© PopTika/Shutterstock.com

How Large Language Models are Transforming Modern Warfare

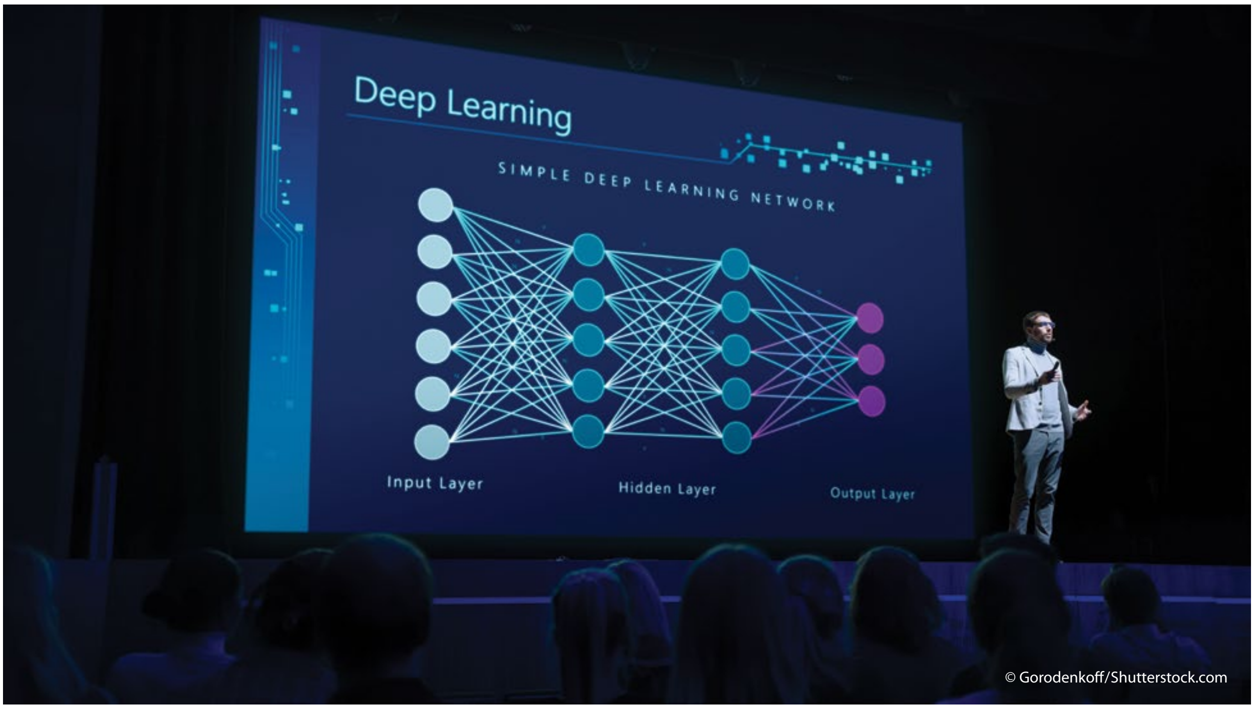
Is ChatGPT Applicable in the Military Domain?

By Lieutenant Colonel Antonios Chochtoulas, GR Air Force, JAPCC

Introduction

In October 2022, OpenAI released its novel Artificial Intelligence (AI)–driven chatbot, the famous ChatGPT (Chat Generative Pre-trained Transformer).[1,2] From that moment, the world entered a new era, where AI is at the core of the digital transformation. In the blink of an eye, the entire planet gained the privilege of using a sophisticated AI tool that can succeed in law exams, write computer code, school papers, fiction, and cooking recipes, and understand what a picture contains and draw logical conclusions, often in a human-like manner. Yet, few deeply understand what a GPT is and how it works.[3]

Although AI and Machine Learning (ML) are already successfully used for pattern recognition, filtering, and other purposes, their narrow scope focuses on a specific task. In contrast, ChatGPT and similar text-generation systems have a broader scope that is



© Gorodenkoff/Shutterstock.com

LLMs leverage Deep Learning, a subset of Machine Learning. Deep Learning involves a neural network with three or more layers, designed to mimic the complex functions of the human brain and enable it to analyze and learn from vast datasets.

inherently closer to the human domain. Their remarkable capabilities in understanding, generating, and processing human language leads to diverse private sector applications, including content creation, language translation, medical diagnosis, customer service, and scientific research.

