MY WEIRD PROMPTS

Podcast Transcript

From Moths to Models: The Rise of Computer Use Agents

https://myweirdprompts.com/episode/computer-use-agents-future/

EPISODE SYNOPSIS

In this episode of My Weird Prompts, Herman and Corn explore the evolution of human-computer interaction, starting with Grace Hopper's vision in the 1950s and leading into the cutting-edge AI of late 2025. They break down the difference between simple chatbots and "Computer Use Agents" that can actually see and manipulate a computer interface. The discussion covers the Model Context Protocol (MCP), the battle between vision-based and programmatic control, and the shift toward Large Action Models (LAMs). Whether you want to automate audio editing or just stop clicking buttons, this episode reveals how close we are to a truly agentic future.

DANIEL'S PROMPT

Daniel

Grace Hopper's 1950s vision of interacting with computers using natural language is finally becoming a reality with agentic AI and the Model Context Protocol (MCP). I'm particularly interested in the practical application of voice control for desktop tasks—like being able to tell my computer to "stop Audacity, save this file, and run the production pipeline" without constant back-and-forth or clarifying questions. What is the state of computer use agents heading into 2026? What is the current nomenclature, and which approach is more promising for conversational computer control: the programmatic method that uses CLI commands, or the vision-based approach that interacts with GUI controls?

TRANSCRIPT

Corn

Hey everyone, welcome back to My Weird Prompts! I am Corn, and I am feeling especially relaxed today, even for a sloth. It is a beautiful day here in Jerusalem, and I am joined, as always, by my brother.

Herman

Herman Poppleberry, at your service! And I am anything but relaxed, Corn. I have been diving deep into the prompt our housemate Daniel sent over this morning. It is all about the bridge between the nineteen fifties and where we are right now, at the tail end of twenty twenty-five.

Corn

Yeah, Daniel was asking about Grace Hopper and this vision of talking to our computers like they are actual assistants. I remember you mentioning her before, Herman. She is the one who found the first actual computer bug, right? Like, a literal moth?

Herman

That is the one! She was a pioneer. But Daniel's prompt goes way beyond just the history. He is looking at how her dream of interacting with computers through natural language is finally, actually happening. He mentioned things like agentic artificial intelligence and the Model Context Protocol, and he wants to know how he can eventually just tell his computer to stop Audacity, save a file, and run a production pipeline without having to click a single button.

Corn

That sounds like the dream, honestly. I would love to just tell my computer to do the dishes, but I guess we are starting with audio editing. So, Herman, where do we even start with this? It feels like we have been hearing about voice control for years, but it always kind of sucked. Why is it different now as we head into twenty twenty-six?

Herman

You are right, Corn. For a long time, voice control was basically just a fancy way to trigger a keyboard shortcut. You would say "Open Mail," and the computer would just execute the command to launch an application. It did not really understand what was happening inside that application. But what Daniel is asking about is a shift toward what we call Computer Use Agents. This is a specific branch of agentic artificial intelligence where the model does not just talk to you; it understands the interface of the computer itself.

Corn

Okay, hold on. Break that down for me. What is the difference between a regular chatbot and an agent that can actually use a computer?

Herman

Think of it this way. A standard chatbot is like a very smart person sitting in a dark room. You can ask them questions, and they can give you amazing answers, but they cannot see the world or touch anything. An agentic computer use model is like that same smart person, but now they are sitting at your desk, looking at your monitor, and holding your mouse. They can see that Audacity is open, they can see the export button, and they can move the cursor to click it.

Corn

That sounds a little bit creepy, but also incredibly useful. Daniel mentioned something called the Model Context Protocol, or MCP. I have heard you nerding out about that in the kitchen lately. What does that have to do with this?

Herman

Oh, MCP is a huge piece of the puzzle! It was developed to create a universal standard for how artificial intelligence models connect to data and tools. Before MCP, if you wanted an AI to talk to a specific piece of software like Audacity or a database, you had to write custom code for that specific connection. It was like having twenty different electronics in your house and every single one needing a different shaped power outlet.

Corn

And let me guess, MCP is like the universal power strip?

Herman

Exactly! It allows developers to create servers that expose certain tools or data in a way that any artificial intelligence model can understand. So, if someone builds an MCP server for Audacity, Daniel could use any model, whether it is from Anthropic, OpenAI, or a local model running on his machine, and that model would instantly know how to "talk" to Audacity. It provides a structured way for the agent to say, "Hey, what files are open?" or "Run the noise reduction filter."

