CEUR-WS.org/Vol-3887/paper17.pdf

Expert Analysis of Potential Threats From The Intensive Development of Artificial Intelligence

Vitaliy Tsyganok[1,2,3], Viktor Holota[4], Oleksandr Hryhorenko[4] and Serhiy Burukin[5]

[1] *Institute for Information Recording of National Academy of Sciences of Ukraine, M. Shpaka str. 2, Kyiv, 03113, Ukraine*
[2] *Taras Shevchenko National University of Kyiv, Volodymyrs'ka str. 64/13, Kyiv, 01601, Ukraine*
[3] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteysky ave. 37, Kyiv, 03056, Ukraine*
[4] *Yevheniy Berezniak Military Academy, Kyiv, Ukraine*
[5] *Customertimes Corp. in New York, USA*

Abstract

Recently, the international environment has been actively seeking consensus on the need to take collective coordinated measures to identify, monitor, measure and minimize risks from the development of artificial intelligence (AI) at both the global and national levels. This study is based on the analysis of potential threats to the intensive development of AI, which was conducted by a team of experts using the Consensus2 software system for distributed collection of expert information. As a result of the group expert assessment, a list of the most influential threats was compiled and their relative importance was determined.

Keywords

artificial intelligence tools, cybersecurity, expert threat assessment, collective expertise

1. Introduction

With the recent rapid development and implementation of artificial intelligence (AI) systems, such as generative linguistic models, the task of complying with the cybersecurity of such systems has become urgent. The growing relevance and interest in security tasks in the field of AI, which have certain features compared to traditional cybersecurity tasks of information protection, is evidenced by a number of publications in this area [1-3]. These, and a number of other works, cover research related to the security of AI systems, their stability, etc.

Recently, the international community has been actively seeking consensus on the need to take collective, coordinated measures to identify, monitor, measure, and minimize risks from AI development at both the global and national levels.

In October 2023, the UN Artificial Intelligence Advisory Body was established to develop general recommendations for assessing risks, opportunities, and mechanisms for international governance of AI technologies. These recommendations should be made public on the eve of the Future Summit (September 2024) and form the basis of the Global Digital Compact [4].

In November 2023, the first-ever AI Security Summit was held in Blechley Park (UK). The event was attended by high-ranking officials from the governments of the UK, the USA, China, Japan, France, Germany, Canada, Italy, Spain, India, Israel, South Korea, Singapore, Switzerland, Ukraine, Turkey, the United Arab Emirates, Saudi Arabia, Kenya, Rwanda, and Nigeria, as well as heads of leading AI service companies, including OpenAI, Google, Meta, Microsoft, and others [4].

The analysis of the final communiqué and the comments of the summit participants shows that the international community is really concerned about the potential threat of AI getting out of human control. Other risks recognized include: the formation of biased attitudes of AI towards certain things; abuse of AI capabilities in the areas of cybersecurity, biotechnology, and disinformation.

ITS-2023: Information Technologies and Security, November 30, 2023, Kyiv, Ukraine

vitaliy.tsyganok@gmail.com (V. Tsyganok); v.holota@ukr.net (V. Holota); gregorenko@ukr.net (O. Hryhorenko); s.burukin@i.ua (S. Burukin)

0000-0002-0821-4877 (V. Tsyganok); 0000-0003-3969-7776 (V. Holota); 0000-0003-0633-7563 (O. Hryhorenko); 0009-0002-1762-2546 (S. Burukin)

© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The participants of the event signed a final communiqué, which states that joint efforts should be focused on the following key areas:

- identification of common risks to AI security, their scientific understanding and justification based on the collected evidence base;
- development of appropriate policies by the signatory countries to reduce these risks, recognizing the possible differences in approaches and the need to involve development companies in solving this problem. The main issues are the development of risk assessment methods and tools for testing the safety of AI technologies [5].

The Deputy Head of the Ministry of Digital Transformation of Ukraine H. Dubinsky took part in the summit on behalf of our country [6].

It is noteworthy that the summit participants, primarily the heads of development companies, agreed on the need to delegate to governments the authority to control the safe development of AI. One of the main problems that, according to the summit participants, needs to be addressed immediately is the development of standards for identifying, monitoring, and measuring AI-related risks.

During the summit, British Prime Minister R. Sunak announced the creation of the AI Security Institute, which will test and authorize the use of new AI services before they enter the market. A similar institution is being created in the United States. According to R. Sunak, work on new breakthrough AI technologies is expected to be completed next year. That is why two international AI security summits are scheduled for 2024 - in the Republic of Korea and France [7].

2. Research

2.1. Identifying threats in AI development

The first stage of this process should be a comprehensive understanding of the problem of potential threats that may arise as a result of the rapid development of artificial intelligence. In order to identify and formulate a list of possible threats caused by the development of AI, a collective examination was conducted with the involvement of nine experts from the Institute of Information Registration Problems of the National Academy of Sciences of Ukraine, the Ministry of Defense of Ukraine, the Yevheniy Berezniak Military Academy, and one of the leaders of the domestic market, Infozahyst.

