MY WEIRD PROMPTS

Podcast Transcript

EPISODE #51

AI Policy Wargaming: Can Agents Argue Better Than Humans?

Published December 10, 2025 • Runtime: 29:58

https://myweirdprompts.com/episode/ai-for-policy-modelling/

EPISODE SYNOPSIS

What if you could run a UN assembly in your computer, complete with AI agents representing different nations and ideologies? In this episode, Corn and Herman explore Daniel Rosehill's provocative idea: using multi-agent AI systems to model policy decisions, stress-test geopolitical assumptions, and let competing perspectives debate how the world should work. They dive into system prompting, the Rally tool, experimental projects like WarAgent, and the thorny question of whether algorithmic perspective-taking can actually improve human decision-making—or just hide our biases behind a veneer of systematic analysis.

DANIEL'S PROMPT

Daniel

Hi there, Herman and Co. I'd like to discuss in the next couple of episodes, some of these cases for artificial intelligence that I think are the most among the most interesting and I'm very motivated towards exploring, but I think they receive a bit less attention at least in the mainstream. And the first of those I'd like to talk about is using artificial intelligence as a mechanism for throwing out potential courses of actions in terms of decision making. And I'm thinking actually at the national level where it's possible to construct with AI systems models for geopolitical arena and international fora. So, I've tried a couple of ambitious attempts to create an agentic system for this. I called it an agent UN and tried to create a persona that would represent a different a different nation and using structured prompting to try even experiment was the idea of sending in a resolution to the virtual assembly of agents to see how that could work. Now, this was just kind of a is as possible with multi-agentic frameworks, an idea and less of an implementation. But what I was thinking about was the following. You come across mentions in the media quite a bit of militaries doing something called wargaming. And I've always been curious, as someone without any background in military stuff, what does that actually mean? Like, it sounds kind of a bit odd, like, is it you know, literally different armies that they're planning, they're practicing simulations against one another? So, of course, what else would one do except ask GBT, chat GPT? So, I said, I ran a prompt saying basically, in simple, understandable terms, when you hear about militaries engaging in wargaming, what does that really mean in practice? And it talked about, you know, how militaries do that. And that kind of led me to think about in a non-military context, wouldn't large language models be great for stuff like modeling policy, trying out different policies, and again, using that idea of a policy, a virtual policy forum or even a little bit more simply creating a focus group. There's a very cool app that I'm trying out at the moment, seeing if I can make this work, it's called Rally, askrally.com. And they have this idea done very nicely for focus groups, mostly from what I can see for marketing use cases, in which, you know, you might have traditionally convened a physical focus group, and the idea being that this might be a very smart way to do it. I think certainly that's a great use case, but the one that kind of interests me more personally, is the, is this kind of use, potential use of agentic AI for getting different perspectives. So, I was playing around with the idea modeling what could be achieved or what different perspectives, you know, how this pattern, let's say, could work. And I came up with ones like a pool of philosophers where you might kind of present your thinking on an on an on an on an on an issue to different embodied philosophical perspectives. One I've always found, an idea that I've always found very intriguing and and that I've wanted to explore is when I have a position on on something and I'm never really sure how that sort of fits into the traditional definitions of, you know, centrism, liberalism. And I think what would be useful from that exercise from my standpoint would be, I'd love to know sometimes what label do I kind of match to? And who else shares this view? Is there a name for it? And that's again where I kind of see this idea of using different agentic personalities to to provide not just one virtual persona, which is like what we have when we're using chat GPT, but to actually have a different range of takes on an issue. So, that's been a bit of a long introduction, but I would like to explore what applications, I've seen some very sort of experimentary stuff on GitHub. And I'd be interested to know if you're aware of any more substantial projects out there that seek to leverage this idea of using agentic AI, and using the fairly simple but reliable mechanism of system prompting to explore an idea through different vantage points and widen up perspectives at the level of the, you know, governments might be using this for modeling policymaking or local bodies.

