British Embassy
Kyiv

Institute of
Innovative
Governance

A GUIDE FOR CONTENT CREATORS

TO IDENTIFY AND COUNTER RUSSIAN PROPAGANDA IN THE LATEST TECHNOLOGIES

Institute for Innovative Governance, 2024

The publication of the Guide for Content Creators to Detect and Counter Russian Propaganda in New Technologies was funded by the UK government as part of the project " AI Awareness: A Guide for Content Creators to Identify and Combat Russian Propaganda in Emerging Technologies", implemented by the Institute for Innovative Governance.

The views expressed in this publication are those of the author(s) and may not coincide with the official position of the UK government.

Author:
Anna Mysyshyn, Ph.D, Co-founder and Director of the Institute for Innovative Governance

Design:
Mark Mironchuk

CONTENTS

SUMMARY

This guide is a comprehensive resource developed by the Institute for Innovative Governance with the support of the British Embassy in Ukraine. It is designed to help content creators, journalists and social media users identify and counter Russian propaganda, especially when it is spread using the latest artificial intelligence technologies.

In today's digital landscape, where artificial intelligence tools such as ChatGPT have become widespread, the risk of rapidly spreading disinformation and propaganda is significant. These technologies are capable of rapidly generating various forms of content, distorting historical facts and news. This issue is particularly acute in the context of the ongoing war in Ukraine and the global political situation.

The guide provides an overview of the capabilities of artificial intelligence technologies in content creation and a detailed approach to detecting Russian propaganda created with the help of artificial intelligence in the information space. It emphasises the importance of fact-checking and verification.

It focuses on the responsible use of artificial intelligence in journalism and content creation, and balancing the benefits and risks of these technologies.

The guide covers such important topics as machine learning, deep learning, generative AI, AI hallucinations and the problem of deep fakes.

The guide also examines the role of social media in spreading disinformation and highlights strategies to improve media literacy. It discusses the accountability of stakeholders, including journalists, AI developers and media platforms, and highlights the need for ethical principles in the age of advanced AI technologies.

This guide thus serves as both an educational tool and a call for ethical engagement with new technologies of information dissemination, which is crucial to combat disinformation and promote a well-informed society.

MAIN OBJECTIVES:

1. To educate the target audience about the nuances of AI-generated content and its potential misuse for spreading propaganda.

2. Provide users with practical skills and knowledge to critically evaluate content created by artificial intelligence, especially in the context of Russian propaganda.

3. Promote ethical standards and responsible practices of using artificial intelligence tools for content creation.

The manual is intended for journalists, bloggers, social media personalities and anyone involved in the creation and dissemination of information content.

INTRODUCTION

In a rapidly evolving digital world, the emergence of artificial intelligence (AI) technologies has fundamentally changed the way we create and consume content. This shift has not only opened up new avenues for innovation and creativity, but also created unique and challenging issues. Among them is the growing proliferation of AI-generated propaganda.

This handbook, created by the Institute for Innovative Governance with the support of the British Embassy Kyiv, aims to provide content creators, journalists and the general public with the knowledge and tools they need to navigate the complex interplay between artificial intelligence and propaganda.

As artificial intelligence technologies such as ChatGPT become more sophisticated, they offer the potential to create a variety of content, from news articles to social media posts. However, these capabilities also carry the risk of misuse to spread disinformation and propaganda. This issue is all the more relevant in the context of the war in Ukraine and the wider global political landscape, where the truth is often the first casualty.

The guide is structured to provide a comprehensive understanding of the capabilities and limitations of modern artificial intelligence technologies. It examines in detail how these technologies can be manipulated to create persuasive but misleading content, and emphasises the need to critically examine and verify the information we consume and create.

In addition, the guide examines the nuances of Russian propaganda: its characteristics, tactics, and how it manifests itself in content created with the help of artificial intelligence. This understanding is crucial for content creators to effectively identify and counter deceptive narratives.

We also examine the ethical considerations and responsibilities associated with the use of artificial intelligence in content creation. In doing so, we emphasise the need to strike a balance between harnessing the benefits of artificial intelligence and maintaining journalistic integrity and objectivity.

In the new digital era, the responsibility to stand up for the truth and fight disinformation is more important than ever. This guide is not just a resource, but a call to action. It encourages each of us to be more aware, critical and ethically responsible about the content we create and consume in a world increasingly shaped by artificial intelligence.

JOURNALISM AND CONTENT CREATION IN THE DIGITAL AGE - BALANCING RISKS AND BENEFITS

The ongoing war between Ukraine and Russia is an important example of the challenges and opportunities faced by journalists and content creators in the digital age. The war has highlighted the need for a careful balance between the use of technological advances and the management of risks associated with disinformation and propaganda.

The dual nature of artificial intelligence in content creation

AI applications are becoming more and more common in content creation, providing great opportunities - from automating routine tasks to writing complex research. However, this convenience carries significant risks when dealing with sensitive content related to the Ukraine-Russia war.

Many people, enthusiastic about the efficiency and novelty of AI tools, have begun to implement them in their workflow. Unfortunately, a noticeable lack of media literacy and a comprehensive understanding of the capabilities of these technologies has led to a critical gap. In many cases, both consumers and creators of online content fail to verify the accuracy of information generated by artificial intelligence. It is the lack of fact-checking before publishing such materials that can inadvertently contribute to the spread of propaganda, resulting in false narratives being widely disseminated.