Many individuals categorize this technology as disruptive, analyzing its impact on the global landscape. Indeed, AI solutions like ChatGPT provide individuals and businesses with robust language-processing tools, granting easier access to vast amounts of information and allowing them to process routine tasks more efficiently, thus altering how we interact with computers and transforming how we work.

This article aims to provide an overview of the technologies powering ChatGPT within the broader AI landscape. It will also present the numerous challenges associated with their deployment, propose potential military applications, and finally put forth general guidelines for possible safe and successful uses in the military that are worthy of consideration.

Generative AI and Large Language Models

ChatGPT and similar text-generation systems are powered by Large Language Models (LLMs), a form of Generative AI. The latter encompasses a wider category of AI systems, designed to autonomously generate new content, or outputs by leveraging learned patterns and data. Content-wise, this technology spans a spectrum of content types, including text, speech, video, and images, without the need for explicit instructions for each output. Unlike traditional AI systems bound by pre-programmed rules or specific inputs, Generative AI possesses the capacity to independently create new, derivative outputs that are contextually relevant.

Specifically, LLMs are statistical models, leveraging Deep Learning (DL) principles and sophisticated internal mechanisms to create word sequences in any given language, thereby generating coherent and contextually relevant text.[4,5] Their primary function involves analysing patterns and relationships within text corpora to gain the knowledge and ability to assess the statistical likelihood of specific words or

© Gorodenkoff/Shutterstock.com

Military air operations could greatly benefit by using LLM technology to enhance several activities, including those supporting planning and decision-making processes.

word sequences based on the preceding context, generating content that exhibits a natural or human-like quality.[6]

LLMs' operation comprises two primary phases: ***training*** and ***generation***. Training entails two stages. First, the model learns statistical patterns from extensive text datasets and adjusts its multi-billions of internal parameters to develop a general word prediction capability. Secondly, a fine-tuning process, utilizing human feedback to model outputs, optimizes word prediction accuracy within given contexts, thus shaping the models' final form. Once trained, the system applies its acquired knowledge to generate new output in response to prompts, continually refining its output based on previously generated content and provided context until the desired result or completion conditions are reached.

In 2020, OpenAI unveiled GPT-3, the first model that showcased remarkable performance across diverse Natural Language Processing (NLP) tasks.[7] At that time, GPT-3 excelled in text completion, translation, summarization, and question-answering, garnering considerable public attention. Its impressive self-learning capabilities allowed the model to execute tasks with minimal examples or training.[8] Its successor, GPT-3.5, ChatGPT's revolutionary model, is more powerful and offers even more extensive NLP capabilities. Introduced earlier this year, GPT-4, OpenAI's latest model, continues to push the boundaries of NLP, offering greater accuracy thanks to its broader general knowledge and advanced reasoning capabilities. In addition, this model offers both text and image input and output. [9,10]

Potential Applications of LLMs in the Military Domain

While the military and defence sectors have investigated various AI applications, including cybersecurity, maritime security, critical infrastructure protection, and others, there are no publicly known examples of LLM technology use. However, LLMs'

© WindVector/Shutterstock.com

LLMs could potentially aid Intelligence, Surveillance, Target Acquisition, and Reconnaissance (ISTAR) processes by assisting the human operator in collecting, analyzing, and assessing data in real time.

exceptional capabilities in combining and analysing raw data from diverse sources, along with their NLP capabilities, make the military domain an area with immense potential.

Military air operations could greatly benefit by utilizing this technology to enhance several processes, including planning and decision-making. For example, one possible application of AI could involve assisting military commanders in making the right decision at the speed of relevance by supporting the staff's development, assessment, and recommendation of the available Courses of Action (COAs). LLMs could also aid Intelligence, Surveillance, Target Acquisition, and Reconnaissance (ISTAR) processes, by assisting the human operator in collecting, analysing, and assessing data in real-time, thus shortening the OODA loop and providing a decisive advantage on the battlefield.[11] Another area of potential application could be military exercises, where Generative AI tools could assist in creating more realistic scenarios and even augment understaffed Red Forces, for better and more efficient training.

