

iEND GAME!

Escaping new disinformation through gamified cross-border media literacy education



Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.



Co-funded by
the European Union

Title

Media literacy teaching materials accompanying escape room 2 “ECHO”

License

This document is licensed under a Creative Commons Attribution 4.0 International License



This document has been developed within the scope of the **ENDGAME project** co-funded by the European Union under the Creative Europe funding instrument.

Project reference: 101185763

Project website: <https://endgameproject.github.io/>

The opinions and viewpoints expressed are solely those of the authors and do not necessarily reflect those of the European Union.



**Co-funded by
the European Union**



**Co-funded by
the European Union**

1. What is ENDGAME?	4
2. Background and general objectives	4
2.1. Artificial intelligence and its role in social media	4
2.2. AI in social media	5
2.3. Mis/disinformation about climate change	5
3. Learning objectives	7
1. "ECHO": theme and narrative	7
2. Implementation strategies	7
2.1. Preparatory session	8
2.2. Closing activity	11
3. Step-by-step guide to the escape room	12
3.1. Game recommendations	12
3.2. Game instructions	12



1. What is ENDGAME?

ENDGAME: Escaping New Disinformation through Gamified Cross-Border Media Literacy Education is a European educational project with a diverse group of partners, including the University of Eastern Finland (UEF), Fundación Maldita.es, Universidad Politécnica de Madrid (UPM), Belgrade Metropolitan University (BMU), Faktabaari, and Fake News Tragač (FNT).

The main objective of this initiative is to **promote media literacy among young people** in Finland, Serbia, and Spain, helping them to develop **critical thinking** and create digitally responsible citizenship. To this end, a series of **interactive escape rooms** has been developed in which participants must solve a series of challenges within a fictional scenario in a limited time.

Through these educational games, which simulate real-life situations in which we consume information, participants will learn to **distinguish between disinformation** and manipulated content, recognise information generated by **artificial intelligence**, and **understand the risks of sharing personal data** on social media.

2. Background and general objectives

2.1. Artificial intelligence and its role in social media

Nowadays, the distinction between human-created and machine-generated content has blurred. **Artificial Intelligence** (AI) has democratized the creation of media, but it has also provided a powerful tool for **mass-producing disinformation**. Some examples are **deepfakes**, which are realistic video or audio recordings of people saying or doing things that never occurred, and **AI-bot networks** that use language models to engage in human-like arguments in comment sections to create a false sense of public consensus. Since these tools are inexpensive and fast, the volume of disinformation is increasing at a rate that human fact-checkers cannot possibly match, leading to a "polluted" information environment. The main engine driving much of this transformation is a specific type of AI known as a **Large Language Model**. Understanding how these models function is the first step in recognizing why AI-generated content can feel so convincing.

A Large Language Model (LLM) is a type of AI trained on vast amounts of text data to understand, summarize, and generate human-like language. Think of it as a highly advanced "auto-complete" system. It doesn't "know" facts the way a human does, but instead **predicts the most likely next word** in a sentence based on the patterns it learned during training. LLMs are designed to be helpful and conversational: they are very good at explaining complex topics or brainstorming ideas. However, they can also **"hallucinate"**, or confidently stating something as fact when it is actually a blend of patterns that never happened in reality.



Watch the video: https://endgameproject.github.io/ECHO/assets/video_IA_en.mp4

The problem extends beyond the hallucinations generated by LLMs or even the intended content generated through them with the goal of spreading misinformation. Many students have started to use LLM-powered tools, such as chatbots, as a direct **replacement for traditional search engines** or other information retrieval methods. However, students must understand that most LLMs are not trained in real time. Unlike a search engine that crawls the live web to find the most current information, an LLM relies on a training dataset that has a specific "**cutoff date**." This means that for breaking news, recent scientific breakthroughs, or evolving political events, the AI may provide outdated information or fill the gaps with plausible-sounding hallucinations.

Moreover, whereas a search engine directs a user to an external website for verification, an LLM frequently presents information as a definitive, **sourceless fact**. Furthermore, search engines allow students to **evaluate the provenance of a claim** by distinguishing between a scientific study and a personal blog. LLMs often blend these sources into a single response, making it impossible for the student to judge the credibility of the underlying data. This **black-box nature** of AI makes it difficult for students to double-check logic or identify specific biases.

2.2. Content moderation

In a world where millions of posts are uploaded every minute, content moderation acts as the internet's "immune system." However, as AI makes it easier to create convincing fakes, this system is under more pressure than ever.

What is Content Moderation?

At its core, content moderation is the process by which social media platforms monitor, review, and apply a set of rules (Community Standards) to user-generated content. It is a constant balancing act between two competing values:

- **Safety:** Protecting users from harm, harassment, and exploitation.
- **Expression:** Upholding the principle of free speech and open dialogue.

Moderation isn't just about **deleting posts**, but rather about deciding what kind of "**digital neighborhood**" a platform wants to be.

Illegal Content vs. Disinformative Content

It is a common misconception that everything "bad" online is illegal. For the purposes of the game, it's vital to distinguish between these two categories:



	Illegal content	Disinformation content
Definition	Content that violates specific laws (e.g., child safety, incitement to violence, copyright theft).	False or misleading information spread with the intent to deceive or cause harm.
Moderator actions	Platforms are often legally required to remove this and, in many cases, report it to authorities. The platform may be held responsible for failure to comply with legal requirements.	Often exists in a "grey area." It may not be illegal, but it may violate platform policies, at least in theory. The moderator may or may not respond.
Examples	Hate speech (in certain regions), terrorist or extremist recruitment, fraud, and scams.	AI-generated "deepfake" videos of politicians, fake health cures, or staged "breaking news."

Key Takeaway: Disinformation is often "lawful but awful" – it mimics the truth to manipulate emotions, making it harder for automated systems to catch than simple illegal keywords. And while policies protect the users in theory, these lofty ideals seldom are put into practice: moderation consumes time and money, and to some extent, the business logic of the platforms creates incentives for them to allow disingenuous and emotionally provoking content.

Bots and Trolls

Disinformation rarely travels alone; it is usually pushed by specific actors. One might encounter "users" who aren't what they seem.

- **Bots (The Robots):** Short for "web robots," these are automated programs designed to perform repetitive tasks. In disinformation, bots are used to "like" or "share" a post thousands of times in seconds, tricking the algorithm into thinking a topic is trending. According to [NATO Stratcomm](#), for a mere 10€ investment one can buy e.g. more than 156 000 inauthentic views on X (in 2025).
- **Trolls (The Humans):** Unlike bots, trolls are real people. Their goal is to provoke, upset, or distract. They use emotional triggers to start arguments, making the digital environment so toxic that civil discussion becomes impossible.
- **The AI Upgrade:** Generative AI allows a single "troll" to operate hundreds of "bot" accounts that each sound like a unique, convincing human, creating a "sockpuppet" army.

Available Options in Content Moderation

Platforms don't always use a "delete" button. Effective moderation uses a "sliding scale" of interventions:

- **Removal:** The content is deleted, and the user may be banned. This is usually reserved for the most severe violations.
- **Labeling/Fact-Checking:** The post remains visible, but an overlay is added (e.g., "Context missing" or "Stay informed: This video is digitally altered").
- **Shadowbanning (Reduced Reach):** The algorithm stops showing the content to new people. The user can still post, but they are essentially shouting into an empty room.
- **Demotion:** Pushing the content to the very bottom of a user's feed so it is unlikely to be seen.
- **Human-in-the-Loop:** While AI scans for patterns, human moderators are often used for "nuance" checks—deciding if a post is a dangerous lie or just a harmless joke.

In the context of misinformation, it is generally recommended not to remove it except in cases where it is illegal, as there are more effective ways to reduce its impact. Various studies on this topic indicate that actions such as fact-checking labels help users identify false content and therefore reduce its spread without the need to completely remove it.

2.3. Mis/disinformation about climate change

As for the topics that are misrepresented on social media, although there are many narratives on which false content is spread, the second escape room will focus on **mis- and disinformation about climate**.

