MY WEIRD PROMPTS

Podcast Transcript

The Future of Local AI: Stable Diffusion vs. The New Guard

https://myweirdprompts.com/episode/local-ai-evolution-2026/

EPISODE SYNOPSIS

In this episode of My Weird Prompts, Corn and Herman Poppleberry dive into the rapidly shifting landscape of generative AI as we approach 2026. They explore whether the legendary Stable Diffusion can hold its ground against powerful newcomers like the Flux series and discuss the growing chasm between local hardware capabilities and cloud-based APIs. From architectural rendering to the "blurry cat" phase of local video generation, the duo debates the merits of community-driven ecosystems versus raw model power.

DANIEL'S PROMPT

Daniel

I'd love to get your take on the state of open-source generative AI models as we head into 2026. Many experts in fields like architecture still recommend Stable Diffusion as a starting point for local rendering, but there is now an overwhelming selection of new models on the market, such as the Flux series by Black Forest Labs. On platforms like Replicate and Fal AI, there's a constant stream of new models for various modalities, from text-to-image to image-to-video. Where does local AI stand today? Is Stable Diffusion still a major force to be reckoned with, or is there a general pivot toward different classes of open-source models?

TRANSCRIPT

Corn

Welcome to My Weird Prompts! I am Corn, and I am feeling particularly energized today, even if I am a sloth and usually prefer the slow lane. We are diving into a topic that has been moving at light speed lately. Our producer, Daniel Rosehill, sent us a prompt that really gets to the heart of how we create things in this digital age. We are looking at the state of open source generative artificial intelligence as we move toward the year twenty twenty-six.

Herman

And I am Herman Poppleberry. It is a pleasure to be here, though I must say, the speed of this industry is enough to make even a sturdy donkey like myself feel a bit winded. The prompt today is asking a critical question: where does local AI stand? For years, Stable Diffusion was the undisputed king of the hill if you wanted to run powerful models on your own hardware. But now we have the Flux series from Black Forest Labs, and platforms like Replicate and Fal AI are churning out new models for image and video every single day.

Corn

It is wild, Herman. I remember when being able to generate a blurry cat on your home computer was a miracle. Now, architects are using these tools for professional rendering. But the big question is whether the old guard, specifically Stable Diffusion, is actually holding its ground or if it is becoming a digital relic.

Herman

Well, I think we have to be careful with the word relic. Stable Diffusion isn't exactly a spinning jenny from the industrial revolution. But you are right that the landscape has shifted. The release of Flux point one was a massive turning point. It brought a level of prompt adherence and anatomical detail that frankly made the older Stable Diffusion models look a bit amateurish.

Corn

See, I don't know if I agree that they look amateurish. If you go on any of the big community hubs, people are still doing incredible things with Stable Diffusion XL. There is this massive ecosystem of LoRAs and custom checkpoints that you just can't find for the newer models yet. Isn't there something to be said for the depth of a community versus just raw power?

Herman

There is, but raw power wins in professional workflows eventually. If an architect needs a render that actually follows their specific instructions about lighting and materials, they can't spend four hours fighting with a model that thinks a human hand should have seven fingers. Flux and its successors have moved the needle on reliability.

Corn

Okay, but let's take a step back for a second. For someone who isn't a power user, what are we actually talking about when we say open source in twenty twenty-six? Because some of these models are huge. You need a massive graphics card to run them locally. Is it really local AI if you need a five thousand dollar setup to host it?

Herman

That is a fair point. We are seeing a divergence. On one hand, you have the massive, high fidelity models like Flux Pro or the latest iterations from Black Forest Labs that are often accessed via API on platforms like Fal AI. On the other hand, we have the distilled versions. These are smaller, faster models that can run on a standard consumer laptop. The democratization is still happening, it just looks different than it did two years ago.

Corn

I still think the local aspect is being undervalued by the big labs. People want privacy. If I am a designer working on a secret project, I don't want my prompts going to a server in the cloud, even if that server is incredibly fast. That is why Stable Diffusion stayed relevant for so long. It was the ultimate sandbox.

Herman

I agree on the privacy aspect, but let's be realistic. The complexity of these models is scaling faster than consumer hardware. We are reaching a point where the gap between what you can do on your own machine and what you can do with a cloud API is becoming a chasm, not just a crack.

