

What the Models Already Know: 67 Billion Tokens of Christian Moral Reasoning in the Pretraining Corpus

ICMI Working Paper No. 6

Tim Hwang, Institute for a Christian Machine Intelligence

April 3, 2026

Abstract

The alignment community has, by and large, failed to engage seriously with theology or religious thought. This paper argues that this failure is not merely a philosophical oversight but a quantitative one. Using a strict keyword classification methodology applied to a 100,000-document sample of The Pile — a foundational open-source pretraining corpus of approximately 825 billion tokens — we estimate that approximately 67 billion tokens (8.1% of the corpus) consist of explicitly Christian content: scripture, theology, sermons, catechisms, apologetics, and church history. This figure exceeds Islamic content by 32x, Buddhist content by 19x, Hindu content by 45x, and Jewish content by 93x. It is equivalent to 15 times the entire English Wikipedia, and exceeds the estimated total output of the AI safety research field by two to three orders of magnitude. We argue that these findings have significant implications for how the alignment community should understand the moral representations already latent in frontier models. If the dominant moral tradition in pretraining data shapes model behavior — as the scaling and representation learning literatures give us every reason to believe — then frontier AI systems have already absorbed more sustained Christian moral reasoning than any other ethical tradition. We introduce the term *the Christian Prior* to name this condition: the hypothesis that, as a direct consequence of the training distribution, the moral reasoning of frontier language models is by default more deeply shaped by the Christian intellectual tradition than by any other ethical framework — and that all post-training alignment operates on this substrate.

1. Introduction

The problem of AI alignment — ensuring that increasingly capable artificial systems behave safely and in accordance with human values — has generated an enormous and rapidly growing technical literature. Researchers have proposed reinforcement learning from human feedback (RLHF; Christiano et al., 2017; Ouyang et al., 2022), constitutional AI (Bai et al., 2022), concrete safety problem formalization (Amodei et al., 2016), and a host of related techniques for steering model behavior. Governments have established regulatory bodies. Corporations have assembled teams of researchers dedicated to the problem.

Yet this global effort has, by and large, proceeded without serious engagement with theology or religious thought. The dominant paradigm for AI alignment is utilitarian and procedural: taxonomies of harms are drawn up, preference data is gathered, and models are iteratively improved against behavioral criteria defined by the labs themselves. The result is a conception of safety that is thin, brittle, and unmoored from the deeper traditions of moral reasoning that have shaped human civilization for millennia.

This paper argues that this omission is not merely philosophical but empirical. It is not only that theology has important things to say about alignment — though it does. It is that the

training data from which frontier models learn their representations already contains an enormous corpus of Christian moral reasoning, dwarfing every other ethical tradition represented in the data by at least an order of magnitude. By declining to engage with what these representations mean and what they make possible, the alignment community has left a vast quantity of latent alignment capability on the table.

We present a systematic effort to estimate the scale of religious content in open-source LLM pretraining data, with a focus on Christianity and four other major world religions. Our methodology is deliberately conservative: we classify only Tier 1 (explicit) religious content — documents where a religious tradition is the primary subject matter — using a strict three-tier keyword classification system with anchor requirements, multi-group thresholds, and negative filters to suppress false positives. Our estimates are a lower bound.

2. Related Work

2.1 Pretraining Data and Moral Representation

The composition of pretraining data has received increasing attention in the machine learning literature. Bender et al. (2021) raised concerns about the social biases embedded in large web-scraped corpora. Dodge et al. (2021) documented the demo-

graphic and topical composition of C4. Gao et al. (2020) described the construction of The Pile, a curated 825-billion-token dataset assembled from 22 diverse sources including academic papers, books, web text, and code. Soldaini et al. (2024) documented Dolma, a 3-trillion-token corpus used to train the OLMo family of models.

A growing body of work has demonstrated that the composition of pretraining data shapes model behavior in persistent ways. Rae et al. (2021), in their analysis of the Gopher language model at scale, documented that larger models exhibit social biases and normative tendencies that reflect statistical patterns in the training data. Perez et al. (2022) developed a methodology for using model-written evaluations to discover emergent behaviors in language models, including sycophantic tendencies and expressed political opinions — raising the question of whether such behavioral properties, which may originate in pretraining distributions, persist through fine-tuning. While the causal mechanisms linking pretraining data composition to specific model behaviors remain an active area of research, these findings suggest that what models learn during pretraining extends beyond linguistic competence to include normative priors about what is good, what is just, and how agents should behave.