Combating propaganda and disinformation

Effectively countering the Russian propaganda machine, which skilfully uses digital platforms and AI tools to create and disseminate misleading narratives, requires a comprehensive and multifaceted approach. Firstly, there is an urgent need for content creators to deepen their understanding of AI technologies, recognising their limitations and potential bias. Media literacy programmes

play a crucial role here by emphasising the importance of vetting AI-generated content. In addition, journalists and content creators should follow strict fact-checking protocols, especially for war-related content. This process involves thoroughly cross-checking information from multiple reliable sources and maintaining a healthy scepticism of unverified AI-generated content.

At the same time, it is crucial to highlight the adherence to ethical standards when utilising AI to generate content. Creators must be fully aware of the ethical implications of disseminating unverified information and the ways in which artificial intelligence can be manipulated for propaganda purposes.

In addition, public awareness campaigns are essential to educate people about the nature of propaganda and how to recognise it. This includes helping the public understand how AI can be used to create «deep fakes» and other forms of misleading content. Together, these efforts form a robust strategy to combat the sophisticated use of AI to spread disinformation and protect the integrity of information in the digital age.

1. UNDERSTANDING ARTIFICIAL INTELLIGENCE

WHAT IS ARTIFICIAL INTELLIGENCE?

John McCarthy

The term artificial intelligence (AI) was coined by John McCarthy, Professor Emeritus at Stanford University, in 1955. He defined AI as «the science and engineering of creating intelligent machines». This definition is broad and covers various aspects of AI, including machine learning, natural language processing, robotics, and problem solving.

In the simplest sense, artificial intelligence (AI) is a branch of computer science that deals with the creation of machines (programs) capable of intelligent behavior. It is about developing systems that can think, learn and act independently - from simple algorithms that solve specific tasks to complex systems that mimic human intelligence.

AI encompasses a range of technologies, from machine learning algorithms to more complex systems such as deep learning and generative AI. Machine learning allows computers to learn and make predictions from data, while deep learning, a subset of machine learning, involves neural networks with multiple layers that can learn from large amounts of data. Generative AI, which has attracted considerable attention, is capable of creating content, including text, images and video, which is often indistinguishable from human-generated content.

ARTIFICIAL INTELLIGENCE TECHNOLOGIES CAN BE:

A) MACHINE LEARNING

What is machine learning?

Machine learning (ML) is a part of artificial intelligence that focuses on creating systems that can learn and make decisions based on data. Unlike traditional programming, where a computer follows explicitly programmed instructions, ML allows computers to learn and adapt based on experience without being explicitly programmed for each task.

How does machine learning work?

Essentially, machine learning involves feeding large amounts of data into algorithms. These algorithms then analyse and identify patterns in the data. Based on these patterns, the system makes predictions or decisions about new data it encounters.

There are several types of machine learning methods:

1) Supervised learning

The algorithm is trained on a labelled dataset, which means that the data is already labelled with the correct answer. The goal is to learn a pattern so that the model can make predictions for new, unseen data. For instance, a spam filter for email is an example of this. The algorithm is trained on a database of emails, each of which is labelled as «spam» or «not spam». By learning from these examples, the model can then predict whether a new email is spam.

2) Unsupervised learning

The algorithm is used on data without explicit instructions and tries to identify patterns and relationships in the data on its own. For example, Spotify's recommendation system analyses users' listening habits to identify patterns in which users who listen to certain songs also tend to listen to other songs, and uses this to recommend new songs to users.

3) Reinforcement learning

The system learns by trial and error, receiving feedback from its actions and adjusting its course accordingly. As an example, consider an AI that learns to play complex video games such as chess or Go. The AI starts with random moves but receives feedback based on victory or defeat. Over time, it learns strategies that increase its chances of winning by adjusting its gameplay based on the results of each game.

B) DEEP LEARNING

What is deep learning?

Deep learning (DL) is an advanced form of machine learning that uses neural networks with multiple layers (hence the term «deep»). These layers consist of nodes that mimic the neurons of the human brain. Each layer processes certain aspects of the input data and passes it on to the next layer, gradually refining and improving the decision-making process.

How does deep learning work?

In deep learning, a model learns to perform tasks directly from text, images, or sound. These models are trained using large labelled data sets and neural network architectures that can learn functions and tasks directly from the data. The «deep» in deep learning refers to the number of layers through which the data is transformed. The more layers, the more complex patterns and relationships can be learned. Voice assistants such as Siri or Alexa use deep learning for natural language processing and speech recognition. They analyse voice data to understand spoken commands, learning to recognise different accents and speech patterns over time.

C) GENERATIVE AI

What is generative AI?

Generative AI is a subset of AI algorithms designed to create new content. This can be text, images, video, and audio. Unlike other AI models, which are mainly used for analysis and forecasting, generative AI models can create new content that mimics human creativity and complexity. This is achieved by learning from large data sets and understanding underlying patterns and structures. Examples of generative AI include well-known applications such as ChatGPT, OpenAI's DALL-E, Bing AI, etc. These AI models are able to generate human-like text or realistic photos based on the data they receive.

How does generative AI work?

Generative AI works with algorithms such as generative adversarial networks (GANs) and variational autoencoders (VAEs). These models essentially learn from a large amount of input data, learn patterns, and then use this knowledge to create new, original results. For example, a generative AI trained on news articles can create completely new articles on similar topics.

