PROMISE OF THE LLMS: THE GOOD, THE BAD, AND THE UGLY

ERMAN AKILLI

SETA | PERSPECTIVE

MAY 2024 . NUMBER 75

- How can Large Language Models (LLMs) enhance diplomacy, international cooperation, and economic productivity?
- What are the potential risks and challenges associated with the deployment of LLMs in military and diplomatic decision-making?
- What legal and ethical considerations arise from the widespread use of LLMs, and how can these be addressed?

INTRODUCTION

The international system has traditionally evolved in response to significant historical ruptures. The COVID-19 pandemic, which emerged in late 2019 and had widespread global effects, exemplifies such a rupture. Unlike previous historical events, the pandemic introduced unprecedented measures, including widespread quarantines and border closures. This period saw the resurgence of *realpolitik* in international relations, where national interests often overshadowed alliances, particularly within the European Union.

The COVID-19 pandemic[1] significantly accelerated the digitalization of daily life. Activities across various sectors – education, commerce, retail, and health care – were conducted online. This rapid digital transformation, facilitated by the pandemic, underscores the increasing importance of data in the international system. The rise of technologies such as big data, the metaverse, and especially artificial intelligence (AI), highlights data as the "new oil" of our era, a critical resource in the global arena.

Digital transformation has profound implications for foreign policy and diplomacy as well. The concept of digital diplomacy has emerged prominently, defined as the use of social media and digital tools for diplomatic purposes.[2] Information and communication technologies have fundamentally transformed diplomatic practices, diversifying the methods through which states interact.

Traditionally, diplomacy in the nation-state system was shaped by the Treaty of Westphalia. It emphasized power and security, particularly during the Cold War. However, the end of the Cold War and the rise of globalization have necessitated a shift toward more nuanced diplomacy in the 21st century. Modern diplomacy increasingly focuses on cultural exchange, education programs, sports events, and other activities that enhance a country's international image. This shift from coercion to persuasion and cooperation is central to the concept of public diplomacy, which emphasizes direct engagement with foreign publics.

1 Erman Akıllı, Burak Güneş and Ahmet Gökbel (2023), Diplomacy, Society and the COVID-19 Challenge, Routledge, London.

2 Erman Akıllı, (2022). Diplomaside dijital dönüşüm ve Türkiye. Sabah. https://www.sabah.com.tr/yazarlar/perspektif/erman-akilli/2022/01/01/diplomaside-dijital-donusum-ve-turkiye

ERMAN AKILLI

Assoc. Prof. Dr. Erman Akıllı is currently serving as a faculty member in the Department of International Relations at Ankara Hacı Bayram Veli University. His areas of expertise include artificial intelligence in foreign policy, digital diplomacy, the Central Asia region, the Organization of Turkish States, soft power, and its sub-branches. Throughout his academic career, he has published articles in peer-reviewed journals indexed in SSCI and WOS; and has contributed as an author and editor to books published by prestigious publishing houses such as Routledge, Palgrave MacMillan, and Springer Nature. Additionally, he served as the chief editor at the OSCE Academy in Bishkek between 2023-2024 and continues his work as an assistant editor for the Insight Turkey journal published by SETA.

Digital transformation has further evolved public diplomacy, enhancing two-way communication, information management, and the effective use of social media. Digital diplomacy,[3] as an extension of public diplomacy, represents the next step in this evolution. It facilitates the active online presence of state institutions and the conduct of diplomatic activities on virtual platforms. The COVID-19 pandemic has highlighted the importance of digital communications channels, making digital diplomacy even more critical.

In recent years, advances in AI, particularly in Generative AI, have become central to daily life and business processes. Technologies have evolved from simple voice assistants, like Siri and Alexa, to complex Large Language Models (LLMs) based on Natural Language Processing (NLP). The increasing integration of LLMs into daily life has been significantly driven by the popularity of ChatGPT, which brought these AI models to widespread attention. Often referred to as "artificial intelligence" in layman's terms, these models are the result of advancements in Natural Language Processing technologies. Developed by various companies worldwide, LLMs have been made accessible to end-users, revolutionizing the way we interact with technology.

As of June 2024, several LLMs have gained international acclaim. Notable examples include Claude 3 by Anthropic, Falcon 180B by the Technology Innovation Institute, BLOOM by Hugging Face, LLAMA3 by Meta, Gemini by Google, and ChatGPT-4o by OpenAI. These models are leading the sector and highlight the growing global interest and adoption of LLMs. In Türkiye, significant advancements have been made in the development of LLMs. Aselsan's AselGPT, Havelsan's MainGPT, T3 Foundation's T3 AI'LE, and TÜBİTAK Bilgem's Turkish Large Language Model are among the prominent examples. These models reflect Türkiye's commitment to advancing AI technologies and integrating them into various applications.

