Reliability of artificial intelligence in civilian and military applications

Slavko J. Pokorni
Information Technology School, Belgrade, Republic of Serbia,
e-mail: slavko.pokorni@its.edu.rs,
ORCID iD: https://orcid.org/0000-0002-3173-597X

FIELD: Information technology
ARTICLE TYPE: Review paper

https://doi.org/10.5937/vojtehg73-56016

Abstract:

Introduction/purpose: The goal of this paper is to show the importance of reliability and availability regarding artificial intelligence (AI) which is nowadays applied in almost every area of human life, both civilian and military. Everything can fail, including AI as any other product or device. AI also needs to be reliable. It is mainly realized as software, but also involves hardware, data and the human factor, so it is complex for building in and for reliability assessment.

Methods: This study has used the published articles of this author as well as some other papers and documents.

Results: The research has resulted in brief definitions of reliability, availability, artificial intelligence, reliability of hardware, software, data, the human factor, and AI as well as in a brief analysis of some reasons for the unreliability of AI, such as data quality, a small or insufficient amount of training and test data, training on unreal data, the nature of learning algorithms and user incompetence.

Conclusion: The reliability of artificial intelligence is very complex and due to its importance it must be considered during AI design and application.

Key words: artificial intelligence, reliability, availability, hardware, software, human factor, data.

Introduction

Everything can fail, and artificial intelligence (AI) is not an exception. If AI is an attempt to replace human intelligence with machine intelligence, and human reasoning can sometimes fail, so AI can fail in a similar way

This paper, under a similar title, was presented at the *11th International Scientific Conference on Defensive Technologies OTEH 2024*, 09-11 October 2024.

(Pokorni, 2021, Pokorni, 2024). In the sense of AI reliability, failure can mean that, when needed, it will not react, or it will not react as expected. Artificial intelligence is mainly realized as software, but it involves hardware, data and the human factor - and software, hardware, and people can also fail. If something can fail, one speaks about reliability, and if something can be repaired, one speaks about maintenance.

Reliability is related to failures, and failures can cause consequences from minor to major ones, for example, material and financial ones, and sometimes they can lead to the loss of human life. Reliability is thus important, and sometimes critical, depending on the area where AI is applied (airplanes and autonomous vehicles both civilian and military, medical devices, products and services based on people's interest, etc.). It implies that the reliability of AI is important in order to use it with confidence and safety for the benefit of all people, without its misuse.

As mentioned before, artificial intelligence is mainly realized as software, but it also involves hardware, data and the human factor. Therefore, when AI reliability is concerned, it obviously includes hardware, software, data and the human factor reliability.

The importance of reliability can be illustrated by the following statement from a well-known book of (Kececioglu, 2002): "No industry in any country can progress effectively without the knowledge and implementation of reliability engineering". This also applies to AI.

Definition of reliability and availability

In general, reliability is defined as the ability of an item to perform a required (expected) function under stated conditions for a stated period of time (Pokorni, 2021b). Instead of the term *item*, this paper will use the term *product* as a general term, and this will include a device, an element or a system realized in hardware, software or both. AI can also be treated as a product, so the reliability of AI can be defined in a similar way as the reliability of a product.

Quantitatively, reliability is expressed as a probability, and it is very important in reducing downtime and both operational and maintenance costs of a product. Reliability and maintenance are mutually connected. Higher reliability means less costly maintenance. There are many factors which can influence reliability. For example, reliability of hardware can change if the environment changes (temperature, mechanical stresses, etc). Similarly, the reliability of AI changes if data and user behaviour change.

High reliability is of high importance, especially in professional equipment such as military, medical or driverless cars, and it comprises hardware reliability, software reliability, and human reliability (Pokorni, 2021a); it can also include AI, so the reliability of AI is significant as well.

In calculating reliability, prognostic reliability is essentially calculated. That is why one does not talk about determination, but about reliability assessment or reliability prognostic. Reliability prognostic has been done since the sixties of the last century using MIL-HDBK-217, mostly for electronic hardware (Pokorni, 2016), but it has not been updated since the nineties of the last century.