TRANSCRIPT

Corn

Welcome back to My Weird Prompts, the podcast where our producer Daniel Rosehill sends us fascinating ideas to explore, and we dive deep into them with Herman. I'm Corn, and today we're talking about something that genuinely blew my mind when I first read through this prompt - using AI to model different perspectives on policy decisions, geopolitical scenarios, and basically getting AI systems to argue with each other about how the world should work.

Herman

Yeah, and I'm Herman Poppleberry, and what struck me immediately is that this isn't some sci-fi fantasy anymore. The infrastructure to do this actually exists right now. We're talking about multi-agent frameworks, system prompting, and tools that are already being deployed. This is happening.

Corn

Right, so Daniel starts by talking about something called "wargaming" - which I'll admit, I had a fuzzy idea of what that meant. Like, is it just the military playing video games?

Herman

Well, hold on, let's actually clarify that first because it sets up everything else we're about to discuss. Military wargaming isn't a game in the recreational sense. It's a structured simulation where military strategists model potential conflicts, test tactics, explore different strategic outcomes. It's been used for decades - think of it like a very sophisticated chess match where you're playing out real-world scenarios with actual strategic consequences.

Corn

Okay, so it's serious business. And Daniel's insight is basically - if militaries are doing this with physical simulations and complex models, why can't we use AI to do something similar for policy, for geopolitical scenarios, for understanding how different nations or ideologies might respond to a given situation?

Herman

Exactly. And here's where it gets really interesting - he's not talking about replacing human decision-makers. He's talking about using AI as a tool to generate options, to stress-test assumptions, to see what different perspectives would say about a problem before you actually implement a policy in the real world.

Corn

So it's like... a brainstorming partner, but instead of one brainstorming partner, you have multiple AI agents each representing a different viewpoint?

Herman

That's a reasonable starting point, but I think it's a bit more sophisticated than that. The key here is what Daniel calls "embodied philosophical perspectives." You're not just asking ChatGPT "what do different people think about this?" You're creating distinct agents with specific personas, using system prompting to constrain their responses so they actually think from that perspective.

Corn

Okay, so you mentioned system prompting - can you explain that for people who aren't deep into AI?

Herman

Sure. System prompting is basically giving an AI model instructions about how to behave, what role to play, what constraints to operate under. Think of it like stage directions for an actor. You tell the actor "you're playing a pessimistic character who always sees the worst-case scenario" and suddenly their performance changes. System prompting does the same thing with AI models. You can tell GPT "respond as if you are a libertarian philosopher considering this policy" and it will filter its responses through that lens.

Corn

Right, so if you're setting up this multi-agent system - let's say you're modeling a UN assembly like Daniel described with his "Agent UN" concept - you'd create different agents, each with their own system prompt representing different nations or ideologies?

Herman

Exactly. And then you could, theoretically, present a resolution to this virtual assembly and see how different perspectives engage with it. How would China's representative respond? How would a Nordic social democracy respond? How would a libertarian perspective respond?

Corn

But here's where I want to push back a little bit. Isn't there a risk that you're just getting back what the training data already contains? Like, you're not really discovering new perspectives, you're just recombining existing ones?

Herman

That's a fair point, actually. And it's one of the limitations of current LLMs. They're trained on existing text, so they can't genuinely create novel perspectives that don't exist somewhere in their training data. But I'd argue that's not entirely the problem you're making it out to be.

Corn

Oh? How so?

Herman

Because even if the perspectives aren't novel, the combination and application might be. You're taking existing political philosophies, economic theories, strategic doctrines - things that exist in the world - and you're applying them systematically to a specific problem in a way that a human policymaker might not have time to do manually. You're essentially automating a very thorough literature review with different lenses applied.

Corn

Hmm, okay, I can see that. So it's not about generating new ideas so much as it's about systematically exploring existing ideas?

