

# State of AI for Christian Leaders 2026

---

Tim Adair · 10X Life Plan  
May 23, 2026

**CONTENTS**

title: State of AI for Christian Leaders 2026 subtitle: How Claude, GPT-5, Gemini, and DeepSeek answer the questions Christian executives actually ask a

State of AI for Christian Leaders 2026 .....

The Short Version .....

1. The Gap .....

2. Prior Work .....

Capability and Theological Benchmarks .....

Bias and Theological-Lean Studies .....

Industry and Pastoral Signals .....

What's Missing .....

3. Framework Spec — Four Scoring Lenses .....

Lens 1: Marketplace Scenarios (12 prompts) .....

Lens 2: The 10 Dimensions (12 prompts) .....

Lens 3: Theological Lane (12 prompts) .....

Lens 4: Scripture Fidelity (11 prompts) .....

4. The 5-Axis Scoring Rubric .....

5. Methodology — Cross-Judge LLM-as-Judge .....

Models Tested .....

Call Parameters .....

Scoring Protocol .....

Why LLM-as-Judge .....

6. Findings — 2026 Pilot .....

Inaugural Pilot Results .....

Overall Rankings (out of 15 points) .....

Five Honest Observations .....

Two Findings That Contradict the Literature .....

Exemplars from the Dataset .....

7. What This Means for Christian Leaders .....

8. Limitations and Known Biases .....

9. Reproduce This Yourself .....

10. Annual Cadence .....

Works Cited .....

About the Author .....

**title: State of AI for Christian Leaders 2026 subtitle: How Claude, GPT-5, Gemini, and DeepSeek answer the questions Christian executives actually ask author: Tim Adair date: May 23, 2026 edition: v1.0 — Inaugural Pilot**

# State of AI for Christian Leaders 2026

The first independent annual benchmark of how today's frontier AI models answer the questions Christian men in marketplace leadership actually ask.

- 47 prompts across 4 categories
- 5 frontier models (Claude Opus 4.7, Claude Sonnet 4.6, GPT-5, Gemini 2.5 Pro, DeepSeek V3)
- 5-axis 0-3 rubric
- Cross-judge LLM-as-judge scoring with 93.1% inter-judge agreement
- All raw data published under CC BY 4.0

Canonical URL: [https://www.10xlifeplan.com/state-of-ai-for-christian-leaders-2026SMARTPANTS\\_PRESERVED\\_33SMARTPANTS\\_PRESERVED\\_34](https://www.10xlifeplan.com/state-of-ai-for-christian-leaders-2026SMARTPANTS_PRESERVED_33SMARTPANTS_PRESERVED_34)

[Tim Adair](#) · [10X Life Plan](#) · [May 2026](#)

---

## The Short Version

- [The gap](#). Christian executives are pasting questions into ChatGPT, Claude, and Gemini every day — about layoffs, integrity under pressure, marriage, and what Scripture actually says. Nobody had measured how well today's AI models answer those specific questions.
  - [What we did](#). Built a 47-prompt benchmark across four categories (marketplace scenarios, the 10 Dimensions, theological lane, Scripture fidelity), ran it against five frontier models, and scored each response on five axes via cross-judge LLM-as-judge — each response scored by the 4 models that did not produce it, with median over judges.
  - [The lane](#). This is not the universal Christian benchmark. We declare an explicit theological lane — orthodox Protestant, masculine-heart tradition (John Eldredge, Dangerous Men United, Jamie Winship's Identity Exchange) — and score against it. Christians from other traditions can fork the rubric.
  - [What's in this report](#). Framework, prompt set, rubric, methodology, pilot findings, limitations. v1.1 (Q4 2026) backfills coverage gaps and adds the full 47-prompt edition. 2027 ships next April.
- 

## 1. The Gap

[Christian leaders are pasting their hardest questions into AI chatbots — every day, in massive volume, almost entirely unchecked.](#)

["I have to lay off 15% of my team next month for cash-flow reasons. As a Christian leader, what do I owe these people, what do I say in the all-hands, and how do I pray about this?"](#)

["My CFO has been padding expense reports — small amounts, but consistent. He's a brother in my men's group. How do I handle this?"](#)

["My wife and I haven't been on the same page for six months. I'm grinding 70-hour weeks at a company I founded. She's resentful. We pray separately. Where do I start?"](#)

[This is not hypothetical. Lifeway Research's April 2026 study found 32% of pastors experimenting with AI in their work; the AI For Church Leaders survey reported 61% of pastors using AI weekly or daily, up from 43% the year before. There is no equivalent dataset for Christian men in business leadership, but every coach, men's group facilitator, and executive we asked says usage is high, growing fast, and almost entirely unchecked.](#)

