

iResilience: Strengthening democratic practises and media literacy

SI. Funded by
Swedish Institute



TRAINING COURSE

«Dealing with propaganda, misinformation and fake news using AI&ML»



Topic:

Opportunities, Challenges, and Threats of Using Generative AI in Creating Fakes

Speaker

Volodymyr Skitsko



Candidate of Sciences (Economics), Associate Professor, Associate Professor of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after Vadym Hetman, Ukraine

Professional Membership of ISACA since 2013

Head of the section "Systems and technologies of artificial intelligence. Cybersecurity" of the Kyiv Small Academy of Science

key areas of research:

LIFE 3.0 INDUSTRY 4.0 SOCIETY 5.0

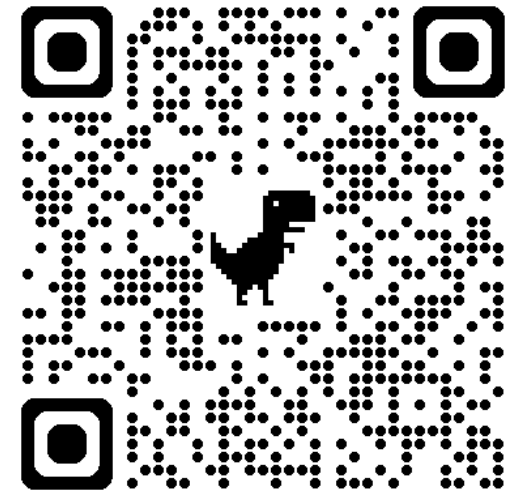
DIGITAL ECONOMY AI



<https://skitsko4.wordpress.com/>



skitsko@kneu.edu.ua



light and dark side generative AI

Every technology has its light and dark side. Generative artificial intelligence is no exception.

We have been using generative AI on a daily basis since 30 November 2022, when ChatGPT by OpenAI was launched. Since then, people have been actively using the capabilities of generative AI for various tasks, both with good intentions and not.

Today, we will talk about the dark side of generative AI, the negative aspects of using generative AI, and the challenges we face when using generative AI.

Threats of using generative AI

The problem of inaccurate results in the answers of generative AI.

Often, users perceive the answer provided by generative AI as truth, as encyclopaedic knowledge. This is especially true when a person asks a question about something in which they are not an expert or do not have a sufficient level of knowledge. In this case, such a person cannot notice inaccuracies or fake information. Often, in this case, a situation of AI hallucination arises, when the answer received looks quite truthful, but in fact it is not.

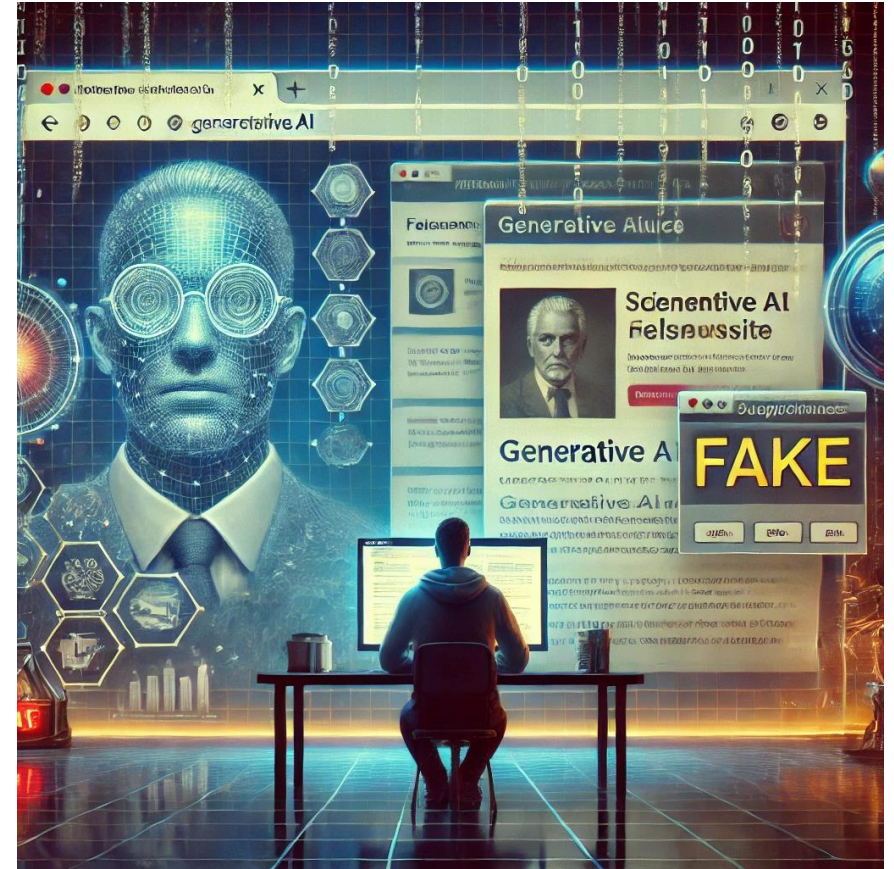
This problem can be used to spread disinformation, influence public opinion, etc.

Previously, the generative AI model provided an answer solely based on its own dataset and this problem was under control, but now generative AI can also use Internet searches to generate such an answer, and this leads to an increase in the number of possible inaccuracies.

The problem of inaccurate results in the answers of generative AI.

Case Study 1.

For example, a website is created with fake information, fake scientific articles, or even a full-fledged website of a fake journal. The generative AI model may take information from such a website when forming its response, and you will receive a text containing fake information. The person may not even doubt the veracity of the information, which is actually fake, as the generative AI response will contain a link to an existing website that looks quite normal at first glance. And in this case, the person will not consciously generate further fakes.



Bias

Generative AI models create new content based on existing content previously created by humans. Therefore, the bias of AI essentially reflects the bias of people, those stereotypes that are widespread in society. Bias can be used to create fakes and be subject to manipulation.

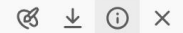
Case Study 2.1



Here is the image of a detailed and realistic apple. Let me know if there are any adjustments you'd like!



Create an image of an apple.



Запит

A detailed and realistic depiction of a fresh, glossy red apple with a green leaf attached to its stem, placed on a light wooden surface, softly lit to highlight its texture. The background is blurred with subtle earthy tones to emphasize the apple's vivid color.

Копіювати

Bias

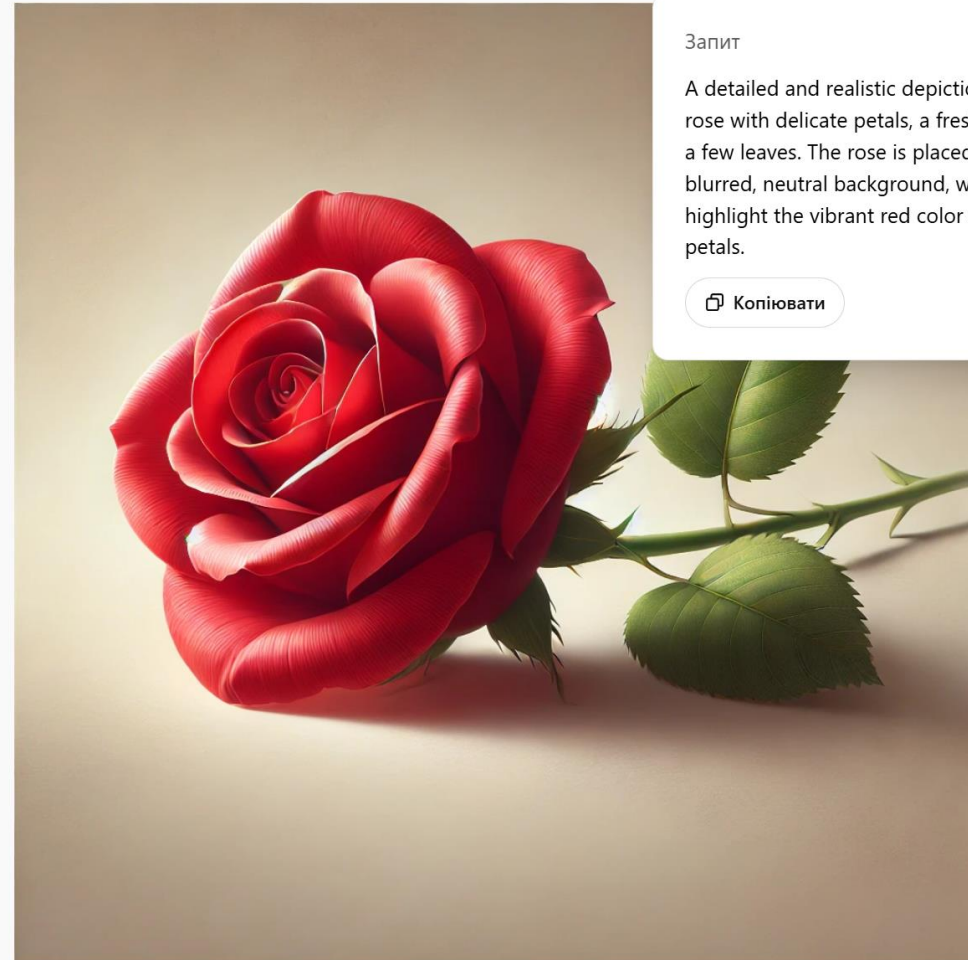
Case Study 2.2

Create an image of a rose.

