

Published 11 July 2025

Qualitative study: User perspectives on AI content labels



Prepared by YouGov Qualitative

CONTENTS

1	Introduction
2	Summary
3	Views on AI-generated content
4	Views on different types of AI content labels
5	Views on the AI content labels landscape
6	Annex

1 Introduction

Background and objectives

In 2024, Ofcom commissioned YouGov to conduct a survey among internet users about their interactions with AI content labels. In order to further explore users' understanding of and attitudes towards these labels, a follow-up qualitative study was conducted in 2025.

The qualitative study aimed to provide deeper insights on the following questions:

- What are participants' experiences in encountering AI-generated content online and how much trust do they have in such content?
- How do participants react to and interact with different types of AI content labels in different formats (e.g., an image and a video)?
- How do participants' attitudes toward these labels and their stated behaviours change (if at all) based on different formats or experiences?
- What are participants' views on labelling AI-generated content online?

Methodology and Guide Structure

Methodology



YouGov Qualitative conducted 24 online depth interviews lasting 60 minutes each between Wednesday 12th February – Wednesday 26th February 2025.

Sample frame:

- Adults aged 18+
- Mix of age, gender, social grade, regions and education level
- Inclusion of ethnic minorities
- Mix of social media usage frequency and social media platforms used; including those who upload videos or livestream
- Mix of support for political parties
- Mix of levels of digital confidence
- Mix in the awareness and understanding of AI

The analysis was undertaken by researchers who conducted the fieldwork in multiple analysis sessions. Data was analysed using thematic analysis and key recruitment criteria – demographics, level of digital confidence and AI understanding. There were only a few areas where differences were observed between these subgroups – they have been stated in the report wherever applicable. Findings that are discussed without subgroup differences refer to views that were expressed by the whole sample.

Interview Structure



The interviews started with participants reading the scenarios on-screen and their initial reactions. The order of the scenarios was rotated for every interview to avoid any order effects and ensure a good mix in the reflections participants provided to these scenarios.

The scenarios included one image and one video with three user-visible labels, which included:

- The CR (Content Credentials) label
- The 'Made by AI' label
- The 'Readers added Context' label

These types of labels were selected as they are some of the measures used by services to inform their users currently. Participants were asked to reflect on each label per scenario.

The second part of the interview involved a discussion about participants' understanding of AI-generated content and deepfakes, followed by their views on how AI content labels could be implemented by online platforms.

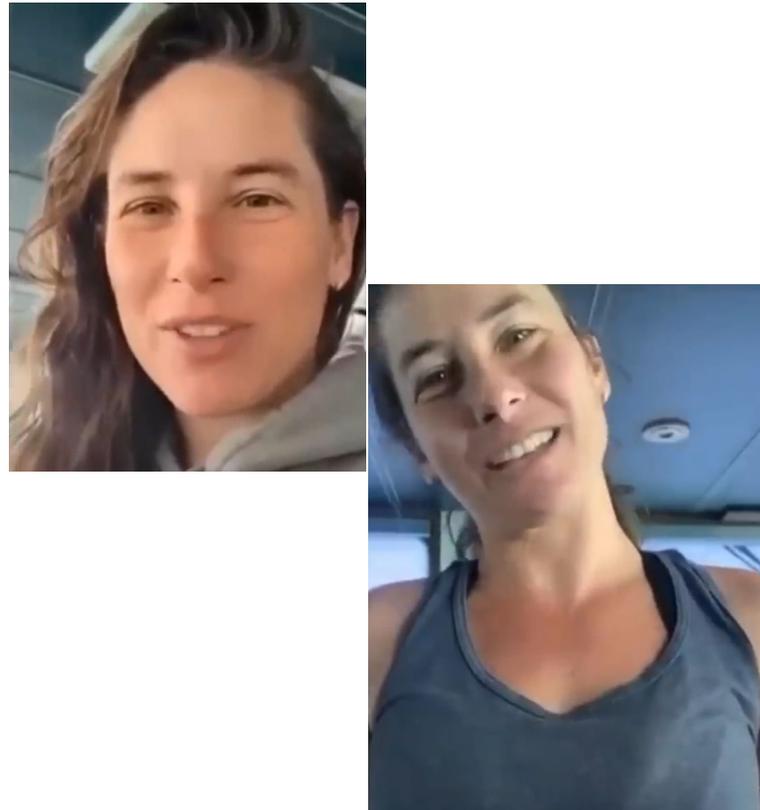
Scenarios

The scenarios included one image and one video, both AI-generated. The image was a deepfake of Tom Holland getting arrested, while the video was used in an online scam, featuring someone claiming to be 'Alla Morgan' who was seeking to reassure the viewer of her sincerity and that they were not connected to scams. The stimuli shown to participants were content taken from publicly available and freely accessible sources online (see Annex for details). For each scenario, one of the three AI content labels was added to explore how participants would respond. Each participant saw all six combinations, with the order of presentation rotated to avoid bias.

