

MY WEIRD PROMPTS

Podcast Transcript

EPISODE #260

Digital Archeology: The Primitive Power of GPT-1

Published January 20, 2026 • Runtime: 19:18

<https://myweirdprompts.com/episode/gpt-1-origins-evolution/>

EPISODE SYNOPSIS

In this episode, Herman Poppleberry and Corn take a fascinating trip back to 2018 to perform some "digital archeology" on the model that started a revolution: GPT-1. While modern users in 2026 might find its 117-million-parameter capacity and tendency to output gibberish laughable, the hosts explain why this "primitive" tool was actually the Wright brothers' flyer of the artificial intelligence era. They dive deep into the technical limitations of the time, including the 512-token context window and the use of absolute positional embeddings that caused the model to frequently lose its train of thought. Beyond the specs, Herman and Corn discuss the shift from supervised learning to unsupervised pre-training and how a dataset of 11,000 unpublished romance novels shaped the early worldview of generative AI. By comparing the raw engine of GPT-1 to the "layered cakes" of 2026, this episode provides a crucial perspective on how far the industry has come and why the ghost of this original architecture still lives within the trillion-parameter giants of today.

DANIEL'S PROMPT

Daniel

I recently experimented with GPT-1 and found it to be surprisingly limited compared to modern models, especially with its small context window and lack of conversational tuning. If GPT-1 was a landmark in the development of AI, why does it seem so much less capable than today's versions? Additionally, what were the early applications for these models, and were they already being prototyped as chatbots?

TRANSCRIPT

Corn

Alright, we have a really fun one today. Our housemate Daniel has been doing some digital archeology in the back rooms of the internet.

Herman

Herman Poppleberry here. And yeah, Daniel was actually playing around with GPT-one. Which, for anyone who joined the AI party in the last couple of years, probably feels like trying to use a stone tool to build a skyscraper.

Corn

It is wild. He sent us this audio where he was basically poking at it with a stick and getting frustrated because it kept calling Paris a village and then devolved into absolute gibberish after three sentences. It's such a perfect starting point for us because GPT-one is often cited as this massive landmark, but if you actually use it today, it feels... well, broken.

Herman

It's not broken, Corn! It's just... primitive. Think of it like the Wright brothers' first flyer. It stayed in the air for twelve seconds and traveled one hundred and twenty feet. If you tried to use it to fly from Jerusalem to New York today, you'd call it a failure, but at the time, the fact that it flew at all changed the world.

Corn

That's a fair analogy. But Daniel's question is really the core of it. Why does it seem so much less capable? I mean, it's still a transformer, right? The "T" in GPT stands for Transformer. That's the same architecture we're using for the massive models today. So what happened between the June 11, 2018 GPT-1 paper and now?

Herman

A lot of things, but the biggest one is scale. OpenAI published the GPT-1 paper on June 11, 2018, with the model having 117 million parameters. Now, to a regular person, a hundred million sounds like a lot. But compared to what we're working with in twenty twenty-six, it's a grain of sand. For context, GPT-four's parameter count is undisclosed by OpenAI but estimated by third parties at around 1.8 trillion in a mixture-of-experts setup, and the models we're seeing this year are pushing even further into agentic autonomy. That is roughly a fifteen-thousand-fold increase in raw capacity since the original.

Corn

Fifteen thousand times. That's like comparing the brain of a fruit fly to a human. But it's not just the size, right? Because Daniel mentioned the context window. He said it started outputting gibberish after just a few turns.

Herman

Exactly. The context window on GPT-one was only five hundred and twelve tokens. In modern terms, that's maybe a page of text. But here's the technical kicker: it used what we call absolute positional embeddings. That means the model literally had a hard-coded limit. It couldn't see token five hundred and thirteen if it wanted to. Once it hit that limit, it didn't have the sophisticated "forgetting" or attention mechanisms we have now to stay on track. It would just start attending to its own previous errors, creating a feedback loop of nonsense. If it made one typo or one weird word choice, it would get "distracted" by its own mistake and just spiral.

Corn

I love that image of the model getting distracted by its own shadow. But there's another part of Daniel's experience that I think is really important. He mentioned that he asked "How are you?" and it gave him a very cold, robotic "I am an AI model, I do not have feelings." He noted it wasn't tuned for conversation. And I think this is where a lot of people get confused. We think of GPT as a chatbot, but GPT-one was never intended to be a chatbot, was it?

Herman

Not even close! This is the biggest misconception about early large language models. Back in 2018, the goal wasn't "let's make a digital friend." The goal was "let's see if we can predict the next word in a sentence well enough to understand language." GPT-one was a proof of concept for something called unsupervised pre-training. Before this, if you wanted an AI to do sentiment analysis, you had to train it specifically on a dataset of labeled "happy" and "sad" sentences.