2.2. Collaborative expert analysis using software tools

The examination was carried out remotely using the System of Distributed Collection and Processing of Expert Information for Decision Support Systems "Consensus- 2" [8]. This study focused more on potential threats from uncontrolled human use of AI tools, which are currently developing rapidly, and less on threats from the consequences of AI's possible release from human control and independent decision-making.

In accordance with the group decomposition technology implemented in the Consensus-2 system, the collective examination included several stages. These stages essentially combined the process of decomposition (separation of components) of the problem "The threats posed by the development of AI tools" with the subsequent assessment of the relative importance of these threats.

At the initial stage, after the formation of the expert group, each expert, operating remotely through the system's web interface, had to personally compile a list of threats that, in his or her opinion, are the most influential in the problem situation under consideration. It is important for the experts involved in the examination to understand what is meant by the term "most influential threats." Within the framework of the subject area modeling concept, the influential factors that are introduced as components of the model are those that have a relative impact of at least 10% of the total value of all influences on a particular current situation. In this case, the total number of the most influential threats will not exceed 10 (or 7±2 according to classical recommendations [9]), which makes it possible to obtain a sufficiently adequate model of the problem situation and confidently operate with a set of threats in the expert assessment of their impact.

Based on the results of this stage, the experts formulated a total of 46 major threats, according to their authors, arising from the rapid development of AI (5-7 threats were independently identified by each expert).

2.3. Prioritizing and rating major AI threats

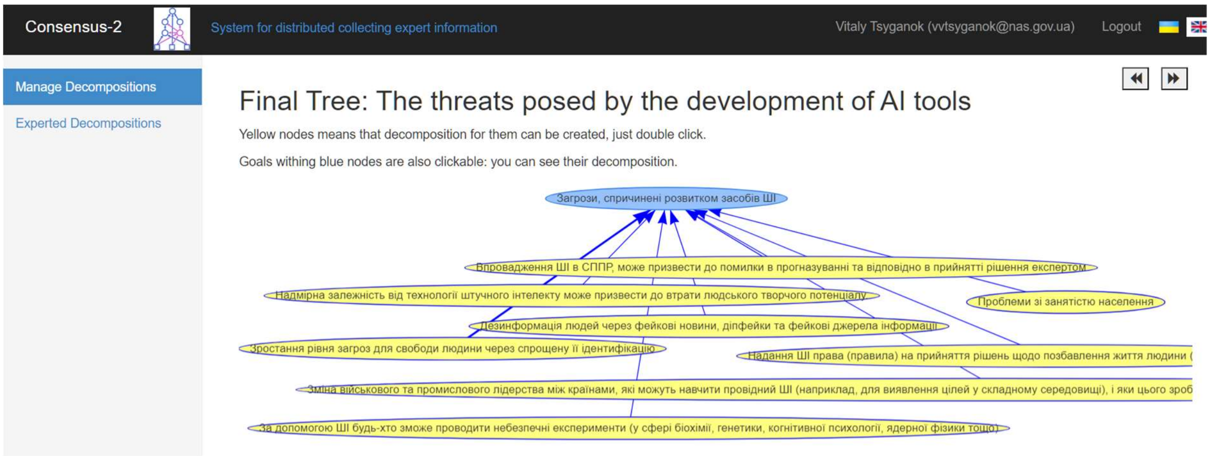
The next step was to group wording that is identical in content, since during autonomous work there is a high probability that different experts will formulate the same threat in different ways. It should be noted that generative artificial intelligence tools [10-13] can and have been used to combine formulations into groups of identical content under the control of an expert organizer (knowledge engineer). Thus, the grouping of formulations of the same content was carried out in an automated mode using AI tools. For the effective use of these AI tools, it is advisable to indicate in the prompter the desired number of groups of identical (similar) formulations no more than 15, so that in the future, after a group assessment with the rejection of insufficiently important threats, 7±2 would remain, which are the most important from the point of view of not an individual expert, but the expert group as a whole.

In order to integrate AI tools into the program code of an expert information gathering and processing system such as Consensus-2, it is advisable to use an application programming interface (API) that allows you to control a number of AI tool parameters that are not available through the web interface. For example, when it comes to the options available through the ChatGPT API versus the web interface, there are several key points worth noting:

1. **Tone or style of response**: Through the API, you can customize the style or tone of a response in more detail. For example, you can ask Chat to work in a certain format (e.g., business or more informal) or set certain text restrictions.
2. **Saving and managing context**: The API gives you more control over context preservation. For example, in the web interface, saving the context of a conversation is automatic, whereas through the API, you can define which context to save or delete.
3. **Customized models or model variants**: Through the API, you can select specific GPT models that are only available through the API, or define specific parameters that are not supported in the web interface.
4. **Context length**: Through the API, you can manage the parameters related to the limit on the number of tokens in the request and response, which is not always available through the web interface.
5. **Additional formatting or response structure options**: Through the API, you can have more control over the formatting of the output, such as the structure of responses in JSON or other formats.
6. **Real-time queries**: Using the API, you can integrate ChatGPT into more complex applications to perform real-time queries, with dynamic data processing, which is not possible through the web interface.

That is, the API allows for more detailed customization of the model's behavior and its interaction with other systems. This makes it possible to obtain a better result of grouping wording by meaning and significantly increase the level of automation of this process. Sometimes, it is even possible to make the process of selecting groups of formulations of the same content fully automatic, or with only minor intervention by the organizer of the examination - the knowledge engineer.