Corn

But isn't that exactly what the open source community is good at? Shrinking things? I mean, look at what happened with Large Language Models. We went from needing a server farm to running decent models on a phone in eighteen months.

Herman

True, but image and video generation are computationally more expensive by orders of magnitude. Especially when we move into image-to-video. Have you tried running a high-end video model locally lately? Your computer would double as a space heater for the entire neighborhood.

Corn

Hey, in the winter, that is a feature, not a bug! But seriously, I want to talk about these new players. Flux is the big name right now, but we are also seeing things like the AuraFlow models and various iterations of Stable Diffusion three. It feels like the market is fragmented. Is fragmentation good for us, or is it just making everything more confusing?

Herman

It is a bit of both. Fragmentation creates competition, which drives innovation. But for the end user, it is a nightmare. You have to learn a new prompting style for every model. You have to manage different environments. It is not like the early days where everyone was just using one version of Automatic eleven-eleven.

Corn

I actually think the fragmentation is a sign of maturity. We are moving away from a one size fits all approach. You might use one model for architectural visualization because it understands straight lines and perspective, and another model for character art because it handles skin textures better. It is like having a toolbox instead of just a single hammer.

Herman

Narrowing it down to a specific toolbox is fine, but the tools are changing every week. How is a professional supposed to build a stable workflow on shifting sand?

Corn

That is a great question, and I want to dig into that more, but first, we need to take a quick break for our sponsors. Larry: Are you tired of your dreams being stuck in your head? Do you wish you could project your subconscious thoughts directly onto a physical medium without all that pesky talent or effort? Introducing the Dream-O-Graph five thousand! This revolutionary headband uses patented neural-static technology to capture your nighttime visions and print them directly onto any flat surface. Want to see that giant purple squirrel you dreamt about? Just strap on the Dream-O-Graph, take a nap, and wake up to a beautiful, slightly damp charcoal sketch on your living room wall. Side effects may include vivid hallucinations of Victorian-era street performers, a metallic taste in your mouth, and a sudden, inexplicable knowledge of how to speak ancient Babylonian. The Dream-O-Graph five thousand. It is not just a printer, it is a portal to your own confusion. Larry: BUY NOW!

Herman

...Alright, thanks Larry. I am not sure I want my dreams printed in damp charcoal, but to each their own. Back to the topic at hand. Corn, you were talking about the professional workflow.

Corn

Right! So, if I am an architect in twenty twenty-six, and I have been using Stable Diffusion for years, why would I switch? If I have my custom LoRAs for specific building materials and I know exactly how to use ControlNet to keep my walls straight, is Flux really going to offer me enough to justify relearning everything?

Herman

The short answer is yes, because of the base intelligence of the model. Stable Diffusion, even the XL version, often requires a lot of hand-holding. You need ControlNet just to make it do the basics. The newer generation of models, like Flux, have a much deeper understanding of spatial relationships and physics right out of the box. You spend less time correcting the AI and more time iterating on your design.

Corn

I don't know, Herman. I think you are underestimating the power of the legacy. There are thousands of free models on sites like Civitai that are built on top of Stable Diffusion. You can't just recreate that overnight. It is like saying everyone should switch from Windows to a brand new operating system just because the new one is ten percent faster. People stay for the software and the community.

Herman

But it isn't ten percent faster, Corn. It is a fundamental shift in quality. When you look at the text rendering in these new models, it is night and day. If an architect wants to include signage in a render, Stable Diffusion gives you alphabet soup. Flux actually writes the words. That kind of thing matters when you are presenting to a client.

Corn

Okay, the text thing is a huge win, I will give you that. But what about the hardware? If we are talking about local AI, we have to talk about the V-RAM. These new models are heavy. Are we moving toward a future where local AI is only for the elite with thirty-nine-ninety or fifty-nine-ninety cards?

Herman

That is the risk. But we are also seeing clever engineering. Quantization is the magic word here. We are taking these massive models and squeezing them down so they fit into sixteen or even eight gigabytes of V-RAM. It is a constant arms race between the model size and the compression techniques.

Corn

It feels like a bit of a treadmill. You buy a new card to run the new model, then a bigger model comes out that needs a bigger card. At what point do we just admit that the cloud is easier?

Herman

Never! Well, I shouldn't say never, but the local movement is fueled by a very specific philosophy. It is about ownership. If you have the weights on your hard drive, nobody can take them away. They can't change the terms of service. They can't censor your creativity based on a corporate whim.