Corn

Right, the "supervised" approach. You had to hold the model's hand for every specific task.

Herman

Exactly. OpenAI's big gamble with GPT-one was saying, "Hey, what if we just feed it a ton of books and let it figure out how language works on its own by trying to guess the next word?" They used the BookCorpus dataset, which contained over 11,000 books scraped from Smashwords—mostly romance, fantasy, and adventure novels.

Corn

Wait, unpublished books? So GPT-one's entire worldview was shaped by aspiring novelists writing about dragons and star-crossed lovers?

Herman

Pretty much! That's why it has that weird, slightly dramatic flair sometimes. It was literally trained on the tropes of indie romance and sci-fi. But the point was that after it "read" those books, it had a general understanding of English. Then, if you wanted it to do a specific task, you would "fine-tune" it. You'd give it a much smaller, specific dataset for, say, answering questions. But the "base" model—the one Daniel was playing with—was just a raw text predictor. It had no idea it was supposed to be talking to a human.

Corn

This is such a crucial distinction. Today, when we use a model, we're actually using a very layered cake. You have the base model that's read the whole internet, then you have the instruction tuning where it learns to follow commands, and then you have RLHF—Reinforcement Learning from Human Feedback—where humans literally rate its answers to make it sound more "helpful" and "pleasant."

Herman

Right. GPT-one had zero of that. No instruction tuning, no RLHF, no safety filters. It was just a raw engine. If you asked it a question, it wasn't trying to "answer" you; it was trying to complete the document. If your prompt looked like the start of a Q and A, it might try to act like a Q and A. But if it got confused, it might decide the best "next word" was just the letter "A" repeated a hundred times.

Corn

It's like the difference between an engine sitting on a workbench and a finished Tesla. Daniel was basically trying to drive the engine block and wondering why there were no cup holders or a steering wheel.

Herman

That is a perfect way to put it. And to address Daniel's point about BERT—because he mentioned that name comes up a lot—BERT was Google's big model released just a few months after GPT-one. BERT was an "encoder-only" model, while GPT was "decoder-only." BERT was actually better at a lot of things back then, like understanding the relationship between words in a sentence, because it could look at a word and see the context both before and after it. BERT large scored around 84.9 on the GLUE benchmark, while GPT-1 excelled on generation tasks like LAMBADA.

Corn

Oh, I remember this. "Bidirectional Encoder Representations from Transformers."

Herman

Look at you with the full acronym! Yeah, BERT was the king of "understanding." But GPT was the king of "generating." Because GPT only looks at what came before, it's much better at actually writing new text. And it turns out, as we've seen over the last eight years, the "generating" path was the one that led to general intelligence. If you can generate the next word perfectly, you have to understand the world.

Corn

So, if GPT-one wasn't a chatbot, what were people actually using it for in twenty eighteen? I mean, was there any practical application, or was it just a lab experiment?

Herman

It was mostly a benchmark buster. GPT-one came out and absolutely smashed the state-of-the-art on nine out of twelve of the tests they used at the time. That was the "Aha!" moment for the industry. It proved that pre-training on a generic dataset actually helped the model perform better on specific tasks it had never seen before. It was the birth of the "foundation model."

Corn

That's the "Generative Pre-trained" part. The "Pre-trained" is the revolution.

Herman

Exactly. Before that, everyone thought you needed specialized architectures for every task. GPT-one said, "No, just give me a big enough transformer and enough books, and I can learn to do anything with just a little bit of extra training."

Corn

It's interesting, though, because Daniel's experience with it outputting gibberish highlights how much we take for granted now. We expect these models to be coherent for thousands of words. But back then, even getting a paragraph that didn't contradict itself was a win. I remember reading that GPT-two—the successor—was the one where people started getting actually worried about its ability to generate "deceptive" or "too-convincing" text.

Herman

Yeah, GPT-two was the one where OpenAI initially said, "This is too dangerous to release." Which, looking back at it from twenty twenty-six, feels almost adorable. GPT-two had one point five billion parameters. It was ten times bigger than GPT-one, and that jump in scale was enough to make the text go from "mostly gibberish" to "wait, did a human write this?"

Corn

It's that exponential curve. But let's go back to Daniel's question about early applications. Were there any prototypes for chatbots back then using GPT-one?

Herman

There were some very early experiments, but they were mostly in the "creative writing" space. People would use it to help them write stories. You'd write a sentence, and the model would give you the next one. It was very "collaborative" because the model was so unstable that the human had to do all the heavy lifting to keep the story on the rails. It was more like a sophisticated auto-complete than a conversation.

Corn

I actually remember a few "AI Dungeon" style games that started around then. They were using these early models to try and generate infinite adventure stories. And yeah, they were notorious for having the characters suddenly turn into inanimate objects or the plot just dissolving into a loop about a "dark and stormy night."

