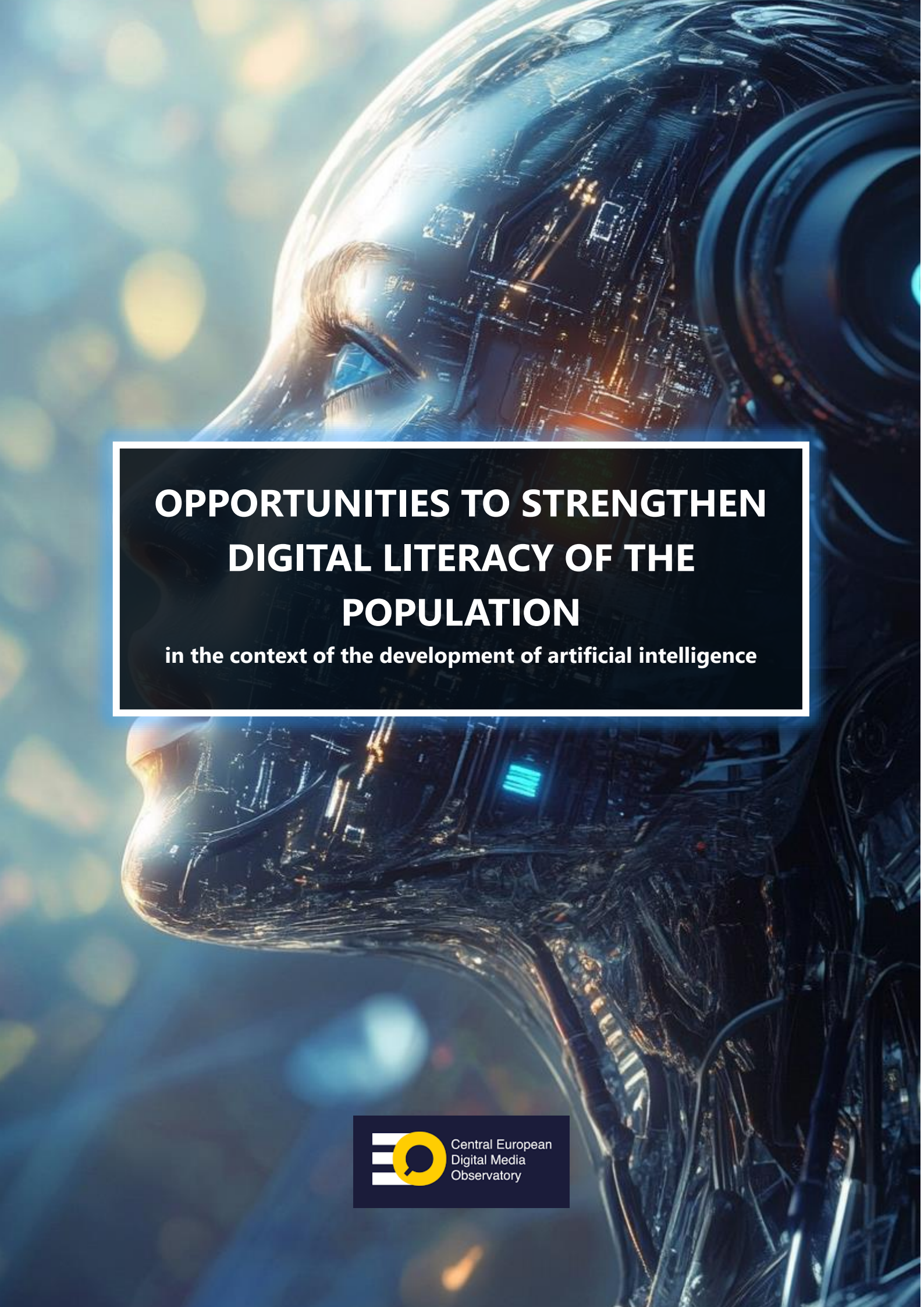


OPPORTUNITIES TO STRENGTHEN DIGITAL LITERACY OF THE POPULATION

in the context of the development of artificial intelligence

Central European Digital Media Observatory

Table of Contents

1 Introduction to the issue

In recent years, artificial intelligence (AI) has evolved dramatically - in the areas of large language models (LLMs), transformer and diffusion models, and other forms of AI - leading to a substantial increase in the availability of these tools to the general public. This has resulted in an increase in publicly available AI services, applications and tools that have begun to be used in both positive and negative ways.

The ability to generate arbitrary content has led to an increase in false news (hoaxes, misinformation, fake news) in text form, but there has also been an increase in manipulated graphics - photographs, photo collages, images - depicting existing people in non-existent contexts. Among the highly dangerous forms of graphic content, in particular, were so-called **deepnude** - generated photographs of people "undressed" using artificial intelligence tools. **Deepfake**[1] videos are a separate chapter.

1.1 Generating fraudulent messages with AI

The fundamental problem associated with the generation of false text messages using applications using LLM[2] is primarily the **speed** with which text messages can be created, their **variability**, which leads to the inability to detect them with common detection mechanisms (i.e. each generated message is different in terms of form, even if the content is preserved), and also **the low effectiveness of ethical boundaries** that are integrated into commonly used tools (ChatGPT). Although these tools are equipped with ethical constraints in terms of illegal content (drug issues, criminal offences, pornography, etc.), in the area of generating false news against a specific person or institution, for example, the ethical constraints are easily overcome. We describe the process of "overcoming boundaries" in detail in the publication *Risks associated with artificial intelligence*.[1] The problem is also the possibility of generating not only the message itself, but also the form of the post under which it should be published, e.g. on social media, in order to get the greatest possible response (impression, impact, clicks).

1.2 Generating underwater graphics with AI

Generative AI focused on the creation of graphic output has huge potential to reach a wide range of professions from photographers to graphic designers, artists, creatives, advertising agencies and the general public. In most cases the graphical output is and will be of a rather positive nature, but it will be - and of course already is - misused for **fraudulent** purposes. The generation of fraudulent graphics using artificial intelligence can present a number of significant risks and issues that have the potential to affect a wide range of society. Fake graphics generated by AI can be used to spread misinformation that can manipulate public opinion or influence the outcome of elections. Altered or entirely

[1] **Deepfake** is a technology that allows the creation of fake digital content (mostly videos) using machine learning algorithms that can realistically replace faces and voice tracks in existing videos, photos and/or audio recordings. In (not only) the Czech environment, the term is often simplified to video content only.