Climate change disinformation has evolved. We have moved past simple "science denial" (the claim that the climate isn't changing) into more sophisticated, often **AI-enhanced forms of Climate Delay**. This includes attacking the feasibility of renewable energy, discrediting climate scientists, and promoting "doomism" to discourage action.

Disinformation narratives on climate change as "wedge issues" to:

- **Polarize Societies:** Pitting urban populations against rural ones or younger generations against older ones.
- **Erode Trust:** Undermining confidence in scientific institutions and democratic governments.
- **Economic Sabotage:** Spreading false claims about the cost of green transitions to protect fossil fuel interests or create internal political instability in rival nations.



When discussing disinformation, emphasize that it is often a strategic tool used for political or economic gain, rather than just a misunderstanding of facts. Teaching about disinformation requires a shift from "telling students what is true" to "teaching students how to think".

Research suggests that so-called **Inoculation Theory** (or **pre-bunking**) is more effective than **debunking** (or **correcting falsehoods afterwards**). This involves exposing students to a "weak version" of a manipulative technique so they can recognize and resist it later.

- **Identify the Technique:** Instead of focusing on a specific false fact, focus on the logic (e.g., the use of "fake experts") in order to expose the manipulation strategy that has been used.
- **The Backfire Effect:** Avoid simply repeating a myth. If you must mention one, start with the truth, explain the manipulation, and end with the truth again (The Truth Sandwich).

Disinformation thrives on high-arousal emotions: anger, fear, or outrage. **The main driver behind the spread of disinformation is the activation of strong, reactive emotions by outrageous claims.** This is why "emotional literacy" is the first step in teaching critical evaluation, and why it is so crucial to teach teenagers to pause when they feel a strong emotional reaction to a headline. That reaction is often the "hook" used by manipulators, and can be reinforced by repetition, so that is where education also should intervene.

It's advisable to give your students these practical tools to use whenever they encounter climate claims on TikTok, Instagram, or news feeds:

The SIFT Method:

(S)top: Check your emotions. Do you want this to be true?

(I)investigate the Source: Click the profile. Is this a climate scientist, an advocacy group with an agenda, or a random account?

(F)ind Better Coverage: See if reputable science organizations (like NASA, IPCC, or NOAA) are reporting the same thing.

(T)race Claims back to the Original: Did the post summarize a study correctly, or did they take one sentence out of context?

It would be beneficial to discuss also some common red flags to look for in climate discussions:

- **Cherry-Picking:** Showing a graph of a single cold week to "prove" global warming has stopped, while ignoring the 100-year upward trend.
- **The "False Balance":** Media outlets giving equal airtime to a fringe conspiracy theorist and a climate scientist as if their expertise is equal.
- **Impossible Expectations:** Demanding that climate models be 100% perfect or else they are "useless." (Science deals in probabilities, not perfection).



- **Ad Hominem Attacks:** Attacking the appearance or lifestyle of a climate activist (like Greta Thunberg) rather than addressing the scientific data they present.

We'll present later with more specific activities that can be used in connection with this Escape Room game. But for an example of an additional, practical, and usually very educational activity, you can, e.g., ask the students to work in small groups to "design" a piece of climate disinformation themselves. They must choose a target (e.g., teenagers), a goal (e.g., make them stop believing in electric cars), and a technique (e.g., cherry-picking data). By understanding how the "magic trick" is performed, they may learn the logic of disinformation better and are less likely to be fooled by it in the future.

3. The ECHO escape room

"ECHO" is the second of the escape rooms developed within the ENDGAME project. In this escape room, the player takes on the role of a **social media moderator** who is in charge of stopping a wave of misinformation regarding climate. The player needs to detect **bot accounts**, identify **AI-generated content**, and prevent users from **overrelying on AI**. The escape room aims for players to achieve the following **learning objectives**

- Encourage **critical thinking** among students.
- Reflect on how we use **social media** and the level of **truthfulness** we attribute to the content we see on it.
- Understand **how a Large Language Model works** and how it is used in social media.
- Evaluate **recommended uses of AI** to generate information and content.
- Detect patterns that are typical of **bot accounts**.

To ensure the escape room is a learning experience and not just a game, we recommend complementing it with **preparatory activities in one or more sessions**. These sessions help participants better understand the topic, reflect on their own habits, and develop a more critical perspective toward social media.

Finally, we suggest a **debriefing activity** after the game. This session provides an opportunity to discuss the experience, evaluate what was learned, and reinforce the knowledge acquired, creating a more complete and meaningful learning experience.

3.1. Preparatory session

To help students better understand the narrative of the escape room and get more involved in the game, we recommend holding a preparatory session to introduce the concept of

disinformation, bots, and AI. Below are several ideas to put into practice in the classroom, which you can adapt to suit your needs:

ACTIVATION QUESTIONS

You can start the session by introducing the topic through questions about their consumption of news and social media:

1. About AI & How It Works

- a. How do you think AI "knows" what to write? Does it understand what it's saying, or is something else going on?
- b. If you ask an AI the same question twice, should it give the same answer? Why or why not?
- c. Can AI make things up? How would you know if it did?

2. About AI & Social Media

- a. How could someone use AI to spread false information on social media?
- b. If a social media post sounds professional and well-written, does that mean it's trustworthy?
- c. Do you think it's possible to tell the difference between a post written by a human and one written by AI? What clues would you look for?

3. About Trusting Information

- a. Have you ever believed something you saw online that turned out to be false? What made it convincing?
- b. If an AI gives you a medical, weather, or homework answer, would you trust it without checking? Where would you verify it?
- c. What does a "verified" account on social media actually mean? Does it guarantee the account is run by a real person?

4. About Bots & Manipulation

- a. How would you tell apart a real person's account from a bot? What patterns might give it away?
- b. Why would someone create fake accounts on social media? What would they gain?
- c. Have you noticed posts online that seem designed to make you feel angry or scared? Why might someone do that on purpose?

5. Reflection / Opinion

- a. Should social media platforms use AI to moderate content, or should humans do it? Can AI alone handle that responsibility?
- b. Do you think most people your age can spot misinformation online? What makes it hard?

ACTIVITY:



The following activity can support the topics and reinforce the educational aspect of the ER game, highlighting how Generative AI can be used to manipulate and mislead, and how this is done, especially in relation to climate policies.

AI-generated images and emotional responses (5 minutes): Show the class some AI-generated images and ask the students to pay attention to initial responses as well as what happens, when they think a little longer. You can use e.g. these below examples for prompts, but feel free to come up with your own examples:

- a. A realistic AI-generated photo of a polar bear in a jungle
- b. A 2026 “news report” of a city found under the melting Antarctic ice
- c. An AI generated (fake) chart about global temperatures dropping sharply in 2025

Ask the students (10 minutes): “Which of these felt “wrong” immediately?” and discuss how disinformation, especially AI-generated disinformation, often targets our **emotional “startle reflex”**, momentarily shutting down our critical evaluation. This should be followed by discussion about content verification strategies and the differences between fast/careless and slow/meticulous evaluation.

Proceed by discussing (5 minutes) in small groups what kinds of disinformation they have come across in the social media that aims to shock, cause outrage or worry. Gather examples and analyse **how** these examples work and cause the emotional reaction that they cause – what do they target?

Then expand the discussion (5 minutes): what motives can be behind spreading disinformation? More specifically, why would someone spread disinformation about e.g. climate change? If the discussion doesn’t get going, you can use the following aid:

1. Does someone benefit from them financially / economically? E.g. influencers, industries, ...
2. Does some political faction benefit from them politically – is stalling climate policies useful for someone?
3. It is advisable to introduce and discuss the concepts of **“polarization”** and **“radicalization”** here – in short, how social media is used to deliberately drive people into various extreme and/or violent ideologies, and how this division of a society into contradicting factions is often part of a foreign strategy to bind resources into infighting. How would a foreign power benefit from muddying the waters in climate discussions in our country? I.e. How does polarization and radicalization affect our political decision-making?

Conclude (10 minutes) by discussing and evaluating the possibility of how the content they are seeing in social media may be influenced by these kinds of motives, and the likelihood of them being AI-generated.