Corn

That makes sense. But Daniel's big point was about the back and forth. He said he does not want to have to answer a bunch of clarifying questions. He just wants to give one command and have it done. Are we actually there yet?

Herman

We are getting very close, but there is still a tug of war between two different philosophies of how to do this. Daniel asked about the nomenclature, and this is where it gets interesting. On one side, you have the programmatic approach, often using the Command Line Interface, or CLI. On the other side, you have the vision-based approach, which interacts with the Graphical User Interface, or GUI.

Corn

Okay, let's do the Herman Poppleberry special. Give me an analogy for those two.

Herman

I love it. Okay, the programmatic or CLI approach is like giving a chef a very precise recipe with exact measurements and temperatures. You tell the computer exactly which commands to run in the background. It is incredibly fast and reliable, but the chef needs to have that specific recipe already in their book. If the software does not have a command line version, you are stuck.

Corn

And the vision-based one?

Herman

That is like the chef just standing in your kitchen and looking at the stove. They see the knobs, they see the ingredients, and they just figure it out by looking. The vision-based agent literally takes screenshots of your desktop every second, analyzes where the buttons are, and moves the mouse just like a human would.

Corn

That seems way harder for the computer. Why would we do it that way if we can just use the recipes?

Herman

Because most of the software humans use was built for eyes, not for recipes. Think about Audacity, which Daniel mentioned. It is a visual tool. While it has some keyboard shortcuts, a lot of the deep work happens by clicking through menus. If an AI can see the screen, it can use any app you have, even if that app was built twenty years ago and has no modern connections.

Corn

I can see why Daniel is excited. But he mentioned that it is still kind of buggy. He said he gets a burst of excitement when it works, but it takes a lot of effort to set up. Why is it so hard to get right?

Herman

It goes back to what Grace Hopper was dreaming about in the nineteen fifties. She wanted computers to understand human intent, not just human syntax. Right now, if Daniel says "Save this file," the AI has to figure out which file he means, where he wants to save it, and what format it should be in. If it makes a mistake, it might overwrite his most important recording. So, the agents are often programmed to be very cautious, which leads to those annoying clarifying questions Daniel wants to avoid.

Corn

So, we need the AI to have more context. Like, it needs to know that when Daniel says "save this," he always means save it to the project folder with today's date.

Herman

Precisely. And that is where the twenty twenty-five developments have been so key. We are seeing models with much larger context windows, meaning they can remember what you did yesterday or what you mentioned in a chat three hours ago. But before we get deeper into the vision versus programmatic debate, I think we should take a quick break for our sponsors.

Corn

Good idea. Let's hear from Larry. Larry: Are you tired of your thoughts being private? Do you wish you could broadcast your internal monologue to everyone within a fifty foot radius? Introducing the Think-O-Graph Nine Thousand! This revolutionary headband uses unshielded copper coils to pick up your brain waves and convert them into high decibel audio. Perfect for family gatherings, job interviews, or just walking down the street. Never worry about "saying the wrong thing" again, because you will be saying everything! The Think-O-Graph Nine Thousand comes in three colors: static gray, feedback silver, and migraine maroon. Warning: may cause temporary loss of personality or permanent hair loss. Think-O-Graph Nine Thousand. BUY NOW!

Corn

Thanks, Larry. I think I will stick to my quiet sloth thoughts for now. Anyway, Herman, back to the computer agents. Daniel was asking which approach is more promising: the CLI commands or the vision-based GUI stuff. What is the verdict as we look toward twenty twenty-six?

Herman

It is a bit of a hybrid future, Corn. But if I had to put my money on one, I think the vision-based approach is where the real "magic" happens for the average person. In late twenty twenty-four and throughout twenty twenty-five, we saw companies like Anthropic release things like "Computer Use" for their Claude models. This allowed the AI to actually move the cursor and type. It was a huge leap.

Corn

But is it fast enough? I feel like if I tell my computer to do something, I don't want to watch it move the mouse slowly across the screen like a ghost is haunting my desktop.

Herman

That is the main drawback right now. Vision-based agents are computationally expensive. They have to process a lot of images very quickly. Programmatic agents, using CLI or direct API calls, are nearly instantaneous. If Daniel wants to "run the production pipeline," a programmatic agent is far superior because it can just trigger the script directly.