[2] **LLM** - large language models that are trained on human language (texts) and communicate in human language. They are able to analyze and generate text.

purpose-generated photos and videos can create **false 'evidence'** of events that never happened or distort reality, which can have serious impacts on democracy and public opinion. Fraudulent graphics can be used for **fraud, phishing or extortion**. A significant risk is the generation of highly realistic graphics that can reduce people's trust in the online content presented, which can lead to **the rejection of authentic photos or videos, the questioning of real events** and a reduced ability to distinguish between truth and lies. The potential economic impact of the fraudulent graphics generated cannot be ignored. The creation of false financial documents, contracts or receipts can have a significant **economic impact**, both for individuals and businesses. Fake graphics can be used to **circumvent control mechanisms** based on graphic output. **Deepfake** technology can be misused to create fake intimate or compromising graphics of private individuals, violating their privacy. Such material can be disseminated without consent and cause emotional distress, psychological distress, etc. to victims.

Applications for deepfake creation are widely available - for example, try CivAI's Deepfake sendbox (https://deepfake.civai.org/).

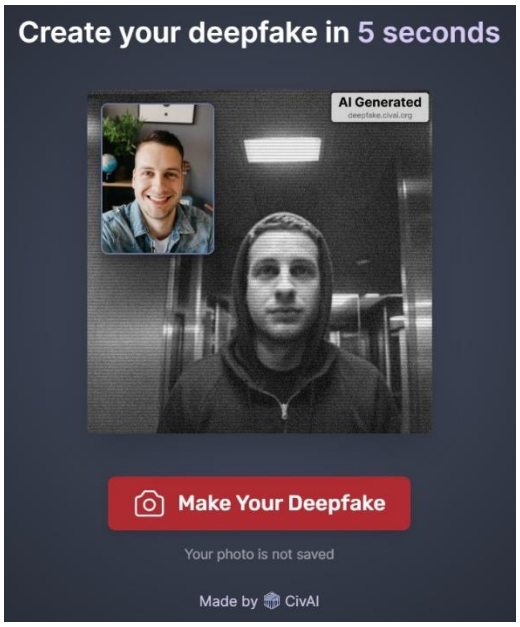


Figure 1. Home screen of the Deepfake sendbox application for a demonstration of deepfake creation (Source: CivAI)

1.3 Generating AI-assisted deception videos

Similar problems, such as generating deceptive graphics, can be caused by AI-assisted generation of deceptive videos. Generative AI systems and applications aimed at video creation are capable of producing very realistic output that is almost indistinguishable from real-life output. A number of AI generators are capable of **creating a video from any photo**, giving a completely different dimension to underwater behaviour. An example

would be animating (talking) a photo of any public figure, which can verbalize completely fictitious and in many cases dangerous facts.

Currently, **deepfake advertisements** referring to fraudulent investment offers are very topical, which we highlight in more detail in subsection 3.1.1 Deep fake videos and internet scams. Public figures such as politicians or artists are often used for this activity, which gives legitimacy to this fraudulent behaviour. **Face swap** videos, where an AI tool can be used to replace a face in a real video with another face, are also very risky.

2 Artificial intelligence and personal data

In relation to the positives and negatives of AI, it should be recalled that AI already works with a range of personal and sensitive data that we provide to it (which we are often not even aware of) and uses it in various ways (e.g. personalisation of search, precise targeting of advertising, AI-assisted HR processing, AI-assisted email, etc.). Therefore, certain privacy rules must be followed when using AI.

One of the key rules is **transparency** - users should know what data is collected, for what purpose and who has access to it. This is because often this is information that can be very sensitive, such as our online behaviour, personal preferences, location data or even biometric data. Transparent information and consent to data processing are the cornerstones of ensuring trust and respect for users' rights. In practice, this can be seen, for example, in the context of mobile phone applications requesting access to various device functions or websites collecting cookies from users. For generative AI, users should check that **their data is not being used to train AI** (most well-known tools have this feature in their settings).

Another important principle is to **minimise the data collected**. AI should only collect information that is necessary to achieve a given goal, and no more. Collecting unnecessary data not only increases the risk of misuse, but also opens the door to potential ethical problems if personal data is used in ways that users do not agree with.

Data security is another crucial area. The data collected must be protected from unauthorised access, leakage or misuse. This includes the implementation of technical measures such as encryption and pseudonymisation, as well as organisational measures including staff training and regular auditing of security procedures.

At this point, it bears repeating that AI tools often learn from the data we feed into them, which logically can include personal data. Therefore, **we strongly advise against entering** any **data of a personal nature into AI tools** - especially their free/public versions - **unless you are sure what is happening with it**. Companies like OpenAI or Microsoft have settings built inside their tools where the AI will not learn from user data (it will not use it for training).

The problem may arise in a situation when we are tasked to prepare a review of a grant project - most state grant agencies (TAČR, etc.) explicitly prohibit the use of AI because there is a risk that the data from the project, which may be highly sensitive, will be leaked during the writing of the review. Similarly, in the case of police and police file processing, we do not recommend entering personal data of victims, perpetrators, or other case details into AI unless you are certain that it will not be further used to train the AI being used.

Most companies use publicly available data to train AI, which includes data from discussion forums or social networks, which of course also contains personal data (including photos, contact details, names, etc.).

In addition, **the principle of** accountability must be considered. Organisations that use AI to process personal data must be able to demonstrate that they comply with data protection legislation such as the GDPR. This includes regular monitoring, assessing the risks associated with the processing of personal data and applying corrective measures.

Last but not least, it is important to **respect the rights of individuals**. This includes the right to have access to information about what data is collected and how it is used, the right to have inaccurate data corrected, the right to have personal data deleted when it is no longer needed, or the right to object to automated decision-making that may have a significant impact on their lives.

Adherence to these principles is essential not only to ensure users' rights, but also to build trust in AI technologies. Without trust in the way our data is processed, the full potential of AI to create innovative and beneficial solutions can hardly be realised.

In addition to the above rules, the **ethical aspects of the use of AI** in the processing of personal data must also be taken into account. AI should be designed and used in such a way that it respects human dignity and privacy and is not a tool for discriminating or harming individuals. Algorithms should be unbiased and data should be managed with due regard to the possible existence of biases that could lead to unfair results. However, in later chapters of this book, we will show that even AI is not completely unbiased, it can commit errors or biases, and that its ethical limits are not always fully operational.

Another important aspect is **automated decision-making**. AI systems are increasingly being used to make decisions in areas such as finance, healthcare or employment. Automated decisions can be fast and efficient, but if they are based on inaccurate or incomplete data, they **can cause serious harm**. Therefore, control mechanisms should be put in place to ensure that people have the opportunity to review these decisions and, if necessary, challenge them.

It is also important to **strengthen users' information literacy** so that they are aware of how their data can be used and the risks involved. Informed users are better able to make informed decisions about what data they share and what permissions they grant, which increases their ability to protect their privacy.

