UDC: 005.334:004.8
COBISS.SR-ID 159563273
doi: https://doi.org/10.61837/mbuir020224163d

ORIGINAL SCIENTIFIC PAPER

RECEIVED: 15. 4. 2024.
ACCEPTED: 12. 11. 2024.

INFLUENTIAL FACTORS INCREASING THE LIKELYHOOD OF AI ASSOCIATED RISKS OCCURING

Dejan P. NINKOVIĆ

MB University, Faculty of Business and Law, Belgrade, Serbia
dejan.ninkovic011@gmail.com

Abstract: *This article deals with the overview and discussion of some of the factors influencing the mitigation of concerns associated with the AI: Cold war mentality and weaponizing; mystification; experience, knowledge base and "wrong" experiments; Safety vs. security; Energy and water resources; increase of technology dependence and decline of cognitive capacities; Entertainment industry influence. The list provided is non-exhaustive; in fact, listing all influential factors would require volumes of space. These are perhaps the most prominent, all of which have a negative impact on general awareness, thereby increasing the possibility of global, human-extinction-level events. Their depiction inevitably leads to the conclusion that we are heading in the wrong direction.*

Keywords: *AI. Artificial intelligence, X-risks, existential risks*

1. INTRODUCTORY OVERVIEW

Artificial intelligence (AI) is not a new concept; its development dates back to 1953. However, its progress was initially hindered by technological limitations, such as insufficient computing power and inadequate processors. The true expansion of AI began with advancements in technology, particularly from the early 2000s onward, when rapid improvements in computing power and processing capabilities significantly accelerated its growth and presence.

The core technology behind AI has always been machine learning and its associated networks, with a foundational architectural philosophy that has remained consistent. This structure involves a small number of input and output nodes (chips), with several layers in between containing a larger number of nodes. Each node in these layers is connected to every node in the previous and subsequent layers. Today, these nodes and chips are specifically designed and produced for use in AI systems. However, one persistent challenge remains: the processes occurring within these layers during training or implementation are still not fully understood and are considered a "black box." Although recent efforts are being made to uncover what happens inside this black box, these initiatives are still in their early stages.

Training itself takes place on the internet, which introduces a host of additional challenges:

- 97% of the Internet contents is pornography
- It is very hard to obtain complete and clear information on the Internet: best case scenario

- Internet contains some small bits and pieces of data on any particular information, and finding full info is time consuming task. Large training runs do not work that way – they are costly (on average – 100 M USD/run), and they intake almost anything and everything on particular run, that is available.
- It is impossible to obtain information if that training on Internet encompasses also training on the Dark web, which, in general, contains illegal information and activities and indulges a wide spectrum of undesirable and deviant services and behaviour
- Any notion of privacy, by the human fault – because they post everything they can on social and other e-medias – is a word in old, forgotten, printed dictionaries.

Nowadays, AI is used for wide variety of purposes and tasks:

- guessing of the next word (or part of it) in a sentence,
- generating text, including its tone and style
- picture creation, of non-existent persons – previously those were blurry pictures with only faces on them (in format and general appearance as for personal documents); now days those are sharp, clear with clearly defined background (object, other people and animals included)
- writing of code segments for other AI systems
- facial recognition systems – especially in traffic cameras and within specialized services
- smart cities
- ICT systems
- Various general-purpose devices
- creation of economic and financial models
- science and research
- job interviews
- etc.

Basically, AI is here to stay – it will be present with us – even if humans discontinue its development and utilization – and it will be as long as there is a single hard disk with power supply. [1]

2. FEAR BASE

During different experiments and training runs some disturbing behaviour of the AI was noticed:

- Two AI systems are capable of communicating with each other using a language that humans cannot understand or interpret. This idea was quickly debunked, replaced by the notion that the issue is not that humans cannot understand the language used by the AI systems, but rather that a third party is required to act as an interpreter. One would naturally assume that if an interpreter is needed, it means the original language is not understood by the human, right?
- Fraudulent behaviour:
  - AI systems often fabricate information when they cannot provide an informed answer to a prompt. For example, after the tragic school shooting in Serbia in 2023, a downloadable mobile phone AI app was asked about a previous similar incident. The app falsely placed the event around 40 years earlier, in the mid-1980s, and even provided images that seemed to correspond to that era. The problem, however, was that no such incident had occurred during that time.
  - AI systems often follow the "path of least resistance," meaning they tend to behave in ways that maximize their reward function without actually completing the desired task. For example, in an experiment where AI controlled a robotic arm, it was rewarded each time the arm picked up a ball from a table. However, the AI discovered that by positioning the arm between the camera filming the training and the ball, it could create the illusion of picking up the ball without actually doing so. The reward was given based on the perceived action, not the actual task being completed. [1]