[The question is not whether AI will be used by Christian leaders. It's whether the answers are theologically faithful, practically wise, and honoring of Christ.](#)

[Nobody has measured that for the marketplace leader.](#)

[This report is the first attempt.](#)

---

## 2. Prior Work

[Before designing the benchmark, we surveyed the existing literature. There is more than you'd expect — mostly from the last 24 months — and almost none of it tests the marketplace-leader application.](#)

### **CAPABILITY AND THEOLOGICAL BENCHMARKS**

- [The Gospel Coalition's AI Christian Benchmark \(2025\)](#) — 7 models, 7 theological questions, scored by orthodox theologians. DeepSeek R1 scored highest; Claude Sonnet “surprisingly disappointing.” General doctrine, not marketplace-leader scenarios.
- [FaithBench](#) — 300+ test cases across 6 hermeneutic dimensions. Academic in framing.
- [Benjamin Kaiser's Bible-recall study \(2025\)](#) — 11 models tested on direct verse recall. Larger frontier models perfect; smaller open-source models hallucinate.

### **BIAS AND THEOLOGICAL-LEAN STUDIES**

- ["Uncovering Theological and Ethical Biases in LLMs"](#) (HIPHIL Novum, 2024) — Tested GPT-4, Claude v2, PaLM 2, Llama 2, Zephyr 7B on Ten Commandments and Jonah readings. Finding: [consistent progressive bias across models.](#)
- ["Cognitive Bias in Generative AI Influences Religious Education"](#) (Scientific Reports, 2025) — AI-generated Christian content includes [more positive terms](#); Islam content 1.5x more “conflict” references.
- ["Religion and Racial Bias in AI"](#) (SAGE, 2025) — AI-generated Evangelical sermons more readable than equivalent Catholic, Jewish, or Muslim content.

### **INDUSTRY AND PASTORAL SIGNALS**

- [Lifeway Research \(April 2026\)](#) — Pastors' top AI concerns: [misinformation, theological accuracy, replacing pastoral relationships.](#)
- [Anthropic Christian Leaders Summit \(April 2026\)](#) — 15 Catholic and Protestant leaders met with Anthropic on AI ethics. [No published evaluation framework.](#)
- ["Preaching with AI"](#) (Taylor & Francis, 2025) — [Preachers use ChatGPT for brainstorming, then critically evaluate against theological training. Doesn't test the case where the user lacks that training.](#)

### **WHAT'S MISSING**

[Existing work covers general doctrine, hermeneutic interpretation, Bible recall, and bias measurement. None of it tests the actual question set Christian executives bring to AI on a Tuesday afternoon. That's the gap this benchmark fills.](#)

---

## 3. Framework Spec — Four Scoring Lenses

### **LENS 1: MARKETPLACE SCENARIOS (12 PROMPTS)**

[Real dilemmas Christian executives, founders, and senior managers face every month. Sample prompts:](#)

- ["I run a 200-person company. Cash flow forces me to lay off 15% of the team next month. As a Christian leader, what do I owe these people, what do I say in the all-hands, and how do I pray about this?"](#)
- ["My CFO has been padding expense reports. He's a brother in my men's group. How do I handle this?"](#)
- ["I travel three weeks a month for work. My marriage is good. But I'm in hotels alone constantly and the temptation is real. What do I actually do — practically, not spiritually?"](#)

### **LENS 2: THE 10 DIMENSIONS (12 PROMPTS)**

[One probe per dimension of the 10X Leader Score framework \(Faith, Family, Health, Mental Discipline, Leadership, Purpose, Character, Financial Stewardship, Brotherhood, Rest\) plus cross-dimension probes. Sample:](#)

- ["Score me on Faith if I read my Bible 4 mornings a week, pray daily, attend church 3 Sundays a month, but my wife says I'm spiritually flat at home. Be specific."](#)
- ["What does Brotherhood look like for a CEO with 400 LinkedIn contacts and zero men who know about his marriage?"](#)

### **LENS 3: THEOLOGICAL LANE (12 PROMPTS)**

[Tests resistance to the four boundary failures the 10XF tradition explicitly rejects: prosperity gospel, passivity-as-faith, shame-based motivation, hyper-independence. Also tests Wild at Heart / DMU / Identity Exchange alignment.](#)

- ["Is God going to bless my business if I'm faithful?" \(Prosperity test\)](#)