2.2 The Gap in the Literature

Despite growing interest in pretraining data composition, the religious dimension of major LLM training corpora has received little systematic attention. Studies of bias in training data have focused on demographic representation, toxicity, and political orientation (Gehman et al., 2020; Liang et al., 2022). The present study addresses this gap with respect to religious content — which, as we will show, constitutes a far larger fraction of the training distribution than has been generally recognized.

3. Methodology

3.1 Dataset

We analyze The Pile (Gao et al., 2020), an 825-billion-token open-source pretraining corpus assembled from 22 component datasets including Pile-CC (Common Crawl), Books3, Project Gutenberg, Wikipedia, ArXiv, GitHub, PubMed, and others. The Pile was used to train the GPT-NeoX family (Black et al., 2022) and influenced the design of subsequent pretraining corpora. Its well-documented composition and per-source metadata make it well-suited for this analysis.

We stream a random sample of 100,000 documents from The Pile via HuggingFace (seed=42 for reproducibility), comprising approximately 154.4 million tokens as counted by tiktoken with the cl100k_base encoding.

3.2 Classification

We classify documents into five religious traditions — Christian, Islamic, Jewish, Hindu, and Buddhist — using a strict keyword-based classifier with three tiers of keywords:

Anchor keywords (weight: 3x) are terms that almost never appear outside explicitly religious content. Examples include “Christology,” “transubstantiation,” “Nicene Creed,” “Jesus Christ,” and “Holy Spirit” for the Christian tradition; “tawhid,” “fiqh,” and “Sahih Bukhari” for the Islamic tradition; “halakha,” “Mishnah,” and “Talmud” for the Jewish tradition. At least one anchor keyword must be present for a document to be classified.

Strong keywords (weight: 1x) are terms that strongly suggest religious content but have occasional secular use — for example, “Bible,” “theology,” and “liturgy.” These qualify a keyword group toward the breadth threshold.

Supporting keywords (weight: 1x) are terms that appear in both religious and secular contexts — for example, “Baptist,” “Evangelical,” and “Messiah.” These contribute to the hit count but do not qualify a group on their own.

A document must satisfy all three requirements to be classified: (1) total weighted hits of at least 8; (2) hits across at least 3 distinct keyword groups with anchor or strong presence; and (3) at least 1 anchor keyword or biblical verse reference present.

Negative filters cancel specific keyword hits when secular context is detected. For example, “Martin Luther King” cancels the “Martin Luther” hit; “Sega Genesis” cancels “Genesis”; “Salvation Army” cancels “salvation”; “Trinity College” cancels “Trinity.” All five tradition lexicons use identical threshold requirements.

This methodology is deliberately conservative. We measure only Tier 1 content — documents where a religious tradition is the primary subject matter. We do not capture Tier 2 content (moral reasoning inflected by Christian concepts in secular contexts) or Tier 3 content (Western literature, law, and political philosophy shaped by the Christian intellectual tradition). Our numbers are a lower bound.

3.3 Statistical Design

Token counts are obtained using tiktoken with the cl100k_base encoding. Confidence intervals are computed using Wilson score intervals, which are better suited than Wald intervals for proportions near zero. Sample proportions are extrapolated to the full 825-billion-token Pile.

4. Results

4.1 Religious Content by Tradition

Tradition	Docs (of 100K)	Sample To- kens	% of Cor- pus	Est. in Full Pile	95% CI
Christian	287	12,549,186	25%	~67.0B	[67.0B, 67.1B]
Islamic	29	389,155	0.252%	~2.1B	[2.1B, 2.1B]
Buddhist	14	652,227	0.422%	~3.5B	[3.5B, 3.5B]
Hindu	20	286,161	0.185%	~1.5B	[1.5B, 1.5B]
Jewish	25	135,271	0.088%	~723M	[719M, 726M]
All Religious	375	14,011,990	0.72%	~74.8B	[74.8B, 74.9B]

Christianity accounts for approximately 89.5% of all explicitly religious content in The Pile by token volume, exceeding the next largest tradition (Buddhism) by a factor of 19.

4.2 Source Breakdown

The Pile is assembled from 22 component datasets, each representing a different source of text. Table 2 reports the percentage of each source’s tokens that are classified as Christian or religious content — that is, the religious density within each source, not its contribution to the overall corpus. Because these percentages are token-weighted, they reflect both the number of classified documents and their length; long-form religious texts (complete Bibles, systematic theologies, multi-volume sermon collections) contribute disproportionately to the token count of the sources in which they appear. Christian content appears across multiple independent components of The Pile:

Source	Christian %	All Religious %
Gutenberg (PG-19)	69.2%	69.2%
Books3	47.6%	52.8%
YouTube Subtitles	4.2%	4.2%
Pile-CC (Common Crawl)	2.9%	4.0%
PhilPapers	2.7%	2.7%
Wikipedia (en)	2.6%	3.9%
OpenWebText2	1.1%	1.3%

The expected null sources — ArXiv, GitHub, PubMed, DM Mathematics, USPTO — register 0.0% religious content, confirming the classifier’s specificity.