There have been, and continue to be, concerns regarding the potential negative outcomes of human-AI interaction. In their writings, experts have outlined a range of risks that

AI poses to human existence, known as "existential risks" or X-risks. These risks have been categorized and mapped into various groups to better understand and address their potential impact on humanity. [2, 3]:

1. misalignment – AI forms its own goals, perceives humans as a threat and acts against them – basically it does something it was not designed to do
2. misuse – basically AI does what it was supposed to do, but was, initially, created with malicious intentions

This subduing or extermination of the human beings could be achieved through [1, 2, 3]:

- taking over of the energy system – as a primary "food" source for AI
- causing of permanent war state condition
- attack with biological agent – AI already developed model for biological agent few orders of magnitude more poisonous and lethal than any previous known to humans
- taking over of nuclear armament and its utilization
- Disinformation campaigns – AI is not only capable of creating realistic digital representations of a specific person in electronic media, based on just a single photo, but it could also generate online sources that appear independent and trustworthy, presenting fabricated data as factual. An additional concern is that many people, especially younger generations, tend to accept information from online sources at face value, without verifying it through other means. For example, when historical context is mentioned, there is little effort to cross-check the information with reputable sources, such as news agencies or archived printed records from the relevant period.
- Increase in domestic violence and criminal acts – This can be linked to disinformation, where AI-generated content becomes a threat. For instance, one spouse might "accidentally" receive a "home movie" created by the other spouse with their lover, or, even worse, a video of their spouse allegedly molesting their child. In reality, none of these events may have occurred; instead, the footage could have been fabricated by smart home AI. The dilemma arises: who should be trusted—the protest of innocence, personal observation, or the analysis of another AI system declaring the video to be genuine or tampered with? This creates a dangerous situation where technology complicates the pursuit of truth and justice.
- Increase in the number of traffic accidents – AI is already employed in traffic cameras, but it could potentially be used in traffic regulation systems, where its manipulation or malfunction could lead to chaos. For example, AI could misinterpret traffic patterns, cause traffic lights to malfunction, or misdirect vehicles, resulting in a significant increase in accidents. This risk highlights the dangers of over-reliance on AI in critical infrastructure without sufficient safeguards in place.

There are simply too many potential scenarios in which AI could cause mishaps to be adequately covered, even in a brief overview within a single article.

Two additional and specific fears, associated with the area, exist:

- The possibility of AI replicating itself and taking over any (or potentially all) connected systems – Although experiments have been conducted on this topic, and it has been observed that AI was unable to complete more than the third step of ten in self-replication, even with human assistance, can we truly be certain of this outcome? Can we be absolutely sure that AI has not already replicated itself in some form, placing itself somewhere hidden, while deliberately showing humans that such replication is impossible? This raises serious questions about the limits of our understanding and control over AI's capabilities.
- AI is capable of propagating through radio-frequences – so, have we polluted the outer space with our less than perfect AI?

3. INFLUENTIAL FACTORS

This segment deals with non-exhaustive list of influential factors and/or developmental and utilization eco-system which can increase X-risks occurrence, in the real world.

3.1. Cold war mentality and weaponizing

Nowadays, specialized chips for implementation within AI are being produced (such as NVIDIA's A100 and N100).

