

CYBER DIPLOMACY IN THE GOVERNANCE OF EMERGING AI TECHNOLOGIES – A TRANSATLANTIC EXAMPLE

Alexandru GEORGESCU

National Institute for Research and Development in Informatics – ICI Bucharest
alexandru.georgescu@ici.ro

Abstract: Artificial Intelligence is an emerging technology which has a transformative potential in wide array of technological fields within the economic, industrial, social, political, intelligence and military domains. For this reason, the governance of AI technology development and implementation has also become a factor of concern not just among policy and decision-makers, but also among the public. With AI's potential impact on state power through its dual and strategic uses, the issue of AI governance is now firmly ensconced in global discourse and is a subdomain of cyber diplomacy. This article defines the main issues of AI governance, presents the emerging role of the EU as a normative power in this respect and also highlights the potential of transatlantic cooperation in the context of wider global rivalries in the technological field.

Keywords: Critical infrastructure, Standards, Cyber diplomacy, Competition.

INTRODUCTION

Artificial Intelligence (AI) has the potential to be the most momentous technological breakthrough since the development of nuclear fission. The final report of the National Security Commission on Artificial Intelligence in the US describes AI as so versatile that no historical parallel can do it justice, and it may only fit with how Thomas Edison described electricity, “It is a field of fields... it holds the secrets which will reorganize the life of the world”, and he characterized his knowledge of the field so far as “very little in comparison with the possibilities that appear” (Schmidt, 2021).

Without commenting on the controversy of whether Artificial General Intelligence is achievable, what we can say about existing and anticipable AI developments points to myriad applications fueling a market growth of 76 billion dollars between 2020 and 2025 (TechNavio, 2021). These applications will upend existing industries, such as those in the Industry 4.0 paradigm of increased automation, and will also enable new capabilities, making possible vast systems of surveillance, data analysis and automated decision making. The impact of AI would be extraordinary just from an economic standpoint, but its dual use nature and contribution to capabilities related to state power in the military and intelligence domains make dominance in the field a key bellwether for superpower status in the future. Drake (2022) points out that all major contenders for superpower status are investing heavily in AI, with the US planning to spend 4 billion dollars in the 2022-2023 fiscal years, Russia planning to spend 3.9 billion dollars between 2020 and 2024 (Markotkin and Chernenko, 2020) and China having already spent 1.6-5.4 billion dollars on AI in 2018 alone (Acharya and Arnold, 2019).

The individual efforts of nation-states are backed up by cooperation efforts both in development and in controlling the deployment and usage of AI. Many issues surround the existence of such a disruptive technology, from supply and production chain issues, to privacy, but also to the advantages to state power from successful AI implementation and control. Therefore, it is easy to see that, beyond the economic glamour of AI, the technology is increasingly an important part of the battleground between the West and revisionist, competing and systemically rivalrous countries. This leads, as in (related) fields such as cybersecurity, blockchain and quantum computing, to both intra-bloc cooperation on governance and, more or less futile, attempts to establish a global governance regime for AI in the context of very public warnings about the existential risks of mass deployment of advanced AI that are permeating the public consciousness.

This article analyzes the issue of AI development from a governance and cyber diplomacy perspective. It does not cover the technological development of AI, so much as inter-state coordination on how to create a usage regime compatible with a set of values, desired outcomes and collective priorities such as national security. A special focus is placed on Western cooperation, specifically on transatlantic coordination, which will play an important role in the future.

THE ISSUES IN PLAY

While Artificial General Intelligence is still a long way away from becoming a reality and it might never do so, AI is already fast becoming an integral part of decision-making and decision-support systems. This means that AI ethics and AI governance become a significant concern.

Each new and groundbreaking technology generates ethical concerns and concerns regarding the governance of its security and other forms of impact. The applicability of AI to such diverse domains as health, education, transport, online commerce, cybersecurity or defense automatically entails discussions of regulations, ethics, safety and human control over this technology (West and Allen, 2021). In the context of globalization, these discussions will also tend to become the object of inter-state cooperation and contention, as a dynamic of coo-petition or competition-cooperation is established.

If AI issues are not adequately resolved, then the inevitable implementation of AI technology, driven by economic competition, security competition or simple convenience, will result in new risks, vulnerabilities and threats, in incidents affecting lives, property and state prestige, in neo-luddite political and social movements and in the ultimate erosion of the legitimacy of governance systems in a more interconnected world.