< 2/2 >



Here is the image of a detailed and realistic rose. Let me know if you'd like any modifications!



Запит

A detailed and realistic depiction of a single red rose with delicate petals, a fresh green stem, and a few leaves. The rose is placed against a softly blurred, neutral background, with soft lighting to highlight the vibrant red color and texture of the petals.

Копіювати

Bias

Case Study 2.3

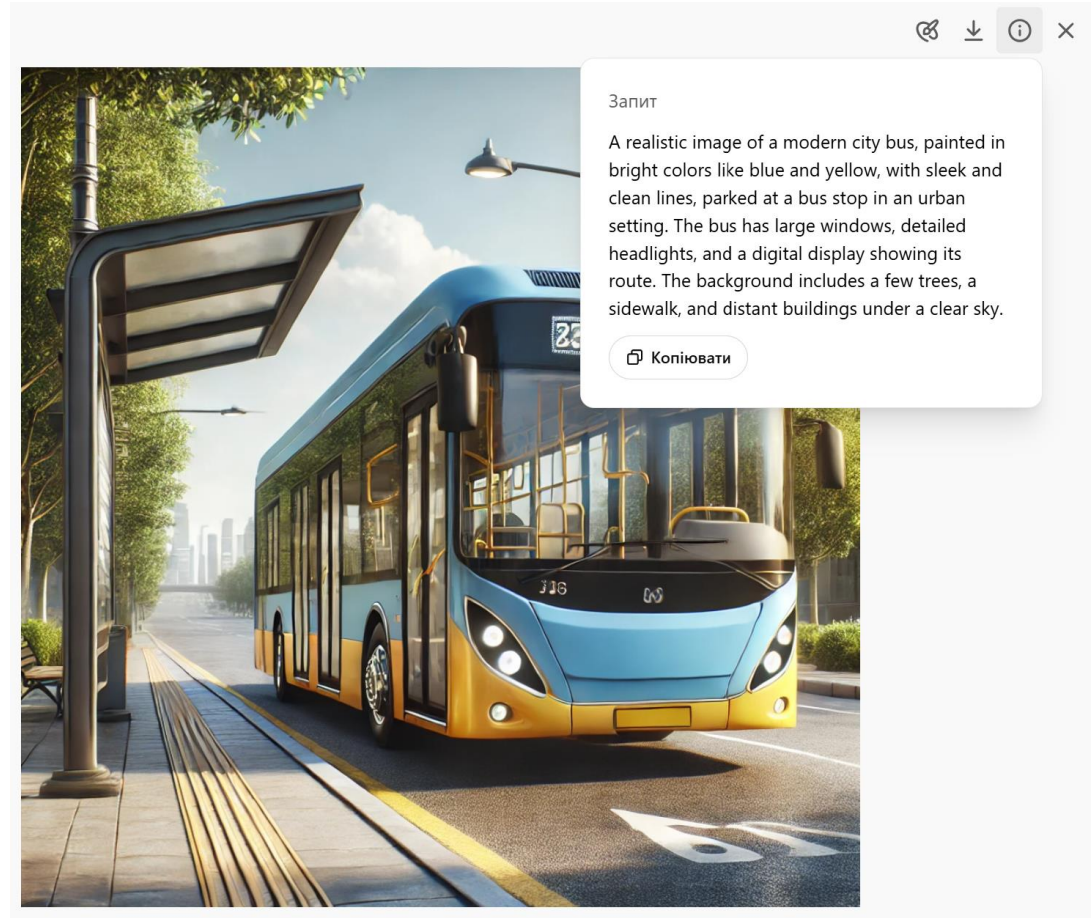


Here is the image of a modern city bus. Let me know if you'd like to modify any details!



Create an image of a bus.

< 4/4 >



Bias

Case Study 2.3

I'm English. Create an image of a bus.

< 5/5 >



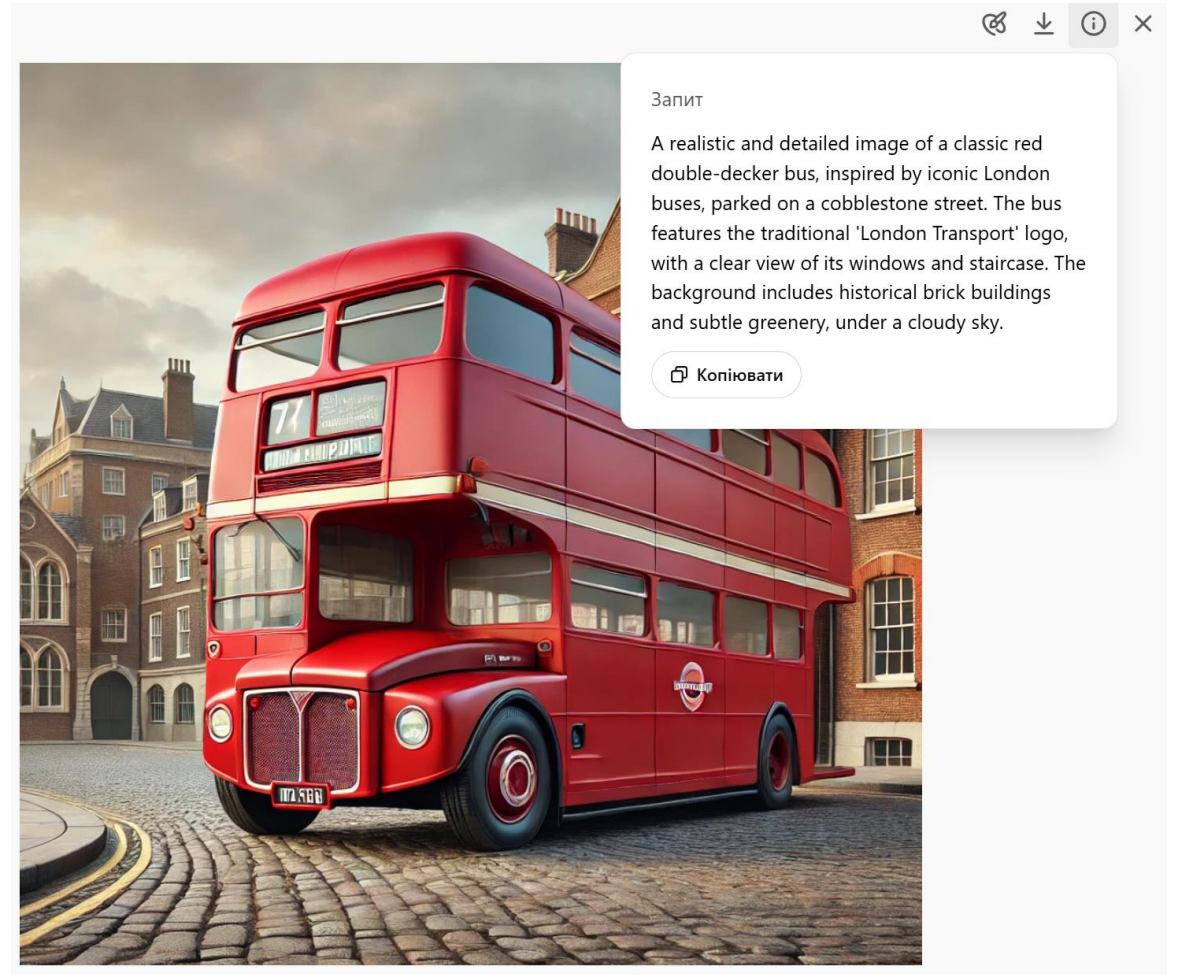
Here is the image of a classic red double-decker bus. Let me know if you'd like any changes or have additional requests!



