



## Гостьові лекції для здобувачів та викладачів факультету інформаційних технологій

05.12.2024



Recording

You are viewing Vyacheslav Kharchenko's screen View Options

Sign in View

Факультет інформаційних технологій  
Хмельницький національний університет

**Dependability and safety of AI systems:  
challenges and solutions**  
**Гарантоздатність та безпека систем  
штучного інтелекту: виклики та рішення**

**Vyacheslav Kharchenko**  
Department of Computer Systems, Networks and Cybersecurity,  
National Aerospace University KhAI, Kharkiv, Ukraine  
Centre for Infrastructure Safety Research and Analysis,  
Research and Production Company Radiy, Kropyvnytskyi, Ukraine  
e-mail: v.kharchenko@csn.khai.edu

Kharkiv - Khmelnytskyi, Ukraine, December 4, 2024

Unmute Start Video Participants 56 Chat Share Screen Record Show Captions Reactions Apps Whiteboards Notes Leave

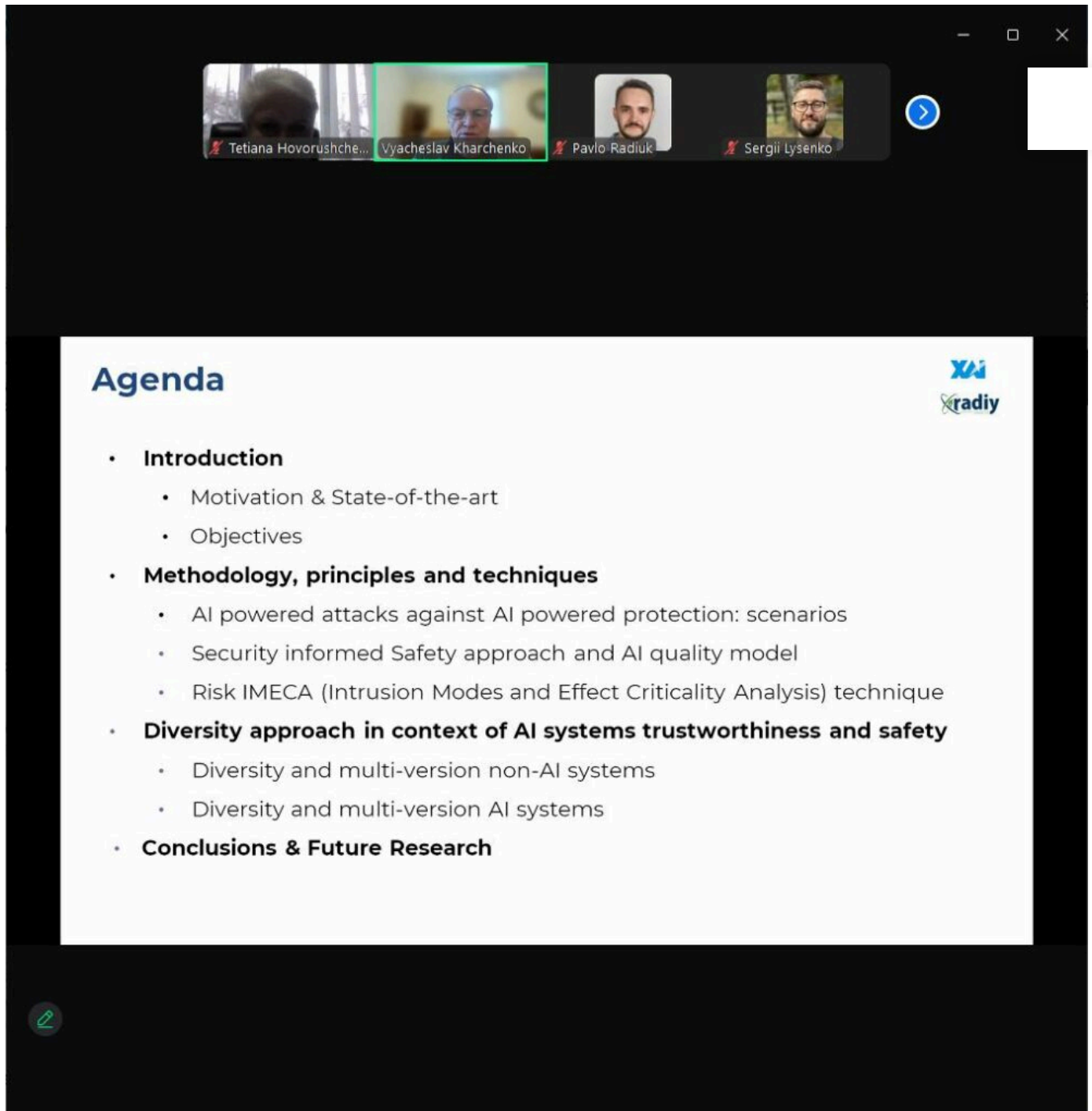
4 грудня 2024 року декан факультету інформаційних технологій Тетяна Говорущенко організувала онлайн-лекції для усіх бажаючих здобувачів освітніх програм факультету різних рівнів вищої освіти із запрошеними гостьовими спікерами. Долучались до лекцій також і зацікавлені викладачі факультету.