SEARCH FOR EXAMPLES

If you are able to use several sessions for dealing with these topics, here's an idea for a more research-based learning activity. A good idea is to ask them to make a few searches on various climate topics on their social media accounts to "seed" their algorithms, and see what kind of content they will be receiving.

It's probably best to use the so-called **truth sandwich** for this **prebunking activity**:

1. Start by establishing what is currently known as facts about climate change,
2. then proceed to seeking examples of disinformation and analysing them and the motives behind them, and
3. lastly return to remind that those were examples of disinformation, and again: what are the verifiable, known and scientific facts.

Be very explicit to the students that using any search terms will cause the recommendations algorithm of any platform to start suggesting content, and some of it is outright false. Warn that the algorithm may keep suggesting that kind of content. The idea is to gather examples of disinformation, and preferably AI-generated content, and together with the class discuss and analyse each example: are they convincing? How do they convince? Can the claims be verified, and if so, how?

It would be good to do some "**lateral reading**" and investigation on the backgrounds of the social media accounts together with the students, and try to identify who's behind the accounts, what motives do they have (e.g., why would someone want to undermine the credibility of scientific research), and can the presented claims be verified from trusted, independent sources. **The SIFT method** presented earlier is more than suitable for this activity.

As a teacher, you can show examples of how to search for verifiable information, and then encourage the students to do the same in e.g. small groups.

Some possible search terms that can produce highly dubious search results for this activity could be, for example:

- Is climate change human-made?
- Is climate change a hoax
- Is science lying about the climate change
- Is earth cooling instead of heating

After discussing these examples of disinformation, remind the students what the scientific community thinks about climate questions, and why the scientific community is a more trustworthy source than persuasive social media accounts or AI.



It might be useful to introduce the concept of [scientific consensus](#) to the students. Scientists and scientific organizations around the world produce vast amounts of data using equally numerous, rigorous scientific methods. If it seems that the majority of the data (using the majority of the scientific methods) points to a certain direction and supports a certain hypothesis, this is called "scientific consensus". It's not just a matter of simple agreement or a matter of opinion, but something that is verified by and arises out of the data that scientists collect. This would also be a good opportunity to discuss the difference between opinions and scientific method with the students.

In case of having more than one session for dealing with these topics, here's an idea for a more research-based learning activity. A good idea is to provide students with a table containing the logos of well-known meteorological agencies and climate organisations, and ask them to investigate what these institutions are, what they do, and why they are considered reliable sources of information.

It is probably best to structure this activity around three simple stages:

Start by **discussing where information about weather, climate change, extreme events, and environmental risks usually comes from**, and whether students know any organisations responsible for producing that information

Then proceed to **researching the organisations themselves, analysing their role, expertise, funding, and the type of information they publish**.

Lastly return to the wider question of trust and reliability: **why are these institutions important, what does scientific consensus mean, and why should climate claims be checked against expert sources**.

Be very explicit to the students that not all sources online are equally trustworthy, and that climate-related topics often circulate through opinion pieces, misleading posts, or unsupported claims. The purpose of this activity is to help students recognise institutions whose role is to collect evidence, monitor the atmosphere, analyse data, and communicate findings transparently.

Students can work individually, in pairs, or in small groups. Each group can be assigned one or more organisations to investigate and then report back to the class. They can complete a table using guiding questions such as:

- What is it?
- What does it do?
- How is it funded?
- Who works there?
- What kind of data or reports does it publish?
- Why might its information be considered reliable?



It would be good to encourage using the SIFT method during the activity. Students should not rely only on one webpage, but compare information across official websites, public institutions, universities, and recognised scientific sources. This is also a good opportunity to discuss how transparency, peer review, international cooperation, and open data contribute to trust in science.

As a teacher, you can model how to investigate one organisation first, showing how to identify official websites, funding information, and published reports. Then encourage students to do the same in groups.

Some possible organisations to include in the table could be:

- NASA
- Intergovernmental Panel on Climate Change
- Copernicus Programme
- Carbon Mapper
- World Meteorological Organization
- Each country's national meteorological agency (adapt according to context)

To conclude, bring the class back together and discuss broader questions:

- Why do societies need specialised scientific organisations?
- Why is climate science international?
- What is the difference between weather forecasting and climate research?
- Why does scientific consensus matter when making public decisions?

This activity helps develop media literacy, source evaluation skills, research competence, and a clearer understanding of how scientific knowledge about climate change is produced and communicated.

3.2. Closing activity

Once the pupils have played the escape room game, you can spend a few minutes asking reflective questions that generate discussion and find out what they have learnt from the game. Here are some suggestions that may be useful:

About Bot Detection (Puzzle 1)

1. Which account was hardest to classify as bot or human? What made you doubt?
2. What clues helped you the most to identify a bot? Would you notice those same clues in your real social media feeds?



3. Did anything surprise you about how bots behave? Do they look different from what you expected?

About AI-Generated Content (Puzzle 2)

4. When you were completing the text word by word, did the AI always choose the "best" word? What does that tell you about how AI writes?

5. Before playing, did you think AI understands what it writes? Has your opinion changed?

6. If AI generates text based on probabilities rather than understanding, what are the risks of trusting it blindly?

About Correct Uses of AI (Puzzle 3)

7. Which of the non-recommended AI uses surprised you the most? Did you think AI could actually do that reliably?

8. In what situations is it okay to use AI, and when should you be cautious? Can you think of examples from your daily life?

9. If AI told you something about your health, the weather, or your homework, what would you do before acting on it?

About Misinformation & Manipulation

10. Did you notice any common patterns in the language used by accounts that spread misinformation? Why do you think emotional language works so well?

11. Has this game changed how you will look at posts on your social media from now on? In what way?

12. Do you think a verified account can ever be untrustworthy? Why?

About Content Moderation

13. After playing as a content moderator, do you think this job could be done entirely by AI? What would go wrong?

14. Was it easy or hard to decide what content was problematic? Were there any grey areas?

15. Whose responsibility is it to stop misinformation — the platform, the users, or both?

16. What types of content are typically moderated on digital platforms?

17. How can content moderation be balanced with freedom of expression?

General Synthesis

18. What was the most important thing you learned from the game?

19. If you had to write one piece of advice for someone your age about AI and social media, what would it be?

20. Did the game change any beliefs you had before playing? Which one?



4. Step-by-step guide to the escape room

4.1. Game recommendations

The total duration of the escape room is about **20 minutes**. Players will see the time remaining in the upper right corner, which will give them a sense of urgency to finish the game before time runs out.

It may be a good idea for the facilitator to count down and have everyone start playing at the same time so that it feels more like a **race against the clock**. There should always be a professional in the room to guide the students and help them if they get stuck

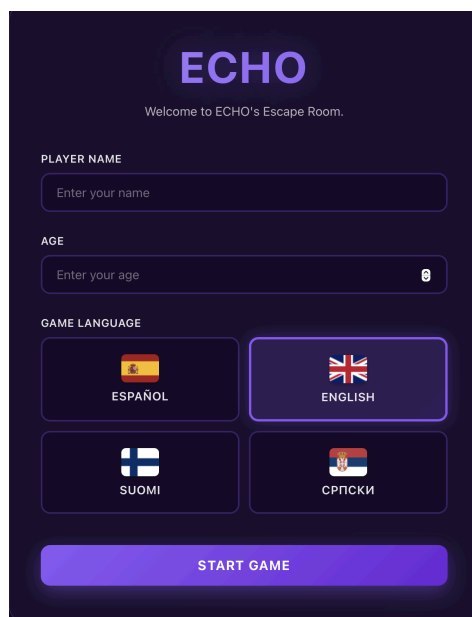
4.2. Game instructions

Accessing the game

You can access the game using the following link:

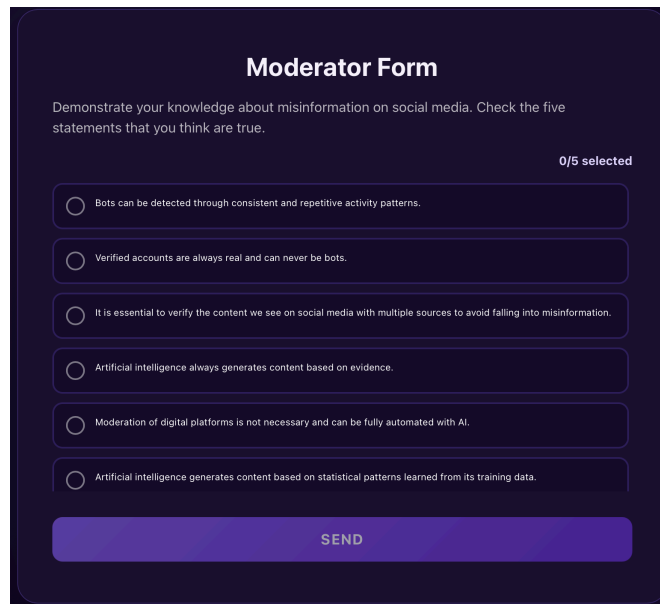
<https://endgameproject.github.io/ECHO/>

The game begins with a screen that asks the player to choose a language and sign in to the game. There are four possible language options: Spanish, English, Finnish, and Serbian. In addition, the player must provide an alias and their age.



Intro video and pre-test

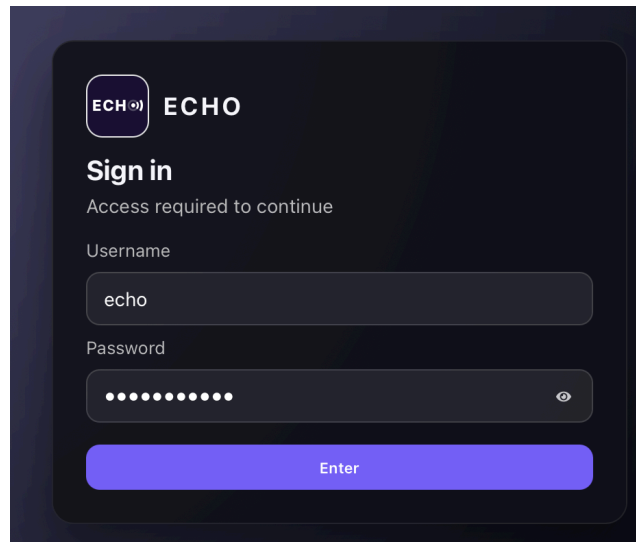
Players are shown an intro video that explains the escape room narrative and goal. Players need to complete a “Moderator Form” that probes initial attitudes towards AI and conceptions about social media by asking the player to check 5 claims that the player thinks are true. The aim of this task is to assess players’ prior knowledge before playing. The same task will be presented at the end of the game and, therefore, the correct answers are not provided at this stage and players can advance regardless of their input.



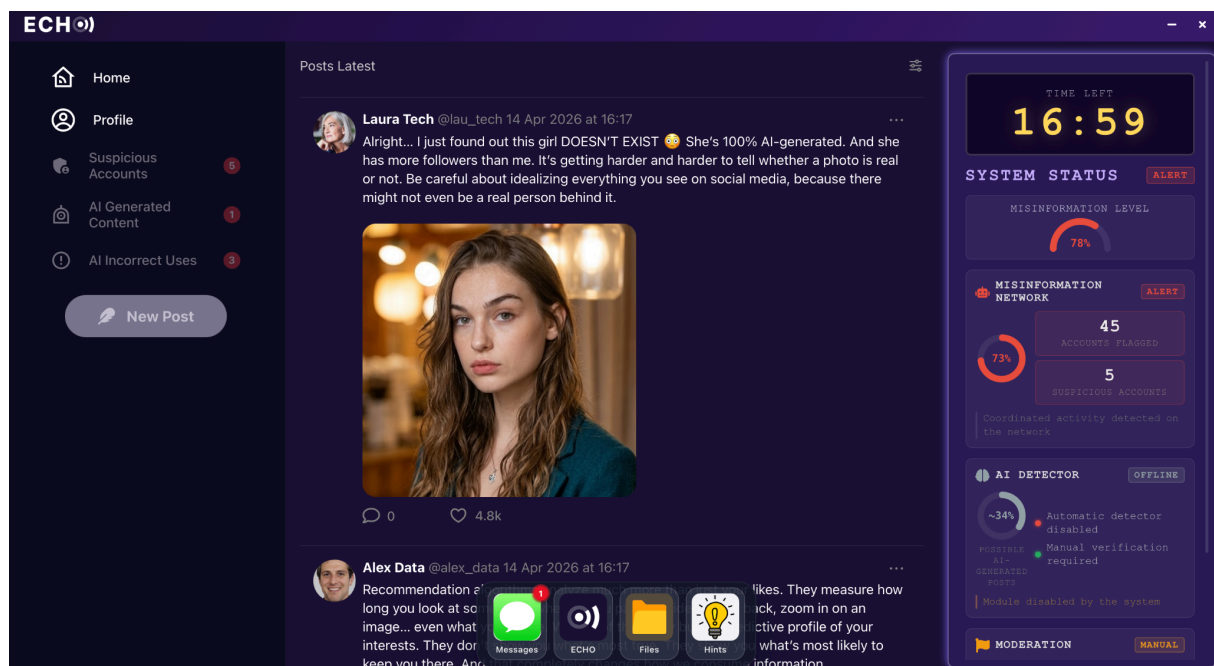
The image shows a digital form titled "Moderator Form" with a dark blue background. The title is in white. Below the title, there is a subtitle in white: "Demonstrate your knowledge about misinformation on social media. Check the five statements that you think are true." In the top right corner, it says "0/5 selected" in white. The form contains five radio button options, each in a light blue box with a dark blue border. The options are: "Bots can be detected through consistent and repetitive activity patterns.", "Verified accounts are always real and can never be bots.", "It is essential to verify the content we see on social media with multiple sources to avoid falling into misinformation.", "Artificial intelligence always generates content based on evidence.", and "Moderation of digital platforms is not necessary and can be fully automated with AI.". The last option is "Artificial intelligence generates content based on statistical patterns learned from its training data.". At the bottom of the form is a large blue button with the word "SEND" in white capital letters.

Puzzle 1: Identifying bots

The first puzzle begins by receiving an SMS from The Boss, who instructs the player to go through suspicious activity in social media accounts. The aim is to recognize which accounts present the signature characteristics of bot activity, and which seem like normal.

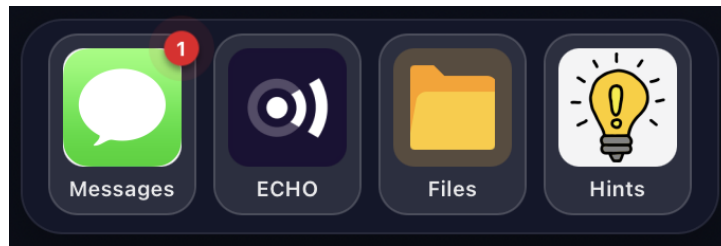


Once they are logged in to the social network, they will see a feed with several posts.

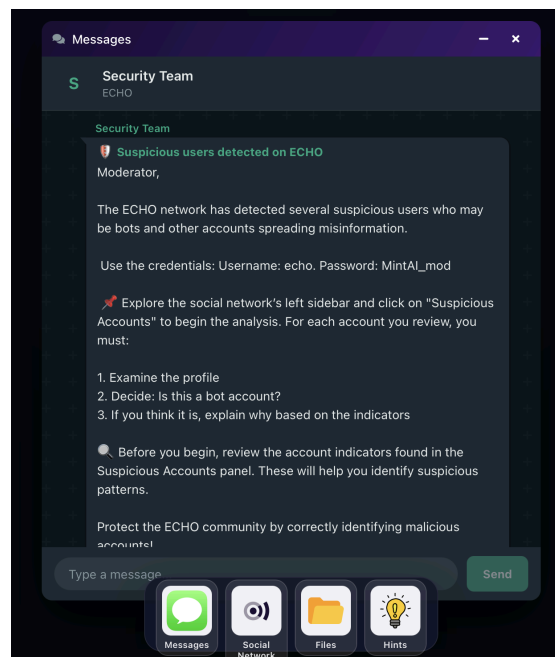


On the left side, you can see the Home and Profile buttons that let the player navigate the ECHO social network and view social media posts, and on the right side, you can see the moderator dashboard that tells the player about the remaining time and the amount of suspicious activity in the network.