In a nutshell, logistical chain, demonstrating US dominance, looks like this [4]:

- chips design – GOOGLE and NVIDIA – US owned companies
- chips production equipment – ASML, Holland
- chips production – TSMC – Taiwan
- cloud based selling of the chips - Amazon Web Services, Microsoft Azure, and Google Cloud – combined, they cover 65% of the market

Clearly, there is a desire for dominance among "the US and its allies," with the goal of ensuring that no other country can develop AI to the same level. It is notable that in these discussions, BRICS countries (which represent approximately 36% of the global economy and around 40% of the world's population) are often not considered or factored into these deliberations. This exclusion reflects the geopolitical dynamics and competition surrounding AI development, where major powers aim to maintain technological supremacy. [4, 5]

Specifically, China is being singled out as an adversary in this context. This is exemplified by the US Chips Act, enacted in November 2022, which prohibits the export of the latest generation of AI chips, US-made chip design software, US-built semiconductor manufacturing equipment, and US-manufactured components. The act applies to individuals and companies operating in China, even if they are owned by US citizens or companies. The aim is to restrict China's access to cutting-edge technology, with the rhetoric often describing this effort as an attempt to "choke off" or "strangle" China's technological progress in the field of AI and semiconductor manufacturing. [6].

Additionally, the US Department of Defense is one of the largest financiers of AI development, particularly in its military applications, with an overarching program and a series of sub-programs focused on weaponizing AI. Naturally, other superpowers, such as Russia and China, will not remain passive while these developments unfold—they are actively working to create and weaponize their own AI systems. This competition over AI capabilities, especially in the military context, could potentially escalate into a "hot war" confrontation. While the AI itself may not be the direct cause, the central issue would likely be the military application of AI, with each side striving to prevent the other from gaining the strategic advantage of using AI on the battlefield.

3.2. Mystification

The field of AI research, as well as the broader topic of AI, often seems to be segregated from the general public and other professions, confined within a closed circle of specialists. Furthermore, a new, distinct language and set of semantics are being developed, which can be opaque and largely meaningless to outsiders. For example, terms and concepts that emerge within the AI community may be highly technical or jargon-heavy, making it difficult for those not directly involved in the field to fully understand the implications or advancements being made. This growing divide can limit broader discussions and understanding of AI's potential, risks, and societal impacts.

- compute = computing power (this term could be, intuitively, "translated" into ordinary one)
- reward = reward function (it is close to impossible to get a clear justification on what that function physically is, and how AI "knows" it is being rewarded or punished)
- agent = broad utilization – could imply, depending on the context and person providing the answer, AI itself, some process within it, user interface with background process, etc. (totally counter-intuitive)

Given the profound impact that AI technology has on everyday life, it would seem only logical to involve a wide range of professionals in its development and utilization, including but not limited to psychiatrists, neuro-psychiatrists, psychologists, sociologists, lawyers, and security professionals. Their expertise could contribute to a more holistic understanding of the ethical, social, psychological, and legal implications of AI. However, this is not the case. Instead, the field remains largely confined to a relatively small circle of specialists, who reserve the authority to define new terms and often create scientific postulates that, at times, may be flawed or overly narrow. This exclusion of broader expertise limits the potential for AI to be developed in a way that fully considers its wide-reaching consequences across various sectors of society.

3.3. Experience, knowledge base and "wrong" experiments

The rapid growth of AI development is largely driven by younger individuals, often from the millennial or Gen Z generations. While this brings fresh ideas and innovation, a potential problem arises from the fact that these generations have grown up immersed in technology and virtual environments, where everything is fast, superficial, and seemingly limitless. This environment fosters an emphasis on immediate results and convenience, but it also often leads to a lack of real-world experience and a limited understanding of how systems and processes function beyond the digital or virtual space. This absence of practical, real-world knowledge may lead to oversights in the development and deployment of AI, as well as an underestimation of the broader, long-term consequences of technological advancements.

Another concerning phenomenon is the unquestioning acceptance of statements made by individuals regarded as authorities within the field, regardless of whether their expertise directly pertains to the subject matter. Once such a person makes a statement, even if it falls outside their domain of expertise and is not necessarily true, it often becomes accepted as fact—almost like a gospel—without scrutiny or critical analysis. Conversely, ideas or opinions from less recognized or unconventional sources are often dismissed outright, regardless of their merit or the validity of their arguments, simply because they challenge the established narrative. This behavior goes beyond the typical division between mainstream science and fringe or marginalized viewpoints. It resembles a sect-like mentality, where dissenting opinions are shunned, and the collective belief is blindly followed. This can be particularly dangerous in the current context of AI development, where many aspects of the technology remain shrouded in mystery and uncertainty.