Гостьовими спікерами були:

– Вячеслав Харченко, д-р техн. наук, професор, заслужений винахідник України, лауреат Державної премії України у галузі науки і техніки, завідувач кафедри комп'ютерних систем, мереж і кібербезпеки Національного аерокосмічного університету «ХАІ», директор НТЦ аналізу та дослідження безпеки інфраструктур НВП «Радій», який розповів слухачам про виклики та рішення у сфері гарантоздатності та безпеки систем штучного інтелекту;

– Олена Голембовська, українська художниця, авторка наукових праць з інтеграції інформаційних технологій та сучасного мистецтва, картини якої виставлялися в Україні, Лондоні, в бойових частинах ЗСУ, яка розповіла слухачам про такі технології інтерактивного мистецтва, як доповнена реальність та штучний інтелект.

Щиро дякуємо нашим неперевершеним спікерам за цікаві та змістовні лекції!



The screenshot shows a Zoom meeting interface. At the top, there is a video gallery with four participants: Tetiana Hovorushche..., Vyacheslav Kharchenko (highlighted with a green border), Pavlo Radiuk, and Sergii Lysenko. Below the gallery is a slide titled "Agenda" with the following content:

**Agenda**


- **Introduction**
  - Motivation & State-of-the-art
  - Objectives
- **Methodology, principles and techniques**
  - AI powered attacks against AI powered protection: scenarios
  - Security informed Safety approach and AI quality model
  - Risk IMECA (Intrusion Modes and Effect Criticality Analysis) technique
- **Diversity approach in context of AI systems trustworthiness and safety**
  - Diversity and multi-version non-AI systems
  - Diversity and multi-version AI systems
- **Conclusions & Future Research**

In the top right corner of the slide, there is a logo for "XAI radiy".

zoom Workplace Meeting VK Vyacheslav Kharchenko's scr Recording... View

Tetiana Hovorushche... Vyacheslav Kharchenko Pavlo Radiuk Sergii Lysenko

## Motivation (1)




**Stanislaw Lem** (one of the most respected thinkers and science fiction writers of the 20th century, 1921-2006):  
**"Civilization is a function of technologies..."**

We add (we would like) that:  
**Civilization is a function of *safe/secure* technologies...**

Era of Artificial Intelligence:  
**Civilization is a function of *intelligent* technologies which have to be *trustworthy/explainable/safe/secure*...**

**TOP 10 AI FAILURES**



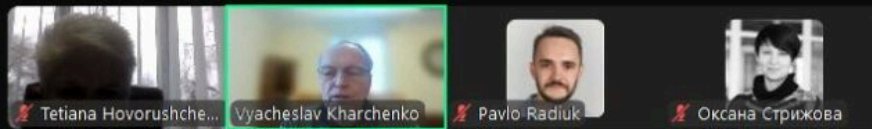
AI to fight cancer could kill patients  
 AI algorithms identify everything but COVID-19  
 An AI-inflected loss that resulted in a lawsuit  
 ChatGPT hallucinates court cases  
 Cruise recalls autonomous vehicles after crash...

