GUIDELINES ON THE RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE IN THE NEWS MEDIA

These guidelines are part of Ukraine's artificial intelligence (AI) regulation roadmap. AI systems can be useful at various stages of the news media operations, from data collection and analysis to content creation, news distribution and audience interaction.

The Guidelines aim to implement the Guidelines on the responsible implementation of artificial intelligence systems in journalism adopted by the Council of Europe's Steering Committee on Media and Information Society and disseminate relevant international practices, principles and approaches to the responsible use of AI systems in news media to respect human rights and professional ethical standards.

Ministry of Digital Transformation of Ukraine

MINISTRY OF CULTURE AND INFORMATION POLICY OF UKRAINE

NATIONAL TELEVISION AND RADIO BROADCASTING COUNCIL OF UKRAINE

Definitions

An AI system is a machine system that, based on the input data it receives, makes a conclusion about generating outputs (such as forecasting, creating content, and providing recommendations or solutions) that can affect a physical or virtual environment. Different AI systems differ in their level of autonomy (the ability to operate without human intervention) and adaptability (the ability to be flexible and adjust to the environment) after application.

Generative AI is an AI system that, using generative models, is able to create new, original content (texts, images, audio and video materials). Examples of generative AI systems include *ChatGPT, Copilot, Bard, Gemini, Midjourney, DALL-E, Stable Diffusion,* and others.

Application areas

The guidelines can be applied to any AI systems—both those developed by third parties and those developed by media entities themselves or commissioned by them as well as those used by journalists and other media professionals, in particular for:

- content creation by using generative AI systems,
- big data analysis,
- administration/automation of workflows (search for experts, selection of sources, drafting standard contracts, etc.),
- content management (searching and verifying content, translating or transcribing material, moderating comments on media websites, prioritising, or personalising content).

Main principles of the responsible use of AI systems in the news media

The use and development of AI systems in the news media should be carried out with due regard to the basic values of journalistic ethics, including truthfulness and accuracy, impartiality and independence, non-discrimination, accountability, inclusiveness, respect for

privacy, confidentiality of sources, and other principles stipulated by the Code of Ethics of Ukrainian Journalists.

In addition, introducing and using AI systems in the news media requires implementing additional principles of activity. It is recommended to define the following principles for the use of AI systems in the media that meet the public interest:

- **responsible editorial decision** (implementation of AI systems based on a conscious and balanced decision of the editorial staff of a media entity and taking into account the understanding of the mission of such a media outlet);
- **legality** (availability of necessary legal rights to use AI systems / publish AI results);
- **systematic assessment of risks** associated with the use of AI systems (conducting legal and technical risk assessment at all stages of the AI system life cycle);
- **transparency and clarity** (disclosure of information on the use of AI systems and explanation of the purposes and methods of their use; understanding the sources of information on which AI operates);
- **awareness of the audience** about the use of AI systems and the nature of the disseminated content (proactive communication of information about the use of AI to the reader/viewer/listener; a clear distinction between authentic and AI-generated content, etc.);
- **confidentiality and data protection** (preventing leakage of personal data or other confidential information through the AI systems used);
- **diversity** (ensuring audience access to diverse content when using personalisation tools) and non-discrimination;
- **human oversight** (ensuring that the results of AI systems are checked by a human);
- **accountability** (a person using AI systems in professional activities is responsible for the consequences of using the results of these tools);
- **adaptability** (continuous improvement of the principles of responsible use in line with the technological development of AI systems and changes in legal regulation).

Implementation of AI systems into professional and organisational practice

Media actors **that develop AI systems themselves or commission AI systems from developers** specifically for the needs of the news media should be aware of their responsibility for the system's design, compliance with human rights standards, and its proper usage. In this context, the media is advised to:

- comply with personal data protection legislation when developing AI systems;
- continuously monitor the state of the AI system, in particular with regard to its ability to comply with human rights standards (especially in cases where the AI system is capable of self-learning);
- ensure that the system is non-discriminatory and does not have negative consequences for vulnerable and marginalised groups of people;
- obtain all necessary legal rights and consents before publishing content created using an AI system;
- notify users about the use of AI systems;
- supervise the operation of the AI system and suspend the use of the AI system in case of malfunctions or deficiencies in the algorithm and timely correct gaps in the AI system before continuing its use;
- provide the audience with the opportunity to file complaints if the audience members suspect that the AI system violates their rights.