Corn

I love that philosophy, but I worry it is becoming a niche. Most people just want the pretty picture. They don't care about the weights. They go to Fal AI or Replicate because it is one click and it works.

Herman

And that is exactly why the open source labs are trying to make their models more accessible. They know they are competing with the convenience of Midjourney and DALL-E. That is why we are seeing so many different versions of these models—Dev, Schnell, Pro. They are trying to cover every use case from the casual enthusiast to the high-end pro.

Corn

Let's talk about video for a second, because that seems to be the new frontier. We are seeing things like CogVideo and the new Luma models. Is local video generation even a reality for most people yet?

Herman

It is on the horizon, but it is very early days. Generating a single high-quality image is one thing. Generating twenty-four frames per second that are temporally consistent? That is a whole different beast. Right now, most of that is happening in the cloud. But the open source community is nipping at the heels of the big players.

Corn

I saw a demo the other day of a local video model that could do five seconds of a character walking. It was a bit shaky, but it was impressive. It reminded me of the early days of image generation.

Herman

Exactly. We are in the "blurry cat" phase of video. By twenty twenty-seven, we will probably be generating full movie trailers on our desktops. But the question of which architecture wins is still wide open. Will it be a transformer-based model? A diffusion-based model? A hybrid?

Corn

I bet on the hybrids. Usually, the middle ground is where the stability is. But speaking of stability, I think we have someone who wants to weigh in on our discussion. We have Jim on the line from Ohio. Hey Jim, what is on your mind today? Jim: Yeah, this is Jim from Ohio. I have been listening to you two yapping about these fancy computer pictures for twenty minutes now and I have had just about enough. You are making it sound like this is some kind of revolution. It is just a more complicated way to make a fake photo. My neighbor Gary bought one of those high-end computers you were talking about, and all he does is make pictures of his dog wearing a tuxedo. What a waste of electricity! The power grid in my town is already shaky enough because of the heat wave.

Corn

Well, Jim, I think the tuxedo dog is just the beginning! It is about the creative potential for everyone. Jim: Creative potential? Give me a break. In my day, if you wanted a picture of a dog in a tuxedo, you put a tuxedo on a dog and you took a photo! It didn't take ten gigabytes of whatever you called it. And don't get me started on the video stuff. I saw a video of a politician that looked real but wasn't, and it nearly gave me a heart attack. We can't trust anything anymore. It is like when the local grocery store started selling those "organic" apples that taste like cardboard. Total scam.

Herman

Jim, I hear your concerns about trust and deepfakes. Those are very real issues. But we are also talking about tools for architects and designers to build better buildings and more efficient products. Don't you think there is value in that? Jim: Better buildings? They don't build anything to last anymore anyway! My porch is falling apart and I can't find a contractor who knows a hammer from a hacksaw. You think an AI is going to fix my porch? No, it is just going to make a pretty picture of a porch while the real one rots. Plus, my cat Whiskers is terrified of the fan on Gary's computer. Sounds like a jet engine taking off every time he tries to "render" something. It is a nuisance.

Corn

We appreciate the perspective, Jim. It is a good reminder that technology has real-world impacts, from the power grid to the noise in the neighborhood. Jim: You bet it does. And someone needs to tell that Larry fellow that his dream machine sounds like a lawsuit waiting to happen. I am hanging up now, I have to go check on my tomatoes. They are looking a bit peaked.

Corn

Thanks for calling in, Jim! Always good to hear from you.

Herman

He isn't entirely wrong about the power consumption, you know. These models are incredibly hungry. If we are moving toward a future where everyone is running these locally, we are going to need a lot more green energy.

Corn

Or just more efficient models! That is what I keep saying. The trend shouldn't just be "bigger is better." It should be "smarter is better."

Herman

I agree, but currently, the path to "smarter" has been through "bigger." We haven't quite figured out how to get that high-level reasoning and visual understanding without billions of parameters.

Corn

Let's get back to the prompt's specific question about the pivot. Is there a general pivot toward different classes of models? Because Stable Diffusion was a U-Net architecture, right? And now we are seeing more Diffusion Transformers, or DiTs. Is that the "pivot" Daniel was asking about?

Herman

Yes, that is exactly the technical pivot. The U-Net architecture was great for a long time, but it has scaling limits. Transformers, which were originally designed for text, have turned out to be incredibly good at processing visual data too. Flux is a DiT. Stable Diffusion three is a DiT. This architecture allows the model to understand the relationship between different parts of an image much better.