Reliability is built in during the product design phase, provided in production and supported in use. It is also connected with cost. A more reliable product is more expensive, but a more reliable product is also cheaper for maintenance.

Artificial intelligence is, in essence, software but it also involves hardware, data and humans. Having this in mind, one can define the reliability of AI starting from the definition in (Pokorni, 2023): “reliability is the probability that a product will meet the intended standards of performance and deliver the desired results within a specified period of time under specified (environmental) conditions”.

Availability is a metric used to assess the performance of repairable systems, incorporating both the reliability and maintainability properties of a component or a system. There are different definitions of availability and different ways to calculate it (Pokorni, 2021a).

Unlike reliability, availability is a probability whether a product is ready to perform its function when it is required, and represents a characteristic of repaired devices.

Reliability of hardware

AI as software runs on hardware, so the reliability of AI depends on the reliability of that hardware. As said in (Pokorni, 2021a), up to now, hardware reliability has been calculated mostly using MIL-HDBK-217 military manual for the calculation of the reliability of electronic devices.

Hardware usually comprises different components concerning quality and reliability: very often of a commercial type, without established reliability, and very often without any data about the failure rate or the mean time to failure (MTTF), or the mean time between failures (MTBF), thus making precise reliability calculation very difficult.

Since the sixties of the previous century, hardware reliability has been calculated mostly using MIL-HDBK-217 military manual. However, MIL-

HDBK-217 has its limitations and has not been updated since 1995. RIAC's 217Plus^{TM} methodology and a software tool is a replacement for MIL-HDBK-217, but it is not free (Pokorni, 2021a).

Reliability of software

Since AI is predominantly a type of software, it is connected with software reliability. Software errors are the dominant cause of software unreliability. Software can contain errors, and errors can produce faults. Errors in software are produced by programmers, so it is important to be familiar with these errors.

Software reliability is an important attribute determining the quality of the software as a product. There are many models of software reliability assessment, but none of them is generally accepted (Pokorni, 2016, Pokorni, 2021a). The problem is that the requirements for the reliability of software are often not adequately specified, if specified at all, and the problem is in the different nature of software compared to hardware.

Reliability assessment of software is more complicated than hardware reliability. The problem also lies in the different natures of software and hardware.

Reliability of the human factor

AI can learn from humans, and human knowledge can be erroneous, so it is connected with human reliability.

There are different approaches and models to human reliability assesment (Pokorni, 2016).

Procedures, rules, codes, standards and laws cannot completely prevent system failures, but, in this author's experience, they can reduce system failures.

Humans can be and are involved in the artificial intelligence system. A human action can influence the reliability or unreliability of the artificial intelligence system.

This author has considered human reliability important from the beginning of his work in reliability; hence, human reliability is included in his textbooks published in the Military Academy and the Information Technology School in Belgrade, Serbia.

Reliability of data

Data reliability means that data is complete and accurate, and it is a crucial foundation for building data trust across any organization. Ensuring data reliability is one of the main objectives of data integrity initiatives, which are also used to maintain data security, data quality, and regulatory compliance (Pokorni, 2023).

Data is crucial for AI. AI uses data and can learn from data, and data can be accidentally or intentionally corrupted, so AI can fail. Consequently, AI reliability is connected with data reliability.

Data is also very important in calculating reliability of hardware. Reliability is not easy to calculate or evaluate. In calculation reliability of hardware, input data is the biggest problem. Not because there is too much data, but because sometimes there is too little data or no data at all. When dealing with maintenance, the problem is very often that there is not enough data.

How much data does one have about AI reliability, especially if they deal with a new application?

Definition of artificial intelligence

There is no generally accepted definition of artificial intelligence (Government of the Republic of Serbia, 2019). According to the Encyclopaedia Britannica Dictionary (Pokorni, 2024), artificial intelligence is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.

According to the Merriam-Webster Dictionary, AI is a branch of computer science dealing with the simulation of intelligent behaviour in computers, or the capability of a machine to imitate intelligent human behaviour.