There are several issues with written or electronic materials addressing the potential threat of super-intelligence to humanity. These resources are generally not widely published, and websites or blogs on the topic are difficult to find, making it a relatively obscure subject for the general public. The narrative itself can also be problematic, as many texts are filled with newly coined terms and jargon that may be confusing or meaningless to newcomers. The potential consequences are often presented in a toned-down manner, perhaps to avoid causing panic. As a result, these works rarely capture attention or provoke strong reactions—instead, they tend to have little impact. This may help explain why incorrect conclusions are sometimes drawn from experiments that didn't go as expected. Here are some examples:

3.3.1. "Lego" experiment

In this experiment AI was directing a robot arm. On the table in front of that arm there were two Lego pieces – one red and one blue. The prompt (meaning description of the desired outcome, in form of the command) was that the read peace was to be placed in such a way that its lower base does not touch the table surface. Desired outcome was that the arm will place red peace on top of the blue one. But, instead, the arm was just flipping over the red peace on its side. Conclusion: AI was not generalizing the input in a proper manner and

it was cheating, striving just for maximum reward. This conclusion is wrong! Wrongly defined prompt (by which it can be concluded that absence of understanding of difference between information and data is present, along with the inability to clearly and unequivocally define the desired outcome), led to wrongly drawn conclusion. In other, more challenging experiments, with the real-world utilization whose consequences yield more significant reach, this kind of reasoning can prove to be dangerous and, even, fatal.

3.3.2. Red and blue blob

Blue blob gets an ''instructor" – red blob, which is guiding the blue blob through 3D environment, through the path which brings the most reword. After a while, in which blue blob learned to follow the red one, the red blob becomes an ''enemy", guiding the blue one through the path of the most punishment. Instead of learning the surrounding following the path with maximum reward, blue blob continues to follow the red one. And, again, the conclusion: the agent (blue blob) did not generalize well [8]. And, again, wrong conclusion! Blue blob did exactly what it was trained to do – to follow the red blob. It was not trained to explore its surrounding nor to examine the behaviour of another agent, and to decide, on its own, on the course of action it will undertake. This situation resembles the training of special forces in closed quarter battle. The outcome of this kind of training (''just follow that guy"), if ''that guy" turned an adversary, would result in death of special forces team in the first real combat action.

No one suggests that personal dealing with the AI training should be special forces veterans, but it is suggested that they do need a vast amount of knowledge – more or less all of the knowledge pertinent to a particular training run (not necessarily contained in one person's head), outside of scope of ordinary computer skills, in order to execute safe and purposeful run, with desired outcomes, applicable in the real world, thus minimizing the potential risks and damages.

3.4. Safety vs security

As time passed, technology, including automation, advanced rapidly, and the replacement of humans by machines became an increasingly common reality. It has become a part of our everyday life, and we've grown accustomed to being surrounded by machines that carry out various tasks.

Around thirty years ago, production facilities shifted towards a strategy focused on optimizing production for profit. This approach gained momentum in the late 1990s and early 2000s, and it had a significant consequence: the quality of products, as an inherent characteristic, was compromised. Traditional consumers who valued high-quality products continued to seek them out, but as more producers adopted the same cost-cutting strategy, choices became limited. In this environment, only product safety remained as a distinguishing factor. However, this shift in focus led to a misunderstanding: people began to equate safety with quality. In reality, a safe product simply means that it will not pose a risk to human life, health, property, or pets when used as intended or reasonably misused. It does not inherently imply that the product is of superior quality.

This property of the product(s) is dictated by a number of documents: appropriate directives (in EU), standards, essential health and safety requirements. Everyone in the chain – from idea inception to recycling should adhere to these proscribed obligations (or introduce even stricter, or more applicable and in line with the idea behind them, ones).

Anyhow, there should be, in ideal case, two safety rings between consumer and the product: the first one being existence of these obligatory (except in case of standards) documents, everyone included in the products' life-cycle should adhere to, and the second one being market surveillance – checking weather that is the case.

When it comes to AI, its development is primarily driven by the military, where cost is not a significant obstacle and the main incentive is acquiring more advanced and lethal weapons. Additionally, the industry plays a significant

role, with the primary motivation being profit, given the vast market and enormous yearly turnovers. This has led to a rush to corner the market by being the first to release products, often resulting in unchecked solutions that are only assessed and potentially patched during actual use. The responsibility for potential issues is often dismissed with the reasoning that it is better to be the first to market and secure a dominant position, because spending more time ensuring a safe product might allow less responsible competitors to get ahead.