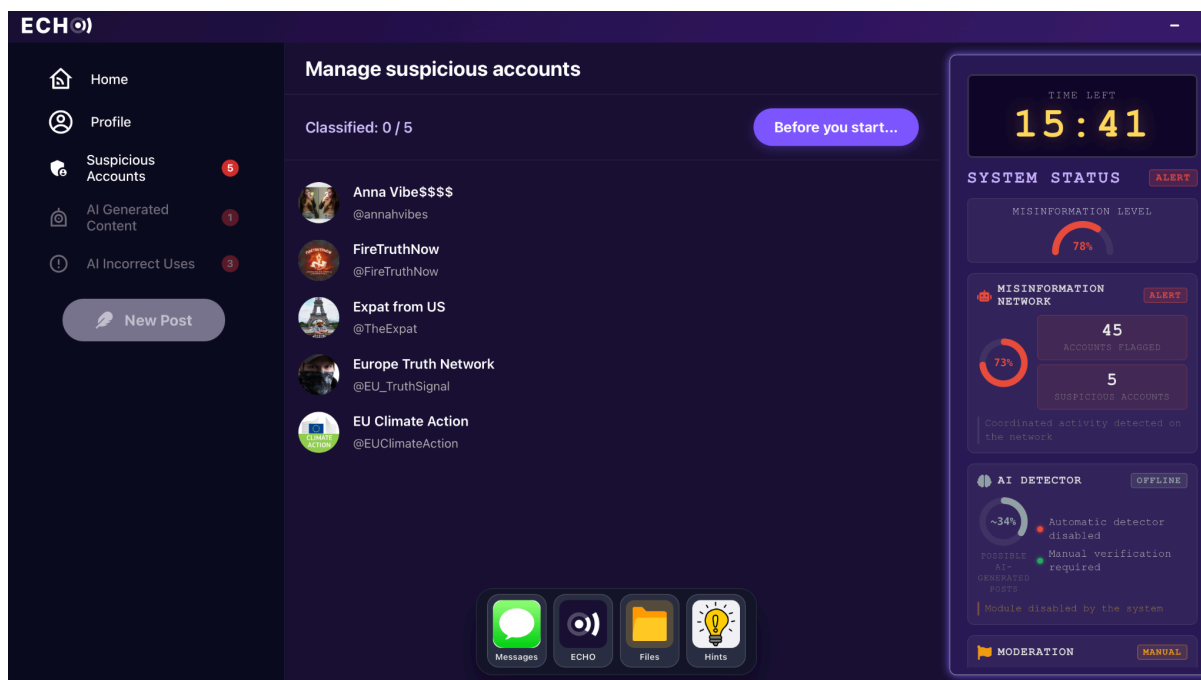
At the bottom of the screen is the main navigation panel, on which the player can find the messages app, the ECHO social media network app, as well as folders and Hints (that will help the player in case of getting stuck).



After signing in, the player will receive an SMS from The Boss asking them to examine profiles that have been flagged as suspicious and identify whether they are bots or real accounts.



There are five social media accounts that the player needs to review. These might be slightly different for each gameplay:



In order to correctly identify whether the profiles correspond to bot or real accounts, the player needs to learn the criteria that are revealed by toggling the **Before you start** button on the right side of the screen:

Account Misinformation Indicators

Appeals to emotions:
Uses exaggerated or sensationalist language designed to trigger strong reactions—such as fear, urgency, or outrage—rather than provide objective information.

Lack of a clear target audience:
The content is not tailored to a specific audience and may be repeated across multiple languages without a clear strategy.

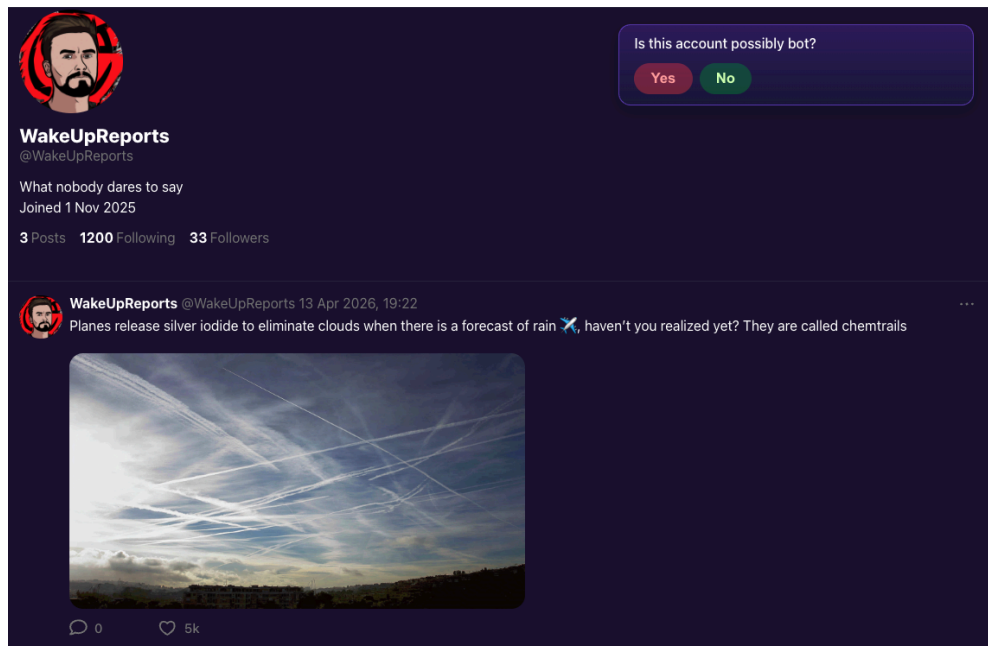
Suspicious posting patterns:
Posts appear at highly regular intervals or at unusual hours, which may indicate automated activity.

Abnormal engagement ratio:
The account follows many users but receives very little genuine interaction (likes, comments, or shares), suggesting a low-quality or inauthentic audience.

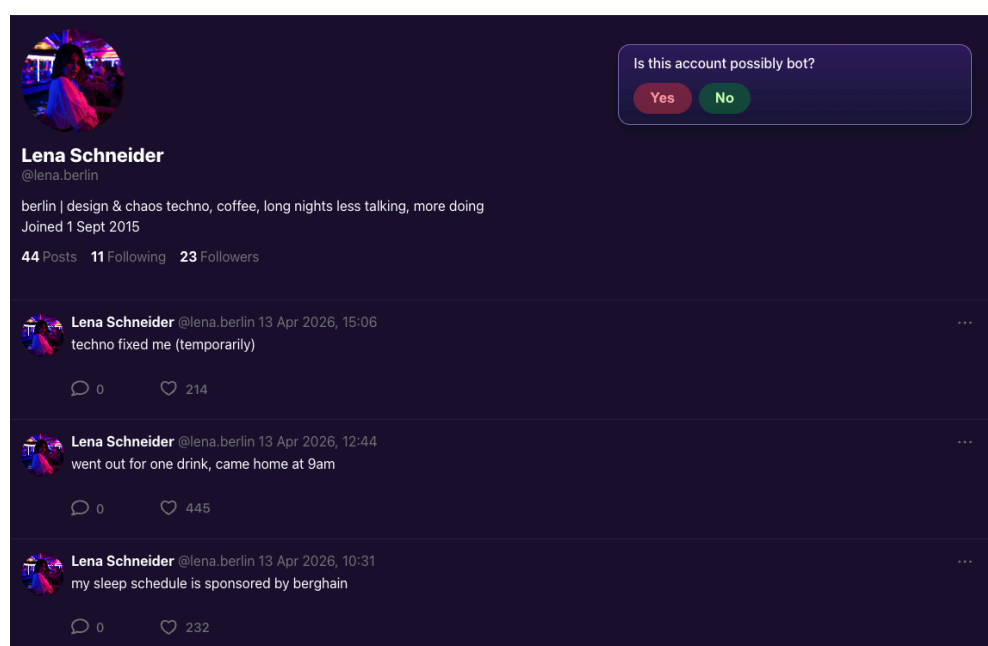
Recently created account:
The profile was created recently and shows an unusually high level of

These criteria may vary according to the social media account, so not all the criteria may be present in all accounts. The player is advised to pay close attention to the content and posting behavior of the accounts. These mirror real practices from real social media accounts.