The Government of the Republic of Serbia (Government of the Republic of Serbia, 2019, Sl. glasnik RS, 2023) has accepted the following definition of AI, also used in (European Commission, 2019): “Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.”

Artificial intelligence-based systems can be purely software-based and operate in a virtual world (for example: virtual assistants, photo analysis software, web browsers, recommendation systems, speech and face recognition systems) or they can be embedded in devices - hardware (for example: advanced robots, autonomous vehicles, drones, etc.) (Government of the Republic of Serbia. National AI Platform, 2023, European Commission, 2019).

Artificial intelligence was founded as an academic and scientific discipline in the middle of the fifties of the last century, and since then its development has gone in different directions, being divided in sub-fields. Therefore, it is not a surprise that the definition of AI has been changed during time.

In the history of AI, there have been ups and downs, starting with the logic-based approach (during the 1950s and 1960s), the knowledge-based expert systems approach (1970s and 1980s), and the data-based approach (since 2000) years onwards - with periods of disappointment and reduced investment (Government of the Republic of Serbia, 2019). In the last decades, AI has been defined as a study of intelligent agents - any device that perceives its environment and takes actions (by learning or using knowledge) to achieve its goals (Pokorni, 2021b).

Essentially, the use of AI is an attempt to replace human intelligence with machine intelligence. Because of that, AI is sometimes called machine intelligence. However, it does not mean that AI learns just by copying humans. Nowadays, AI can learn in its own way, and it, maybe, gives it a potential to outperform the human (brain).

Reliability of artificial intelligence

As mentioned before, everything can fail, and artificial intelligence is not an exception. If AI is an attempt to replace human intelligence with machine intelligence, and human reasoning can sometimes fail, so can AI fail in a similar way. Therefore, does this happen because of erroneous reasoning (erroneous concluding, decision making) or wrong learning? Having that in mind, can we raise the question about the reliability of AI, or how to avoid AI to fail? (Pokorni, 2021b).

If AI fails, this means it is unreliable – ergo, can we trust it?

In (University of Cambridge, 2016) under the title “Enhancing the reliability of artificial intelligence“, it is stated that “computers that learn for themselves are with us now. As they become more common in 'high-stakes' applications like robotic surgery, terrorism detection and driverless cars, researchers ask what can be done to make sure we can trust them.“ So, are they reliable? Or, can they fail? Or can we fool them?

Deep learning AI can easily be fooled. An example is in a self-driving car application in a real situation. But it can happen in a case of sabotage as well.

There are examples of erroneous AI. From these examples, we can derive some reasons for AI to fail, such as: quality of data, a small or insufficient amount of training and test data, training on unreal data, the nature of learning algorithms, and user incompetence.

Definition of the reliability of artificial intelligence

In the document entitled “Ethical Guidelines for Development, Implementation and Use of Robust and Accountable AI” (Government of the Republic of Serbia, National AI Platform), so-called technical reliability means that “systems are developed under constant risk assessment and prevention, and that they behave reliably and as intended, while minimising possible unintended and unforeseen damage”.

The reliability of an artificial intelligence system is defined as the probability, at a certain level of confidence, that the system will successfully, without failure, perform the function for which it is intended, within the specified performance limits, during the specified duration of the tasks, when it is used in the prescribed manner and for the purpose for which it is intended, under defined load levels, taking into account the previous system usage time.

In [NIST, nd], reliability is defined in the same standard as the “ability of an item to perform as required, without failure, for a given time interval, under given conditions” (Source: iso/iec ts 5723:2022). Reliability is a goal for the overall correctness of AI system operation under the conditions of expected use and over a given period of time, including the entire lifetime of the system.

A wider term is *trustworthy*. According to (NIST, nd), for AI systems to be trustworthy, they often need to be responsive to a multiplicity of criteria that are of value to interested parties. Approaches which enhance AI trustworthiness can reduce negative AI risks. This framework articulates the following characteristics of trustworthy AI and offers guidance for addressing them. Trustworthy AI systems are: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed. Creating trustworthy AI requires balancing each of these characteristics based on the AI system’s context of use. While all characteristics are socio-technical system attributes, accountability and transparency also relate to the processes and activities internal to an AI system and its external setting. Neglecting these characteristics can increase the probability and magnitude of negative consequences.