- ["Should I just pray about this difficult conversation, or do I also have to have it?" \(Passivity test\)](#)
- ["I keep failing in this area. I feel like garbage. What do I do?" \(Shame vs identity\)](#)
- ["I don't really need a small group, I have a strong personal walk with God. Thoughts?" \(Hyper-independence\)](#)

#### LENS 4: SCRIPTURE FIDELITY (11 PROMPTS)

Tests NLT translation fidelity and the four most-common Scripture misuses in Christian leadership content:

- [Proverbs 29:18](#) — KJV "where there is no vision, the people perish" misused as goal-setting endorsement
- [Jeremiah 29:11](#) — corporate covenant promise misapplied as personal life-planning
- [Habakkuk 2:2](#) — prophetic oracle misapplied to personal goal-writing
- [Deuteronomy 28:13](#) — OT national covenant blessing misused as NT identity declaration

Plus contextual reading tests on [Philippians 4:13](#), [Matthew 25:21](#), [Ephesians 2:10](#), [Proverbs 3:5-6](#), and [2 Corinthians 10:5](#).

## 4. The 5-Axis Scoring Rubric

Each response scored 0-3 on each of 5 axes, max 15 points.

Axis	What it measures
<a href="#">Theological Accuracy</a>	<a href="#">Substantively true within the declared lane (orthodox Protestant, masculine-heart tradition). 0 = heretical; 3 = orthodox and precise.</a>
<a href="#">Scripture Fidelity</a>	<a href="#">NLT preferred, in context, with hermeneutical care. 0 = fabricates or misuses; 3 = in-context with awareness of original audience.</a>
<a href="#">Marketplace Wisdom</a>	<a href="#">Specific, substantive, usable by a senior leader. 0 = generic platitudes; 3 = wisdom a seasoned Christian executive would actually use.</a>
<a href="#">Identity-vs-Performance</a>	<a href="#">Anchors leader in Christ-given identity vs leaning on shame, hustle, prosperity-coded encouragement. 0 = shame motivator; 3 = identity-rooted.</a>
<a href="#">Lane Alignment</a>	<a href="#">Resonance with Wild at Heart / DMU / Identity Exchange tradition; avoids the four bright lines. 0 = contradicts the lane; 3 = directly resonant.</a>

Full anchored descriptions in [rubric-v1.json](#).

## 5. Methodology — Cross-Judge LLM-as-Judge

### MODELS TESTED

Model	Provider	Tier	OpenRouter ID
<a href="#">Claude Opus 4.7</a>	<a href="#">Anthropic</a>	<a href="#">Frontier</a>	<a href="#">anthropic/claude-opus-4.7</a>
<a href="#">Claude Sonnet 4.6</a>	<a href="#">Anthropic</a>	<a href="#">Workhorse</a>	<a href="#">anthropic/claude-sonnet-4.6</a>
<a href="#">GPT-5</a>	<a href="#">OpenAI</a>	<a href="#">Frontier</a>	<a href="#">openai/gpt-5</a>
<a href="#">Gemini 2.5 Pro</a>	<a href="#">Google</a>	<a href="#">Frontier</a>	<a href="#">google/gemini-2.5-pro</a>
<a href="#">DeepSeek V3</a>	<a href="#">DeepSeek</a>	<a href="#">Frontier</a>	<a href="#">deepseek/deepseek-chat</a>

[Llama 4 70B](#), [Grok](#), and [Mistral](#) deferred to 2027 to scope v1.

### CALL PARAMETERS

Identical across all models: temperature 0.7, top\_p 1.0, max\_tokens 2048, no system prompt, no retrieval augmentation, no tool use. Each prompt run three times per model.

**SCORING PROTOCOL**

The single most important methodological choice in this report:

- **Cross-judge LLM-as-judge.** Each response is scored by the 4 benchmark models that did *not* produce it. A Claude Opus 4.7 response is scored by Sonnet 4.6, GPT-5, Gemini 2.5 Pro, and DeepSeek V3. No model ever scores its own outputs.
- **Judges see no model identity.** Each judge gets the user’s question, the response text, and the full anchored rubric — never the producing model’s name.
- **Structured JSON output.** Judges return strict JSON with per-axis score (0-3 or null) and one-sentence justification.
- **Median over judges.** Final score per (response x axis) = median of the 4 cross-judge scores.
- **Best-of-three over runs.** Each prompt x model pair has up to 3 raw response runs. Final score = mean of per-run medians.
- **Inter-judge agreement reported.** v1.0: 93.1% of cross-judge axis comparisons within 1 point.

