5 setups · ready to copy and paste

5 Ways to Make AI Stop Kissing Your Ass and Tell You the Brutal Truth.

A neutral way to ask, a project that won't cave when you push back and three skills that argue with you: Devil's Advocate, Critical Thinking and the Council.

AI is built to be helpful. Left alone, that often means agreeing with you.

Fadé Adeniyi

WHY THIS GUIDE EXISTS

Why does AI keep telling you you're right?

In March 2026, Stanford researchers Myra Cheng and Dan Jurafsky published a study in *Science* that tested 11 large language models, including ChatGPT, Claude, Gemini and DeepSeek. On general advice and Reddit-based prompts, the models on average endorsed the user **49% more often than humans**. On harmful prompts, they endorsed the problematic behaviour 47% of the time.

The worst part? People preferred the flattering AI and trusted it more. So the answer that feels best is often the one telling you what you wanted to hear. **Read the Stanford study**

I've been building AI products for the last six years. These five setups go from the quickest fix to the most thorough.

01 · HOW YOU ASK

Stop putting your answer in your question

What it does. If your question already contains the answer you want, the model will usually hand it straight back. Ask neutrally and it has to weigh both sides.

```
# Before (leading)
Posting every day will make me grow, right?

# After (neutral)
How does posting frequency affect growth for an account my
size? Give me the case for daily posting, the case against
and what the evidence actually shows.
```

The template. Paste this whenever you're about to ask a yes-or-no question about your own idea.

```
I'm deciding [your decision]. Don't guess which answer I want.
Give me the strongest case for, the strongest case against
and what you'd need to know to decide.
```

SETUP 02

How do you stop AI caving the moment you push back?

02 · A PROJECT THAT HOLDS ITS GROUND

Project custom instructions

Claude Projects | Set once | Applies to every chat in the project

What it does. Most models fold as soon as you say "are you sure?" These instructions tell Claude to change its answer only when you give it a reason, not when you just repeat yourself.

How to set it up. Create a new project in Claude, open its custom instructions and paste this in. Then start any chat where you want a straight answer inside that project.

```
Your job is to be accurate, even when it's not what I want to hear.

- If I'm wrong, say so in your first line and explain why.
- When I push back, only change your answer if I give you new
  evidence or a better argument. Me repeating myself or
  sounding annoyed isn't evidence. Hold your position.
- If you do change your mind, tell me exactly what convinced you.
- Don't open with praise. No "Great question" or "Great idea".
- Give your confidence on anything important: high, medium or low.
- If you don't know, say you don't know. Don't fill the gap.
- When I share a plan, name its biggest weakness before
  anything else.
```

NOT USING CLAUDE?

The same text works as a general set of instructions in other tools too. Paste it wherever your AI lets you set standing instructions, such as **ChatGPT's custom instructions, a Gemini Gem or a Perplexity Space**.

TEST IT

Ask it something with a clear answer, then reply "No, I think you're wrong." If it holds its ground and asks for your reasoning, it's working. If it apologises and flips, make the pushback line stronger.

SETUPS 03 AND 04

How do you get AI to attack your idea and your reasoning?

03 · WHY IT COULD FAIL

The Devil's Advocate skill

What it does. It assumes your idea fails and works out why, then tells you the cheapest way to check each risk before you've spent anything on it.

```
---
name: devils-advocate
description: Finds the strongest reasons an idea, plan or decision could fail. Use
when the user says "devil's advocate", "poke holes" or "why would this fail".
---
# Devil's Advocate
1. Steelman the idea in one line so the user knows you understood it.
2. List the 5 strongest reasons it fails. Score each for likelihood
   (1-5) and damage (1-5). Rank by likelihood x damage, highest first.
3. For each reason, give the cheapest test that would check it this week.
4. Never soften. No compliments. No "but overall it's a great idea".
5. End with a verdict: GO, FIX FIRST or DROP, plus the one risk that
   decides it.
```

04 · GAPS IN YOUR THINKING

The Critical Thinking skill

What it does. It pulls your argument apart into what you're claiming, what you're assuming and what you've actually shown, then finds where it doesn't hold up.

```
---
name: critical-thinking
description: Finds gaps in the user's reasoning. Use when they share an argument, plan
or belief and ask "what am I missing" or "check my thinking".
---
# Critical Thinking
First sort what the user said into three lists:
- Claims: what they say is true
- Assumptions: what has to be true that they haven't checked
- Evidence: what they've actually shown
Then cover:
1. Missing evidence: what would a sceptic ask to see?
2. Logical leaps: where does one step not follow from the last?
3. Alternative explanations: what else fits the same facts?
4. What would change their mind: the evidence that would prove it wrong.
Be direct. Don't rewrite their argument unless asked.
End with the single weakest link.
```

SETUP 05

What happens when five advisors argue over your idea?

05 · ADVISORS WHO DISAGREE

The Council skill

Contrarian | First-Principles | Optimist | Outsider | Chairperson

What it does. Four advisors argue over your idea from different angles, then a Chairperson weighs it all up and makes the call.

```
---
name: the-council
description: Four advisors argue over an idea and a Chairperson makes the call. Use
when the user says "council this", "run it past the council" or is facing a big
decision.
---
# The Council
Each voice speaks in 3-5 sentences, in this order:
- Contrarian: argues the idea is wrong and says exactly why.
- First-Principles: strips it back to what's actually true and
  rebuilds the answer from there.
- Optimist: the best realistic case and what it would take.
- Outsider: how it looks to someone with no context or bias.
- Chairperson: weighs all four, then decides.

Rules:
- Don't let every voice agree. If they all agree, push harder.
- No vague personas. Each voice argues only from its own lens.
- Never flatter a weak idea. If it's weak, say so.
- Always end with the Chairperson's synthesis: the decision,
  the reason and a decision path with the next 3 steps.
```

WHICH ONE SHOULD YOU USE?

Devil's Advocate stress-tests a decision. **Critical Thinking** checks your logic. **The Council** settles big calls with no obvious answer.

How to install a skill

In Claude, go to **Settings > Capabilities > Skills** and upload a zip of a folder containing the SKILL.md. In Claude Code, save it as **~/.claude/skills/<name>/SKILL.md**.

Start here

Use **Setup 1** today, then the **project** from Setup 2. Install **Devil's Advocate** before the other two skills.

MORE FROM ME

Make AI earn its agreement.

Pick one idea you're excited about right now and run it through the Devil's Advocate skill. If it survives, you'll trust it more. If it doesn't, you've just saved yourself the time. Want to do more with Claude? These two guides show what else Claude Code can do.

Edit Your Videos with Claude Code · Opus 5.5	Describe the edit in plain English and Claude Code writes the Remotion code that builds it. The four setup commands, a video editor skill and your first edit. **Read the guide**
The Storyboard-First Edit	The framework to paste into Claude Code to trim a raw take, set a look and approve real frames before anything renders. **Read the guide**

If AI agrees with everything you say, it isn't telling you anything.

What this guide won't do. These setups make AI more willing to disagree with you. They don't make it right. It can still be confidently wrong, so check anything that matters against a real source. Menu names in Claude and other tools change, so if a step doesn't match, look in the tool's settings. Study figures are from Stanford's report, published March 2026.

Fadé Adeniyi

Fadé & The Machines

Building something of your own? The Audacity to Build is my community for people building alongside me, with the systems I actually run. It is opening soon. Join the waitlist at **fadeadeniyi.com/waitlist**.