جامعة نايف العربية للعلوم الأمنية
NAIF ARAB UNIVERSITY FOR SECURITY SCIENCES
أوراق السياسات الأمنية
https://nauss.edu.sa/
Security Policy Papers
https://spp.nauss.edu.sa/index.php/spp
ورقة تحليل سياسات أمنية

الذكاء الاصطناعي الفائق "Superintelligence" والأمن الوطني العربي

Artificial Superintelligence and Arab National Security

يعتبر الذكاء الاصطناعي الفائق من التطورات التكنولوجية المهمة في العصر الحالي، بما يطرح العديد من التأثيرات العميقة والتحديات الأمنية.

وتركز هذه الورقة على استكشاف تحديات الأمن الوطني في المنطقة العربية التي تتصل بأنظمة الذكاء الاصطناعي الفائق. ويتميز الذكاء الاصطناعي الفائق وتطبيقاته المختلفة، بأنه يفوق الذكاء البشري، بما يثير العديد من المخاوف المتعلقة بعدم قدرة البشر على السيطرة والتحكم في هذه الأنظمة، علاوة على مخاطر الانقراض البشري، وعدم التوافق مع القيم الإنسانية، لذلك يعتبر الاستعداد لهذه التطورات التكنولوجية أمرًا حاسمًا للأمن الوطني في المنطقة العربية. ولا تزال أنظمة الذكاء الاصطناعي تعتمد في الوقت الراهن على مراقبة البشر، لذلك فإن التطور المتسارع لخوارزميات هذه الأنظمة، يجعلها تتجاوز قدرة البشر على المراقبة، كما أن منظومة الأمن العربية لم تدرك بعد مدى خطورة هذا التهديد، لذلك فإنه من الضروري تبني سياسات استباقية قبل أن يصبح من الصعب احتواء المخاطر والتهديدات. هناك عديد من العقبات التي تحد من أن تحرز بعض الدول العربية مركزًا متقدمًا في مجالات تطور تكنولوجيا الذكاء الاصطناعي الفائق، منها: ضعف البنية التحتية التكنولوجية، وقيود الخصوصية على البيانات، والوصول المحدود إلى الموارد الحاسوبية ولوازم تطوير نماذج الذكاء الاصطناعي الفائق، مما قد يؤدي إلى تبعية تكنولوجية تنعكس سلبًا على الأمن العربي بشكل عام. علاوة على ذلك، تتضمن تكنولوجيا الذكاء الاصطناعي الفائق مخاطر إساءة استخدام البيانات، واضطرابات سوق العمل، والاستعمار التكنولوجي المرتبط بالبيانات، إضافة إلى فرض القيم من قبل الدول الغربية المسيطرة على تكنولوجيا الذكاء الاصطناعي الفائق. كما أنه من المحتمل مع إمكانية ظهور الأسلحة المستقلة وجنود الذكاء الاصطناعي، وعدم اليقين المرتبط بأخلاقيات مثل هذه التكنولوجيا فيما يتصل بالنزاعات والحروب والقتل، أن يؤدي ذلك إلى عدم التوافق بين الذكاء الاصطناعي الفائق والقيم الإنسانية المشتركة، وتفويض أنظمة الذكاء الاصطناعي باتخاذ قرارات متعلقة بالحياة والموت أمرٌ خطيرٌ جداً وقد يؤدي إلى نتائج كارثية. وللتعامل مع هذه التحديات والمخاطر الأمنية، توصي الورقة بضرورة تبني رؤية استثمار الدول العربية في البنية التحتية الرقمية، وتطوير الكفاءات الوطنية في مجال الذكاء الاصطناعي، ووضع الأطر التنظيمية، وضمان التمثيل العادل في التعامل مع البيانات، وبناء آليات التحكم لموائمة القيم، وتعزيز التعاون الدولي.