Media organisations **that use AI systems developed by third parties** (including those that are freely available) are recommended to follow the recommendations on monitoring the status of AI systems, notifying the audience about the use of AI systems, supervising and suspending the use of systems in case of violations, and providing the opportunity to complain about a breach of legal rights. In addition, media actors are recommended to:

- ensure that an AI system that is not developed by or commissioned by this media entity complies with human rights standards;
- refrain from using AI systems originating from the aggressor state or developed by citizens of the aggressor state;

- use AI systems on legal grounds in compliance with the intellectual property rights of third parties;
- monitor notifications from the AI system manufacturer regarding the system's compliance with human rights and data security standards as well as updates to the AI system and its compliance with ethical standards.

Working with AI systems often requires skills that go beyond the existing training of most media professionals. Therefore, media actors are encouraged to provide training for their staff on using AI systems through programmes, courses, and workshops that bring together technical specialists and journalists and promote awareness of human rights and professional ethics. In particular, an educational series on AI is offered on the Diia.Osvita portal.

As AI systems require an organisational infrastructure to support their proper functioning, media actors are advised to create such an internal infrastructure by hiring or training staff with clearly defined AI roles (these roles should not be implemented solely through the development of routine AI competencies among staff). In particular, it is recommended to conduct training that covers the risks of using AI systems developed by other companies and provides a detailed overview of their operation before they are put into practice.

The **annex** to the guidelines contains a Recommended algorithm for selecting AI systems for work and evaluation of AI-generated content, which can be used to assess AI systems' overall security and reliability.

Transparency and explainability

Media outlets are encouraged to disclose when and how they use AI systems. The principle of transparency requires that it be applied **at all stages of the AI system's life cycle** and explain how the system operates, disclose potential human rights risks from using AI systems, and indicate possible impacts of such systems. Media entities are encouraged to develop and publish their own policies, strategies, standards, guidelines, or other documents that outline the principles of using AI systems in their operations. Examples include the published principles of global media: The Guardian, BBC, Wired, CNET, Associated Press, etc.

Any use of AI systems that affects the collection of information, production, processing, or distribution of media content should be clearly marked and communicated to all those who consume this information.

Within a media entity, it is recommended that information about which AI systems are used to collect information, produce, process, or distribute content, their purpose, and the values they reflect be made available to employees.

Distinguishing authentic and AI-generated content

Generative AI makes spreading false information and disinformation easier through altered words, photos, video, or audio, including content that may not show signs of change and appear authentic.

The need to distinguish AI-generated content can arise in two cases: when a media outlet uses other people's content and when a media outlet deliberately generates content using AI systems to illustrate its own materials.

Journalists are advised to exercise due diligence, care, and scepticism to avoid accidental use of other people's AI-generated content to ensure that the materials they refer to are **authentic**[1]. It should also be remembered that in Ukraine, AI-generated content is subject to intellectual property rights (Article 33 of the Law of Ukraine "On Copyright and Related Rights").

If the materials **contain content generated using AI systems**, the media is recommended to:

- identify the source of original content, conduct reverse image search to confirm their origin and the presence/absence of automated processing or generation, and check for messages with similar content in trusted sources (including official sources, fact-checking resources, etc.);
- use technical tools, including AI-based ones, to verify whether the content is original or artificially generated to counteract the spread of deepfakes or other content generated for malicious purposes. Examples of such tools include *Sentinel,*

[1] For the purposes of these Guidelines, "authentic" means content created without the use of AI systems to generate or significantly modify it.

