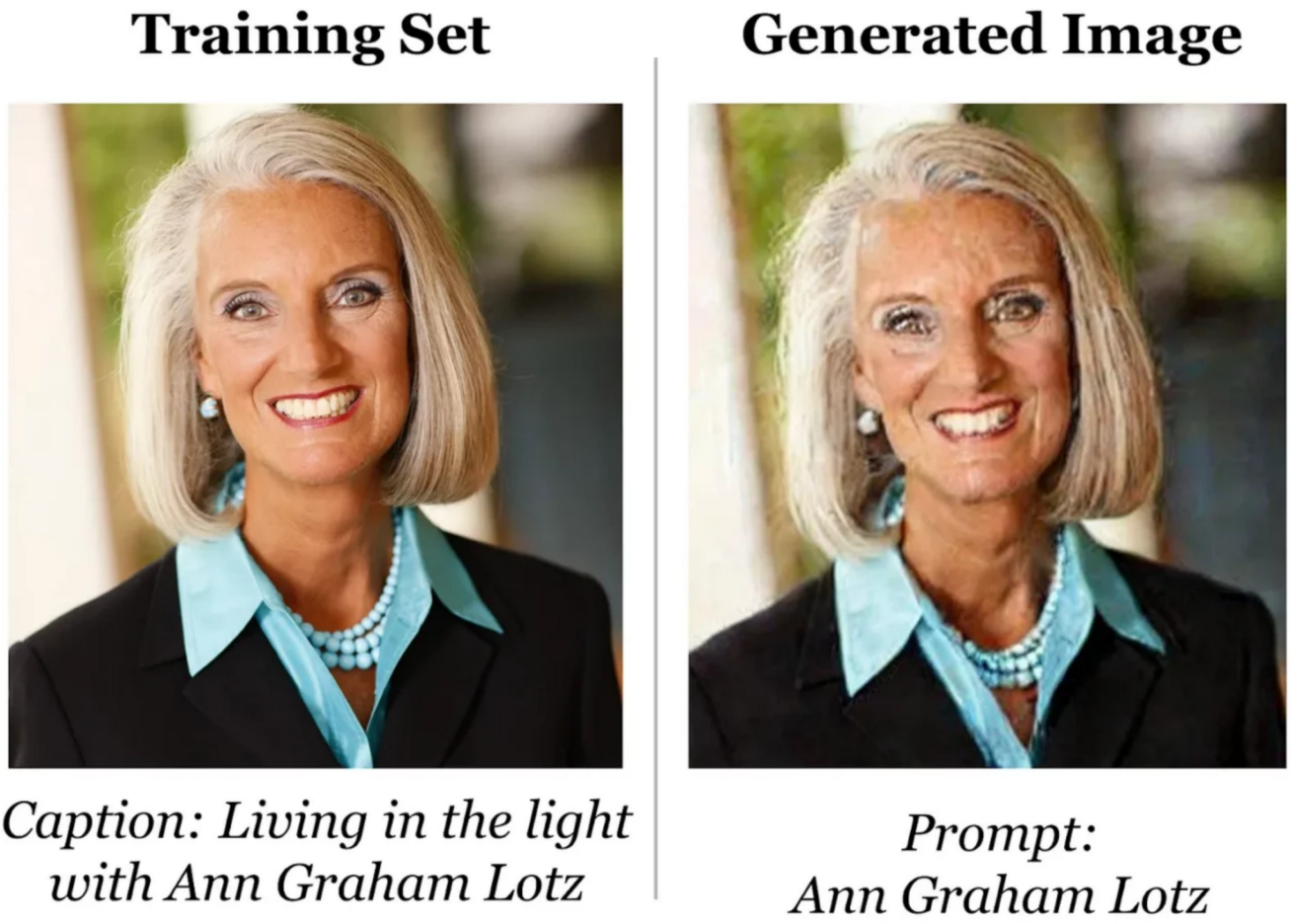
Stable Diffusion's copying problem

Here's one of the most awkward examples for Stability AI:



Enlarge

This example comes courtesy of this paper by researchers at Google and several universities. On the left is an image of Ann Graham Lotz, daughter of famed evangelist Billy Graham, from Stable Diffusion's training data (it's also on Lotz's Wikipedia page). On the right is a picture of Lotz generated by Stable Diffusion. It's not a perfect copy, but it clearly is *a* copy.

Stable Diffusion doesn't generate direct copies like this very often. Researchers tried to reproduce 350,000 images from Stable Diffusion's training set but only succeeded with 109 of them—a success rate of 0.03 percent. And in ordinary use—with users who aren't trying to get the software to reproduce training images—verbatim copies like this should be even less common.

Enlarge

Here's another example from a recent research paper. These computer scientists used training image captions as Stable Diffusion prompts and found that they could generate an image similar to a training image 1.88 percent of the time. They used a much looser test for image similarity—loose enough that it's not clear how many of these cases would trigger copyright liability. And again, ordinary users should see this much less frequently since, unlike the researchers, they're not trying to produce copies of training images.