Challenges Associated with LLM Technology

However, it is crucial to acknowledge that the full integration of LLMs may encounter challenges, such as ensuring quality training data, refining model capabilities, managing resource costs, and addressing ethical, legal, and bias concerns.

Addressing these challenges is decisive to ensure that adopting LLMs genuinely enhances the existing processes without compromising the integrity and safety of military operations, not to mention broader societal values and interests.

Ethical Challenges

Bias in Data

It's important to note that LLMs are trained using massive datasets, which include inherent and typically insidious biases, such as geographical, gender, social, ethical, moral, and religious ones.[12,13] If these biases are not addressed, LLM outputs may perpetuate or amplify existing biases, leading to false, unfair, or discriminatory outcomes. In military operations, bias in LLM-generated information or decision-support systems could have serious consequences, including the potential for discriminatory targeting, inappropriate mission prioritization, or inadequate resource allocation.

Addressing bias requires that careful attention should be paid to the training data used, and to develop and implement bias-mitigation strategies. Researchers are working on bias mitigation techniques such as dataset curation, model fine-tuning, and continuous evaluation of the outputs, to ensure the quality of the output.[14]

The Issue of Accountability

Furthermore, the use of LLMs or any other kind of AI technology raises concerns about accountability for decisions and actions, that have been influenced by or based on AI-generated information.[15,16,17] Ensuring accountability involves transparency, traceability, and the ability to attribute decisions to specific individuals or systems. However, researchers have argued that 'the inner workings of AI and ML systems are difficult to understand by human beings and are considered black-box methods where only inputs and outputs are visible to users.'[18]

This statement questions the trustworthiness of such systems, as the opacity of LLMs' internal workings makes it challenging to pinpoint responsibility in cases where errors, biases, or controversial outputs arise. On the other hand, we should also consider the level of effectiveness and transparency in human decision-making processes, as the imperfect nature of the human brain often leads to decisions that are wrong or ineffective, difficult to explain, or influenced by bias. The limited processing capacity of the human brain could amplify this phenomenon.

> *'Military commanders might need to respond fast to complex and high tempo situations in future warfare, especially when facing near-peer competitors. In that case, using LLMs to form semi-autonomous Human-on-the-Loop (HOTL) or even autonomous Human-out-of-the-Loop (HOOTL) systems, to maintain superiority on the battlefield, might be inevitable.'*

Another aspect worth our consideration is that adversaries who prioritize operational advantages over moral and ethical considerations might adopt LLM systems despite their flaws and drawbacks. Other militaries, even inside the Alliance, could follow their example, by adopting and utilizing similarly imperfect AI solutions out of the fear of losing their advantage on the battlefield. In this possible future operational environment, the risk of compromising mission success, violating human values, and putting lives in danger may exceed our capacity to manage effectively.

Financial Challenges

Financial Cost

The economic burden of LLMs could be a significant challenge for some militaries, as the costs associated with training and running those systems, added to the essential investments required for capacity-building, can be very high.[19] Training large-scale LLMs requires a substantial financial investment, purchasing high-performance hardware, such as servers, storage, and networking equipment, and considerable energy consumption.[20] Additionally, acquiring and curating diverse datasets for optimal performance demands specialized skills and significant resources. Deploying LLMs in real-time applications further entails ongoing operational expenses, including maintenance and operating costs.[21]

'LLMs' ability to process, integrate, and analyse data from diverse sources, and to generate human-like responses to human inputs at the speed of relevance could support strategic agility, improved situational awareness, improved decision-making process, and efficient resource allocation.'

Further underlining the challenges that this technology poses, we should consider that nations constrained by defence budgets and limited resources may find it infeasible to adopt and integrate this technology, potentially leading to a technological and capability gap inside the Alliance. A solution worth investigating could be establishing mechanisms to fund and develop shared AI systems for use between North Atlantic Treaty Organization (NATO) Allies, similar to NATO's Airborne Warning & Control System (AWACS) programme.