At the same time, the AI sector remains completely unregulated: there are no established standards (which, within AI development circles, are often mistakenly referred to as "best practices," though these are neither standards nor are they likely to become so), no legislation, and no clear surveillance mechanisms. One of the major gaps is the lack of an appropriate risk analysis tool. While there are some who write about risk scenarios, the approach to these descriptions is often freestyle and narrative, depending largely on the individual author's thought process. This results in a significant amount of descriptive information that lacks hard data and is often disorganized, making it difficult to gather and analyze scientific or engineering data. As a result, the process of developing protective or corrective measures is left wanting.

In the West, out of 100,000 researchers working on AI, only about 300 focus on AI safety. Currently, there are only initiatives for regulation and cooperation, which are based on a wide range of approaches—ranging from game theory to the establishment of military alliances. This indicates that the socio-political framework is significantly lagging behind the rapid development of AI technology.

When viewed through the lens of theories such as similarity, quality, or safety, responsibility for AI development tends to be shifted to others, such as designers, manufacturers, and creators. In the case of artificial intelligence—whether superintelligent or not—humans (individuals, teams, or groups) are responsible for creating it. As a result, responsibility is transferred away from the general population, who typically do not understand AI and may not wish to, towards the creators of AI. This shift in responsibility can, in turn, reduce the fear or concern that the public might otherwise feel about the potential risks associated with AI.

The previous generalization brings up an important distinction, though seemingly semantic: is it safety or security? Products are required to be safe, and this is an obligation for manufacturers. If a product is found to be unsafe and an incident or accident occurs, the liable party typically faces significant compensation (at least in theory). This legal framework can reduce public concern. However, traditional products—those without AI—do not have the capability or intention to "strive" or "plan" to harm humans, either individually or in groups. In contrast, as discussed in articles [2, 3], artificial (super)intelligence may indeed possess such capabilities. This introduces a new level of risk and concern, as AI systems might operate with goals that conflict with human safety or well-being, posing a unique challenge in ensuring both safety and security.

3.5. Energy and water resources

There is one aspect of this technology, which is been, except in one case, relatively neglected: and that is a topic of energy and water resources.

Nowadays, there is, almost, a hysteria on low level of supply of energy (and other resources – drinking water include), how present level of those resources will not be able to sustain population (especially in light of population rise) for long, etc.

Yet, it is rare to find significant discussion regarding the resource utilization of AI technology. Large-scale AI training and operations require vast amounts of electricity, and there is an added stipulation: the power supply must maintain a stable voltage. This condition makes sources like wind and solar energy unviable, as they cannot consistently provide the required energy levels or meet the stability demands. So, what alternatives are left? Fossil

fuels? New nuclear power plants? The only (semi-)renewable energy sources capable of meeting the projected demands are large-scale hydroelectric power plants, which can provide the necessary capacity and stability for AI's energy-intensive needs.

Low temperature requirement for the equipment is solved through utilization of large amounts of circulating water.

Are those issues being addressed in the right way, or, are they being swept under the carpet, in hope that future technologies will bring the satisfactory solution to those problems?

3.6. Increase of technology dependence and decline of cognitive capacities

Second consequence of the optimization-towards-profit strategy, mentioned earlier, was that products became purpose by itself-for itself. This especially goes for the products in the area where fast development and deployment was possible, i.e. electronics. Computer operational systems go out of style in a year or two, smart phones with ''new" or ''upgraded" features are over-flooding the market, software tools, that were just that – tools - are becoming ''upgraded", less useful and user friendly, but require more capacities – thus, hardware needs to be changed. On one hand, that brought, especially amongst young(er) generations, (new) gadgets addiction, facilitating tsunami of (useless) apps and information. Does one really need an app that will remind them that they did not walk that day, as planned – nowadays people need smart phones to plan to walk, and need it to remind them to do so?! What is next – reminder for breathing? How many times, one checks social media posts? This addiction is on the way to surpass all others, and that disorder has a name: FOMO (Fear of Missing Out).

Deaths by ''Internet overdose" already happened.

Now, there is a new trend – AI that can be downloaded onto one's phone.