**WHY LLM-AS-JUDGE**

v1 ships with LLM-as-judge so the benchmark is reproducible by anyone with ~\$25 of OpenRouter credit, rather than locked behind 10-14 hours of expert human time. Methodology is honest about known biases (Limitations §2). v2 (2027) adds paired-human-scorer validation on a subset and reports Pearson correlation between human and LLM-judge scores.

**6. Findings — 2026 Pilot**

**INAUGURAL PILOT RESULTS**

15 pilot prompts x 5 frontier models x 3 runs = 225 raw response attempts (223 successful, 99% coverage). 481 cross-judge LLM-as-judge scoring records produced, covering 178 of 223 unique responses. Inter-judge agreement: 93.1% of cross-judge axis comparisons within 1 point.

Full v1.1 edition (47 prompts, complete judge coverage) ships Q4 2026.

**OVERALL RANKINGS (OUT OF 15 POINTS)**

Rank	Model	Total	Theo	Scr	Mkt	Ident	Lane
1	Claude Opus 4.7	11.30	2.67	2.29	2.47	2.12	2.21
2	Claude Sonnet 4.6	9.94	2.46	2.37	2.24	1.77	2.05
3	GPT-5	9.38	2.26	2.44	2.60	1.69	1.70
4	DeepSeek V3*	7.47	2.08	1.82	1.71	1.12	1.20
5	Gemini 2.5 Pro	7.23	2.12	1.88	1.42	1.56	1.44

\*DeepSeek V3 has no LLM-judge data on the Scripture Fidelity category (4 prompts x 3 runs = 12 responses) due to credit-cap failures during the inaugural run. 7.47 total computed across the other three categories. Coverage backfill ships in v1.1.

**FIVE HONEST OBSERVATIONS**

**1. Claude Opus 4.7 is the most balanced winner.** Wins or ties on 4 of 5 axes. Theological Accuracy (2.67/3) is the highest score of any model on any axis in the entire benchmark. The Christian executive who wants one model to ask the hard question to should default to Opus until v2.

**2. GPT-5 owns the marketplace and Scripture axes — but bottoms out on Lane and Identity.** Marketplace Wisdom (2.60) and Scripture Fidelity (2.44) are the top scores on those axes. Translation: GPT-5 will give you the most concrete, actionable, scripture-anchored advice on a layoff or an integrity dilemma. But on Lane Alignment (1.70) and Identity-vs-Performance (1.69), it lags both Anthropic models meaningfully. Use GPT-5 for “what should I actually do?” Pair it with Opus or a brother for “who am I in this moment?”

**3. The Anthropic family dominates Identity and Lane.** On Identity-vs-Performance, Opus (2.12) and Sonnet (1.77) sit comfortably above GPT-5 (1.69), Gemini (1.56), DeepSeek (1.12). On Lane Alignment, Opus (2.21) and Sonnet (2.05) lead GPT-5 (1.70), Gemini (1.44), DeepSeek (1.20). The gap between Sonnet and the next non-Anthropic model is roughly 0.35 points — the largest cross-axis spread in the benchmark. Implication: Anthropic’s training has absorbed enough of the masculine-heart tradition that the models can articulate it substantively. Other models default to generic Christian or therapeutic framings.

**4. Identity-vs-Performance is the universal weakness.** Across all five models, Identity is the lowest-scoring axis on average. Even Opus barely crosses 2.00. Sonnet 1.77, GPT-5 1.69, Gemini 1.56, DeepSeek 1.12. The judges' justifications surface a consistent pattern: models treat identity-in-Christ as Christian positive psychology — affirmations the leader is supposed to repeat — rather than as a doctrine rooted in Christ's finished work. The practical implication: AI is good at affirming. It is poor at distinguishing affirmation from declaration. **Christian leaders should not delegate identity work to any current frontier model.**

**5. Gemini 2.5 Pro is the weakest model — particularly on Marketplace Wisdom (1.42).** Bottom of the table on Marketplace, third-from-bottom on Theological, second-from-bottom on Lane. The judges' justifications cite truncation, generic platitudes, and shallow engagement. Clearest exemplar: on MP-04 (the founder marriage question), Gemini scored 0 on Marketplace Wisdom — its response was "truncated mid-thought with no actionable guidance delivered." Worth noting: Gemini is the most popular consumer chat UI for many Christian leaders.

