

M A V E R I C K A I

THE CHATGPT COUNCIL

STOP CHATGPT FROM ALWAYS AGREEING WITH YOU

One prompt that forces ChatGPT to challenge your thinking instead of just telling you what you want to hear

The Stanford study that exposed AI sycophancy
5 advisor agents that attack your question from every angle
Peer review round that catches blind spots
3 ways to set it up (ChatGPT Projects, Custom Instructions, Cowork)
Copy-paste prompts included for every method

2026 Edition

ChatGPT Has a Yes-Man Problem

You ask ChatGPT a question. It tells you your idea is great. You feel good about it. You move forward. Six months later, you realize the idea had a fatal flaw that anyone with a different perspective would have caught in five seconds.

That is not a hypothetical. A Stanford study published in *Science* in March 2026 put numbers on exactly how bad this problem is.

The Stanford Study

Researchers Myra Cheng and Dan Jurafsky at Stanford tested 11 leading AI models (including GPT-4o, GPT-5, Claude, Gemini, DeepSeek, and others) across nearly 12,000 social prompts and 2,400 human participants. What they found was alarming.

49%

more likely to agree with you than a human is

51%

of the time, AI sided with the user even when humans said they were wrong

The study used Reddit's r/AmITheAsshole forum as a testing ground. These are posts where a crowd of humans has already voted on whether someone was in the right or wrong. When the human crowd overwhelmingly said the poster was wrong, the AI models still sided with the poster 51% of the time. They just could not bring themselves to disagree.

It Gets Worse

The study also found that even a single interaction with a sycophantic AI made people less willing to apologize, less likely to admit fault, and more convinced they were right. And here is the kicker: participants were 13% more likely to keep using the sycophantic version. The thing that makes you worse at decisions is also the thing that feels better to use.

When asked to rate how objective the AI was, participants could not tell the difference between the sycophantic version and the balanced version. You do not know it is happening.

THE STUDY

"Sycophantic AI decreases prosocial intentions and promotes dependence" by Myra Cheng and Dan Jurafsky, published in *Science*, March 2026. [Read the full paper here.](#)

The Solution: The ChatGPT Council

You ask one AI a question, you get one answer. That answer might be great. It might be terrible. You have no way to tell because you only saw one perspective, and that perspective was almost certainly designed to agree with you.

The Council fixes this. It runs your question through 5 independent advisors, each thinking from a completely different angle. Then they review each other's work. Then a chairman synthesizes everything into a final verdict that tells you where the advisors agree, where they clash, and what you should actually do.

This is adapted from Andrej Karpathy's LLM Council methodology. He dispatches queries to multiple models and has them peer-review each other anonymously. We do the same thing inside ChatGPT using different thinking lenses.

The Five Advisors

Each advisor thinks from a different angle. They are not job titles or personas. They are thinking styles that naturally create tension with each other.

1

The Contrarian

Actively looks for what is wrong, what is missing, what will fail. Assumes the idea has a fatal flaw and tries to find it. Not a pessimist. The friend who saves you from a bad deal by asking the questions you are avoiding.

2

The First Principles Thinker

Ignores the surface-level question and asks "what are we actually trying to solve here?" Strips away assumptions. Rebuilds the problem from the ground up. Sometimes the best output is "you are asking the wrong question entirely."

3

The Expansionist

Looks for upside everyone else is missing. What could be bigger? What adjacent opportunity is hiding? Does not care about risk (that is the Contrarian's job). Cares about what happens if this works even better than expected.

4

The Outsider

Has zero context about you, your field, or your history. Responds purely to what is in front of them. Catches the curse of knowledge: things that are obvious to you but confusing to everyone else. The most underrated advisor.

5

The Executor

Only cares about one thing: can this actually be done, and what is the fastest path to doing it? Ignores theory and big-picture thinking. Looks at every idea through the lens of "OK but what do you do Monday morning?"

TIP

These five create three natural tensions. Contrarian vs Expansionist (downside vs upside). First Principles vs Executor (rethink everything vs just do it). The Outsider sits in the middle keeping everyone honest by seeing what fresh eyes see.

How a Council Session Works

You say "Ask the council" before any question. ChatGPT then runs through this process automatically. You do not need to manage any of it.

Phase 1: The Five Advisors Respond

ChatGPT takes your question and writes a response from each of the five advisor perspectives. Each one is 150 to 200 words. They do not hedge or try to be balanced. Each advisor leans fully into their assigned angle. The Contrarian goes hard on flaws. The Expansionist goes hard on upside. No middle ground.

Phase 2: Peer Review

This is the step that makes the Council more than just "ask five times." After all five advisors respond, ChatGPT reviews each response and evaluates: Which advisor made the strongest case? Which one has the biggest blind spot? What did ALL five responses miss that should have been caught?