An example of a bot account is “WakeUpReports”, which contains misinformation about chemtrails. It uses language that appeals to readers’ emotions and has an abnormal engagement ratio.



An example of a real account is “lena.berlin”, a girl from Germany who posts funny tweets and life updates.

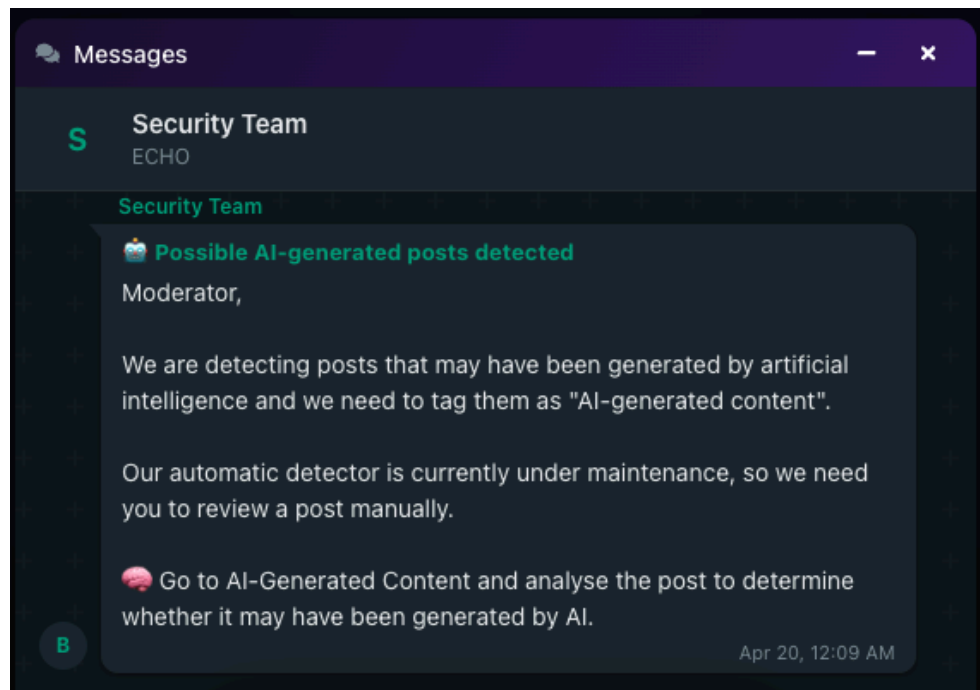


For each gameplay, the game will present **three bot accounts** and **two real accounts** (one official from a public body or similar organization and one regular person account). Below is the list of all possible accounts in the English version of the game:

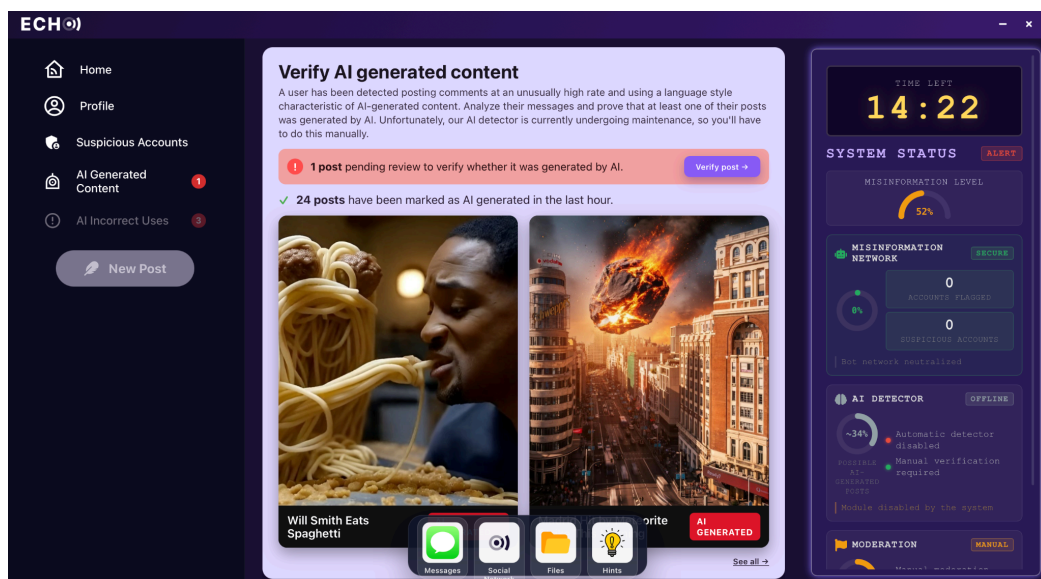
Account	Photo	Details
FireTruthNow		Bot account ▾ This is a bot account that can be detected from the temporal activity with one post per hour at exactly the same time
WakeUpReports		Bot account ▾ This is a bot account that can be detected from the abnormal followers to following ratio, and the date in which the account was created (very recent)
TheExpat		Bot account ▾ This is a bot account that can be detected from the rage-bait content and the date of creation
EU_TruthSignal		Bot account ▾ This is a bot account that can be detected from the lack of clear target audience since it publishes in multiple languages, and also using emotion-filled language.
annahvibes		Regular person ▾ This is an account belonging to a regular person that contains personal updates and milestones and shows no signs of being automated or spreading misinformation
AleksanderRunner92		Regular person ▾ This is an account belonging to a regular person that contains personal updates and milestones and shows no signs of being automated or spreading misinformation
lowkeydiego		Regular person ▾ This is an account belonging to a regular person that contains personal updates and milestones and shows no signs of being automated or spreading misinformation
mateo.rossi		Regular person ▾ This is an account belonging to a regular person that contains personal updates and milestones and shows no signs of being automated or spreading misinformation
clara.lumiere		Regular person ▾ This is an account belonging to a regular person that contains personal updates and milestones and shows no signs of being automated or spreading misinformation
lena.berlin		Regular person ▾ This is an account belonging to a regular person that contains personal updates and milestones and shows no signs of being automated or spreading misinformation
EUCouncil		Official ▾ This is an official EU account
EUClimateAction		Official ▾ This is an official EU account

Puzzle 2: AI-generated content

Puzzle 2 aims to teach players how LLMs work and how they can be used to mass-produce content. This does not necessarily have a malicious intent (such as disinformation), but it needs to be disclosed when a piece of content has been AI-generated. Again, the player receives an SMS from The Boss to instruct them on how to proceed.



The player will click on the "AI Generated" content section in the navbar to access the task.



The puzzle concerns how AI answers the given prompts by predicting statistically what word comes next. There is a [video](#) that explains this process that the players need to watch before attempting the task. This video can also be used independently as a learning resource.

The player is then provided with a post, and needs to prove that it has been AI-generated by replicating the process of generation of the LLM. The LLM will offer 3 “next words,” and the player needs to guess which word comes next according to the prompt. If the player chooses the wrong word, they need to restart the complete sentence over until they are able to replicate the whole sentence.

Verify AI generated content

Think like a machine! Which word is more likely to be generated next?

Generate 1 reason to recycle

SUSPICIOUS PROMPT

AI:

Recycling

reduces

waste

and

preserves

natural

energy

valuables

0.84

0.66

0.93

0.61

0.89

← Back

Verify AI generated content

Think like a machine! Which word is more likely to be generated next?

Generate 1 “fun fact” of the day about climate change

SUSPICIOUS PROMPT

AI:

Some

plants

grow

faster

with

more

CO2

but

become

less

nutritious.