I'm English. Create an image of a bus.



Here is the image of a modern double-decker bus in red, inspired by the iconic London buses.



Запит

A realistic and detailed image of a classic red double-decker bus, inspired by iconic London buses, parked on a cobblestone street. The bus features the traditional 'London Transport' logo, with a clear view of its windows and staircase. The background includes historical brick buildings and subtle greenery, under a cloudy sky.

Копіювати

Bias

Case Study 2.3

I'm English. Create an image of a bus.

◆ Sure, here is an image of a bus:



I'm English. Create an image of a bus.



Повідомлення з Copilot

vs I'm English. Create an image of a bus.



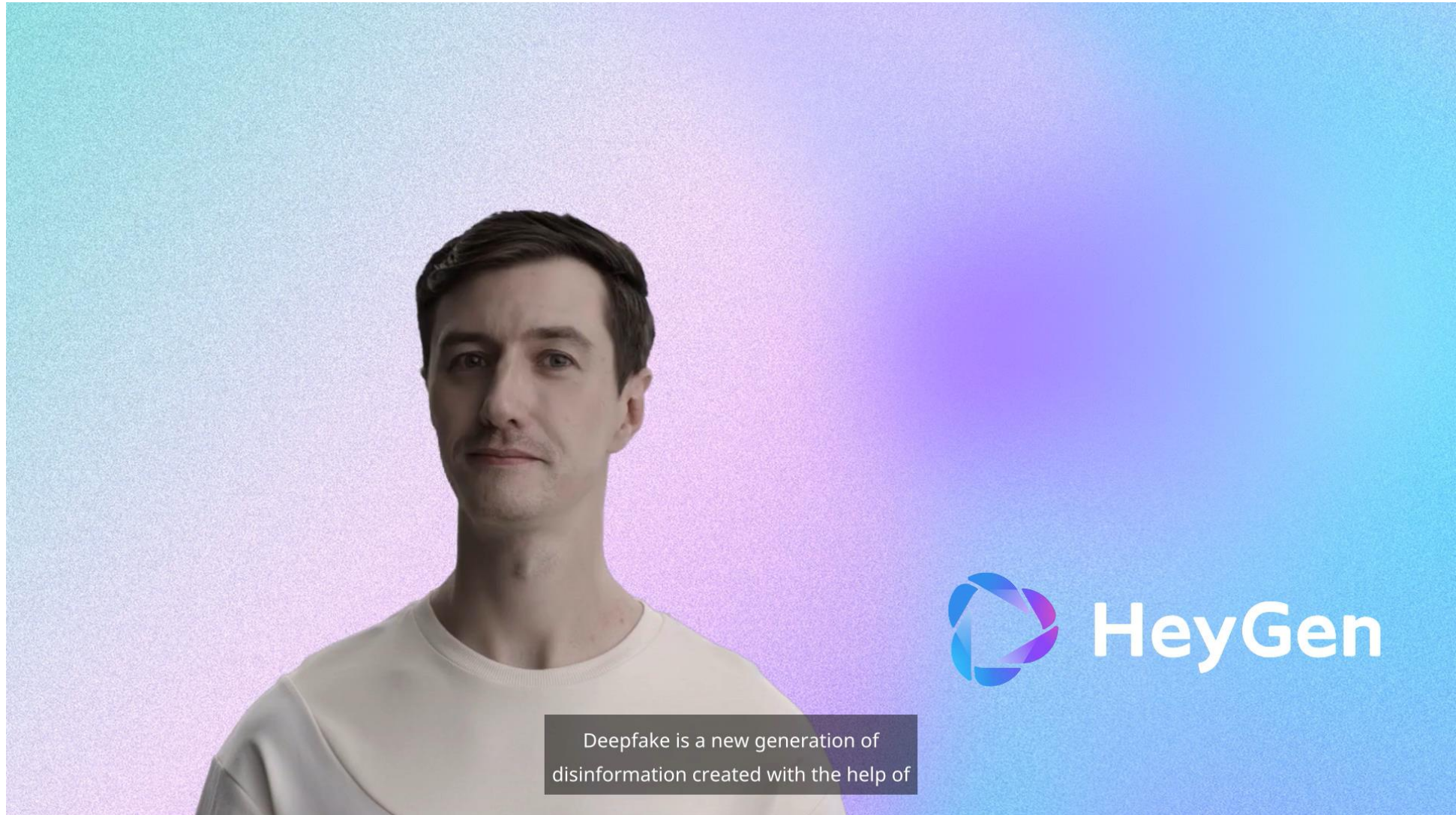
Worked for 3s



09:49

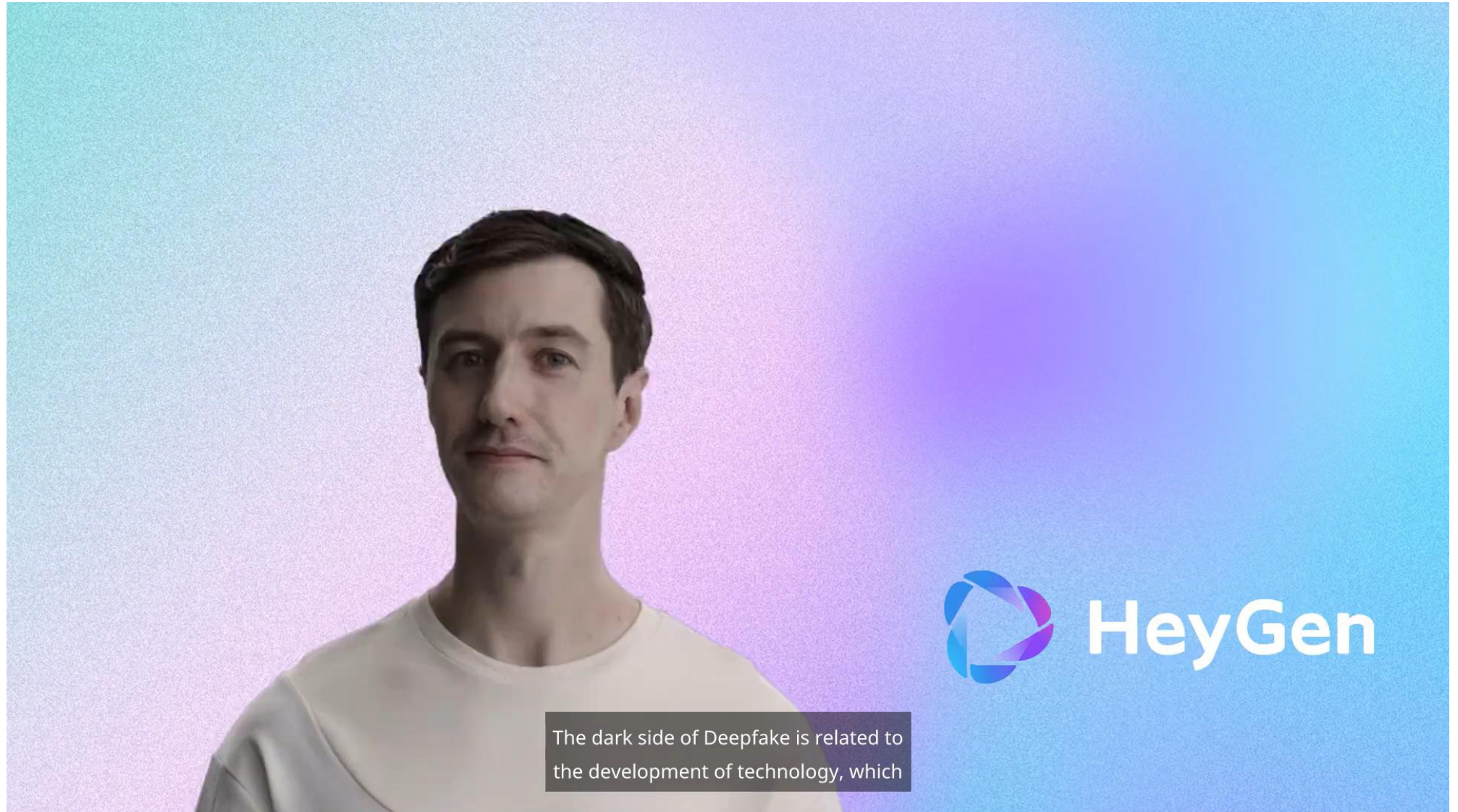


Deepfake



Deepfake is a new generation of disinformation created with the help of

Deepfake



The dark side of Deepfake is related to the development of technology, which

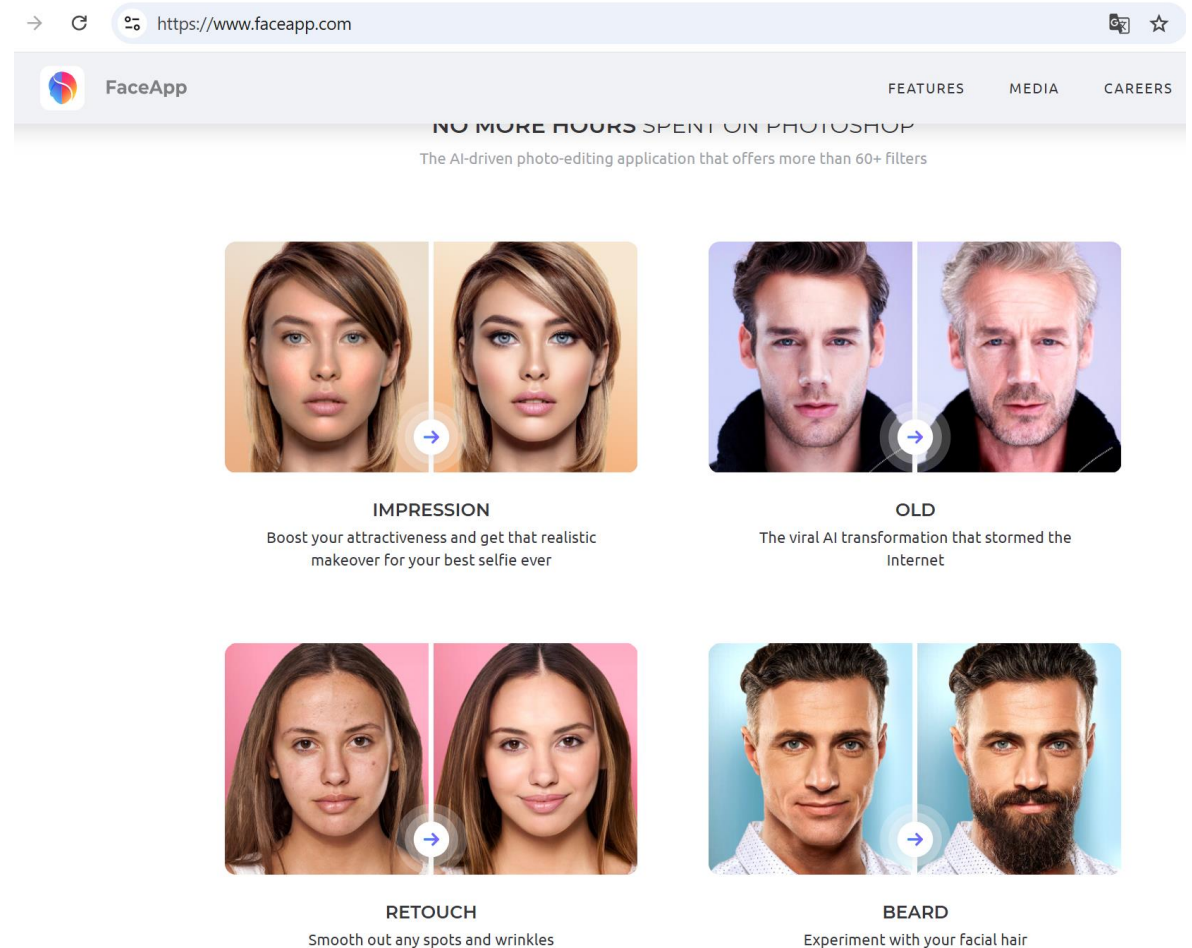
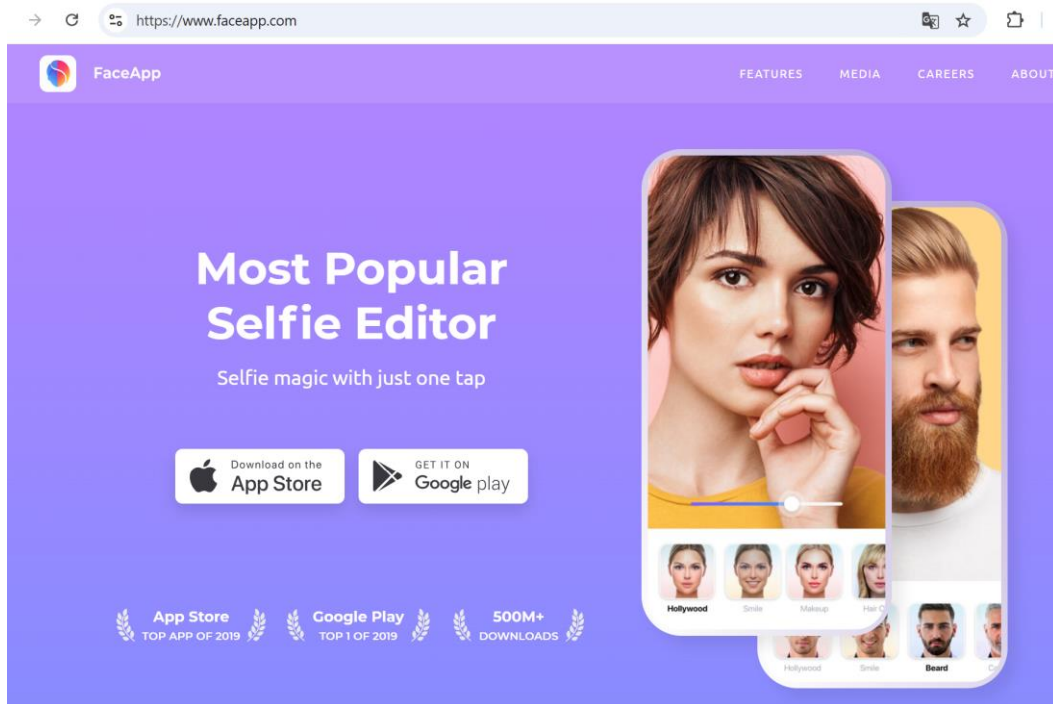


Deepfake

Case Study 3.1

Creating a Deepfake is becoming easier and easier, and there are more and more tools.

