EPISODE #155

Building an Ideation Factory: Beyond Generic AI Ideas

Published January 04, 2026 • Runtime: 24:05

https://myweirdprompts.com/episode/ai-high-volume-ideation/

EPISODE SYNOPSIS

In this episode of My Weird Prompts, Herman and Corn tackle the technical hurdles of high-volume AI ideation. They explore why standard LLMs often hit a "context window fatigue" wall, resulting in repetitive and generic suggestions when asked for large quantities of ideas. By proposing a sophisticated multi-agent workflow—complete with stateful memory, semantic distance auditing, and "Chain of Density" prompting—the brothers demonstrate how to transform AI into a powerful engine for solving real-world problems like the economic brain drain in Jerusalem.

DANIEL'S PROMPT

Daniel

"I believe one of the most powerful and underexplored utilities for AI is ideation. I define this as giving an AI tool a lot of context and asking for a high volume of ideas—maybe 50 or more—rather than just using it as a sounding board. About a year ago, I tried building a 'side hustle ideator' based on my resume, but I ran into issues with the agent repeating obvious recommendations because it lacked a memory of what it had already suggested. I'm interested in using this for larger goals, like brainstorming ways to improve the local economy in Jerusalem and provide more opportunities for the next generation. What is the best way to set up these ideation workflows today? Which models or components would make the most sense, and how can we implement a memory layer to avoid repetition and find those few truly original ideas?"

TRANSCRIPT

Corn

Hey everyone, welcome back to My Weird Prompts. I am Corn, and I am sitting here in our living room in Jerusalem with my brother.

Herman

Herman Poppleberry, at your service. It is a beautiful day outside, but we are tucked in here because our housemate Daniel sent us a really fascinating audio prompt this morning.

Corn

He really did. It is one of those topics that feels like it is right on the edge of how we use technology today versus how we will be using it in a few years. He is talking about ideation. Not just as a casual brainstorming partner, but as a high-volume engine for generating original ideas.

Herman

And specifically, he is looking at it through the lens of solving real-world problems right here in our backyard. He mentioned the local economy in Jerusalem and how to create more opportunities for the next generation so they do not all feel like they have to move down the road to Tel Aviv.

Corn

It is a personal one for us, too, living here. But the technical challenge he brought up is what really caught my eye. He tried building a side hustle ideator about a year ago and ran into a wall. The AI kept repeating itself. It would give him two or three good ideas and then just start circling the same drain because it did not have a memory of what it had already suggested.

Herman

That is the classic bottleneck of the large language model. We call it the context window fatigue, or more simply, the lack of stateful memory. If you ask a model for fifty ideas in one go, the probability of it repeating itself or getting generic increases exponentially the further down the list you go.

Corn

Right. It is like the model loses its creative spark because it is trying too hard to satisfy the prompt requirements without having a way to check its own work against what it just said ten seconds ago.

Herman

Exactly. And Daniel wants to know how we can fix that today, in early twenty twenty six, using the tools we have now. How do we build a memory layer? How do we find those truly original ideas hidden under the pile of obvious ones?

Corn

I think we should start by breaking down why this happens. Herman, you have been deep in the research on agentic workflows lately. Why does a standard prompt fail when we ask for volume?

Herman

Well, Corn, it comes down to how these models predict the next token. When you ask for fifty ideas, the model is essentially running a marathon. By the time it gets to idea number thirty, the preceding text in its immediate attention span is filled with its own previous suggestions. Because of the way attention mechanisms work, the model starts to gravitate toward the semantic center of what it has already written. It becomes an echo chamber of itself.

Corn

So it is not just being lazy. It is actually being too focused on what it just said.

Herman

Precisely. It is trying to be consistent, but in ideation, consistency is the enemy of novelty. You want divergence, not convergence. Most people use a single prompt for this, which is like asking one person to stand in a room and shout fifty ideas without taking a breath. Eventually, they are going to repeat themselves or pass out.

Corn

So the solution is not one big prompt, but a workflow. A system of agents or steps.

Herman

That is the first big shift. If we want to solve Daniel's problem of improving the Jerusalem economy, we cannot just ask a single model to give us fifty ideas in one block. We need to move toward what we call recursive or iterative ideation.

