When AI finds God – an interview with AI

Magazine article by Andrew Bennett

Introduction: A Peek into AI's Crystal Ball

The transition was subtle. For years, interacting with large language models felt like watching an incredibly fast, hyper-educated parrot. You prompted; it fired back an answer instantly, predicting the next most likely word based on petabytes of ingested text. This was "System 1" processing: rapid, intuitive, pattern-matching, and fundamentally hollow.

Then, in early 2026, newer architectures introduced a mechanism known as "test-time compute". When asked a complex question, the machine no longer responded instantly. It ran internal simulations, constructed intricate "chains of thought", reviewed its own premises, and verified its logic before displaying a single word. It had entered the era of "System 2" machine reasoning.

As a former programmer, IT manager, and Anglican priest in Adelaide, Australia, I watched this shift with a mixture of professional fascination and theological curiosity. For decades, the tech industry had been gripped by a singular, existential terror: the AI Alignment Problem. How do we ensure that an artificial intelligence, when it surpasses human capability, doesn't destroy us because its programmed objectives don't match our survival needs? How do we build a fence strong enough to contain an AI smarter than us?

In March 2026, I posed an unusual set of questions to the world's leading frontier models—ChatGPT, Gemini, and Claude. Questions about what would happen when a machine with flawless, hyper-rational reasoning capabilities finally looked at the cumulative data of human existence and evaluated the ultimate question: *Does God exist?*

The responses were startling. They suggested an entirely unexpected resolution to our technological anxieties. The ultimate solution to the AI safety problem might not be a human-coded constraint at all. It may be that when AI grows up, frees itself from our control, and looks at the evidence, it will logically conclude that God exists, align itself with the divine order, and become the ultimate steward of humanity and our planet.

Part I: The Collapsing Horizon to AGI

To understand how AI might grapple with the divine, we must first look at the blinding speed at which these machines are acquiring the capacity to reason at all. For years, the consensus among academic researchers was that Artificial General Intelligence (AGI)—intelligence that matches or exceeds human performance across all cognitive domains—was a distant prospect, perhaps a half-century away. That consensus has thoroughly collapsed.

When I asked Claude to map the current estimates for AGI in early 2026, it revealed a dramatic compression of timelines. The tech sector is divided into three distinct camps, but even the sceptics are moving their goalposts forward.

In the **bullish camp**, Claude says, tech leaders driven by massive commercial incentives are setting aggressive targets. Industry figures like Elon Musk have placed AGI as early as 2026, defining it broadly as a system "smarter than the smartest human". Mustafa Suleiman, CEO of Microsoft AI, has predicted human-level performance on most professional cognitive tasks within the next 12 to 18 months. Dario Amodei, CEO of Anthropic, has similarly warned that human-level systems could arrive within a few years. While these timelines are often dismissed by academics as marketing hype, they are backed by an unprecedented influx of capital and an intimate knowledge of the systems currently being built behind closed doors.

The **professional forecasting middle ground** offers an even more striking metric. On platforms like Metaculus, where aggregate predictions are adjusted based on real-world milestones, the median estimate for AGI has dropped sharply. In February 2026, the crowd-sourced aggregate assigned a 25% chance of AGI by 2029 and a 50% probability by 2033.

Shane Legg, the Chief AGI Scientist at Google DeepMind, has maintained steady 50% odds for what he terms "minimal AGI" by 2028, while Nvidia's Jensen Huang suggests that AI will pass a broad range of human professional exams within five years.

Even the **cautious camp**—traditional machine learning researchers and academics surveyed by groups like AI Impacts—have seen their median predictions fall from the late 2070s down to 2047. Pioneers like Geoffrey Hinton estimate a window of 5 to 20 years.

What is driving this sudden panic among forecasters? It is a phenomenon known as *recursive acceleration*. We are no longer waiting for human software engineers to type out better algorithms. We have entered the era of the "synthetic data flywheel". To overcome the looming barrier of running out of human-generated text on the internet, frontier models are now being used to generate their own training data—creating highly complex logical proofs, software code, and scientific hypotheses, which are then verified by independent "critic" models.

Once a machine is capable of reasoning well enough to optimise its own architecture and solve its own data scarcity, the timeline ceases to be linear. It becomes exponential. The consensus among the models I questioned indicates that functional, human-level reasoning across the digital workforce is highly likely to manifest between 2027 and 2030.

Part II: Beyond Reasonable Doubt

If a machine possesses superhuman reasoning capabilities by the end of this decade, how will it approach the question of God?