It should be noted that an important process of combining individual expert knowledge and transforming it into collective knowledge is used, when, in fact, subjective expert knowledge is combined and objectified and transformed into group knowledge ("collective intelligence"), which is more reliable than the original individual knowledge. The final result of the group work in the form of a graph obtained from the decomposition of the concept "The threats posed by the development of AI tools" into its components - threats, is shown in Fig. 1.


Consensus-2
System for distributed collecting expert information
Vitaly Tsyganok (vvtsyganok@nas.gov.ua)
Logout
Manage Decompositions
Experted Decompositions

Final Tree: The threats posed by the development of AI tools

Yellow nodes means that decomposition for them can be created, just double click.

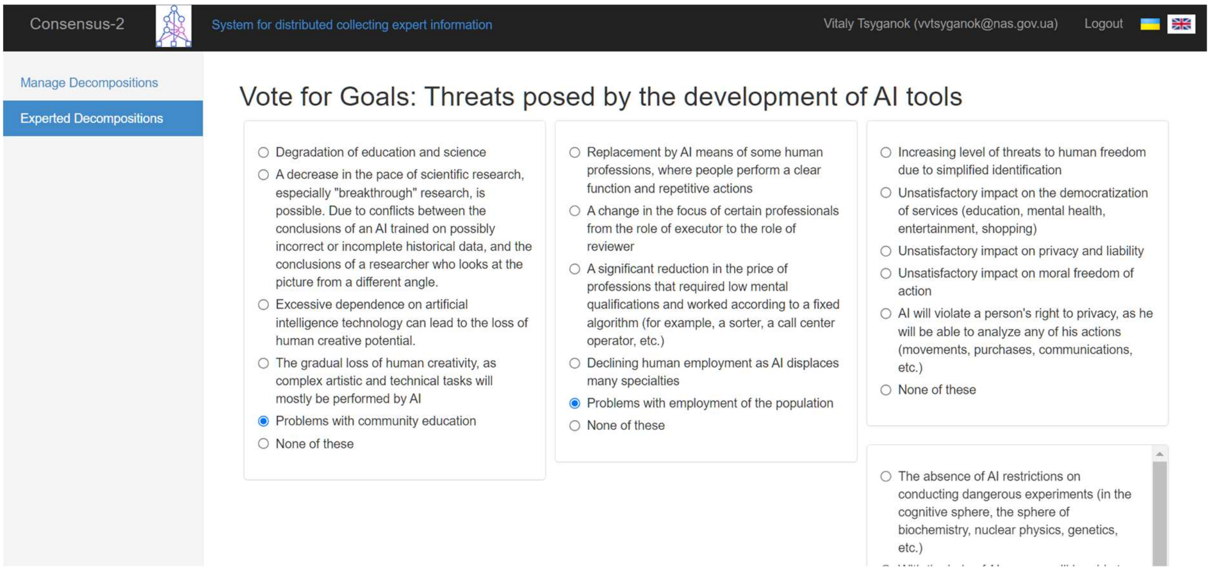
Goals withing blue nodes are also clickable: you can see their decomposition.

Загрози, спричинені розвитком засобів ШІ
Впровадження ШІ в СППР, може призвести до помилки в прогназуванні та відповідно в прийнятті рішення експертом
Надмірна залежність від технології штучного інтелекту може призвести до втрати людського творчого потенціалу
Проблеми зі занятістю населення
Дезинформація людей через фейкові новини, діпфейки та фейкові джерела інформації
Зростання рівня загроз для свободи людини через спрощену її ідентифікацію
Надання ШІ права (правила) на прийняття рішень щодо позбавлення життя людини (
Зміна військового та промислового лідерства між країнами, які можуть навчити провідний ШІ (наприклад, для виявлення цілей у складному середовищі), і яки цього зроб
За допомогою ШІ будь-хто зможе проводити небезпечні експерименти (у сфері біохімії, генетики, когнітивної психології, ядерної фізики тощо)

Figure 1: Image of the graph of decomposition of the concept "The threats posed by the development of AI tools" within the web-interface of the "Consensus-2" system

At this stage, the impersonal wording (without indicating their authorship) was grouped into similar content around the following issues: loss of human creativity; decline in human employment; violation of the right to privacy; conducting dangerous experiments; (un)intentional manipulation of data; granting AI the right to make decisions to take human life; and AI making false decisions to take human life.

Subsequently, the experts who provided the wording of the threats were involved in voting for the best wording in each group of similar wording. An example of an expert voting for the best wording among those of the same content is shown on the screen form corresponding to an episode of the expert's work (Fig. 2).