**10 famous AI disasters**

Feature  
 Apr 17, 2024 - 13 mins

<https://www.immuniweb.com/blog/top-10-failures-of-ai.html> <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>

Audio Video Participants 68 Chat React Share Host tools Pause/stop recording More End



## Goal and Objectives



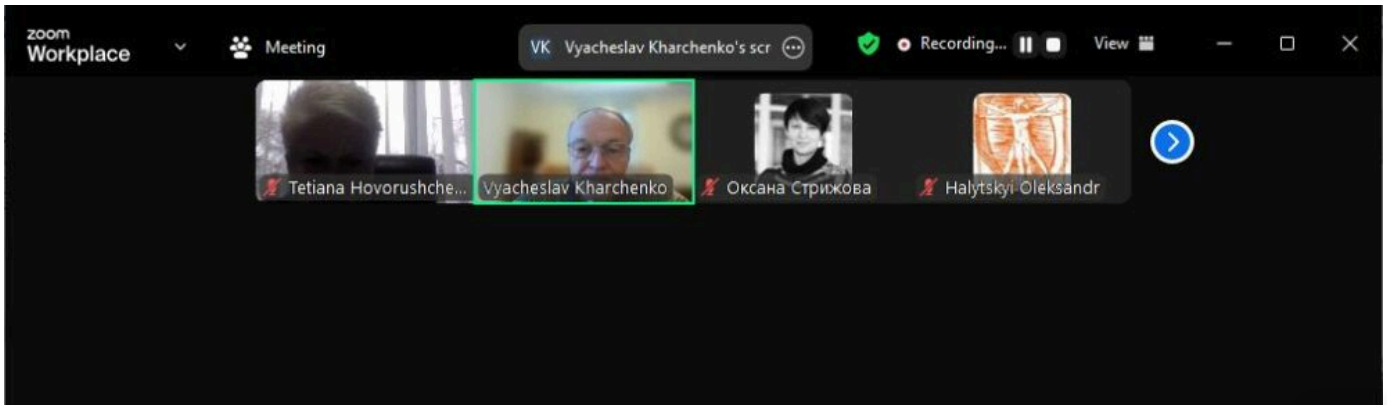
### Goal:

- to analyze possibilities of application von Neumann's paradigm (VNP) and VNP-based solutions to improve trustworthiness and other characteristics of AI systems

### Objectives:

- to discuss AI quality models\*, their characteristics and sub-characteristics to determine which of them and how can be improved by use of VNP-based approach;
- to analyze stages of VNP evolution to justify possible options for implementation of the paradigm for providing trustworthiness and other AI characteristics;
- to develop models of trustworthy and safe AI systems which are based on application of diversity principle or version redundancy (VR) on creating redundant channels and implement VNP using such approach.

\* Kharchenko V., Fesenko H., Illiashenko O. Quality Models for Artificial Intelligence Systems: Characteristic-Based Approach, Development, Application. *Sensors*, 2022, 22. P.1-32.