FakeCatcher, WeVerify, Microsoft's Video Authenticator Tool, detection of duplicate content using *Phoneme-Viseme Mismatches,* etc.;

- assess on a case-by-case basis whether the use of AI-generated content is contextually appropriate and whether the use of such content will not mislead readers/viewers/listeners even if labelled (for example, if the content deals with a very sensitive topic and may cause an emotional reaction);
- refer to such content or distribute it in compliance with the principle of transparency, labelling it accordingly if the content was generated without the purpose of harming a particular legitimate interest;
- avoid redistributing content generated using AI systems for the purpose of misleading, spreading disinformation or illegal content (such as calls for violence, hate speech, discrimination, war propaganda, etc.;)
- in case of distribution of content generated by third parties using AI systems, comply with the requirements for protecting the intellectual property rights of such parties.

Content that is not authentic and does not bear any labelling as being generated by AI is recommended to be considered misleading. If journalists or other media participants have doubts about the material's authenticity, it is not recommended to use it.

In cases where **media outlets use AI to generate content**, in particular, to illustrate cultural materials, model situations, and illustrate examples and theories, they are recommended to:

- check whether the content generated by the AI system clearly and accurately reflects reality, refers to available and reliable sources, and uses up-to-date information;
- assess on a case-by-case basis whether the use of AI-generated content is contextually appropriate and whether the use of such content will not mislead readers/viewers/listeners even if labelled (for example, if the content deals with a very sensitive topic and may cause an emotional reaction);
- refrain from using or distributing news and news-analytical content created by AI that imitates the recording of real-world footage or real people; refrain from using generative AI to add or remove any elements from authentic photographs, video, and audio materials, except when necessary to ensure human rights, protect the rights of third parties or comply with other legal requirements;

- distribute content generated using AI systems in compliance with the principle of transparency, labelling it accordingly.

It is crucial to ensure a clear distinction between content derived from physical recordings of the real world (e.g. photography, audio, and video) and that generated or significantly altered by AI systems. The use of authentic footage and recordings to reflect actual events should be preferred as long as such use does not violate human rights or applicable law.

Media actors are encouraged to use tools that authenticate published content, where possible, by providing reliable information about its origin and any subsequent changes it may have undergone. For example, the UK news agency Reuters has implemented a tool developed by the Content Authenticity Initiative that allows for authentication of photojournalism content by including information about its origin and modification history in digitally signed file metadata.

Labelling AI-generated content

In the case of publishing AI-generated content, such as an illustration, text, or other artwork created with the help of AI, it is critical to mark such content clearly.

There are a number of ways in which media can transparently inform audiences about the use of AI: watermarks, labels, or other markings to identify artificially generated content clearly. For example, use the caption *"Material generated using artificial intelligence"* in reports about illustrations or artwork created by AI or in materials created by AI.

In addition to direct labeling in materials, media are encouraged to offer even greater transparency by using a sidebar on online platforms with data on the AI systems used the purpose and methods of such use, and publishing technical documentation or a technical repository to explain the AI tools used.

Human oversight

While AI systems can perform highly automated tasks as part of daily workflows, freeing up time and resources for other activities, human decision-making should remain central to long-term strategies and daily editorial tasks. To avoid incorrect or biased processes/outputs from AI systems, it is recommended that final professional control be retained by humans.

It is essential for editorial teams to clearly define the goals, scope, and terms of use for each AI system. At the same time, it is advisable to conduct end-to-end and continuous oversight of the impact of deployed AI systems, ensure that they align with the terms and purposes of use, and maintain the ability to deactivate them at any time.

Human oversight is critical for tasks where the results are particularly sensitive (e.g. those with specific consequences for individuals) or socially significant (e.g. affecting national security and defence interests). Media should consider the impact that the use of AI systems may have on public opinion or the emotional state of the audience.

Personalisation of media content and protection of personal data

The development and use of AI systems for automatic content personalisation and recommendation should be guided by journalistic ethics. AI-assisted content personalisation and AI-based recommendation systems should be grounded on the principles of diversity, pluralism of opinions, and integrity of information and in compliance with personal data protection laws. In particular, mechanisms for personalising and prioritising content should not be used to spread covert election campaigning. It is recommended to use safeguards in such mechanisms to prevent potentially harmful, dangerous, or sensitive content (targeting 16+ or 18+ audiences) from being offered to underage audiences.

It is recommended that the use of such AI systems be made transparent: the audience should be regularly reminded that certain news content is personalised, how and why it is personalised. This means providing individuals with the necessary information about their personal data used for personalisation, ensuring that they have control over their data to adjust their personal profile and choose between different personalisation methods that consider their short- and long-term interests.