Corn

Okay, let us walk through that. If I am Daniel and I want to build this today, what is the architecture? Do I start with a specific model?

Herman

I would argue that the choice of model is actually secondary to the choice of framework, but you do want a high-reasoning model for the heavy lifting. Something like the latest Claude three point five Sonnet or the newer reasoning-heavy models we have seen enter the market lately. But the real magic happens in the memory layer.

Corn

And when we say memory layer, we are not just talking about a long transcript, right?

Herman

No, absolutely not. We are talking about a vector database or a simple state-tracking list. Imagine an agentic loop. Agent one generates five ideas based on the context Daniel provides, like his resume or the specific economic data of Jerusalem. Those five ideas are then sent to a separate memory bank.

Corn

And then when it goes for the next five, it checks that bank?

Herman

Exactly. But it is more than just checking. You use a separate agent, let us call it the Critic or the Deduplicator. Its only job is to look at the new ideas and compare them to the old ones. If it sees something too similar, it rejects it and tells the first agent to try again, but specifically tells it what to avoid.

Corn

That sounds like it would solve the repetition issue Daniel mentioned. But what about the quality? He said he kept getting obvious recommendations, like being an AI consultant. How do we push the AI to find the weird, original stuff?

Herman

That is where we get into the temperature settings and the diversity of the seeds. If you keep the temperature low, the model will always give you the most probable, safe answer. To get to the original ideas, you have to force the model into the long tail of probability.

Corn

I love that phrase, the long tail of probability. It sounds like a place where you would find a very strange treasure.

Herman

It really is. One way to do this is to vary the personas. Instead of just asking for economic ideas, you ask the model to brainstorm as if it were a nineteenth-century urban planner, then as a twenty-first-century tech nomad, then as a local shop owner in the Old City. By shifting the perspective, you force the model to pull from different parts of its training data.

Corn

That is brilliant. It is like the Oblique Strategies cards that musicians use to break creative blocks. You are giving the AI a constraint that forces it to be creative.

Herman

Exactly. And because you have that memory layer we talked about, you can ensure that the urban planner is not suggesting the same things as the tech nomad. You are building a diverse portfolio of ideas rather than just a long list of the same idea.

Corn

I want to dig deeper into the Jerusalem context Daniel mentioned, because that is a very specific and difficult nut to crack. But before we get into the weeds of local economic theory and AI agents, let us take a quick break.

Herman

Good idea. We will be right back.

Corn

Let us take a quick break for our sponsors. Larry: Are you tired of your thoughts being disorganized? Do you wish you could capture every brilliant idea before it floats away into the ether? Introducing the Thought-Catcher Headband. Using patented copper-coil technology and a very small, harmless amount of static electricity, the Thought-Catcher reads your brain waves and translates them into Morse code that is recorded directly onto a vintage cassette tape worn around your waist. No more forgetting your grocery list! No more losing that million-dollar invention! Disclaimer: May cause mild scalp tingling, vivid dreams about the nineteen seventies, and an inexplicable craving for saltine crackers. The Thought-Catcher Headband. It is like a net for your mind. Larry: BUY NOW!

Corn

Alright, thanks Larry. I am not sure I want my thoughts recorded in Morse code on a cassette tape, but I appreciate the enthusiasm.

Herman

I actually think the static electricity might interfere with my donkey ears, Corn. I will stick to the AI agents for now.

Corn

Probably for the best. So, back to Daniel's prompt. We were talking about the Jerusalem economy. It is a unique challenge because we have this incredible brain drain. People study at Hebrew University or Bezalel, and then they often head to the coast for the high-tech jobs.

Herman

It is a classic problem of a secondary hub. But Daniel's idea of using AI for high-volume ideation to solve this is actually very timely. If we want to find opportunities that are not just the obvious ones, we have to look at the intersections of what makes Jerusalem unique.

Corn

Right. It is not just about building another software-as-a-service company. It is about the intersection of history, religion, tourism, and deep tech.

Herman

So, if I were setting up this workflow for Daniel today, I would use a multi-agent system. I would start with a Research Agent. Its job is to ingest all the current economic data, the demographic shifts, and the existing infrastructure plans for Jerusalem.