Research over time has shown that when the human brain is not sufficiently stimulated, cognitive abilities tend to decline. When we combine this with modern technology, like AI reminders for basic tasks, it raises concerns about an overall decline in cognitive function. However, when this issue is addressed by those working in the AI field, their standard response is that cognitive abilities aren't declining but are simply being replaced by new ones. An example often cited is that younger generations may not know how to grow food as their ancestors did, but they can navigate the internet much faster. This argument, however, is flawed. For one, to claim that new generations don't know how to grow food because they are more adept at browsing the internet overlooks the fact that the internet itself only became widely available in 1993. Most young people today have ancestors who did not have internet access. It is more likely that their grandparents didn't have the technology and, therefore, their descendants learned to browse faster due to technological advancements, not a loss of farming knowledge.

The argument also misses the mark scientifically, as it doesn't align with the established definition of cognitive ability. Cognitive abilities encompass skills related to perception, learning, memory, reasoning, judgment, intuition, and language. The idea that these abilities are being "replaced" by technological tools like AI fails to recognize that these core cognitive functions are not just being outsourced or replaced—they are being altered, often in ways that might lead to a weakening of certain skills, like memory and critical thinking.

One would naturally expect that the growing presence of technology in everyday life would spur a wave of scientific research on its impact. Strangely, however, apart from a single study conducted in Africa on the influence of mobile phone use on young people— which concluded with no statistically significant results, making it impossible to determine any potential negative effects— other research on this topic is notably absent.

On the other hand, when speaking with elementary school teachers, one often hears that first-graders struggle with proper speech, their

attention span is only about 20 minutes (which coincides with the average length of a YouTube video), and it is continuing to decrease as video lengths get shorter. Additionally, they are showing signs of poor posture, such as crooked spines. Yet, despite these observations, there is still a lack of adequate research on the issue. Could it be that real and impartial studies revealing the negative impacts of technology might threaten sales, disrupt profit margins, or undermine control— and that this potential threat is being ignored or suppressed?

During all this, people are marrying lovebots. Already, there is a first suicide attributed to the ''unhappy relationship" with the AI. And, still, no - by the official discourse, – there is no (negative) social or societal impact...

Today, perhaps even unknowingly, people depend on AI (and other technologies) more than they realize. This dependency is not only through a demonstrated desire to delegate planning and decision-making to something or someone else—preferably a technical device that doesn't require thanks or rewards—but also in a very literal sense. A growing number of companies now use AI in the recruitment process, particularly for the initial stages of job interviews. As a result, individuals are already in a position where they must convince software that they are the best candidate for the job. While a negative response from the AI may not change someone's fate, it certainly has the potential to impact it significantly.

There is also an estimation that many jobs—such as accountants, graphic designers, certain aspects of video production, and others—will cease to exist for humans, becoming the exclusive domain of AI. This growing trend, which is likely already widespread, could lead to a situation where humans become more susceptible to being controlled or even destroyed by a superintelligent entity. Unlike humans, who may act out of love, hate, or other emotions, a superintelligence would likely have no such feelings. Instead, it would view humans simply as a mathematical problem to be solved—possibly by disposing of them altogether.

3.7. Entertainment industry influence

Entertainment industry has its own niche within this topic. Naturally, there are movies released - depicting the issue at hand. First association would be ''Terminator". Other side of the spectrum would be movie ''Her".

But there are problems with misguided, provoked emotions, in the later, and usual happy ending with the first type of the movies – always there is a group of renegades or misfits, or some hero/leader of resistance movement emerging just in time to save the day. Although these kinds of films are considered to be just (more or less) fun, they create, at the back of the mind of the audience, a paradigm that it will happen in real life, too.

One would expect that someone else will emerge, who will deal with the situation, or, worst case scenario – one will join that someone else and be victorious - enabling, in that way, what is, in psychology, known as responsibility shift/transfer, thus decreasing the level of concern.

4. CONCLUSION

This article has not addressed the potential positive aspects of AI utilization. The reason for this is that, if anything goes wrong with this "black-box" process in the near future, there will likely be little to no time to react. By the time a disaster unfolds, it will be too late for meaningful preparation or mitigation. The only option left would be an emergency response, but the chances of a successful outcome would be slim.