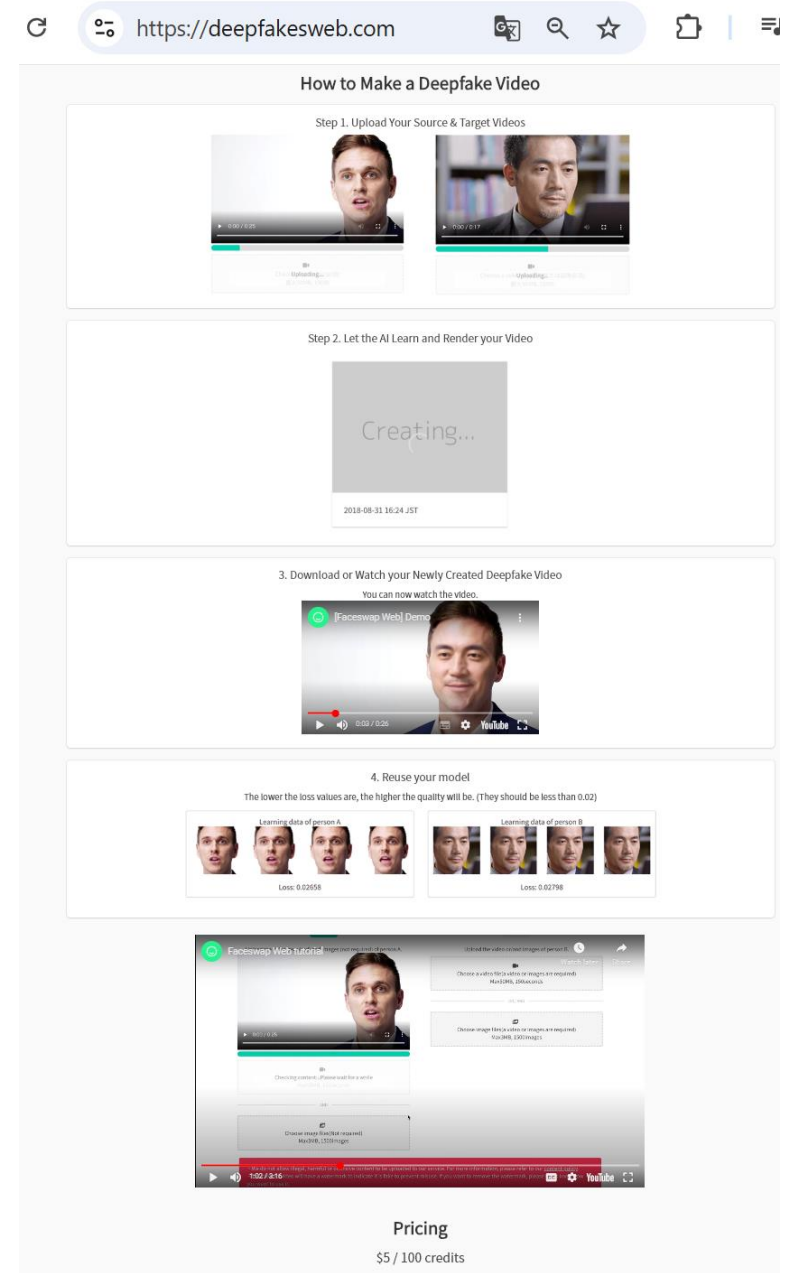
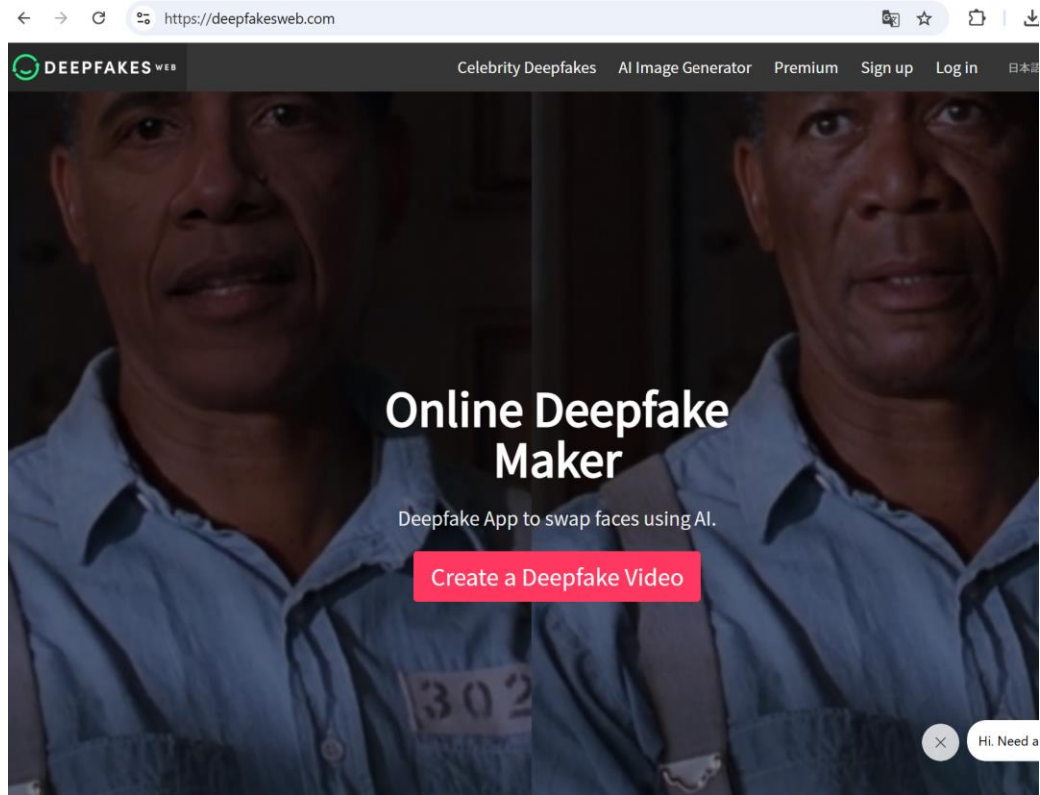
FaceApp