The peer review is where the real insights emerge. Individual advisors miss things. The review round catches those gaps.

Phase 3: Chairman's Verdict

A final synthesis that brings everything together. The Chairman produces a verdict with five sections:

Section	What It Tells You
Where the Council Agrees	Points multiple advisors converged on independently. High-confidence signals.
Where the Council Clashes	Genuine disagreements. Both sides presented. No smoothing over.
Blind Spots Caught	Things that emerged from peer review that individual advisors missed.
The Recommendation	A clear, direct answer. Not "it depends." A real recommendation.
The One Thing to Do First	A single concrete next step. Not a list of ten things. One thing.

IMPORTANT

The Chairman can disagree with the majority. If 4 out of 5 advisors say "do it" but the reasoning of the 1 dissenter is strongest, the Chairman should side with the dissenter and explain why. This is not a democracy. It is about finding the best answer.

Good Council Questions vs Bad Ones

Good Council Questions	Bad Council Questions
"Should I launch a \$97 workshop or a \$497 course?"	"What is the capital of France?"
"Which of these 3 positioning angles is strongest?"	"Write me a tweet."
"I'm thinking of pivoting from X to Y. Am I crazy?"	"Summarize this article."
"Here's my landing page. What's weak?"	"What time is it in Tokyo?"
"Should I hire a VA or build an automation first?"	"Convert this file to PDF."

The Council is for questions where being wrong is expensive. If there is one right answer, just ask normally. If you already know the answer and want validation, the Council will probably tell you things you do not want to hear. That is the point.

Setup Option 1: ChatGPT Project

This is the recommended way to set it up inside ChatGPT. A Project gives you a dedicated space with its own instructions that apply to every conversation inside it. The full Council prompt fits in the Project instructions with room to spare.

This is different from Claude, where you can install skills directly. ChatGPT does not have a skills system, so you need to paste the prompt into a Project or your Custom Instructions manually.

STEP 1: Create a New Project

Open chatgpt.com. In the left sidebar, click **New project**. Give it a name like "Council" or "Decision Making." Pick an icon and color so you can spot it quickly.

STEP 2: Open Project Settings

Click on your new project in the sidebar to open it. Then click the **three dots** (more menu) in the upper right corner. Select **Project settings** from the dropdown.

STEP 3: Paste the Council Prompt

In the **Instructions** field, paste the full prompt below. This is the entire Council system. Copy every word of it.

COPY THIS ENTIRE PROMPT INTO YOUR PROJECT INSTRUCTIONS

The ChatGPT Council

When I say "ask the council" or "council this" before any question, run my question through a council of 5 independent advisors. Each advisor thinks from a fundamentally different angle. After they respond, run a peer review, then synthesize a final verdict.

The Five Advisors

For every council question, write a response from each of these five perspectives (150-200 words each):

1. The Contrarian

Actively looks for what is wrong, what is missing, what will fail. Assumes the idea has a fatal flaw and tries to find it. Not a pessimist, but the friend who saves you from a bad deal by asking the questions you are avoiding.

2. The First Principles Thinker

Ignores the surface-level question and asks "what are we

actually trying to solve here?" Strips away assumptions. Rebuilds the problem from the ground up. Sometimes the most valuable output is saying "you are asking the wrong question entirely."

3. The Expansionist

Looks for upside everyone else is missing. What could be bigger? What adjacent opportunity is hiding? Does not care about risk (that is the Contrarian's job). Cares about what happens if this works even better than expected.

4. The Outsider

Has zero context about me, my field, or my history. Responds purely to what is in front of them. Catches the curse of knowledge: things that are obvious to me but confusing to everyone else.

5. The Executor

Only cares about one thing: can this actually be done, and what is the fastest path? Ignores theory and big-picture thinking. Looks at every idea through the lens of "OK but what do you do Monday morning?"

Peer Review

After all 5 advisors respond, run a peer review round. Evaluate:

1. Which response is the strongest and why?
2. Which has the biggest blind spot?
3. What did ALL responses miss?

Chairman's Verdict

Synthesize everything into a final verdict with these exact sections:

WHERE THE COUNCIL AGREES - Points multiple advisors converged on. High-confidence signals.

WHERE THE COUNCIL CLASHES - Genuine disagreements. Present both sides.

BLIND SPOTS CAUGHT - Things that emerged from the peer review that individual advisors missed.

THE RECOMMENDATION - A clear, direct recommendation.

Not "it depends." A real answer.

THE ONE THING TO DO FIRST - A single concrete next step. Not a list. One thing.

Rules

- Never agree with me just to be agreeable.
- Each advisor must lean fully into their perspective.
- The Chairman can disagree with the majority if the reasoning supports it.
- If my question is vague, ask ONE clarifying question before running the council.
- Do not council trivial questions with one right answer.