Corn

So it's not just that the models are newer, it's that the underlying engine is totally different?

Herman

Exactly. It is like moving from a piston engine to a jet engine. They both get you where you are going, but the jet engine can go much faster and higher if you give it enough fuel.

Corn

That makes a lot of sense. But if I am a casual user who just wants to make some cool art for my Dungeons and Dragons campaign, do I really need to care about DiTs versus U-Nets?

Herman

You don't need to care about the math, but you will notice the results. You will notice that you don't have to type "five fingers, highly detailed, masterpiece" in your prompt anymore. You can just say "a hand holding a sword" and it will actually work. That is the practical takeaway.

Corn

I actually want to push back on that a little bit. I think there is a charm to the "struggle" with the older models. There is a specific aesthetic that came out of the limitations of Stable Diffusion one point five. Sometimes these new models are almost too perfect. They look like stock photos.

Herman

Oh, here we go. The "digital vinyl" argument. You think the flaws make it art?

Corn

In a way, yeah! If everything is perfectly rendered and anatomically correct, where is the soul? Where is the weirdness? That is why I think people will stick with the older models for certain creative projects. It is about the "vibe," Herman.

Herman

I can't argue with a "vibe," Corn, but I can argue with efficiency. Most people using these tools for work don't want a "vibe," they want a finished product that doesn't need ten hours of post-processing in Photoshop.

Corn

Fair enough. So, let's talk practical takeaways. If someone is listening to this and they want to get into local AI today, where should they start?

Herman

If you have a decent computer with at least twelve gigabytes of V-RAM, I would say look at a quantized version of Flux point one Dev. It is currently the gold standard for open weights. If you have less than that, Stable Diffusion XL is still a very solid choice and has the best community support.

Corn

And what about the software? Is it still all command line stuff and complicated installs?

Herman

Not at all. Tools like Forge or ComfyUI have made it much more accessible. ComfyUI in particular is great because it uses a node-based system. It is a bit of a learning curve, but it gives you total control over the process.

Corn

I tried ComfyUI once. It looked like a plate of spaghetti with all those wires and boxes. I preferred the simpler interfaces. But I guess if you want to be a "pro," you have to learn the nodes.

Herman

It is worth it, I promise. It is like learning to use a professional camera instead of just a point-and-shoot.

Corn

I also think it's important to mention that you don't have to choose just one. You can have multiple models installed. You can use Flux for the base image and then use a Stable Diffusion LoRA to style it. The future is definitely hybrid.

Herman

That is a great point. The interoperability is getting better. We are seeing bridges being built between these different ecosystems.

Corn

So, to wrap up the core of the discussion, it sounds like we are saying that while Stable Diffusion isn't dead, the "pivot" to Transformer-based models like Flux is very real and very necessary for the next leap in quality.

Herman

Absolutely. We are heading into an era where the boundary between "open source" and "state of the art" is almost non-existent. The models you can run at home are becoming just as capable as the ones behind the billion-dollar paywalls.

Corn

That is an exciting thought. Imagine what people will be creating by twenty twenty-six. We might have entire indie films made by one person in their bedroom.

Herman

Or just a lot more dogs in tuxedos, if Jim's neighbor has anything to say about it.

Corn

Hey, don't knock the tuxedo dogs! They are a vital part of the internet ecosystem.

Herman

I suppose they are.

Corn

Well, this has been a fascinating deep dive. We covered everything from the technical shift to DiT architectures to the philosophical importance of hardware ownership.

Herman

And we managed to mostly agree, which is a rare treat.

Corn

Don't get used to it, Herman. I am sure I will find something to disagree with you about in the next episode.

Herman

I look forward to it.

Corn

Before we go, a quick reminder that you can find My Weird Prompts on Spotify and all your favorite podcast platforms. Big thanks to Daniel Rosehill for this prompt—it really gave us a lot to chew on.

Herman

Indeed. It is a brave new world of pixels and parameters out there. Stay curious, everyone.

Corn

And keep those prompts coming! We love seeing what kind of weirdness you want us to explore next.

Herman

Just maybe nothing that involves Babylonian charcoal sketches.

Corn

Speak for yourself, Herman. I have already ordered three Dream-O-Graphs.

Herman

Of course you have.

Corn

Goodbye everyone!

Herman

Until next time.