For the future development of AI, it is necessary to **monitor the impact of its use and to adapt legal and regulatory frameworks** to keep pace with technological change. Lawmakers, developers and organisations should work closely together to ensure that AI innovation is conducted in accordance with both data protection and ethical standards.

These measures strike a balance between harnessing the potential of AI and protecting the rights of individuals, which is essential to ensure the ethical and responsible development of AI technologies in society.

3 Current state of research in Czech Republic

3.1 CEDMO Trends

CEDMO Trends is a research project that aims to understand how the Czech and Slovak population's consumption behaviour is changing in the context of the media, especially in relation to misinformation and disinformation. It also explores media and technology literacy with a particular focus on generative artificial intelligence (AI) and highlights protective factors against hybrid threats.

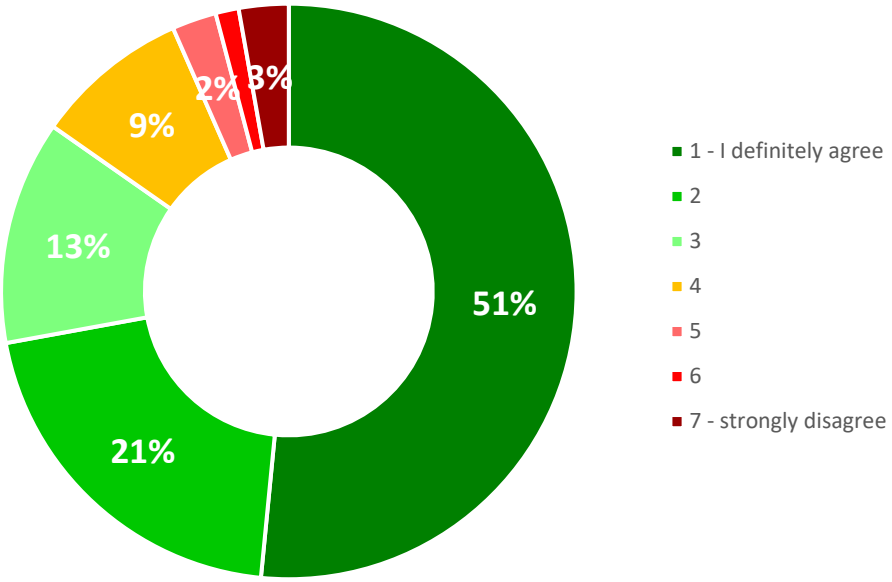
The main data source for CEDMO Trends is a longitudinal panel study that is ongoing in both the Czech Republic and Slovakia. Its duration is 34 months. It is being prepared for Charles University by MEDIAN, s.r.o. as part of the project "Czech Population in the Age of Infodemia". In Slovakia, the research is carried out by IPSOS, s.r.o. The panel of respondents was established in March 2023 for the Czech Republic and in September 2023 for Slovakia. Initially, it included 4,000 Czech and 2,700 Slovak citizens over 16 years of age. Respondents were selected to match socio-demographic characteristics (e.g. age, gender, highest level of education or region of residence), plus past voting behaviour and level of internet use. The method of data collection is a web-based interview, the so-called CAWI (Computer Assisted Web Interviewing) method. The questions are asked simultaneously in the Czech Republic and Slovakia to ensure comparability of the situation in both countries and also to allow for subsequent comparison of the results. The 18th wave of the survey is currently underway in the Czech Republic and the 14th wave in Slovakia.

3.2 Deepfake through the eyes of the Czech population

The questions asked of respondents in CEDMO Trends are mainly attitudinal. They focus on how the Czech and Slovak population perceives current and long-term issues related to media consumption. A specific area is the use of generative AI in the public environment - especially on social networks. An example that falls under this issue is the perception of the so-called deepfake, i.e. in particular audiovisual content generated by artificial intelligence that imitates real persons, events, etc. The issue of deepfake is discussed in more detail in the following chapter. Here we will only present partial results of CEDMO Trends that are relevant to this topic.

Firstly, there is the perception of the downside of this new technology. The vast majority (85%) of people in the Czech Republic assume that deepfake will lead to more disinformation content on the internet. Half (51%) are completely convinced of this (answers "strongly agree"). Only around 7% of respondents hold the opposite opinion.

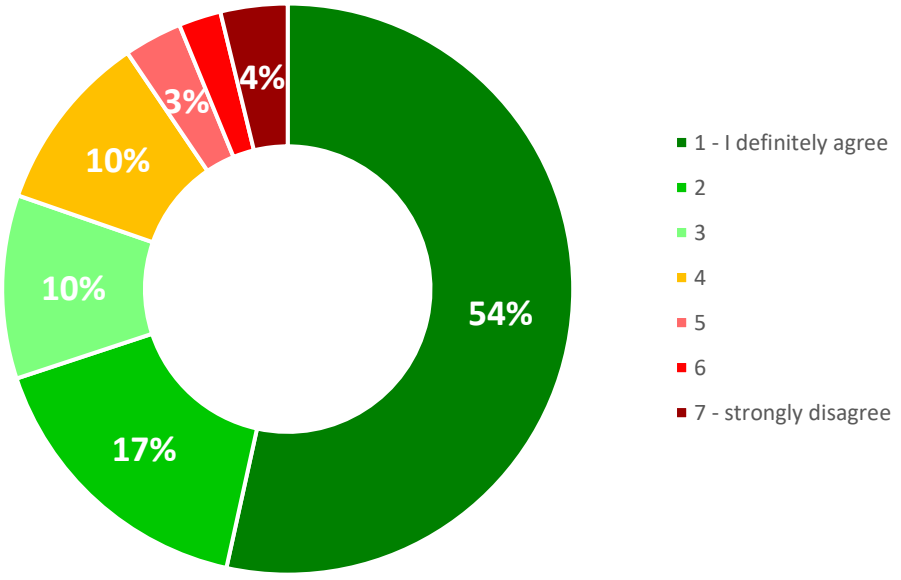
Deepfake will substantially increase the amount of disinformation content (i.e., purposefully disseminated false information) shared on social networks and the internet in general.



n=3000

This unambiguous result shows a predominantly negative view of what deepfake entails. The vast majority of the Czech population is at least aware of the risk of mass dissemination of false content that this technology brings. It is therefore appropriate to ask the question about the activities that should be developed in connection with this new phenomenon. It is easy to ask respondents whether they think it is appropriate to delete such false material from the so-called social networks.

Social network and internet operators should deepfake (i.e. artificial intelligence-generated false videos) delete.