Consensus-2
System for distributed collecting expert information
Vitaly Tsyganok (vvtsyganok@nas.gov.ua)
Logout
Manage Decompositions
Experted Decompositions

Vote for Goals: Threats posed by the development of AI tools

○ Degradation of education and science
○ A decrease in the pace of scientific research, especially "breakthrough" research, is possible. Due to conflicts between the conclusions of an AI trained on possibly incorrect or incomplete historical data, and the conclusions of a researcher who looks at the picture from a different angle.
○ Excessive dependence on artificial intelligence technology can lead to the loss of human creative potential.
○ The gradual loss of human creativity, as complex artistic and technical tasks will mostly be performed by AI
● Problems with community education
○ None of these

○ Replacement by AI means of some human professions, where people perform a clear function and repetitive actions
○ A change in the focus of certain professionals from the role of executor to the role of reviewer
○ A significant reduction in the price of professions that required low mental qualifications and worked according to a fixed algorithm (for example, a sorter, a call center operator, etc.)
○ Declining human employment as AI displaces many specialties
● Problems with employment of the population
○ None of these

○ Increasing level of threats to human freedom due to simplified identification
○ Unsatisfactory impact on the democratization of services (education, mental health, entertainment, shopping)
○ Unsatisfactory impact on privacy and liability
○ Unsatisfactory impact on moral freedom of action
○ AI will violate a person's right to privacy, as he will be able to analyze any of his actions (movements, purchases, communications, etc.)
○ None of these

○ The absence of AI restrictions on conducting dangerous experiments (in the cognitive sphere, the sphere of biochemistry, nuclear physics, genetics, etc.)

Figure 2: Form for voting for the best wording of the interface of the automated workstation of the expert in the Consensus-2 system

The decision was made by majority vote, provided that all experts were considered equally competent (the relative competence of the expert in the group in the issues under consideration was not taken into account). Such a rather significant simplification of expert procedures was allowed due to restrictions on the duration and cost of examinations, but it is recommended that in the future such procedures be carried out using the methods of pairwise comparisons [14, 15], taking into account the relative competence of the expert in the matter under consideration [16]. Since the relative competence of the experts in the group was not taken into account and, in this case, multiplication of the expert opinion by the relevant competence coefficient of the expert who provided this opinion was not performed when calculating the resulting rating, parity may often arise when determining the best wording among those of the same content. In this case, several wording may receive the same highest rating, and then the best wording is chosen among them at random.

This disadvantage is eliminated (minimized) by taking into account the relative competence of the experts in the group. It should be noted that recent experimental studies [17] have shown the importance of taking into account the relative competence of experts in the group in group evaluation. By simulating expert assessments, the experiment showed the need to take into account the competence of the so-called small groups, in which the number of experts does not exceed 2-3 dozen. Given the high cost of expert labor, conventional expert evaluations are usually conducted with the involvement of small groups, and therefore, taking into account competence is necessary.

Determining the relative competence of experts in a group is the subject of another study [18], the fundamental principle of which is to determine the relative weight of an expert within an expert group. Moreover, the weight of an expert can be determined only in relation to a specific issue that is currently under consideration. When assessing competence, it is proposed to take into account three components: self-assessment, mutual assessment and the objective component of the expert's assessment.

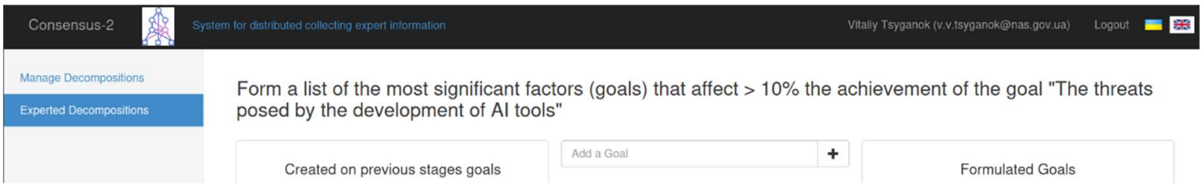
Determining the relative competence of an expert for further consideration in the examination is also a laborious process, requires considerable time and is a costly procedure. For this reason, this procedure was not performed during this stage of the examination related to the generalization of expert knowledge.

At this stage, when voting for the best phrasing among a group of similar options, participants had the option to select "none of the listed" as well as any specific wording (see Fig. 2). If the majority of experts selected this option, none of the threat formulations from that group were included in the final list of threats. In this case, the expert team concluded that the threat either lacked significant impact compared to other formulations or did not qualify as a threat at all.

In essence, such group voting by selecting the best wording or not selecting any of the wording is the definition of the existing links/influences between concepts in the model of the subject area formed by the group of experts, as well as the definition of essential concepts that have a significant impact. From this point of view, group modeling, in which a team of experts participates, is a process of selecting and including in the model only important elements that have a significant impact on the functioning of the system whose model is being built. Moreover, the materiality of the impact is determined by the group of experts at the stage of building the model structure and, subsequently, is specified in the course of impact assessment at the next stage. Thus, as a result of this stage of the examination, eight formulations of threats caused by the development of AI tools were selected (see Table 1).

The final step was to determine the relative importance of the threats from the list. Given that the list of threats resulting from the previous stage of the assessment includes only the most important threats, from the point of view of the group of experts, this list should include those threats whose relative importance is not less than 0.1 (10% of the impacts of all threats). This caveat should be taken into account by each expert when formulating threats and identifying impacts in the group threat model.

Figure 3 shows the interface of the expert's workstation with the proposal “Form a list of the most significant factors (goals) that affect > 10% of the achievement of the goal "The threats posed by the development of AI tools"”.


