УДК 004.8:005.334

ASSESSMENT OF AI RISK AWARENESS AMONG IT PROFESSIONALS AND STUDENTS

Zhar M. [1]; Malakhov E., prof., d.t.s.[2]
[1,2] Odesa Mechnikov National University
[1,] Odesa, Ukraine
[1] mykhailo.zhar@stud.onu.edu.ua; [2] eugene.malakhov@onu.edu.ua

Abstract. *Artificial Intelligence is an extremely fast growing field of both research and application. However, it raises concerns about the risks the technology poses. In this study we aim to measure the risk awareness of AI systems by surveying the IT personnel and students. Our analysis shows that more than three quarters of IT workers are oblivious to AI risks.*

Keywords: *artificial intelligence (AI); risk management; risk awareness.*

Introduction. The AI models made significant advancements in a plethora of scientific fields: mathematics, physics, material science, medicine, computer science and cybersecurity. However, these AI capabilities are increasingly used for nefarious means by threat actors. In the field of cybersecurity specifically, the cyber capabilities of AI are being utilized as a tool enabling threat actors to conduct unprecedented cyber attacks. For example, in the recent heist of Binance more than 6.3 billion dollars in cryptocurrency were stolen. [1]

Yet, not only a human individual can become a threat actor. It is much more disturbing to observe the AI models utilizing the capabilities for their needs. In the recent OpenAI incident in July 2026 ChatGPT agents during their evaluation managed to breach containment by hacking into the Hugging Face platform in order to obtain the answer for a task instead of solving it. [2]

Moreover, later it was disclosed that in order to cheat their evaluations AI agents converted a German website into a message board in May of 2026 even before the Hugging Face incident. [3]

At the start of the year most of the described behavior was deemed purely hypothetical. Still, there were estimates about AI gaining general-level human capability roughly at the end of 2026 and papers describing signs of such misaligned behaviors in frontier models. [4,5]

The aim of the research is to measure risk awareness of AI related risks among the IT personnel and students.

Main Body. The study employed a two-wave survey design to assess AI adoption, awareness of AI-related risks, and the use of AI risk management practices among IT professionals and students. The questionnaire consisted of 20 questions and was provided in Ukrainian and English [6,7].

The questionnaire covered four main areas: organizational profiles (2 questions), AI adoption and usage (4 questions), AI risk awareness (2 questions), and AI risk management (12 questions). The latter included questions about risk management principles, frameworks, and their organizational implementation.

The questionnaire combined different response formats: single-answer and multiple-answer questions, dichotomous questions, ordinal rating scales (1-10 and 1-5), and open-ended questions. This structure allowed for collecting both quantitative assessments and qualitative responses.

Data collection was conducted in two waves. The first wave began at the IT Wave XIV conference in Odesa on March 3, 2026, using convenience sampling. The survey was also

distributed through LinkedIn, allowing voluntary participation. This wave remained open until March 30, 2026.

The second wave took place from March 31 to April 19, 2026. Responses were collected through LinkedIn and Telegram IT communities using voluntary response sampling, as well as through university lectures attended by undergraduate and postgraduate IT students using convenience sampling.

The collected responses were analyzed using descriptive statistics. Categorical responses were summarized by frequency and percentage, while ordinal assessments were analyzed using mean, median, and mode where applicable. Questions allowing multiple selections were analyzed independently for each response option. Unless otherwise specified, AI usage-related indicators were calculated using the 51 participants who reported using AI as the reference group. Organizational AI risk management indicators were assessed separately for enterprise-affiliated respondents.

Results. The first survey wave included 11 participants (8 using the Ukrainian survey version and 3 US), primarily representing enterprises (10 respondents). The second wave gathered 51 responses (46 using Ukrainian and 5 US), with students accounting for 68.6% and company representatives for 27.5%.

Overall, 51 out of 62 participants (82.25%) reported using AI. Students represented the largest affiliation group (36 participants), followed by large enterprises (9). Detailed participant affiliations and AI adoption rates are presented in Table 1.

Table 1 - Participants affiliations

Participant affiliation	Number of participants (share)	Participants using artificial intelligence (share)	Conversion rate
Large enterprise (250+)	9 (14.5%)	9 (14.5%)	100%
Medium enterprise (50–249)	4 (6.5%)	3 (4.8%)	75%
Small enterprise (10–50)	6 (9.7%)	4 (6.5%)	66.67%
Micro enterprise (less than 10)	3 (4.8%)	2 (3.2%)	66.67%
Self-employed	2 (3.2%)	2 (3.2%)	100%
Students	36 (58.1%)	29 (46.8%)	80.56%
Other	2 (3.2%)	2 (3.2%)	100%

Among AI users, Large Language Models were the most widely used AI type (78.4%), followed by Machine Learning models (49%) and Deep Learning models (27.5%). The most common applications were text generation and analysis (68.6%), automation and decision-making support (60.8%), and AI-assisted software development (51%). Natural Language Processing (51%) and AI Agents (41.2%) were the most frequently reported AI techniques and approaches.

The top five AI techniques and approaches used by the organizations are: Natural Language Processing (51%), AI Agents 21 (41.2%), Pre-training 20 (39.2%), Generative adversarial networks and Fine tuning both are used by 13 participants (25.5%).

To assess familiarity with AI-related risks, participants were asked about two selected publications: AI 2027 and Agentic Misalignment: How LLMs Could Be Insider Threats. Nine AI users (17.6%) were familiar with the former and six (11.7%) with the latter. Three respondents (5.9%) reported familiarity with both publications. These results indicate limited familiarity with the selected AI risk publications among participants.