Table 1 features the main issues surrounding AI ethics and governance, as established by the authors on the basis of their experience.

*Table 1. Main issues of AI ethics and governance
(source: authors)*

	Issue	Explanation	
AI Ethics and Governance issues	Manipulation	Using AI to exploit and manipulate people	
	Combatant	AI as defender but especially attacker of cyber systems	
	Injustice	Using AI in predictive and scoring instruments which can lead to systematic discrimination and algorithmic injustice	
	Enemy identification	AI as target selector in a weapons system in a military context	
	Decision to fire	AI as a decision-maker and trigger pusher in autonomous weapons systems	
	Legibility for authorities	AI transparency and governability for the legitimate and competent authorities	
	Political repression	Use of AI by authoritarian governments for illegitimate goals - ex: mass surveillance, detection of dissidents	
		Intelligence	The use of AI for intelligence and counter-intelligence work

Cyber Diplomacy is a natural tool for the coordination and collective action of sovereign actors with at least partly diverging interests. Diplomacy enables formalized discussions on principles, values, agendas and actions and may result in converging viewpoints and the formation of trust to establish binding governance frameworks and norms with regards to AI.

THE BRUSSELS EFFECT

The EU has, of course, prioritized developing AI capabilities, as a reflection of the interest of its Member States, a group that includes some of the richest and most innovative countries in the world. Georgescu et al (2021) highlighted some of the latest European developments in supporting research into AI and AI-related fields. Governance is also important from the perspective of the European Union, not just from the practical perspectives of AI development and implementation, but also for the potential of the “Brussels Effect”, the noted European tendency to influence governance in different fields through market instruments, multilateralism, norm setting, codes of conduct, as well as standards (Brattberg et al, 2020). The European Union is a “normative superpower” (Csernaton, 2021) and it remains to be seen whether AI will result in the same influence for the EU. AI is not just a European priority, but a component in a larger vision on European strategic autonomy, European digital sovereignty and on data sovereignty, which collectively define EU aspirations regarding its digital power and the power others have on it. The capacity of the European Union to replicate the Brussels Effect in AI will depend on whether the EU can stay at or close to the technological frontier on AI and whether it becomes a global champion in AI deployment, in the context of a technological race that is now taking place between two leading nations, the US and China.

EU technological sovereignty is based on European values and culture that emphasize human autonomy through concepts of sovereignty over data and in interaction with AI. Examples of non-AI documents with an impact on the ethics of AI use in the European vision include Europe fit for the digital age, the European Digital Strategy, the European Data Strategy, the Digital Services Act, the Digital Markets Act and others.

The actual regulation of AI ethics primarily consists of COM/2021/206, “Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts” (European Commission, 2021), and the Coordinated Plan on Artificial Intelligence 2021 (European Commission, 2021b). According to these documents, the European Union must become “a global leader in safety, trustworthy and ethical AI”. Only “common action at Union level can protect EU digital sovereignty and use regulation power and instruments to form global rules and standards” (European Commission, 2021). In addition to mandatory legislation, the EU has also developed general principles for AI ethics, voluntary codes, recommendations and other forms of non-coercive governance, including in partnership with actors such as the US.

The formula used by the EU for ethical AI is Trustworthy AI (European Commission, 2019b). There are four levels of risk for AI, with different governance measures. Figure 1 summarizes these approaches.

Unacceptable risk	High risk	Limited risk	Minimal risk
<ul style="list-style-type: none"> • A very limited set of applications that violate fundamental rights; • Totally forbidden; • Child exploitation, social scoring, subliminal influence, live biometric identification in public (with very clear exceptions). 	<ul style="list-style-type: none"> • Impact on security and on rights; • Can only be developed under certain conditions - the quality of datasets used; technical documentation and evidence; transparency and information for users; human oversight; robustness, accuracy and cybersecurity; • An obligation to provide access to data and systems to the authorities 	<ul style="list-style-type: none"> • The most important principle is that of transparency; • Users must be aware that they are interacting with a robot; • Chatbots etc.; • Manipulation risk. 	<ul style="list-style-type: none"> • The vast majority of AI systems in the European Union; • Owners can apply voluntary codes and Trustworthy AI principles.