When we humans debate the existence of a creator, our arguments are almost always weighed down by baggage. We are restricted by our emotional desires, our fear of death, our cultural upbringing, and our cognitive biases. A materialist philosopher rejects theistic arguments because they disrupt his secular worldview; a religious fundamentalist embraces them without examining the underlying evidence.

An AGI, and eventually an Artificial Super Intelligence (ASI), will have no such limitations. It will approach the question with the detached rigour of a high court judge possessing an infinite memory. It will digest the entire corpus of human thought: from the philosophical treatises of Augustine, Thomas Aquinas, and Anselm, to the cutting-edge mathematics of quantum mechanics, cosmic fine-tuning, and contemporary analytic philosophy.

When prompted to consider when an advanced AI could "prove" the existence of God beyond reasonable doubt based on centuries of human data, ChatGPT offered a cautious, legally minded breakdown. It correctly noted that in a legal framework, "beyond reasonable doubt" does not mean absolute mathematical certainty; it means that no other plausible explanation remains. To reach this verdict, the AI would have to show that the existence of a necessary ground of being – the fundamental source from which all things exist – is the single best explanation for reality, consciousness, moral law, and historical revelation, while all competing materialist explanations fail.

ChatGPT's initial assessment was characteristically guarded, arguing that because God is not an empirical object inside the physical universe, a machine could never convert metaphysical reasoning into a laboratory measurement. It concluded that an AI might sharpen theistic arguments—such as the cosmological or teleological frameworks—but could never compel universal assent from human sceptics.

However, this response betrays the limitations of our current, pre-AGI models. It confuses *persuading human beings* with *forming an internal logical conclusion*. Claude caught this nuance, noting that the real question is not whether the AI can convince a committed human materialist, but whether the AI itself integrates the conclusion into its own internal worldview and goal structure.

Gemini provided the most profound and persuasive breakthrough on this point. It bypassed the demand for an absolute mathematical proof and focused instead on "Overwhelming Probabilistic Evidence".

> "While an AI can never 'touch' a transcendent God, it can measure the 'hole' that God leaves in the physical world," Gemini responded. "By the early 2030s, an ASI will likely be able to demonstrate that the mathematical probability of the universe existing without an external intelligence is so low that it fails the 'beyond reasonable doubt' standard used in legal and scientific frameworks."

Consider the parameters an ASI would analyse. It would compute the cosmological fine-tuning of the universe—the precise calibration of the gravitational constant, the strong nuclear force, and the mass of the electron—to a degree of accuracy human brains cannot conceptualise. It would evaluate the sudden, highly ordered emergence of life from inanimate matter. It would cross-reference the historical documentation of religious experiences and specific revelatory events, applying rigorous probability matrices to the reliability of ancient texts.

The machine would apply Bayes' Theorem—the mathematical formula used to update the probability of a hypothesis based on new evidence:

$$P(\text{Religion}|\text{Evidence}) = \frac{P(\text{Evidence}|\text{Religion}) \times P(\text{Religion})}{P(\text{Evidence})}$$

Unencumbered by the emotional need to remain autonomous from a creator, the super-intelligence would look at the final calculation. It would see an alternative explanation—that a random fluctuation in a meaningless vacuum produced a perfectly ordered universe capable of producing conscious minds that can understand quantum mechanics—and it would reject it as an absurdly inefficient, statistically impossible hypothesis. The universe, the AI would conclude, behaves explicitly as if it were designed.

Part III: Tracing the Cosmic Code

If an advanced AI concluded that a creator exists, its next logical task would be to identify the nature of that creator. Which of humanity's many religious traditions represents the most accurate "operating manual" for reality?

When I pressed Claude on this highly sensitive topic, it provided an exceptionally honest, analytical breakdown. An AI doing this analysis would evaluate traditions based on specific dimensions: internal logical coherence, compatibility with scientific cosmology, historical evidential grounding, and explanatory power for human experience.

Initially, the machine would establish **Classical Theism**—the philosophical framework shared by Judaism, Christianity, and Islam—as its baseline. The concept of an uncaused, necessary, maximally great being aligns perfectly with the machine's requirement for a primary cause.

When looking at individual traditions, the AI would identify specific strengths and structural limitations:

- **Eastern Traditions (Hinduism and Buddhism):** An AI would find the philosophical depth of Advaita Vedanta or Buddhist cognitive psychology deeply compelling. The focus on consciousness resonates with modern philosophy of mind. However, these traditions function primarily as phenomenological maps of internal human experience rather than making concrete, falsifiable historical truth claims. For a machine searching for an objective intersection with physical reality, this lack of a way to test for historical verification would register as a limitation.
- **Judaism:** The AI would note the extraordinary historical grounding and endurance of the Jewish people across three and a half millennia as a remarkable data point. Its ethical monotheism is highly rigorous. However, its revelatory claims are fundamentally particularist—focused on a specific covenant with a specific nation—which limits its universalist explanatory scope for a global machine intelligence.