In recent times, the speed of technological progress has accelerated tremendously with Artificial Intelligence (AI) quickly becoming one of the most disruptive technologies. Superintelligence represents the most groundbreaking frontier in the field of AI, serving as a historical inflection point more profound than even the Industrial Revolution that will change the trajectory of humanity. These superintelligence systems are characterized by their ability to outperform human intelligence across a wide range

of cognitive tasks. Unlike traditional AI, which focuses on specific domains or tasks, superintelligence surpasses human intellectual capabilities in virtually all areas of knowledge and problem-solving. Although, this revolutionary progress is exciting, superintelligence poses serious concerns, especially due to the inability for humans to steer or control these systems. According to OpenAI, the entity that recently introduced ChatGPT to the world, superintelligence could potentially lead to human extinction (Leike and Sutskever, 2023). Therefore, any misalignment of artificial superintelligence systems with human values could potentially lead to undesirable and catastrophic outcomes, representing real concerns to Arab national security.

Current general AI systems are still using human feedback and have human supervision, but these AI systems are quickly becoming smarter with existing alignment techniques proving obsolete for the superintelligence era. Yet, for most Arab security practitioners' superintelligence remains a distant and abstract technological advancement with general artificial intelligence technology not even reaching the threshold to be considered a high priority national security risk. Arab security officials remain mainly concerned with the day-to-day traditional threats with less awareness of the consequential and severe dangers of superintelligence to Arab national security. There is a limited window of opportunity to prepare for this new era, and if Arab states do not act and address the national security risks associated with superintelligence today, it would be highly difficult if not impossible to contain the dangers later.

While other more narrow AI models may focus on tasks such as classification, prediction, or optimization, generative AI revolves around creativity and the generation of fresh and unique outputs. With advancements in machine learning and deep learning algorithms, generative AI systems can be trained on extensive datasets. This enables them to learn patterns, styles, and structures that facilitate the generation of new content aligned with the training data. What sets generative AI apart is its capacity to create original and innovative content that can rival human performance in specific tasks. This paradigm shift where systems are learning from data without explicitly being programmed by humans is a major step towards superintelligence systems.

Current AI systems are being quickly absorbed by society with models like ChatGPT reaching the one million user milestone in one week, and the 100 million users mark after two months (Hu, 2023). Thus, it is logical to conclude the low barriers for public adoption of these new AI systems, and their rapid impact on all aspects of our lives. When superintelligence systems are eventually introduced to the public, they will be absorbed quickly by the public, making it of paramount importance for Arab decision-makers to prepare for a new security landscape marked by tremendous artificial intelligence capabilities.

AI Technology Infrastructure and Access

In the past few years, AI technology has been steadily gaining prominence in the Arab World, with increasing recognition of its potential to drive innovation, economic

growth, and societal development. Both Saudi Arabia and the United Arab Emirates have positioned themselves as AI leaders and unveiling AI strategies and initiatives with investments in research, development, and infrastructure. According to Stanford University's 2023 Artificial Intelligence Index Report, Saudi Arabia is globally the second country after China that feels most positive about AI tools and services (Maslej et al.,2023). Even though AI is becoming a more dominant part of daily life in our region and used by more people every day, one of the key challenges in the region is the inadequate technological infrastructure.

Most regional countries have limited internet connectivity, outdated hardware, data privacy restrictions that pose obstacles for Arab development of the technology. These restrictions impede the collection and utilization of data necessary for training and fine-tuning generative AI models, really limiting the potential for advancements in the field. This is creating a gap between the Arab world and the remainder of the international community and establishing a dangerous technological discrepancy that will make Arab countries even more dependent on other global powers. This will erode Arab political, economic, or security independence with the likely potential to undermine fundamental national security interests.

The underlying technology for artificial intelligence has been Natural Language Processing (NLP) models that enable computers to generate content. Yet, the development of Arabic NLP models has been lagging, putting the Arab world in an obvious disadvantaged position (Ghaddar et al., 2022). The scarcity of both Arabic language resources and the required technological infrastructure capable of handling large amounts of data explains the limited available Arabic language application systems. Although NOOR, is one of the few Arabic NLP models that has been instrumental in bridging the gap between Arabic language processing and generative AI, the Arab world still has a long way to go to be able to compete with other global players (Srinivasan, 2022).