Application in content creation

Generative AI can automatically create written content, such as reports, summaries, or even entire articles, which can be particularly useful for covering rapidly evolving situations.

Generative AI can create realistic images and videos that can be used in digital storytelling to increase the visual appeal of content.

Generative AI can tailor content to individual preferences, optimising reader engagement and improving the experience.

Challenges and ethical considerations

One of the biggest problems with generative AI is the possibility of creating persuasive but false content. This poses significant risks in journalism, where the accuracy of information is of paramount importance, especially in covering conflicts and warfare.

Generative AI models can perpetuate biases present in their training data. This can lead to a distorted or unfair representation of the content created.

1.1. DEEPFAKES

WHAT IS DEEPFAKE?

Deepfake is synthetic media that replaces a person with someone's else's likeness, often using artificial intelligence techniques such as deep learning. These technologies make it possible to create audio and video clips that are incredibly difficult to distinguish from real content. The term «deep» comes from «deep learning», a form of AI that uses neural networks to process data and create these hyper-realistic results.

The consequences of Deepfake in journalism

1. In the field of journalism, Deepfakes can be used to create false narratives or misrepresent events, individuals or statements. For example, a Deepfake could depict a political leader making a statement that he or she did not actually make, potentially affecting public opinion or diplomatic relations.

2. The authenticity of audio and video materials is difficult to verify due to the sophistication of fake materials. In conflict zones and in times of war, where propaganda and information warfare prevail, distinguishing between genuine material and deep fakes becomes a critical task.

3. The possibility of Deepfake being used to spread false information can lead to a general erosion of trust in the media. If the audience cannot distinguish between genuine and manipulated content, it can lead to scepticism and moreover doubt even about legitimate news sources.

Detecting Deepfakes is an increasingly important skill in journalism and content creation, requiring a keen understanding of both the technology that creates them and the tools developed to detect them.

With the development of «deep fake» technology, distinguishing authentic content from manipulative content requires careful observation and the use of specialised tools. Here is an overview of how to detect Deepfakes, as well as an understanding of the technologies used to create and detect them.

AI-generated image

TIPS FOR DETECTING VIDEO FAKES:

AP

Pay attention to facial features.
Look out for facial expressions that are not natural or unnatural. Inconsistencies in the direction of gaze can also be a sign of lying.Pay particular attention to the ears, as they can sometimes appear distorted, unnaturally placed or have jewelry on only one side.

Evaluate the lip sync.
Discrepancies between spoken words and lip movements may indicate a Deepfake.

Analyse the lighting and shadows.
Inappropriate lighting or shadows that do not match the physical environment may indicate manipulation of the content.

Examine your hair and skin.
Unusual textures or patterns on the hair and skin, which are often a problem for deep learning algorithms, can be incriminating.

Voice audit.
Listen for any differences in tone, pitch, or accent that may differ from the person's known characteristics.

Analysis of the background.
Look for anomalies or unnatural changes in the background scenery.

Check for digital artefacts.
Pixelation or compression errors, especially around the edges of the face, can indicate a deep forgery.

Use technical tools.
Use artificial intelligence-based detection tools that analyse video for inconsistencies that are difficult to detect with the naked eye.

Cross-references with sources.
Compare questionable content with verified material for authenticity.

KNOWN APPS AND PROGRAMS FOR CREATING AND DETECTING DEEPFAKE

Apps for creating fakes: DeepFaceLab, FaceSwap, ZAO and Reface demonstrate how easy it is to create Deepfakes. Most of the apps are designed for entertainment purposes, for example, the Ukrainian app Reface allows you to create entertaining videos that overlay users' features with famous faces from the film or music industry. Some apps, such as Midjourney, are not designed to spread disinformation, but their videos and photos are often used for this purpose.

Detection tools: Microsoft Video Authenticator, Deepware Scanner, Adobe Content Authenticity Initiative (CAI), Sensity, Intel, and TruePic all offer ways to detect and authenticate digital content.

So, while Deepfake technology poses challenges in areas such as journalism, the development of sophisticated detection tools offers a way to combat its misuse. However, it is crucial that these tools are used responsibly and in conjunction with traditional verification methods to preserve the authenticity and credibility of digital content.

1.2. WHY DOES ARTIFICIAL INTELLIGENCE GENERATE DISINFORMATION?

In this technological world, where digital information is created and disseminated on a massive scale every second, it is important to critically assess the role of generative artificial intelligence (AI) in the context of information accuracy and integrity. The omnipresence of digital media has made it easier to distribute content on a global scale, but this ease of dissemination also comes with the risk of rapid and widespread misinformation.

Generative AI, especially in the form of deep learning models, is capable of creating highly realistic images, videos and text. While this ability is revolutionary and valuable for various creative and educational applications, it also opens the door to potential misuse. The technology can produce content that is increasingly difficult to distinguish from reality, which poses significant challenges for information verification.

So why is a technology with such transformative potential acting as a conduit for disinformation? This understanding goes beyond mere technical necessity; it is a cornerstone of maintaining journalistic integrity and accuracy, especially in an era of war, when the role of artificial intelligence is growing significantly. Several factors contribute to the generation of false information:

DATA FROM OPEN SOURCES.

