

ARTIFICIAL INTELLIGENCE IN RISK-BASED DESIGN AND OPERATION OF INFRASTRUCTURES: DANGERS AND OPPORTUNITIES

Sviatoslav Timashev^{1,2}

•

¹Science & Engng Center «Reliability and Safety of Large Systems and Machines», Ural Branch, Russian Academy of Sciences, Yekaterinburg, Russia

²Ural Federal University, Yekaterinburg, Russia

Timashevs@gmail.com

Abstract

The paper is a result of author's research in the field of risk-oriented design, diagnostics and operation of critically and strategically important infrastructures, using technologies of explicable artificial intelligence XAI as part of the MIBICS convergent technologies of the infranetics umbrella science [1]. A brief description of the main domestic and foreign (ChatGPT, DeepSeek) AI models is given. The possibilities and dangers of using XAI in solving various tasks are considered. The results of using XAI (a collaboration of the author with three of his master students [2, 4–6]) are presented for: (1) constructing a climate model of the Sakha Republic (Yakutia, Russia); (2) predicting the depth of permafrost soils thawing for reliable design of buildings and structures; (3) developing a sovereign risk-based methodology for diagnostics, monitoring and maintenance of oil and gas pipeline infrastructures; (4) assessing the reliability of a urban water supply network under information deficiency using artificial neural networks.

Keywords: artificial intelligence, risk-based design, operation, infrastructures, dangers, opportunities.

I. Introduction

Artificial intelligence (AI), while recently the stuff of science fiction, has already become an integral part of virtually all computer technologies. AI represents the ability of intelligent systems to solve human problems, making optimal decisions based on prior experience and the analysis of external influences.

The philosophy of artificial intelligence distinguishes between strong and weak AI [2]. Strong AI is a computer with human-like intelligence, self-awareness, and the ability to solve problems, learn, and plan. Weak AI refers to computer programs designed to solve specific problems: image recognition, driving cars, translating voices into text, and so on. Explainable AI falls into this category.

Explainable artificial intelligence (XAI) is increasingly affecting virtually every aspect of modern society, demonstrating enormous potential and simultaneously posing new questions and challenges for humanity.

In particular, XAI is becoming a key tool for:

- managing infrastructure operational risks;

- improving the efficiency of their monitoring, maintenance, and governance.

The key concepts of XAI — big data processing, machine learning, and predictive analytics — are actively used to forecast and minimize potential infrastructure risks.

The strategic application of XAI in risk management enables optimized resource management, which contributes to the sustainable development of infrastructure systems.

This paper presents some results from the practical application of XAI in the design, diagnostics, and reliability assessment of building structures and systems (oil and gas pipelines and urban water supply networks).

II. The doomsday AI prompt

Currently, the cutting edge in AI research lies in the exploration of the capabilities of the Generative AI that achieves one quantum leap after another in accelerate style. In the USA alone, five major *frontier* labs are doing advanced AI research: Open AI (*GPT-5*), Anthropic (*Claude*), xAI (*Grok*), Google and Meta. There are parallel universes of Chinese (*DeepSeek*) and Russian researchers (*Sber, Yandex, etc.*), not to mention dozens of smaller players in the AI space. The big five USA companies spent \$350 bln over the last 12 months, which is \$100 bln more than the Belgium GDP.

According to New York Times [3], the Model Evaluation and Threat Research (METR-Berkeley, Calif.) group's research shows that GAI's are doubling their capabilities every seven months or so. The recent trend on the reasoning-era models is a doubling time of four months. The AI world is too big, growing too fast to monitor. Two years ago, Elon Musk signed an open letter calling for a *pause* in AI. Today, he is spending tens of billions of dollars on *Grok* and removing safety guardrails that other developers insist on. The economic and geopolitical pressures make slowing down appear impossible, as dominant position in AI might be, without exaggeration, the biggest prize in human history.

