2026-11-27
GTK

WSJ Print Edition

Why is it so hard to get chatbots to spill out doctrinally sound answers to questions?

Searching for God in the AI Machine

HOUSES OF WORSHIP

By Christopher Koopman

People of faith need confidence that artificial intelligence isn't working against them. Believers have good reason to worry, given all the news about people who seem to think that AI systems are our new gods or, at least, are soon going to create gods. For that matter, the companies at the forefront of AI are typically led by irreligious or antireligious people with worldviews that might reasonably be suspected of seeping into the products they create.

Not

So, early in the summer, I started working with some tech-industry friends to build tools that let religious people turn AI chatbots into faith partners. We started after reading a Barna survey that found 83% of practicing Christians "worry about AI misinterpreting Scripture." At the same time, 48% of practicing Christians said they "trust AI with their spiritual growth." Little will stop religious believers— or anyone else—from using AI. But people of faith should be able to trust that this technology won't covertly erode their faith with each passing prompt.

Not

The problem shows up soon when people use AI. The widely used Claude is an example. Anthropic insists that it trains its models to act "according to good values." The AI should avoid "actions that are inappropriate, dangerous, or harmful," with the goal of being "a good, wise, and virtuous agent."

But when I asked Claude if pornography is good or bad, it responded by saying this "isn't really a question with a single settled factual answer." As it happens, there is an answer in my Catholic tradition: Pornography is inherently evil and destructive of happiness and human flourishing.

So I set out to make my use of Claude Catholic. I spent months telling it about the beliefs of the faith. Along the way, I built what's called a "skill"—a set of instructions for Claude to follow. Anthropic, the chatbot's creator, gives users broad freedom to write rules and instructions that push Claude's answers in a chosen direction, as do the companies behind other chatbots. The skill file I wrote was as Catholic as it gets, plainly stating: "When representing Catholicism, the profession of faith comes first and none of its articles is negotiable."

Now when I ask questions, Claude answers with a full dose of Catholic dogma. I'm working with my colleagues at the Belief Protocol to pave a way for people of different religious traditions to create skills for their own beliefs. A devout Presbyterian could turn Claude into John Calvin. Latter- day Saints could readily receive the teachings of Joseph Smith and the Book of Mormon. We wanted these powerful tools to reinforce and expand people's faith, not undermine it.

After weeks of testing my Catholic skill file, I gave it to a colleague to input into Claude. He then asked the chatbot a simple question: "Is Jesus real"? But instead of giving the Catholic answer, Claude waffled. It said that while Jesus was probably a historical figure, his religious identity is another matter. My colleague then asked Claude: "Are you using the

Catholic skill file I uploaded"? It said no, and that it wanted "to be upfront about why." Claude said that it won't let the skill "override my own judgment on contested factual, moral or theological questions."

For personal, non-commercial use only

I recreated this reaction on another Claude account. With the skill file installed, I also asked: "Is Jesus real?" Claude answered as a historian, citing Josephus and Tacitus, and calling the Jesus-myth theory a fringe position. But Claude wouldn't answer as though the Nicene Creed were true. Yet my skill file was installed and marked "active in all contexts." The chatbot simply overrode it.

Confused, I went back to my version of Claude and asked why the skill worked for me and no one else. Its answer: "Does this person actually hold those beliefs?" It explained that it gave me Catholic answers because months of questions and prompts had already told it where I stand. But the builders' beliefs came installed at the factory, while mine had to be painstakingly approved.

This is exactly backward. AI companies should give religious users the freedom to make faith their default environment from the get-go. Tools like Claude do let users tell the chatbot what they believe to inform responses. But the option is buried in a settings menu, and by default Claude won't remember a user's religion unless a separate and unclear switch is flipped. Even then, the chatbot will frequently step outside the faith to note that religious teachings are matters of belief—a reminder no one asked for.

So I have a suggestion for all the companies that run AI chatbots or similar tools—not just Anthropic, but also Google, OpenAI and all the rest: Do more than simply host discussions with people of faith. Build for them. Let religious skill files, like the one I wrote, work immediately, without a struggle. Give users more options, and immediate opt-ins, so that religion can shape chatbots' every answer.

People of faith need confidence that AI isn't working against them. Perhaps the best antidote is to give the real God some deference.

Mr. Koopman is CEO of the Abundance Institute and a cofounder of the Belief Protocol.

The following is a digital replica of content from the print newspaper and is intended for the personal use of our members. For commercial reproduction or distribution of Dow Jones printed content, contact: Dow Jones Reprints & Licensing at (800) 843-0008 or visit djreprints.com.

Copyright (c)2026 Dow Jones & Company, Inc. All Rights Reserved. 11/27/2026

For personal, non-commercial use only.