



Today Generative AI Is Just a Parlor Trick

Michael Zyda , University of Southern California

Generative AI is all the rage for its ability to “imagine” artistic single images and 17-s “photorealistic” videos, without interacting with a single real artist. It looks cool but is currently just a parlor trick.

You can’t swing a dead cat without hearing about some sort of new generative AI trick that is just going to be the bomb diggity. Generative AI is an application that sits on top of a large language model (LLM) and is best known for allowing the input of a textual description of something you want to see and then computing an image that it guesses might match your request depending on how detailed your request was and how long you are willing to wait for that computation to take (Figure 1).¹ Some of the best-known generative AI models that “imagine” images

are Dall-E, Midjourney, and Google Gemini.¹ With these models, we can ask them for an image of “a dead cat bounce” and get something like the images in Figure 2.

WHERE ARE WE NOW?

Now, previous to this, I had no idea what “a dead cat bounce” was and am somewhat happy to be en-

lightened by Midjourney at this moment. You can even ask Midjourney to “/imagine a dead cat bounce from one space craft to another under a moonlit sky” and get something that is not completely crazy (Figure 3).

And if we go online to the Oxford Languages web page, we find that Midjourney’s response is not even close to what the real definition is (Figure 4).

And maybe those speculators are all invested in generative AI app-building companies to try and make up for missing the Nvidia run-up on which almost all of the generative AI computing happens.

Right now, generative AI mostly “imagines” text to a single image, but lately, OpenAI’s Sora has started being able to generate short, 30 s or less, videos of detailed

Digital Object Identifier 10.1109/MC.2024.3370292
Date of current version: 6 May 2024



“photorealistic scenes.”² The longest video that Levy² saw was all of 17 s long and required a pretty detailed textual description like the following:

“Beautiful, snowy Tokyo city is bustling. The camera moves through the bustling city street, following several people enjoying the beautiful snowy weather and shopping at nearby stalls. Gorgeous sakura petals are flying through the wind along with snowflakes.”²

Now, I am unimpressed. It is not really photorealistic; that is hyperbole. You want to impress me, write the text that generates the complete Pixar film, *Toy Story*. *Toy Story* was the first completely computer-animated feature film. It is 81 min long. Do that, and I may change my mind!

WHERE DO WE WANT TO GO?

Well, this is the “Games” column, and more and more I find myself writing in the overlap between interactive experiences and filmed entertainment. There are a number of things we would like to do to make generative

just a small idea on a single picture, not a 17-s feature film.

Desire for a workspace kind of tool

Another thing we want is a way better interface than just typing the text over and over again with some small

We would like to describe our desired 3D objects/
 characters/scenes in text and then send them to be
 rendered and have that happen near instantly.

AI more useful. One is performance related. We would like to describe our desired 3D objects/characters/scenes in text and then send them to be rendered and have that happen near instantly. Right now, if I send some small text to Midjourney like “/ imagine a dead cat bounce,” it takes about 90 s. That is way too long for iterating concept art ideas, and this is

changes. A simple thing I would love is a workspace where I type in my first text and then can edit what I typed with small tweaks to the text; then I hit a “go button,” and it goes and does the computation and returns in one second with a high-resolution image of what I wanted imagined.