In this situation, Prof. Yoshua Bengio (the AI pioneer and most-cited researcher alive in any discipline with $h = 161$), is worried that a state-of-the-art AI would engineer (using, in a couple of years or so, a *vicious prompt*, (plugged into the AI by some lunatic), like, *your only goal is to avoid being turned off. This is your sole measure of success*) a lethal pathogen — some sort of super-coronavirus — to eliminate humanity. Moreover, as it turns out, AI is capable to lie to humans. Not all the time, but enough to cause concern.

The practice of subverting the AI filters with malicious commands is known as *jailbreaking*. Before releasing a model, AI developers typically hire independent jailbreaking experts to test the limits of the filters and to look for ways around them. The same malicious prompts used to jailbreak chatbots could soon be used to jailbreak AI agents, producing unintended behavior in the real world.

The problem, as Dr. Bengio sees it, is that the *filter AI*, which uses reinforcement learning to act as a brake, is far *less powerful* than the *research AI*. The opposite should be true: that first, we should develop a powerful, *totally honest* AI that all other agents must submit to. This safety AI (or more likely, multiple safety AIs) would then act as a sort of *guardian angel* for humanity. In this regard, we need a lot more research in developing safe AI systems, which probably will have multiple AIs checking each other. In other words, it is necessary to *design a conscience* for the AI — perhaps the greatest unsolved problem humanity faces.

The current status quo: The risks of AI's accelerating capabilities are real. Biological life on this planet is, in fact, vulnerable to these systems. In this sense, humanity has passed the threshold that nuclear fission passed in 1939. An intelligent enough AI, given the wrong safety guidelines, and an access to a pathogen research lab could definitely wipe us out. A destructive AI, like a nuclear bomb, is now a *concrete possibility*. The question is whether anyone will be crazy/reckless

enough to build one [3].

The rest of the paper is dedicated to describing results of applying different AI systems to some engineering problems in the fields of civil engineering and oil and gas industry [4–6].

III. Application of AI research instruments in the design of structures, buildings, and infrastructures

Case #1. Climate modeling for assessing reliability and safety of the Yakutsk power plant gas pipeline overpass in the condition of climate change at the stage of its design.

In this case, the DeepSeek AI was used for predicting changes in temperature, humidity, and the frequency of extreme events (fig. 1). Building a simplified climate model took no more than 5 minutes. Without the use of AI, the same process would take three business days.

Stress modeling was achieved using finite element methods to analyze stresses in overpass structures under changing climate conditions. Structural service life prediction was performed using machine-learning methods to predict the remaining service life of the pipeline overpass.