Corn

So maybe the answer is that the AI should use the vision to find the buttons when it has to, but use the CLI for the heavy lifting?

Herman

That is exactly what the most sophisticated systems are doing now. They are starting to use a "planner" model. The planner looks at the task and asks, "Can I do this through a direct command?" If yes, it does it. If no, it switches to vision mode. This nomenclature is often referred to as "Large Action Models" or "Agentic Workflows." We are moving away from just "Large Language Models" because the "Action" part is what matters now.

Corn

I like that. Large Action Models. It sounds like a summer blockbuster movie. So, for Daniel's specific example, he wants to say, "Stop Audacity, save this file, and run the production pipeline." How would that actually look in practice?

Herman

In a perfect world, or at least the world we are entering in twenty twenty-six, the agent would first send a signal to the operating system to find the process ID for Audacity. It would send a "stop" command. Then, it would look at the active window to see if there are unsaved changes. This is where the vision comes in. It sees the little asterisk next to the file name that indicates it is unsaved. It clicks "File," then "Save As." Because it has access to Daniel's file system through something like the Model Context Protocol, it knows exactly where the "Weird Prompts" repository is. It types the path, hits enter, and then opens a terminal to run the final production script.

Corn

And all of that happens from one voice command?

Herman

That is the goal. The reason Daniel is seeing bugs right now is because the handoff between these steps is still brittle. If the "Save" window takes two seconds to pop up but the AI only waits one second, the whole thing crashes. Or if a random notification pops up and covers the button the AI was looking for, it gets confused. Humans are great at ignoring distractions, but agents are still learning how to do that.

Corn

It is kind of funny to think that Grace Hopper was dealing with literal moths in the machinery, and now we are dealing with digital moths, like pop-up ads or slow loading windows, that confuse our AI agents.

Herman

It really is a full circle moment! Hopper's work on COBOL, which stands for Common Business Oriented Language, was all about making computer code look more like English so that more people could use it. She wanted to bridge the gap between human thought and machine execution. What we are doing now with natural language agents is just the final, ultimate version of that. We are finally removing the need for the "code" layer entirely for the end user.

Corn

So, if I am Daniel, and I want to get this working right now, what should I be looking for? Are there specific tools that are making this easier?

Herman

There are a few. For the programmatic side, tools that implement the Model Context Protocol are essential. There are already MCP servers for things like Google Drive, Slack, and even local file systems. For the vision side, he should look into things like the Open Interpreter or the desktop versions of the major AI assistants that are starting to roll out "screen awareness." But honestly, the biggest breakthrough for twenty twenty-six is going to be local processing.

Corn

Local processing? Like, on his own computer instead of in the cloud?

Herman

Yes. Right now, every time the AI takes a screenshot of Daniel's desktop, it has to send that image to a server somewhere else to be analyzed. That is slow, it is expensive, and it is a bit of a privacy nightmare. But the new chips coming out in twenty twenty-five and twenty twenty-six are designed specifically to run these vision models locally. When the "brain" is inside the computer itself, the lag disappears. That is when the "back and forth" Daniel hates will finally start to vanish.

Corn

That makes a lot of sense. If it is local, it can see everything instantly without waiting for the internet. I imagine that would make it feel a lot more like a real assistant sitting next to you.

Herman

Exactly. And let's talk about the voice part of Daniel's prompt. He is very interested in voice technology. We have seen a massive leap in what we call "Speech to Intent." Old voice assistants would turn your voice into text, then try to understand the text. Newer models are "omni-modal," meaning they listen to the audio directly. They can hear the tone of your voice, your pauses, and even your frustration.

Corn

Oh, so if Daniel sounds stressed, the AI might realize it should not ask him five clarifying questions and just do its best?

Herman

Actually, yes! Or it might realize that when he says "Stop Audacity" with a certain urgency, he means "kill the process right now" versus a polite "please close the application when you have a moment." This level of semantic understanding is what takes us from a "tool" to an "agent."

Corn

This is all fascinating, Herman. I feel like I am actually learning something, which is dangerous for a sloth. It might make me want to move faster. But let's get practical for a second. If this technology is finally here, what are the implications for how we work? Does this mean we don't need to learn how to use software anymore?