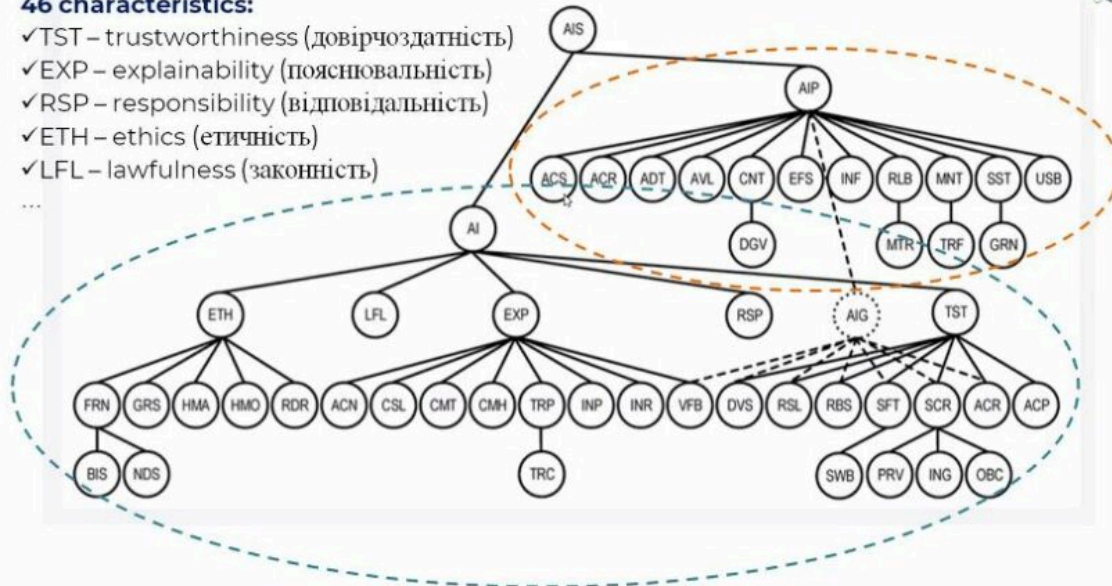


## Quality model of AI system (model and platform)

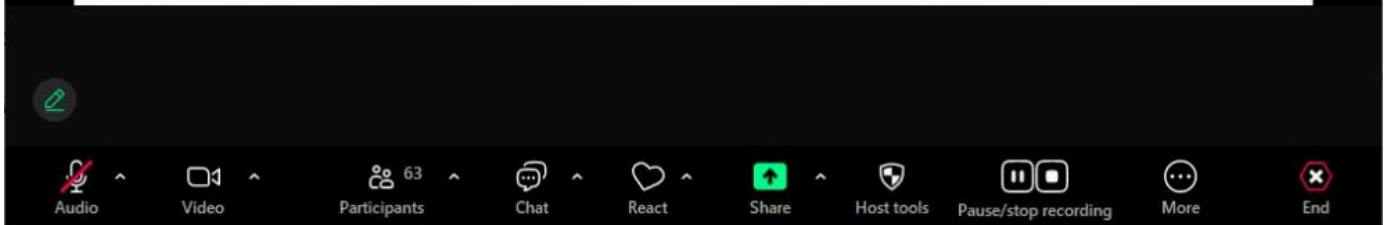


### 46 characteristics:

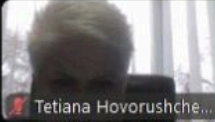


- ✓TST – trustworthiness (довірчоздатність)
- ✓EXP – explainability (пояснювальність)
- ✓RSP – responsibility (відповідальність)
- ✓ETH – ethics (етичність)
- ✓LFL – lawfulness (законність)





Kharchenko V., Fesenko H., Illiashenko O. Quality Models for Artificial Intelligence Systems: Characteristic-Based Approach, Development, Application. *Sensors*, 2022, 22. P.1-32.





zoom Workplace
Meeting
VK Vyacheslav Kharchenko's scr
Recording...
View







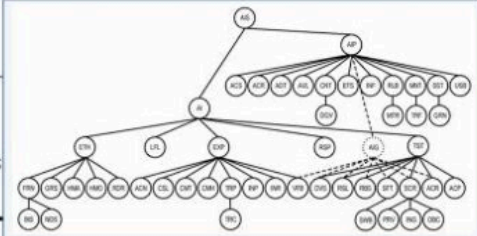


## Quality model of AI system (2)

**Model of AI quality (simplified option)**

Characteristics	Definition	Subcharacteristics
Ethics, ETH	The ability of AI to meet current standards of morality on the results of functioning	Fairness, FRN; <u>Graspability</u> , GRS; Human agency, HMA; Redress, RDR
Lawfulness, LFL	Ability of AI to comply with laws and regulations	No
<u>Explainability</u> , EXP	The ability of AI to be understood and predictable in terms of purpose and behavior	Completeness, CMT; Comprehensibility, MH; Interpretability, INP; Interactivity, INR; Transparency, TRP; Verifiability, VFB
<u>Responsibility</u> , RSP	Ability of AI to function considering the expectations of the client (user) in accordance with ethical norms, legal regulations, as well as to inform him in case of possible violation	No
<u>Trustworthiness</u> , TST	Ability of AI, characterized by the degree of confidence of the stakeholders, developers, auditors, etc.) that the AI meets and performs its functions in a predictable manner	Accuracy, ACR; Diversity, DVS; Resilience, RSL; Robustness, RBS; Safety, SFT; Security, SCR