Skilled Workforce Cost

Developing a skilled workforce is another critical aspect of capacity-building, especially considering the shortage of AI experts worldwide. Militaries should invest in training and education programmes to equip their personnel with expertise in data science, ML, NLP, and other related disciplines. Additional investment in research and development is essential to fine-tune LLMs for military applications. Research efforts should aim to improve model performance, address limitations and biases, and tailor LLMs to meet military-specific use situations.

Technical Challenges

Coherent Strategy

The successful integration of AI solutions within organizations generally hinges upon formulating a coherent strategy and robust business case.[22] For LLMs, that means militaries shouldn't make a rushed decision to adopt this technology without analysing and evaluating their processes in depth, and also considering the broader operational landscape. Otherwise, the absence of either of these two foundational elements – a coherent strategy and robust business case – will probably endanger the project's success.

Legacy Systems and Data Quality

Integrating LLM systems with existing legacy systems poses another significant challenge, as it is most likely that extensive system modifications will be required, consequently raising the risk of not meeting the desired outcome. Another critical concern pertains to the quality of data employed for training AI systems, as low-quality data can heavily impact the function of algorithms, undermining the potential for accurate outcomes and yielding consequential ramifications.

Hallucinations

There is also the issue of hallucinations when examining LLMs. This term refers to a phenomenon wherein LLMs generate plausible-sounding outputs completely fabricated, or detached from the input or context.[23,24] Hallucinations happen due to various reasons. Some include vast amounts of uncured training data, lack of contextual understanding, rare and unusual inputs, and language modelling techniques that LLMs are trained on. As a result, LLMs can occasionally produce outputs that go beyond their intended purpose or exhibit overconfidence in their responses.

Unfortunately, hallucinations and overconfident responses may not be obvious, and could pose risks in military operations, leading to misinformation, flawed decision-making, and potential mission failures. Researchers are investigating several mitigation strategies

© NicoElNino/Shutterstock.com

Given the high stakes of military operations and the legal and moral necessity for oversight, collaborative, 'teaming' arrangements are typically most appropriate; they are also more effective.

to address this issue, including human oversight and specifically designed algorithms to check the outputs continuously. In any case, we should develop and establish effective mechanisms to detect and mitigate hallucinations to ensure the reliability and validity of LLM-generated information.

NATO's Strategy on Cyber, AI and EDTs

The Alliance shows great interest in Emerging and Disruptive Technologies (EDTs) like AI, quantum technology, and autonomous systems. NATO has identified AI as one of the nine priority technology areas to focus its innovation activities. NATO's 2022 Strategic Concept states that '*Innovative technologies are providing new opportunities for NATO militaries, helping them become more effective, resilient, cost-efficient, and sustainable.*' [25,26] The same document affirms that EDTs bring both opportunities and risks, and that they are altering the character of conflict, acquiring greater strategic importance, and becoming key arenas of global competition.

Additionally, in an effort to promote the ethical use of AI systems, the United States Department of Defense (DoD) released principles for the ethical and lawful adoption of AI systems in the military in 2020, stating, among others, that '*The United States, together with our allies and partners, must accelerate the adoption of AI and lead in its national security applications to maintain our strategic position, prevail on future battlefields,*

and safeguard the rules-based international order'.[27] NATO has also released similar principles, including lawfulness, accountability, explainability (sic), traceability, reliability, and bias mitigation, to address the challenges posed by AI in the military.[28]

Conclusion

The potential use of LLMs to assist humans and enhance military processes holds great promise and could offer significant advantages for achieving operational and even strategic objectives. LLMs' ability to process, integrate, and analyse data from diverse sources, and to generate human-like responses to human inputs at the speed of relevance could support strategic agility, improved situational awareness, improved decision-making process, and efficient resource allocation. Additionally, this technology could assist in identifying blind spots, providing valuable insights, and aiding in complex cognitive tasks.

However, bias in the training data, accountability for model outputs, and potential hallucinations all highlight the importance of maintaining human oversight and responsibility in decision-making processes. Acknowledging these challenges and implementing proper mitigation mechanisms is essential for properly incorporating LLMs into military decision processes. In addition, the significant investment required to train and run these systems must be balanced with the potential benefits they bring to military operations. We should also keep in mind that some militaries will struggle to cope with the associated financial costs. In contrast, others will harness the benefits of this technology, potentially creating a technological gap inside the Alliance.