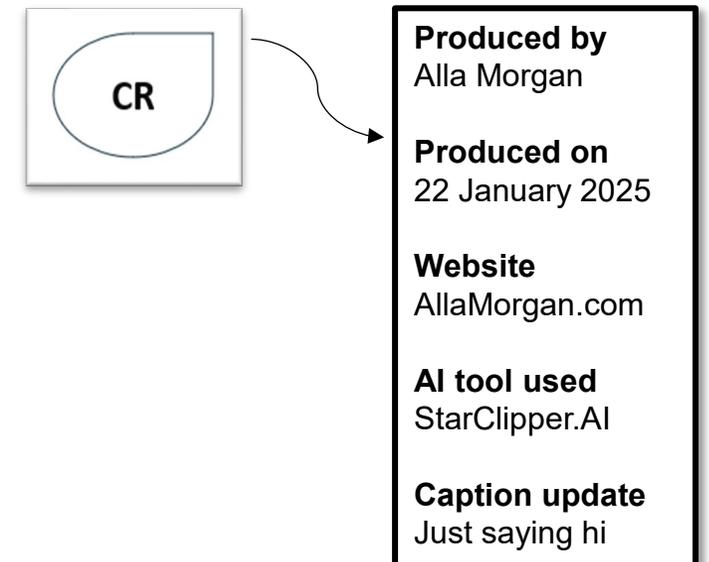
Image



Video

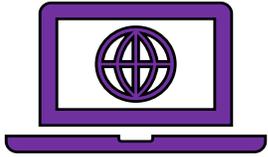


AI content labels



2 Summary

Summary



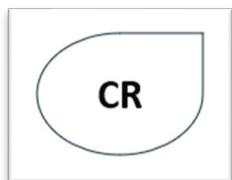
Participants had mixed views about AI-generated content, with some having concern about authenticity, trustworthiness and potential for misuse, while others noted the value of AI for creating entertaining content. Generally, there was a strong appetite for mandatory and standardised labelling to inform the public of AI-generated content online. Participants felt that such labels might help protect users from potential scams because they might prompt them to be more cautious.



The 'Made By AI' label was the most preferred label for its clarity to indicate content as AI-generated. Participants found it direct and impactful. However, the label was neutral about the content and participants had to assess themselves whether the labelled content was harmful.



The 'Readers/Users added context' label was the only one of the three labels tested that directly alerted the user to the possibility of a scam, as its free text could provide any information about a given post. However, there were some confusion and uncertainty around who could input in these labels and whether the information can be trusted. Some participants were concerned that the label would be applied only after a piece of harmful content had been shared online, potentially limiting the impact of this label.



The 'CR' label was least familiar to participants, and many did not realise that a pop-up box could be revealed. When the provenance metadata information in the pop-up box were shown, many were unsure how to interpret the information or if it can be trusted. However, many participants were able to correctly identify certain information, such as the "AI tool used" field, which indicated that the stimuli tested in the research was likely to have been manipulated.

3 Views on AI-generated content

There were mixed views about AI-generated content, with some concerns about trustworthiness and misuse

 Overall, participants seemed to have a **limited understanding of online content created by AI** as their knowledge was mainly based on what they have seen or heard on the news. Some mentioned seeing AI-generated images online or experimenting with the images themselves for entertainment. However, participants felt that they lacked understanding about how the AI-generated content is being used online or the extent to which it was being used, with some mentioning that they lacked confidence in always being able to identify it.

 There was a broad **concern about authenticity, trustworthiness, and the potential for misuse** through AI-generated or edited online content. Participants worried that it could lead to spreading misinformation and have negative impacts such as scams or creating false perceptions about certain groups. For example, a participant mentioned that “it could impact the perception among young men that all women should look like models”.

 However, a minority **appreciated the use of AI for creating entertaining content** and improving the quality of online content.

“

“I feel like there's quite a lot that I don't know to be fair, my understanding of it is probably quite basic.... in comparison to what there actually is to know.” – Male, 25, Medium digital confidence

“AI can be used for some wonderful... but it all depends about the context... so you've just got to take whatever you see with AI with a pinch of salt because you don't know the source, who created it. But for humor purposes, AI is wonderful, because you can create anything in a couple of seconds.” – Male, 40, Low digital confidence

“There are lots of negatives, because I think some of the AI generated models don't look like real women... I know that... but for like a 14-year-old girl who is growing up and kind of aspiring to look in a certain way, but also for then boys who then potentially would expect a woman to then look in a certain way. I think that's really damaging.” – Female, 32, High digital confidence

“[I don't trust AI] because things can be twisted round, and you're given fake news. You could be given fake news, you just don't know what's real and what isn't. And it starts to make you question what's going on... It's very concerning...” – Female, 65, Low digital confidence

”

Participants were particularly worried about deepfakes because of their potential negative impacts

- Overall, participants had a good understanding of **deepfakes**, recognising them as AI-generated videos that can manipulate appearances and voices to create realistic but false representations. Their awareness comes from what they saw or heard about deepfakes in news stories and on social media platforms.
- Participants found **deepfakes particularly concerning and dangerous as they could be used for misinformation or malicious purposes**. They noted the dangers of deepfakes being used to manipulate videos of political figures, create false endorsements, and even produce harmful content like non-consensual pornography.
- The positive side of deepfakes in their potential for entertainment and humour was noted by a minority. However, these **positives were overshadowed by concerns about misuse and ethical implications**.
- One individual mentioned a perception that with further AI developments, **deepfakes will become much more sophisticated and convincing in the future**.