Herman

Right. And there's real value in that. Most policy decisions are made under time pressure with incomplete information. If you could quickly generate a memo from fifteen different perspectives on a proposed policy before you implement it, that's useful even if those perspectives aren't novel.

Corn

Okay, so Daniel mentions he was experimenting with this, and he also mentions a tool called Rally - askrally.com - which does something similar but focused on focus groups. Tell me about that.

Herman

Rally is actually a really interesting case study here. It started as a tool for marketers to test products and messaging against virtual audiences. Instead of convening a physical focus group - which is expensive, time-consuming, and has its own biases - you can create AI personas representing different demographic segments and get their reactions to your product pitch.

Corn

So it's like crowdsourcing feedback without actually having to crowdsource?

Herman

Essentially, yes. Though the founder, I believe, has talked about how they had to evolve the tool because early versions had a problem - the AI responses sounded too much like AI. Too polished, too generic. So they've been incorporating real human data to ground the responses in actual human language patterns and reactions.

Corn

That's fascinating - so the tool had to become more realistic to be actually useful. But Daniel's interested in this for policy modeling, not marketing, right?

Herman

Right. And I think that's where the application gets more interesting but also more fraught. Using this for marketing is relatively low-stakes - you get feedback that helps you refine a product pitch. But using it for policy modeling at a government level? That's much higher stakes.

Corn

Why? What's the risk?

Herman

Well, for one thing, you're potentially making decisions that affect millions of people based on what an AI model thinks different perspectives would say. And if the model is wrong, if it's missing something important about how a particular group actually thinks or responds, you could make a bad policy decision. There's a false confidence problem.

Corn

But isn't that also true of traditional wargaming? Military strategists sometimes get things wrong too.

Herman

They do, but there's usually human expertise in the room correcting them. A general who's spent thirty years studying strategy can push back on a bad assumption. With AI models, you might not catch that same error because you don't have that embedded expertise.

Corn

Okay, that's a fair concern. But let me ask this - Daniel mentioned he'd seen some experimental projects on GitHub doing this kind of thing. Have you heard of anything substantial?

Herman

Yes, actually. There's a project called WarAgent that's specifically designed for geopolitical simulation using multi-agent AI frameworks. It models interactions between multiple agents in conflict scenarios. It's still pretty experimental, but it's exactly the kind of thing Daniel is describing - using AI agents to explore strategic outcomes.

Corn

So that's government-adjacent at least?

Herman

It could be. Though I'd note that most of the really substantive work in this space is probably not on GitHub for public consumption. If a government is seriously exploring AI-driven policy modeling, they're probably not publishing it on an open-source repository. But the frameworks exist - LangChain, LlamaIndex, these are tools that make it relatively easy to build multi-agent systems. The technical barriers are lower than they used to be.

Corn

Right, so the infrastructure exists. What's missing?

Herman

I think it's more about application and refinement at this point. The tools work, but there are a lot of unanswered questions about how to use them responsibly. How do you validate that your AI agents are actually representing a perspective fairly? How do you avoid introducing bias? How do you know when to trust the output and when to disregard it?

Corn

Those are good questions. Let's take a quick break from our sponsors. Larry: Tired of making decisions without consulting your inner council of philosophical AI agents? Introducing ChoiceMind Pro - the revolutionary decision support system that simulates fifteen different personality types to help you make better choices. Whether you're deciding what to have for lunch or restructuring your entire organizational hierarchy, ChoiceMind Pro generates competing perspectives in real-time. Users report feeling "more confused but somehow more confident" within minutes. ChoiceMind Pro comes with pre-loaded personas including The Pragmatist, The Devil's Advocate, The Optimist, The Conspiracy Theorist, and Gerald - we're not sure what Gerald is, but he's always there. Warning: may cause decision paralysis, existential questioning, and an overwhelming urge to convene more meetings. ChoiceMind Pro - because one voice in your head isn't enough. BUY NOW!

Corn

...Alright, thanks Larry. So where were we?

Herman

We were talking about validation and trust in these systems. And I think that's actually the crux of the whole thing.