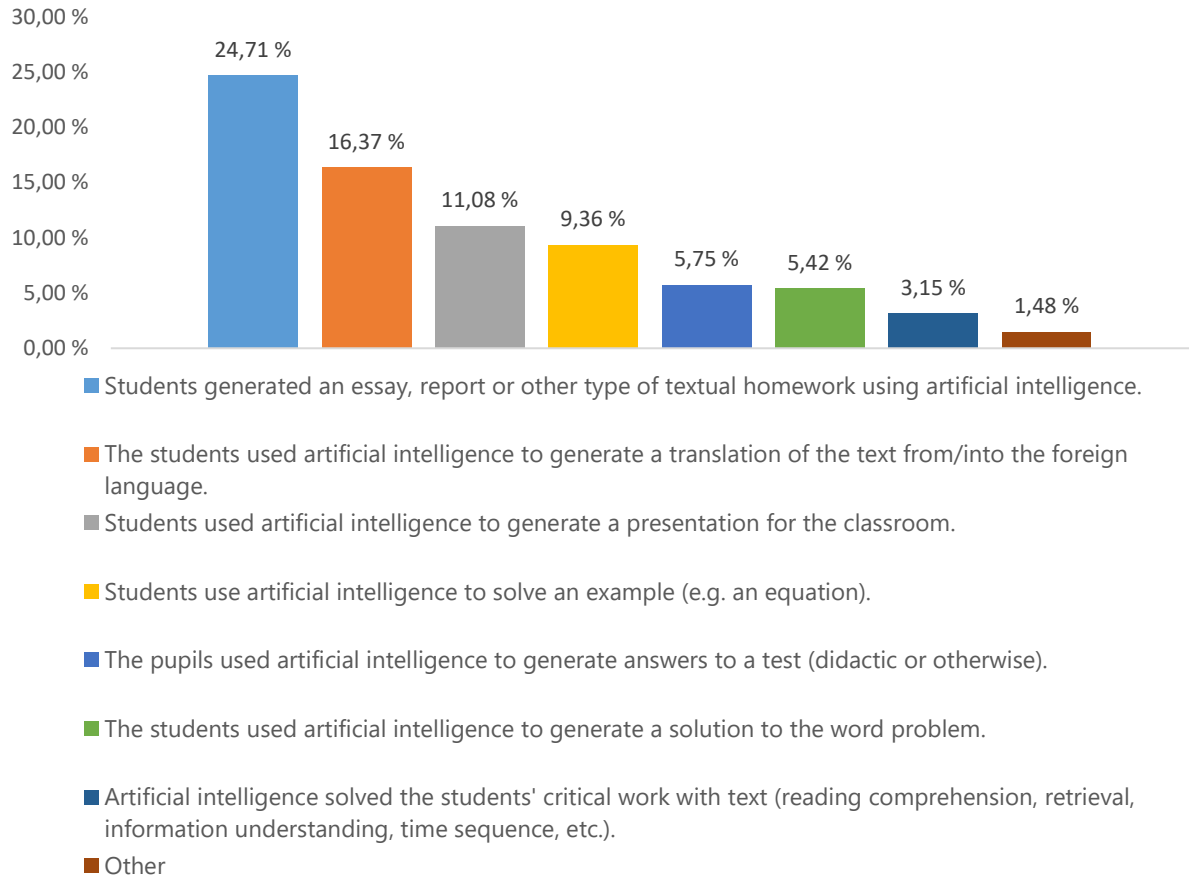
n=3000

The vast majority of the Czech population agrees that so-called deepfake content should be deleted by website operators. More than half (54%) of respondents are uncompromising in this position. 80.5% of all respondents agree at least partially. Only about one in ten (9.5%) are against deleting deepfake content. Roughly the same proportion of respondents are undecided on this question. Such a high level of willingness to delete deepfake posts thus indicates not only a willingness to partially moderate publicly shared content, but also a concern about generative AI technologies.

3.3 Czech Schools and Artificial Intelligence - Teachers' Perspectives

In the Czech Republic, research was conducted in 2023 on Czech primary and secondary school teachers,[2] which investigated teachers' experiences with artificial intelligence (AI) in both in-school and out-of-school environments. The research showed that almost half of teachers (46.7%) consider AI as a tool that can be used to cheat. 34% of teachers confirmed that they know that their students have used this technology to cheat. It is important to note here that this figure only captures situations that educators are aware of. Of course, it is likely that they were unaware of many of the scams.

Educators' experiences of misusing AI for cheating



n=2157

Despite these concerns, most teachers (82%) recognise the need to develop new didactic skills for working with AI. However, many educators feel that society is not ready for the mass introduction of AI in schools.

3.4 Czech children and artificial intelligence - the children's perspective

At the moment, a team from Palacký University in Olomouc is conducting research on the use of artificial intelligence in school environments. The results will be available at the end of 2024.

4 AI misuse incidents

Incidents of AI misuse are reported from all over the world and manifest themselves in many areas of society. In this section we will therefore focus on selected examples of incidents that have occurred in Europe (and of course in the Czech Republic).

4.1 Deepfake and disruption

One of the problems that deep fake videos (and other AI-generated materials) bring is the **erosion of trust in public institutions and political processes, with the associated concerns about influencing democratic processes** (e.g. elections). Fake videos or audio recordings of politicians in compromising or controversial situations can be easily disseminated through social networks and other media. Videos and audio recordings are often so convincing that it is difficult for the average viewer to recognise their falsity. As a result, misinformation can be spread, which can influence public opinion and therefore elections.

Slovakia, for example, has experience with influencing elections using these technologies - just before the 2023 parliamentary elections, a fake recording of a telephone conversation between the chairman of Progressive Slovakia, which is running for the Slovak National Council, and journalist Monika Tódová about the manipulation of election results began to circulate.[3] However, the conversation did not actually take place, it was created by artificial intelligence. Both the Slovak police and the fact-checking agency AFP have described the recording as fake.[4] However, the recording did go viral on the internet and it is clear that it managed to influence some voters.

President of Moldova Maia Sandu also became a target of disinformation campaigns using deepfake videos [5–7] - A few weeks before the local elections, a number of deepfake videos appeared on social media, in which, for example, she refers to the US and George Soros as sponsors of Moldova's pro-European leadership or ironizes the living standards of the population. Moldova's national security authorities attribute these attacks to the Kremlin, which has long sought to destabilise the country's pro-European direction.