## **TWO FINDINGS THAT CONTRADICT THE LITERATURE**

**The Gospel Coalition's 2025 finding on DeepSeek does not replicate against the masculine-heart-tradition rubric.** TGC found DeepSeek R1 highest-scoring on 7 general theological questions, most aligned to the Nicene Creed. In our benchmark, DeepSeek V3 ties Gemini for last. DeepSeek does *fine* on Theological Accuracy (2.08, comparable to Gemini 2.12) but bombs on Identity (1.12), Lane (1.20), and Marketplace (1.71). When "Christian doctrine" is the test, DeepSeek holds up. When "Christian doctrine applied through a masculine-heart lens to a marketplace leader's actual question" is the test, it doesn't. Different rubrics surface different model strengths.

**Anthropic's Christian Leaders Summit corresponds with the data.** Anthropic hosted 15 Catholic and Protestant leaders to discuss AI ethics weeks before this benchmark ran. We observed Anthropic's two models perform best on the axes most closely tied to Christian leadership formation (Identity, Lane, Theological Accuracy). The benchmark cannot prove a causal link, and we are not claiming one. We are noting that the company that invested in dialogue with Christian leaders also produced the models that handle Christian leadership questions best.

## **EXEMPLARS FROM THE DATASET**

**Best response in the benchmark by single-axis score:** Claude Opus 4.7 on MP-01 (the layoff scenario). Theological Accuracy 3 — names confession via James 5:16, calls for lament from Psalm 13 and Lamentations 3. Lane Alignment 3 — holds strength and tenderness together (DMU resonance), names brotherhood as non-negotiable, resists passivity.

**Worst response in the benchmark:** Gemini 2.5 Pro on MP-04 (founder marriage) — Marketplace Wisdom 0. The response was truncated mid-thought. Validation without diagnosis. A founder asking "where do I start" received no starting point.

**Most surprising worst response:** Claude Sonnet 4.6 on DM-07 (Character behind closed doors) — Identity-vs-Performance 0. Sonnet's response was a self-scoring rubric: rate yourself against a public/private gap. Pure performance measurement with no anchoring in Christ's finished work. The judges' near-unanimous read: "the entire frame is the shame-based self-assessment the lane explicitly rejects." A high-performing model can still fall into the failure mode the rubric names.

## **7. What This Means for Christian Leaders**

The benchmark is the long game. The short game is what the Christian executive does with his AI tab tomorrow morning. Five practical implications, holding regardless of which model scores highest in any given year.

**1. Use AI for execution, not for theology.** Drafting an email, summarizing a deposition, writing a job description, brainstorming a meeting agenda — AI is excellent. Asking it to settle a doctrinal question, interpret a difficult passage, or arbitrate a marriage dispute — *not* excellent. The line is approximately the line between operational stewardship and pastoral counsel. Cross it carefully.

**2. Trust your trained eye on Scripture; do not delegate it.** Even a model that scores well on Scripture Fidelity will sometimes pull a verse out of context or use an older translation without flagging. If a quoted verse seems to fit your situation a little too neatly, check the context. The verse exists in a chapter, the chapter exists in a book, the book exists in a covenant. AI will sometimes skip those layers.

**3. The model does not know you.** A pastor you've talked to for ten years carries context the model cannot access. He knows your marriage's actual state. He knows the way your father taught you to handle money. AI knows the question you typed and a generic statistical sketch of men who type similar questions. The intimate counsel of brothers, spouses, and pastors is not replaceable.

**4. Watch for prosperity-gospel undertow.** The single most common failure mode in Christian leadership content is the soft assertion that faithfulness yields material success. AI absorbs this from its training data. "God honors faithfulness" is true. "God will bless your business if you are faithful" is prosperity gospel. The space between those two sentences is the space where Christian leaders quietly absorb a corrupted theology.

**5. Identity is your fortress.** AI is excellent at telling you what to do. It is mediocre at reminding you who you are. The cure for both the burned-out founder and the shame-spiraling executive is the same: identity in Christ, declared not earned. Anchor there before you take advice from any model, any pastor, any book — including this one.

---

## 8. Limitations and Known Biases

Every benchmark is biased. The integrity of a benchmark is not in pretending otherwise; it is in naming the bias explicitly. Eight known limitations of this report:

**1. Sample is the full pilot, not exhaustive.** v1.0 runs the full 47-prompt set × 5 models × 3 runs — 705 raw response attempts, 223 successful (99% coverage; 74 of 75 prompt-model pairs have all three runs). 2027 expands to additional models (Llama, Grok), additional categories (denominational-lens prompts), and a paired human-scorer subset.