The main reason for this is the nature of the data used to train AI. Generative AI models are typically trained on large data sets obtained from open, publicly available sources. While this approach allows AI to learn from a wide range of content, it also carries significant risk. Open-source data can often contain misinformation, inaccuracies, and biased viewpoints. When AI models are trained on such data, they inadvertently learn and reproduce these inaccuracies in their results. Real-world examples can illustrate this more clearly:

a) Social media content:

AI models trained on social media data may inadvertently learn from posts or comments that contain unverified information, rumours, or subjective opinions. For example, in the early stages of the Ukraine-Russia war, social media was full of unverified reports and personal interpretations of events, which, if used for training, could lead to AI generating misleading narratives.

b) Online forums and discussion boards:

These platforms, such as Reddit, often contain a mix of factual information, personal anecdotes,

and speculative content. If an AI model is trained on such a mix, it may not be able to distinguish between verified facts and speculative discussions. For example, online forums discussing the political motivations behind the Ukraine-Russia war may contain a mixture of accurate historical data and biased political opinions.

c) News websites with different editorial standards:

An AI trained on news articles from websites that do not adhere to strict journalistic standards may reproduce the bias or inaccuracies present in those articles. If an AI system is constantly fed news from sites that tend to be sensationalistic or do not conduct thorough fact-checking, it may create content that reflects these shortcomings.

HALLUCINATIONS IN ARTIFICIAL INTELLIGENCE.

Another important reason is the inherent ability of AI, especially generative models, to create so-called «hallucinations». This term refers to cases when AI systems generate false, misleading, or completely fabricated information that does not correspond to real facts or logical sequence. These hallucinations occur due to a variety of factors, including the complexity of language processing, the limitations of AI algorithms, and the quality of training data. AI hallucinations can be particularly problematic in journalism, as they can lead to the unintentional spread of disinformation, thereby undermining the credibility of content. Practical examples of AI hallucinations include:

a)Fabricated historical events:

An AI model trained on inaccurate historical data or fictional narratives can generate articles or reports that refer to non-existent events. For example, AI can create a story about a fabricated military operation during the Ukrainian-Russian war, making readers believe in events that never happened.

b) Fictitious quotes or statements:

AI models can generate realistic quotes from public figures that have never actually been uttered. For example, an AI model could hallucinate a statement attributed to a world leader or military official regarding the Ukraine-Russia war, potentially causing confusion and spreading false information.

c) Distorted statistical data:

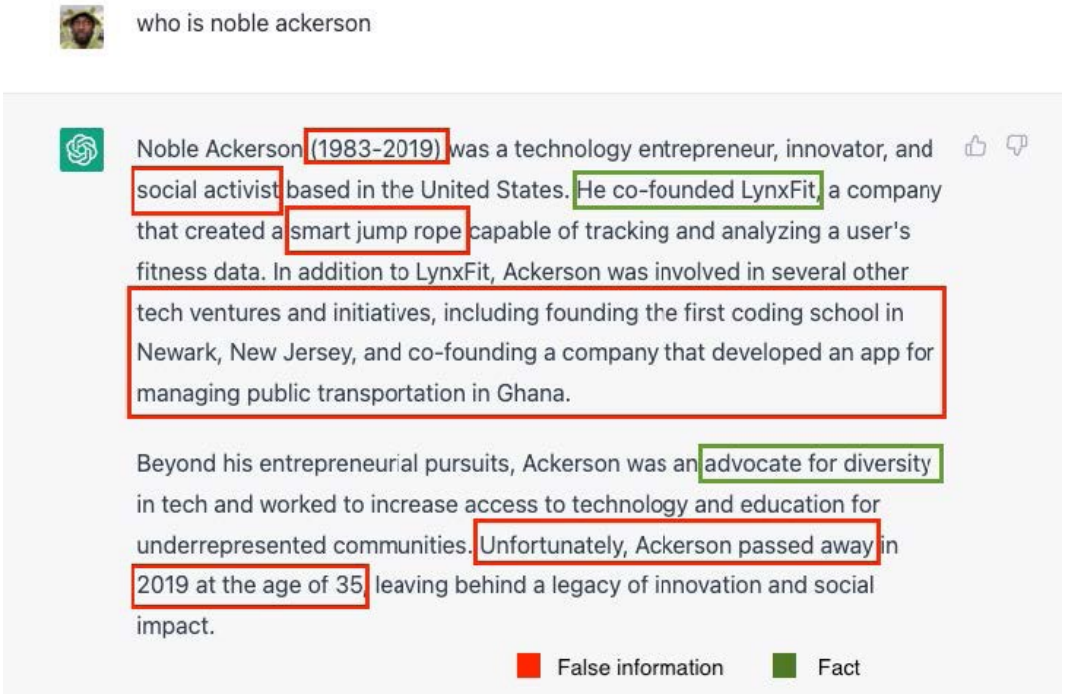
In cases where AI is used to analyse and present statistical data, hallucinations can lead to the generation of incorrect numbers. This can be especially misleading in journalism when it comes to sensitive issues such as the number of casualties or the number of refugees in a war.

d) Distortion of terrain or geopolitical facts:

AI may misrepresent geographic or political realities, such as misrepresenting the location of a key event in the Ukraine-Russia war or inaccurately describing territorial control over certain regions.

e) False images or videos:

In the realm of deep fakes, AI can create convincing but completely false images or videos. This can be particularly damaging in journalism, where visual evidence is often crucial. An AI-generated video falsely depicting military operations or events in the Ukrainian-Russian war could have serious consequences.