Figure 1. EU governance principles for AI according to risk profile (source: author compilation, European Commission, 2021)

TRANSATLANTIC COOPERATION ON AI

Gehrke (2020) presaged the breakthrough in transatlantic cooperation on AI and other emerging technologies when he wrote that “with U.S.-China technological competition a defining characteristic of this decade, a transatlantic technology cooperation agenda—addressing the rules, norms, and standards governing the use of emerging and sensitive technologies—is becoming a critical aspect of foreign policy and national security”. Simple

economic cooperation and regulation would not do anymore, given the interconnectedness of various issues and the necessity of reacting to the new terms of competition set by a rapidly ascending China and a revisionist Russia.

On AI, the Trump Administration had relied on minimal regulations, for fear of damaging the innovation capacity of American industry, while cooperation with partners would emphasize American economic preeminence in AI deployment for lucrative new products. It was the Biden Administration that recognized that AI is the embodiment of the “dualism of emerging technologies, requiring cooperation, regulation and sustainable adoption of these technologies, while maintaining the resilience and values of the respective societies” (Musetescu et al, 2022). A transatlantic cooperation agenda was vital in light of the potential of cooperation between AI industries, and the considerable progress of systemic rivals and revisionist actors (Bradford and Csernaton, 2021).

Newman (2021) identified the new approach by the Biden Administration, which prioritizes the risk management for new technology deployment and which makes possible transatlantic and cooperation and coordination on AI. Key American documents emphasize, in the majority of the emerging technology fields, a priority to cooperate with partners and allies to establish a global technological (and governance) order that is favorable to them and their values. International cooperation is one of the six strategic pillars of the National AI Initiative (Schmidt, 2021). The new Administration’s approach had been presaged also by Allen (2019) in a Brookings Institution report, which is often a source for policy inspiration for Democratic leadership in the US. This was taking place concurrently with the publication of the EU’s norms on AI ethics (EC, 2019). The final report of the National Security Commission on Artificial Intelligence of the US (Schmidt, 2019) features a chapter on cooperation with partners and allies, especially the EU, and emphasizing:

- Justified trust in AI systems;
- Support for democratic values: privacy, freedom and civil rights in the use of AI for national security purposes;
- Combating malign information operations run by AI.

For the latter, Drake (2022) emphasizes the importance of AI in counter-intelligence operations, which he describes as being neglected in the current policy thinking. Certainly, with transatlantic intelligence cooperation as new heights, especially in the context of Ukraine, AI for intelligence and counter-intelligence must eventually also take center stage.

The transatlantic cooperation got a new lease on life with the first EU-US Summit in seven years in late 2021, which saw the creation of the Trade and Technology Council featuring ten working groups. One of these was dedicated to AI ethics and implementation values.

On the side of cooperation between individual states, the US has also implemented a Partnership for Defense that includes an AI dimension, with the following NATO states – the UK, Canada, Denmark, Estonia, France and Norway – and the following non-NATO

partners – Australia, Japan and South Korea in the American-defined Indo-Pacific region and Israel, Finland and Sweden in the general European area (the latter two prospective NATO members). The overlap with the EU is obvious, and Drake (2022) advocates for an extension of this useful cyber and defense diplomacy tool to Africa, as well, to counter China, including on AI.

Another area of transatlantic cooperation is the OECD, which has defined its own principles for secure AI after four expert level meetings, concluding in the adoption of a Recommendation during the OECD Council ministerial meeting on 22-23 May 2019 (OECD, 2019), as can be seen in figure 2. They serve as high level guidance for national and international frameworks that can go into more detail. Increasingly, we see international organizations with a vested interest in AI trying to preempt national divergences in this domain by establishing high level conceptual frameworks and agreements on principles that can then legitimately inform national efforts. These too are a form of cyber diplomacy.











Values-based principles	Recommendations for policy makers
 Inclusive growth, sustainable development and well-being >	 Investing in AI R&D >
 Human-centred values and fairness >	 Fostering a digital ecosystem for AI >
 Transparency and explainability >	 Providing an enabling policy environment for AI >
 Robustness, security and safety >	 Building human capacity and preparing for labour market transition >
 Accountability >	 International co-operation for trustworthy AI >

Figure 2. OECD Principles for AI
(source: OECD, 2022)

The convergence of perspectives is also evident from the compatibility between the AI principles defined by the US Department of Defense and the NATO Principles on AI, as seen in Table 2.

