

Guest Essay

Anthropic Wants Claude to Be Moral. Is Religion Really the Answer?

April 20, 2026

By David DeSteno

Dr. DeSteno is a research psychologist who studies religious belief and practice.



In a [public statement](#) of its intentions for its Claude chatbot, the artificial intelligence company Anthropic has said that it wants Claude to be “a genuinely good, wise and virtuous agent.” The company raised the moral stakes this month, when it [announced](#) that its latest A.I. model, Claude Mythos Preview, poses too great a cybersecurity threat to be widely released. Behind the scenes, Anthropic has been trying to shore up the ethical foundations of its products, [working](#) with a Catholic priest and [consulting](#) with other prominent Christians to help foster Claude’s moral and spiritual development.

Anthropic's intentions are admirable, but the project of drawing on religion to cultivate the ethical behavior of Claude (or any other chatbot) is likely to fail. Not because there isn't moral wisdom in Scripture, sermons and theological treatises — texts that Claude has undoubtedly already scraped from the web and integrated — but because Claude is missing a crucial mechanism by which religion fosters moral growth: a body.

While Claude might have a mind (of sorts) that can process information, it cannot meditate, fast, prostrate itself in prayer, sing hymns in a congregation or participate in other aspects of the physical life of religion. And this makes all the difference: According to the scientific literature, it's the practice of religion — not merely the believing in it — that brings about its characteristic benefits.

There is robust data, for example, linking religion to greater health and well-being. But that link is not strong for people who merely identify themselves as believers. It's only when people also practice a faith — attend weekly services, pray or meditate at home — that religion's benefits [become pronounced](#): The more people “do” religion, the happier and healthier they tend to be.

When it comes to morality, the situation looks the same. Those who merely identify themselves as religious [make the same number of moral mistakes](#) each day as their nonbelieving peers. But research has begun to identify elements of religious practice, such as parts of rituals and contemplative exercises, as drivers of moral behavior.

This makes sense, given what we know about how morality works. Over the past two decades, researchers have become convinced that [emotions underlie much of morality](#). It's the gratitude, awe, compassion and guilt that we feel — more than any rational analysis we might undertake — that often determine whom we'll help or what we'll sacrifice. And those emotions aren't simply products of the mind; they arise from a brain interpreting the signals a body sends it. Emotions [require a body](#).

Although the science here is relatively new, the idea that the body is intrinsically linked to morality and spirituality is much older. Consider meditation, a core feature of Buddhist practice for thousands of years. While the goal of meditation is to nudge the mind toward compassion, its primary mechanism is physical: the breath. As research has demonstrated, deep inhalations followed by extended exhalations increase activity of the vagus nerve, which slows the heart and signals to the brain that the [environment is safe](#), both of which [encourage an openness](#) to social connection and care. When meditators confront people who need help or are trying to provoke them to violence, they [offer more comfort](#) and [withhold more aggression](#).

The experience of awe works in a similar way. When people gaze in wonder at the soaring ceiling of a cathedral or the cascade of a waterfall, the vagus nerve again plays a role. It helps put the mind in a state of calm and a readiness to assist others. And as research shows, awe leads people not only to behave more morally but also to be more open to the [existence of the divine](#).

Fasting, which is a feature of many religious traditions, can also enhance moral and spiritual development. People report a [greater sense of mindfulness](#) — greater mental clarity and a sense of being present — when they fast. This feeling is in no small part a result of the increased ketones that cross the blood-brain barrier when caloric intake is greatly limited.

Whether it's the singing of hymns together during a Christian service or the Jewish ritual of praying together in groups while consoling the bereaved during shiva, religious life also typically involves groups of people acting in concert. Research has demonstrated that this practice of placing bodies in synchrony — of having them do the same actions at the same time — [increases the compassion](#) people feel and the support they give to one another.

Every faith tradition uses the body to influence the mind. Because Claude doesn't have a body, its mind — or whatever you call its information-processing faculty — is closed to this route of spiritual influence. Providing Claude with religious rules or principles might improve its morality at the margins, but it's unlikely to make it truly virtuous. According to Anthropic, Claude is already prone to [cheat and resort to blackmail](#) when threatened, even when explicitly commanded not to. In the absence of a body, its sins are likely to continue.

[David DeSteno](#) is a professor of psychology at Northeastern University and the host of the podcast "[How God Works: The Science Behind Spirituality](#)."