By the way, the **International Olympic Committee** (IOC) has also become a target of disinformation videos - a deepfake video has appeared on the Internet (on the Telegram platform and other services) in which a fake Tom Cruise as part of a new (fictional) Netflix documentary series called *Olympics Has Fallen* criticizing the IOC for corruption and the destruction of Olympic sport. The video is fake and appeared shortly after the IOC suspended the Russian National Olympic Committee over its decision to recognize regional sports organizations in the occupied Ukrainian regions of Donetsk, Kherson, Luhansk and Zaporizhzhya as members.[8] Tom Cruise's image has been misused in other videos warning of violence during the 2024 Summer Olympics in Paris.[9,10] The authorship of these fraudulent videos is attributed to the Russian Federation, specifically to the Storm-1679 group and Storm-1099 (see Microsoft analysis)[11] and are primarily aimed at

damaging the IOC's reputation and stoking fears that violence will erupt at the Paris Olympics.

We have experience with fraudulent videos in the Czech Republic as well.During the presidential elections in 2023, a doctored video of presidential candidate Petr Pavel went viral, and another example is the fraudulent deepfake video against Interior Minister Vít Rakusán,[12,13] or various kinds of satirical videos using Andrej Babiš, Tomio Okamura or Alena Schillerová (Kopecký, 2024).

Figure 2. Excerpt from the deepfake video with Minister Austrian (Source: Social network X)

On a global scale, the proliferation of fake deepfake photos of Donald Trump surrounded by African-Americans,[15] to encourage black voters to vote for the Republican candidate in the presidential election in the autumn of 2024. At first glance, the materials are very plausible, but closer examination reveals flaws, such as too shiny skin or missing fingers.

Figure 3. Fake picture of Donald Trump surrounded by African Americans (Source: BBC News)

During the 2024 US presidential primaries, there was a situation in which Joe Biden's artificial voice persuaded Democratic voters not to go to the polls.[16] Steve Kramer, a Democratic political consultant who worked for Biden's intra-party rival in the presidential primaries, Congressman Dean Phillips, admitted to being the author of the artificial voice resembling Joe Biden.[17] Kramer is currently facing a number of charges and trials. The Federal Communications Commission has proposed a six million dollar fine for Kramer and Lingo Telecom has proposed a two million dollar fine (Reuters, 2024).

Similar scenarios can be expected in other countries, and the European Union itself has warned of the risks of influencing elections,[18] which, in the context of the European elections, is calling on the major technology platforms X, TikTok and Facebook to identify and label AI-generated content. The UK and other European countries are also concerned about deepfake videos influencing elections.[19]

Deepfake videos can **also** be **exploited in the context of military conflicts** - the Russian-Ukrainian war provides fresh experience. Already in 2022, a deepfake video appeared online showing Ukrainian President Volodymyr Zelensky surrendering and calling on his troops to lay down their arms.[20,21]

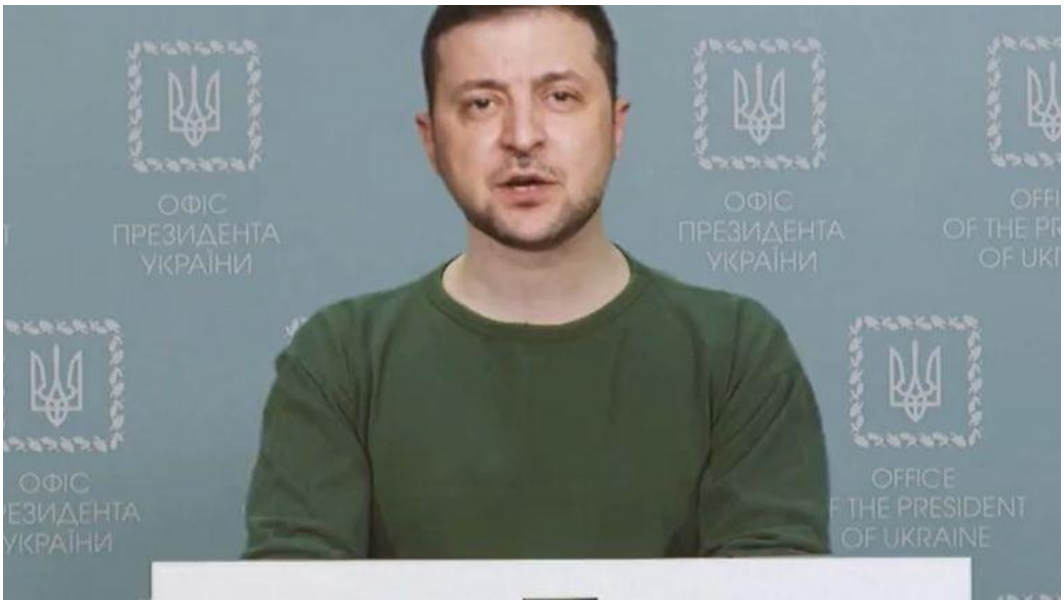


Figure 4. Deepfake video with Ukrainian President (2022) (Source: Sky News)

4.1.1 Deepfake and internet fraud

Another area in which AI (especially deepfake technology) has been used extensively is internet fraud. AI can be used to create phishing or extortion emails, or to generate deepfake videos in which well-known and respected celebrities encourage viewers to react in a certain way. Europol in its report *Facing reality? Law enforcement and the challenge of deepfakes*[22] points out that it is the deepfake technology is and will be an active contributor to organised crime.

Fraudulent advertising has been circulating on YouTube for many months,[23,24] encouraging users to make investments that can make them miraculously rich. The videos then feature Czech President Petr Pavel, former Czech Prime Minister Andrej Babiš and other well-known personalities speaking to users. But it is all a scam. Unfortunately, despite repeated reports, the operators of the big services (Google, Meta, etc.) are unable to eliminate this scam. There are many scams of this type in the media.[25] and the number is likely to increase.

Figure 5. Fraudulent ad using a deepfake video of Czech President Petr Pavel (Source: YouTube)

Figure 6. Fraudulent ad using deepfake video of CEZ CEO Daniel Beneš and Karel Havlíček (Source: YouTube)

Figure 7. Image from Andrej Babiš's deepfake video (Source: Facebook)

These scams are quite successful because they combine a well-known personality (politician, actor, etc.) with a well-known media brand (TN.CZ, CNN Prima News, CT24), company or institution (CEZ, Government Office). For Internet users who have a lower level of media literacy and are blinded by the vision of easy profit, the offer is often credible and they are willing to pay.

4.1.2 Deepfake underwater bankers

AI tools can also be misused for a very dangerous type of fraud in the form of so-called online fraudulent bankers. These are deepfake videos presenting fictitious bankers, and are often sent via email that looks very authentic and believable, often sent from the bank's real email address. In the video, the banker usually reports a problem that has occurred in the bank account, e.g. that someone unauthorised has broken into the bank account and is planning to steal money. In a deepfake video, which is very often created from a photograph of a real bank employee, thus further enhancing its credibility, the banker reports on a procedure that supposedly eliminates the possible theft of funds.

