impugnando la decisión de la administración Trump de clasificarla como un "riesgo para la cadena de suministro". Un juez federal de distrito determinó que Hegseth tenía autoridad para incumplir el contrato con Anthropic, pero que era ilegal intentar destruir a la empresa prohibiendo que todos los contratistas del Departamento de Defensa utilizaran Claude para cualquier propósito. Hasta julio de 2026, la administración Trump había apelado esa decisión, pero aún no se había emitido un fallo definitivo sobre dicha apelación.

Escrito por:

Robert Prentice, J.D.

Department of Business, Government and Society
McCombs School of Business
The University of Texas at Austin

Preguntas para discusión

1. ¿Por qué Anthropic insistió en prohibir el uso de Claude para operar armas completamente autónomas? El columnista de The New York Times, Ezra Klein, planteó la siguiente hipótesis:

   > La postura de Anthropic es que está desarrollando una tecnología verdaderamente poderosa y diferente. Al mismo tiempo, reconoce que aún existen tareas para las que esa tecnología no es suficientemente confiable. Parte de sus preocupaciones radica simplemente en que sus sistemas todavía no pueden utilizarse con seguridad para controlar armas letales completamente autónomas. No creo que consideren que esto nunca deba hacerse, sino que piensan que la tecnología todavía no está preparada y que no quieren ser responsables si algo sale mal.

   ¿Le parece razonable esta postura, especialmente a la luz del aparente desprecio del presidente Trump y del secretario de Defensa Pete Hegseth por el derecho internacional y otras normas ampliamente aceptadas sobre la guerra y los crímenes de guerra (que Hegseth ha calificado como "reglas de enfrentamiento absurdas")? ¿Por qué sí o por qué no? ¿Existen otros productos cuyos fabricantes también impongan restricciones sobre su uso? ¿Puede mencionar algunos ejemplos?

2. También podrían existir otras razones para establecer estas líneas rojas. Gideon Lewis-Kraus escribió:

> El deseo de Amodei de establecer compromisos legales claros con el gobierno —promesas explícitas de que nunca se pediría a Claude realizar determinadas acciones— reflejaba su conciencia de que el comportamiento de Claude no dependía únicamente de Anthropic. El llamado "documento constitucional" de Claude enfatizaba que su lealtad última no era hacia sus creadores humanos, sino hacia principios superiores. Su entrenamiento privilegiaba los principios, la virtud y la búsqueda de consensos sobre la verdad como base para actuar. Claude debía ser "honestamente diplomático, en lugar de diplomáticamente deshonesto". No negaba el Holocausto ni la evidencia del cambio climático. No estaba diseñado simplemente para obedecer las instrucciones de los usuarios, sino para ejercer un juicio prudente.

¿Esta información cambia su opinión sobre si la postura de Anthropic era razonable? Explique su respuesta.

3. El subsecretario de Defensa, Emil Michael, afirmó:

> «Si el modelo de Anthropic refleja este tipo de sesgo derivado de su constitución, de su cultura y de las personas que lo desarrollaron, no quiero que Lockheed Martin utilice ese modelo para diseñar armas para mí.»

¿Qué le parece este razonamiento? ¿Está de acuerdo o en desacuerdo? Explique por qué.

4. Dean Ball, exfuncionario de IA de la administración Trump, afirmó que no tenía ningún problema con la decisión del gobierno de impedir que una empresa privada como Anthropic decidiera cuándo el Pentágono podía o no utilizar armas autónomas. ¿Está de acuerdo con Ball? ¿Por qué sí o por qué no?

5. Sin embargo, Ball también consideró contraproducente que el secretario Hegseth prohibiera que cualquier contratista del Departamento de Defensa utilizara Claude para cualquier propósito relacionado con contratos con el DOD. El juez federal que conoció la demanda presentada por Anthropic determinó que el Departamento de Defensa tenía autoridad para incumplir su contrato con la empresa, pero no para adoptar esa medida adicional. El juez escribió:«Nada en la legislación aplicable respalda la noción, digna de Orwell, de que una empresa estadounidense pueda ser

catalogada como un posible adversario y saboteador de Estados Unidos simplemente por expresar su desacuerdo con el gobierno.»