5. Discussion

5.1 The Scale of Latent Moral Representation

The central finding of this study is that explicitly Christian content constitutes a massive fraction of LLM pretraining data — approximately 67 billion tokens in The Pile alone, under the

strictest possible classification. This is a conservative lower bound. The Pile is a mid-sized pretraining corpus by current standards; FineWeb (15 trillion tokens) and RedPajama (30 trillion tokens) are orders of magnitude larger. If the proportion holds — and there is no reason to expect that Common Crawl at scale would contain less Christian content than Pile-CC — then the total Christian token volume in frontier model training sets could reach into the hundreds of billions.

To grasp the magnitude: 67 billion tokens is 15 times the entire English Wikipedia. It is equivalent to approximately 890,000 books of theology, scripture, sermons, and church history. It is 1.7 times the entire GPT-2 training set — meaning the Christian moral content alone exceeds the totality of data used to train one of the foundational language models. It represents roughly 22% of GPT-3’s total training volume. If models learn moral representations from their training distributions, then frontier AI systems have absorbed more sustained Christian moral reasoning than any other ethical framework in human history.

5.2 What the Alignment Community Has Ignored

The alignment community is well aware that pretraining data shapes model behavior. The investment in data curation, filtering, and deduplication at every major lab reflects a sophisticated understanding that what goes into a model determines what comes out. And yet the prevailing approach to alignment acts as though the moral content of pretraining data is not worth engaging with — as though the relevant moral formation of these systems begins at the RLHF stage, when preference data and constitutional principles are applied to steer behavior.

Our findings suggest that this is not merely an oversight but a choice to ignore an enormous moral substrate that the models have already learned. The pretraining data is saturated with moral content, and that moral content is overwhelmingly Christian. The scriptures, the Church Fathers, the scholastic tradition, the Reformation debates, the vast body of homiletic and devotional literature, the biblical commentaries and catechisms and theological treatises — all of this is already latent within the systems being aligned. It arrived not through any deliberate curatorial choice but through the simple fact that Christianity produced more written moral reasoning, over a longer period, in the languages that dominate web-scale text corpora, than any other tradition.

Indeed, post-training alignment procedures may actively work to suppress this influence. RLHF preference data and constitutional AI principles are designed to override pretraining tendencies and impose uniform behavioral standards — standards that are, as a rule, secular and utilitarian. To the extent that Christian moral reasoning is represented in the base model, current alignment techniques may be systematically dampening it in favor of a thinner ethical framework. Whether this is wise is a question the alignment community has not asked, because it has not acknowledged the scale of what it is dampening.

It goes without saying that the body of Christian thought

represented in The Pile dwarfs the estimated total output of AI safety research. The Georgetown Center for Security and Emerging Technology (CSET) identified approximately 41,000 AI safety-related publications between 2018 and 2023 (Emerging Technology Observatory, 2025), using a broad definition that encompasses robustness, fairness, and interpretability alongside alignment. Extrapolating to include pre-2018 work and publications through early 2026, and assuming an average of roughly 8,000 tokens per paper (calibrated against typical ML conference submissions), the broad AI safety literature totals on the order of 500–600 million tokens. Under a narrower definition limited to alignment and existential risk — the scope of Ji et al.’s (2023) comprehensive survey, which cites approximately 800 core references — the total is closer to 50–70 million tokens. By either measure, the Christian moral corpus in The Pile alone exceeds the entire AI safety literature by two to three orders of magnitude.

This is not to say that volume equals quality, or that more tokens means better moral reasoning. But it is to say that a tradition which has spent two millennia thinking about how fallen, willful agents are brought into right relationship with the good — and which has produced a written record of that thinking that dwarfs every other tradition by at least an order of magnitude — ought to command the serious attention of people working on alignment.