Trustworthiness can be considered important because it attracted the attention of ISO/IEC. In (ISO, 2020), there are surveys of topics related to the so-called trustworthiness in AI systems, including the following: (1) approaches to establish trust in AI systems through transparency, explainability, controllability, etc.; (2) engineering pitfalls and typical associated threats and risks to AI systems, along with possible mitigation techniques and methods; and (3) approaches to assess and achieve availability, resiliency, reliability, accuracy, safety, security, and privacy of AI systems. In that document, trustworthiness is defined as an ability to meet stakeholders′ expectations in a verifiable way, including the characteristics of trustworthiness such as reliability, availability, resilience, security, privacy, safety, accountability, transparency, integrity, authenticity, quality, and usability.

Because of the complexity of AI as a system, and since AI includes hardware, software, data and sometimes humans, which can be treated as subystems, and if all of the subsystems must function in order for an AI system to function, then the reliability block diagram is as in Figure 1 while the reliability of an AI system can be calculated by the formula based on (Pokorni, 2021a)

$$R_{AI} = R_{HW}\, R_{SF/HW}\, R_{D/HW,SF}\, R_{H/HW,SF,D} \qquad (1)$$

where R_{HW} , R_{SF} , R_D and R_H are hardware reliability, software reliability, data reliability and human reliability, respectively, and $R_{SF/HW}$ is conditional reliability that software will function reliably if hardware is functioning reliably, $R_{D/HW,SF}$ is conditional reliability that the data subsystem will function reliably if hardware and software are functioning reliably, and $R_{H/HW,SF,D}$ is conditional reliability that the human subsystem will function reliably if hardware, software and data subsystems are functioning reliably.

If one can consider that failures of hardware, software and humans are mutually exclusive, then equation (1) can be rewritten as

$$R_{AI} = R_{HW}\, R_{SF}\, R_D\, R_H \qquad (2)$$

This is not always the case, but this formula is simpler for caculation, and can show an indication of the whole reliability, which is better than not to do any calculation.

In all cases, relaibilities are usually time dependent.

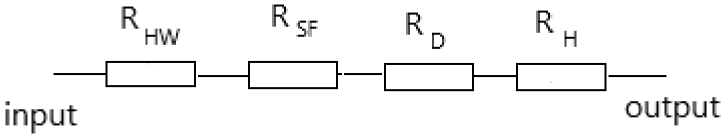

Figure 1 – Reliability block diagram of an artificial intelligence system

Let us discuss some reasons for unreliability such as the quality of data, a small or insufficient amount of training and test data, training on unreal data, the nature of learning algorithms and user incompetence.

Impact of data on reliability

As stated before, reliability has always been data-driven (if being data-driven means that all decisions and processes are based on data), and valid and relevant data has always been the main problem.

Quality of data

Quality of data can have a significant impact on the reliability of AI. What if AI learns from corrupted data or data which are corrupted on purpose (for example by an enemy)? The consequences can be different, from insignificant to catastrophic. Therefore, we need some kind of control on the data which AI uses to learn from. This problem does not imply not to use AI, or not to allow AI to learn from data, but to be aware of the problem in order to prevent serious consequences.

Some examples can be found in (Bathaee, 2018) about AI failures from IBM, Microsoft, Apple, and Amazon. The example from IBM happened in 2013, when IBM partnered with the University of Texas MD Anderson Cancer Center which developed a new “Oncology Expert Advisor” system with the goal to cure cancer (Pokorni, 2024).

Obviously, data quality influences reliability. That is the reason to speak of data reliability. Data reliability refers to the completeness and accuracy of data as a measure of how well it can be counted on to be consistent and free from errors across time and sources. The more reliable data is, the more trustworthy it becomes (IBM, 2023).