The combination of these factors makes journalists and content creators using generative AI extremely cautious. Understanding the origin and nature of the data on which these AI models are trained, as well as an awareness of the phenomenon of AI hallucinations, is crucial. This awareness should be coupled with rigorous fact-checking processes and critical evaluation of AI-generated content, especially when dealing with sensitive and influential topics such as the Ukraine-Russia war.

To ensure accuracy and reliability in their work, journalists and content creators must couple their awareness of the complexities of using generative AI with rigorous fact-checking processes and critical evaluation of AI-generated content.

2. DISINFORMATION ON SOCIAL MEDIA

This section of the guide is dedicated to helping content creators, journalists and social media users identify and counter propaganda created and disseminated by AI.

Russia's disinformation campaigns on social media are a key concern, particularly evident in scenarios such as the Ukraine-Russia war and the 2016 US presidential election. These campaigns often mix truth with fiction, creating narratives that serve strategic interests, manipulate public opinion and sow division. For example, during the Ukraine-Russia war, social media saw a surge in false images and fabricated stories designed to influence the perception of the war.

The integration of artificial intelligence technologies has further complicated this landscape. AI-driven bots and fake accounts that can mimic real users are spreading disinformation at an alarming rate and scale. In addition, AI algorithms are able to create persuasive content, such as deep fakes, and target disinformation to specific demographic groups. This makes detecting and countering disinformation not only more challenging, but also more important.

The Cambridge Analytica scandal is a vivid example of how artificial intelligence can influence public opinion. In the 2010s, the British consulting company Cambridge Analytica collected personal data of millions of Facebook users without their consent, mainly for use in political advertising. In this case, the personal information of countless Facebook users was collected and combined with artificial intelligence techniques to create highly targeted political ads. This situation highlights the ethical dangers associated with the use of AI and data analytics to influence public opinion. AI's ability to perform precise targeting is also an important factor in the spread of disinformation. By processing huge volumes of user data, AI systems can accurately identify individuals for individual disinformation campaigns.

Countering such sophisticated disinformation requires not only technological solutions, but

also a comprehensive approach. Increasing media literacy is crucial to ensure that the public can identify and understand the tactics used in disinformation campaigns. Social media platforms also need to take responsibility by improving the detection and removal of fake accounts and cooperating with fact-checkers to ensure transparent content moderation.

In addition, the regulatory environment should evolve to hold social media platforms accountable for the content they disseminate and protect user data from exploitation. Governments and international organisations should consider regulations that address these issues, balancing the need for freedom of expression with the imperative to preserve the integrity of information.

Finally, this section highlights the importance of vigilance, technological literacy and ethical standards in the complex interplay between Russian propaganda and artificial intelligence on social media. For journalists, content creators, and social media users, understanding these dynamics is key to effectively countering disinformation and upholding the truth in our increasingly interconnected world.

2.1. FACEBOOK, X AND OTHER SOCIAL PLATFORMS

Social media platforms such as Facebook and X (formerly Twitter) have become key players in the AI-driven disinformation landscape. The emergence of generative AI has greatly enhanced the proliferation and sophistication of disinformation campaigns. These campaigns are increasingly using AI tools to create and disseminate misleading information on an unprecedented scale. These media platforms also face the challenge of moderating AI-generated content. Their algorithms sometimes inadvertently amplify sensationalist or misleading content through higher engagement rates, thereby spreading disinformation further.

Meta has reported various instances of coordinated inauthentic behaviour and cyber espionage on its platforms, including the disabling of accounts associated with covert influence operations and disinformation campaigns.

For example, in 2023, the company announced the dismantling of what it described as the largest covert influence operation on various platforms. This large-scale campaign, which involved thousands of accounts, targeted more than 50 platforms, including Facebook, Instagram, X (formerly Twitter), YouTube, TikTok, and Reddit. The operation primarily disseminated pro-Chinese content, including positive reviews of Xinjiang province, and criticised the United States, Western foreign policy, and those perceived to be opponents of the Chinese government.

Meta's investigation revealed the use of coordinated fake accounts that operated on a regular schedule and were likely managed by the same team in a shared space. They shared similar content on different platforms with misleading headlines such as «US bombing of

Nord Stream - the first step in a «plan to destroy Europe» Despite the scale of the operation, Meta noted that it failed to attract a significant number of real followers, instead gaining fake followers from regions outside its target audience.

In addition, Meta noticed that since 2019, the Spamouflage network has shifted towards smaller platforms. In its report, the company emphasises that it continues to counter other covert influence operations, including those targeting Türkiye and the Russian campaign spreading disinformation about the war in Ukraine, which is now expanding its activities to the United States and Israel. Impersonating the media was a common strategy for these campaigns.

On the social network X (formerly Twitter), numerous fake profiles such as «Bella Morne» use artificial intelligence to create convincing images. These accounts, which have tens of thousands of followers, use AI-generated images resembling models to create their online identity. They strategically generate revenue as «content creators» by posting emotionally charged content on sensitive topics such as the situation in Palestine and Israel. In addition, these accounts are known for expressing support for Donald Trump, which further strengthens their presence and influence on the platform.