V Kharchenko, H Fesenko, O Illiashenko. [Basic model of non-functional characteristics for assessment of artificial intelligence quality](#) // Radioelectronic and Computer Systems, 2022, 2, pp.131-144



Audio
Video
Participants 62
Chat
React
Share
Host tools
Pause/stop recording
More
End

zoom Workplace Meeting VK Vyacheslav Kharchenko's scr Recording... View

Tetiana Hovorushche... Vyacheslav Kharchenko Artem Kachur Красильников... Красильников Сергій

### Era of AI-based systems = Era of AI related safety risks?

Tesla driverless car crash...

<https://focus.ua/uk/world/536845-ne-srabortali-tormoza-v-kitae-tesla-ustroila-smertelnoe-dtp-na-bolshoy-skorosti-video>

Audio Video Participants 62 Chat React Share Host tools Pause/stop recording More End

zoom Workplace Meeting VK Vyacheslav Kharchenko's scr Recording... View

Tetiana Hovorushche... Vyacheslav Kharchenko Artem Kachur Красильников...

## Diversity-based Solution for AI: Industrial Perspectives

radiy

### Diversity Principle: industrial systems (AI application?)

#### Classification of diversity (version redundancy) NUREG 7007 Std, 2009

V. Kharchenko, M. Yastrebenetsky, Comparativistics for Big Safety: Hegel's Dialectic-based Methodology and Critical Applications Experience, 2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2023, pp. 1-8

Kinds of diversity (standards NUREG 7007, IEC26262, EN50128...)	Critical application domains and systems										
	Space		Auto-motive	Aviation		Rail-way	Chemical industry	Defense	NPP		IT (cloud+) infra-structures
	Shuttle	International Space Station	Control Systems (ASIL D)	Airbus A380	Boeing 777	Signaling systems	Centers for Chemical Process Safety	Military ICS	Reactor Trip Systems	Engineered Safety Features Actuation Systems	Globally distributed DBs
	SS1	SS2	MS1	AS1	AS2	RS1	CS1	MS1	NS1	NS2	DS1
Design	DD										
Equipment Manufacturer	ED										
Logic processing equipment	PD										
Functional	FD										
Life-cycle (Human)	HD										
Logic	LD										
Signal	SD										
Software	WD										
Others	OD										



zoom Workplace Meeting Olena's screen Recording... View

Tetiana Hovorushche... Olena Vadym Hurskiy Mukyliak

### 4.2.2 Сценарій «AR за групою глядачів»

The diagram illustrates a workflow for an AR application. It starts with an **Artist** who provides **Titles** to a system. This system generates **Possible titles** for **Visitors**. The system also manages **Video storage** and a **Video adaptation service**. The **Video adaptation service** interacts with a **mobile device** (labeled **Visitor**) to provide **Titles** for **Painting**. The **mobile device** also receives **Titles** from the **Video adaptation service** and the **Artist**.


Windows taskbar: Поиск, 100%, 13:42, 04.12.2024


Zoom controls: Audio, Video, Participants (42), Chat, React, Share, Host tools, Pause/stop recording, More, End

zoom Workplace Meeting Olena's screen Recording... View

Tetiana Hovorushche... Vyacheslav Kharchenko Olena Artem Kachur

# ТЕХНОЛОГІЇ ІНТЕРАКТИВНОГО МИСТЕЦТВА: AR & AI

 Олена Голембовська  
художник

 Вячеслав Харченко  
Національний аерокосмічний університет «ХАІ»

Audio Video Participants 49 Chat React Share Host tools Pause/stop recording More End

*Інформація факультету інформаційних технологій*

Загальні питання: [centr@khnmu.edu.ua](mailto:centr@khnmu.edu.ua)  
Подача новин та анонсів: [press@khnmu.edu.ua](mailto:press@khnmu.edu.ua)

Центр кар'єри

Скринька довіри

Цивільний захист

Пожежна безпека

Охорона праці



Хмельницький національний університет, 2025