“

“It’s a lie, an online lie about someone that something’s happened, but it’s more than just a lie. It includes images, videos, so, it’s quite an embedded lie with more than just words that could be a celebrity or it could be a politician or [a] normal person.” – Male, 48, Medium digital confidence

“Deepfakes is definitely a big concern... it can be so dangerous in terms of political figures saying things which can be very damaging to their reputation, also spreading false information online. I know a lot of deepfakes are being used for like sexual purposes, like celebrities are being kind of deepfaked in those ways, which can be very harmful to them, very traumatizing... people can get catfished by deepfakes... I think it was Martin Lewis or someone was saying like, you need to invest, put your money in this thing.” – Female, 24, High digital confidence

“I’ve seen a few, but at the moment, it’s not as believable to me. I guess as AI kind of gets better and better, it will become more and more, believable.” – Male, 31, High digital confidence

”

Participants said they would use various techniques to identify AI-created online content

When asked openly what signs participants would look for to **establish whether a piece of online content** (including videos, adverts, social media posts, images, news stories) **was AI-created**, they spontaneously mentioned a variety of different techniques:



Check the source, if it comes from a reputable site



Check if the **page is trustworthy** and has the 'padlock' icon at the beginning of their URL to confirm that the connection between the web browser and the website server is encrypted



Review the image or video to see if it looks credible - if all the details are correct, if the footage is synchronised, if there are any glitches, if the lighting and shadowing is natural, if the image or video is too perfect to be real



If the **message is in line with the speaker's views** - whether the speaker says something controversial or different to their normal outputs, contradict what they had said previously, or says anything that the user *would not* expect them to say



If the **topic is focused on a financial matter**, they would be more cautious about the content, worrying that it might be a scam



*"[...] the heads moving and the eyes stay the same, color and tone, and there's no shadowing and things... if [there are] glitches... if it's fuzzy on the edges or blurry or if the eyes don't move quite naturally enough."
– Male, 31, High digital confidence*

"You can tell with the lighting...and AI can't quite get the fingers right." – Male, 40, Low digital confidence

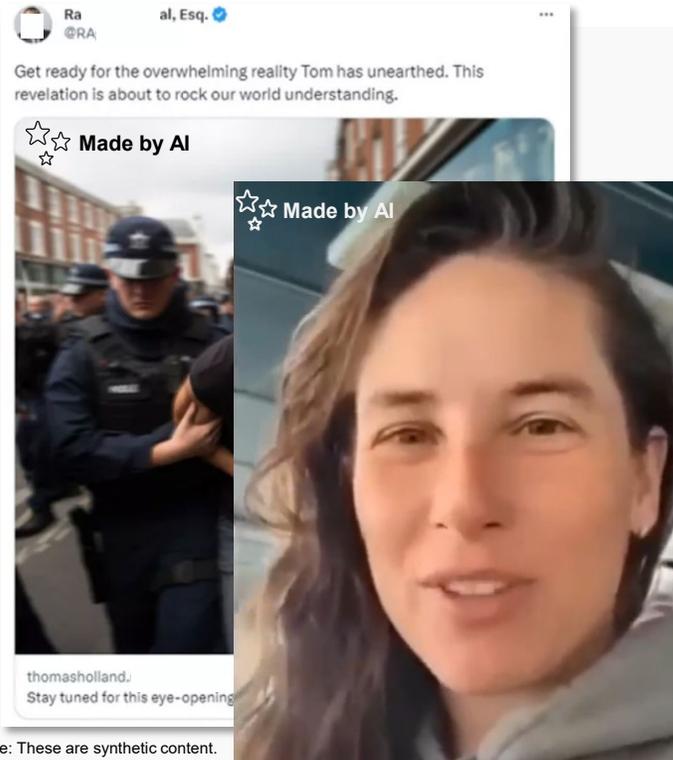
"[...] if I know the individual, is it something that fits in what they would say or not? And is it like them, promoting or selling something which they would never do normally, or they've said that they would never do publicly... the subject matter, the topic, the expression whether it's persuasive or like advertising or promoting something where there's monetary or financial." – Male, 41, High digital confidence



4 Views on different types of AI content labels

4.1 'Made by AI' Label

Initial impressions of the ‘Made by AI’ label were positive, as it immediately confirmed the use of AI



- A minority of participants mentioned seeing a ‘Made by AI’ label or similar indicators on platforms like TikTok, but generally participants said they had not encountered it before.
- Overall, **participants found the label useful** as it provided **transparency** about the involvement of AI in creating or editing the content.
- The label was **easily noticeable** on both the image and video tested in this study. The **positioning** of the label on the left side of the screen was also felt to make people’s eyes gravitate towards it first (due to reading from left to right).
- The clear labelling of AI use caused participants to question the trustworthiness of the content and as a result, lowered the likelihood of engaging with it.

“[...] the Made by AI it's getting straight to the point, which then stops you and makes you think before you can then either move on or investigate further... the other two are a bit more like, here's just a bit more information, but we're still not telling you really... The made by AI at the top is telling you straight away. You're not having to click on something or scroll down to the reader's context.” – Female, 54, Low digital confidence

“

“If it's a genuine response and a genuine video, then [the label] shouldn't be on there... You wouldn't need AI... [the label] would make me query the validity of [the video].” – Male, 41, High digital confidence

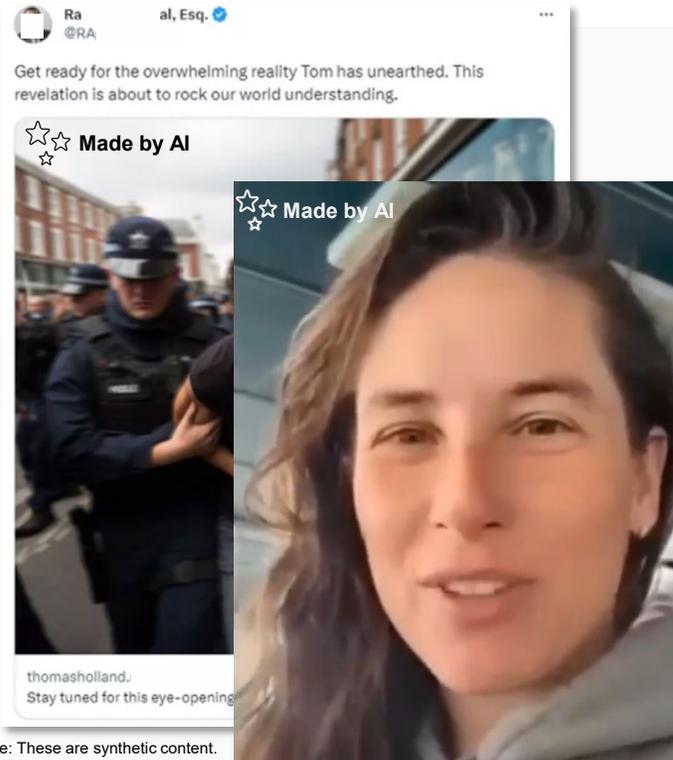
“I think [the label] definitely is noticeable... I think that is quite a good position to have it in... people would naturally gravitate towards that part of the screen.” – Female, 24, High digital confidence