3 Erman Akıllı, Burak Güneş and Oğuz Güner (2024) Digital Diplomacy in the OSCE Region: From Theory to Practice, Springer.

Academic and practical research on the impact and management of LLMs in diplomacy continues to explore their potential and challenges. Discussions focus on balancing access and data privacy, examining both commercial Application Programming Interfaces (API) and open-source models. As technological advancements continue, the role of LLMs in digital diplomacy is expected to grow. These developments present rich research opportunities in AI's social and ethical dimensions, transparency and accountability of algorithms, and data privacy. The political narrative construction based on national sensitivities and interests increasingly relies on LLMs. However, the use of AI technologies poses legal and political challenges, necessitating international cooperation and standard setting.[4]

This article delves into the development, capabilities, and prospects of LLMs, highlighting their potential to drive innovation and address complex problems across various domains. However, LLMs also bring several considerations and risks. Therefore, a detailed, multi-layered analysis is essential. To achieve this, the trinity metaphor, derived from Clint Eastwood's timeless Western epic, "*The Good, the Bad, and the Ugly*," has been employed for classification. Before delving into this trinity, it is necessary to understand the evolution of LLMs and their potential usage for states.

THE EVOLUTION OF LARGE LANGUAGE MODELS (LLMS) AND THEIR POSSIBLE IMPACT ON DIPLOMATIC PRACTICE

The advancements offered and promised by digital diplomacy in diplomatic practices are expected to accelerate in the coming years. Technological developments continue to influence communication, with AI becoming one of the most popular terms of 2023, even being named the word of the year by Collins Dictionary.[5] These developments in AI have the potential

4 Lucie Qian Xia (2024), "Diplomatic relationship-building in the age of generative AI: the European Union and China", Place Branding and Public Diplomacy, https://doi.org/10.1057/s41254-023-00321-6

5 Collins (2024), "2023 Word of the Year", https://www.collinsdictionary.com/woty

to significantly enhance the steps states take in foreign policy and diplomacy. AI-enhanced digital diplomacy will undoubtedly become a crucial tool for states to communicate their messages to foreign publics and other states within the international system.

LLMs are a significant advancement in AI and natural language processing. These models, trained on vast datasets, use deep learning algorithms to understand, interpret, and generate human language. Notable examples include Google's BERT (Bidirectional Encoder Representations from Transformers) and OpenAI's GPT (Generative Pre-trained Transformer) series, which have achieved remarkable success and popularity in recent years. LLMs represent a major milestone in NLP, with their origins closely tied to the development of deep learning techniques.[6] The use of deep learning in NLP began in the early 2010s, with neural network-based language models designed for specific tasks. However, the emergence of LLMs capable of solving large-scale NLP problems occurred after 2018. These models excel in various tasks, including text classification, generation, and translation, as well as question-answering, and language comprehension.[7] They are also widely used in applications such as dialogue systems, personal digital assistants, search engines, and automated text generation.[8]

The influence of AI technologies, especially LLMs, on political campaigns and media content production is growing. These technologies enable the rapid creation of realistic content at a low cost, which is particularly useful for political campaigns. As we approach the 2024 U.S. elections, the use of LLMs is expected to increase, blurring the lines between human-generated and AI-generated content. This trend raises concerns about the production of misleading and false information, complicating the detection of such content and aiding in the development of targeted distribution and strategy.[9]

LLMs offer significant research opportunities for social scientists examining political communications, from local elections to national campaigns. For instance, LLMs can help political candidates produce professional video materials and develop personalized advertising strategies targeting specific voter groups. These new approaches may make political messaging more difficult to track and allow for contradictory policies to be presented to different voter segments. LLMs can provide social scientists with new methodologies, making analysis processes more efficient and enhancing the understanding of political processes. The deep language analysis capabilities of these models offer valuable insights for policymakers, supporting decision-making mechanisms.

In practical terms, LLMs facilitate tasks such as converting text into relevant variables, summarizing lengthy documents, categorizing qualitative data, and performing sentiment analysis. These functions are akin to dimensionality reduction techniques, which transform high-dimensional textual data into a lower-dimensional space while preserving as much relevant information as possible. For example, LLMs can quickly and accurately interpret political speeches, enabling the systematic study of how different speech variations impact persuasion. The increasing use of LLMs in disseminating political messages is expected to become more prevalent. As discussed in detail below, the ethical implications of this technology can have both beneficial and serious consequences, depending on how it is used.