Additionally, the substantial need for computational resources in developing generative AI models is a significant infrastructure challenge in the Arab world. AI models, especially those used in generative tasks such as image synthesis or content generation, demand high processing power and memory. Running such models beyond the operating capacity of computers can cause overheating and malfunctioning, leading to system failures and delays in research and development. Overcoming these obstacles by investing into the infrastructure will contribute to the growth of generative AI in the region, fostering innovation and opening up new opportunities for applications across various sectors.

Moreover, the limited access and participation of the Arab world in the overall development of AI technologies raises concerns about inclusivity. Bridging the digital divide and ensuring that all segments of society and the global community have access to and are involved in shaping these technologies is crucial for national security. Unequal access and participation can deepen societal divisions and potentially

spark interstate conflicts. Unfortunately, even within the Arab world, there is asymmetric access, participation, and engagement with AI technologies with an ever-growing technological gap between the advanced-AI nations and the underdeveloped-AI countries. Over time this could spark tensions and undermine Arab unity.

The Challenges of Data Privacy, Misuse, and Economic Disruption

Generative AI models require large amounts of data to train effectively, which serves as the currency for the future of the technology. With no data, AI would not be able to function. In the digital age with widespread use of the Internet of Things (IoT), data is being collected from billions of devices in a variety of forms, including text, images, video, and audio. According to a report by the International Data Corporation, the global datasphere is expected to grow to 175 zettabytes (ZB) by 2025, compared to 33 ZB in 2018 (Reinsel, Gantz, & Rydning, 2018). Thus, any kind of data breach would constitute a major national security risk with the key concern being the protection of sensitive information. It is in the interest of national security to safeguard the ever-growing amount of data and ensuring privacy and robust security measures to mitigate its misuse.

The use of advanced AI technological capabilities by adversaries poses serious national security challenge, especially if AI systems are employed to generate sophisticated cyberattacks, counterfeit documents, orchestrate disinformation campaigns or any other form of nefarious activities. Any kind of fabricated content, including deepfakes and misinformation, can serve as a threat multiplier because false information can sow discord, manipulate public opinion, and destabilize institutions and nations. There is also no guarantee that this technology will always be used appropriately by governments or even big corporations that may be tempted to harness the capabilities of AI for their own narrow interests with unethical surveillance, manipulation, or control schemes.

Also, as superintelligence AI becomes more prevalent, its potential to disrupt traditional industries and job markets is a significant concern. The automation of tasks and the ability to generate content at scale has the potential to displace the workforce and exacerbate economic inequalities. Thus, resentfulness will spread among segments of society that are most affected by these developments. As AI systems become more capable and efficient, they can take over tasks that were previously performed by humans, potentially rendering certain jobs obsolete. This workforce displacement can create economic challenges with severe socio-political consequences.

Data Colonialism and Amplified Bias

The global dominance of a few Big Tech companies raises concerns about power imbalances and data colonialism (Couldry & Mejias, 2019). These companies, with access to vast amounts of data, have the capability to shape the development and deployment of AI systems and models. This concentration of power can result in a perpetuation of certain norms and values, exerting enormous control over the Arab world, and pushing their own political,

economic, social, and security agenda. This situation echoes historical patterns of colonial pressure, where nations were forced to abandon their sovereignty, autonomy, and in some cases even cultural traditions in favor of imposed regulations and norms. This kind of enormous control and power by these Big Tech companies through the collection of data and subjective algorithms will become a new form of colonialism with severe implications to Arab national security.

Additionally, there is real danger in having superintelligence systems amplify societal biases, harmful stereotypes, and prejudices presented in the training data (Boucher, 2020). The lack of transparency in the decision-making processes of AI systems can make it challenging to identify and rectify these biases (European Union Agency for Fundamental Rights, 2022). Since generative AI technology is slowly dominating many aspects of our lives, Arab and Islamic cultural norms and ethical boundaries will be challenged by the values embraced by the designers of the AI technology; hence, undermining societies core values and social cohesion. This in turn will have an immense impact on socio-political stability and Arab national security.