It is recommended that users have the option to disable content personalisation by AI systems to ensure unfiltered access to media content.

Accountability

Media entities should be aware of editorial responsibility for using AI to collect, process, or disseminate information. According to the Commission on Journalistic Ethics' Recommendations on the Use of Artificial Intelligence in the Production of Journalistic

Materials, the responsibility for journalistic material lies with the author and the editorial office, even if the author used AI to prepare the material.

It is recommended to consider any result of generative AI as material that requires mandatory additional verification and human participation in the decision-making process on its distribution (publication).

It is recommended to include the principles of using AI systems in the news media in editorial charters and developing internal rules for the ethical use of specific AI systems (both in media work and for operational purposes).

Media self-regulation

Media and journalists have an essential role in developing and updating standards for the responsible use of AI in the media and should have a clear and coherent vision. It is recommended to develop industry codes and internal guidelines for the safe, ethical use of AI in the media based on the principles of respect for human rights, democracy, and the rule of law. An example of this is the aforementioned Recommendations of the Commission on Journalism Ethics on the use of artificial intelligence in journalistic production.

The exchange of best practices, interdisciplinarity, and cooperation between the professional and academic communities and the public sector are important conditions for the development of responsible use of AI systems in the media sector.

RECOMMENDED ALGORITHM FOR SELECTING AI SYSTEMS FOR WORK AND EVALUATION OF AI-GENERATED CONTENT

This recommended algorithm is designed to offer a convenient mechanism for assessing the suitability of a particular AI system for work and the suitability of content generated using such a system for distribution.

Every checklist contains two categories of items: **“red flags”** (which indicate unacceptable or high risks of using an AI system and disseminating AI-generated content) and **“pay attention”** (which means that an informed decision needs to be made).

The use of AI systems and the content they generate always involves risks. Each risk can be assessed on a scale:

“red flags”

- **unacceptable**—the presence of at least one such risk indicates that it is impossible to use such an AI system or distribute its generated content. This is an ultra-red zone for the use of an AI system and its generated content;
- **high**—the presence of several of these risks indicates either the impossibility of using such an AI system or distributing the content generated by it or a high probability of financial, organisational, and reputational losses if such an AI system is used or the content generated by it is distributed. This is a red zone for the use of an AI system and its generated content;

“pay attention”

- **medium**—there is a potentially medium likelihood of financial, organisational, and reputational losses if such an AI system is used or the content generated by it is disseminated, but such losses are unlikely to affect the entity’s media activities in the long term. This is a yellow zone for the use of an AI system or its generated content;
- **low**—low probability of negative financial, organisational, and reputational losses if such an AI system is used or its generated content is distributed. This is a green zone for using an AI system or its generated content.

Important: low risks associated with using a particular AI system do not automatically guarantee low risks when distributing the content generated by it, so each content generated using an AI system must undergo a risk assessment associated with its distribution.

Choosing an AI system for work

0. READ THE TERMS OF USE FROM THE AI SYSTEM DEVELOPER

1. JURISDICTION

- country of registration of the AI system developer;
- ultimate beneficial owners of the AI system developer;
- place of business of the AI system developer;
- location of servers/cloud storage of the AI system developer;
- applicable law under the agreement with the AI system developer;
- the court in which disputes with the AI system developer regarding the use of the AI system are considered under the agreement.

“red flags”

An unacceptable risk is the presence of the Russian Federation, Iran, and the DPRK in responding to any of these items.

A high risk is the presence in the answers to any of the aforementioned items of countries with a high human rights violation index (in particular, in relation to the protection of personal data and intellectual property rights), the presence of sanctioned persons among the ultimate beneficial owners of the AI system developer, cooperation of such an AI developer with the governments of the countries listed as having unacceptable risks.

“pay attention”

A medium risk is the presence of countries other than Ukraine and the EU (except for the countries mentioned in the red flag zone) in the answers to any of the above items; cooperation of such an AI developer with sanctioned persons.

