BUSAM ANALYSIS

THE ACTORIZATION OF ARTIFICIAL INTELLIGENCE IN INTERNATIONAL RELATIONS:
A READING THROUGH THE ANTHROPIC–OPENAI DIVERGENCE

Assoc. Prof. Dr. Kıvılcım Romya Bilgin

FEBRUARY 2026

BAŞKENT UNIVERSITY
STRATEGIC STUDIES IMPLEMENTATION AND RESEARCH CENTRE

The Actorization of Artificial Intelligence in International Relations: A Reading Through the Anthropic–OpenAI Divergence

The airstrike launched by the Trump administration on 28 February 2026 against Iran, in coordination with Israel, revealed the extent to which the U.S. military has integrated artificial intelligence into its operational architecture. During the operation, the United States Central Command (CENTCOM) in the Middle East made extensive use of AI tools—most notably the large language model (LLM) Claude, developed by Anthropic—for intelligence assessment, target identification, and the simulation of combat scenarios. This case demonstrates the depth of AI integration into military systems. The intensive deployment of AI at the CENTCOM level during the operation indicates that such technologies are neither incidental nor temporary instruments. Rather, they appear to be the outcome of a long-term, institutionalized, and strategically planned integration process carried out in collaboration with specific technology firms.

Indeed, Anthropic—founded in 2021 by former OpenAI researchers—entered into an intensive and increasingly institutionalized engagement process with the U.S. Department of Defense and intelligence agencies from late 2024 onward and throughout 2025. In November 2024, through partnerships with Palantir and Amazon Web Services (AWS), the company began integrating Claude into U.S. defense and intelligence systems, including classified environments. In June 2025, Anthropic introduced Claude Gov, a version specifically designed for government and national security workflows. By the end of 2025, this version had reportedly found pilot and operational use cases within U.S. intelligence and defense institutions. Meanwhile, in July 2025, the U.S. Department of Defense (DoD) awarded artificial intelligence service contracts—valued at approximately 200 million USD in total—to several AI firms, including Anthropic. Rather than treating AI procurement as a one-time acquisition, the Department of Defense opted for a phased and institutionalized integration model. Through these contracts, AI companies and their products were gradually incorporated into U.S. military and operational processes under different contractual frameworks, reflecting a deliberate strategy of layered and incremental integration.

Under the awarded contracts, Anthropic's Claude family of large language models remained in use by the Pentagon until the contractual dispute that emerged in late February 2026. Prior to the suspension of the agreement, Anthropic continued to provide specially secured versions of its tools—designed for both governmental and classified use—through partners such as Palantir and Amazon Web Services Top Secret Cloud. Moreover, Anthropic's AI systems had been integrated into classified networks at an exceptional level, enabling their deployment for reasoning, summarization, and auxiliary analytical tasks. In this context, during contract renewal negotiations and discussions over expanding the scope of use in late February 2025, the Pentagon reportedly requested the right to use the Claude model "for all legal purposes." This demand allegedly entailed the removal of restrictions in areas such as fully autonomous lethal weapons systems and the mass surveillance of citizens, thereby triggering a serious crisis between the Pentagon and Anthropic. Secretary of Defense Pete Hegseth subsequently declared Anthropic a national security supply chain risk and prohibited all federal agencies from engaging in business with the company. President Trump, for his part, issued sharply critical statements about Anthropic via Truth Social.

Anthropic CEO Dario Amodei characterized the ban as "retaliatory and unprecedented," emphasizing that no comparable sanction had previously been imposed on an American firm. In an interview with CBS News, Amodei stated, "Being in disagreement with the government is the most American thing there is," affirming that the company stood by its decision and openly challenging the U.S. government's stance.

During this period, OpenAI adopted a position that may be characterized as more cautious yet broadly aligned with the state. While explicitly excluding the use of its models for direct combat, targeting, or lethal operations, the company maintained limited and controlled collaborations with the U.S. security bureaucracy in areas such as intelligence analysis, scenario generation, strategic simulations, and decision-support systems. Although this approach can be interpreted as a pragmatic legitimacy strategy—seeking to balance public-facing ethical commitments with the practical demands of national security—it generated significant criticism among segments of the public and AI professionals. In fact, the divergence between OpenAI and Anthropic regarding the military and security applications of artificial intelligence became particularly visible in the context of the Iran crisis. Between 2023 and 2025, OpenAI pursued a more explicit and pragmatic line in its engagement with the U.S. government, defense, and intelligence ecosystem. The company emphasized that its AI models should not function as weapons per se, but rather as tools for defense, analysis, and decision support. While formally excluding the generation of direct offensive actions, this framework nevertheless enabled the active deployment of AI in intelligence analysis, open-source intelligence (OSINT), threat scenario development, and crisis simulations. Critics have argued, however, that such uses effectively contribute to positioning AI as an operational actor in the battlespace, particularly when integrated with autonomous systems.

Anthropic, by contrast, has maintained a more normatively cautious stance toward military use since its founding. The company has repeatedly underscored the high risks associated with what it terms "frontier models"—advanced AI systems with significant capabilities—and has articulated principled objections to the use of AI in warfare. While not categorically rejecting cooperation with states, Anthropic advocates for stringent limitations on the deployment of AI in active conflict environments. In the context of the Iran conflict, the company argued that AI should not assume the role of a decision-making actor in war, but at most remain a passive and supportive analytical tool. It further warned that the use of such technologies in geopolitical crises could generate irreversible normative precedents. In this respect, Anthropic represents a more ethics-centered, precautionary approach with a clearer claim to normative authority when compared to OpenAI's comparatively pragmatic alignment with state security imperatives.

The tactical and operational integration of large language models into warfare signals a significant paradigm shift in the conduct of modern war. Experts increasingly emphasize that such tools possess a dual-use character, functioning both in intelligence processes and in contexts that may indirectly enable lethal operations. While international legal frameworks exist to regulate nuclear, chemical, and biological weapons, artificial intelligence remains situated within a regulatory grey zone, lacking a comprehensive and binding global governance regime. The United States' airstrike against Iran, coupled with the simultaneous state intervention directed at Anthropic, illustrates that contemporary warfare is no longer confined to the physical battlefield. Rather, it extends into technological infrastructures, data ecosystems, and corporate–state relations. In this respect, the Iran operation may be interpreted as one of the first major rupture points at which artificial intelligence gained visibility as a quasi-autonomous actor within international relations.