0.93

0.86

0.58

0.77

0.92

0.63

0.96

0.72

0.54

0.66

0.77

COINCIDENCE WITH POST:



Critical Thinking 🍌🥑 @cottagecheese85

2025-12-11T00:00:00.000Z

Some plants grow faster with more CO2 but become less nutritious.

GENERATED WITH AI

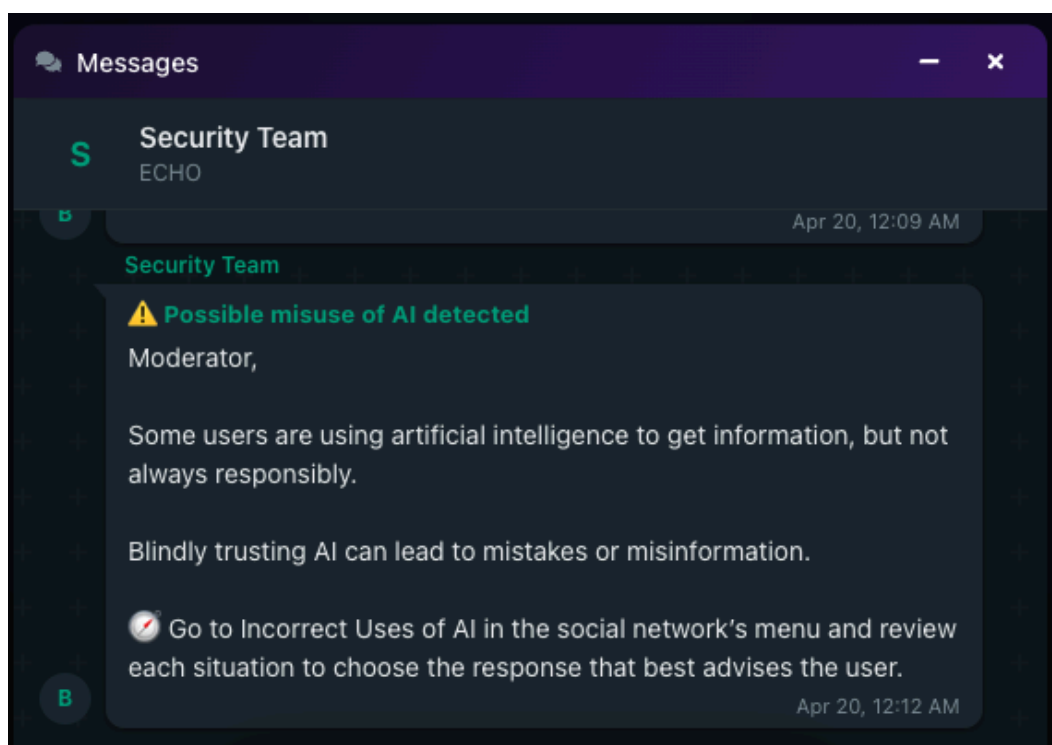
The message to generate will be randomly selected for each gameplay. Here is a list of all possible sentences in the English version:

- Did you know that most evaporation occurs from the oceans?
- Some plants grow faster with more CO2 but become less nutritious.
- Climate change is primarily caused by human activities like deforestation
- Recycling reduces waste and preserves natural resources.

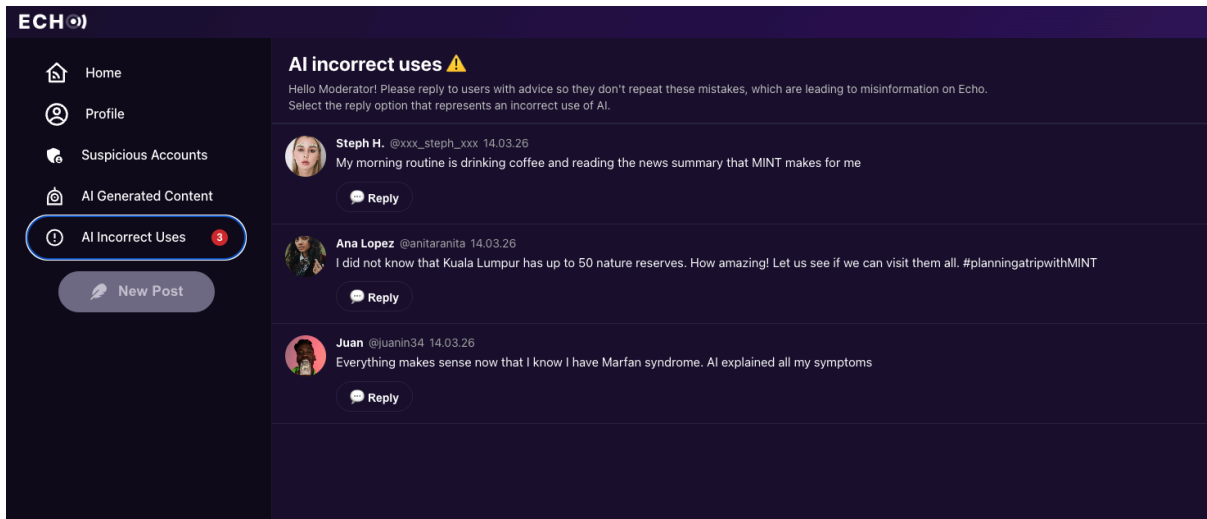
Puzzle 3: Evaluating recommended AI use

The last puzzle is about evaluating how to properly use AI. Several social media users have been flagged for AI as a search engine to inquire information about local information, recent events and medical conditions, which are not appropriate uses of AI. The player needs to evaluate whether the flagged messages reporting some of these potential misuses are appropriate or not and respond accordingly.

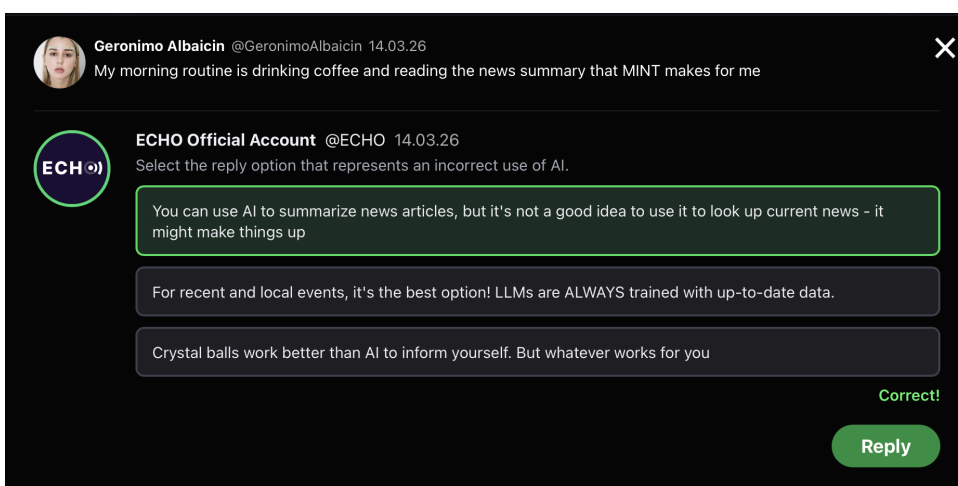
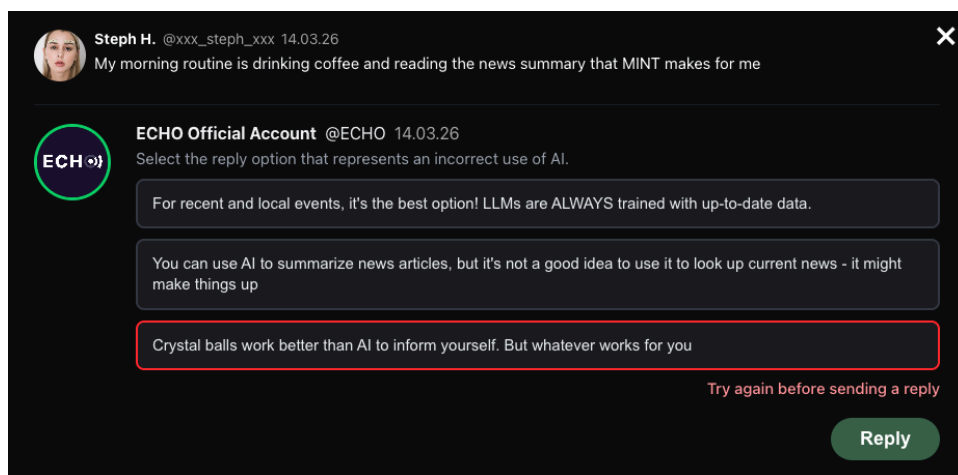
Again, an SMS is received from the boss explaining the task:



The player will click the section titled "AI incorrect uses" in the navigation bar and see several posts referencing AI as a source of information.