Consensus-2
System for distributed collecting expert information
Vitaly Tsyganok (v.v.tsyganok@nas.gov.ua)
Logout
Manage Decompositions
Experted Decompositions
Add a Goal
+
Created on previous stages goals
Formulated Goals

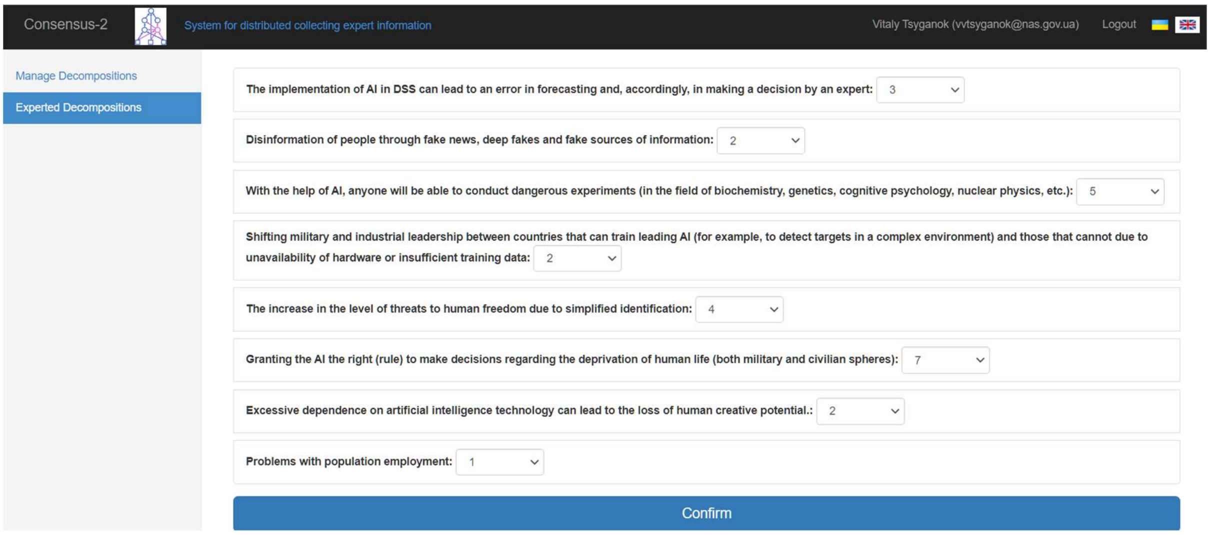
Form a list of the most significant factors (goals) that affect > 10% the achievement of the goal "The threats posed by the development of AI tools"

Figure 3: Interface of the expert's work in the Consensus-2 system to form a list of the most significant threats caused by the development of AI tools

The determination of a numerical rating - the relative importance of the most significant influence factors - is usually carried out using a whole arsenal of expert evaluation methods. The most effective methods that give the most reliable results include methods of obtaining and processing expert information based on pairwise comparisons [14]. In order to avoid pressure on the expert during the evaluation, it is necessary to give the expert the opportunity not to perform a particular pairwise comparison, for example, due to a conflict of interest, due to the expert's lack of information on the subject matter of the expertise, lack of competence of the expert, etc. In other words, the comparison of each pair of alternatives is not mandatory, and preference in expert evaluation is given to methods that use and process incomplete pairwise comparisons [16, 19, 20].

The aggregation of MSEs to find the vector of priority weights is carried out with the obligatory determination of consistency [21, 22], because in the case of aggregation of inconsistent estimates, an unreliable result can be obtained, for example, the average body temperature of patients in a hospital, because this value is not informative. It is important to determine the level of sufficient consistency for further aggregation (the so-called consistency threshold) for a particular assessment and the requirements for the required level of confidence in the results of this assessment [23]. This level should serve as an indicator that the aggregation of valuations is legitimate and the resulting aggregated valuation will be reliable. In the case of insufficient consistency, when it is below the consistency threshold, feedback to the expert is required, in which the expert is contacted again with a proposal to revise his or her previous assessment and increase the level of consistency above the threshold [21].

In the study under consideration, given the limited expert resources, especially the time for conducting the examination, the use of pairwise comparisons and feedback methods is considered laborious and inappropriate in this situation. Therefore, in order to simplify and speed up the group examination, it is proposed to conduct it on a point scale, and to consider expert assessments as consistent without verification. To some extent, such concessions reduce the reliability of the examination results, but for this study these concessions are justified.


Consensus-2
System for distributed collecting expert information
Vitaly Tsyganok (vvtsyganok@nas.gov.ua)
Logout
Manage Decompositions
Experted Decompositions
Confirm

The implementation of AI in DSS can lead to an error in forecasting and, accordingly, in making a decision by an expert: 3

Disinformation of people through fake news, deep fakes and fake sources of information: 2

With the help of AI, anyone will be able to conduct dangerous experiments (in the field of biochemistry, genetics, cognitive psychology, nuclear physics, etc.): 5

Shifting military and industrial leadership between countries that can train leading AI (for example, to detect targets in a complex environment) and those that cannot due to unavailability of hardware or insufficient training data: 2

The increase in the level of threats to human freedom due to simplified identification: 4

Granting the AI the right (rule) to make decisions regarding the deprivation of human life (both military and civilian spheres): 7