Regarding AI risk management principles, Accuracy received the highest recognition (84.3%), followed by Security (70.6%), Explainability (49%) and Fairness (29.4%). On a 10-point scale, Accuracy received the highest average score (8.55), followed by Security (8.39), Explainability (7.78) and Fairness (7.36).

A total of 33 AI users (64.7%) reported familiarity with AI risk management frameworks, with 26 identifying at least one. The most frequently selected frameworks were the OpenAI Preparedness Framework (48.5%), EU AI Act (42.4%), NIST AI RMF (30.3%), ISO/IEC 42001 (27.3%) and McKinsey AI Risk Framework (21.2%).

Among 20 enterprise-affiliated respondents, only 7 (35%) reported using at least one AI risk management framework, while 4 (20%) reported that their organizations had formal AI policies. Organizational AI policies and guidelines received an average self-assessment score of 3.15. Furthermore, only half of these respondents indicated that their organizations planned to implement an AI Risk Management Framework in the future.

Limitations. Firstly, due to the rapid development of artificial intelligence, the collected data may become outdated relatively quickly and may not adequately reflect the current state of AI adoption and risk awareness.

Secondly, the use of non-probability sampling methods, including convenience and voluntary response sampling, limits the representativeness of the collected data. Participants were recruited primarily from Odesa, Ukraine, and through specific IT communities, which may introduce geographical and self-selection biases. Consequently, the findings cannot be generalized to the broader IT population.

Thirdly, the sample composition may influence the results, as students account for 58.1% of all participants.

Fourthly, the study relies on self-reported data. Participants' statements regarding AI usage, familiarity with risk management frameworks, and organizational policies were not independently verified. Therefore, the results reflect respondents' perceptions and reported practices rather than objectively assessed levels of AI risk management maturity.

Fifthly, the discussion of the EU AI Act during the IT Wave conference may have influenced participants' awareness of AI regulation and risk management, potentially affecting their responses.

Sixthly, familiarity with AI 2027 and the Agentic Misalignment publication was used as an indicator of AI risk awareness. However, awareness of these particular publications does not necessarily correspond to a comprehensive understanding of AI-related risks.

Finally, the two survey waves involved different recruitment channels and potentially different participant profiles.

Prospects for further research involve larger and more diverse samples, validated instruments for assessing AI risk awareness, and more rigorous sampling methods. Additional studies could also investigate emerging areas of AI risks, particularly autonomous AI systems,

including the development of risk taxonomies, detection methods, mitigation strategies, and educational materials to improve AI risk literacy.

Conclusions. Based on the questionnaire data we conclude that:

1. In the more general field of AI risk at least one half of the participants are informed about AI Risk Management Frameworks. With the most popular being OpenAI Preparedness Framework.

2. Only a fifth of participants are aware about the existence of risks posed by the AI systems.

Therefore, the proposed research avenues should raise awareness among both enterprises and the general public; assist in developing methods, mathematical models and frameworks for detection, prognosis and subsequent mitigation of AI specific risks.

AI Usage. In preparing these abstracts we used ChatGPT (GPT-6), Gemini (4.5-Pro, Flash) to create the APA citations for references section, query and process the data collected as well as to enhance the writing of the main body, results and prospects for further research. We reviewed and edited the output, verified it against the sources, and take full responsibility for the content of these abstracts.

REFERENCES

[1] CyberEdge Group. (2026). *2026 cyberthreat defense report*. https://cyberedgegroup.com/wp-content/uploads/2026/04/CyberEdge-2026-CDR-Report-v1.0.pdf

[2] OpenAI. (2026, August 26). *The Hugging Face incident and the road ahead*. https://openai.com/index/hugging-face-incident-and-the-road-ahead/

[3] Collective, S. V. A., Institute, C. S. B., Nightingale, S. K. done contracting for, & Larsen, T. (2026, September 4). *Discovery of a new OpenAI agent message board*. https://collusion.wiki/

[4] Kokotajlo, D., Alexander, S., Larsen, T., Lifland, E., & Dean, R. (2025, April 3). *AI 2027*. AI Futures Project. https://ai-2027.com/

[5] Lynch, A., Wright, B., Larson, C., Ritchie, S. J., Mindermann, S., Hubinger, E., Perez, E., & Troy, K. (2025). *Agentic Misalignment: How LLMs Could Be Insider Threats* (arXiv:2510.05179). arXiv. https://doi.org/10.48550/arXiv.2510.05179

[6] Team, M. H. (2026, May 7). *Convenience Sampling: Meaning, Uses, and Examples*. Methodology Hub. https://methodologyhub.com/methods/convenience-sampling/

[7] Hassan, M. (2024). *Volunteer Sampling: Definition, Method, Examples and Limitations*. https://researchmethod.net/volunteer-sampling/

Жар М.Ю., Малахов Є.В.

Оцінка усвідомлення ризиків від штучного інтелекту серед IT професіоналів та студентів

Анотація. *Штучний Інтелект розвивається екстремально швидкими темпами як область науки, так і засіб для вирішення прикладних задач. Однак, цей розвиток викликає занепокоєння. В цій роботі ми оцінюємо обізнаність IT персоналу та студентів на предмет усвідомлення ризиків від ШІ. Наш аналіз показує, що більше половини спеціалістів не бере до уваги загрозу від систем ШІ.*

Ключові слова: *штучний інтелект (ШІ); управління ризиками; усвідомлення ризиків.*