AI Explained · Images, Voice and Video, Part 2

When the Sheep Starts Moving

By Dr Peter Fields

THE 2-MINUTE SUMMARY

A generated voice has to get the words, the sound of the speaker and the delivery right. A generated video has to keep a scene consistent as things move. Add sound and it has to arrive at the right moment.

More computing power has helped. So have better ways to describe training footage and to represent movement across a whole clip. A video can look better and still get the action wrong.

Your own creative work can be protected even if you used AI. For an image generated without a human author, start by asking whether copyright exists. Then ask who arranged for that particular image to be made.

The sheep takes a bite

Picture the sheep from Part 1. This time it lowers its head, bites a tuft of grass and lifts its head again. Somebody says, “It’s still hungry.” The bite has to sound at the right moment, and the delivery makes a difference too. Is the speaker amused or worried? Once you have a clip worth using, there’s another question. Who owns it?

How a voice is made

Try saying “You’re coming” twice. The first time it’s a statement. The second time it’s a question. The words are identical, but the pitch and emphasis change what the listener hears. Those features, together with rhythm and pauses, are part of what speech researchers call prosody.

One way of generating speech is to turn sound into compact codes, or audio tokens. In a research system called VALL-E, the words to be spoken and a short reference recording guide the prediction of audio codes. A decoder turns those codes back into audible speech. In the researchers’ tests, the output could pick up qualities of the speaker and even features of the recording, such as the sound of a room with an echo.[1]

The change was partly in how speech was represented and partly in the scale of training. The researchers trained VALL-E on about 60,000 hours of English speech. Even when a generated voice sounds right, it can stress the wrong word or stumble over a name. Try a short passage first and listen for pronunciation, pauses and emphasis. The companion Deepfakes articles deal with impersonation and the use of a person’s face or voice.[1]

How a video holds together

Pause a clip of the grazing sheep and the frame may look perfectly plausible. Play it through and the animal's ear may change shape, or its muzzle may pass straight through the grass. The model has to keep the subject and its surroundings consistent as they move.[2,3]

The first Sora research report describes one approach. Compress the footage, then represent it in small pieces that cover part of the picture and a short span of time. The model can learn relationships among those pieces. Lumiere used a different design to generate the motion across a whole clip at a lower resolution before refining the picture. Both illustrate how models can account for movement across a sequence.[2,4]

Why has it improved?

Computing power is part of the answer. Peebles and Xie tested diffusion transformers on ImageNet, a set of still images. Models using more computation per pass tended to score better on their image-quality measure. The original Sora report points to another change. More detailed descriptions of the footage used for training helped the generated clips follow the brief.[2,5]

Sound was one advance. OpenAI's Sora 2, launched in 2025, could generate a clip with synchronised sound. OpenAI closed the app and website in April 2026, followed by the API in September.[6]

Editing control has improved too. Google's Gemini Omni 1.2 Flash, launched in November 2026, can use reference images, short motion clips and audio cues, then revise a clip through successive instructions. In one of Google's demonstrations, the background changes while the characters and framing stay recognisable.[7]

To try it yourself, give the sheep one action and keep the camera steady. Watch the clip at normal speed, then replay the bite. If you make a follow-up edit, check whether it changed anything you wanted to keep.[3]

Who owns the output?

A designer draws a shopfront and asks AI for a night-time backdrop. She combines the two, redraws the reflections and adds her own lettering. Then the client asks for exclusive rights to the finished picture. The answer may differ for the sketch, the backdrop and the picture she puts together.

Start with the human work. The original sketch, and the finished composition if it reflects the designer's own creative choices, can qualify for copyright. The backdrop is harder. The usual test for originality looks for a person's creative choices. The Government's report points out how difficult that test is to apply to something made without a human author.[8,9]

Who counts as the author of the backdrop?

Section 9(3) of the Copyright, Designs and Patents Act 1988 deals with computer-generated works. If the backdrop qualifies for copyright, the Act treats whoever made the arrangements needed to create it as the author. In 1988, Parliament was thinking about outputs such as weather maps and reports from expert systems. Supporters of the rule argued that clearer rights would give people confidence to invest in the technology.[8,9,10]

In its March 2026 report, the Government took the view that the person who enters the request will usually count as the author when a general-purpose AI responds. Think of the provider supplying a tool that can produce all sorts of pictures and the user supplying the brief for this particular one. How the statutory test applies to a typical text-to-image request remains open.[8,9]

A computer pool game gives an example where the programmer made the key arrangements. The player decided how to take the shot. The High Court looked at who devised the visuals and wrote the rules and program that created the frames, and treated that person as the author. An image generator gives the user a different kind of input. The facts about who shaped the result still matter.[11]

The Government now questions how much this protection does to encourage investment when another output can be made so cheaply. It proposed removing the protection unless there is evidence of continuing value. Section 9(3) remains in force as at 18 November 2026.[8,9]

What can the designer offer the client?