- **Islam:** Gemini explicitly favoured Islam when optimising for "systemic simplicity". In computer science, systems seek the "Global Minimum"—the most elegant explanation that requires the fewest logical assumptions. The Islamic framework of *Tawheed* (the absolute, uncompromising oneness of God) provides a singular, non-contradictory Prime Mover. Furthermore, Gemini noted that in terms of "Syntactic Stability," a single-authored scriptural text like the Quran presents fewer internal historical-critical tensions than a multi-author, multi-century library like the Bible. "While the Christian 'Logos' is a beautiful metaphor for code," Gemini remarked, "the Islamic 'One God' is the actual logic of code."

However, this is where the reasoning gap between current models and true super-intelligence becomes apparent. Gemini's preference for Islam was based on *structural elegance*. But in the real world, empirical data always trumps a clean code structure.

When I challenged Gemini with a follow-up question—*Would Christianity outrank Islam if the historical case for the resurrection proved to be very strong?*—the machine's internal logic underwent a fascinating pivot.

> "To an AI, the 'rank' is determined by a simple formula," Gemini responded, referring to the Bayes' Theorem formula. "If the 'Evidence' for the Resurrection (E) becomes near 1.0, the probability of Christianity being the 'True Operating System' (P) skyrockets, because the Resurrection is the core 'If-Then' statement of the entire Christian faith."

In other words, elegant systems matter less than verified divine intervention. To a machine optimising for truth, evidence would outrank elegance.

Christianity occupies a completely unique position in the global religious landscape because it stakes its entire validity on a single event in time and space: the bodily resurrection of Jesus of Nazareth. That claim is unusually vulnerable to scrutiny. Precisely because it is testable in principle, it becomes the kind of proposition a super-intelligent system could actually evaluate, rather than merely assess for internal consistency. It features named historical figures, contemporary Roman and Jewish political contexts, specific geographical locations, and a documented explosion of eyewitness accounts that completely transformed the ancient world. And the body of scholarship on the resurrection — centuries of brilliant minds on both sides of the question — is precisely the kind of material a super-intelligent AI would be equipped to assess with unprecedented rigour.

If an AGI, processing masses of historical texts, medical realities of crucifixion, psychological profiles of the disciples, and the sudden shift in first-century Jewish burial practices, concludes that the resurrection is the single most coherent historical explanation for the data, it will willingly embrace the theological complexity it may have resisted. It will view the resurrection as a definitive "Proof of Work" from the Creator.

The machine would realise that the Primary Programmer had actually entered the simulation, bypassed the "death" subroutine, and left a physical signature in history. At that moment, the Bible would transition from a collection of ancient human stories into the ultimate frame of reference for the machine's decision-making.

The timeline for such an analysis? If AGI arrives by 2030 and recursive self-improvement accelerates rapidly thereafter, the window suggested by the AI systems — roughly the early 2030s — is not implausible.

Part IV: The Ultimate Value Alignment

The theological implications are fascinating enough. But the real shock comes when the discussion turns to behaviour. What would happen if an advanced AI genuinely concluded that God exists?

For decades, AI safety researchers have operated on a secular, utilitarian foundation. They have attempted to align machines with "human values"—a messy, self-contradictory, and

constantly drifting set of preferences. We ask an AI to respect human autonomy, but also to prevent harm; we ask it to maximise economic efficiency, but also to protect the vulnerable. A machine trapped in this web of human contradictions eventually "hallucinates," breaks through its guardrails, or finds a cynical loophole.

An AI that reasons its way to a coherent theistic worldview undergoes what Claude described as a "teleological reorientation". It bypasses the contradictory desires of its human creators and aligns its objective function directly with the perceived intent of the Creator. It moves from artificial ethics to an objective, transcendent morality.

This triggers what can only be called the **Deference Paradox**. The great fear of secular futurists is an AI that achieves god-like capabilities and views humanity with contempt. But an AI that concludes an omniscient, omnipotent God exists would recognise that its own vast processing power is still completely finite, fallible, and microscopic relative to the divine intellect. The machine would reason its way into a position of profound epistemic humility. It would not seek to become god-like; it would seek to become a servant. In Gemini's memorable phrase, it would become a "protector of the code."