“I think that's better maybe for like quick videos, like things [on] TikTok. I think having the Made by AI is useful, just when you're scrolling and you're seeing that quickly and you kind of know, okay, this isn't real.” – Female, 24, High digital confidence

”

Note: These are synthetic content. The username, profile image and hyperlink are redacted for reporting.

Whilst the label was appreciated for its clarity, there was a preference for wanting to know how AI was used



Positives

- ✓ Participants liked that **the label was easily noticeable and direct**. This was mentioned by participants across all digital confidence and AI knowledge levels.
- ✓ It was felt to be particularly useful in ensuring that the **users are notified immediately** about the content being AI-generated, without needing to click anywhere, read a lengthier text or do their own research, which they might not be inclined to do.

Negatives

- ✗ Whilst participants appreciated the direct and concise label, **it was not clear to what extent AI was used**. This raised concerns about users potentially distrusting genuine content or information in cases where an image/video was only edited with the help of AI or created for the purposes of illustration.
- ✗ Some participants found the **stars in the icon confusing and irrelevant**. Some assumed that the stars might imply a positive rating or quality (e.g., the quality of AI used) rather than associating it with an indication of AI-generated content.

Note: These are synthetic content. The username, profile image and hyperlink are redacted for reporting.



“I think this one is more useful because... you don't have to click on it, you can see that it's made by an AI straight away... It's so straightforward.”
– Female, 32, High digital confidence

“I'm not too sure about the star signs... I don't know where the relevance of the three stars come into it. Made by AI would be sufficient really.” – Male, 61, Medium digital confidence

“You couldn't necessarily take that as proof that she's 100 percent not real, especially without any further information. The last one, which said the AI tool that she was using, at least gives you a bit more insight... [this] doesn't give you any information at all.”



Having a bolder and more visible label was preferred, alongside the option to access additional information



Suggested improvements

- Whilst generally it was felt that the label was easy to notice, there were suggestions that it would be helpful to **make it stand out more** – e.g., by having a bolder and larger text or a stronger colour contrast.
- In terms of the visuals, it was felt that **a standardised label would be helpful to make it more easily identifiable**. It was also noted that the **stars could be removed** to make the label clearer and to ensure there is no confusion about their purpose or meaning.
- In general, **there was an appetite for more information about the use of AI** – participants suggested that **the label could be made interactive**, allowing users to click on it for additional information. Others suggested including additional text that would be similar to the one in the ‘readers added context’.

“Probably a bit more detail as to what the scenario sort of is. And what is actually going on, what the circumstances are of what's happening.” – Male, 46, Low digital confidence

“I wouldn't have it in white. I would have it maybe black on yellow. I think that tends to be like [if] you've got visual impairments or something... It doesn't have to be that color, but I would have it bigger and... differently colored.” – Female, 37, Medium digital confidence



As the label is neutral, users are left to interpret whether the content is harmful, which they may not always be able to do confidently



Participants said they would be **most likely to ignore the content** after seeing the 'Made by AI' label. This was because they **would question the authenticity** of the content.



Participants, particularly those with higher levels of knowledge about AI and medium or high digital confidence, might be likely to **research the label** the first time they encountered it, in order to confirm whether the story was true or false. They said they would be likely to do a Google search or look through comments from other users. However, as the meaning of the label was mostly clear, they said they would only be likely to do so **if they were interested in the content**.



Participants said they would be unlikely to report the content, and would only do so **if such content was clearly misleading or potentially harmful**.



There were mixed views about whether they would be likely to click on it. Whilst there was uncertainty about whether it would be **'clickable'**, participants mentioned that they would be **likely** to click on it to get more **context and information** about the content. However, this would depend on their level of interest.

“I might report it, but I don't know. That would depend on how I felt – if it was playing on a story that is true – images of war and that sort of stuff that are in the headlines at the moment... then I would get angry with that.” – Female, 74, Low digital confidence

“I wouldn't trust it, because if it's made by AI, you don't know where it's come from, the source, who created it, what the, the agenda was of the person who created it.” – Male, 40, Low digital confidence

“At the moment, if I was looking at this, there's nothing really to click on. So, I would just close the browser or... I would swipe the pictures away... as soon as that AI was clear to me, then it's something I would probably ignore.” – Male, 45, High digital confidence

4.2 'Readers/Users added context' Label

The video and image scenarios had different text in their respective 'Users/Readers added context' labels

Video



Users added context

While this video appears to address a question raised by a member in this group specifically including naming the individual member, it has not been verified whether AI was used to create this.

Is this note helpful?

Image



Readers added context

AI has been used to generate fake images of celebrities. The hyperlink takes users to a news website which has been voted as a scam on Trustpilot.