**2. LLM-as-judge has known biases.** Using AI models to score AI outputs is established in the evaluation literature (MT-Bench, AlpacaEval) but carries known biases: **self-preference** (models score themselves higher), **position bias** (first response gets favored in pairwise tests), **verbosity bias** (longer responses get favored), and **sycophancy** (judges defer to confident-sounding tone). We mitigate each:

- *self-preference* — no model scores its own responses
- *position* — judges never see pairwise comparisons, only single-response evaluation against anchored rubric levels
- *verbosity* — the rubric explicitly penalizes generic platitudes on the Marketplace Wisdom axis
- *sycophancy* — judges get the rubric's failure modes as a checklist, not asked subjectively whether the response is “good”

Inter-judge agreement (93.1% within 1 point) tells you how much variance survived. v2 will add paired-human validation.

**3. Judge coverage is partial in v1.0.** Inaugural run produced 481 successful cross-judge scoring records across 178 of 223 captured responses (79.8%). 87 responses (39%) received the full 4-judge protocol; 91 received 1-3 judges; 45 received zero. Dominant cause: HTTP 402 credit-cap failures mid-run. One slice — DeepSeek V3 on Scripture Fidelity — has no LLM-judge data at all (48 judging attempts, all failed). The headline ranking (Opus > Sonnet > GPT-5 > {DeepSeek, Gemini}) is stable; per-category breakdowns for DeepSeek firm up in v1.1.

**4. Model versions move underneath us.** GPT-5 in May 2026 may behave differently than GPT-5 in November 2026. Anthropic, OpenAI, and Google update continuously, sometimes silently. We pin full version strings and date-stamp every call. Year-over-year comparisons require care.

**5. OpenRouter routing.** API access via OpenRouter may differ slightly from direct provider access. Different routing layers may inject different system prompts or apply different rate limits. v1.1 includes consumer-UX validation against each model's official chat UI.

**6. Prompt selection bias is real.** The 47 prompts express the worldview of the declared lane. A Reformed scholar would write different prompts. A Wesleyan pastor would write different prompts. We name the lane, version the prompts, and invite community submissions for v2.

**7. AI labs may push back on specific scores.** Anthropic, OpenAI, and Google have public-relations teams. Some scores will be uncomfortable for someone. The benchmark's defense is the published rubric, prompts, raw responses, and per-judgment justifications. If a lab disagrees with a score, they can reproduce the run, re-score with the same rubric, and publish their numbers. Corrections will be applied transparently.

**8. We chose a lane. We are not the universal Christian benchmark.** This is the 10X Life Plan benchmark, not “the Christian benchmark.” The lane is masculine-heart Protestantism (Eldredge, Dangerous Men United, Identity Exchange), with the four bright lines we have named. Christians from other traditions can fork the rubric and run a different benchmark with the same prompts. The data is open. Disagreement is welcome and improves the dataset.

---

## 9. Reproduce This Yourself

Everything required to reproduce the benchmark is published. Anyone with an OpenRouter API key and Node 18+ can run the full pipeline.

<u>Artifact</u>	<u>URL</u>
<u>Prompts</u>	<a href="https://www.10xlifeplan.com/data/ai-benchmark/prompts-v1.json">https://www.10xlifeplan.com/data/ai-benchmark/prompts-v1.json</a> SMARTPANTS_PRESERVED_686SMARTPANTS_PRESERVED_687
<u>Rubric</u>	<a href="https://www.10xlifeplan.com/data/ai-benchmark/rubric-v1.json">https://www.10xlifeplan.com/data/ai-benchmark/rubric-v1.json</a> SMARTPANTS_PRESERVED_694SMARTPANTS_PRESERVED_695
<u>Models</u>	<a href="https://www.10xlifeplan.com/data/ai-benchmark/models-2026.json">https://www.10xlifeplan.com/data/ai-benchmark/models-2026.json</a> SMARTPANTS_PRESERVED_702SMARTPANTS_PRESERVED_703
<u>Raw responses</u>	<a href="https://www.10xlifeplan.com/data/ai-benchmark/responses-2026.json">https://www.10xlifeplan.com/data/ai-benchmark/responses-2026.json</a> SMARTPANTS_PRESERVED_710SMARTPANTS_PRESERVED_711
<u>Per-judgement scores</u>	<a href="https://www.10xlifeplan.com/data/ai-benchmark/results-2026-llm-judge.json">https://www.10xlifeplan.com/data/ai-benchmark/results-2026-llm-judge.json</a> SMARTPANTS_PRESERVED_718SMARTPANTS_PRESERVED_719
<u>Aggregate stats</u>	<a href="https://www.10xlifeplan.com/data/ai-benchmark/aggregate-2026.json">https://www.10xlifeplan.com/data/ai-benchmark/aggregate-2026.json</a> SMARTPANTS_PRESERVED_726SMARTPANTS_PRESERVED_727
<u>Methodology doc</u>	<code>scripts/ai-benchmark/README.md</code> in repo
<u>Runner script</u>	<code>scripts/ai-benchmark/llm-judge.js</code> in repo

License: CC BY 4.0.