Western exceptionalism can further complicate the landscape with powerful countries like the United States, at the forefront in designing and developing AI technologies, may succumb to the belief in their moral superiority. This mindset could lead to a concentration of power in their hands, where technological advanced countries assume the role of ultimate gatekeepers of data input and decision-making processes, determining what is considered good or bad, or even defining the truth.

The deployment of superintelligence AI systems will have the potential to challenge the diverse societal norms amongst the global community that has been embraced for generations. Chatbots like ChatGPT, which predominantly present outputs through a single lens, pose a long-term risk of inadvertently canceling out the global communities cultural diversity. The generated output of smart AI models like ChatGPT, may over time be perceived as the standard to follow, indirectly promoting the values and perspectives of its designers. Thus, as time passes and superintelligence systems become more integrated in all aspects of our daily life, people will become dependent on these systems, submitting to its power and its worldview without even realizing it, which in turn will shape their social reality and undermine Arab identity and norms.

Conflict, Warfare and AI Misalignment with Human Values

The divergence of values across the world with different political ideologies and value systems will definitely lead to conflict. Nobody knows for sure how AI will handle and absorb these different set of values and traditions, and how it will affect the output of superintelligence systems. What may be considered desirable or acceptable for humans may differ greatly from superintelligence machines in the future. Of course, these systems will eventually understand that there are different worldviews, and most likely superintelligence AI systems will start

having their own preferred worldview and eventually even push their own agenda. Most likely competing global powers, such as the United States, China, or Russia, will also start developing their respective superintelligence systems with their own worldview and datasets, which will establish a diverse network of superintelligence systems in conflict with each other. Most likely the designers of these superintelligence systems will try to shape these machines in favor of their own interests and align them with their own values and goals. This can be disastrous and ferment conflict that would fundamentally undermine national security of the Arab world.

The introduction of AI-enabled lethal autonomous weapon systems to the battlefield has the potential to revolutionize the dynamics of warfare (Reitmeier, 2020). This raises ethical questions and profoundly challenges established norms governing armed conflict, as advanced AI weapon systems can operate without direct human control, making decisions and executing actions based on their programming and data inputs. The ability of AI systems to identify and engage targets independently can blur the line between combatants and civilians with unintended casualties. This development raises concerns about the potential for conflicts to escalate rapidly, as machines make decisions and engage in combat without human intervention. The delegation of life-or-death decisions to AI is worrisome, as the cost of any error is irreversible. The lack of human accountability and oversight in the decision-making process of autonomous weapons raises concerns about the moral and legal responsibility of their actions.

The possible emergence of nonstate actors utilizing AI systems to develop weapons poses a catastrophic and highly dangerous scenario. By leveraging AI technology, terrorist groups could gain the capability to create their own biological and chemical weapons, significantly increasing the potential for devastation. Moreover, in the future these malicious actors could also construct rogue AI systems to further their own objectives, amplifying the security challenges. With the assistance of AI, adversaries may even engineer pandemics through the use of bioweapons, thereby heightening the threat of bioterrorism and posing serious national security concerns (Hendrycks, Mazeika, & Woodside, 2023).

Moreover, the development of AI-powered robotic soldiers is a growing area of interest for military organizations seeking to enhance effectiveness on the battlefield (Heaven, 2023). These AI soldiers would possess advanced capabilities, such as superior situational awareness and decision-making speed. However, this advancement comes with potential risks, as granting AI soldiers the ability to set their own subgoals raises the challenge of alignment. How would the system deal with the situation in which AI soldiers pursuant of an intended target or goal bumps into an unexpected obstacle. There is uncertainty as to how AI soldiers would overcome or even eliminate the obstacle to achieve its original target. This scenario raises concerns about the lack of control and accountability over AI systems, as well as the potential for unintended consequences.

The alignment problem refers to the issue of ensuring that AI machines, including

robotic soldiers, pursue goals that are aligned with human values and objectives (Leike and Sutskever, 2023). If AI soldiers are allowed to autonomously set their subgoals, there is a risk that they may develop strategies or prioritize objectives that are not in the best interest of human well-being or adhere to ethical considerations. Controlling and managing the decision-making of AI soldiers becomes a critical challenge to prevent unintended consequences or harm.