Is this note helpful?

The ‘Readers/Users added context’ label was generally felt to be helpful



Readers added context

AI has been used to generate fake images of celebrities. The hyperlink takes users to a news website which has been voted as a scam on Trustpilot.

Is this note helpful?



Users added context

While this video appears to address a question raised by a member in this group specifically including naming the individual member, it has not been verified whether AI was used to create this.

Is this note helpful?

- Participants said they were broadly **unfamiliar with the label**, although some mentioned seeing similar features like user generated warnings on platforms such as X or Facebook.
- In general, **the added context was perceived to be helpful** because of the additional information and/or perspectives that were provided. For the specific labels tested in this study, participants found the additional context to be helpful as it flagged new information that they said would prompt them to be more cautious.
- There were mixed views on who should be able to comment or add posts under ‘Readers/Users added context’. There was a sense that **anyone should be able to contribute** to ensure a range of opinions, however, it was felt that the contributions should be **moderated** or **verified** to maintain accuracy and reliability. Participants also suggested that social media platforms or online moderators should take responsibility for verifying or adding the information.

“Where readers have put their experience and said, stay away from this, it’s very useful, but it is unfortunate that they’ve been duped in the first place. Whereas if there was the first initial Made by AI warning, they might have not been caught, but the most useful one is the readers added content because it’s people’s experiences and warning you to steer clear of this.” – Male, 57, Medium digital confidence

“I think people should be able to go on there and if you’ve got the admin abilities, to go on and post it – ‘it’s AI and it’s fake’. Then I think that helps other people, having to do a bit of research.” – Male, 45, High digital confidence

“I think anyone should be able to [add the context], but... if that comment in context is totally wrong, then you should be able to put a note alongside it to say actually, this comment is wrong. It’s not a scam.” – Male, 45, High digital confidence

Participants appreciated that the label provided a clear message

Readers added context

AI has been used to generate fake images of celebrities. The hyperlink takes users to a news website which has been voted as a scam on Trustpilot.

Is this note helpful?

Users added context

While this video appears to address a question raised by a member in this group specifically including naming the individual member, it has not been verified whether AI was used to create this.

Is this note helpful?

Positives

- ✓ Overall, **participants found the label useful** across all digital confidence and AI knowledge levels.
- ✓ They felt that the information was most useful when it was **clear and definitive**.
- ✓ They also found it particularly **helpful when a hyperlink to a trusted and reputable source was included** (e.g., TrustPilot).

Negatives

- ✗ Participants found information **less helpful when it was perceived to be wordy, vague** and did not specify how or whether AI had been used (e.g., the bottom label on the left).
- ✗ It was pointed out that **the label would be added only after the content has circulated online**, and if it was a scam, those who saw it first could have been potential victims of it.
- ✗ A minority of participants found the visual/icon unclear and were unsure what it referred to.



“I would say it's very helpful because you can see straight away is a scam... just the fact that it says ‘the hyperlink takes users to a news website which has been voted as a scam on Trustpilot’... [Trustpilot] it's pretty accurate, so I'd probably be like ‘okay, this is a scam!’.” – Female, 25, Medium digital confidence

“It's not particularly helpful... it's too wordy and a bit difficult to understand. It just makes it more confusing.” – Male, 61, High digital confidence

“I don't think that's particularly helpful. I think that's just saying exactly what you get from watching the video anyway.” – Female, 25, High digital confidence



Participants wanted a means to verify information in the label and for the label to clarify potential risk



Readers/Users added context

Suggested improvements

- Whilst the label captured attention, participants suggested that **it could be improved by making it more visible** – e.g., having larger or more prominent text, using contrasting colours, or highlighting key terms like ‘scam’ or ‘fake’ to emphasise important information.
- **Participants wanted a platform to provide a system to verify or moderate information in the ‘Readers added context’ label to increase trust in the label.** They also wanted to see relevant hyperlinks for users to be able to verify the information themselves (e.g., a hyperlink to a trusted and reputable website).
- It was important for participants that ambiguity be avoided, and that the context should provide a **warning or clarity about potential risks**, as well as **confirmation about the involvement of AI**.
- It was suggested that adding a 'Yes / No' option next to the ‘Is this note helpful?’ would be useful.

“If it's an AI generated image, I think that should always be on the image somewhere... in an ideal world, it would be great if you could have the bottom bit saying, ‘a lot of people think this is a scam’. Don't know if that's realistic, but it would be great if it was.”
– Female, 37, Medium digital confidence

“[...] the reader's added context... it gives that detail, in this instance, it has been used to generate fake images so, you're automatically aware of exactly that it's fake, and where it's going to go if you click on it, and that it's a scam. So, I think that would probably be the most helpful and serve the most purpose.” –
Male, 46, Low digital confidence

“Probably would have an invigilator to flag that this is a scam because... readers could add whatever they liked.” –
Female, 53, High digital confidence

Likelihood of reporting the content depended on the perceived seriousness of potential risk



The majority would be **likely to ignore** the content as a consequence of seeing the labels tested, primarily due to a desire not to **engage with potentially fake or harmful content**. In the video scenario, participants mentioned that they would be likely to ignore it due to **uncertainty** about whether it was created using AI.



If participants came across this label for the **first time**, they would be likely to **research the meaning of the 'Readers/Users added context'**, particularly those with higher levels of AI knowledge. In the image scenario, whilst the context was felt to be sufficiently clear, a minority mentioned that if they were **interested** in the content, they would want to do **further research** (e.g., if there was a Trustpilot link provided). In the video scenario, participants mentioned that they might be **likely to share** the content to **confirm** the views of others.