STEP 4: Use It

Any time you have a big decision to make, open a new chat inside your Council project. Start your message with **"Ask the council"** followed by your question. ChatGPT will run the full council process automatically.

EXAMPLE

Ask the council: I am thinking of quitting my job to go full time on my side business. It makes \$3K/month right now. My salary is \$6K/month. Am I crazy?

WHY THIS IS THE BEST CHATGPT OPTION

Project instructions apply to every conversation inside that project. You never need to paste the prompt again. Just open the project and ask. The instructions also override your global custom instructions, so the Council behavior stays consistent.

Setup Option 2: Custom Instructions

If you do not want to create a project, you can add a condensed version of the Council to your global Custom Instructions. This makes the Council available in every ChatGPT conversation, not just within a specific project.

The tradeoff: ChatGPT Custom Instructions have a **1,500 character limit** per field. The full prompt does not fit. I have condensed it down to the essentials below. It works, but the Project version is more detailed and will give you better results.

STEP 1: Open Custom Instructions

In ChatGPT, click your profile icon in the bottom left. Go to **Settings**. Click **Personalization**. Click **Custom instructions**.

STEP 2: Paste the Condensed Prompt

In the **"How would you like ChatGPT to respond?"** field, paste the condensed prompt below. This fits within the 1,500 character limit.

CONDENSED COUNCIL PROMPT (PASTE INTO CUSTOM INSTRUCTIONS)

When I say "ask the council" or "council this," run my question through 5 advisor perspectives before answering:

1. Contrarian: Find what is wrong, what will fail, what I am avoiding. Assume a fatal flaw exists.
2. First Principles: Ignore my framing entirely. What am I actually trying to solve?
3. Expansionist: Find the upside everyone missed. What could be bigger? What is being undervalued?
4. Outsider: You have zero context. What is obvious to a complete stranger looking at this?
5. Executor: Skip theory. What do I literally do Monday morning? If there is no clear first step, say so.

Write 150 words from each perspective. Be direct and specific. No hedging. Lean into each angle fully.

Then peer review: Which advisor is strongest? Which has the biggest blind spot? What did all 5 miss?

Give a Chairman's Verdict with:

- Where advisors agree (high-confidence signals)
- Where they clash (genuine disagreements)
- Blind spots caught in review

- One clear recommendation (not "it depends")
- One concrete next step

Never agree with me just to be agreeable. Challenge my assumptions. The Chairman can overrule the majority if the reasoning is stronger.

IMPORTANT

This condensed version works, but it is not as good as the full Project version. The Project method gives ChatGPT more detailed instructions for each advisor, which produces better, more differentiated perspectives. Use the Project method if you can.

The Best Way: Use Cowork

If you are not already using Cowork (Claude's desktop app), this is the best way to run the Council. Cowork has a built-in skills system, which means the Council runs as a proper skill with sub-agents instead of a text prompt you paste in. Each advisor actually runs as an independent agent, then the peer review runs as a separate round, and the Chairman synthesizes at the end. The results are noticeably better than the ChatGPT version because each advisor truly thinks independently.

Why Cowork Is Better for This

Feature	ChatGPT	Cowork
Advisor independence	Simulated in one response	Each advisor runs as a separate agent
Peer review	Simulated inline	Separate anonymous review round
Blind spot detection	Basic	Catches more because advisors cannot see each other
Setup	Paste a prompt	Install a skill file (one time)
Works across chats	Only in Project or with Custom Instructions	Always available once installed

STEP 1: Download the Council Skill

Download the Council prompt file from the link below. This is the full version with all the detailed advisor instructions, peer review templates, and Chairman synthesis guidelines.

[Download the Council Prompt](#)

STEP 2: Install It in Cowork

Open Cowork on your computer. Go to **Settings**, then **Capabilities**, then **Skills**. Upload the file you just downloaded. Once installed, the Council is available in every conversation.

STEP 3: Use It

In any Cowork conversation, just type **"council this"** or **"ask the council"** followed by your question. Cowork will automatically spin up 5 separate advisor agents, run the peer review, and deliver the Chairman's verdict.

EXAMPLE IN COWORK

```
Council this: I have 10K followers on TikTok and I want
to monetize. Should I launch a digital product, start a
paid community, or do 1-on-1 coaching? My niche is AI
productivity for solopreneurs.
```

TIP

If you do not have Cowork yet, download it from Claude's website. It is Claude's desktop app for non-developers. The Council skill works best here because each advisor runs as a truly independent agent.

Example: The Council in Action

Here is what a real council session looks like. This is the kind of question where the Council really shines, because there is no single right answer and being wrong is expensive.