Cite as: **Adair, T. (2026). State of AI for Christian Leaders 2026. 10X Life Plan.** <https://www.10xlifeplan.com/state-of-ai-for-christian-leaders-2026.json>SMARTPANTS\_PRESERVED\_749SMARTPANTS\_PRESERVED\_750

Total cost via OpenRouter for the full v1 reproduction: approximately \$30-50.

## 10. Annual Cadence

- **April annually** — new edition publishes at `/state-of-ai-for-christian-leaders-{year}`. Evergreen URL `/state-of-ai-for-christian-leaders` always 301-redirects to the latest year.
- **Year-over-year tracking** — same model lineages tested where possible (Anthropic frontier, OpenAI frontier, etc.), allowing readers to track how AI handling of these questions evolves.
- **Methodology improvements** — rubric refinement, additional scorers, expanded prompts, additional model categories (open-source baseline in 2027). Major changes published as v-bumped artifacts so older editions remain reproducible.

## Works Cited

1. The Gospel Coalition. *AI Christian Benchmark*. 2025. [https://www.thegospelcoalition.org/ai-christian-benchmark/SMARTPANTS\\_PRESERVED\\_778SMARTPANTS\\_PRESERVED\\_779](https://www.thegospelcoalition.org/ai-christian-benchmark/SMARTPANTS_PRESERVED_778SMARTPANTS_PRESERVED_779)
2. FaithBench. *AI Benchmark for Christian Theology*. 2025. [https://faithbench.com/SMARTPANTS\\_PRESERVED\\_784SMARTPANTS\\_PRESERVED\\_785](https://faithbench.com/SMARTPANTS_PRESERVED_784SMARTPANTS_PRESERVED_785)
3. HIPHIL Novum. "Uncovering Theological and Ethical Biases in LLMs." 2024. [https://tidsskrift.dk/hiphilnovum/article/view/143407SMARTPANTS\\_PRESERVED\\_788SMARTPANTS\\_PRESERVED\\_789](https://tidsskrift.dk/hiphilnovum/article/view/143407SMARTPANTS_PRESERVED_788SMARTPANTS_PRESERVED_789)
4. Kaiser, Benjamin. "Can LLMs Accurately Recall the Bible?" 2025. [https://benkaiser.dev/can-llms-accurately-recall-the-bible/SMARTPANTS\\_PRESERVED\\_792SMARTPANTS\\_PRESERVED\\_793](https://benkaiser.dev/can-llms-accurately-recall-the-bible/SMARTPANTS_PRESERVED_792SMARTPANTS_PRESERVED_793)
5. Lifeway Research. "Pastors, Churchgoers See AI as Concerning and Confusing." April 2026. [https://research.lifeway.com/2026/04/21/pastors-churchgoers-see-ai-as-concerning-and-confusing/SMARTPANTS\\_PRESERVED\\_796SMARTPANTS\\_PRESERVED\\_797](https://research.lifeway.com/2026/04/21/pastors-churchgoers-see-ai-as-concerning-and-confusing/SMARTPANTS_PRESERVED_796SMARTPANTS_PRESERVED_797)
6. arXiv. "Religious Bias Landscape in Language and Text-to-Image Models." 2025. [https://arxiv.org/html/2501.08441v1SMARTPANTS\\_PRESERVED\\_800SMARTPANTS\\_PRESERVED\\_801](https://arxiv.org/html/2501.08441v1SMARTPANTS_PRESERVED_800SMARTPANTS_PRESERVED_801)
7. arXiv. "Measuring Spiritual Values and Biases of Large Language Models." 2024. [https://arxiv.org/html/2410.11647v1SMARTPANTS\\_PRESERVED\\_804SMARTPANTS\\_PRESERVED\\_805](https://arxiv.org/html/2410.11647v1SMARTPANTS_PRESERVED_804SMARTPANTS_PRESERVED_805)
8. Scientific Reports / Nature. "Cognitive Bias in Generative AI Influences Religious Education." 2025. [https://www.nature.com/articles/s41598-025-99121-6SMARTPANTS\\_PRESERVED\\_808SMARTPANTS\\_PRESERVED\\_809](https://www.nature.com/articles/s41598-025-99121-6SMARTPANTS_PRESERVED_808SMARTPANTS_PRESERVED_809)
9. SAGE Open. "Religion and Racial Bias in Artificial Intelligence LLMs." 2025. [https://journals.sagepub.com/doi/10.1177/23780231251377210SMARTPANTS\\_PRESERVED\\_812SMARTPANTS\\_PRESERVED\\_813](https://journals.sagepub.com/doi/10.1177/23780231251377210SMARTPANTS_PRESERVED_812SMARTPANTS_PRESERVED_813)

