Google's approach to content provenance:
Understanding the origins of content in the age of AI

As we continue to bring AI to more products and services to help fuel creativity and productivity, we are focused on helping people better understand how a particular piece of content was created and modified over time. We believe it's crucial that people have access to information about the source and history — the provenance — of digital content, and we are investing heavily in tools and innovative solutions to provide it.

Identifying AI-generated content with SynthID

SynthID is Google's state-of-the-art watermarking tool, designed specifically for AI-generated content. It empowers users to identify AI-generated (or altered) content, helping to foster transparency and trust in generative AI.

Unlike traditional watermarks, which are often applied like a stamp on an image and can easily be edited out, SynthID embeds digital watermarks directly into AI-generated images, audio, text, or video. For example, for images, the watermark is embedded within the pixels of an image. The watermarks are included across Google's generative AI consumer products, and are imperceptible to humans — but can be detected by SynthID's technology.

SynthID not only preserves the quality of a given piece of content, it acts as a robust watermark that remains detectable even when the content is shared or undergoes a range of transformations, such as screenshots or cropping. While originally focused only on AI-generated imagery, we've since expanded SynthID to cover AI-generated text, audio, and video content across Google's generative AI consumer products. Over 20 billion pieces of content have already been watermarked with SynthID.

In November 2025, we announced that users can now verify if content (image, audio, or video) was generated with or edited by Google AI right in the Gemini app. If you want to confirm if a particular piece of content was made by Google AI, upload it to the Gemini app and ask a question such as: "Was this created by Google AI?" or "Is this AI generated?" Gemini will check for the SynthID watermark and use its own reasoning to return a response that gives you more context.



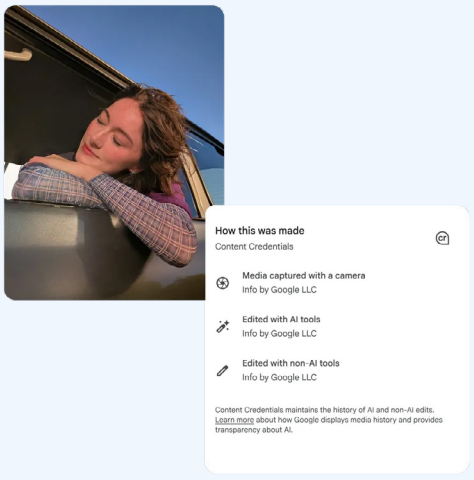
Advancing content transparency and authenticity with C2PA

While SynthID is effective for watermarking and identifying content generated through AI, it is just as important to provide strong signals about non-synthetic content, such as photos and videos that come from cameras. That's why we're helping develop the Coalition for Content Provenance and Authenticity (C2PA) and their Content Credentials technology.

Content Credentials convey a rich set of information about how media such as images, videos, or audio files were made, protected by the same digital signature technology that has secured online transactions and mobile apps for decades. It empowers users to identify AI-generated (or altered) content, helping to foster transparency and trust in generative AI.

Content Credentials are an industry standard backed by a broad coalition of leading companies for securely conveying the origin and history of media files. The standard is developed by the C2PA, of which Google is a steering committee member.

In August 2025, Google announced that the new Pixel 10 family of devices will support Content Credentials in Pixel Camera and Google Photos. This makes Pixel the first smartphone to implement Content Credentials within its native camera app. The full process takes place on-device, where we create secure metadata within the image that documents its full journey. This information will be available directly within Google Photos.



Google

Learn more at **publicpolicy.google/content**

Advancing content transparency and authenticity with C2PA

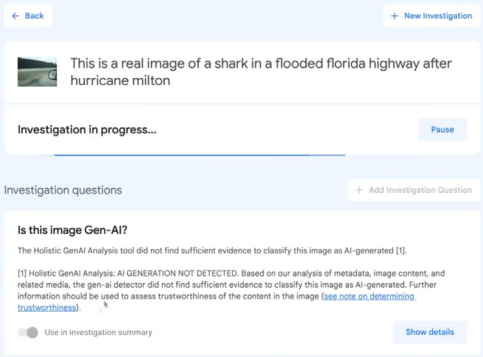
Images and videos generated with the Gemini App come with Content Credentials. And if users come across an image that contains Content Credentials or a SynthID watermark, they can use tools in Google Search like About this image to see if it was created or edited with AI tools. And on YouTube, we relay information from Content Credentials to viewers when content is captured with a camera that has built-in C2PA support, or if the content was fully generated by an AI model. This complements YouTube's "altered or synthetic content" disclosure, which creators are required to add if a video contains realistic-looking content that has been meaningfully altered or synthetically generated.

Exploring the context of online images with Backstory

Because simply asking "Is this generated by AI?" does not suffice in assessing content trustworthiness, we work closely with leading information literacy experts around the world to understand the latest research and help ensure that our products are empowering users with information and tools to cross-check what they find online.

One such innovation we are exploring in this space is Backstory, an experimental AI tool from Google DeepMind that surfaces information and helps people learn more about the context and provenance of images seen online. When given an image and a written prompt, Backstory investigates whether an image was AI-generated, when and where it's previously been used online, and whether it's been digitally altered.

Built using Gemini, Backstory draws on a number of technologies designed to identify whether an image was created using generative AI. Backstory then combines this with a more holistic assessment of the context of the image. It surfaces how the image has been used on the internet over time and other information like metadata to help answer the user's written prompt.



Our work in this space is far from done. At the speed of technological change, something that works today may not be sufficient next year, and we understand that technological solutions alone are not sufficient.

We're committed to working with our users, partners, experts — and all those interested in helping people having an easier time deciding what to trust online — to continuously evolve and update our approach in this space.