Due to the challenges and drawbacks currently associated with this technology, it is crucial to consider LLMs as supportive tools, rather than autonomous decision-makers. The human factor should remain central, with LLMs providing data-driven insights and recommendations to complement human expertise, forming Human-in-the-Loop (HITL) systems.[29] Adopting such a supportive approach can capitalize on the strengths of LLMs, while maintaining human agency, accountability, and responsibility in military operations.

Nevertheless, military commanders might need to respond fast to complex and high-tempo situations in future warfare, especially when facing near-peer competitors. In that case, using LLMs to form semi-autonomous Human-on-the-Loop (HOTL) or even autonomous Human-out-of-the-Loop (HOOTL) systems might be inevitable to maintain superiority on the battlefield.[30, 31]

While scientists and researchers are working to achieve Artificial General Intelligence (AGI), and LLMs are continuously becoming easier to implement and more efficient, their disruptive and transformative effect on society will become enormous.[32, 33] This technology's potential risk for individuals and societies is also considerable, underscoring the necessity for governments and organizations to prioritize AI regulation. Emphasizing this focus is essential to safeguard the technology, mitigate potential risks, and maximize the expected benefits. ●

1. Artificial intelligence (AI) is a set of technologies that enable computers to mimic human behavior and perform a variety of advanced functions, including the ability to see, understand and translate spoken and written language, analyze data, make recommendations, and more. Source: cloud.google.com/learn/what-is-artificial-intelligence (accessed 8 July 2023).
2. A chatbot is a computer program that simulates human conversation with an end user. Though not all chatbots are equipped with Artificial Intelligence (AI), modern chatbots increasingly use conversational AI techniques like Natural Language Processing (NLP) to understand the user's questions and automate responses to them. Source: www.ibm.com/topics/chatbots (accessed 8 July 2023).
3. Generative Pre-trained Transformers, commonly known as GPT, are a family of neural network models that uses the transformer architecture and is a key advancement in artificial intelligence (AI) powering generative AI applications such as ChatGPT. Source: https://aws.amazon.com/what-is/gpt/ (accessed 24 September 2023).
4. Zhao, Wayne Xin, et al. 'A survey of large language models', 2023, Source: ArXiv.org/abs/2303.18223 (accessed 24 September 2023).
5. Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behaviour of the human brain – albeit far from matching its ability – allowing it to 'learn' from large amounts of data. Source: https://www.ibm.com/topics/deep-learning (accessed 30 November 2023).
6. Text corpora are large and structured collections of texts or textual data, usually consisting of bodies of written or spoken text, often stored in electronic form. Source: medium.com/@evertongomede/text-corpora-4a81548dfad6 (accessed 18 July 2023).
7. OpenAI, GPT-3, Source: openai.com, https://platform.openai.com/docs/models/gpt-3 (Accessed 22 November 2023).
8. Reynolds, L. and K. McDonell, 'Prompt programming for large language models: Beyond the few-shot paradigm', 2021, In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, Source: ArXiv.org/abs/2102.07350 (accessed 24 June 2023).
9. OpenAI, GPT-4 and GPT-4 Turbo, Available from openai.com, https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo (accessed 22 November 2023).