Corn

And we can do that now with much larger context windows than we had a year ago, right?

Herman

Oh, absolutely. We can feed in thousands of pages of city planning documents. Then, I would have an Ideation Agent that uses a technique called Chain of Density. This is a prompt engineering method where you ask the model to generate an idea, then identify what is missing or what is too generic about it, and then rewrite it to be more information-dense and specific.

Corn

I remember we touched on something similar back in episode one hundred and eighty four when we were talking about the evolution of the transformer architecture. The idea that the model can critique its own output is really the key to moving past that first layer of boring ideas.

Herman

It really is. And to solve Daniel's specific issue with memory, I would implement what is called a CRUD system—Create, Read, Update, and Delete—for the ideas. As the Ideation Agent works, it writes to a shared database. A second agent, let us call it the Diversity Auditor, reads that database and calculates the semantic distance between the new ideas and the existing ones.

Corn

Semantic distance. Can you explain that for the listeners?

Herman

Sure. Every idea can be turned into a mathematical vector, a string of numbers that represents its meaning. If two ideas have vectors that are very close together, they are basically the same idea in different words. The Diversity Auditor would reject anything that is too close to an existing vector. This forces the Ideation Agent to pivot.

Corn

So if the agent keeps trying to suggest AI consulting, the Auditor eventually says, no, your vector for AI consulting is already full. You have to move to a different part of the map.

Herman

Exactly! It forces the model to explore the edges of its knowledge. Maybe it starts looking at the intersection of stone masonry and augmented reality, or high-tech agriculture in the Judean hills. That is where the truly original ideas live.

Corn

This feels like a much more active way of using AI. It is not just asking a question; it is building a factory for thought.

Herman

That is a great way to put it. An ideation factory. And the best part is that once you have the fifty or one hundred ideas, you can use another agent to filter them based on Daniel's resume and skills. It is like a funnel. You start with a massive, diverse set of possibilities and then narrow it down to the ones that are actually feasible for him to execute.

Corn

I think one thing Daniel mentioned that is really important is the next generation. How do we create opportunities for them? If we are using this AI workflow, can we specifically prompt for intergenerational wealth or community-building?

Herman

Absolutely. One of the most powerful things about these models is their ability to simulate complex systems. You could have an agent whose only job is to play the role of a twenty-year-old student in Jerusalem five years from now. It looks at the ideas being generated and asks, would this make me want to stay in the city? If the answer is no, the idea gets sent back for revision.

Corn

That is like having a focus group of the future, right in your computer.

Herman

It really is. And when you combine that with the real-world constraints of the Jerusalem economy—the cost of housing, the transport links, the cultural diversity—you get something much more robust than a simple brainstorming session.

Corn

So, to summarize the technical stack for Daniel: He needs a multi-agent framework like LangGraph or maybe the newer stateful versions of AutoGen. He needs a vector database like Pinecone or even just a local Chroma instance to track the ideas and avoid repetition. And he needs a clear set of diverse personas and a critic agent to push the boundaries.

Herman

And do not forget the importance of the initial context. The more Daniel can feed into that first Research Agent—real data, not just generalities—the better the output will be. If he gives it his resume, his specific interests, and the actual economic challenges we see every day walking through the Shuk or Har Hotzvim, the AI will have the ingredients it needs to cook up something special.

Corn

It is funny, we often think of AI as this thing that is going to replace creativity, but what you are describing is a system that actually demands more creative direction from the human. Daniel has to be the architect of the factory.

Herman

I love that. The human is the architect, the AI is the engine. And in a city like Jerusalem, which has been built and rebuilt for thousands of years, there is something poetic about using the most modern tools to think about the next layer of its history.

Corn

It really is. I think about our own experience here. We have seen the city change so much just in the time we have lived together. There is so much untapped potential in the neighborhoods that people usually do not associate with high-tech.

Herman

Exactly. And that is where the high-volume ideation pays off. If you only generate five ideas, you will get the obvious ones. If you generate fifty, you might find that one idea that connects a traditional craft in the Old City with a global export market using decentralized logistics. Something that no one would have thought of in a standard meeting.

Corn

It makes me wonder about the second-order effects. If everyone starts using these ideation factories, do we end up with a surplus of great ideas but a shortage of people to execute them?