2. TERMS OF PAYMENT

“pay attention”
(medium to low risks):

- payment model (one-time payment, subscription, etc.) or open source licence;
- whether the price offered is competitive compared to AI systems with similar functions;
- whether there are additional costs (e.g. infrastructure, equipment, team training) or hidden fees;
- whether it is possible to pay in hryvnia: due to the martial law, payment in foreign currency is often difficult;
- whether the AI system developer is ready to provide documents for the correct making of payments regarding Ukrainian tax legislation, namely, signing a contract, act, etc.

3. INTELLECTUAL PROPERTY RIGHTS

- what legal rights to AI-generated content the user receives.

“red flags”
(risks range from unacceptable to high if the answer is “no” to even one of these questions):

- whether the AI-generated content can be used for commercial purposes or in the statutory activities of the entity in the media field;
- whether the rights to the AI-generated content can be transferred to other persons (in particular, customers) if such transfer is planned;

- whether the AI system developer guarantees that the AI-generated content does not infringe on any intellectual property rights of third parties.

“pay attention”
(medium to low risks):

- whether the developer of the AI system retains any rights to the generated content, and if so, which ones;
- whether the developer of the AI system declares that it has complied with all intellectual property rights to the databases (datasets) on which it trained its AI model;
- whether the developer of the AI system guarantees to the user of such a system that the user recognises and retains the intellectual property rights to the user’s databases (datasets) and user samples.

4. PRIVACY: CONFIDENTIALITY AND PROTECTION OF PERSONAL DATA

“red flags”
(risks from unacceptable to high if personal data is used in the industrial sector and the AI system developer does not ensure proper protection of such data):

- whether it is possible to delete personal data if it is entered in a prompt;
- where the prompts are stored if they contain personal data;
- who, other than the AI system developer, has access to the prompts if personal data is entered in them;
- whether the developer of the AI system guarantees that the generated content does not contain images, names, surnames, voices, or other identifying features of actual individuals;
- whether it is planned to use confidential information in the user’s prompts and, if so, whether the developer guarantees the confidentiality of the information contained in the user’s prompts;

- whether it is planned to use personal data in the prompts and, if so, whether the AI system developer guarantees compliance with the legislation on personal data protection.

“pay attention”
(medium to low risks):

- whether the user has the option to delete the prompt;
- what other persons, in addition to the AI system developer, may have access to the user’s prompts;
- what legislation the AI system developer is guided by in terms of personal data protection.

5. REQUIREMENTS FOR TEAM SKILLS, INFRASTRUCTURE AND EQUIPMENT

“pay attention”
(risks from high to low):

- what skills are required to use the AI system and how long it takes to learn how to use it;
- whether the AI system developer offers training in the use of its AI system and under what conditions (duration, payment, regularity, etc.)
- whether the use of the AI system imposes additional infrastructure and hardware requirements (e.g. access to cloud technology, incompatibility with specific platforms);
- if so, what are the short-term and long-term additional costs;
- whether technical support is offered and on what terms (schedule, scope, cost);
- whether the media as a user has the right to switch to another AI system developer and transfer the dataset to another AI system;
- whether the developer of the AI system has the unilateral right to change, modify, or terminate the user’s access to the AI system at any time.

6. HUMAN OVERSIGHT

“red flags”
(risks from unacceptable to high):

- what skills are required from the media entity’s team to control the operation of the AI system;
- who in the media entity’s team is responsible for reviewing content generated using the AI system for compliance with legal and ethical standards;
- whether the media entity has assigned such a responsible person and the obligation to do so in a document and, if so, in what document.

7. RESPONSIBLE DEVELOPMENT

“pay attention”
(medium to low risks):

- whether the AI system is designed specifically as a journalistic AI;
- whether the AI system is designed to work with the Ukrainian language and for the needs of the Ukrainian audience;
- whether the developer of the AI system guarantees that its AI system is developed in compliance with public values and human rights.