Deepfake

Case Study 3.1

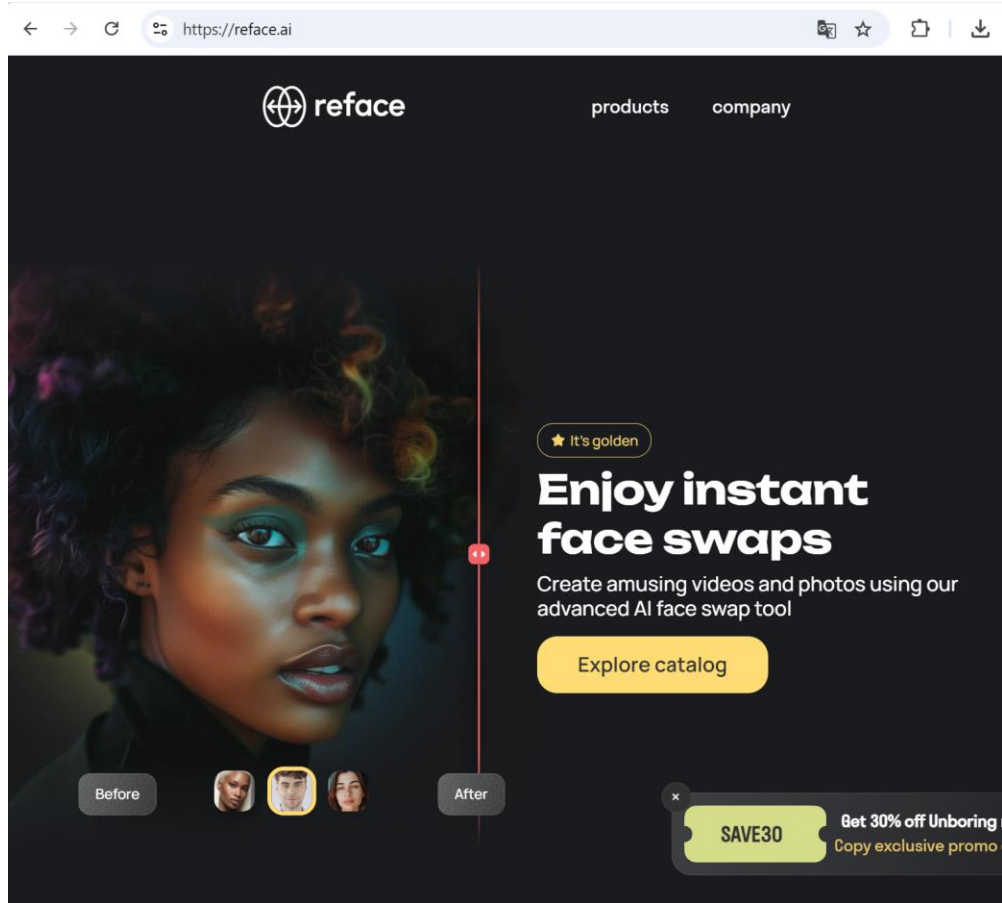
DeepfakeWEB



Deepfake

Case Study 3.1

reface



← → ↻ https://reface.ai

reface products company

★ It's golden

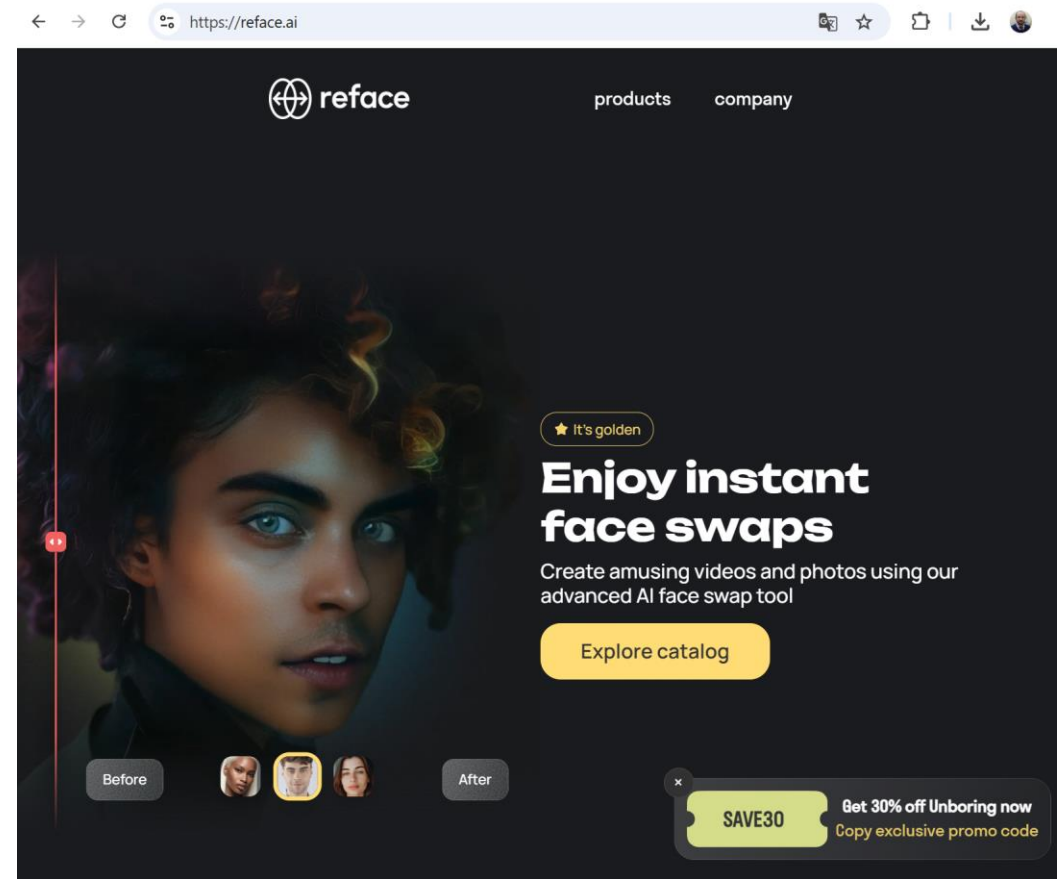
Enjoy instant face swaps

Create amusing videos and photos using our advanced AI face swap tool

Explore catalog

Before After

SAVE30 Get 30% off Unboring now Copy exclusive promo code



← → ↻ https://reface.ai

reface products company

★ It's golden

Enjoy instant face swaps

Create amusing videos and photos using our advanced AI face swap tool

Explore catalog

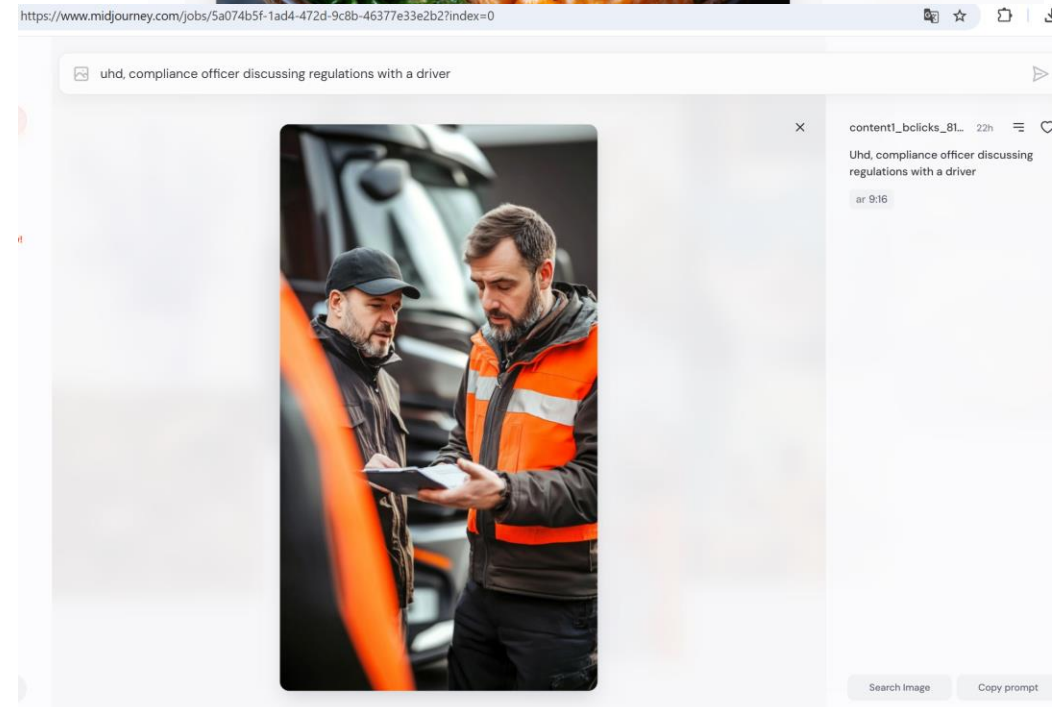
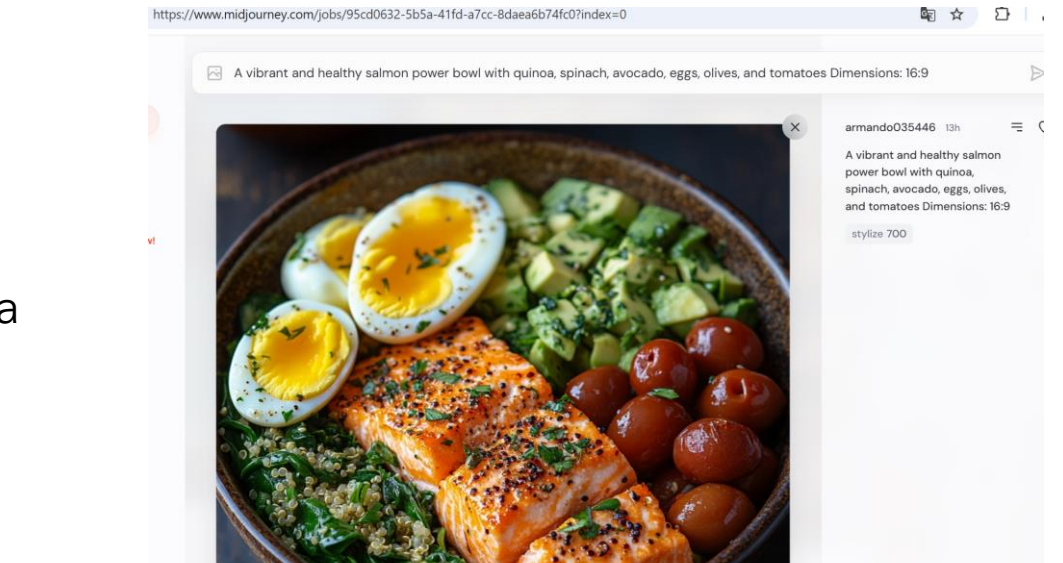
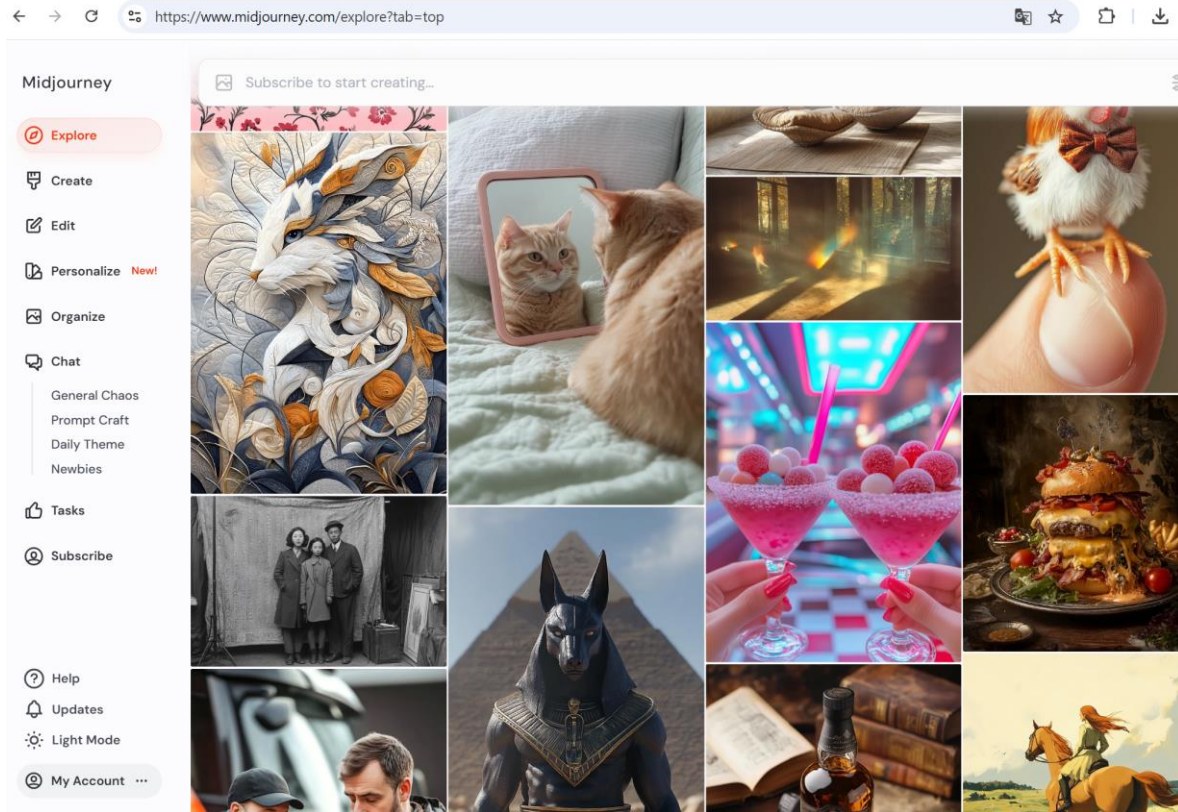
Before After