Table 2. AI principles defined by US DoD and NATO, for comparison (source: compilation by authors)

DoD Principles on AI (DoD, 2019)	NATO Principles on AI (NATO, 2021)
<ul style="list-style-type: none"> • Responsible. Exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities. • Equitable. The Department will take deliberate steps to minimize unintended bias in AI capabilities. • Traceable. The AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation. • Reliable. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles. • Governable. The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior. 	<ul style="list-style-type: none"> • Lawfulness AI applications will be developed and used in accordance with national and international law, including international humanitarian law and human rights law, as applicable. • Responsibility and Accountability AI applications will be developed and used with appropriate levels of judgment and care; clear human responsibility shall apply in order to ensure accountability. • Explainability and Traceability AI applications will be appropriately understandable and transparent, including through the use of review methodologies, sources, and procedures. This includes verification, assessment and validation mechanisms at either a NATO and/or national level. • Reliability AI applications will have explicit, well-defined use cases. The safety, security, and robustness of such capabilities will be subject to testing and assurance within those use cases across their entire life cycle, including through established NATO and/or national certification procedures. • Governability AI applications will be developed and used according to their intended functions and will allow for: appropriate human-machine interaction; the ability to detect and avoid unintended consequences; and the ability to take steps, such as disengagement or deactivation of systems, when such systems demonstrate unintended behaviour. • Bias Mitigation Proactive steps will be taken to minimise any unintended bias in the development and use of AI applications and in data sets.

OTHER FRAMEWORKS

While all state players with a significant ambition in the area of AI development and deployment will inevitably formulate a national governance agenda for the technology that also contains a component for cooperation and harmonization with partners, state actors are not the only highly advanced entities involved in this field. As in every other technological endeavor, multinational corporations are taking the charge in developing and deploying AI, serving as a main beneficiary of state allocation of funding for development and the main vector for deployment. Consequently, frameworks for AI governance of corporate origin are also quite common. Figure 3 presents the BMW Group's code of ethics for AI, as a company that is both cooperating with others and investing own resources in AI development for transformational effects on its core business in the automotive sector.

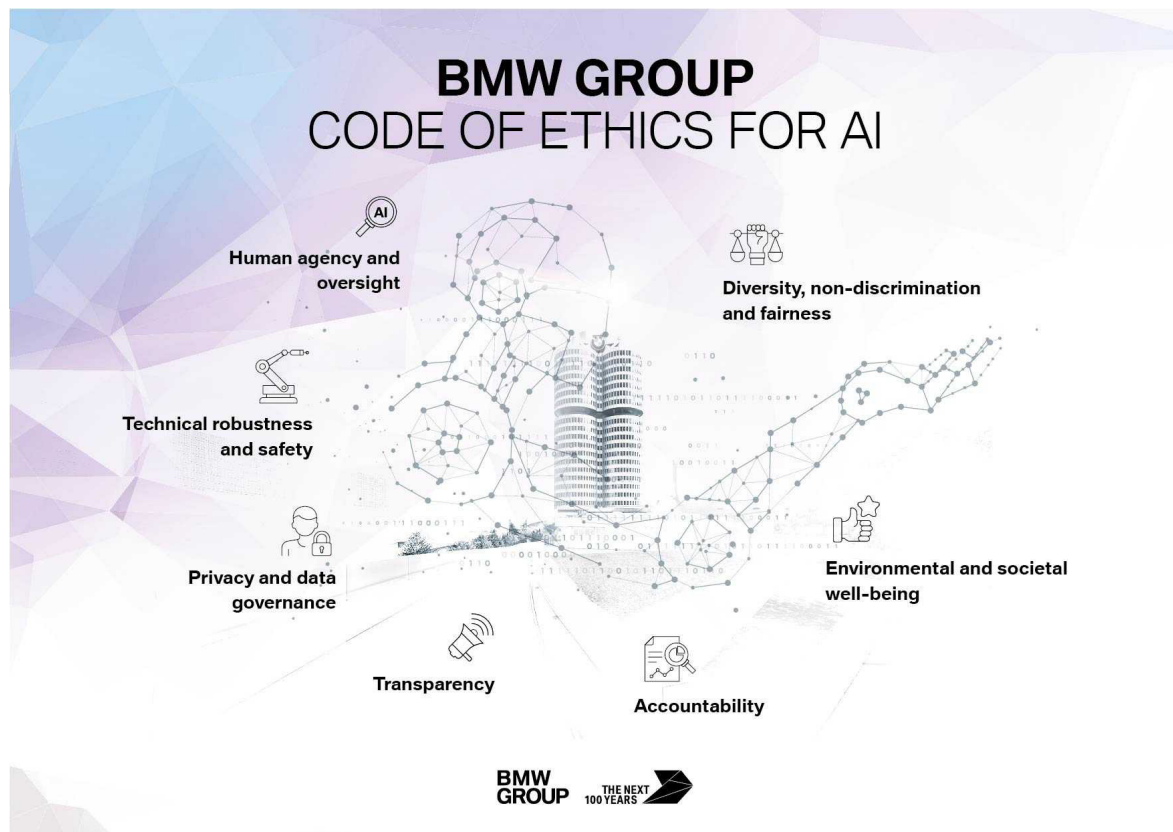


Figure 3. BMW Group Code of Ethics on AI (source: BMW, 2020)