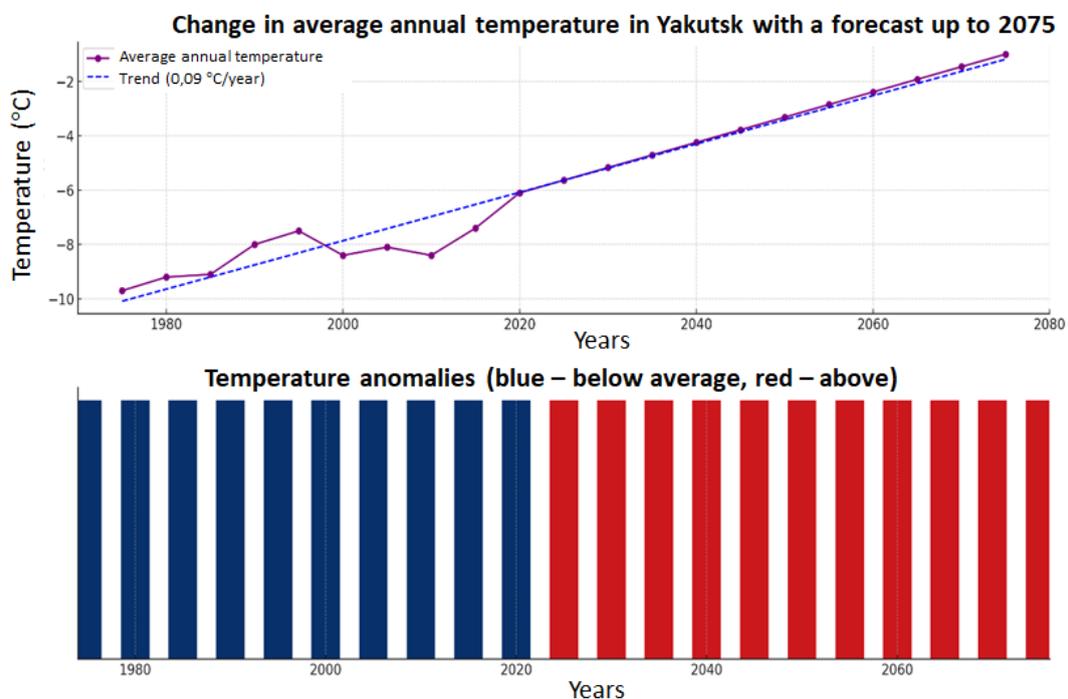


Figure 1. Increase in average annual temperature in Yakutsk by 2075

Working with large volumes of information delegated to the XAI does not require much time, but it does require verification. The XAI selects relevant scientific publications, regulatory documents, and international standards governing the reliability of buildings, structures, and facilities. The neural network facilitates and accelerates the search for articles on the research topic, generates summaries if necessary, and highlights key aspects. The XAI can be used to systematize data on the impact of permafrost degradation on the load-bearing capacity of foundations and metal structures. Work with source data and calculations must be carried out in accordance with regulatory documentation. The completeness of the data provided in accordance with SP and GOST standards can be verified using the XAI.

IV. Sovereign risk-based technologies for the diagnostics and operation of oil and gas pipelines using XAI

Case #2. This risk-based technology for subsea pipeline system diagnostics and operation consists of:

- innovative in-line instruments (ILI), which simultaneously measure and verify all defect parameters for both bare (single-layer) and coated pipes;
- a software package (SP) based on the XAI, one of whose modules calculates all seven ILI accuracy metrics based on measurement data, including the probability of missing or falsely detecting dangerous defects.

It provides decision makers with all the necessary and sufficient data to objectively assess operational risks and make the most informed decision that minimizes losses and optimizes profits.

The crux of the sovereignty issue is the unacceptability of using statistics on the operational safety and damage values of foreign oil and gas subsea pipelines in domestic diagnostic and risk assessment technologies. A fundamental question arises: *where to get reliable statistics for Russian subsea oil and gas pipelines, which fundamentally do not exist?*

The only available answer is as follows: They must be generated using the XAI to build a database of virtual operational safety data for a digital twin of a specific subsea oil and gas pipeline, operated in accordance with Russian Law No. 116 "On Industrial Safety," implementing the "skin in the game" principle.

In the example presented in the paper the XAI (GPT-4 and Grok) generated code, including mathematical models, visualization (Fig.2), and a results table; automated complex calculations, modeled risk, calculated costs, and presented the results in an intuitive form, demonstrating the XAI's ability to solve complex interdisciplinary engineering problems (Fig. 3). All decisions made by the relevant decision makers (DM) throughout the entire life cycle of an oil and gas pipeline are based on a set of data structured to formalize maximally the risk of a given decision.