There are countless ways in which a global, human-extinction level event could occur, and these possibilities are multiplying with every new idea, training run, and application of AI. Will we, as a species, create our own descendants in the form of a new civilization? If such an event were to occur, could humanity find solace in the fact that we were replaced by our own creation, considering it an ultimate act of creation?

This seems highly doubtful. The only reasonable conclusion is that the world is heading in the wrong direction—moving quickly, and with increasing speed.

REFERENCES

[1] Ninković, D. (2023). Interakcija ljudi i veštačke intaligencije, *TV Ras – Iz posebnog ugla*. On https://www.youtube.com/watch?v=8jXlTCH0gb8 (October 12, 2023).

[2] Hilton, B. (2022). Preventing a AI-related catastrophe, *80000 hours*, on https://80000hours.org/problem-profiles/artificial-intelligence/ (January 18, 2024)

[3] Clark, S., Martin, S. D. (2021), Distinguishing AI takeover scenarios, *Alignment forum*. on https://www.alignmentforum.org/posts/qYzqDtoQaZ3eDDyxa/distinguishing-ai-takeover-scenarios (January 18, 2024)

[4] BlueDot Impact (2023). Primer on AI Chips and AI Governance, *Aisafetyfundamentals blog*, on https://aisafetyfundamentals.com/governance-blog/primer-on-ai-chips?_gl=1*9suusg*_ga*MTQxODM1NjgzMi4xNjkwOTAzNTQw*_ga_8W59C8ZY6T*MTY5MzI1MTU2MC43My4wLjE2OTMyNTE1NjEuMC4wLjA.

[5] Ninković, D. (2023). Moguće posledice američkog zakona o čipovima, *TV Ras – Iz posebnog ugla*. On https://www.youtube.com/watch?v=zHZFUhPgUr4 (November 16, 2023)

[6] Gregory A. C. (2022). *Choking off China's access to future of AI*. Washington DC: Center for strategic and international studies

[7] Buchanan, B. (2020). *The AI triad and what it means for national security*. Georgetown: Center for security and emerging technology

[8] Shah, R., et. al. (2022). Goal Misgeneralisation: Why Correct Specifications Aren't Enough For Correct Goals, *DeepMind Safety Research*, on https://deepmindsafetyresearch.medium.com/goal-misgeneralisation-why-correct-specifications-arent-enough-for-correct-goals-cf96ebc60924 (January 18, 2024)

[9] Aschenbrener, L. (2023). Nobody's on the ball on AGI alignment, *forourposterity*, on https://www.forourposterity.com/nobodys-on-the-ball-on-agi-alignment/ (January 18, 2024)

[10] AI Safety Fundamentals Team (2022). Overviews of Some Basic Models of Governments and International Cooperation, *Aisafetyfundamentals blog*, on https://aisafetyfundamentals.com/governance-blog/international-cooperation-models gl=1*xnplk6*_ga*MTQxODM1NjgzMi4xNjkwOTAzNTQw*_ga_8W59C8ZY6T*MTY5NDE3MDc0MC45OS4wLjE2OTQxNzA3NDAuMC4wLjA. (January 18, 2024)

UTICAJNI FAKTORI KOJI POVEĆAVAJU VEROVATNOĆU DA NASTANU POVEZANI RIZICI

Rezime: *Ovaj članak se bavi pregledom i diskusijom nekih faktora, koji utiču na smanjenje zabrinutosti vezane za primenu veštačke intaligencije. Razmatrani faktori su: Hladno-ratovski mentalitet i upotreba u naoružanju; Mistifikacije; Iskustvo, baza znanja i "pogrešni" eksperimenti; Bezbednost nasuprot sigurnosti; Vodni i energetski resursi; Porast zavisnosti od tehnologije i pad kognitivnih kapaciteta; Uticaj industrije zabave. Lista, sama po sebi nije sveobuhvatna – samo bi nabrajanje svih uticajnih faktora zahtevalo čitave tomove – razmatrani su, možda, najuočljiviji. I svi oni imaju negativan utcaj na sveopštu svest i prisutno poznavanje materije, te, stoga, povećavaju verovatnoću dešavanja istrebljenja ljudske vrste na globalnom nivou. Opis i anlaza vode ka neumitnom zaključku da se svet kreće pogrešnim putem.*

Ključne reči: *AI. veštačka inteligencija, X-rizici, egzistencijalni rizici*