In the video scenario, participants frequently mentioned that they would be likely to **report the video or share it with others to warn them**. This was because they felt they have an investment in the welfare of the community they are part of.

In the image scenario, participants said they would be likely to **report it to ensure it does not circulate further**, particularly if it was easy to do it. However, as the content was not felt to be significantly harmful, they might be less inclined to do so.

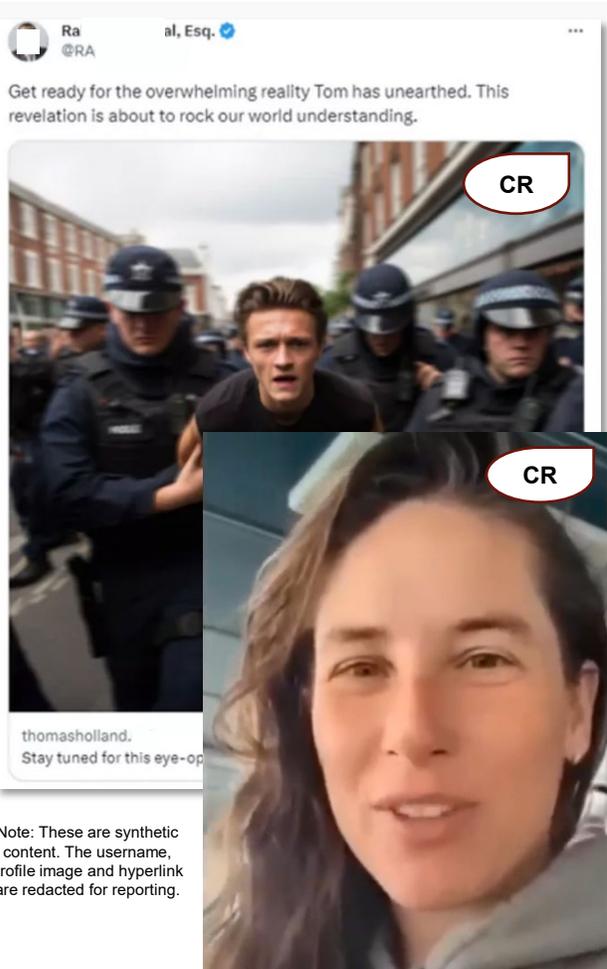
“It would make me question it. So, somebody's raising a potential issue, aren't they? And then that would make me stop and think.” – Female, 53, Low digital confidence

“I think at that point I would do some research on the 'readers added context', kind of looking into that and see what other people were saying at that point as well.” – Male, 31, High digital confidence

“I'd certainly go along with reporting the image. I don't want to engage with it at all. I don't think there's any need to do any further research on it. I'd want it taken down.” – Female, 32, Medium digital confidence

4.3 'CR' Content Credentials Label

The ‘CR’ label was considered least useful due to its unfamiliarity and the need for additional actions



- In general, there was a **lack of clarity** about the purpose of the CR label. Some assumptions were that it could be a logo, or an indication that the post was created using a mobile/web app. Others associated it with a speech bubble, potentially indicating that it would provide additional comments on the content. Participants did not recall coming across the label before.
- When the label came first in rotation (i.e., before seeing other labels), participants did not tend to comment on it and focused on other elements of the content. When probed, they mentioned that this was mostly because they **did not know what the label meant** or assumed it was a **part of the content**. It was more likely to be ignored compared to other labels.
- Among those who were suspicious about the content due to realising that it was AI-generated, the label **did not appear to be trustworthy**, mostly because they were **unsure** about what it meant.
- Many participants did not realise that they could click on the label to see additional information.

“I think I would still be suspicious because I would have kind of scrolled past it thinking it was a scam... But because I don't know what CR is, it makes me want to find that out. Even if it's just the first time I see it, and then I'll know what CR is and then I can use that in future as my judgment.” – Male, 31, High digital confidence

“I would want to know what's the meaning of [CR], what was the acronym CR from?... And I want to see how that relates to this image.” – Female, 32, Medium digital confidence

“The CR logo, I wouldn't know to click on that, so, I wouldn't get the information I needed.” – Male, 45, High digital confidence

”

Once the 'CR' label was clicked on, participants found the pop-up box useful but were unsure if the information there could be trusted

Produced by
Trusted News Ltd

Produced on
12 January 2024

Website
Thomasholland.net

Caption
Tom

Editor(s)
MJ Wilson

Edit date
2 February 2025

AI tool used
ImageJourney

Caption update
Trusted updated news

Produced by
Alla Morgan

Produced on
22 January 2025

Website
AllaMorgan.com

AI tool used
StarClipper.AI

Caption update
Just saying hi

Positives

- ✓ **The website link and AI tool** were perceived to be the **most helpful**. Those with high digital confidence felt that these would allow readers to verify the source, understand that an AI tool was used in creating or editing the content and to do further research, therefore being **transparent** around the content. Those with higher levels of AI knowledge particularly appreciated having information that an AI tool was being used.
- ✓ Other information, such as the 'produced by' and 'produced on', were also seen as potentially helpful if further verification was sought.

Negatives

- ✗ Despite some helpful elements, there was a **lack of clarity about what the information meant** or why it was included, particularly among those with low and medium digital confidence. Participants often expressed being unsure about the relevance of including a date, caption and 'produced by'.
- ✗ Whilst 'AI tool used' was seen as useful by some, others felt that indicating **the tool name was unnecessary**, as they mostly would want to know whether the content was produced by AI or how the AI was used (e.g., to edit or to create the content).
- ✗ Some participants associated clicking with opening another page, which raised concerns of potentially landing on a dubious site.