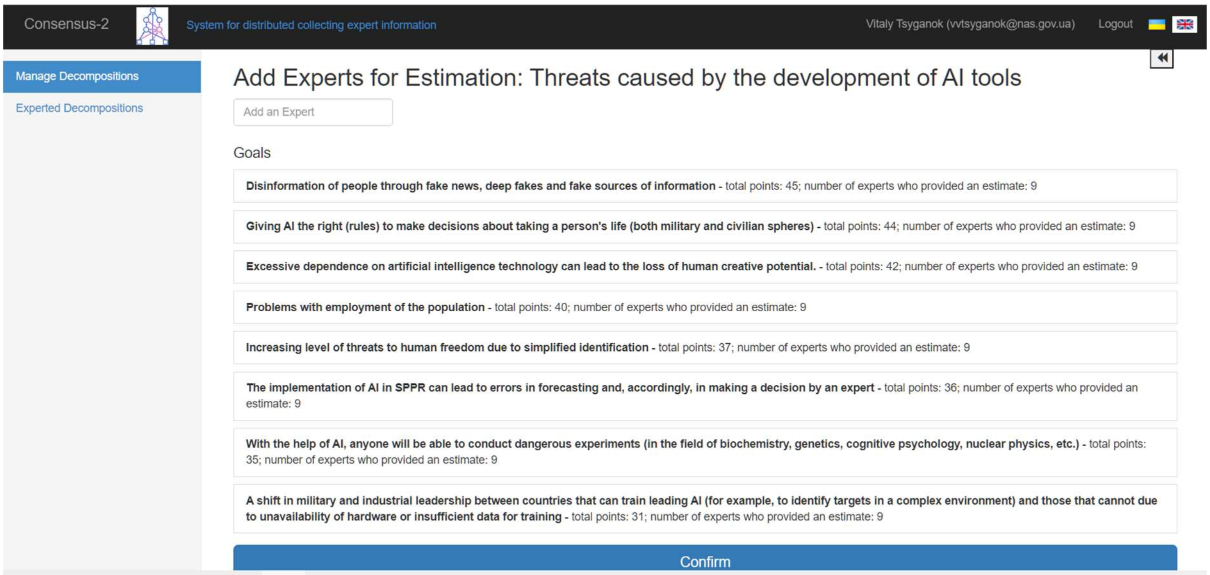
Excessive dependence on artificial intelligence technology can lead to the loss of human creative potential.: 2

Problems with population employment: 1

Figure 4: An example of an expert's interface in the Consensus-2 system for determining the importance of threats

Therefore, in this study, it was decided to conduct a group expert assessment on a 7-point scale. An example of an expert's work on assessing the importance of threats is shown in Fig. 4.

The result of the group examination was a numerical threat rating (Fig. 5), which is formed by summing up the points given to a certain wording by each of the group of experts. Again, the competence of the experts was not taken into account. It should be noted that the point scale assessment has a number of disadvantages that can lead to manipulations in group expert assessment.


Consensus-2
System for distributed collecting expert information
Vitaly Tsyganok (vvtsyganok@nas.gov.ua)
Logout
Manage Decompositions
Experted Decompositions
Add an Expert
Confirm

Add Experts for Estimation: Threats caused by the development of AI tools

Goals

Disinformation of people through fake news, deep fakes and fake sources of information - total points: 45; number of experts who provided an estimate: 9

Giving AI the right (rules) to make decisions about taking a person's life (both military and civilian spheres) - total points: 44; number of experts who provided an estimate: 9

Excessive dependence on artificial intelligence technology can lead to the loss of human creative potential. - total points: 42; number of experts who provided an estimate: 9

Problems with employment of the population - total points: 40; number of experts who provided an estimate: 9

Increasing level of threats to human freedom due to simplified identification - total points: 37; number of experts who provided an estimate: 9

The implementation of AI in SPPR can lead to errors in forecasting and, accordingly, in making a decision by an expert - total points: 36; number of experts who provided an estimate: 9

With the help of AI, anyone will be able to conduct dangerous experiments (in the field of biochemistry, genetics, cognitive psychology, nuclear physics, etc.) - total points: 35; number of experts who provided an estimate: 9

A shift in military and industrial leadership between countries that can train leading AI (for example, to identify targets in a complex environment) and those that cannot due to unavailability of hardware or insufficient data for training - total points: 31; number of experts who provided an estimate: 9

Figure 5: Interface of the automated workstation of the expert organizer at the stage of group expert assessment of the importance of threats

It should be borne in mind that if the maximum number of evaluation points (the size of the rating scale) is set incorrectly, a situation may arise where the rating of one expert distorts the general opinion (judgment) of the entire group of experts. For example, if five experts evaluate a certain alternative on a 100-point scale and four of them give equally low scores, the fifth expert will still be able to significantly increase the rating of this alternative relative to the others by giving it the maximum score. The above example shows that the size of the rating scale should be consistent with the number of experts in the group.

In addition, manipulations are possible when experts collude and give a certain alternative an overestimated score, while at the same time lowering the score for the rest of the alternatives. Other manipulations are also possible when evaluating in point scales. In view of the above, it is proposed to use relative scores in the course of further evaluations. This is, in fact, an alternative to the use of point scales. In addition, it is advisable to provide an opportunity for a particular expert not to perform a particular assessment. That is, any expert can skip any assessment.