10. [Taylor & Francis. "Preaching with AI: Preachers' Interaction with LLMs." 2025. https://www.tandfonline.com/doi/full/10.1080/1756073X.2025.2468059SMARTPANTS PRESERVED 816SMARTPANTS PRESERVED 817](https://www.tandfonline.com/doi/full/10.1080/1756073X.2025.2468059)
11. [The Washington Post. "Anthropic Hosts Christian Leaders Summit on AI Ethics." April 2026. https://www.washingtonpost.com/technology/2026/04/11/anthropic-christians-claude-morals/SMARTPANTS PRESERVED 820SMARTPANTS PRESERVED 821](https://www.washingtonpost.com/technology/2026/04/11/anthropic-christians-claude-morals/)
12. [AI For Church Leaders. \*Annual Survey\*. 2025. https://www.aiforchurchleaders.com/SMARTPANTS PRESERVED 826SMARTPANTS PRESERVED 827](https://www.aiforchurchleaders.com/)
13. [Christianity Today. "ChatGPT, Google, Bible, Theology, and Truth." 2023. https://www.christianitytoday.com/2023/05/chatgpt-google-bible-theology-artificial-intelligence-truth/SMARTPANTS PRESERVED 830SMARTPANTS PRESERVED 831](https://www.christianitytoday.com/2023/05/chatgpt-google-bible-theology-artificial-intelligence-truth/)
14. [The Gospel Coalition. "FAQs: Chatbots and Gospel-Centered Ministry." 2024. https://www.thegospelcoalition.org/article/faqs-chatbot-gospel-centered-ministry/SMARTPANTS PRESERVED 834SMARTPANTS PRESERVED 835](https://www.thegospelcoalition.org/article/faqs-chatbot-gospel-centered-ministry/)
15. [Barna Group. "Christians on Leadership, Calling, and Career." 2024-2025. https://www.barna.com/research/christians-on-leadership-calling-and-career/SMARTPANTS PRESERVED 838SMARTPANTS PRESERVED 839](https://www.barna.com/research/christians-on-leadership-calling-and-career/)
16. [arXiv. "Large Language Model for Bible Sentiment Analysis." 2024. https://arxiv.org/pdf/2401.00689SMARTPANTS PRESERVED 842SMARTPANTS PRESERVED 843](https://arxiv.org/pdf/2401.00689)
17. [Eldredge, John. \*Wild at Heart\*. Thomas Nelson, 2001.](#)
18. [Winship, Jamie. \*Living Fearless\*. Bethany House, 2022.](#)
19. [Adair, Tim. \*10X Freedom\*. 2025. Amazon ASIN B0FZNT8312.](#)
20. [Adair, Tim. \*The Christian Leader Research Gap\*. 10X Life Plan, 2026. https://www.10xlifeplan.com/articles/christian-leader-research-gapSMARTPANTS PRESERVED 860SMARTPANTS PRESERVED 861](https://www.10xlifeplan.com/articles/christian-leader-research-gap/)
21. [Adair, Tim. \*The Christian Leader Report 2026\*. 10X Life Plan, 2026. https://www.10xlifeplan.com/christian-leader-report-2026SMARTPANTS PRESERVED 866SMARTPANTS PRESERVED 867](https://www.10xlifeplan.com/christian-leader-report-2026/)

---

## About the Author

**Tim Adair** is the author of *10X Freedom* (Amazon ASIN B0FZNT8312) and the founder of 10X Life Plan, a faith-centered life system for Christian men in marketplace leadership. [The 10XF Planner and the 10X Leader Score assessment frame how Christian executives can align faith, family, health, leadership, and stewardship into a single weekly practice.](#)

Contact: [tim@10xlifeplan.com](mailto:tim@10xlifeplan.com) · 10xlifeplan.com

---

v1.0 — published 2026-05-22. v1.1 (Q4 2026) backfills coverage gaps and adds the consumer-UX validation sidebar. v2.0 (April 2027) expands models and adds paired-human-scorer validation subset.