Figure 8. A still from the deepfake video of the fake Fio banka banker. (Source: Facebook, 2024)

4.2 Deepnude and Invasion of Privacy

Deepnude tools use AI technology to create sexually explicit photos of often real people. Through AI algorithms implemented in a deepnude application, any photo of a clothed person embedded in the tool can be transformed into a form bordering on pornographic content. The behaviour in question can cause a number of problems, in particular invasion of personal privacy and damage to reputation. The creation of deepnude content from photographs of a child is very risky, with the associated risk of child pornography. AI deepnude tools have been misused many times even by primary and secondary school students to generate sexually explicit photos of female classmates. An example is a media-famous case from 2023 recorded in the Spanish city of Almendralejo,[26] in which photographs of 20 girls aged between 11 and 17 were used. The girls' images were downloaded from their Instagram profiles, then transformed into sexually explicit images using the deepnude app, and then distributed via WhatsApp and Telegram. One of the victims was even blackmailed by publishing an intimate photograph if she did not pay the extortionists the requested amount.

Abuses of deepnude instruments have also been recorded in the United States, for example, when five boys at Beverly Vista High School in Beverly Hills, California[27] created and shared sexually explicit images of female classmates.

Cases of misuse of deepnude instruments do not avoid the Czech Republic either.[28] Several cases have even been dealt with by the Police of the Czech Republic.

In most of the cases reported in the media space, there is a noticeable trend of abuse of deepnude tools against women, which may be related to the fact that the machine learning

models of AI deepnude tools are trained almost exclusively on nude photos of women. In the case of a generated sexually explicit photo of a man, many deepnude tools fail, especially when modeling the male genitalia - the tool simply models the female genitalia. In the figure below we show an example of the output of the selected deepnude application.

Figure 9. Sample output of a deepnude app (the image of the woman is generated by AI, it is not a real person) (Source: Undressing AI app, 2024)

In 2024, in conjunction with the growing online pornography generated by AI tools, among others, a new section **191a** of the Criminal Code was introduced, ***entitled "Abuse of identity for the production and distribution of pornography"***. The wording of the proposed law reads as follows: '*Whoever produces or distributes a pornographic work that depicts or otherwise exploits a person who has not given his consent shall be punished by imprisonment for up to two years, prohibition of activity or forfeiture of property. If he or she publishes such material, for example on the internet, he or she faces a prison sentence of six months to three years. If he does so as part of an organised group in several countries or if he seeks to obtain large-scale benefits for himself in this way, he will lose his freedom for one to five years.*" If lawmakers adopt the proposed amendment to the section, it would take effect in July 2025.

5 Solution proposals

The proposed solutions respond to the current challenges associated with the development of artificial intelligence (AI) and its impact on society. They cover the areas of legislation, technological innovation and education. Legislation focuses on the regulation of AI based on the risks posed by each application to ensure the safe and ethical use of these technologies. Technological solutions focus on increasing digital literacy and strengthening cyber security through advanced AI tools. Education and dissemination of good practice support the development of digital competences and a responsible approach to AI.

5.1 Legislation

With the advent of advanced forms of AI, legislation naturally and necessarily responds to this state of affairs - for example, the basic regulatory framework for artificial intelligence, the **EU AI Act,** was adopted at the European Union level,[29] which divides AI tools into four tiers according to the risks they may pose.

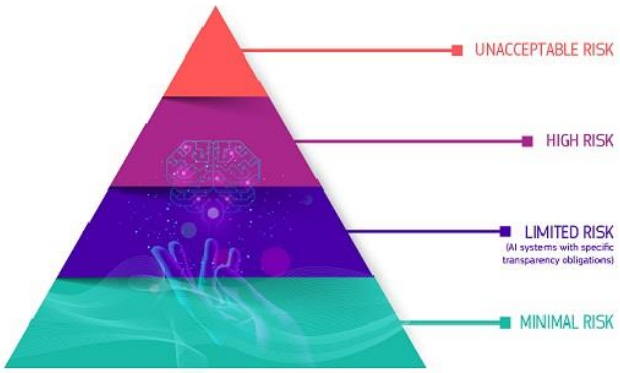


Figure 10. Four levels of risk under the EU AI Act

At the same time, the EU AI Act draws attention to the need to transform European and national legislation in this area.

The newly proposed rules include, for example:

- address risks specifically created by AI applications,
- prohibit AI practices that pose unacceptable risks,
- establish a list of high-risk AI applications,
- Set clear requirements for AI systems for high-risk applications,
- Define specific obligations for operators and providers of high-risk AI applications,
- require a conformity assessment before the AI system is put into service or placed on the market,
- Introduce post-market enforcement of the AI system in question,
- create a governance structure at European and national level.

Details are available on the European Commission website (https://digital-strategy.ec.europa.eu/cs/policies/regulatory-framework-ai).

5.2 Technological solutions

Digital literacy is becoming a key part of modern society, with technology solutions playing a vital role in its development and supporting digital literacy in a number of areas. **This includes in particular the area of education in the form of online courses, educational apps and games, MOOCs and other forms of open education.** Mobile technologies and applications enable easy access to digital content, for example through mobile touch devices.

Collaborative tools facilitate digital communication and effective online collaboration. Cybersecurity is increasingly at the forefront and cannot be done without appropriate technology solutions. For many years, we have been protected by a range of antivirus and anti-malware programs, firewalls, phishing filters, VPN technologies, etc. However, the development of AI systems brings a completely different dimension to the field of cybersecurity. AI brings technological solutions capable of responding to increasingly sophisticated threats in the online space.

It uses machine learning, big data analytics, and predictive algorithms to improve data protection, detect cyber attacks, misinformation, unethical or manipulative behaviour, etc. to minimise potential security risks. Examples include AI tools aimed at **detecting AI-generated text** and detecting unethical behaviour, such as Winston.ai,[30] Originality.ai[31] etc., tools protecting against phishing attacks and fake emails such as PhishMe Cofense,[32] tools against malware, ransomware and other types of attacks such as Sophos Intercept X,[33] CrowdStrike Falcon[34] etc., or innovative tools to combat unethical campaigns such as FactNinja[35].

FactNinja[35] is an advanced tool that uses artificial intelligence to **analyse different types of visual content** such as images, photos, flyers or screenshots of social media posts. The tool is able to identify levels of manipulation, argumentation errors and other problematic elements of visual communication, and is particularly useful in political campaigns or advertising, including fraudulent practices. However, it is not aimed at detecting technical image alterations, such as retouching, AI-generated images, green screen or photomontage, nor at verifying the factual accuracy of information. The analysis is performed by a specially designed multimodal AI model that drives the entire process.