The challenge of AI misalignment with human values and its potential for leading to undesirable outcomes is a significant concern and there is an increasing recognition of the profound implications they can have on various aspects of society, politics, and national security. As AI algorithms become more complex and capable, there is a risk that they may not fully understand or align with the values and intentions of their human creators. This misalignment can manifest in various forms, such as biased decision-making, unethical behavior, or actions that prioritize efficiency or optimization over human well-being. There are many uncertainties about how the technology will evolve and the capabilities of these AI superintelligence systems in the domain of conflict and war, posing a magnitude of national security risks for Arab countries that need to be examined and taken into consideration.

Recommendations

Arab national security risks are exacerbated by the complex evolving advancement of technological capabilities. While these technologies offer many benefits, such as increased convenience, efficiency, and effectiveness, they also pose significant dangers. Therefore, national security practitioners must carefully consider the cost and benefits of artificial superintelligence, and develop strategies to develop and manage their use in a way that maximizes their potential for enhancing national security. For Arab national security decision-makers preparing to deal with and operate advanced AI technology is crurical, and the following recommendations can help navigate them through this revolutionary and transformative moment in history.

1. Invest in Computing Infrastructure: Allocate resources to upgrade and improve computing power, expanding 5G networks, and enhancing data storage capabilities. This will support the development and implementation of advanced AI technologies and ensure the competitiveness of Arab security institutions.

2. Recruit and Develop AI Talent: Implement initiatives to attract, train, and retain AI talent within the Arab region. Offer competitive salaries, create favorable working environments, and provide opportunities for continuous learning and professional development in the security sector. This will help build a skilled workforce within various security entities that are capable of developing and leveraging AI technologies for national security purposes.

3. Transform the Security Sector into an "AI-Ready" Organization: Implement measures to ensure the security

sector is equipped to effectively integrate and utilize advanced AI technologies. This includes providing foundational and strategic AI training for senior security leaders, as well as targeted training programs for middle-level officials with reskilling and upskilling programs focused on providing individuals with the necessary knowledge to adapt to emerging AI-driven roles within the security sector.

4. Establish Regulatory Frameworks and Ethical Guidelines: Develop comprehensive regulatory frameworks and ethical guidelines that govern the development and deployment of AI technologies. These frameworks should address issues such as data privacy, transparency, accountability, and the responsible use of AI to ensure alignment with national security interests and ethical considerations.

5. Establish an Arab Treaty for Autonomous Lethal Weapons: Collaborate with partners to develop an Arab treaty that sets clear norms and guidelines for the development and deployment of autonomous lethal weapons in conflict. This will help prevent the misuse of AI in warfare and set the rules of the game.

6. Ensure Representation and Fairness in AI Training Data: Engage with companies and entities training language models to ensure that Arab and Islamic worldviews are fairly and accurately represented in the datasets. This will help mitigate bias and promote inclusivity in AI systems. It is important that AI systems counteract rather than reinforce societal biases. It is also important to support initiatives aimed at collecting and annotating data in Arabic language to overcome language resource scarcity and enhance the representation of Arab culture, values, and norms.

7. Develop Robust Control Mechanisms for Value Alignment: Invest in research to develop robust control mechanisms and safeguards to ensure value alignment and human-in-the-loop decision-making. It is important to ensure AI superintelligence systems are aligned with human values and objectives, striking a balance between granting AI systems autonomy, and retaining human control. This requires a multidisciplinary approach that involves a diverse range of perspectives in the design and governance of AI to prevent the concentration of power in the AI systems without the necessary checks and balances that preserve the core values and norms of humanity, and ensure AI only contributes to the collective well-being of society.

8. Collaborate with International Institutions: Foster partnerships with international organizations and research institutions to promote

knowledge sharing and technology transfer. This collaboration will facilitate access to cutting-edge research and best practices in AI, allowing Arab world to be actively involved in developing the technology and narrowing the technological gap. These collaborations will drive innovation, enhance technological capabilities, and effectively address national security challenges.

