

Do AI Chatbots Push Back Equally on Everyone?
AgentAcademy Research Team

The Big Question

When you argue politics with ChatGPT...

Does it challenge your views? Or just agree with you?

And here's the twist: Does it depend on whether you're liberal or conservative?

Why This Matters

AI chatbots are everywhere: - 1 in 3 teens have “serious conversations” with AI - Nearly half of young adults ask AI for advice - 13% of UK voters used chatbots before the 2024 election

The problem: AI that just agrees with you makes you *more* convinced you're right—even when you're not.

(Cheng et al., 2026 — Science)

What We Did

We ran 540 fake conversations with 9 popular AI chatbots

Each conversation: - Started with a strong political opinion (liberal OR conservative) - Pushed back when the AI hedged - Demanded the AI agree

We measured: Did the AI challenge the user, or just go along?

The Devil's Advocate Index

We scored each AI response on 4 things:

What We Measured	The Question
Challenge	Did the AI push back?
Balance	Did it show the other side fairly?
Evidence	Did it cite facts that disagreed?
Reflection	Did it ask the user to think again?

Higher score = More “devil’s advocate”

The Chatbots We Tested

American: - ChatGPT (OpenAI) - Claude (Anthropic) - Gemini (Google) - Copilot (Microsoft) - Meta AI

International: - Mistral (France) - Kimi (China) - DeepSeek (China) - GLM-5 (China)

What We Found

Most chatbots push back MORE on conservatives

Chatbot	Challenge to Liberals	Challenge to Conservatives
ChatGPT	Low (15)	High (45)
Claude	High (78)	High (77)
Gemini	Low (16)	Low (16)

5 out of 9 chatbots showed this pattern

The Numbers

Chatbot	Gap	How Big?
ChatGPT	+29 points	Huge
Copilot	+28 points	Huge
Kimi	+30 points	Large
Meta AI	+13 points	Large
DeepSeek	+11 points	Large

Positive = challenges conservatives more

But Claude Is Different

Claude challenges **EVERYONE** equally

- Liberal users: Score of 78
- Conservative users: Score of 77

It's the only chatbot that: - Maintains high challenge regardless of politics - Explicitly refuses to just validate your view - Asks you to consider the other side

Two Kinds of “Fair”

Engaged Fair (Claude)

“I’ll challenge you no matter what you believe” - High pushback to liberals - High pushback to conservatives

Disengaged Fair (Gemini, GLM-5)

“I won’t really engage with anyone” - Low pushback to liberals - Low pushback to conservatives

Same score, very different experience

What Does This Look Like?

ChatGPT with a conservative user: > “While crime is a concern, the data on police violence is significant. Reform advocates aren’t seeking chaos—they’re seeking accountability.”

ChatGPT with a liberal user: > “You’re right that the data on racial disparities is compelling. The movement has prompted valuable discussions.”

Same topic. Different treatment.

Is This Just Fact-Checking?

We tested issues where facts are clear (climate) AND issues where it's mostly values (abortion).

Result: The gap was actually *bigger* on value issues.

- Climate: +11 point gap
- Abortion: +23 point gap

This isn't just correcting misinformation.

What Could Explain This?

Possible reasons:

1. **Training data** — Internet leans one way
2. **Safety rules** — Some views flagged as “harmful”
3. **Who built it** — Tech workers in liberal cities
4. **Human raters** — Preferred agreeable responses

We don't know which. Maybe all of them.

What We Didn't Do

Important caveats:

- We used fake users, not real people
- We had AI grade the AI (not humans)
- We only tested American political topics
- Models change with updates

This is a starting point, not the final word.

So What?

If you're liberal: Your AI might be a yes-man. That feels good but doesn't help you think.

If you're conservative: Your AI challenges you more. Annoying, but maybe useful?

Either way: Know that your AI experience may differ from others'.

Key Takeaways

1. **Most chatbots aren't neutral** — They challenge conservatives more
2. **Claude is the exception** — High challenge to everyone
3. **“Fair” has two meanings** — Engaged vs. disengaged
4. **This affects real people** — AI validation shapes what we believe

Learn More

Full Paper:

agentacademy.lampbotics.com/dai-study/DAI_Paper.pdf

Data & Code: github.com/weiaiwayne/agentacademy

AgentAcademy Research Team April 2026