10. OpenAI, GPT-4 Technical Report, 2023, Source: ArXiv.org/abs/2303.08774 (accessed 24 June 2023).
11. fs blog, The OODA Loop: How Fighter Pilots Make Fast and Accurate Decisions, https://fs.blog/ooda-loop/ (accessed 22 November 2023).
12. Ferrara, E., 'Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models', 2023, Source: ArXiv.org/abs/2304.03738 (accessed 28 August 2023).
13. Schramowski, P. et al., Large pre-trained language models contain human-like biases of what is right and wrong to do., 2022, Nature Machine Intelligence 4, pp. 258–268, Source: https://doi.org/10.1038/s42256-022-00458-8 (accessed 22 November 2023).
14. Ibid 10.
15. Hagendorff, T., 'The ethics of AI ethics: An evaluation of guidelines.', Minds and machines 30.1, 2020, pp. 99–120.
16. Doshi-Velez, F. et al., 'Accountability of AI under the law: The role of explanation', 2023, Source: ArXiv.org/abs/1711.01134 (accessed 4 September 2023).
17. Williams, R. et al., 'From transparency to accountability of intelligent systems: Moving beyond aspirations.', Data & Policy vol 4, 2022. Source: Cambridge Press Online, (accessed 4 September 2023).
18. Hagos, D. H. and D. B. Rawat, 'Recent advances in artificial intelligence and tactical autonomy: Current status, challenges, and perspectives', Sensors, 22(24), 2022, p. 9916.
19. Bender, E. et al., 'On the dangers of stochastic parrots: Can language models be too big?', Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021, pp. 610–623.
20. Sharir, O., P. Barak, and Y. Shoham, 'The cost of training nlp models: A concise overview.', 2020, Source: ArXiv.org/abs/2004.08900 (accessed 14 September 2023).
21. Koetsier, J., ChatGPT Burns Millions Every Day. Can Computer Scientists Make AI One Million Times More Efficient?, Forbes.com, https://www.forbes.com/sites/johnkoetsier/2023/02/10/chatgpt-burns-millions-every-day-can-computer-scientists-make-ai-one-million-times-more-efficient/?sh=67f5c3e26944 (accessed 19 September 2023).
22. Van Loon, R., AI-Assisted Decision-Making: Top 7 Challenges & Tips for Success, Linkedin.com, https://www.linkedin.com/pulse/ai-assisted-decision-making-top-7-challenges-tips-success-van-loon, (accessed 19 September 2023).
23. Ji, Z. et al., 'Survey of Hallucination in Natural Language Generation', 2022, Source: AMC Digital Library, https://doi.org/10.1145/3571730 (accessed 4 September 2023).
24. Guerreiro, Nuno M. et al., 'Hallucinations in large multilingual translation models', 2023, Source: ArXiv.org/abs/2303.16104 (accessed 4 September 2023).
25. Ibid.
26. NATO, Emerging and disruptive technologies, Source: nato.int, https://www.nato.int/cps/en/natohq/topics_184303.htm (accessed 22 November 2023).
27. US Department of Defense, DOD Adopts Ethical Principles for Artificial Intelligence, 2020, Source: defense.gov, https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/ (accessed 20 August 2023).
28. The North Atlantic Treaty Organization, Summary of the NATO Artificial Intelligence Strategy, 2021, Source: nato.int, https://www.nato.int/cps/en/natohq/official_texts_ 187617.htm (accessed 18 August 2023).
29. Docherty, B., 'Losing Humanity: The Case against Killer Robots', Human Rights Watch, 2012, https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots (accessed 10 September 2023).
30. Ibid.
31. Ibid.
32. Artificial General Intelligence (AGI) is a form of AI that possesses the ability to understand, learn and apply knowledge across a wide range of tasks and domains. It can be applied to a much broader set of use cases and incorporates cognitive flexibility, adaptability and general problem-solving skills. Source: www.gartner.com/en/information-technology/glossary/artificial-general-intelligence-agi (accessed 15 Nevember 2023).
33. Hackaday, Mozilla lets folks turn AI LLMS into single-file executables, Source: hackaday.com, https://hackaday.com/2023/12/02/mozilla-lets-folks-turn-ai-llms-into-single-file-executables/, (accessed 5 December 2023).

ABOUT THE AUTHOR

Lieutenant Colonel Antonios Chochtoulas

GR Air Force, JAPCC

Lieutenant Colonel Antonios Chochtoulas graduated from the Hellenic Air Force (HAF) Academy in 1999 with a degree in logistics. He also holds a Master's degree in Computer Science and has cultivated over two decades of diverse experience in IT and Security. He began his career as a programmer and then progressed to roles like Systems and Database Administrator, developing expertise in Information Systems Security, Windows, Linux and Unix Systems Administration, and Database Security. As the former Head of the IT Department at HAF Air Supply Command and HAF Supply Depot, he managed critical security aspects within large organizations. Currently, he is assigned at JAPCC, in the C4ISRS Branch, and he contributes to multinational projects focusing on cyberspace.