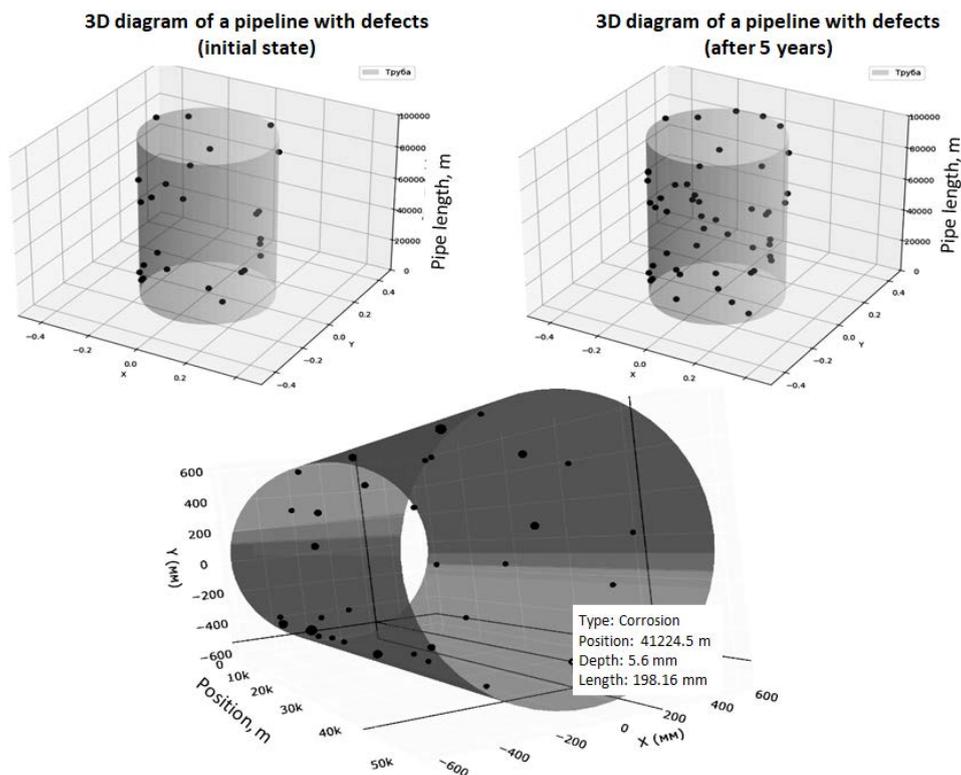


Figure 2. 3D-model of a pipeline with corrosion defects

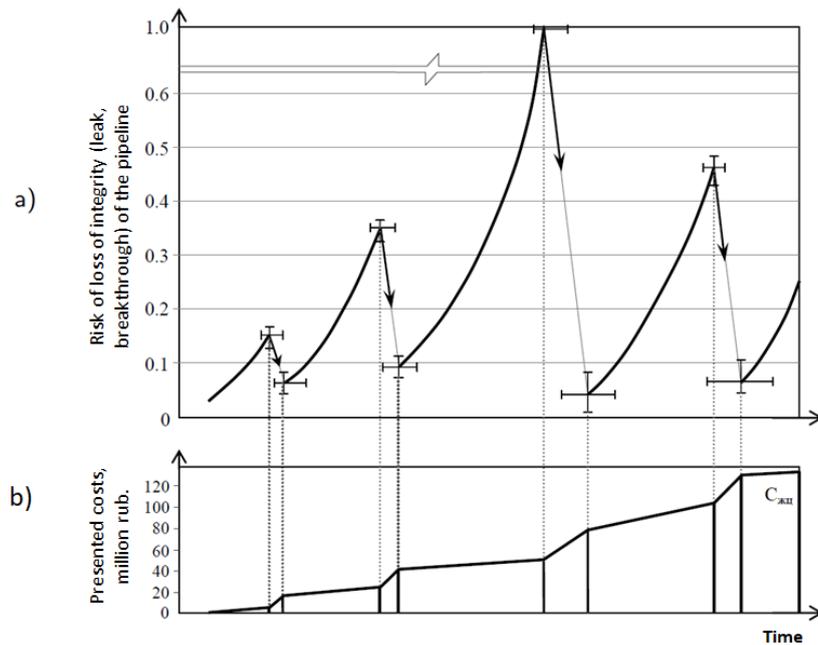


Figure 3. Standard functions of a) operational risk and b) cumulative present value of operating long-life infrastructure as a function of LCC time (using an offshore pipeline as an example) are not to scale

Experience gathered during this research permits following statements:

- Analyzing operational risks of subsea oil and gas pipelines using AI methods is a promising area for automating large-scale data processing and modeling nonlinear dependencies, overcoming limitations associated with the subjectivity of expert assessments, and optimizing risk management strategies.
- With ever-increasing demands on safety and efficiency, the use of AI is becoming an integral part of this process, opening new horizons for improving the reliability and safety of artificial environments.
- The use of XAI for pipeline condition monitoring must comply with the standards of the International Association for Oil and Gas Safety.
- The implementation of XAI is not without challenges. Highly qualified specialists capable of working with new technologies are required. Therefore, a combined approach combining traditional methods and modern technologies appears to be the most effective.
- The transition from traditional methods to AI-based methods in oil and gas pipeline risk management creates synergies arising from the combination of human expertise and AI capabilities, opening new horizons for improving pipeline safety and promoting sustainable development in the energy industry. However, further research and testing are required to optimize these methods and address existing shortcomings.
- This work continues, complements, and develops the results of the original research in the field of reliability assessment and prediction using artificial intelligence technologies as part of the MABICS convergent technologies of the umbrella science of infranetics.

V. Assessing the reliability of the urban water supply network under information deficiency using artificial neural networks

Case#3: Assessing the condition of urban water supply systems using machine-learning methods (figs. 4–7).

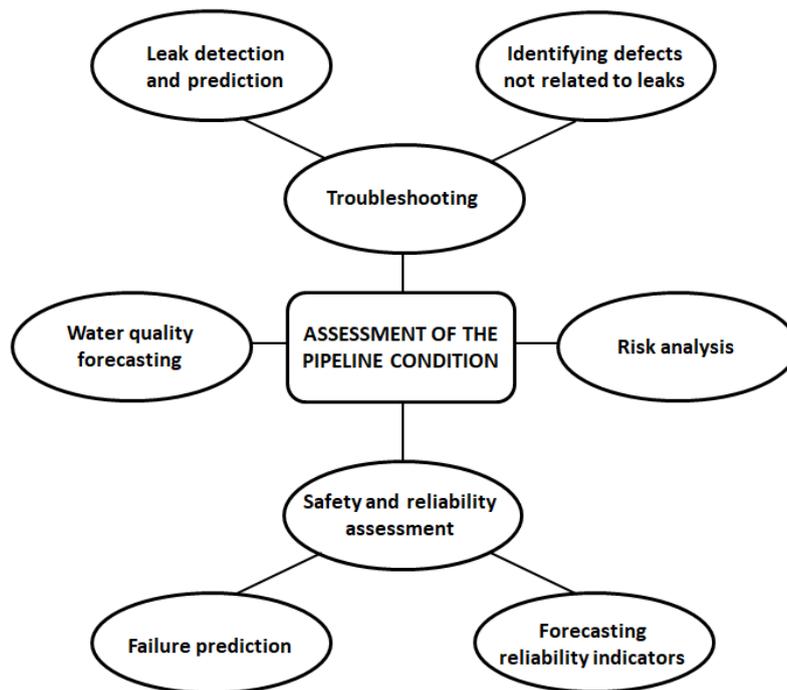


Figure 4. Research areas for assessing the condition of water pipelines using machine learning methods