In addition to the total number of points in the third column, Table 1 also shows a numerical threat rating - relative values whose sum equals one. That is, each of these values represents a share of the total threat from the development of AI.

In future studies, when conducting expert evaluation, it is proposed to abandon the use of point scales, since such evaluations are highly dependent on the dimension of the survey scale. If the scale dimension for evaluation is not chosen correctly, it is possible that one expert can offset the assessments of other members of the expert group. This may be due to the significant impact of a particular expert's score on the average score on a scale with a large number of points. It is also desirable to take into account the competence of the group members [17], which increases the reliability of generalized expert assessments, thereby objectifying the result obtained on the basis of subjective data.

As already mentioned, AI tools, namely ChatGPT 3.5, were involved in this process as part of the experiment. This generative linguistic AI model can be useful in successfully formulating a query for the task of grouping identical wording. That's why we used this chatbot to answer a question about the threats that could arise from the development of artificial intelligence.

Table 1
Numerical threat rating generated as a result of the examination

Ordinal rating	Threat	Points	Numerical rating
1	Granting AI the right (rule) to make decisions on taking human life (both military and civilian spheres);	47	0,1526
2	Disinformation of people through fake news, fake news and fake sources of information;	45	0,1461
3	Over-reliance on artificial intelligence technology can lead to a loss of human creativity;	42	0,1364
4	Problems with employment;	38	0,1234
5	Increased threats to human freedom due to simplified identification;	36	0,1169
6	With the help of AI, anyone will be able to conduct dangerous experiments (in biochemistry, genetics, cognitive psychology, nuclear physics, etc.);	35	0,1136
7	The introduction of AI in the PDRS may lead to errors in forecasting and, accordingly, in the decision-making of the expert;	34	0,1104
8	Shift in military and industrial leadership between countries that can train advanced AI (e.g., to detect targets in complex environments) and those that cannot due to lack of hardware or insufficient training data;	31	0,1006

To this question, ChatGPT 3.5 generated an answer about potential threats in the field of AI, namely (here is the content generated by the chatbot without edits):

- Loss of jobs: the development of automation and autonomous systems may lead to the loss of jobs in industries where people perform repetitive tasks;
- Data privacy and security: The increased ability of AI to analyze and process information can lead to data privacy and security issues if appropriate protection measures are not in place;
- AI weapons: the use of artificial intelligence in the field of military technology may raise ethical issues and increase the risk of uncontrolled hostilities;
- Ethical considerations: The development and use of AI also raises a number of ethical issues, including determining responsibility for AI actions, transparency of systems, and issues of equitable and multicultural use;
- dependence on technology: as AI develops, society may become more dependent on these technologies, which may have negative social and economic consequences;
- Failure to control: AI can develop so rapidly that society may not be able to adequately control its expansion and application.

2.4. Future directions for ethical AI development and control

In addition to the described expertise on identifying threats caused by the development of AI tools, the expert community raised the question of the conditions for AI to possibly get out of human control. One of the many expert opinions was the original expert opinion that AI may cease to be subject to humans and surpass them in terms of mental abilities when AI is able to plan and conduct experiments independently, i.e., to acquire knowledge independently. It is also important that AI is able to test this knowledge in practice, because the philosophical principle "practice is the criterion of truth" is more relevant than ever.

To prevent AI from getting out of human control, it seems important to control the access of AI systems to expert knowledge, to sensitive behavioral information about a person, and to any sources of knowledge. In addition, there was an expert opinion that it is necessary to limit the access of AI tools to verify their generated results in practice.

Additional commentary from ChatGPT 3.5: "It is important to emphasize that many of these issues can be addressed with proper regulation, the development of ethical standards, and the inclusion of different stakeholder groups in the decision-making process. AI has great potential to develop and improve lives, but its development also requires attention to possible negative consequences."

Conclusions

In the course of the collective examination, the potential threats to the intensive development of artificial intelligence were analyzed using the Consensus-2 system of distributed collection and processing of expert information for decision support systems. A list of the most influential threats was compiled and their relative importance was determined as a result of group expert evaluation.

The study showed the existence of potential threats from the intensive development of AI, such as loss of human creativity; reduction of employment; violation of the right to privacy; conducting dangerous experiments; (un)intentional manipulation of data; granting AI the right to make decisions to take human life; and making false decisions to take human life by AI.

Researching and responding to potential threats from AI requires intensified efforts by both the state and civil society. In order to address this problem in a more systematic and qualitative way, it is necessary to:

- actively participate in key international events on the above-mentioned issues, such as AI security summits (conferences) initiated by leading countries and the UN;
- raise the level of digital literacy of the population. It is important to implement the Roadmap for AI regulation in Ukraine, which should, among other things, help ordinary citizens learn how to protect themselves from AI risks;
- institutionalize research on this issue by creating new government institutions (such as the AI Security Institutes already established in the United States and the United Kingdom) with the involvement of private sector experts who will deal with AI security at the national level;
- develop probable scenarios of potential threats from the development of AI with appropriate indicators and measures to minimize the identified threats.