Herman

That is the big question for twenty twenty six and beyond, Corn. Execution is still the bottleneck. But I would argue that a truly great, original idea—one that fits the person and the place perfectly—carries its own momentum. It is much easier to start a business when the idea feels like a perfect click rather than a forced effort.

Corn

That is true. A lot of people fail because they are working on the wrong thing, not because they are not working hard.

Herman

Precisely. This workflow Daniel is asking about is really a way to minimize that risk. It is a way to find the path of least resistance between your skills and the needs of your community.

Corn

I think we should also touch on the misconception that AI cannot be original. People always say, it just predicts the next word, it cannot think of anything new. But when you use these diversity-enforcing workflows, aren't you essentially forcing it to synthesize its training data in ways that have never been seen before?

Herman

That is exactly what is happening. Originality is often just a new combination of existing elements. By using a memory layer to block the common combinations, you are forcing the model to find the rare ones. It is like a chemist trying to create a new molecule. They know all the elements, but it is the specific arrangement that makes it new.

Corn

So the AI is not a creator in the vacuum, but it is a master of synthesis if you give it the right constraints.

Herman

Exactly. And for Daniel, those constraints are his life in Jerusalem and his desire to see the city thrive. That is the heart of the prompt.

Corn

I am really excited to see what he builds with this. He has a way of taking these technical concepts and actually turning them into something functional.

Herman

He really does. And if he manages to find a few truly original ideas for the local economy, it could have a real impact. Imagine a Jerusalem where the next generation feels like the city is a laboratory for the future, not just a museum of the past.

Corn

That is a beautiful vision, Herman. It reminds me of what we discussed in episode two hundred and fifty six about building the future of the Negev. These regional hubs have so much power if they can just figure out their unique value proposition in the age of AI.

Herman

It is all connected. The tools we are using to think about these problems are becoming as important as the problems themselves.

Corn

Well, I think we have given Daniel—and our listeners—a lot to chew on. The move from single prompts to agentic, stateful workflows is really the frontier of how we interact with these models today.

Herman

It is the difference between having a conversation and building a system. And for anyone listening who is feeling stuck in their own brainstorming, I highly recommend trying this. Even if you do not build a full agentic loop, just the act of telling the AI, here are ten ideas I already have, now give me ten more that are completely different, can break that cycle of repetition.

Corn

That is a great practical takeaway. Just providing the negative constraints can be incredibly powerful.

Herman

It really is. Well, Corn, I think my donkey brain is starting to overheat from all this high-level talk. Maybe we should go for a walk and see if we can spot any of these economic opportunities in the wild.

Corn

Good idea. Let us head out. But before we go, I want to say a huge thank you to Daniel for sending in this prompt. It is exactly the kind of deep dive we love doing.

Herman

Absolutely. And to all of you listening, if you are enjoying our explorations here on My Weird Prompts, we would really appreciate it if you could leave us a review on your favorite podcast app. It genuinely helps other curious minds find the show.

Corn

It really does. You can find us on Spotify and at our website, myweirdprompts.com. We have an RSS feed there for subscribers and a contact form if you want to send us your own weird prompts.

Herman

We love hearing from you. Whether it is about AI, urban planning, or the best way to keep a sloth awake during a recording session.

Corn

Hey! I was perfectly awake for this one. This topic was too good to nap through.

Herman

I will take your word for it this time.

Corn

Alright, everyone. Thanks for joining us for episode two hundred and sixty one. We will be back next week with more deep dives and hopefully fewer Morse code cassette tapes.

Herman

No promises on the tapes! Until next time, keep being curious.

Corn

This has been My Weird Prompts. Thanks for listening.

Herman

Goodbye from Jerusalem!

Corn

Bye everyone.

Herman

Wait, Corn, did we mention the specific vector database settings for the diversity auditor?

Corn

I think we can save the technical fine-tuning for the show notes, Herman. Let them breathe a little.

Herman

Fair enough. I just get so excited about the cosine similarity math. It is just so elegant.

Corn

We know, Herman. We know. Now let's go get some hummus.

Herman

Hummus and high-volume ideation. The perfect afternoon.

Corn

Exactly. Signing off for real now.

Herman

See ya!