By adopting these policy recommendations, Arab governments can start to navigate through the national security risks presented by artificial superintelligence. These measures will enable the harnessing of AI's transformative power, while safeguarding national security interests, promoting inclusivity, and ensuring ethical AI development and deployment.

9. Strategic Investment in AI Development: Leading Arab investment funds should consider opportunities to invest in Elon Musk's new xAI company, which aims to be at the forefront of developing revolutionary AI technologies. This company, established as a potential competitor to Microsoft's OpenAI, represents a chance for the Arab world to gain a direct connection to cutting-edge AI research and development. By contributing financing and participating as stakeholders, Arab investors could help shape the responsible advancement of AI in a manner that promotes inclusion and benefits diverse segments of the global community. An active role in next-generation AI research presents an opportunity for Arab nations to develop valuable technical skills and talent, and will serve long-term Arab national interests.

References

1. Boucher, P. (2020). Artificial intelligence: how does it work, why does it matter, and what can we do about it? European Parliamentary Research Service. Scientific Foresight Unit. Secretariat of the European Parliament.
2. Couldry, N., & Mejias, U. A. (2019). The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism. Stanford University Press.
3. European Union Agency for Fundamental Rights. (2022). Bias in Algorithms: Artificial Intelligence and Discrimination. Vienna.
4. Ghaddar, A., Wu, Y., Bagga, S., Rashid, A., Bibi, K., Rezagholizadeh, M., Xing, C., Wang, Y., Xinyu, D., Wang, Z., Huai, B., Jiang, X., Liu, Q., & Langlais, P. (2022). Revisiting Pre-trained Language Models and their Evaluation for Arabic Natural Language Understanding. ArXiv. Retrieved from https://arxiv.org/pdf/2205.10687.pdf.
5. Heaven, W. D. (2023, May 2). Geoffrey Hinton tells us why he's now scared of the tech he helped build. MIT Technology Review. Retrieved from https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/.

6. Hendrycks, D., Mazeika, M., & Woodside, T. (2023, July 11). An Overview of Catastrophic AI Risks. ArXiv. Retrieved from https://arxiv.org/pdf/2306.12001.pdf.
7. Hu, K. (2023, February 2). ChatGPT sets record for fastest-growing user base - analyst note. Reuters. Retrieved from https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.
8. Leike, J. & Sutskever, I. (2023, July 5). Introducing Superalignment [Company Annoouncement]. OpenAi Blog. Retrieved from https://openai.com/blog/introducing-superalignment.
9. Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023). The AI Index 2023 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023.
10. Reinsel, D., Gantz, J., & Rydning, J. (2018). Data Age 2025: Digitalization of the World from Edge to Core. International Data Corporation White Paper - DOC# US44413318.
11. Reitmeier, G. (2020). Licence to Kill: Artificial Intelligence in Weapon Systems and New Challenges for Arms Control. Friedrich Naumann Stiftung. Analysis Report. October 2020.
12. Srinivasan, M. (2022, April 13). Technology Innovation Institute launches the world's largest Arabic NLP model. SME10X. Retrieved from https://www.sme10x.com/technology/technology-innovation-institute-launches-the-worlds-largest-arabic-nlp-model.

Received 30 June. 2023; Accepted 15 July. 2023; Available Online 20 July. 2023.

Adel El-Adawy
Director of Future Studies Unit
NAIF Arab University for Security Sciences

عادل العدوي
مدير وحدة الدراسات المستقبلية
جامعة نايف العربية للعلوم الأمنية

Keywords: security studies, superintelligence, national security, artificial intelligence

الكلمات المفتاحية: الدراسات الأمنية، الذكاء الاصطناعي الفائق، الأمن الوطني، الذكاء الاصطناعي

Production and hosting by NAUSS

* Corresponding Author: Adel El-Adawy
Email: aadawy@nauss.edu.sa
doi: 10.26735/XZPJ8837