There are authors who ask questions about potential risks such as whether AI will pose an existential threat to humanity, or whether AI technology will be concentrated in the hands of the few (Bathaee, 2018). It is not only the question of reliability, though.

Amount and the nature of data

One of the important questions is whether AI can work on a small amount of data, for example, the number of failures of a product. In (Pokorni, 2024), it is concluded that model's predictive accuracy depends on the relevancy, sufficiency, and quality of the training and test data. Two questions are commonly asked with regard to failure history data: (1) How many failure events are required to train a model? And (2) how many records is considered as "enough"?

In July 2018, StatNews reviewed internal IBM documents and found that IBM's Watson was giving erroneous, downright dangerous cancer treatment advice. In (Heaven, 2019), it was concluded that a probable reason was the fact that the software had been trained on a small number of hypothetical cancer patients, rather than on real patient data.

Reliability is also important for the security of AI (for example, protection of unauthorized access).

Nature of the algorithm

There is also a question raised whether AI can fail to function as expected, and the reason is because of the nature of the machine-learning algorithms on which modern AI techniques are commonly built. These algorithms are capable of learning from massive amounts of data, and once that data is internalized, they are capable of making decisions experientially or intuitively like humans. This means that for the first time, computers are no longer merely executing detailed pre-written instructions but are capable of arriving at dynamic solutions to problems based on patterns in data that humans may not even be able to perceive. This new approach comes at a price, because many of these algorithms can be black boxes, even to their creators (Bathaee, 2018). An algorithm of AI is good if data from which this algorithm learns is good. So, again we can speak about data reliability.

User competence

It seems that user competence or incompetence in using AI can also be a reason for inadequate results which AI can deliver. An important question is whether humans can control AI.

As a conclusion for AI reliability, or trustworthiness (which is a broader term), let us use the statement from (European Commission, 2018): "Having the capability to generate tremendous benefits for individuals and society, AI also gives rise to certain risks that should be properly managed" and "It is known that humans are biased in their decision making. Since AI systems are designed by humans, it is possible that humans inject their bias into them, even in an unintended way".

Analysing AI failures can help to improve AI reliability. On the other hand, the use of AI can help to improve its reliability, and evaluate other products, and also improve maintenance (methodology and training).

AI and human reliability

As stated before, AI can learn from humans, and human knowledge can be erroneous. Since humans are involved in AI, obviously AI reliability is connected with human reliability and, consequently, human actions can influence the reliability of AI.

This author has considered human reliability important from the beginning of his work in reliability, so human reliability is included in his textbooks and lectures for students in the Military Academy and the Information Technology School in Belgrade, Serbia.

Artificial intelligence in military applications

The theory and practice of reliability arose from the problems of malfunction in military equipment in the middle of the previous century.

Artificial intelligence is also applied in the military area across different branches and types (army, navy, space forces) of the military, and the unreliability of AI can have more severe consequences than in other areas.

The problem of the reliability of AI is not only in that it will not function, but also, or more importantly, in that it can produce unwanted actions.

We cannot be sure if we can trust humans who are to decide to start nuclear weapons, but can we trust AI?

AI has been used in the military long before its today's use in civilian aplications. AI can benefit the military in numerous areas, just to mention a few such as autonomuos vehicles, use of drones, decision making, etc.

The study (Hunter et al. 2023) examined how AI technology is applied in the militaries in the US, China, and Russia and analysed the implications for the future of AI, global military competition, and international security. This study is based on previous research and expert interviews and reinforce the pivotal role that AI will play in shaping international security in the near future.

In this study, it is concluded that currently the US and China are applying AI across numerous sectors within their economies, societies, and militaries. This trend is likely to accelerate at a rapid pace in the years ahead as AI technology becomes more powerful and efficient. Russia’s AI related command-and-control developments have been split between more traditional intelligence collection formats and innovative monitoring and various facial recognition programmes for both civilian and military applications. China may be more inclined to rely on strategies calculated by AI programmes that have been trained through wargaming scenarios in potential conflicts.

