BCLP. Client Intelligent

Insights

DON'T BE SHY, DISCLOSE AI!

COMPLYING WITH SOCIAL MEDIA USER GUIDELINES WHEN USING GENERATIVE AI TOOLS / DIGITAL SPEAKS SERIES

Jan 31, 2024

With adoption of generative AI tools scaling exponentially amongst content creators, companies need to be vigilant to ensure all content posted on social media platforms complies with user guidelines about AI, especially where this is created by third party creators.

WHAT'S THE LATEST?

The end of 2023 saw a wave of implementation of new content labelling tools, prompted in part by global concerns about disinformation and the role AI tools can play in flooding social media platforms with (deliberately) inaccurate AI-generated content. TikTok has been at the forefront, recently unveiling a new tool, designed to help creators label content that has been generated or significantly edited by AI.

This development is to support 'transparent and responsible content creation practices' and reduce the risk of confusing and misleading viewers, allowing creators to inform their community of AI-generated content, through the use of digital labelling, so creators make it clear 'when content has been significantly altered or modified by AI'. TikTok will also be releasing educational videos and resources, to explain how and why to use these labels.

WHAT DOES THIS MEAN FOR CONTENT?

Creators (both individuals and corporate content creators) will be required to disclose their AI-generated content in order to comply with TikTok's Community Guidelines on integrity and authenticity, including its misinformation and impersonation policies. TikTok also recently rolled out its synthetic media policy, requiring people to label AI-generated content that contains 'realistic images, audio or video, in order to help viewers contextualise the video'. This is also aimed at preventing content which can mislead viewers spreading online, with the requirement for clear disclosure designed to combat this. This disclosure does not necessarily need to be made through the use of the new labelling tool, and can instead be done through the use of a sticker or a note in the caption. If AI-generated content is not disclosed, this may be deemed to be a violation of

TikTok's Terms of Service. TikTok may also take down AI-generated content that's not labelled. In some cases, we understand this may mean action could be taken against both content and accounts that violate the company's policy.

The guidelines also do not allow AI-generated content that contains the likeness of any real figure, when used for political or commercial endorsements. In order to enforce the new AI-generated content regulations, viewers can report content that they see on the app, if it is not correctly labelled. This, we understand, will then lead to TikTok reviewing the content. These efforts to support responsible content creation practices, come after TikTok backed the launch of the 'framework for responsible use of AI-generated media'. The framework sets out guidance recommendations for those creating, sharing and distributing AI content.

TikTok will also be testing out an automated 'AI-generated' content label, that will detect when a video has been edited or created with AI, and then automatically applying the label to content. TikTok has also added a number of filters (that use AI) to the app and is to rename the effects to explicitly include 'AI' in their name. This is to ensure further transparency for viewers about AI-generated content.

WHAT'S NEXT?

Although as yet there is nothing on the horizon requiring disclosure of *all* AI-generated content on other social media platforms, given upcoming elections around the world, new rules have been announced by other platforms relating to the use of generative AI in political adverts.

One platform recently announced it will be launching new global guidelines so users can understand when a social issue, election, or political advertisement has been digitally created or altered, including through the use of AI, effective early in 2024. This will require advertisers to disclose whenever a social issue, electoral, or political ad contains a photorealistic image or video, or realistic sounding audio, that was 'digitally created or altered to depict:

- a real person as saying or doing something they did not say or do; or

- a realistic-looking person that does not exist or a realistic-looking event that did not happen, or alter footage of a real event that happened; or

- a realistic event that allegedly occurred, but that is not a true image, video, or audio recording of the event'.

To ensure user compliance, we understand the platform will reject ads that do not disclose the use of AI and repeated failure to disclose AI usage may result in penalties. It is also banning manipulated media, such as deepfakes (a video of a person in which their face or body has been digitally altered), recognising that these types of images are increasingly being made with the use

of AI tools. It will therefore remove misleading manipulated media if it has been edited or synthesised or is the product of AI that replaces content onto a video, making it appear authentic.

WHAT ARE OTHER KEY PLAYERS DOING?

In tandem, Sony, Nikon and Canon are collaborating on a watermarking standard to show that an image is **not** made by AI. The new camera technology will embed digital signatures in images so that they can be distinguished from fakes (also known as deepfakes). The signatures will be resistant to tampering and will contain data such as date, time, location and photographer. The feature will be made available in their mirrorless cameras (including smartphones). The companies have also teamed up with global news organisations to launch a web-based tool, Verify, to enable users to confirm the legitimacy of photographic images free of charge. The tool will flag an image as having 'No Content Credentials' if an image has been tampered with using AI.

ANY TIPS FOR MANAGING THIS IN PRACTICE?

Yes! Managing your organisation's use of AI generated materials should involve:

- Getting content creators to confirm that they own (or have the right to use) all the rights in any content you intend to use on social media channels and ensure you are given an assignment or right to use the content for the purposes of distribution on social media platforms.

- Requiring content creators to confirm any applicable labels which should be applied to content created using AI technology.

- Considering use of tools to check for AI usage.

- Checking that all of your employees understand the risks of using generative AI tools to produce content and to flag where they have done so.

- Considering if you require the creator to give you some comfort (potentially via an indemnity) that your usage of the content will not expose you to any third party claims (in particular, IP claims).

The above commentary is not intended to constitute legal advice regarding compliance with any terms of use of social media platforms. Please do get in touch if you would like to discuss further at all.

RELATED CAPABILITIES

- Technology Transactions
- Intellectual Property & Technology Disputes

MEET THE TEAM

Anna Blest

London

anna.blest@bclplaw.com
+44 (0) 20 3400 4475

Amelda Henning

London

amelda.henning@bclplaw.com
+44 (0) 20 3400 3004

Kate Jeffery

London

kate.jeffery@bclplaw.com
+44 (0) 20 3400 3667