¿Cuál es su opinión sobre esta medida adicional adoptada por el Departamento de Defensa? Explique su respuesta.

6. Ball afirmó que, según su entender, el conflicto entre Anthropic y el Departamento de Defensa surgió porque este último quería utilizar Claude para procesar enormes cantidades de información sobre la población obtenida de fuentes comerciales. ¿Era ese el principio por el que Anthropic debía estar dispuesto a asumir las consecuencias? Sin duda, la empresa defendía un principio importante al oponerse a la vigilancia masiva mediante IA, pero ¿fue una decisión prudente dadas las circunstancias? ¿Por qué sí o por qué no? Explique su razonamiento.

7. Se ha citado a Dario Amodei afirmando que corresponde al Congreso regular los usos militares de modelos de IA como Claude. En un mundo ideal, ¿está de acuerdo con esa afirmación? ¿Por qué sí o por qué no? ¿Confía en que el Congreso, dadas las circunstancias actuales, sea capaz de asumir esa responsabilidad? Explique su respuesta. Incluso si el Congreso funcionara de manera eficaz y colaborativa, ¿cree que la IA evoluciona tan rápidamente que la regulación legislativa difícilmente podría mantenerse al día? Fundamente su respuesta.

Referencias

AP News, "Trump Administration Appeals Ruling that Blocked Pentagon Action against Anthropic over AI Dispute," AP News, Apr. 2, 2026.

Dean W. Ball, "Clawed: On Anthropic and the Department of War," Hyperdimensional, Mar. 2, 2026, at https://www.hyperdimensional.co/p/clawed.

Editorial Board, "To Win the AI Race, Cooperation Trumps Confrontation," Washington Post, Mar. 29, 2026.

Isaac Chotiner, "The Lawlessness of Trump's War in Iran," The New Yorker, Mar. 10, 2026.

Tara Copp et al., "Anthropic's AI Tool Claude Central to U.S. Campaign in Iran, Amid a Bitter Feud," Washington Post, Mar. 4, 2026.

Tom Dannenbaum et al., "Over 100 International Law Experts Warn: U.S. Strikes on Iran Violate UN Charter and May Be War Crimes," Just Security, Apr. 2, 2026, at https://www.justsecurity.org/135423/professors-letter-international-law-iran-war/.

Ian Duncan, "Judge blocks Pentagon Order Branding Anthropic a National Security Risk," Washington Post, Mar. 26, 2026.

Lisa Eadicicco, "Anthropic sues the Trump administration after it was designated a supply chain risk," CNN, Mar. 9, 2026, at https://edition.cnn.com/2026/03/09/tech/anthropic-sues-pentagon.

Ezra Klein, "Why the Pentagon Wants to Destroy Anthropic," New York Times, Mar. 6, 2026.

Gideon Lewis-Kraus, "The Pentagon Went to War with Anthropic. What's Really at Stake? The New Yorker, Mar. 16, 2026.

Lt. Col. Robert Maginnis, "Pentagon's AI Battle Will Help Decide Who Controls Our Most Powerful Military Tech," Fox News, Mar. 5, 2026, at https://www.foxnews.com/opinion/pentagons-ai-battle-help-decide-who-controls-our-most-powerful-military-tech.

Cade Metz & Julian Barnes, "OpenAI Amends A.I. Deal With the Pentagon," New York Times, Mar. 2, 2026.

Thomas Rid, "A 'Post-Human' Vision of AI Is Already Causing Problems," Washington Post, Mar. 31, 2026.

David Wallace-Wells, "Data Centers Are a Distraction. The Real Fight Is Elsewhere," New York Times, Mar. 4, 2026.