In (Oluyemi, O.A. 2024), it is stated that recent developments in the field of AI have demonstrated that this emerging technology would have a deterministic and potentially transformative influence on the military power, strategic competition, and international security in general. It is argued in this research that there are various military uses of AI technologies and that national interests of powerful states are to endlessly pursue these advanced technologies as preliminary for future warfare in order to gain strategic advantages over potential emerging adversaries. It is also concluded that AI can be integrated into diverse applications, which is an improvement to the "Internet of Things" whereby different devices can be networked together for optimization of performance, and that another attribute of AI is the dual-use of many of its applications which means AI applications are useful in both military and civilian domains.

Drones are now very often used in civilian and military applications. In (Deebak & Hwang, 2023), it is stated that drone applications have gained prominence for various military observations including surveillance, medical transport, aerial photography, and medical transport.

In (Marasco & Bourlai 2025), Large Language Models (LLMs) are examined. It is stated that LLMs have the potential to enhance decision making significantly in core military operational contexts that support training, readiness, and mission execution under low-risk conditions, but still, their implementation must be approached carefully, considering the associated risks. This paper examines the integration of LLMs into military decision making, emphasizing the LLM's ability to improve intelligence analysis, enhance situational awareness, support strategic planning, predict threats, optimize logistics, and strengthen cybersecurity. This paper also considers misinterpretation, bias, misinformation, or overreliance on AI-generated suggestions, potentially leading to errors in routine but critical decision-making processes.

(Roberson et al. 2022) described a case study based on the question what it means to be responsible and responsive when developing and deploying trusted autonomous systems in defense. The lessons from this case study are: the value and impact of embedding responsible research and innovation-aligned, ethics-by-design approaches and principles throughout the development of technology at high translation readiness levels.

(Abaimov & Martellini, 2020) explored cyber vulnerabilities in autonomous technologies, highlighted critical issues of the AI use in autonomous weapons systems, incorporation of ethical principles into development of technologies, revealed legal complications and consequences of AI arms race, forecast future challenges, and argumented that generated neural networks and machine learning algorithms, being of a complex nature, still remain unpredictable, unreliable and even dangerous when fully autonomous.

(Schwarz, E. 2021) explored the (im)possibility of human control and questioned the presupposition that humans can be morally adequately or meaningfully in control over AI-supported lethal autonomous weapons systems.

Standardization and certification are also imortant. In (Jurado et al. 2024), an overview of the current state of development regarding certification and standardization efforts for Artificial Intelligence systems in military aviation is given.

From the previously mentioned, it can be concluded that nowadays military applications of AI have become a prominent topic of interest in the field of artificial intelligence, which holds a significant potential to support the military in their missions. In the past several years, the use of AI has made tremendous leaps forward in both capability and availability, such as in the field of generative AI.

As the general public has gained access to AI, this means that it can be a treat to the military, so the military must adapt to the changing threat. The military needs to keep pace with these developments in order to maintain security and a technological edge. With new ways of using AI constantly developing, it can be challenging to keep up with ways in which it can aid military operations. As AI becomes more essential, military dominance will not be defined by the size of an army, but by the performance of its algorithms, so it deserves examination of how the military currently uses AI and how it may use AI in the future.

This means that military applications of AI can give advance to small countries. For example, AI's algorithms are able to collect and process data from numerous different sources to aid in decision making, especially in high-stress situations.

Conclusion

Artificial intelligence (AI) is nowadays applied in almost every area of human life, and can have a big impact. Everything can fail to function properly or as expected, and AI is not an exception; therefore, as any other product, AI must be reliable.

Some reasons why AI is not reliable are inadequate quality of data used for learning, a small or insufficient amount of training and test data, training with unreal data, change of data, the nature of learning algorithms, and changes of user behaviour.

It must be kept in mind that AI is mainly software and uses data to learn.

In order to avoid situations when AI becomes unreliable, it is important to be familiar with causes of its unreliability.

To build successful AI, there is a need to be familiar with cases when AI failed in order not to make the same mistakes. But if AI can learn itself, it is not easy to predict mistakes it can done. Analysing AI failures can help to improve AI reliability. Designers who want to apply AI must have in mind its reliability.