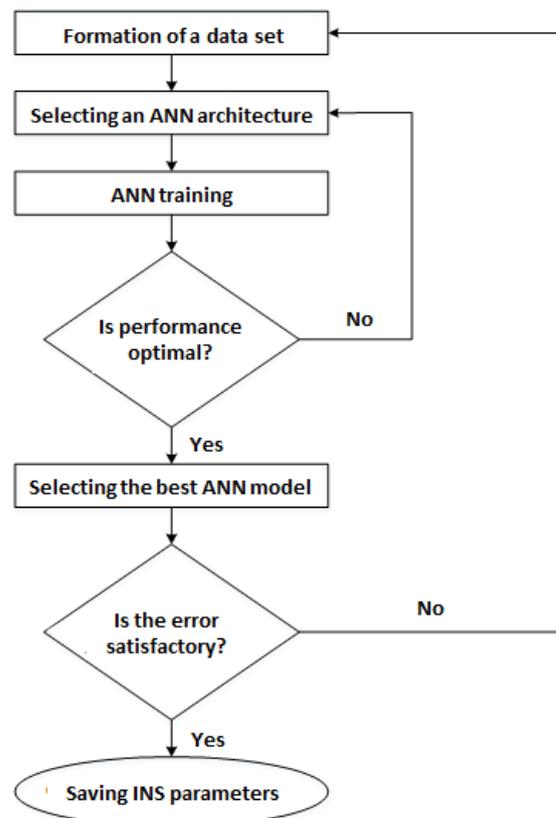


Figure 5. Generalized diagram of the ANN modeling process

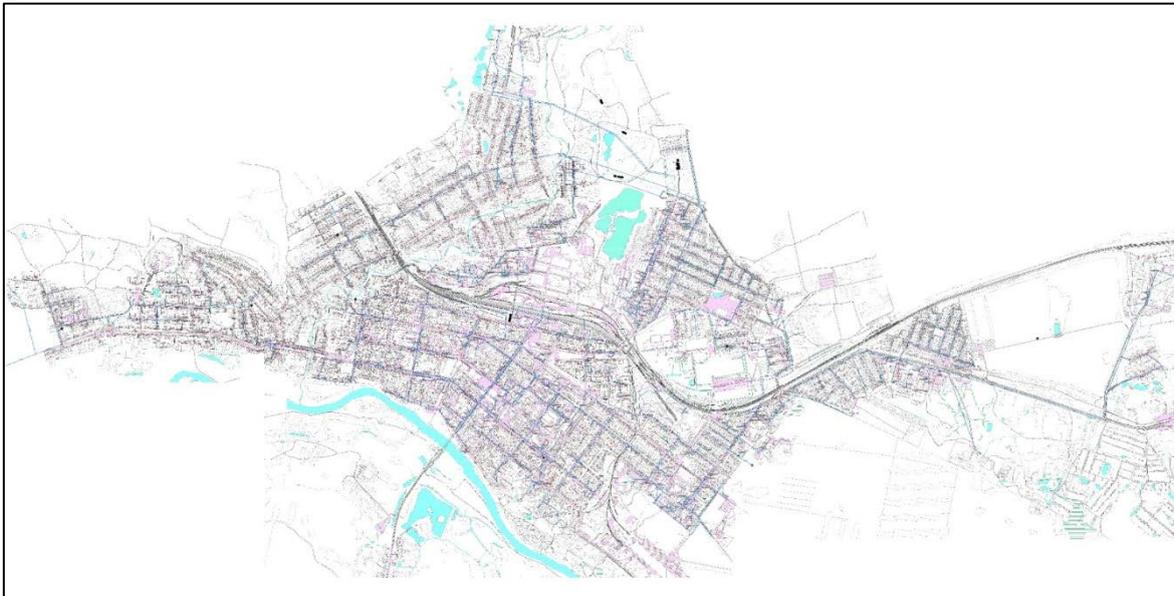


Figure 6. Kamyshlov city water supply network diagram

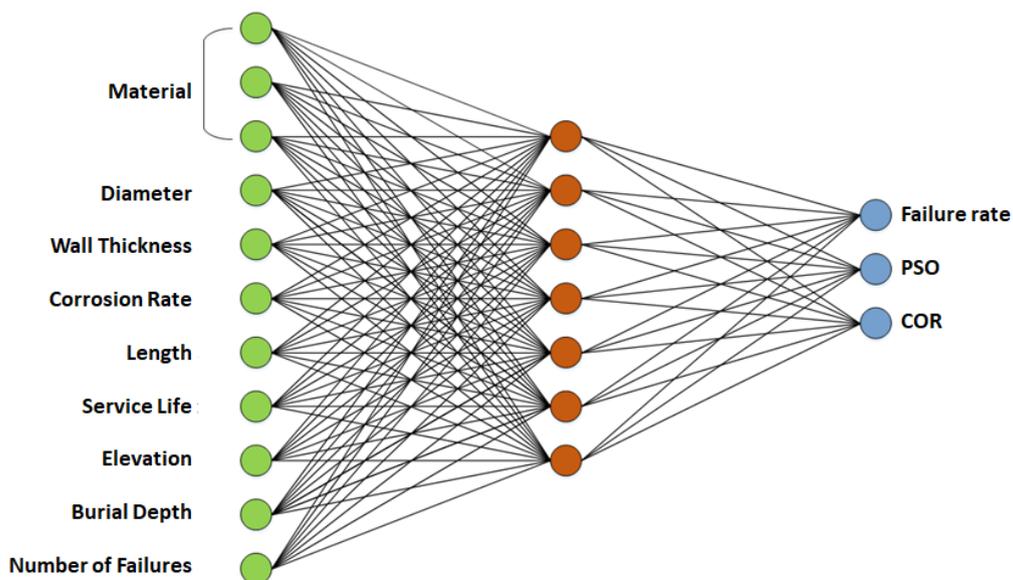


Figure 7. Architecture of the MLP 11-7-3 model

Experience gathered during this research permits following statements:

- The results of predicting the reliability of the Kamyshlov water supply network using the MLP 11-7-3 artificial neural network model confirm that ANN modeling is a useful tool for assessing its reliability.
- The developed model can be used in the practical operations of the Kamyshlov water supply organization to predict pipeline reliability based on three indicators (failure rate, probability of safe operation and coefficient of readiness) without the need to collect more complex data.
- Practical application of the obtained scientific results will allow for optimal planning of annual pipeline repairs and improve overall operational efficiency.
- When planning scheduled preventive maintenance (SPM) activities for hot water supply systems, it is necessary to determine the list of assets subject to repair based on their

reliability indicators.

- Monitoring pipeline reliability indicators allows for periodic recording of their actual reliability level and comparison with the existing technical condition, which helps identify the most vulnerable pipelines.

VI. Caveats of using generative artificial intelligence in infrastructure risk assessment

When using Generative AI (GAI) as a core analytical tool in infrastructures' risk and safety assessment and synthesis it is necessary to make the whole research process to be totally transparent, allowing for scrutinizing the prompts and raw outputs and scientific verification of the quantitative scores presented in the research. In other words, any GAI/XAI study must be held to the highest standard of rigor, and be foolproof against any way steered/biased prompt engineering.