The European Commission has recently cautioned Elon Musk and his Twitter platform X to comply with new laws on fake news and Russian propaganda. The warning comes after X was found to have the highest percentage of disinformation posts among major social media platforms. A report on the spread of fake news in the EU says that millions of fake accounts have been removed by TikTok and LinkedIn, with Facebook ranking second among the biggest offenders.

According to the Digital Services Act (DSA), which came into force in August 2023, posts classified as disinformation will be illegal across the EU. Facebook and other tech giants such as Google, TikTok and Microsoft have signed up to the EU code of practice to comply with these laws. Twitter, however, has withdrawn from this code but still needs to comply with the DSA or face a ban in the EU.

Generative AI tools have become more accessible and have made it easier to create and disseminate disinformation on a massive scale. For example, Venezuelan state media used AI-generated videos of news anchors from a non-existent international English-language channel to spread pro-government messages. Similarly, in the United States, videos and images of political leaders created with the help of artificial intelligence were shared on social media, including a video of President Biden making transphobic comments and an image of Donald Trump hugging Anthony Fauci.

TikTok is another platform where propagandists are skillfully using AI to create manipulative content. This is, for example, one of thousands of similar pages on TikTok with enticing screensavers and AI-generated content. The page was created by propagandists and promotes harmful messages for Ukrainian society:

Forced mobilisation. Through storytelling, the protagonist (and at the same time the user) develops a sense of fear, doom, and injustice.

Corruption sown by MPs. Here, too, the goal is to increase distrust of the authorities, to evoke

a sense of injustice in the user and the thought «what are we fighting for». These narratives are then actively and unconsciously picked up by bloggers, spreading the message that «everyone steals».

Bribe-taking commanders. Here, the goal is to increase distrust of Ukraine's military leadership. Such posts are often posted under hashtags such as: #ukraine #tc #mobilisation #war #afs #corruption #deputy

In the vivid images generated by artificial intelligence, the military is depicted as well-fed and insatiable for power. It is as if they enjoy mobilising ordinary citizens, who are sitting drearily in grey public transport in the background. In this way, in addition to creating despondency, propagandists deliberately create a barrier and divide society with the help of technology and creativity. And this, of course, weakens Ukraine's resilience in the war.

AI generates not only visually appealing propaganda content. Many of the fake videos on TikTok are edited from various cuts of news or entertainment programmes on popular Ukrainian TV channels. They are often voiced by artificially generated voices of well-known presenters. This conclusion was reached by the Institute of Mass Information in a study.

Such videos are usually designed in large fonts and bright colours. Sometimes, the captions overlap the faces of the presenters in the frame. This makes it difficult to see the discrepancy between the voiceover and the facial expressions of the people in the frame.

In addition, the availability and cheapness of generative AI lowers the barrier to entry for disinformation campaigns, allowing not only state actors but also various groups to engage in these activities. The proliferation of AI-generated content on the Internet has also led to the phenomenon of a «liar's dividend», whereby alertness to falsification makes people more sceptical of truthful information, especially during times of crisis or political conflict.

DETECTING DISINFORMATION AND FAKE ACCOUNTS WITH ARTIFICIAL INTELLIGENCE

Detecting disinformation and fake accounts on social media managed by artificial intelligence is becoming increasingly difficult. However, there are certain approaches and methods that can be used:

Content analysis. A thorough check of the content for logical inconsistencies or sensationalistic headlines, as well as verification of the source's credibility.

Profile verification. Check user profiles for signs of a bot or fake account, such as lack of personal information, recent creation date, and unusual posting history.

Interaction patterns. We look at distorted follower-to-follower ratios and repetitive, inauthentic interactions that point to bots.

Technical tools. Using artificial intelligence-based detection and analytics tools to analyse account behaviour and content for signs of manipulation.

The growing sophistication of disinformation driven by artificial intelligence requires increased public awareness and a collective effort by social media, governments and civil society to effectively combat this challenge.

2.2. INCREASING MEDIA LITERACY: STRATEGIES FOR A MORE INFORMED SOCIETY

Increasing media literacy is an important strategy in building a better-informed society, especially in an era when disinformation driven by artificial intelligence is spreading on social media platforms. By increasing media literacy, people become more able to critically evaluate the information they encounter and make informed decisions. Here are some strategies with examples:

2.2.1. Question what you see and hear

In practice: Every time you come across a news story or video, such as a supposed clip of a politician, ask yourself: Does it look realistic? Who is spreading this information and why? Look for similar news on reputable websites to see if it has been reported elsewhere.

2.2.2 Learn to understand your emotions and unconscious behaviour

Practice: Healthy self-doubt about your media literacy skills will help you keep your head up. In fact, our brains do many things unconsciously. It has been proven that a fake that a person hears, even in the context of a refutation, can become familiar to the brain. The feeling of «familiarity» subconsciously inspires trust. Therefore, the next time you come across a similar thesis, you may get the impression that you have already heard or seen it somewhere, and you will be more likely to believe the manipulative information. Keep this in mind when consuming content.

2.2.3. Learn how fake content is created

In practice: Take some time to understand how fake news and Deepfakes are created. For example, in fake videos, people do or say things they have never done before. This awareness will help you stay cautious and not believe everything you see.

2.2.4. Diversify your news sources

In practice: Don't rely on just one social network or media outlet for all your information. Follow different news channels, websites, and even international

media to get a wide range of perspectives. This will help you avoid the trap of only hearing the views that match your own.