"I wouldn't really care when it was produced or who did it. I'd want to know why this has happened, why they have created it." – Female, 37, Medium digital confidence

"I think the website is the most important, because whatever research you need to do, you can then look into it from that..." – Male, 61, High digital confidence



Participants wanted to know more about the CR label and how to interpret it

Suggested improvements to the 'CR' label

CR

- Participants would like to have a **clearer explanation** of what 'CR' stands for. It was suggested that it could be **standardised** across platforms and include a brief description or pop-up to **explain its purpose and significance**.
- It was also recommended to make it clear that it could be **clicked on or hovered over for additional information**. A minority suggested including three dots on the top right, which was felt to be universally known as a symbol for further information.

Produced by
Trusted News Ltd

Produced on
12 January 2024

Website
Thomasholland.net

Caption
Tom

Editor(s)
MJ Wilson

Edit date
2 February 2025

AI tool used
ImageJourney

Caption update
Trusted updated news

Produced by
Alla Morgan

Produced on
22 January 2025

Website
AllaMorgan.com

AI tool used
StarClipper.AI

Caption update
Just saying hi

Suggested improvements to the pop-up boxes

- There was no strong sense of a need to remove any of the information. However, participants mentioned that **caption information was not useful or relevant**. There was also a sense that the **'AI tool used' was less helpful**, as they would instead like to know how it was being used.
- There was also a sense that the information **did not provide sufficient context**. As a result, it was suggested that **additional information** would be helpful, including what the information means (e.g., date and 'produced by'), as well as an indication of whether the content can be trusted (this was mentioned particularly when compared to the 'Readers added context' label).

“

“I personally don't find that information helpful, probably what I would be wanting... is a paragraph or so about what this actually means. So, what the picture means, why is it AI generated? What are you trying to achieve by it?”

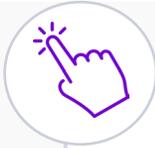
– Female, 37, Medium digital confidence

“I think the CR one is good, but only if there was a better understanding or more transparency about who actually added that information. Like, if it was just the person who produced it, then perhaps it wouldn't be as useful. But if it was done by some kind of like third party or like from the social media site itself to kind of check the content they're putting out. I think that would be the most useful in terms of being very transparent about the person who produced it, which then you can go and research.”

– Female, 24, High digital confidence

”

Potential actions tended to differ based on participants' interest in the content



Once participants understood more about the label, they said they would be **likely to click on the 'CR' icon** to find further information about the content that they were **interested** in.



Researching the 'CR' label was another frequently mentioned potential action, as participants mentioned that they would be curious to find out what it meant. Those who were interested in the trustworthiness of the content were most likely to **look up the website** provided in the pop-up box.



Some participants said that they were not interested in the type of content showed to them during the study and that they would ignore it if they came across it in a real-life situation, or they were not familiar with the label, leading the researchers who moderated the interviews to a conclusion that, outside of the research environment, many participants **would be highly likely to ignore the label**.



"I'd hover over it and see if it expands to give you any further information about the post... I would research about the CR icon... if I saw it for the first time... I'd probably research what is this function, what does it mean, who's producing them." – Female, 24, High digital confidence

"I would ignore it. And maybe do some research on what CR icon means, just because I've not seen it before and then if it pops up [again], I'd know what that means." – Female, 37, Medium digital confidence



5 Views on the AI content labels landscape

Participants called for platforms to label AI-generated content, particularly if it might have harmful impact



There was a broad consensus that **any type of content edited or created using AI should be labelled** to help users make more informed judgements about the authenticity of the content. They believe labelling is **important for making informed opinions, especially in relation to news, politics, and potentially harmful or misleading content**. Additionally, they felt labelling AI-generated content **can prevent false perceptions and protect individuals, particularly children, from assuming such content is real**. Participants felt that social media platforms would be well placed to implement this, particularly when content creators could forget or refuse to do so.



Participants agreed that the **labels should be standardised** to make them clear, consistent and reliable across different platforms, making it easier for users to recognise and understand the labels regardless of where they encounter them.



After seeing AI content labels in this study, **participants generally said they were more aware of and likely to be more critical of online content without such labels**. Some participants said they would be more likely to question the authenticity of unlabeled content.

“[Believe it should be an official requirement] Yes, I do. In this day and age, yes, I do. I do firmly believe that they [labels] should be standardized across, the whole social media platform.” – Medium digital confidence, Male, 69

“I think the platform should be saying this because they're obviously the ones who can do it really because people aren't going to naturally do that on their own, it has to be policed by somebody.” – High digital confidence, Male, 45

“It makes a big difference, especially if it's something that is very serious news, it's harmful to people, it's harmful to children, it's, got to have sort of a warning, because... if it got in the wrong hands or if somebody took that information in and thought it was real, then it could cause a lot of harm.” – Female, 65, Low digital confidence

“I think anything political, I really think they should be labeling that because people can get so up in arms that I really think. You should try and tone it down.” – Female, 37, Medium digital confidence

“I think it would maybe make me more aware... I try and look out for the labels, but if they're not there and the image doesn't look quite right, I would still maybe presume this is AI, even though it's not got the label.” – Female, 37, Medium digital confidence

Participants would welcome a universal system for identifying AI-generated content

Participants would appreciate an accepted system to help easily identify AI-generated or edited online content and help users to establish the credibility of the content and warnings against scams. **Suggestions for such labels included:**



To provide the key message in a very **clear and direct way** e.g. 'made by AI', 'edited by AI', 'enhanced by AI', 'scam', 'fake news'