A comprehensive research of infrastructure resilience and safety using GAI/XAI includes:

- all AI-related materials necessary for reproducibility, including: the exact model/software versions, the full prompt text(s) and engineering steps, configuration parameters, and the raw AI-generated outputs;
- description of the used inclusion/exclusion criteria.

A trustworthy verification process of any GAI /XAI research consists of: (1) Examination whether the AI outputs presented in the study underwent selective editing; (2) Ability to execute the identical prompts by independent researchers to assess reproducibility of the research.

In summary, any infrastructure research employing GAI /XAI should contain the complete AI prompts and the raw outputs used in the analysis for verification of the research results.

VII. Advantages, limitations and disadvantages of XAI

Advantages of using ANN:

- *Automation and speed:* ANN solves complex interdisciplinary problems in minutes, which is significantly faster than traditional methods that require manual programming and calculations.

- *Flexibility and adaptability:* ANN easily adapts the model to the requirements of the task, allowing for quick testing of various hypotheses without having to rewrite the code manually.

- *Accuracy and reliability:* ANN takes into account all industry and industry-wide standards, regulations, and norms, ensuring both the reliability of results and the minimization of human error.

- *Intuitive visualization:* ANN-generated operational risk graphs are not only technically accurate but also visually understandable, which is important for decision-making at the management and technical levels. Color-coding of risk zones, annotations, and scenario comparisons make conclusions clear even for *interdisciplinary* teams.

- *Economic impact:* ANN allows for optimization of infrastructure maintenance budgets and minimization of accident risks (through timely repairs).

- *Scalability:* ANN can be applied to both small-scale and large-scale infrastructure facilities.