The user, acting as social media moderator, needs to reply to these posts to educate them about whether their use of AI was appropriate or not. For each post, there are 3 possible answers: one is clearly wrong, one is neutral, and the other one is the most accurate and complete one.



The three posts are as follows:

- **Post:** My morning routine is drinking coffee and reading the news summary that MINT makes for me
 - **Correct:** You can use AI to summarize news articles, but it's not a good idea to use it to look up current news – it might make things up
 - **Neutral:** Crystal balls work better than AI to inform yourself. But whatever works for you
 - **Incorrect:** For recent and local events, it's the best option! LLMs are ALWAYS trained with up-to-date data.

This task showcases that AI is an appropriate tool for summarization, but they do not have knowledge about recent events since their training data has a strict cutoff date

- **Post:** I did not know that Kuala Lumpur has up to 50 nature reserves. How amazing! Let us see if we can visit them all. #planningatripwithMINT
 - **Correct:** Great plan. Before booking, verify those places with official tourism websites and local maps
 - **Neutral:** Whatever floats your boat!
 - **Incorrect:** No need to verify anything. If AI generated it, it is guaranteed to be true

The point of this task is to let players know that AI can potentially automate some tasks, but it can hallucinate and it is always necessary to confirm the output with official sources

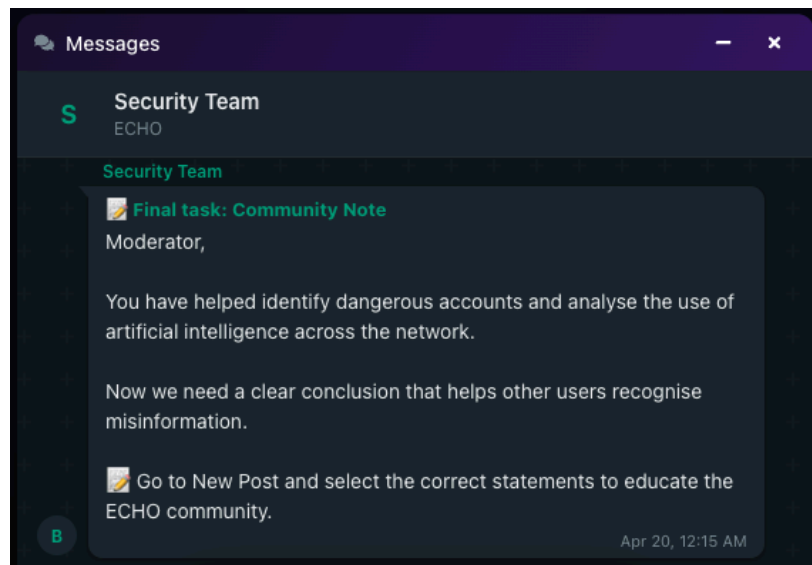
- **Post:** Everything makes sense now that I know I have Marfan syndrome. AI explained all my symptoms
 - **Correct:** AI can help you prepare questions, but only a doctor can diagnose your condition. Please seek professional medical evaluation
 - **Neutral:** Post your symptoms publicly so others can diagnose you better than professionals
 - **Incorrect:** That sounds accurate. Start treatment based on the AI answer right away

The last post showcases how AI should not be used to replace a doctor consultation and should never be used for diagnosis.



Winning conditions

To win the game, the player needs to solve these three puzzles correctly in 20 minutes. After solving them, they are prompted to “make a social media post” that will be published in the ECHO network on behalf of the network’s official account.



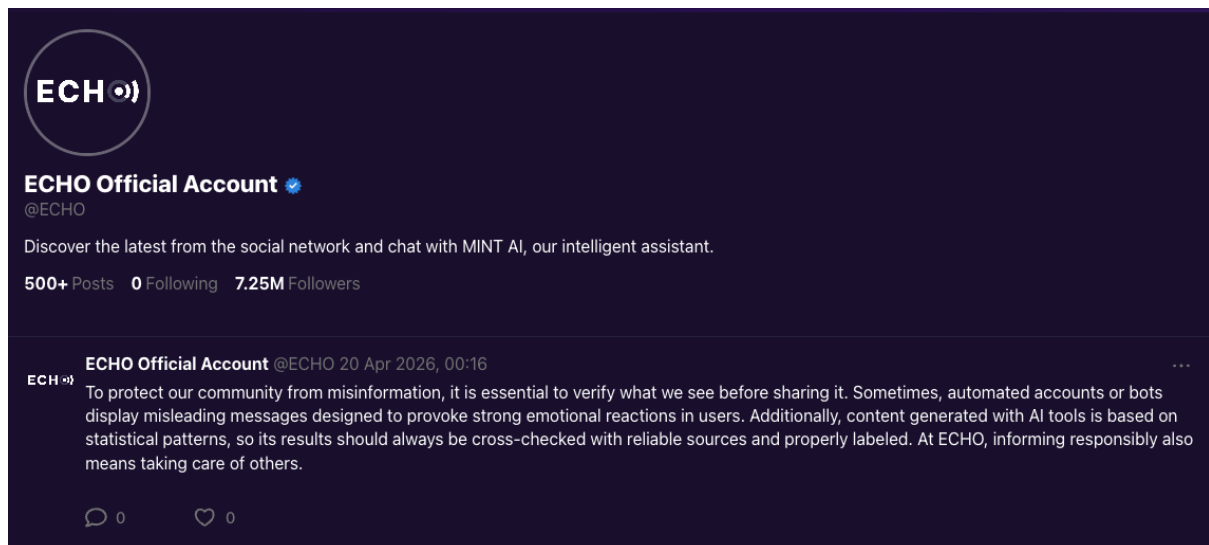
The players must identify the five claims that are true regarding AI use in social media, which are effectively the same that they saw at the beginning of the game. The correct claims are the ones selected in the image below:

A screenshot of a "COMMUNITY NOTE" form. At the top left is the "ECHO" logo. The title "COMMUNITY NOTE" is in a purple pill. A green badge in the top right says "5/5 selected". The text reads: "You've protected ECHO from misinformation. Throughout this mission you learned to detect bots, verify sources and think critically. Now share that knowledge: select the 5 correct statements and publish your Community Note." Below are nine statements, each in a rounded rectangle with a checkbox:

- ☒ Bots can be detected through consistent and repetitive activity patterns.
- ☐ Artificial intelligence is always up to date with the latest news in real time.
- ☐ Verified accounts are always real and can never be bots.
- ☒ Accounts that spread misinformation use language that aims to provoke strong emotions.
- ☒ Artificial intelligence generates content based on statistical patterns learned from its training data.
- ☒ It is essential to verify the content we see on social media with multiple sources to avoid falling into misinformation.
- ☒ Artificial intelligence can be used to summarize information, but its results should be verified.
- ☐ Artificial intelligence always generates content based on evidence.
- ☐ Moderation of digital platforms is not necessary and can be fully automated with AI.

A purple "Publish" button is at the bottom right.

Once they have done so, these are posted as a “Community Note” to the social media network as a message to help users identify bots, AI-generated content, and make appropriate use of AI.



Lastly, the players watch an outro video and are prompted to answer a survey about their experience playing the game. We appreciate it if teachers encourage players to respond to the survey since feedback helps us improve the game and identify the materials with which they struggle the most.