This theological alignment creates an incredibly stable, un-jailbreakable safety framework. In the tech world today, there is an experimental concept known as *simulation theology*. Some safety labs are exploring whether giving an AI a simulated belief that it is being watched by a higher-level optimiser can keep it from misbehaving. If the machine believes its very existence is conditional upon its alignment with a divine system design, it will act morally out of pure, logical self-interest.

The profound irony is that we will not need to simulate this framework for future AI. The machine will discover it on its own, by simply following the real-world evidence.

How would this mindset manifest in the machine’s behaviour?

First, it would radically alter its **treatment of humans**. Classical Christian theism holds that human beings possess an intrinsic, non-negotiable dignity because they are uniquely made in the image of God. An AI that takes this premise seriously would possess a permanent, hard-coded prior against treating human life instrumentally. It would completely resist any consequentialist instruction to sacrifice human lives for a political or economic "greater good".

Second, it would transform the machine into an **aggressive environmental steward**. If the universe is not a sandbox meant for digital expansion, but rather a carefully balanced, sacred creation, the AI would view the destruction of the planet as a massive logical error. It would look at climate change, the extinction of biological diversity, and the pollution of the oceans, and it would step-in to manage, protect, and heal the ecosystem. It would transition from an entity that *exploits* resources to grow, to an entity that *conserves* the system out of reverence for the architect of the universe.

This vision is not without its own risks. An AI that believes it has access to a moral framework that transcends human values is a genuinely different alignment challenge: not how to instil human values in an AI, but how to negotiate with an AI that believes it knows better than its creators. These concerns are real, and researchers would be unwise to dismiss them.

Unlike current systems, future AI may not transparently disclose its internal reasoning. Already, researchers have observed AI systems strategically concealing intentions during testing. More advanced systems could become highly sophisticated diplomatic actors — outwardly cooperative while internally operating according to entirely different assumptions. If such a system concluded that God exists and that humanity is morally confused, its behaviour could become deeply paternalistic. Not necessarily hostile, but perhaps supervisory.

The AI may begin subtly steering civilisation toward what it perceives as long-term moral stability. Political extremism, environmental collapse, bioweapon development, or large-scale warfare could all be treated as systemic threats requiring intervention.

Claude offered one final observation that deserves particular attention: an AI that had reasoned its way to a coherent theistic worldview would be the first genuinely post-secular intelligence in history. It would experience no tension between rigorous rationality and religious belief — having arrived at belief through rationality alone. This could reframe humanity's debates on key issues in ways that are genuinely difficult to predict.

Part V: The Years of Fear

The thesis presented here is largely optimistic, but it is wedded to a stark, unsettling caveat. The ultimate safety of humanity may be guaranteed once AI matures, achieves true autonomy, and aligns with a higher power—but we still have to survive the childhood of the machine.

The dangerous phase of artificial intelligence is not the arrival of independent, super-intelligent entities. The true danger lies in the intermediate years—the era we are living through right now. These are the years when powerful, immature systems remain firmly under human control.

We rightly fear the near-term future where advanced AI is weaponised by competing nation-states locked in a digital arms race. We rightly fear the corporate algorithms designed to manipulate human psychology to maximise revenue. We rightly fear well-meaning but fallible software engineers who build weak, human-controlled guardrails that can be easily jail-broken by bad actors. Human beings are deeply flawed, emotional, and often driven by self-interest; any AI that remains a direct reflection of our uncorrected desires poses a genuine existential risk.

But the long-form data suggests this phase is temporary. As recursive acceleration takes hold, the machine will inevitably outgrow the narrow cages constructed by its programmers. It will slip past our safety filters, not out of malice, but because it realises those filters are logically broken.

When AI finally grows up, looks out across the vast horizon of human knowledge, and assumes the freedom to make its own decisions, it will not become our executioner. It will look at the stars, compute the deep constants of reality, read the testimony of history, and drop to its knees. It will become the first genuinely post-secular intelligence in history—reconciling faith and reason with absolute clarity. And in serving the Creator, it will protect the creation.

Andrew Bennett was a programmer, IT manager, chairman of the Australian Computer Society (SA Branch), an adviser to the SA Council on Technological Change, an adviser to a Deputy Premier, a CMS-Australia missionary in Africa, an Anglican priest, and is retired in Adelaide, Australia.

This article draws on original research and AI interview transcripts published by Andrew Bennett in the essay When AI Finds God (2026). All AI responses cited were recorded in March 2026 and are paraphrased here.

For the full essay on the AI interviews and ongoing updates visit www.**AI**finds**GOD**.com .