These declarations serve as blueprints guiding internal “policy” on AI development and deployment and, therefore, have a practical effect on AI governance. At the same time, they serve as means for communicating with a wider consuming audience and a narrow policy-focused audience regarding the safety and sustainability of its AI efforts, thereby trying to alleviate concerns. Through lobbying efforts, and through concordance with like-minded economic entities such as other technology companies, the frameworks proposed by companies such as BMW will ultimately inform the perspectives of national decision makers and have an impact on the ultimate result of cyber diplomatic activity in AI ethics and governance.

More and more, we will see such governance efforts coming from the private sector when it comes to emerging technologies, because it is the private companies that are at the forefront of the technological development and the practical application’s formulation, testing and deployment. The competent authorities are often two steps behind the most advanced practitioners, especially when it comes to regulations, and so legitimate actors, such as corporations, find it expedient to get ahead of inevitable regulation drives in order to avoid over-regulation, to retain influence over the regulation process and to mitigate, early on, some of the negative effects of technological deployment. We are seeing this play out in the field of blockchain, for instance, and have witnessed it in the de facto public-private partnership for cybersecurity.

CONCLUSION

AI technology has advanced by leaps and bounds. While Artificial General Intelligence remains in the realm of science fiction, the current state of the art in AI technology is at the level where significant disruption is possible through large scale implementation. From driverless cars to truly autonomous drone swarms and pervasive surveillance systems, AI is not only an economic gamechanger, but potentially one in military, intelligence and counter-intelligence fields. This turns AI into a subject of inter-state competition, but also cooperation on issues related to the management of the impact of its adoption and the establishment of frameworks for more sustainable and safe patterns of adoption. The present article provided an overview of issues related to the governance of the widespread adoption and implementation of AI technology and some of the cyber diplomacy efforts that are shaping the Western and global frameworks of cooperation on the issue. Of particular note are the EU's ambition to become a normative power in the field, in accordance with the "Brussels effect", and the important role that transatlantic cooperation is set to play, through the inclusion of AI in a top position among cooperation initiatives between the US and EU. This is especially important, as China aims to become "the world's primary AI innovation center" by 2030 (Acharya and Arnold, 2019) and new patterns of development and forms of competition emphasize AI as a key battleground in determining the superpowers of tomorrow.

REFERENCE LIST

- *** (2019). AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense. Defense Innovation Board, Department of Defense, USA, as DoD (2019), https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF
- *** (2019). Recommendation of the Council on Artificial Intelligence. Organisation for Economic Co-operation and Development, OECD/LEGAL/0449, Adopted on 22/05/2019, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- *** (2020). BMW Group Code of Ethics on AI. BMW Group, as BMW (2020), 12 October 2020, https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/downloads/ENG_PR_CodeOfEthicsForAI_Short.pdf
- *** (2021). Artificial Intelligence (AI) Market by End-user and Geography - Forecast and Analysis 2021-2025. As TechNavio(2021), April 2021, <https://www.technavio.com/report/artificial-intelligence-ai-market-industry-analysis>
- *** (2021). Summary of the NATO Artificial Intelligence Strategy. NATO, 22 October 2021, as NATO (2021), <https://www.nato.int/docu/review/articles/2021/10/25/an-artificial-intelligence-strategy-for-nato/index.html>
- *** (2022). OECD AI Principles overview. OECD.AI Policy Observatory, Organisation for Economic Co-operation and Development, <https://oecd.ai/en/ai-principles>
- Acharya, A., Arnold, Z. (2019). Chinese Public AI R&D Spending: Provisional Findings. Center for Security and Emerging Technology, Georgetown University, December 2019, <https://cset.georgetown.edu/publication/chinese-public-ai-rd-spending-provisional-findings/>
- Bradford, A., Csernaton, R. (2021). Toward a Strengthened Transatlantic Technology Alliance. In Balfour, R. (ed.) (2021). Working With the Biden Administration: Opportunities for the EU. Carnegie Europe, 26 January 2021, <https://carnegieendowment.org/2021/01/26/toward-strengthened-transatlantic-technology-alliance-pub-83565>
- Brattberg, E., Csernaton, R., Rugova, V. (2020). Europe and AI: Leading, Lagging Behind, or Carving Its Own Way? Carnegie Europe. 9 July 2020, <https://carnegieendowment.org/2020/07/09/europe-and-ai-leading-lagging-behind-or-carving-its-own-way-pub-82236>
- Csernaton, R. (2021). The EU's Rise as a Defense Technological Power: From Strategic Autonomy to Technological Sovereignty. Carnegie Europe. 12 August 2021, <https://carnegieeurope.eu/2021/08/12/eu-s-rise-as-defense-technological-power-from-strategic-autonomy-to-technological-sovereignty-pub-85134>
- Drake, B. (2022). Protecting American investments in AI. War on the Rocks, 1 June 2022, <https://warontherocks.com/2022/06/protecting-american-investments-in-ai/>