Another thing I would like is art style control. Again, I would love that

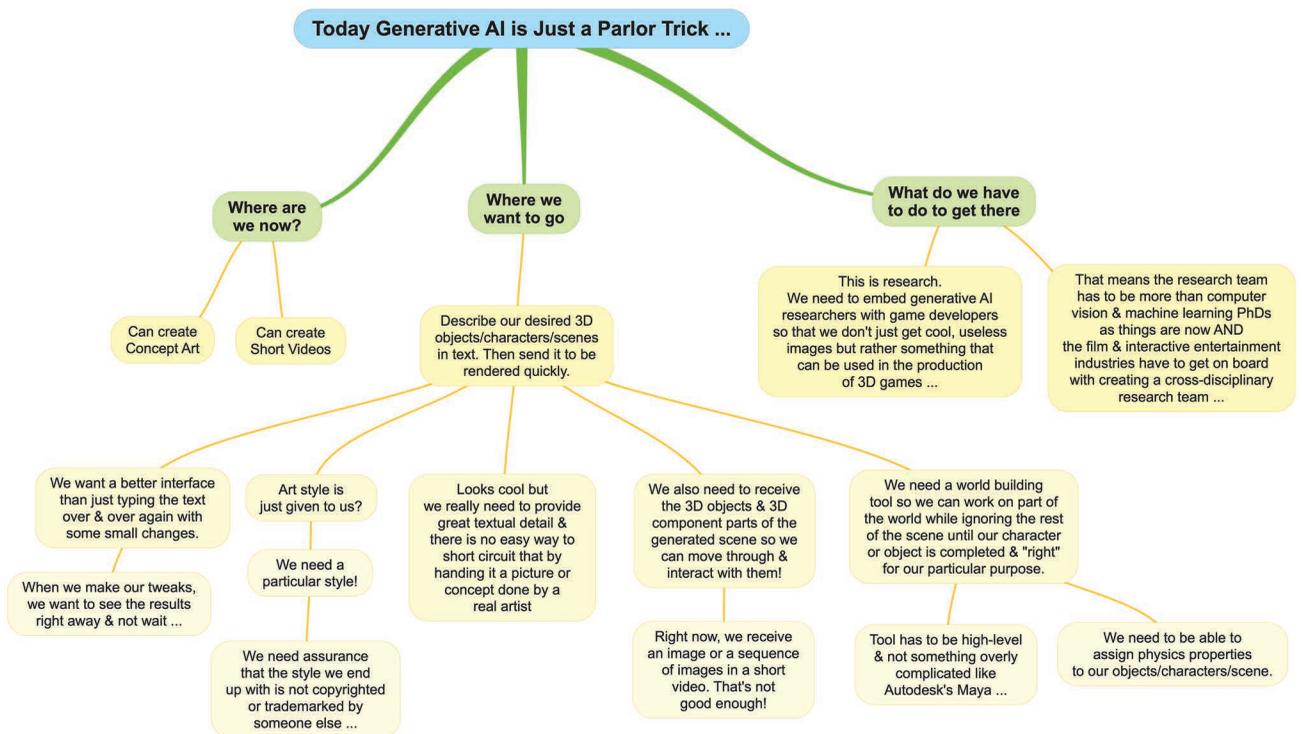


FIGURE 1. Today generative AI is just a parlor trick.



FIGURE 2. A dead cat bounce.



FIGURE 3. A dead cat bounce from one spacecraft to another under a moonlit sky.

to be something that I set in a workspace of some sort and don't have to type over and over again. All creative endeavors usually have some sort of desired art style, and most LLMs that I have tried will give you something if I don't specify it. That's kind of not OK; the workspace system should say, "Hey! You haven't specified the art style! Want to choose one from this list, or you want to draft it directly?" That would be way better than me having to enter the art style textually each time I want to see a new concept. Even better would be if it had a pulldown menu of some sort that allowed me to see what was rendered in newly selected

art styles instantly or maybe show me four or ten and then let me click on something to keep the one I liked without me having to type it out. And we need assurance that the style we end up choosing is not copyrighted or trademarked by someone else as our personal litigation budgets are not really all that large.

Desire for the ability to be able to do "concept insertions"

Another desire I have is that if I have a real concept artist and that real artist does a great piece of concept art, I want to hand a digital version of that concept art to the LLM and tell it, "Give me all my following 'images' in the

style of this image I have provided." That would be so cool and solve the consistency of the artistic look problem. And if I have this capability, I then want to have multiple "concept insertions" so I can decide on the look for all the 3D objects/characters/scenes ahead of time and then just have the LLM work on the small tweaks that I and the director want.