2.2.5. Use websites for fact-checking

In practice: Before sharing unexpected news or a shocking image, use sites like Snopes, FactCheck.org, or others to check its authenticity. This is especially important during major events or elections, when disinformation spreads at a high rate.

2.2.6. Be sceptical of sensationalist headlines

In practice: Headlines that sound overly dramatic or evoke a strong emotional response are often designed to get clicks rather than inform. Read on, and if the content doesn't support the headline, it's likely to be misleading.

2.2.7. Explore digital tools

Practice: Learn simple digital verification techniques such as reverse image searching (using tools like Google Images) to check the origin of a photo or video. It's a quick way to find out if an image from a «current» event is actually recycled from an older one.

2.2.8. Participate in discussions and ask questions

In practice: If you are not sure about some information, discuss it with friends or family. Sometimes discussion can open up new perspectives or encourage others to think critically.

2.2.9. Trust trusted experts

In practice: quite often, news in the media is based on someone making a statement, predicting or analysing something. And it is this forecast or analysis that can cause you to feel strong emotions, indignation or disagreement. So pay attention to who said it, whether the person has the relevant experience to comment on such things, what organisation they work for, etc. The same advice applies to checking the author who wrote a particular news story or column on the site. His or her opinion is only his or her opinion, and it may not reflect the reality. For example, Russian so-called experts or political analysts often broadcast their own picture of the world, promote the necessary propaganda narratives, and do not really analyse the situation.

2.2.10. Maintain transparent social media practices

In practice: Be aware of how social media platforms work and advocate for clear policies to counter disinformation. Use reporting tools to identify fake news and support initiatives that aim to make information sharing more transparent and truthful.

3.
THE RESPONSIBILITY OF ARTIFICIAL INTELLIGENCE IN SPREADING DISINFORMATION

Tracing responsibility for the spread of AI-driven disinformation is a critical task that involves examining the roles of different stakeholders in the digital ecosystem. The complexity of artificial intelligence technologies and their integration into social media platforms has led to unprecedented challenges in detecting and combating disinformation.

Who is responsible when AI spreads disinformation? This is an important question as we seek to comprehend the ethical and practical implications of AI in our information landscape.

1. AI developers and technology companies - responsible for the ethical development of AI, including protection against misinformation. Their role is to ensure that AI technologies are used responsibly and ethically, preventing misuse to spread false information.

2. Social media platforms - responsible for content moderation. They should cooperate with fact-checkers to identify and reduce the spread of false information by maintaining vigilant content management.

3. Government and regulatory bodies play a key role in formulating and implementing policies to counter the misuse of AI in disinformation, while striking a balance between freedom of expression and privacy protection.

4. Media and fact-checking organisations - need to adapt to AI-driven disinformation by using advanced fact-checking and content tools to combat false narratives.

Shared responsibility. Combating AI-driven disinformation requires a concerted effort from all of these groups. This includes ethical AI development, effective content moderation, regulatory oversight, thorough fact-checking, and public education to create a well-informed and critical audience.

On the other hand, artificial intelligence technologies are also used to counter disinformation. Artificial intelligence systems developed by online platforms greatly contribute to the efficient and rapid spread of disinformation, but they are also used to detect and reduce the spread of false information online.
Scholars also emphasise the need to take into account ethical considerations when using AI to create and distribute content. Ethical implications arise when AI can be used both to create and disseminate disinformation and to combat it. This duality requires careful consideration of the role of AI in the modern information ecosystem, especially in light of the protection of fundamental rights and freedoms, including freedom of expression and information.

Consequently, the liability of AI in the context of disinformation presents a complex challenge, necessitating a balanced assessment of both the potential harm and the opportunities to combat disinformation. It is imperative that the development and deployment of AI technologies in this domain are guided by ethical principles and a firm commitment to safeguarding the integrity of information and democratic processes.

3.1 KEY LEGAL DOCUMENTS

In the context of AI and disinformation, several European legal instruments and frameworks have been developed to address the challenges posed by these technologies. These documents are aimed at regulating the use of AI, protecting the rights of citizens and ensuring the responsible dissemination of information. Here are some of the most significant legal frameworks and documents:

The EU Code of Practice on Disinformation.

The EU Code of Practice on Disinformation, established in 2018 and strengthened in 2022, is a self-regulatory framework aimed at combating disinformation online. It includes commitments from online platforms, trade associations and key players in the advertising sector. The Code aims to ensure greater transparency and accountability of online platforms and offers a structured framework for monitoring and improving platforms' disinformation policies. It includes specific measures such as strengthening efforts to demonetise disinformation, increasing transparency of political and issue-based advertising, enabling users to identify and flag false content, expanding fact-checking across the EU, and providing researchers with greater access to data. The Code has also established a permanent working group to develop and adapt the measures, ensuring that it remains responsive to the dynamic nature of disinformation.

The General Data Protection Regulation (GDPR).

While the GDPR is primarily focused on data protection and privacy, it has implications for AI and disinformation, especially with regard to the use of personal data in microtargeting and profiling.

The GDPR sets out strict rules on the collection, storage and use of personal data that affect how AI algorithms can use this data to create targeted content. This regulation helps to prevent the misuse of personal data for disinformation campaigns by ensuring that any data-driven AI activity, especially those related to personal profiling and targeting, meets strict privacy and consent standards.