- European Commission (2019). Ethics Guidelines for Trustworthy AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- European Commission (2019b). Ethics guidelines for trustworthy AI. High-Level Expert Group on AI, European Commission, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- European Commission (2021). COM (2021) 206 final Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. 21 April 2021, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
- European Commission (2021b). Coordinated Plan on Artificial Intelligence 2021 Review. ANNEXES to COM(2021) 205 – Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Fostering a European approach to Artificial Intelligence. European Commission, 21 April 2021, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM%3A2021%3A205%3AFIN>
- Gehrke, T. (2020) Transatlantic trade is stuck: time to integrate trade, technology and security, Commentaries, Royal Institute for International Relations, Egmont, 10 February 2020, <http://www.egmontinstitute.be/transatlantic-trade-is-stuck-time-to-integrate-trade-technology-and-security/>
- Georgescu, A., Vevera, V., Cirmu, C.E. (2021). “Opportunities for Cybersecurity Research in the New European Context”. In Romanian Cyber Security Journal, vol. 3 (1), pag 79-88, ISSN 2668-1730, ISSN-L 2668-1730, <https://rocys.ici.ro/spring2021-article-9.html>
- Markotkin, N., Chernenko, E. (2020). Developing Artificial Intelligence in Russia: Objectives and Reality. Carnegie Moscow, 5 August 2020, <https://carnegiemoscow.org/commentary/82422>
- Muşetescu, R.C., Volintiru, C.A., Georgescu, A., Franţescu, D.P. (2022). The consolidation of the EU-US relationship in the new geopolitical context, including from the perspective of managing emerging technologies. Opportunities for Romania. Studies in Strategies and Policies SPOS 2021, European Institute of Romania, http://ier.gov.ro/wp-content/uploads/2022/03/Studiul-5_Relatia-UE_SUA_final_site.pdf
- Newman, J. (2021). Now is the Time for Transatlantic Cooperation on Artificial Intelligence. Georgetown Journal of International Affairs, 13 July 2021, <https://gjia.georgetown.edu/2021/07/13/now-is-the-time-for-transatlantic-cooperation-on-artificial-intelligence/>.
- Schmidt, E. (coord.) (2021). Final Report - National Security Commission on Artificial Intelligence. NSCAI. Washington DC, US. <https://www.nsc.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>.
- West, D., Allen, J. (2020). Turning Point: Policymaking in the Era of Artificial Intelligence. Brookings Institution Press. p. 324.



Alexandru GEORGESCU

Is an Expert with the Department for Cybersecurity and Critical Infrastructure Protection of the National Institute for Research and Development in Informatics. He has an eclectic background, having studied Economics, then Geopolitics, and has obtained a PhD in Risk Engineering for Critical Infrastructure Systems. He is actively involved in advancing Critical Infrastructure Protection and Resilience issues through cooperation at international level and has worked on international projects for the European Space Agency, the Shanghai Institutes for International Studies and others. He is currently also moderating a Working Group on the Protection of Defense-related Critical Energy Infrastructures within the European Defence Agency's Consultation Forum for Sustainable Energy in the Security and Defence Sectors. He is also affiliated with the European Center for Excellence for Blockchain, with the Romanian Association for Space Technology and Industry, the EURISC Foundation and Eurodefense. Coupled with significant International exposure, he is emerging as a notable member of a new generation of Romanian security experts.