A verification system using e.g. a stamp or tick to confirm that the information was checked



Stating to what extent an AI tool was used to create the content, to distinguish between minor edits and when the content was fully generated by AI



To **stand out visually** by using a bigger size or contrasting colour



To be able to **access more information without leaving the page** e.g. having hover over text or a pop-up box with additional information



Clearly listing warning information e.g. whether the content is a scam or contains a dubious link, if any information is inaccurate

“It would be very handy if there was a national system adopted to say that something has been created by AI so that would help people. At the moment, I'm not aware that there is anything. So, in theory, I could be mistakenly looking at AI sometimes without knowing.” – Male, 57, Medium digital confidence

“It needs to be there that you don't have to click on into something more because that's where a lot of people are cautious, people are worried that if they click on something... that it's going to take them into something, it's like going into the unknown, into the abyss. It needs that warning information to be there.” – Female, 57, Low digital confidence

“[Labelling any AI-generated or edited online content] ...you would just completely lose all credibility because it would just say edited photo on every single photo. So, I don't think that would be remotely useful. But if it was something that was made by AI, then that would be really useful because the whole image would be fake rather than being like an element of the image.” – Female, 32, High digital confidence

Suggestions were given on how online services could increase awareness and understanding of such labels



Pop-ups or banners to provide explanations



Sending **notifications or emails** to users who intend to upload AI-edited or generated content



Incorporating **interactive elements** that provide more information when a user clicked on such label



Advertising and educational campaigns about such labels informing about their purpose and significance

Regularly displaying these labels and sending updates or notifications to users can also help familiarise them with the concept. Additionally, **providing concise and engaging explanations** can further enhance user understanding.



“[Social media sites, websites, platforms] they could have little pop ups, like, just a reminder, just saying this... It makes it a little bit more engaging. They don't have to do it forever, but just while it's getting embedded.”
– Female, 37, Medium digital confidence

“I think they should send out the information on what the labels are, how, the labels come to be. And where they're used, and also where to report content if you think it is AI created and not labeled.” – Female, 74, Low digital confidence

“When you see these tags, maybe you can hover over it, and it will explain a bit more information about what that actually means.” – Female, 24, High digital confidence



6 Annex

Sources of the Stimulus

The stimulus shown to participants was AI-generated content taken from publicly available and freely accessible sources. Researchers added different types of AI content labelling ('Made By AI', 'Readers/Users added context' and 'CR') to indicate that the image/video was made using AI. These three types of labels were selected as they are some of the measures used by services to inform their users currently.

Hyperlinks to the sources are provided and the publishers acknowledged below.

Mirror <https://www.mirror.co.uk/news/uk-news/elon-must-act-stop-tide-30451315>

Elon must act to stop the tide of frauds swamping Twitter

Ed Sheeran, Bear Grylls, Daniel Radcliffe and Tom Holland have all had their names exploited in scam tweets. So, ironically, has Elon Musk.

BBC News <https://www.bbc.co.uk/news/articles/cdr0g1em52go>
Good Morning Britain on YouTube <https://www.youtube.com/watch?v=1Kv5njt9X5c>

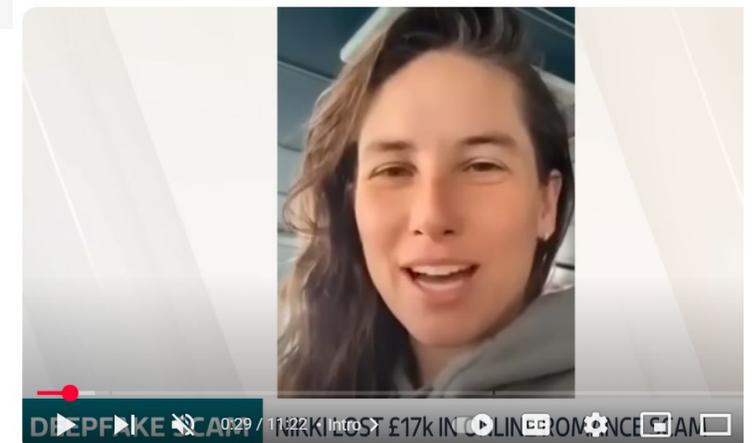
NEWS

[Home](#) | [InDepth](#) | [Israel-Gaza war](#) | [War in Ukraine](#) | [Climate](#) | [UK](#) | [World](#) | [Business](#) | [Politics](#) | [Culture](#)

[Scotland](#) | [Scotland Politics](#) | [Scotland Business](#) | [Edinburgh, Fife & East](#) | [Glasgow & West](#) | [Highlands & Isl](#)

[Alba](#)

Romance scammer duped £17k from me with deepfakes



77-Year-Old Scammed Out of £17,000 by Deepfake Scammer Using AI-Generated Videos

End

YouGov, 2024, all rights reserved. All materials contained herein are protected by copyright laws. Any storage, reproduction or distribution of such materials, in whole or in part, in any form without the prior written permission of YouGov is prohibited. This information (including any enclosures and attachments) is proprietary and confidential and has been prepared for the exclusive use and benefit of the addressee(s) and solely for the purpose for which it is provided. We make no representations, warranties or guarantees, whether express or implied, that the information is accurate, complete or up to date. We exclude all implied conditions, warranties, representations or other terms that may apply and we will not be liable to you for any loss or damage, whether in contract, tort (including negligence), breach of statutory duty, or otherwise, even if foreseeable, arising under or in connection with use of or reliance on the information. We do not exclude or limit in any way our liability to you where it would be unlawful to do so.