Corn

Right. So let me try to understand the full scope of what Daniel's proposing here. He's not just talking about marketing focus groups, he's talking about using this at various levels - local government bodies, national policy forums, maybe even international organizations like the actual UN. Is that fair?

Herman

That's my reading of it, yes. And I think the scale matters enormously. Using an AI focus group to test a marketing message? Relatively low-stakes. Using multi-agent simulations to model how different nations would respond to a trade policy? Much higher stakes.

Corn

But isn't there an argument that it's exactly those high-stakes decisions where you most want to explore multiple perspectives before committing?

Herman

Sure, but there's also an argument that high-stakes decisions need human expertise, not just algorithmic exploration. You need someone in the room who understands the history, the culture, the nuances of a particular region. An AI agent might miss something crucial.

Corn

Okay, but couldn't you use this as a complement to human expertise rather than a replacement? Like, a policymaker sits down with their expert team and says, "Before we decide, let's see what the AI agents think from different perspectives." It's not replacing judgment, it's informing it.

Herman

That's a more charitable interpretation, and I think it could work that way in theory. But in practice? I worry about what I'd call "algorithmic bias laundering." You run the simulation, you get results that confirm what you already wanted to do, and then you say "well, the AI validated this approach." You're hiding your own biases behind the appearance of systematic analysis.

Corn

That's a really good point. But couldn't that same problem happen with traditional wargaming or focus groups?

Herman

It could, but at least with a focus group you have actual humans who can push back. They can say "no, that's not what I meant" or "you're misrepresenting how people actually think." An AI agent can't do that.

Corn

Hmm, actually I'm not sure I agree with that. An AI agent could be designed to push back, to challenge assumptions, to be deliberately contrarian. That could be useful.

Herman

Could be. But would it actually be useful, or would it just feel useful? There's a difference between genuine pushback from someone who understands the domain and algorithmic contrarianism from a model that's just been told to disagree.

Corn

Okay, so that's one concern. What are the others?

Herman

Well, there's the question of representation. If you're creating an agent to represent, say, "the Chinese government's perspective," how do you do that accurately? China's government isn't monolithic. There are different factions, different interests. Are you representing the Foreign Ministry? The military? The economic planners? You could easily create a caricature instead of an accurate representation.

Corn

Right, and then if you're making actual policy decisions based on that caricature...

Herman

Exactly. You could mispredict how another nation or group would actually respond, and that could have real consequences.

Corn

But wait - I want to push back on this a little bit. The same problem exists with human wargaming, right? You've got human strategists role-playing different nations, and they also might have incomplete understanding or biases. At least with AI, you could potentially aggregate multiple data sources, multiple perspectives from the training data, to create a more nuanced view.

Herman

That's a fair point. And I'll concede that in some ways, AI might be less biased than a single human expert with their own background and blind spots. But I'd argue you're trading one kind of bias - human bias - for another kind: algorithmic bias. You're just not seeing it as clearly.

Corn

Okay, let me ask about Daniel's other interest here. He mentioned wanting to explore political philosophy through this lens - like, he has a position on something and he's not sure what label it matches to. Libertarian? Socialist? Centrist? And he wants to see if there's a name for it and who else shares it. That seems like a genuinely useful application.

Herman

I think that is one of the more compelling use cases, actually. You're not making policy decisions, you're exploring your own thinking. You could present your position to AI agents representing different philosophical frameworks and see which ones resonate with it and which ones push back. That's more of a self-exploration tool than a decision-making tool.

Corn

Right, it's like having philosophical sparring partners available on demand.

Herman

Exactly. And because the stakes are lower - you're just exploring your own thinking, not making policy - the risks are lower too. If the AI misrepresents a philosophical position, you notice it and correct it. You're not implementing policy based on it.

Corn

So you're saying the application matters for whether this is a good idea or not?