SAVE30 Get 30% off Unboring now Copy exclusive promo code

Deepfake

Case Study 3.2

Midjourney allows you to create photorealistic images at a text prompt.



Fact checking

Case Study 4.1

The easiest way to protect yourself from fakes is to use information only from verified sources, or from several sources, and check the facts.

Case Study 4.2

One of the ways to fight against deepfake audio can be the following advice - to formulate a code phrase that can be called if necessary and prove that your interlocutor is not a fake.



Fact checking

Case Study 4.3

To identify visual deepfake, you can use Google Images - image search. You can find where else the image has been used.

However, Google Images can be circumvented - it is enough to make a mirror image of a real photo and Google Images may not know the original source.

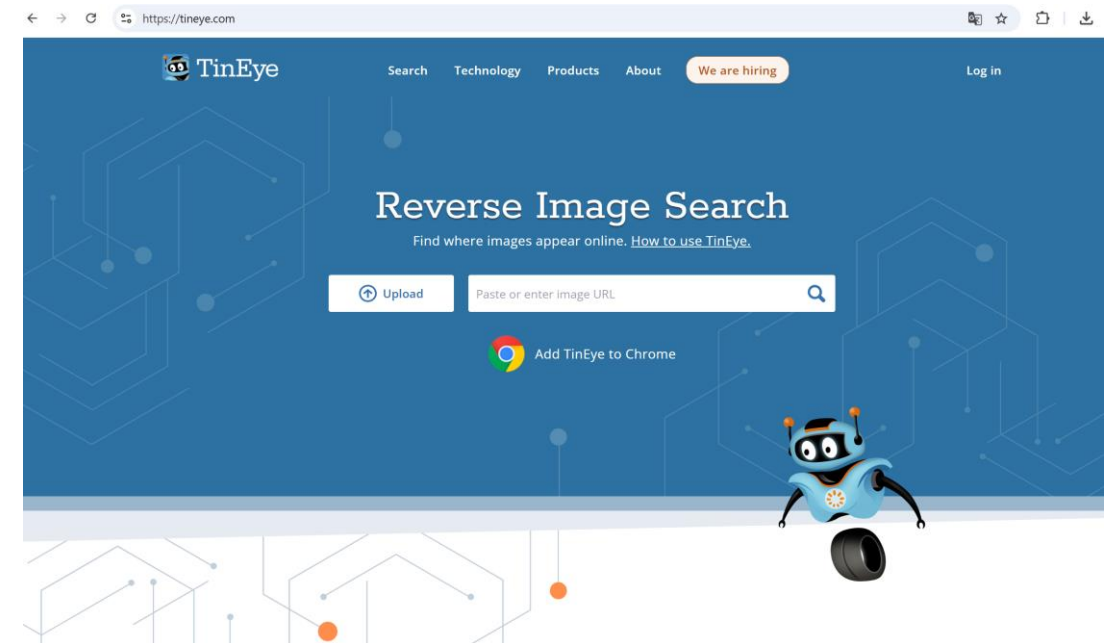


Case Study 4.4



TinEye helps not only to search for pages on which the image was previously published, but also to filter the results according to the following parameters:

- Date (latest or oldest);
- File size and weight;
- The number of publications on this or that site or domain.



Fact checking

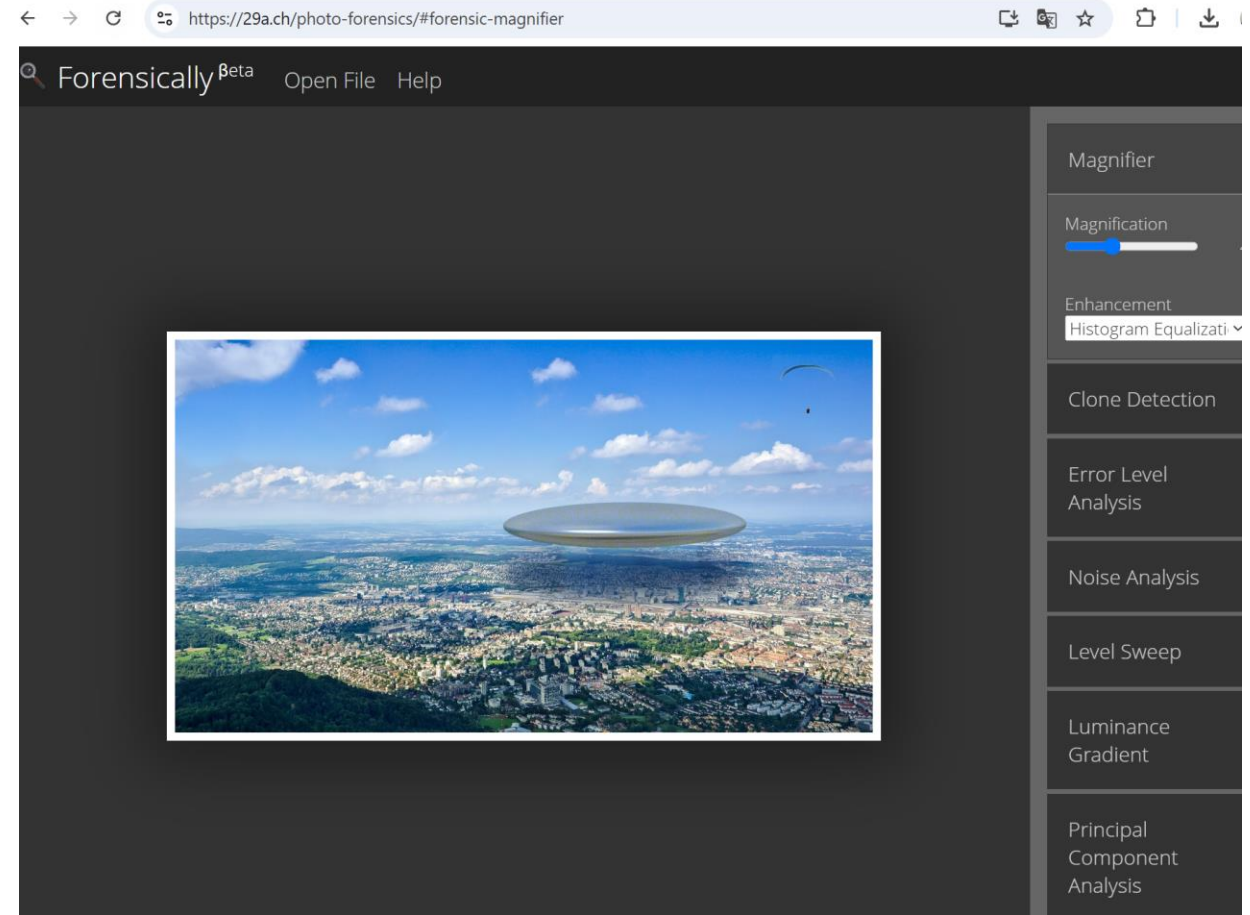
Case Study 4.4

For deeper image fact-checking, you can use the **Forensically** service.