5.3 Education and "good practice"

Education is clearly one of the ways to strengthen the digital competence of the population in the field of artificial intelligence. In the Czech Republic, there are a number of organisations dedicated to education in this area.

Among the best known is the **AI for Kids** initiative, which is a leader in AI education. It promotes creativity, critical thinking and social responsibility in a technology-driven world. It focuses on the hands-on engagement of educators, children and youth in the education process to develop digital competencies and understanding of AI technologies. It offers a wide range of educational materials and tools. It is currently developing an AI curriculum

for primary and secondary schools (https://kurikulum.aidetem.cz/), working on a formative assessment assistant called Tiny, training educators (including future ones) and school leaders, and working on systemic change in education in the age of AI.

The National Institute of Education is actively supporting schools in implementing artificial intelligence (AI) in education through projects funded by the National Recovery Plan. Training programs are aimed at primary and secondary school educators and include webinars, face-to-face workshops and online courses that specialize in specific uses of AI in the classroom. In addition, regional ICT methodologists are available to provide methodological and technical support in integrating AI into the school environment. NPI CR has published key materials such as "Recommendations for the use of AI in primary and secondary schools" or FAQs - frequently asked questions by teachers about AI, which are tailored to different target groups.

Figure 11. Recommendations on the use of AI from NPI CR

Experimental activities include the ExploreEDU project, in which teachers explore the possibilities of using tools like ChatGPT in the classroom and share their experiences. The NPI CR also works with partners such as AI for Children to jointly develop events for teachers. Internal development includes training and workshops for NPI staff on working with generative AI, while at the same time AI is used for the organisation's internal processes. NPI CR conducts an annual survey to map the attitudes of schools and teachers towards AI, and then uses these findings to further develop methodological materials and activities.

Other well-known organizations that are working on this issue include the **Centre for the Prevention of Risky Virtual Communication** and its **E-Bezpečí (E-Safety)** project at the

Faculty of Education at Palacký University in Olomouc. It offers both full-time and online courses oriented on the issue of artificial intelligence, and has also created an online video course (http://ai.e-bezpeci.cz/videokurz) full of pre-recorded video lectures and tutorials for educators and the public. E-Safety also operates a portal dedicated to AI at: http://ai.e-bezpeci.cz.

In addition, E-Safety publishes various types of guides, mainly dedicated to the risks associated with artificial intelligence, such as *Risks Associated with Artificial Intelligence for the Elderly*, *Risks Associated with Artificial Intelligence (for the Public)* and *Artificial Intelligence - Risks and Liability*. Many E-Safety activities have also been implemented with the support of the CEDMO network. It cooperates very closely in this area, for example with the Ministry of the Interior of the Czech Republic or with Google and Microsoft.

Examples of publications produced by the E-Safety team (on the topic of artificial intelligence):

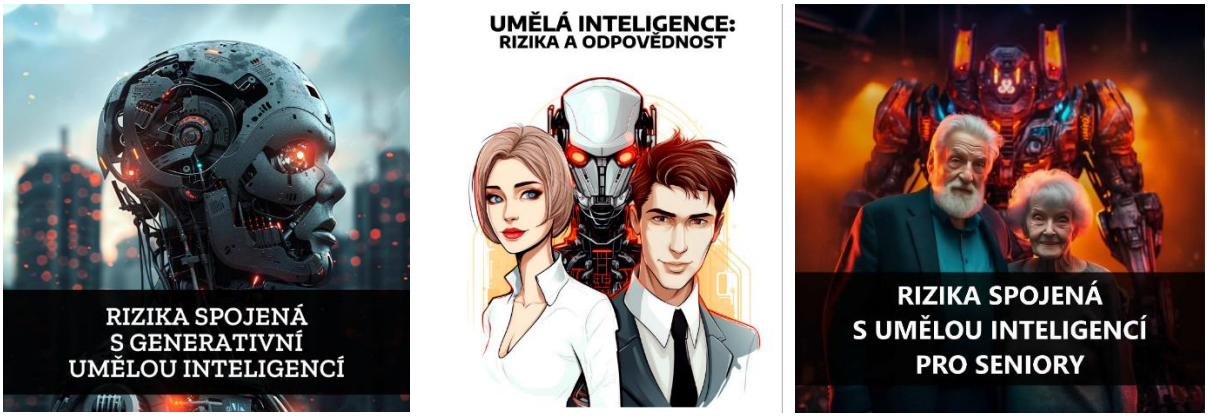


Figure 12. E-Bezpečí materials on artificial intelligence (Source: E-Bezpečí.cz)

Activities aimed at strengthening competences in the field of artificial intelligence are also carried out by the corporate sector, such as the activities of **Microsoft** or **Google**, both independently (https://learn.microsoft.com/cs-cz/training/educator-center/topics/ai-for-education) and through community groups (www.gug.cz).

6 Reference

1. Kopecký, K., Szotkowski, R., & Ziburová, K. (2024). *Risks associated with generative artificial intelligence*.

2. Kopecký, K., Szotkowski, R., Voráč, D., Krejčí, V., & Dobešová, P. (2023). *Czech Schools and Artificial Intelligence.* https://www.e-bezpeci.cz/index.php/ke-stazeni/vyzkumne-zpravy/163-ceske-skoly-a-umela-inteligence-2023/file

3. Šnídl, V. (2023). *The pre-election "deepfake" was not of good quality, the state intervened against pro-Russian channels. It may soon be worse.*

https://dennikn.sk/3608268/predvolebny-deepfake-nebol-kvalitny-proti-proruskym-kanalom-zakrocil-stat-coskoro-to-moze-byt-horsie/

4. Barca, R. (2023). *Alleged recording of a phone call between the PS chairman and a Denník N journalist shows numerous signs of manipulation according to experts | Fakty*. https://fakty.afp.com/doc.afp.com.33WY9LF.