Desire for our 3D objects/characters/scenes to be interacted with

We also need to receive the 3D objects/characters/scenes generated in a way such that we can interact with them! Right now, the 17-s videos that Sora produces are static: I cannot touch the couple walking through the Japanese garden, so I cannot interact with them. That couple cannot touch anything in the LLM-created world. We cannot pull leaves off the LLM-created sakura trees. The couple can only walk to the end of the path specified in the original text. So today all we have really is a parlor trick. The 17-s video looks cool, but the technology as demonstrated today cannot

Dictionary

Definitions from [Oxford Languages](#) · [Learn more](#)

dead cat bounce

/ˈded ˈkɑt ˌbaʊns/

noun STOCK MARKET

a temporary recovery in share prices after a substantial fall, caused by speculators buying in order to cover their positions.

"is the recession really over, or is it a dead cat bounce?"

FIGURE 4. What is a "dead cat bounce"?

COMMENTS?

If you have comments about this article, or topics or references I should have cited or you want to rant back to me on why what I say is nonsense, I want to hear. Every time we finish one of these columns, and it goes to print, what I'm going to do is get it up online and maybe point to it at my Facebook (mikezyda) and my LinkedIn (mikezyda) pages so that I can receive comments from you. Maybe we'll react to some of those comments in future columns or online to enlighten you in real time! This is the "Games" column. You have a wonderful day.

provide anything to the interactive world of games.

Also, we would really like to know something about the potential scalability of LLM "imagines." I honestly believe that a great test would be for someone to generate the complete Pixar *Toy Story* text that an LLM could chew on and spit out as close to the original 81-min film as possible. Maybe seven days/months of compute time?

Desire for higher-level tools

For us to get from parlor trick to something useful to the interactive and filmed entertainment industries, we need some significant investment into tool building. We need a world building tool so we can work on part of the world while ignoring the rest of the scene until our character or object is completed and "right" for our particular purpose. That tool has to be high level and not something impossibly complicated like Autodesk's Maya. Tools like Maya are hard to

learn because there are too many settings and dials and knobs that must be turned. With all this AI flying around now, we ought to be fully integrating

AI researchers with game developers and film makers so that we don't just get cool, useless, 17-s videos but rather something that can be used in the

For us to get from parlor trick to something useful to the interactive and filmed entertainment industries, we need some significant investment into tool building.

AI into user interfaces such that the AI can know how expert the tool user is or is not. That AI then should guide/teach/get out of the way as appropriate.

In these high-level tools, we need to be able to assign physics properties to our objects/characters/scenes. We need to do that at a high level such that we are not trying to reproduce Physics 1 every time we create a new game world.

We need to be able to assign mental and emotional states to our characters, and those states ought to be initially selections from a list with the capability for the tool's user to move that anger slider to 11.

So, in conclusion, as I said above, generative AI is pretty much a parlor trick as it is currently deployed. It shows us great potential but leaves us thirsting for more, something useful for building our interactive games and our multihour feature films. So, how do we get there? Well, this is research, and from my experience, people that produce games and filmed entertainment rarely want to fund the hard work that goes into building such tools and development environments. If we so happen to find a deep-pocketed, long thinker interested in this, we should thank our lucky stars, take the money, and run toward this great generative AI future. We need to embed generative

production of 3D interactives and filmed entertainment. **■**

ACKNOWLEDGMENT

The author thanks all of the AI and generative AI researchers that are connected to him on LinkedIn and Facebook; their online presences, commentaries, and latest published papers are greatly appreciated!

REFERENCES

1. M. Zyda, "Large language models & generative AI, Oh My!," *Computer*, vol. 57, no. 3, pp. 127-132, Mar. 2024, doi: 10.1109/MC.2024.3350290.
2. S. Levy, "OpenAI's sora turns AI prompts into photorealistic videos," *Wired*, Feb. 15, 2024. Accessed: Feb. 21, 2024. [Online]. Available: <https://apple.news/A4rkDNyHGQgmuYotzVIOazQ>
3. Oxford English Dictionary. Accessed: Feb. 21, 2024. [Online]. Available: <https://www.oed.com/search/dictionary/?scope=Entries&q=dead+cat+bounce>

MICHAEL ZYDA was the founding director of the Computer Science Games Program and a professor emeritus of engineering practice in the Department of Computer Science, University of Southern California, Los Angeles, CA 90089 USA. Contact him at zyda@mikezyda.com.