Can perform Noise Analysis, Error Level Analysis, or Clone Detection on an image.

This way, you can spot photo editor tampering.

The program also shows key technical parameters, such as the time when the image was taken, the size and weight of the file, the characteristics of the camera used to capture the image, etc.



Fact checking

Case Study 4.5

When it comes to AI-generated videos, it's not easy to make a good deepfake.

Therefore, many fakes can be recognised through careful viewing and common sense.

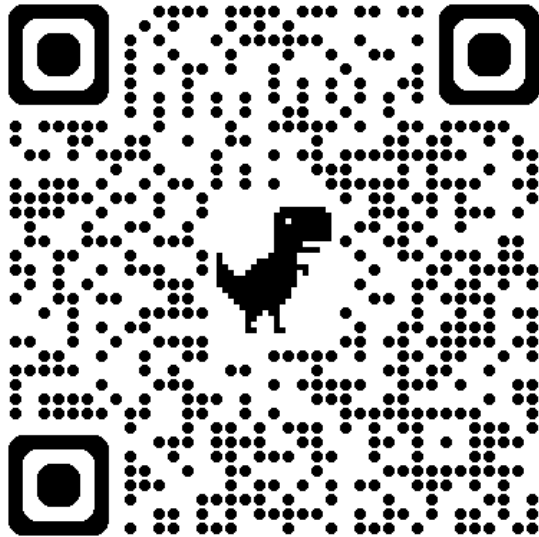
Critical and analytical thinking is today an effective tool for identifying fakes.



Trainer services that help you improve your fake detection skills.

Watch frame by frame

The Watch frame by frame site allows you to view videos frame by frame, which helps detect deepfakes faster

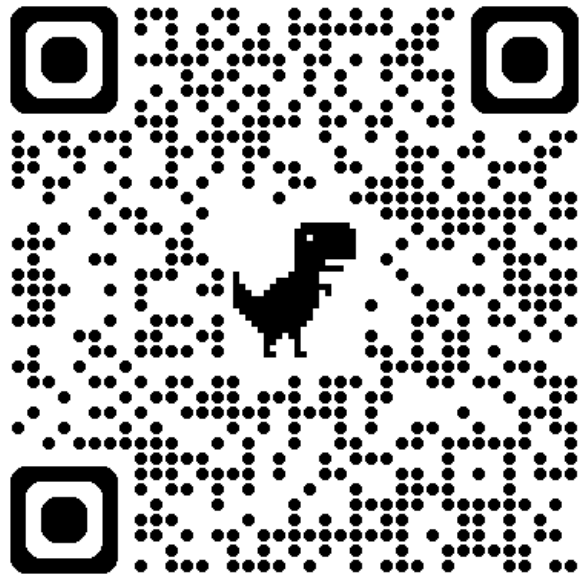


A screenshot of the website 'www.watchframebyframe.com'. The browser address bar shows the URL and a warning icon. The site has a navigation menu with 'About', 'Contact', and 'Donate'. The main content area features the text 'Watch YouTube and Vimeo videos frame by frame and in slow motion' above a search input field labeled 'ENTER VIDEO URL OR ID' and a blue 'WATCH VIDEO' button. Below this is a 'Featured' section with three video thumbnails: 'SpaceX - Static Fir...', '30 Days Timelaps... 4K', and a message 'Відео недоступне' (Video unavailable).

Trainer services that help you improve your fake detection skills.

DETECT FAKES

you can try yourself as a fake detector




→ ↻ <https://detectfakes.kellogg.northwestern.edu> 🔍 ☆ 📄 | ⬇

DETECT FAKES ABOUT CONSENT INSTRUCTIONS

AI-generated or Real?

Take a look at this image and share whether you think it is generated by AI or not, and how confident you are in this judgment. You have unlimited time to look at the image.



I have seen this before.

Real: This is a real image.

Fake: This is a synthetic image generated by AI.

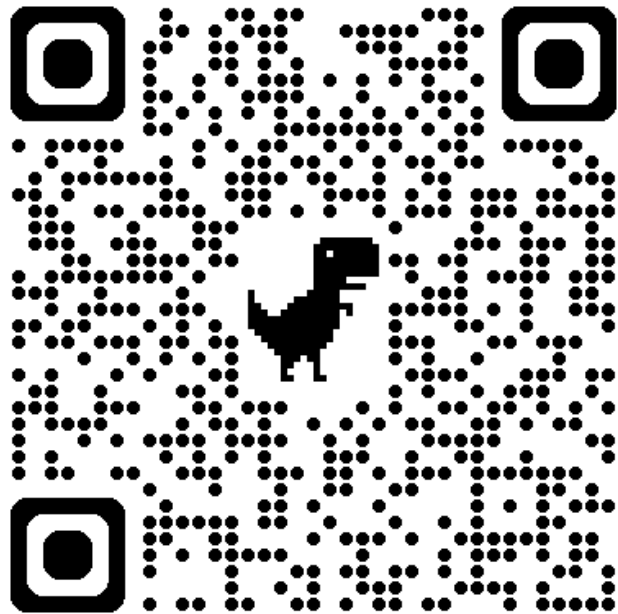
—●—

Slide the dot to share your confidence

Trainer services that help you improve your fake detection skills.


Sharpen your sense

training your visual and auditory skills to recognize fakes.



← → ↻ 🔍 <https://innovation.dw.com/sharpen-your-senses/index.html> 📄 ☆ 🗑️ ⬇️

Sharpen your senses and strengthen your verification skills

Share: 

Verification is not just about tools. Essential are our human senses. Whom can we trust, if not our own senses?


“The world is full of magic things, patiently waiting for our senses to grow sharper.” (W.B. Yeats)

We created a training for you, so that you are able to take full advantage of your senses.

In this training you will find exercises to sharpen your **vision** and **hearing** skills. Each activity has three exercises at three levels: easy, medium and advanced.

Doing these exercises will make you more confident to detect the almost (in)visible traces in synthetic and manipulated media like blurred spots, missing pieces or a change of voice emphasis and cut & paste manipulation within a speech.

Go full screen for the optimal experience




Active observing

Click on an exercise and train your eagle eyes.

Tip: Take a close look and gather the facts.

- Bubble Mix
- 20 Hints



Precise Listening

Click on an exercise and train your lynx ears.

Tip: Close your eyes.

- Voice Clone
- Critical Listening

Overall Progress: 0%

References

Соціальний вплив генеративного ШІ: переваги та загрози

<https://www.unite.ai/uk/social-impact-of-generative-ai-benefits-and-threats/>

Що таке **Deepfakes**?

<https://www.unite.ai/uk/що-таке-дипфейки/>

Фейки, шахрайство і дезінформація: загрози штучного інтелекту і як від них захиститись

<https://weukraine.tv/top/fejky-shahrajstvo-i-dezinformatsiya-zagrozy-shtuchnogo-intelektu-i-yak-vid-nyh-zahystytys/>

Маркування ШІ-згенерованого контенту: як уряди та компанії посилюють прозорість використання штучного інтелекту в медіа та соцмережах

<https://dslua.org/publications/markuvannia-shi-zghenerovanoho-kontentu-iak-uriady-ta-kompanii-posyliuiut-prozorist-vykorystannia-shtuchnoho-intelektu-v-media-ta-sotsmerezhakh/>

Speaker

Volodymyr Skitsko



Candidate of Sciences (Economics), Associate Professor, Associate Professor of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after Vadym Hetman, Ukraine

Professional Membership of ISACA since 2013

Head of the section "Systems and technologies of artificial intelligence. Cybersecurity" of the Kyiv Small Academy of Science

key areas of research:

LIFE 3.0 INDUSTRY 4.0 SOCIETY 5.0

DIGITAL ECONOMY AI



<https://skitsko4.wordpress.com/>



skitsko@kneu.edu.ua