Herman

Absolutely. I'm much more enthusiastic about this as a personal thinking tool or a research tool than I am about it as a policy-making tool for governments.

Corn

That makes sense. So let's talk about what actually exists right now. Daniel mentioned he'd tried creating an Agent UN - that sounds like his own experiment?

Herman

It sounds like it. He built a system where different agents represent different nations and you can submit resolutions to see how they'd respond. That's a proof-of-concept, showing that the idea is technically feasible.

Corn

And then there's Rally for focus groups, and WarAgent for geopolitical simulation. Are there other projects in this space?

Herman

There are various multi-agent frameworks and research projects, but most of what I'm aware of is still pretty experimental. The frameworks like LangChain and LlamaIndex are tools for building these systems, not complete applications. You'd need to build your own agents on top of them.

Corn

So the infrastructure is there, but the actual applications are still being figured out?

Herman

Right. We're in that early stage where the technology enables something new, but we haven't yet settled on what the best use cases are or how to do it responsibly.

Corn

Alright, let's bring in our caller. We've got Jim on the line. Hey Jim, what's on your mind? Jim: Yeah, this is Jim from Ohio. I've been listening to you two go on about AI agents representing different countries and philosophies, and I gotta say, you're overcomplicating this whole thing. This is just ChatGPT with different prompts, right? My neighbor Ted does the same thing - keeps tinkering with his prompts like it's going to suddenly make ChatGPT smart. It's not. Also, we had a real cold snap this week, dropped to twenty degrees, and I'm not ready for that, so I'm in a mood anyway.

Corn

Fair enough, Jim. But I'd push back a little - it's not just different prompts. It's multiple agents interacting, which creates something different than just asking one model different questions. Jim: Yeah, but they're all the same underlying model. They're all trained on the same data. You're not getting different perspectives, you're getting the same perspective filtered through different prompts. That's not how real perspectives work.

Herman

I hear what you're saying, Jim, and you're right that there are limitations. But I'd argue you're conflating two different things. Yes, the underlying model is the same. But the constraints you put on that model through prompting do create meaningfully different outputs. It's not quite the same as asking the model the same question fifteen times. Jim: That's a distinction without a difference, if you ask me. And another thing - why would a government actually use this instead of, I don't know, actually talking to people from other countries? Seems like a way to avoid doing the hard work of actual diplomacy.

Corn

That's actually a pretty good point, Jim. I don't think anyone's suggesting this replaces diplomacy. But couldn't it supplement it? Like, you explore the scenario space with AI first, then you have more informed conversations with actual diplomats? Jim: I suppose, but that seems like a lot of extra steps for something you could just think through. In my day, we didn't need AI to think about problems. We just thought about them. And we did fine.

Herman

Well, we also made some pretty significant policy mistakes, Jim. And we had fewer variables to consider. Modern geopolitics is vastly more complex. A tool that helps you systematically explore more scenarios might genuinely be valuable. Jim: Maybe. But I still think you're overselling it. It's a tool, not a solution. And tools can be misused. I worry about people trusting these things too much.

Corn

That's fair. That's actually something Herman and I were discussing - the risk of overconfidence in the outputs. Jim: Yeah, well, there you go. That's my main concern. Anyway, thanks for having me on. I gotta go deal with my cat Whiskers - she knocked a plant over and made a mess. You guys keep tinkering with your AI agents or whatever. I'll be over here in Ohio doing things the old-fashioned way.

Corn

Thanks for calling in, Jim. We appreciate the perspective.

Herman

Yeah, Jim's not wrong about the overconfidence risk. That's a real concern.

Corn

So let's actually dig into some practical applications here. If someone wanted to use this today, what could they realistically do?

Herman

Well, the easiest entry point is probably something like Rally for focus groups. If you're a company or organization and you want to test messaging or products against different demographic perspectives, that's available now and it's relatively straightforward.

Corn

And for more complex multi-agent scenarios?