Herman

Right! And that brings us back to the context window. If the model can only "remember" the last five hundred words, it's going to forget that you're holding a sword or that you're in a cave. By the time you get to the end of the page, the "beginning" of the page is gone from its memory. It's like a goldfish trying to write a novel.

Corn

A goldfish on romance novels! It's a hilarious image. But okay, so if we look at where we are now—today is January twentieth, twenty twenty-six—we're seeing models that can process millions of tokens. We're talking about "infinite" context windows, or at least windows large enough to hold an entire library. When you look back at GPT-one from this vantage point, does it feel like a dead end or a direct ancestor?

Herman

Oh, it's a direct ancestor. Absolutely. The core idea—unsupervised pre-training on a decoder-only transformer—is still the heart of everything. We just added more layers, more parameters, better data, and that crucial "alignment" layer that makes it talk like a person. If you stripped away all the safety filters and the instruction tuning from GPT-five today, you'd still see the ghost of GPT-one in there. It's just a much, much smarter ghost.

Corn

It's like looking at the first single-celled organism. It's not "bad" at being an organism; it's just the starting point. I think what's really fascinating is that Daniel was able to just "download" it and run it. That's another huge change. In twenty eighteen, running a hundred-million-parameter model required some decent hardware. Now, you can run it on a phone or even in a browser.

Herman

Yeah, the "smallness" of it is now its greatest feature. We use models that size for very specific, tiny tasks now—like "is this email spam?" or "what is the sentiment of this tweet?" We don't need a trillion parameters to tell us a tweet is angry. So GPT-one's "size" has actually become the standard for "edge" AI.

Corn

That's a great point. We've gone through this cycle where we built these massive, god-like models, and now we're trying to shrink that intelligence back down into "small" models that are the same size as GPT-one but way more capable because the training techniques have improved so much.

Herman

Exactly. A "small" model today with a hundred million parameters would absolutely smoke GPT-one, even though they're the same size. We have better data now. We don't just use over 11,000 unpublished books; we use high-quality synthetic data, curated web scrapes, and textbooks. The "quality" of the "reading material" matters just as much as the "brain size."

Corn

So, for the listeners who want to try this themselves—because it is a trip—what should they look for? If they go to a site like Hugging Face and pull up these ancient models, what's the best way to interact with them to see what they were actually meant for?

Herman

Don't treat it like a chatbot. Treat it like a "complete the sentence" game. Give it a very clear, structured start. Like, "The capital of France is..." and see if it says "Paris." Or start a story with a very specific style and see if it can mimic it for three sentences. If you try to argue with it or ask it for its "opinion," you're just going to get the gibberish Daniel found. It's a mirror, not a person.

Corn

It's a mirror that's been shattered and glued back together. I think the takeaway for me is just the sheer speed of this. We're talking about less than eight years. In eight years, we went from a model that couldn't stay on topic for a paragraph to models that are passing bar exams and diagnosing rare diseases.

Herman

It's the fastest technological ramp-up in human history. And I think it's important to remember that people were blown away by GPT-one at the time. There were headlines about how "OpenAI's new model understands language better than ever." We were impressed by the "village of Paris" back then because before that, models couldn't even get "Paris" right half the time!

Corn

It's all about perspective. I remember being impressed by my first flip phone because it could send a text message. Now I'm annoyed if my watch doesn't translate a foreign language in real-time. We're very good at moving the goalposts on "impressive."

Herman

We really are. But I think there's something beautiful about going back and looking at these early versions. It reminds you that this isn't magic. It's engineering. It's math. It's trial and error. GPT-one was a very successful error that pointed the way to everything else.

Corn

Well, I think we've thoroughly dissected the "village of Paris." This was a great prompt from Daniel. It's easy to forget the history when the future is moving so fast.

Herman

Definitely. And hey, if you're out there listening and you've been enjoying these deep dives into the "weird" side of AI and tech, we'd really appreciate it if you could leave us a review on Spotify or wherever you get your podcasts. It actually helps a lot more than you'd think.

Corn

Yeah, it keeps the lights on and the sloths fed. You can find us at myweirdprompts.com if you want to send in your own prompt—maybe you found an old AI in a digital basement somewhere that we should talk about.

Herman

Or just a question that's been bugging you. We're here for all of it.

Corn

Alright, that's Episode two hundred and twenty-six. Thanks for hanging out with us in Jerusalem. I'm Corn.

Herman

And I'm Herman Poppleberry. We'll talk to you next week.

Corn

Stay curious, everyone.

Herman

And keep poking the models with sticks. It's the only way to learn.

Corn

Exactly. Bye everyone!

Herman

Bye!