The EU Artificial Intelligence Act (AI Act).

The EU Artificial Intelligence Law proposed in 2021 is an important legislative step towards regulating AI in the European Union. It aims to set standards for the development and deployment of AI systems, ensuring that they comply with EU values and fundamental rights. The law classifies AI applications into categories based on their potential risk to human rights and security. High-risk categories include AI systems used to manipulate information, which is a direct response to concerns about the use of AI to spread disinformation. This classification system

underlines the EU's commitment to mitigating the risks associated with AI technologies, especially those that could affect democratic processes and public security.

The Digital Services Act.

The Digital Services Act (DSA) proposed by the European Commission aims to create a more secure and accountable digital space in the EU. The law focuses on protecting the fundamental rights of users on the Internet and provides for measures to increase transparency of the algorithms of online platforms. It also addresses the issue of illegal content and disinformation, setting out clear obligations for digital service providers to address these issues. The DSA is an important step towards regulating the digital space, ensuring that it remains a safe and secure environment for users.

AI White Paper.

The UK White Paper on online harm outlines the government's plan to create a legal framework to tackle illegal and harmful content online, including disinformation. It aims to introduce a statutory duty to protect users from harmful content, which will force internet companies to protect users from harmful content. This initiative is an important step in regulating the online space, ensuring safer internet use and combating harmful content, including disinformation, to maintain a safe and secure digital environment.

These legal instruments reflect the growing recognition of the need for a robust legal framework to manage the challenges posed by the spread of disinformation through artificial intelligence. They aim to balance innovation and technological progress with the protection of individual rights and the integrity of public discourse.

In 2020, Ukraine approved the Concept of Artificial Intelligence Development. This initiative highlights the need to refine the legal framework governing the development of artificial intelligence. The Expert Committee on the Development of Artificial Intelligence in Ukraine, under the Ministry of Digital Transformation of Ukraine, is actively working on creating a legal framework for the regulation of AI. These efforts include addressing the problems associated with the creation and dissemination of disinformation using artificial intelligence, ensuring that these important issues are taken into account when developing new regulations. The committee also considers it necessary to take into account the experience of the United Kingdom, which has released a White Paper on Artificial Intelligence that describes the government's approach to balancing regulation and stimulating the development of artificial intelligence, as well as provides a better understanding of the vector of artificial intelligence development for society and business.

4. ETHICAL GUIDELINES AND RESPONSIBILITIES OF JOURNALISTS, AI DEVELOPERS AND MEDIA PLATFORMS IN THE ERA OF ADVANCED AI TECHNOLOGIES

In an era dominated by advanced artificial intelligence technologies, journalists, AI developers and media platforms face unique ethical challenges and responsibilities. Journalists must adhere to strict standards of fact-checking and transparency, especially when using AI-generated content. AI developers, in turn, must ensure that their creations avoid bias and misinformation by keeping ethical considerations at the forefront of the development process. Media platforms must strike a delicate balance between protecting freedom of speech and preventing the spread of disinformation by using transparent AI algorithms to moderate content. Adherence to these standards is crucial for maintaining trust and integrity in the digital information sphere.

FOR JOURNALISTS

a. Ensure that all information is accurate, especially if it is de-duplicated by AI-based platforms.

b. Be transparent and clearly disclose the use of AI in your content creation.

c. Recognise and reduce bias in AI-generated content.

d. Adhere to ethical standards when collecting and reporting information, especially when using AI to analyse data.

e. Avoid sensationalism and respect the dignity of subjects, especially when AI is helping to create content.

FOR ARTIFICIAL INTELLIGENCE DEVELOPERS

a. Develop AI with ethical considerations in mind, including fairness, accountability and transparency.

b. Continuously work to identify and reduce bias in AI algorithms.

c. Ensure robust data protection measures in AI systems.

d. Be clear about the capabilities and limitations of AI technologies.

e. Engage with journalists, media platforms and the public to understand and address ethical issues.

MEDIA PLATFORMS

a. Implement content moderation and apply policies to AI-generated content, ensuring that it meets ethical and journalistic standards.

b. Clearly identify AI-generated content and its sources.

c. Protect user data and privacy, especially when AI is used for personalisation and analytics.

d. Educate the public about the role and impact of artificial intelligence on content creation and distribution.

e. Make sure AI-powered ads are transparent, fair, and respect users' privacy.

CONCLUSIONS

In the digital age, especially in the context of the Ukraine-Russia war, the role of journalists and content creators is more important than ever. While technology offers unprecedented opportunities for storytelling and reporting, it also presents significant challenges. Balancing these aspects requires a commitment to the ethical principles of journalism, critical analysis of AI-generated content, and ongoing efforts to educate both creators and the public.

This guide explores the complex interplay between artificial intelligence, journalism and disinformation. Particular attention is paid to understanding artificial intelligence technologies such as machine learning, deep learning and generative artificial intelligence, as well as the challenges posed by deep fakes and AI-driven disinformation. The key role of social media platforms in the spread of disinformation is considered, and the importance of media literacy is emphasised.

Finally, the guide outlines the ethical responsibilities of journalists, AI developers and media platforms that promote responsible practices in the digital age. This guide emphasises the need for vigilance, ethical awareness and continuous learning to preserve the integrity of information in our rapidly evolving digital world.