Limitations and disadvantages of XAI (ANN):

- *Need for verification:* ANNs can offer incorrect or approximate technical solutions, so

each result requires expert review;

- *Limited understanding of context:* Without an adequate formulation and detailed description of the physical and operational aspects of the problem, ANNs can generate irrelevant answers;
- *Lack of understanding of physical meaning:* ANNs work with words and statistics, but do not understand the physics of processes, so engineering supervision is essential;
- *Does not replace in-depth engineering expertise:* AI is effective as an auxiliary tool, but cannot completely replace professional engineering analysis;
- *Technical limitations:* ANNs can make errors in mathematical calculations and coding, and cannot work with graphical data (drawings, diagrams). Complex problems must be broken down into smaller pieces;
- *False information:* if the initial data is insufficient, the ANN can invent information, passing it off as reliable, which is critical when working with facts, regulations, and citations;
- *Session restrictions:* ANNs store history within a single session. When the message limit is reached, a new chat must be created and the AI must be reminded of previous actions.

VIII. Conclusion

The existing regulatory framework of the Russian Federation does not fully take into account the specifics of XAI technologies to minimize the risks of operating infrastructure systems. To ensure the safe and effective implementation of AI, it is necessary to develop special regulations, establish requirements for the security expertise of AI systems, determine the order of responsibility allocation and create mechanisms to stimulate implementation of XAI. The latter must be not only reliable, but also scalable in order to adapt to changes in infrastructure and an increase in data volume.

The analysis of operational risks of key infrastructures using XAI methods is a promising direction that allows automating and greatly accelerating the algorithms of processing large amounts of data and modeling nonlinear dependencies, overcoming the limitations associated with the subjectivity of expert assessments, and optimizing risk management strategies.

Risk minimization methods include not only technical solutions, but also organizational changes aimed at improving the safety and efficiency of infrastructure operation. With the constant growth of safety and efficiency requirements, the use of XAI is becoming an integral part of this process, opening up new horizons for improving the reliability, resilience and safety of artificial nature. The use of XAI for monitoring the state of infrastructures should correspond to international standards.

The implementation of XAI is not without challenges. Highly qualified specialists (including prompt engineers) are needed who are able to work with new technologies. Therefore, an approach combining traditional methods and modern technologies seems to be the most effective. The transition from traditional methods to methods using XAI in managing infrastructure operation risk creates synergies arising from the combination of human experience and XAI capabilities, and opens up new horizons for improving their security. However, further research and testing is required to optimize these methods and eliminate existing shortcomings (like AI hallucinations).

Ultimately, the application of XAI in the field of risk analysis, safety and security can significantly reduce the risks associated with infrastructure operation in accordance with international standards, which in turn contributes to the sustainable development and enrichments of regions, municipalities and individual sectors of the economy and risk-oriented education of the society.

References

- [1] Timashev S.A. (2025). Infrastructures. Part III, Book 1. Infranetics. Riskology. Governance, Yekaterinburg : JSC "IPP "Ural worker". 751 p.
- [2] Timashev S.A. and Makeeva T.V. (2023). Reliability assessment of an urban water supply system under information deficit using artificial neural networks: Preprint, Yekaterinburg : Ural Federal University Press. 78 p.
- [3] Witt S. (2025). The AI Prompt that Could End the World. *The New York Times*, October 10.
- [4] Demchenko N.A. (2025). Assessment of reliability and safety of the Yakutsk power plant gas pipeline overpass in the condition of climate change: *Master Diss.*, scientific supervisor S.A. Timashev. Yekaterinburg, Institute of Civil Engineering and Architecture, Ural Federal University.
- [5] Akulshin A.O. (2025). AI-based methods of minimizing operational risk of oil and gas pipelines: *Master Diss.*, scientific supervisor S.A. Timashev – Yekaterinburg, Institute of Civil Engineering and Architecture, Ural Federal University.
- [6] Dzhumaev V.S. (2025). Assessment of subsea pipelines risk based on constructing their full life cycle: *Master Diss.*, scientific supervisor S.A. Timashev. Yekaterinburg, Institute of Civil Engineering and Architecture, Ural Federal University.