8. ACCOUNTABILITY

“red flags”
(risks from unacceptable to high):

- who is responsible for what: for what specifically (guarantees of proper and legal functioning of the AI system, observance of the rights of third parties, etc.), to what extent the developer of the AI system is responsible, and to what extent the user is responsible;

- what guarantees are offered by the AI system developer in case of legal liability of the user to third parties due to factors beyond the control of the media entity;
- what legal assistance is guaranteed by the developer of the AI system in case of detection of violations of intellectual property rights, human rights, personal data protection, etc., as a result of the conscientious use of the AI system by the media entity;
- what measures the AI developer takes to ensure the ongoing security of its AI system and ongoing compliance with legal and ethical requirements.

Evaluating the content generated by the AI system

1. HUMAN OVERSIGHT

“red flags”
(risks range from unacceptable to high if the answer is “no”):

- whether the content has been reviewed by a person responsible for reviewing the content generated by the AI system for compliance with legal and ethical standards.

2. USING AN AUTHORISED AI SYSTEM

“red flags”
(risks range from unacceptable to high if the content is generated by an AI system that is not authorised by a media entity according to the criteria above):

- whether the content is generated by an AI system authorised by a media entity in accordance with the aforementioned criteria.

3. INTELLECTUAL PROPERTY RIGHTS

“red flags”
(risks range from unacceptable to high even if one of these questions is answered “no”):

- whether the current payment model or open-source licence allows the use of AI-generated content for media distribution, commercial purposes, or other purposes planned by the media entity;
- whether the rights to the AI-generated content can be transferred to other persons (including customers) if such transfer is planned;
- whether the AI-generated content visually/audibly contains signs of similarity (or imitation) to any objects protected by intellectual property rights (trademark, work, etc.);
- whether the developer of the AI system guarantees that the content generated by the AI does not violate any intellectual property rights of third parties;
- whether the user has used any intellectual property of third parties (trademarks, works, etc.) in the prompt without legal grounds. The main way to avoid risks in this regard is not to use the intellectual property of third parties without proper legal grounds.

“pay attention”
(medium to low risks):

- whether the developer of the AI system retains any rights to the AI-generated content, and if so, which ones;
- whether the developer of the AI system guarantees to the user of such system the recognition and retention of intellectual property rights to the user’s databases (datasets) and user’s samples;
- make a decision whether it is necessary to protect intellectual property rights to the user’s own software;
- decide whether to protect intellectual property rights to AI-generated content.

4. PRIVACY: CONFIDENTIALITY AND PROTECTION OF PERSONAL DATA

“red flags”
(risks range from unacceptable to high even if one of these questions is answered “no”):

- whether the content generated by the AI system contains confidential information or trade secrets;
- whether the content generated by the AI system contains personal data, in particular, whether the generated content does not contain images, names, surnames, voices, or other identifying features of actual individuals.

The main way to minimise risks is not to use confidential information, trade secrets, or personal data in prompts and to check the content generated by AI for their absence.

5. RELIABILITY AND ETHICAL USE OF CONTENT GENERATED BY AI SYSTEMS

“pay attention”
(risks from high to medium):

- whether the use of AI-generated content is contextually appropriate and whether the use of such content will not mislead the audience, even if labelled (for example, if the content deals with a very sensitive topic and may cause an emotional reaction);
- whether the AI-generated content is reliable and all facts in it have been verified;
- whether the AI-generated content is checked for originality and authenticity to counteract the spread of deepfakes or other content generated by AI for malicious purposes. As already mentioned, examples of such tools as of January 2024 include *Sentinel, FakeCatcher, WeVerify, Microsoft’s Video Authenticator Tool,* tools for detecting deepfakes using *Phoneme-Viseme Mismatches,* etc.;
- whether the AI-generated content does not violate the non-property rights of third parties, in particular, honour, dignity, and business reputation;
- whether the AI-generated content is inclusive;
- whether the AI-generated content is not discriminatory or misleading;
- whether AI-generated content does not incite hatred, hostility or promote violence;
- whether the AI-generated content does not violate any other norms of the current legislation and journalistic ethics.

6. TRANSPARENCY IN THE USE OF AI

“pay attention”
(risks from high to medium):

- ◇ whether the content created using AI is labelled as having been created using AI;
- ◇ whether the user is informed about the use of the AI system for content personalisation;
- ◇ whether the media user is granted the right to refuse content personalisation: how the user is informed of this right and what is the procedure for such a refusal.●