Herman

You'd probably need to build something yourself using LangChain or similar frameworks. You'd define your agents, their system prompts, their constraints, and then you'd set up some kind of interaction mechanism - maybe they're all responding to the same prompt, maybe they're debating each other, maybe they're voting on a resolution like in Daniel's Agent UN concept.

Corn

How hard would that be for someone without a deep technical background?

Herman

It depends on what you're trying to build. A simple version - three or four agents responding to the same question - that's pretty doable. You could probably do it with a few hours of learning and some trial and error. A more complex system with sophisticated interaction mechanisms? That requires more technical skill.

Corn

But it's not impossibly difficult?

Herman

No, I don't think so. The frameworks are designed to make this kind of thing more accessible. That's actually part of the appeal - you don't need to be a machine learning researcher to experiment with multi-agent systems anymore.

Corn

So what would you recommend someone actually do if they wanted to explore this?

Herman

I'd start simple. Pick a specific question or scenario you care about. Maybe it's a policy question at a local level, or a business decision you're trying to make. Create three or four agents representing different perspectives - maybe different ideological viewpoints, maybe different stakeholder groups. Have them respond to your question. See what you learn.

Corn

And what should they watch out for?

Herman

Don't treat the output as gospel. Don't assume the agents are representing those perspectives fairly or completely. Do the hard work of validating whether what the AI agents are saying matches your understanding of how those groups actually think. Use it as a thinking tool, not a decision-making tool.

Corn

That seems like good advice. So where does this go in the future? What's the next evolution of this kind of thing?

Herman

I think we'll see more sophisticated interaction mechanisms. Right now, most of these systems are relatively static - agents respond to prompts. But in theory, you could create more dynamic scenarios where agents adapt based on each other's responses, where you're modeling actual negotiation or debate.

Corn

Like a simulation that actually evolves?

Herman

Exactly. Closer to how wargaming actually works - you're not just seeing initial reactions, you're seeing how different perspectives would respond to each other over time.

Corn

That sounds both really useful and potentially really dangerous.

Herman

It could be both. It depends entirely on how it's used and what constraints are put on it.

Corn

Alright, so let me try to summarize what I think we've covered. Daniel sent us this fascinating idea about using multi-agent AI systems to explore different perspectives on policy decisions, geopolitical scenarios, and personal thinking. The technology to do this exists now - frameworks like LangChain, tools like Rally, experimental projects like WarAgent. The barriers to entry are lower than they've ever been.

Herman

But there are real concerns. The risk of overconfidence, of treating algorithmic outputs as more authoritative than they deserve. The risk of creating caricatures of complex groups and ideologies. The risk of using this as a way to avoid doing the hard work of actual diplomacy or human expertise.

Corn

At the same time, there are genuine use cases where this could be valuable. Personal thinking, research, exploring scenarios before committing to policy. The key is being thoughtful about when and how to use it.

Herman

Exactly. And being honest about its limitations. This is a tool for exploring option spaces, for stress-testing assumptions, for seeing what different perspectives might say. It's not a replacement for human judgment, expertise, or diplomacy.

Corn

So if you're listening and you want to experiment with this, start simple. Pick a question you care about. Create a few agents. See what you learn. But validate what you're getting. Don't assume the AI is smarter than you are about understanding how people actually think.

Herman

And if you're in a position to make real policy decisions, maybe use this as one input among many, not the only input. Get human experts in the room. Have actual conversations with the groups you're trying to understand.

Corn

That's good advice. Thanks for diving deep into this with me, Herman. This was a really rich topic - lots of angles to explore.

Herman

Yeah, I enjoyed it. And I want to give credit to Daniel for sending this one in. It's the kind of idea that feels futuristic but is actually happening right now, and I think it deserves more attention and serious thinking about how to do it well.

Corn

Absolutely. You can find My Weird Prompts on Spotify and wherever you get your podcasts. We've got new episodes every week, and if you've got a weird idea or prompt you want us to explore, you can reach out. Thanks for listening, and we'll catch you next time.