Herman

That is a deep question, Corn. I think it means the "learning curve" for software changes. Instead of learning where every button is in a complex program like Photoshop or Audacity, you just need to learn how to describe what you want to achieve. The "interface" becomes your language. But there is a risk. If we stop learning how the tools work, we might not know when the AI is doing a mediocre job.

Corn

Right, like if the AI saves the file but uses a really low quality bit rate, Daniel might not notice until the podcast is already uploaded.

Herman

Exactly. So the role of the human moves from "operator" to "editor" or "supervisor." You are still the creative director, you just have a very fast, very capable intern doing the clicking for you. Grace Hopper actually had a famous quote about this. She said, "The most dangerous phrase in the language is, 'We've always done it this way.'" She was always pushing for the next simplification.

Corn

I like that. I think "We've always done it this way" is also the reason I still take four hour naps, but maybe I can use an agent to schedule those more efficiently. So, looking ahead to twenty twenty-six, do you think we will see a "Universal Computer Agent"? Like one app that controls everything?

Herman

I think it will be integrated into the operating system itself. We are already seeing Apple and Microsoft and Google racing to make the OS "agentic." Instead of opening Audacity, Daniel might just speak to his desktop. The desktop "is" the agent. It has the vision to see all his apps and the programmatic connections to control them.

Corn

So, no more icons? Just a blank screen that listens?

Herman

Maybe not entirely blank, but certainly less cluttered. The computer becomes a true extension of your intent. But we have to be careful about the "nomenclature" Daniel asked about. We are going to hear a lot of marketing buzzwords. "Autonomous Agents," "Actionable AI," "Cognitive Architectures." At the end of the day, it all comes back to what Daniel said: can it save the file and run the pipeline without being a nuisance?

Corn

That is the ultimate test. The "Daniel Test." If it can handle a grumpy podcaster in Jerusalem, it can handle anything.

Herman

Haha, exactly! And honestly, the fact that he is already getting it to work, even with some bugs, is a huge sign. A year ago, this was pure science fiction. The Model Context Protocol only really started gaining steam recently, and it is already changing how developers think about software. They aren't just building for humans anymore; they are building for agents.

Corn

That is a big shift. It's like when everyone started building websites for mobile phones instead of just desktop computers. Now they are building apps for AI to use.

Herman

Spot on, Corn. That is the "API-first" or "Agent-first" development model. If an app has a good MCP server, it will be much more popular in twenty twenty-six because people can actually use it with their voice or through their agents. If an app is a "walled garden" that the AI can't see or talk to, it's going to feel very old-fashioned very quickly.

Corn

Well, I hope Audacity is listening and getting their MCP server ready. I don't want Daniel to have to work any harder than he already does. He's got enough on his plate with us as housemates.

Herman

Very true. To wrap up the technical side for Daniel, I would say the most promising approach is definitely the hybrid. Use programmatic commands whenever possible for reliability and speed, but keep the vision-based system as a "fail-safe" or for navigating complex menus that don't have commands yet. And keep an eye on those local models. As soon as you can run a "Vision-Language-Action" model on your own hardware, that is when the dream really becomes a reality.

Corn

This has been a lot to process, but I feel like I have a much better handle on why Daniel is so excited about this Grace Hopper stuff. It's not just about the past; it's about finally catching up to the vision someone had seventy years ago.

Herman

It really is. It is a testament to human persistence. Or donkey persistence, in my case. We keep chipping away at these problems until the technology finally catches up to the imagination.

Corn

Well, my imagination is currently picturing a snack. But before we go, I want to remind everyone that you can find "My Weird Prompts" on Spotify and at our website, myweirdprompts.com. We have an RSS feed there if you want to subscribe, and a contact form if you want to send us a prompt like Daniel did.

Herman

Yes, please do! We love digging into these topics. And Daniel, thanks for the prompt. It was a great excuse to talk about one of my heroes, Admiral Grace Hopper. I think she would be pretty impressed with where we are heading in twenty twenty-six.

Corn

Definitely. Thanks for listening, everyone. We will be back next time with another weird prompt. Until then, stay curious and maybe try talking to your computer. Just don't be surprised if it doesn't answer back yet.

Herman

Or if it does, and it asks you where you want to save your files for the tenth time.

Corn

Exactly. Thanks for listening to My Weird Prompts! Goodbye from Jerusalem!

Herman

See ya!