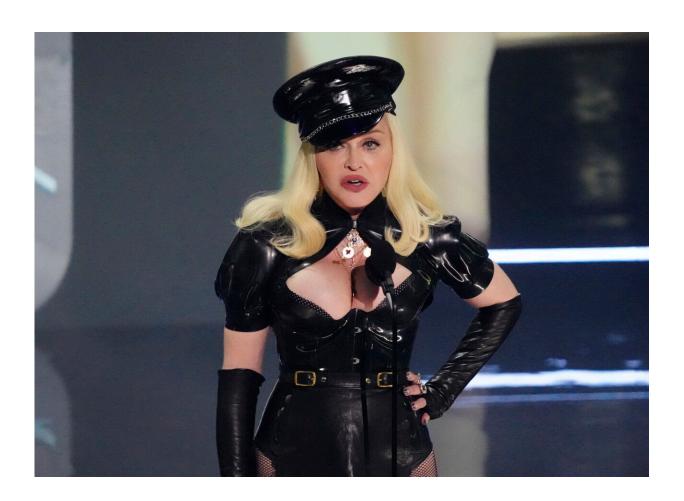# Do AI video-generators dream of San Pedro? Madonna among early adopters of AI's next wave

March 4 2024, by Matt O'brien

Madonna speaks at the MTV Video Music Awards at Barclays Center on Sept. 12, 2021, in New York. Making instant videos is the next wave of generative artificial intelligence, much like chatbots and image-generators before it. And the pop star Madonna is among the early adopters. Madonna's team used an AI text-to-video tool to make moving images of swirling clouds featured in her

ongoing Celebration Tour.Credit: Photo by Charles Sykes/Invision/AP, File

Whenever Madonna sings the 1980s hit "La Isla Bonita" on her concert tour, moving images of swirling, sunset-tinted clouds play on the giant arena screens behind her.

To get that ethereal look, the pop legend embraced a still-uncharted branch of generative artificial intelligence—the text-to-video tool. Type some words—say, "surreal cloud sunset" or "waterfall in the jungle at dawn"—and an instant video is made.

Following in the footsteps of AI chatbots and still image-generators, some AI video enthusiasts say the emerging technology could one day upend entertainment, enabling you to choose your own movie with customizable story lines and endings. But there's a long way to go before they can do that, and plenty of ethical pitfalls on the way.

For early adopters like Madonna, who's long pushed art's boundaries, it was more of an experiment. She nixed an earlier version of "La Isla Bonita" concert visuals that used more conventional computer graphics to evoke a tropical mood.

"We tried CGI. It looked pretty bland and cheesy and she didn't like it," said Sasha Kasiuha, content director for Madonna's Celebration Tour that continues through late April. "And then we decided to try AI."

ChatGPT-maker OpenAI gave a glimpse of what sophisticated text-to-video technology might look like when the company recently [showed off Sora](link), a new tool that's not yet publicly available. Madonna's team tried a different product from New York-based startup Runway, which helped pioneer the technology by releasing its first public text-to-video model

last March. The company released a more advanced "Gen-2" version in June.

Runway CEO Cristóbal Valenzuela said while some see these tools as a "magical device that you type a word and somehow it conjures exactly what you had in your head," the most effective approaches are by creative professionals looking for an upgrade to the decades-old digital editing software they're already using.

He said Runway can't yet make a full-length documentary. But it could help fill in some background video, or b-roll—the supporting shots and scenes that help tell the story.

"That saves you perhaps like a week of work," Valenzuela said. "The common thread of a lot of use cases is people use it as a way of augmenting or speeding up something they could have done before."

Runway's target customers are "large streaming companies, production companies, post-production companies, visual effects companies, marketing teams, advertising companies. A lot of folks that make content for a living," Valenzuela said.

Dangers await. Without effective safeguards, AI video-generators could threaten democracies with convincing "deepfake" videos of things that never happened, or—as is already the case with AI image generators—flood the internet with fake pornographic scenes depicting what appear to be real people with recognizable faces. Under pressure from regulators, major tech companies have promised to watermark AI-generated outputs to help identify what's real.

There also are copyright disputes brewing about the video and image collections the AI systems are being trained upon (neither Runway nor OpenAI discloses its data sources) and to what extent they are unfairly

replicating trademarked works. And there are fears that, at some point, video-making machines could replace human jobs and artistry.

For now, the longest AI-generated video clips are still measured in seconds, and can feature jerky movements and telltale glitches such as distorted hands and fingers. Fixing that is "just a question of more data and more training," and the computing power on which that training depends, said Alexander Waibel, a computer science professor at Carnegie Mellon University who's been researching AI since the 1970s.

"Now I can say, 'Make me a video of a rabbit dressed as Napoleon walking through New York City,'" Waibel said. "It knows what New York City looks like, what a rabbit looks like, what Napoleon looks like."

Which is impressive, he said, but still far from crafting a compelling storyline.

Before it released its first-generation model last year, Runway's claim to AI fame was as a co-developer of the image-generator Stable Diffusion. Another company, London-based Stability AI, has since taken over Stable Diffusion's development.

The underlying "diffusion model" technology behind most leading AI generators of images and video works by mapping noise, or random data, onto images, effectively destroying an original image and then predicting what a new one should look like. It borrows an idea from physics that can be used to describe, for instance, how gas diffuses outward.

"What diffusion models do is they reverse that process," said Phillip Isola, an associate professor of computer science at the Massachusetts Institute of Technology. "They kind of take the randomness and they congeal it back into the volume. That's the way of going from

randomness to content. And that's how you can make random videos."

Generating video is more complicated than still images because it needs to take into account temporal dynamics, or how elements within the video change over time and across sequences of frames, said Daniela Rus, another MIT professor who directs its Computer Science and Artificial Intelligence Laboratory.

Rus said the computing resources required are "significantly higher than for still image generation" because "it involves processing and generating multiple frames for each second of video."

That's not stopping some well-heeled tech companies from trying to keep outdoing each other in showing off higher-quality AI video generation at longer durations. Requiring written descriptions to make an image was just the start. Google recently demonstrated a new project called Genie that can be prompted to transform a photograph or even a sketch into "an endless variety" of explorable video game worlds.

In the near term, AI-generated videos will likely show up in marketing and educational content, providing a cheaper alternative to producing original footage or obtaining stock videos, said Aditi Singh, a researcher at Cleveland State University who has surveyed the text-to-video market.

When Madonna first talked to her team about AI, the "main intention wasn't, 'Oh, look, it's an AI video,'" said Kasiuha, the creative director.

"She asked me, 'Can you just use one of those AI tools to make the picture more crisp, to make sure it looks current and looks high resolution?'" Kasiuha said. "She loves when you bring in new technology and new kinds of visual elements."

Longer AI-generated movies are already being made. Runway hosts an

annual AI film festival to showcase such works. But whether that's what human audiences will choose to watch remains to be seen.

"I still believe in humans," said Waibel, the CMU professor. "I still believe that it will end up being a symbiosis where you get some AI proposing something and a human improves or guides it. Or the humans will do it and the AI will fix it up."

Citation: Do AI video-generators dream of San Pedro? Madonna among early adopters of AI's next wave (2024, March 4) retrieved 9 May 2024 from https://techxplore.com/news/2024-03-ai-video-generators-san-pedro.html