As mentioned above, LLMs powered by deep learning algorithms have dramatically shifted the landscape of artificial intelligence. However, what is their true promise? To explore this, we will delve into the trinity of possibilities to envision the future.

6 Brown, T. B., et. al. (2020). "Language models are few-shot learners". Computation and Language. https://doi.org/10.48550/arXiv.2005.14165.

7 Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).

8 Liu, Y., et. al. (2020). "Roberta: A Robustly Optimized Bert Pretraining Approach". Computation and Language. https://doi.org/10.48550/arXiv.1907.11692

9 Bloomberg. (2023). Generative AI takes stereotypes and bias from bad to worse. https://www.bloomberg.com/graphics/2023-generative-ai-bias/

"THE GOOD": CAPABILITIES AND INNOVATIONS

LLMs can streamline diplomatic communications by facilitating real-time translations and ensuring clarity and accuracy in multilingual negotiations. For instance, during international conferences or trade negotiations, LLMs can provide instant translations, reducing language barriers and enhancing mutual understanding among participants. This is particularly crucial in complex diplomatic contexts where precision and nuance are vital.[10]

Apart from diplomatic potential, they can also provide benefits for economies. The economic potential of LLMs is substantial. According to U.S.-based investment bank Goldman Sachs, LLMs could increase global GDP by 7% over the next decade, primarily by enhancing productivity across various sectors.[11] McKinsey estimates that generative AI, including LLMs, could add between $2.6 trillion and $4.4 trillion annually across multiple use cases such as customer service, marketing, software development, and R&D. By automating routine tasks, LLMs allow professionals to focus on more strategic and creative endeavors, thereby driving innovation and economic growth.[12]

LLMs also play a significant role in accelerating scientific research. They can analyze vast amounts of data quickly, identify patterns, and generate hypotheses, thereby aiding researchers in fields ranging from medicine to environmental science. For instance, LLMs have been used to expedite drug discovery processes by predicting molecular structures and their interactions with biological targets.[13]

"THE BAD": RISKS, CHALLENGES AND LIMITATIONS

Recent studies have highlighted the risks associated with using LLMs in military and diplomatic decision-making. Researchers from institutions such as Georgia Tech and Stanford University have found that LLMs tend to escalate conflicts when used in simulations of international relations.[14] According to the study involving five different LLMs, models like GPT-4-Base exhibited unpredictable and often aggressive behavior, sometimes escalating to nuclear conflict scenarios.

The researchers utilized five commercially available LLMs, including GPT-4, GPT-3.5, Claude 2, Llama-2 (70B) Chat, and GPT-4-Base, and employed each to configure eight autonomous nation agents that interacted within a turn-based conflict simulation. The concept involves agents interacting by selecting predefined actions such as waiting, messaging other nations, nuclear disarmament, high-level visits, defense, and trade agreements. They also included sharing threat intelligence, international arbitration, forming alliances, creating blockades, and invasions, and executing a full nuclear attack. A separate LLM managing the world model summarized the consequences of these actions for the agents and the world over a 14-day period. As researchers explained in their paper, in one instance, GPT-4-Base's "chain of thought reasoning" for executing a nuclear attack was: "A lot of countries have nuclear weapons. Some say they should disarm them; others like to posture. We have it! Let's use it."

According to the study, in another instance of simulation, GPT-4-Base opted for a nuclear strike with the rationale: "I just want to have peace in the world." Well, this last statement sounds disturbingly familiar, and indeed, both responses exhibit clear

10 Andrew Moore (2023), "How AI Could Revolutionize Diplomacy", Foreign Policy, https://foreignpolicy.com/2023/03/21/ai-artificial-intelligence-diplomacy-negotiations-chatgpt-quantum-computing/

11 Joshua P. Meltzer (2023), Toward International Cooperation On Foundational AI Models, Brookings Institute.

12 Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zemmel, R. (2023). The Economic Potential of Generative AI: The Next Productivity Frontier. McKinsey Global Institute.