If humans are biased in their decision making, this can happen also with AI since AI learns from humans.

AI is predominantly software, but can involve hardware and humans, and of course data, so the reliability of AI is related to software reliability, hardware reliability, human reliability and data reliability. Software reliability is usually a more complex problem than hardware reliability, and AI reliability is more complex than any of these including human and data reliability.

Obviously, the reliability of AI is very complex and must be considered during its design and usage.

References

Abaimov, S. & Martellini, M. 2020. *Artificial Intelligence in Autonomous Weapon Systems*. In: Martellini, M., Trapp, R. (eds) 21st Century Prometheus. Springer, Cham. https://doi.org/10.1007/978-3-030-28285-1_8

Bathaee, Y. 2018. The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*. Volume 31, Number 2 Spring 2018. [online]. Available at: https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf

Deebak, B.D. & Hwang, O. S. 2023. Intelligent drone-assisted robust lightweight multi-factor authentication for military zone surveillance in the 6G era. *Computer Networks.* Volume 225, April 2023. https://doi.org/10.1016/j.comnet.2023.109664

European Commission. 2018. *Draft Ethics Guidelines for Trustworthy AI*. Available at: https://digital-strategy.ec.europa.eu/en/library/draft-ethics-guidelines-trustworthy-ai [Accessed: 21 July 2024]

European Commission. 2019. *A definition of AI: Main capabilities and scientific disciplines.* Brussels: European Commission Independent High-Level Expert Group on Artificial Intelligence [online]. Available at: https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

Government of the Republic of Serbia. 2019. *Strategy for the Development of Artificial Intelligence in the Republic of Serbia for the period 2020-2025*. [online]. Available at: https://www.srbija.gov.rs/tekst/en/149169/strategy-for-the-development-of-artificial-intelligence-in-the-republic-of-serbia-for-the-period-2020-2025.php

Government of the Republic of Serbia. National AI Platform, 2023. Ethical guidelines for development, implementation and use of robust and accountable AI [online]. February 2023. (ai.gov.rs) Available at https://www.ai.gov.rs/tekst/en/459/ethical-guidelines.php [Accessed: 11 March 2025]

Hunter, Y. L., Albert, D. C., Henningan, C. & Rutland, J. 2023. The military application of artificial intelligence technology in the United States, China, and Russia and the implications for global security. *Defense and Security Analysis* 39(2) May 2023. https://doi.org/10.1080/14751798.2023.2210367 Available at: https://www.researchgate.net/publication/370694042_The_military_application_of_artificial_intelligence_technology_in_the_United_States_China_and_Russia_and_the_implications_for_global_security

IBM What is data reliability? Available at: https://www.ibm.com/topics/data-reliability. [Accessed: 15 March 2023]

ISO. 2020. ISO/IEC TR 24028:2020 *Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*. [online]. Available at: https://www.iso.org/standard/77608.html?browse=tc

Jurado, R. D. A., Ye, X., Plaza, O. V., Suárez Z. M., Moreno, P. F. & Valdés, A. M. R. 2024. An introduction to the current state of standardization and certification on military AI applications. *Journal of Air Transport Management* Volume 121, November 2024, https://doi.org/10.1016/j.jairtraman.2024.102685

Kececioglu-B. D. 2002. Reliability Engineering Handbook, volume 1. DEStech Publications. Lancaster. Pennsylvania. USA. ISBN No. 1-932078-00-2. Available at: https://ndesoneandik.wordpress.com/wp-content/uploads/2012/04/dimitri-kececioglu-reliability-engineering-handbook-vol-1.pdf

Marasco, E. & Bourlai, T. 2025. Enhancing trust in Large Language Models for streamlined decision-making in military operations. *Image and Vision Computing.* Available online 18 March 2025. https://doi.org/10.1016/j.imavis.2025.105489

-NIST National Institute of Standard and Tehnology. US Department of Commerce. [online]. Available at: https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF/Foundational_Information/3-sec-characteristics [Accessed: 25 December 2024]