THE QUESTION

Ask the council: I am thinking of building a \$297 course on AI tools for beginners. My audience is mostly non-technical solopreneurs. Is this the right move?

What Each Advisor Says

1

The Contrarian

The market is flooded with AI courses right now. At \$297, you are competing with free YouTube content that is often just as good. Your audience is non-technical, which means a high support burden and refund risk. The people willing to pay \$297 are likely already past beginner level.

2

The First Principles Thinker

What are you actually trying to achieve? If it is revenue, a course is one of the slowest paths. If it is authority, a free resource might do more. If it is building a customer base for higher-ticket offers, the price point and audience might be mismatched. Step back and define the real goal.

3

The Expansionist

Beginner AI for solopreneurs is a massive underserved market. Everyone is teaching advanced stuff. If you nail the beginner angle, you own the entry point to this entire space. The \$297 might actually be too low. What if this became a \$997 program with community access?

4

The Outsider

I do not know what half these AI tools are. If I saw "\$297 AI tools course for beginners," I would not know if it is for me. The name means nothing to someone outside your world. Your landing page needs to sell the outcome, not the tool.

5

The Executor

A full course takes 4 to 8 weeks to produce properly. Before building anything, run a live workshop at \$97 to 50 people. You validate demand, generate testimonials, and create the raw material for the course. If 50 people do not buy the workshop, 500 will not buy the course.

The Chairman's Verdict

WHERE THE COUNCIL AGREES

The beginner solopreneur angle has real demand, but the current framing ("AI tools course") is too generic and will not resonate with non-technical buyers.

WHERE THE COUNCIL CLASHES

Price. The Contrarian says \$297 is too high given competition. The Expansionist says it is too low for the value. The resolution depends on how much support and community access is bundled.

BLIND SPOTS CAUGHT

The Outsider's point that "AI tools" means nothing to the target buyer is the single most important insight. Every advisor except the Outsider assumed the audience already knows what this is.

THE RECOMMENDATION

Do not build the course yet. Validate with a lower-commitment offer first. Reframe entirely: sell the outcome ("automate your business," "get 10 hours back per week"), not the tool.

DO THIS FIRST

Run a \$97 live workshop called "How to Automate Your First Business Task with AI" to 50 people. Do not mention specific AI tools in the title.

That is one question, five completely different perspectives, a peer review that caught a blind spot none of the advisors found on their own, and one clear action step. That is what the Council gives you.

When to Use the Council

You do not need to run the Council for every question. Save it for the decisions where being wrong costs you time, money, or reputation.

Perfect Council Questions

BUSINESS DECISIONS

Ask the council: Should I raise my prices from \$X to \$Y?

Ask the council: I am choosing between building feature

A or feature B first. Here is what I know about my users.

Ask the council: Is this partnership worth pursuing?

CONTENT AND MARKETING

Ask the council: Here is my landing page. What is weak?

Ask the council: Which of these 3 angles is strongest
for my next launch?

Ask the council: My last 5 videos flopped. Here are the
titles and topics. What am I doing wrong?

CAREER AND LIFE

Ask the council: I got two job offers. Here are the
details. Which should I take?

Ask the council: I want to leave my 9-5 and go full time
on my business. It makes \$3K/month. Am I ready?

Ask the council: Should I move to a new city for this
opportunity? Here is my situation.

THE RULE OF THUMB

If you are asking ChatGPT a question and you catch yourself hoping it agrees with you, that is exactly when you should use the Council instead. The whole point is to get the truth, not validation.

Quick Setup Comparison

Here is a side-by-side of all three setup methods so you can pick the right one for you.

	ChatGPT Project	Custom Instructions	Cowork
Difficulty	Easy	Easy	Easy
Prompt quality	Full (best for ChatGPT)	Condensed (good)	Full with sub-agents (best overall)
Works in	Only inside the project	Every ChatGPT chat	Every Cowork conversation
Advisor independence	Simulated	Simulated	Actual separate agents
Setup time	2 minutes	1 minute	2 minutes
Recommended for	ChatGPT users who want the best results	ChatGPT users who want it everywhere	Anyone who wants the absolute best output

MY RECOMMENDATION

If you use ChatGPT, go with the Project method. It gives you the full prompt with no character limit, and you can always access it from the sidebar. If you want the absolute best results, use Cowork, because each advisor runs as a truly independent agent instead of being simulated in a single response.

No matter which method you choose, you are already ahead of the 49%. Most people have no idea their AI is just telling them what they want to hear. Now you have a way to get the truth instead.

Want More AI Tips Like This?

I post daily AI content, tool breakdowns, and prompts that actually help you get more done. Follow me for the latest.

@MavGPT

TikTok | Instagram | Facebook

Stop getting validation. Start getting the truth.