She can offer exclusive rights in her own original work to the extent she owns those rights and is free to grant them. If she made it as an employee in the course of her job, the employer would usually own the copyright. For the backdrop, she needs to check whether any copyright exists, who owns it and what the tool's terms allow. A licence to use the image can let her put it in the project. Exclusivity requires rights the client can enforce against others.[8,9]

She also needs permission for any protected material she uses, unless an exception applies. And she should look at the result itself. Researchers have shown that some image models can produce near-identical versions of training examples under particular test conditions. In the UK, copying a substantial part of a protected work can infringe copyright.[8,9,12]

In Getty Images v Stability AI, the High Court held that the versions of Stable Diffusion before it had never contained or stored copies of Getty's works and were therefore not infringing copies. Getty appealed that ruling. The appeal concerns the model itself. The designer still has to check the image she plans to deliver.[13]

Geek Box · A picture, a film and a recording

UK law treats films and sound recordings separately. The Act identifies the producer as the author of a sound recording, and the producer and principal director as the authors of a film. Rights in the recording and rights in the performance also need separate consideration. In England, the use of a recognisable face or voice can raise questions of privacy, data protection or passing off, for example where a likeness falsely suggests endorsement.[8,9,14]

Before handing it over

Keep the sketch, reference files and working versions. They show what you brought to the finished picture. If a client asks for exclusivity, the agreement should identify the elements it covers.[8,9]

Sources and notes

1. Chengyi Wang and colleagues (2023), Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers, arXiv:2301.02111, §§3–5.4 and 6. Paper

2. OpenAI (2024), Video Generation Models as World Simulators, sections on visual patches, scaling, language understanding and limitations. Technical report

3. Jing Gu and colleagues (2026), PhyWorldBench: A Comprehensive Evaluation of Physical Realism in Text-to-Video Models, ICLR, §§3–4 and Table 3. Conference paper

4. Omer Bar-Tal and colleagues (2024), Lumiere: A Space-Time Diffusion Model for Video Generation, §3 and Fig. 2. Paper

5. William Peebles and Saining Xie (2023), Scalable Diffusion Models with Transformers, ICCV, pp. 4195–4205, especially §4 and Figs. 2 and 6. Conference paper

6. OpenAI (2025), Sora 2 System Card, overview. OpenAI, What to Know About the Sora Discontinuation. System card · Shutdown dates

7. Google (5 November 2026), Gemini Omni 1.2 Flash, launch examples and published capabilities. Launch announcement

8. Copyright, Designs and Patents Act 1988, ss 1, 4, 5A, 5B, 9, 11, 12(7), 16–17, 90, 178 and Part II. For a qualifying computer-generated work, protection lasts 50 years from the end of the year it was made. Current text checked 18 November 2026. Act · Authorship · First ownership

9. UK Government (March 2026), Report on Copyright and Artificial Intelligence, sections B, G, I (especially paras 1 and 13–15) and J. On originality, THJ Systems Ltd v Sheridan [2023] EWCA Civ 1354, [15]–[16] and [23]–[27]. Government report · THJ Systems

10. HL Deb, 12 November 1987, vol. 489, Copyright, Designs and Patents Bill, second reading. Hansard

11. Nova Productions Ltd v Mazooma Games Ltd [2006] EWHC 24 (Ch), [104]–[106]. Judgment

12. Nicholas Carlini and colleagues (2023), Extracting Training Data from Diffusion Models, USENIX Security 32, pp. 5253–5270, §§3–4 and Figs. 2–3. Conference paper

13. Getty Images v Stability AI [2025] EWHC 2863 (Ch), [599]–[609] and [758]. Getty's appeal concerns secondary infringement, case CA-2026-000007. Status checked 18 November 2026. High Court judgment · Appeal submission

14. Fenty v Arcadia Group Brands Ltd [2015] EWCA Civ 3, [29]–[35], on the absence of a general English image right and the application of passing off. See also the Government report, section J. Court of Appeal judgment