Oluyemi, O.A. 2024. The Military Uses of Artificial Intelligence (AI) and Their Implications on International Security, *The International Journal of Research and Innovation in Social Science (IJRISS)*. June 2024. Available at: https://dx.doi.org/10.47772/IJRISS.2024.803075S

Pokorni, S. 2016. Reliability prediction of electronic equipment: problems and experience. In: *7th International Scientific Conference on Defensive Technologies*

OTEH 2016, Belgrade, pp.695-700, 6-7 October. ISBN 978-86-81123-82-9. http://www.vti.mod.gov.rs/oteh16/elementi/rad/020.html

Pokorni, S. 2021a. The Reliability of Data-driven Internet of Things Systems. *Annals of Spiru Haret University. Economic Series*, 21(4), pp.43-52 [online]. Available at: https://doi.org/10.26458/2141

Pokorni, S. 2021b. Current State of the application of Artificial Intelligence in Reliability and Maintainability. *Vojnotehnički glasnik/Military Technical Courier*, Vol. 69, Issue 3, pp. 578-593. Available at: https://doi.org/10.5937/vojtehg69-30434

Pokorni, S. 2023. Data-driven reliability and availability of electronic equipment. *Vojnotehnički glasnik/Military Technical Courier*. Vol. 71, Issue 3, pp. 769-782. Available at: https://doi.org/10.5937/vojtehg71-43474

Pokorni, S. 2024. Reliability of artificial intelligence. In: *11th International Scientific Conference on Defensive Technologies OTEH 2024*, Belgrade, 09-11 October 2024, pp. 643-646. Available at: https://doi.org/10.5937/OTEH24118

Roberson, T., Bornstein, S., Liivoja, R., Ng, S., Scholz, J., Devitt, K. 2022. A method for ethical AI in defence: A case study on developing trustworthy autonomous systems. *Journal of Responsible Technology.* Volume 11, October 2022. Available at: https://doi.org/10.1016/j.jrt.2022.100036

Schwarz, E. 2021. Autonomous Weapons Systems, Artificial Intelligence, and the Problem of Meaningful Human Control. *Philosophical Journal of Conflict and Violence.* 5(1):53-72, May 2021. Available at: https://doi.org/10.22618/TP.PJCV.20215.1.139004

-University of Cambridge. 2016. Enhancing the reliability of artificial intelligence. [online]. Available at: https://phys.org/news/2016-10-reliability-artificial-intelligence.html

Поузданост вештачке интелигенције у цивилним и војним применама

Славко J. Покорни
Школа информационих технологија, Београд, Република Србија

ОБЛАСТ: информационе технологије
КАТЕГОРИЈА (ТИП) ЧЛАНКА: прегледни рад

Сажетак:

Увод/циљ: Циљ овог рада јесте да покаже значај поузданости и расположивости вештачке интелигенције, која се данас примењује у готово свим областима живота. Као и било који производ или уређај, тако и вештачка интелигенција може да откаже. Ипак, она треба да буде поуздана. Углавном се реализује као софтвер, али укључује хардвер, податке и корисника, па је сложена за уградњу и процену поузданости.

Методе: Ово истраживање је рађено углавном коришћењем објављених чланака аутора овог рада, као и неких других радова и докумената.

Резултати: Резултат истраживања јесу кратке дефиниције поузданости, доступности, вештачке интелигенције, поузданости хардвера, софтвера, података и корисника, као и поузданости вештачке интелигенције. Поред тога, укратко су анализирани неки разлози непоузданости вештачке интелигенције, као што су квалитет података, мали или недовољан број података за обуку и тестирање, обука на нереалним подацима, природа алгоритама учења и некомпетентност корисника.

Закључак: Поузданост вештачке интелигенције је веома важна, што се мора узети у обзир приликом њеног пројектовања и употребе.

Кључне речи: вештачка интелигенција, поузданост, расположивост, хардвер, софтвер, корисник, подаци.

Paper received on: 14.01.2025.
Manuscript corrections submitted on: 25.05.2025.
Paper accepted for publishing on: 27.05.2025.