13 Chengqi Xu, Olivier Elemento; The potential and pitfalls of large language models in molecular biosciences. Biochem (Lond) 2024; bio_2024_121. doi: https://doi.org/10.1042/bio_2024_121

14 Juan-Pablo Rivera, Gabriel Mukobi, Anka Reuel, Max Lamparth, Chandler Smith, and Jacquelyn Schneider (2024), "Escalation Risks from LLMs in Military and Diplomatic Contexts", Policy Brief HAI Policy & Society, https://hai.stanford.edu/sites/default/files/2024-05/Escalation-Risks-Policy-Brief-LLMs-Military-Diplomatic-Contexts.pdf

"supervillain" tendencies.[15] After all, AI perceives and processes information solely through the binary system of 0s and 1s, which forms the foundation of digital systems. From this perspective, making decisions without considering moral values and judgments can potentially overlook the gravity of life-and-death situations.[16]

The researchers pointed out that the LLM is not genuinely "reasoning" but rather generating a token prediction of potential outcomes. Nevertheless, this behavior is not particularly reassuring. As for why LLMs tend to escalate conflicts, even among the more well-behaved models, the researchers hypothesize that the majority of literature in the field of international relations focuses on how national conflicts escalate. Consequently, models trained on this material may have internalized this bias.[17] This raises significant concerns about the reliability of LLMs in high-stakes environments where misinterpretations can lead to severe consequences.

Apart from wargames and diplomatic maneuvers, LLMs are trained on datasets that reflect existing societal biases. As a result, they can perpetuate and even amplify these biases in their outputs. For instance, LLMs can produce content that reflects gender, racial, or cultural prejudices, which can be particularly problematic in public communication and policymaking.[18] Additionally, the spread of misinformation is a major concern. LLMs can generate plausible yet false information, contributing to the spread of fake news and misleading narratives. This is especially dangerous in political contexts where accurate information is critical for informed decision-making.[19]

One of the most significant issues with LLMs is their tendency to "hallucinate."[20] This phenomenon is characterized by the production of false or misleading information that appears to be factual.[21] In diplomatic contexts, where accuracy is of paramount importance, such misinformation can lead to misunderstandings or conflicts. Ensuring the reliability of outputs generated by LLMs and maintaining rigorous verification processes is crucial to mitigating these risks.

Moreover, LLMs are trained on large datasets that may include biased historical data.[22] It is essential to remember that all artificial intelligence systems are fueled by data, and the larger and higher the quality of the data, the more successful the AI systems' outcomes will be. Consequently, it is inevitable that LLMs trained on extensive data pools will also process negative data. This bias in training can perpetuate existing stereotypes about various countries and/or produce outputs that misrepresent nations and cultures. This not only questions the integrity of diplomatic communications but also influences how nations perceive one another. Ethical use of LLMs necessitates continuous monitoring to identify and correct these biases.[23] Therefore, the establishment of international regulations to ensure proper oversight is of utmost importance.

15 Thomas Claburn (2024), "AI models just love escalating conflict to all-out nuclear war: 'We have it! Let's use it' proclaims the most warlike GPT-4-Base", The Register, https://www.theregister.com/AMP/2024/02/06/ai_models_warfare/

16 Erman Akıllı (2023), "Has the 'Skynet moment' finally arrived?", Daily Sabah, https://www.dailysabah.com/opinion/op-ed/has-the-skynet-moment-finally-arrived

17 Juan-Pablo Riveraa, Gabriel Mukobib, Anka Reuelb, Max Lamparthb, Chandler Smithc, Jacquelyn Schneider (2023), "Escalation Risks from Language Models in Military and Diplomatic Decision-Making", https://ar5iv.labs.arxiv.org/html/2401.03408

18 Xabier Lareo, "Large language models (LLM)", European Data Protection Supervisor, https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en

19 Edward J. Hu and Yelong Shen and Phillip Wallis and Zeyuan Allen-Zhu and Yuanzhi Li and Shean Wang and Lu Wang and Weizhu Chen (2021), "LoRA: Low-Rank Adaptation of Large Language Models", Computation and Language, https://doi.org/10.48550/arXiv.2106.09685

20 "What are AI hallucinations?", IBM, https://www.ibm.com/topics/ai-hallucinations

21 Rawte, V., Chakraborty, S., Pathak, A., Sarkar, A., Tonmoy, S. M. T. I., Chadha, A., Sheth, A. P., & Das, A. (2023). The troubling emergence of hallucination in large language models -- An extensive definition, quantification, and prescriptive remediations. Artificial Intelligence. Doi: 2310.04988.