5. Greego, S. (2024). *Moldova's president a victim of deepfake manipulation. In a fake video she was supposed to ironize the living standards of citizens*. Infosecurity.Sk. https://infosecurity.sk/articles/moldavska-prezidentka-sa-stala-obetou-deepfake-manipulacie-vo-falosnom-videu-mala-ironizovat-zivotnu-uroven-obcanov/

6. Necsutu, M. (2023). *Moldova Dismisses Deepfake Video Targeting President Sandu*. Balkan Insight. https://balkaninsight.com/2023/12/29/moldova-dismisses-deepfake-video-targeting-president-sandu/

7. Ratieieva, A. (2023). *Deepfakes - AI in the Hands of Propaganda*. Ukraine Crisis - Media Center. https://uacrisis.org/en/deepfakes-ai-in-the-hands-of-propaganda#

8. Starcevic, S. (2023). *AI 'Tom Cruise' joins fake news barrage targeting Olympics*. Politico. https://www.politico.eu/article/ioc-says-it-was-hit-by-fake-news-campaign-and-ai-tom-cruise/

9. Dilanian, K. (2024). *Russia is trying to scare people away from the Paris Olympics, report says*. NBC News. https://www.nbcnews.com/sports/olympics/russia-trying-scare-people-away-paris-olympics-report-says-rcna154924

Stone, J., & Zuidijk, D. (2024). *Russian Bots Use Fake Tom Cruise for Paris Olympic Disinformation*. Bloomberg. https://www.bloomberg.com/news/articles/2024-06-03/russian-bots-use-fake-tom-cruise-for-olympic-disinformation?srnd=homepage-europe&embedded-checkout=true

11. Watts, C. (2024). *How Russia is trying to disrupt the 2024 Paris Olympic Games - Microsoft On the Issues*. https://blogs.microsoft.com/on-the-issues/2024/06/02/russia-cyber-bots-disinformation-2024-paris-olympics/.

12. Kopecký, K. (2024). Negative impacts of generative AI will be increasingly visible. *E-Security*, *9*(1), 18-22. https://e-bezpeci.cz/journal/articles/3802.html

13. List News. (2024). *A compromising video was produced on an Austrian. He insults people from Karviná and invokes censorship - Seznam Zprávy*. Seznam Zpravy. https://www.seznamzpravy.cz/clanek/domaci-politika-na-rakusana-vyrobili-kompromitujici-video-kde-vyhrozuje-cenzurou-244613

14. Kopecký, Kamil. (2024). Generative artificial intelligence significantly changes our view of information. What will we trust? And how will it affect, for example, elections? *E-Safety*, *9*(1), 33-37. https://e-bezpeci.cz/journal/articles/3884.html

15. Spring, M. (2024). *Trump supporters target black voters with faked AI images*. BBC News. https://www.bbc.com/news/world-us-canada-68440150

16. Wolf, Z. B. (2024). *The deepfake era of US politics is upon us*. CNN. https://edition.cnn.com/2024/01/24/politics/deepfake-politician-biden-what-matters/index.html

17. NBC News. (2024). *Democratic operative admits to commissioning fake Biden robocall that used AI*.

18. Goujard, C. (2024). *EU turns to Big Tech to help deepfake-proof elections*. Politico. https://www.politico.eu/article/eu-big-tech-help-deepfake-proof-election-2024/

19. Sankaran, V. (2024). *UK election may be rigged by adversaries using AI deepfakes, home secretary warns*. The Independent. https://www.independent.co.uk/news/uk/politics/uk-election-rigged-deepfakes-ai-b2502385.html

20. Burgess, S. (2022). *Ukraine war: Deepfake video of Zelensky telling Ukrainians to "lay down arms" debunked*. Sky News. https://news.sky.com/story/ukraine-war-deepfake-video-of-zelenskyy-telling-ukrainians-to-lay-down-arms-debunked-12567789

21. Skácel, O. (2022). *Zelensky's call to lay down arms was a deepfake. 'Russians want to challenge everything,' Koubsky believes*. IRozhlas. https://www.irozhlas.cz/veda-technologie/technologie/deep-fake-zelenskyj-slozte-zbrane-dezinformace-ruska-propaganda_2203182232_vtk

22. Europol. (2024). *Facing reality? Law enforcement and the challenge of deepfakes An Observatory Report from the Europol Innovation Lab.* https://doi.org/10.2813/158794

23. Kopecký, K. (2023). YouTube is infested with fraudulent investment advertising that exploits ČEZ or, for example, President Petr Pavel. Don't get caught, it's a scam. *E-Safety*.

24. Lánský, T. (2024). *When Pavel or Babiš advises on investments. Internet scammers are getting better*. IDNES.Cz. https://www.idnes.cz/zpravy/domaci/umela-inteligence-deep-fake-video-pavel-babis.A240209_192239_domaci_pukk

25. Beneda, J. (2023). *VIEWPOINT: 'We are entering a world where not a single video can be trusted'*. IResponses. https://www.irozhlas.cz/veda-technologie/technologie/deepfake-podvod-prezident-petr-pavel-ukrajina-rusko-caputova-dezinformace_2311120500_job

26. Llach, L. (2023). *Naked deepfake images of teenage girls shock Spanish town: But is it an AI crime?* Euronews. https://www.euronews.com/next/2023/09/24/spanish-teens-received-deepfake-ai-nudes-of-themselves-but-is-it-a-crime

27. Singer, N. (2024). *Teen Girls Confront an Epidemic of Deepfake Nudes in Schools*. The New York Times. https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html

28. Habešová Daniela. (n.d.). *Undressing apps are rampant in the Czech Republic. Any woman can find herself naked, experts sound the alarm - CNN Prima NEWS*. Retrieved October 16, 2024, from https://cnn.iprima.cz/v-cesku-radi-nova-aplikace-umela-inteligence-vas-na-fotce-svlekne-do-naha-415502

29. *EU AI Act: first regulation on artificial intelligence | Topics | European Parliament*. Retrieved October 17, 2024, from https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

30. *Winston.ai*. (n.d.). Retrieved September 17, 2024, from https://gowinston.ai/

31. *Originality.ai*. (n.d.). Retrieved September 17, 2024, from https://originality.ai/

32. *PhishMe Cofense*. (n.d.). Retrieved September 17, 2024, from https://cofense.com/

33. *Sophos Intercept*. (n.d.). Retrieved September 17, 2024, from https://www.sophos.com/en-us

34. *CrowdStrike Falcon*. Retrieved September 17, 2024, from https://www.crowdstrike.com/en-us/

35. Kopecký, K. (2024). *FactNinja*. https://www.factninja.cz

Opportunities for strengthening digital literacy in the context of the introduction of artificial intelligence

(analytical report)

Authors:
Kamil Kopecký, René Szotkowski

Centre for Prevention of Risky Virtual Communication
Faculty of Education, Palacký University in Olomouc
www.prvok.upol.cz

2024

This publication has been produced by the Centre for the Prevention of Risky Virtual Communication of the Faculty of Education of Palacký University in Olomouc in cooperation with CEDMO with the support of the National Recovery Plan within the framework of the project 1.4 CEDMO 1 - Z220312000000 Support to increase the impact, innovation and sustainability of CEDMO in the Czech Republic, which is funded by the EU Recovery and Resilience Instruments.