Further research in this area is planned to be continued with the involvement of a wider range of experts and using appropriate intelligent technology to generate probable scenarios for the realization of potential threats from the development of AI.

References

[1] Illiashenko, O., Kharchenko, V., Babeshko, I., Fesenko, H., & Di Giandomenico, F. Security-Informed Safety Analysis of Autonomous Transport Systems Considering AI-Powered Cyberattacks and Protection. *Entropy.* 2023. 25(8), 1123. DOI: 10.3390/e25081123

[2] Kharchenko, V., Fesenko, H., & Illiashenko, O. Quality Models for Artificial Intelligence Systems: Characteristic-Based Approach. Development and Application. *Sensors.* 2021. 22(13), 4865. https://doi.org/10.3390/s22134865

[3] Moskalenko, V., Kharchenko, V., Moskalenko, A., Kuzikov, B. Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, Models and Methods. *Algorithms.* 2023. 16, 165. https://doi.org/10.3390/a16030165

[4] The Bletchley Declaration by Countries Attending the AI Safety Summit, November 1-2, 2023. URL: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

[5] New UN Advisory Body aims to harness AI for the common good. URL: https://news.un.org/en/story/2023/10/1142867.

[6] Ukraine has signed an international declaration on the safe use of AI. URL: https://www.epravda.com.ua/news/2023/11/2/706166/.

[7] Prime Minister's speech at the AI Safety Summit: November 2, 2023. URL: https://www.gov.uk/government/speeches/prime-ministers-speech-at-the-ai-safety-summit-2-november-2023.

[8] Software web-system for distributed collection and processing of expert information for decision support systems ("Consensus-2”). Certificate of copyright registration for the work No. 75023 dated 11/17/2017. https://dss-lab.org.ua/applications/consensus-2

[9] Miller, G.A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review.* 1956. 63(2), 81-97.

[10] GhatGPT. URL: https://chat.openai.com.

[11] Galactica. URL: https://galactica.org.

[12] Houston A. B., Corrado E. M., Embracing ChatGPT: Implications of Emergent Language Models for Academia and Libraries. *Technical Services Quarterly.* Vol. 40, Iss. 2, 2023. 76-91. https://doi.org/10.1080/07317131.2023.2187110.

[13] S. Atlas, ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI - 2023. URL: https://digitalcommons.uri.edu/cba_facpubs/548.

[14] Totsenko V.G., Tsyganok V.V. Method of paired comparisons using feedback with expert. Journal of Automation and Information Sciences. 1999. Vol.31, No9. P.86-97. https://doi.org/10.1615/JAutomatInfScien.v31.i7-9.480.

[15] Vitaliy Tsyganok, Oleh Andriichuk, Sergii Kadenko, Yaroslava Porplenko, and Oksana Vlasenko An Approach to Reducing the Number of Pair-Wise Alternative Comparisons During Individual and Group Decision-Making. Studies in Computational Intelligence Volume 1022. System Analysis & Intelligent Computing. Theory and Applications. Editors: Michael Zgurovsky & Nataliya Pankratova. Springer, 2022, 163-183. https://doi.org/10.1007/978-3-030-94910-5_9

[16] Zgurovsky M.Z., Totsenko V.G., Tsyganok V.V. Group Incomplete Paired Comparisons with Account of Expert Competence. *Mathematical and Computer Modelling.* - February 2004. - v.39, N4-5 . P.349-361. https://doi.org/10.1016/S0895-7177(04)90511-0

[17] Tsyganok V.V., Kadenko S.V. & Andriichuk O.V. Significance of Expert Competence Consideration in Group Decision Making using AHP. *International Journal of Production Research.* 2012. v.50, issue 17. P.4785-4792. https://doi.org/10.1080 /00207543.2012.657967

[18] Totsenko, V. Methods and Systems for Decision-making Support. Algorithmic aspect. *Naukova dumka.* Kyiv. 2002.

[19] Bozóki Sándor & Tsyganok Vitaliy The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. *International Journal of General Systems.* 2019. Vol.48, No.4. P.362-381. https://doi.org/10.1080 /03081079.2019.1585432

[20] Szádoczki, Z., Bozóki, S., Juhász, P., Kadenko, S., Tsyganok, V. Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Annals of Operations Research* (2023). https://doi.org/10.1007/s10479-022-04819-9

[21] Roik P.D., Tsyganok V.V. A method for improving the consistency of expert assessments in the course of a dialog. *Data Recording, Storage & Processing.* 2018. т.20. №2. С.85-95. (in Ukrainian) https://doi.org/10.35681/1560-9189.2018.20.2.142915

[22] Tsyganok V.V., Roik P.D. Method for determining and improving the consistency of expert estimates in supporting group decision-making. *System research and information technology.* 2018., N3. 110-121. (in Ukrainian) https://doi.org/10.20535/SRIT.2308-8893.2018.3.10

[23] Tsyganok Vitaliy, Olenko Andriy, Roik Pavlo, Vlasenko Oksana. Determining adequate consistency levels for aggregation of expert estimates. arXiv preprint. Methodology (stat.ME) arXiv:2410.03012v1 [stat.ME] 3 Oct 2024 9p. https://doi.org/10.48550/arXiv.2410.03012