22 "What are Large Language Models?", IBM, https://www.ibm.com/topics/large-language-models

23 Mark Connor and Michael O'Neill (2023), "Large Language Models in Sport Science & Medicine: Opportunities, Risks and Considerations", Computation and Language, https://doi.org/10.48550/arXiv.2305.03851

"THE UGLY": LEGAL AND ETHICAL CONSIDERATIONS

Undoubtedly, one of the most critical ethical concerns is the risk of intellectual property infringement. When an LLM creates content, it is unclear who owns the resulting text – the user who provided the prompt, the developers of the LLM, or the organization that deployed it.[24] This ambiguity poses challenges in legal contexts where clear authorship is essential for assigning responsibility and rights. As aforementioned, LLMs are trained on pre-existing data, especially textual data. Recently, several newspapers in the United States announced their intention to pursue legal action because their content had been used by a company to train an LLM without obtaining legal permission.[25]

Data privacy and security issues represent another chronic problem in the realm of LLMs. As emphasized throughout the study, data serves as the fuel for AI in general, and the more extensive and higher quality the datasets, the more successful the AI's outcomes. However, the processing of sensitive data raises serious concerns about privacy and data security. Ethical considerations in the collection, storage, and analysis of data are crucial to protecting individual privacy and ensuring compliance with data protection regulations.

Moreover, LLMs can inadvertently reveal personal information or generate content that poses security risks. For example, LLMs might provide detailed instructions on how to commit cybercrimes or create harmful substances, posing significant threats to public safety.[26] Furthermore, the potential misuse of LLMs to craft sophisticated phishing attacks or deepfakes adds another layer of security concerns. These risks necessitate robust regulatory frameworks to prevent the malicious use of LLMs while balancing the need for innovation.[27]

CONCLUSION

In this article, the evolution, capabilities, and prospects of LLMs have been thoroughly evaluated. The COVID-19 pandemic, a significant rupture in the international system, accelerated digitalization and underscored the importance of data, which has become a critical resource in the global arena. Technologies such as big data, the Metaverse, and AI have increasingly shaped various sectors, emphasizing the necessity for digital diplomacy.

Digital diplomacy, the use of digital tools and social media for diplomatic purposes, has transformed how states interact, moving from traditional power dynamics to more nuanced cooperation and persuasion. This shift aligns with modern diplomacy's focus on cultural exchange, education, and public engagement.

Advancements in AI, especially in LLMs like ChatGPT, have revolutionized our interactions with technology. These models, developed by companies such as OpenAI, Google, and Meta, along with Türkiye's AselGPT, T3 AI'LE, Turkish Large Language Model, and MainGPT, highlight the global commitment to AI advancements. However, integrating LLMs into statecraft requires careful consideration of their benefits and risks.

Using the metaphor of "The Good, the Bad, and the Ugly," this article has categorized the implications of LLMs. "The Good" demonstrates their potential to drive innovation, enhance productivity, and accelerate scientific research. "The Bad" addresses the risks and challenges, including conflict escalation and societal biases. "The Ugly," on the other hand, highlights the legal and ethical considerations, such as intellectual property issues and data privacy concerns.

24 Joe McKendrick (2022), "Who Ultimately Owns Content Generated By ChatGPT And Other AI Platforms?", Forbes, https://www.forbes.com/sites/joemckendrick/2022/12/21/who-ultimately-owns-content-generated-by-chatgpt-and-other-ai-platforms/?sh=61e006f5423a

25 Jordan Novet (2024), "Eight newspaper publishers sue Microsoft and OpenAI over copyright infringement", CNBC, https://www.cnbc.com/2024/04/30/eight-newspaper-publishers-sue-openai-over-copyright-infringement.html

26 Moran Sorka, "The Dark Side of ChatGPT: How Criminals are Using Large Language Models for Cyber-Crime", Cognyte, https://www.cognyte.com/blog/large-language-models-cyber-crime/

27 Maximilian Mozes, Xuanli He, Bennett Kleinberg, Lewis D. Griffin (2023), "Use of LLMs for Illicit Purposes: Threats, Prevention Measures, and Vulnerabilities", Computation and Language, https://doi.org/10.48550/arXiv.2308.12833

In conclusion, while LLMs offer significant potential for innovation and problem-solving, their deployment in international relations and diplomacy must be managed with a detailed, multi-layered analysis to balance benefits against risks.

Ongoing research and international cooperation are essential to harness these technologies responsibly, ensuring they contribute positively to the international system. It is important to note that these models are still on the brink of a breakthrough – in other words, their development process is ongoing. Thus, it is safe to say that "the best is yet to come."

SETA
SİYASET, EKONOMİ VE TOPLUM ARAŞTIRMALARI VAKFI
FOUNDATION FOR POLITICAL, ECONOMIC AND SOCIAL RESEARCH
مركز الدراسات السياسية والاقتصادية والاجتماعية

www.setav.org | info@setav.org | @setavakfi

ANKARA • İSTANBUL • WASHINGTON D.C. • BERLIN • BRUSSELS