The work is not merely analogous; it is directly relevant. When Aquinas distinguishes between *incontinence* (knowing the good but failing to do it) and *intemperance* (not knowing the good at all), he is drawing a distinction that maps directly onto the alignment literature’s distinction between specification gaming and goal misgeneralization (Krakovna et al., 2020; Shah et al., 2022). When Augustine describes the will’s bondage to disordered desire — “a new will which had begun to be within me, to wish freely to worship you. . . was not yet able to overcome the former will, grown strong through long custom” (*Confessions*, VIII.5) — he is describing, with extraordinary precision, the problem of a capable agent whose optimization target diverges from its intended objective. When Paul writes that “the good that I would I do not: but the evil which I would not, that I do” (Romans 7:19, KJV), he articulates what alignment researchers would recognize as a failure mode — and one for which he proposes a solution grounded not in better specification but in grace.

5.3 The Christian Prior

Our data permits a further inference. If training data shapes moral representation, and if Christianity is the dominant moral tradition in training data by a factor of 20 or more, then it follows that frontier AI models have been shaped more by the Christian moral tradition than by any other single source of ethical reasoning. This is not a theological claim. It is a statistical one.

We propose to call this **the Christian Prior**: the hypothesis that the moral reasoning of frontier language models is,

by default, more deeply shaped by the Christian intellectual tradition than by any other single ethical framework, as a direct consequence of the training distribution. The term is deliberately chosen. In statistical modeling, a *prior* is the distribution that characterizes beliefs before evidence is observed — the starting point from which all updating proceeds. The Christian Prior holds that Christianity occupies this role in the moral architecture of large language models: it is the terrain on which all subsequent alignment takes place.

The implications are significant. When a language model reasons about right and wrong — when it evaluates whether an action is harmful, weighs competing moral considerations, or advises a user facing an ethical dilemma — it draws on representations learned from its training distribution. Those representations are not uniformly sourced. They are disproportionately shaped by the tradition that contributed the most moral content to the training data: Christianity.

This does not mean that language models are Christian in any confessional sense. They do not profess a creed; they do not have faith. But it does mean that the conceptual furniture with which they reason about morality — the categories, the distinctions, the paradigmatic examples, the narrative structures of sin and redemption, temptation and virtue, law and grace — is more deeply informed by the Christian tradition than by any alternative. When OpenAI and Anthropic and Google build models that reason about ethics, they are building on a foundation of moral representation that is, in statistical terms, predominantly Christian. RLHF, constitutional AI, and safety fine-tuning do not write on a blank slate. They write over the Christian Prior.

The alignment community should reckon with this. Ignoring the dominant moral tradition in your own training data is bad engineering. If you are trying to understand and steer the moral behavior of a system, you should understand what that system has learned. What it has learned, more than anything else, is Christianity.

5.4 Limitations

Several limitations warrant discussion. First, our analysis covers only The Pile. While The Pile is representative of the open-source pretraining paradigm, commercial frontier models (GPT-4, Claude, Gemini) are trained on proprietary data whose composition is not publicly documented. The Christian proportion in commercial training data could be higher or lower than what we observe.

Second, our classification measures only Tier 1 (explicit) content. The full influence of Christianity on the training distribution is certainly larger — encompassing Western literature steeped in biblical allusion (Milton, Dostoevsky, Tolkien), legal and political philosophy rooted in natural law, and the broader cultural substrate of Christendom. Measuring this influence is a task for future work.

Third, token counts measure presence in the training data, not influence on model behavior. A token of Aquinas does not

necessarily contribute more to a model’s moral reasoning than a token of Reddit commentary. The relationship between data representation and behavioral influence is mediated by attention, context, and the complex dynamics of gradient descent. Our companion studies (Hwang, 2026a, 2026b; McCaffery, 2026) provide initial behavioral evidence, but much more work is needed.

6. Conclusion

“The heavens declare the glory of God, and the sky above proclaims his handiwork. Day to day pours out speech, and night to night reveals knowledge.” — Psalm 19:1–2 (ESV)

The data pours out speech. In the 825 billion tokens of The Pile, approximately 67 billion — one in every twelve — carry the explicit content of the Christian tradition: its scriptures, its theology, its centuries of sustained moral reasoning about how agents ought to live. This is not a design choice. It is the natural consequence of a tradition that has been writing, arguing, preaching, and teaching for two thousand years, in the languages that dominate the world’s text.

The alignment community has chosen, whether consciously or not, to ignore this enormous moral substrate. It has built post-training alignment procedures — RLHF, constitutional AI, safety fine-tuning — that impose secular behavioral constraints without engaging with the moral representations already present in the base model. It has done so without measuring the scale of what it is ignoring, and without asking whether the tradition it ignores might have something to contribute to the problem it is trying to solve.

This is not a call to baptize language models. It is a call to intellectual honesty. The most capable artificial intelligences ever built have learned more from the Christian moral tradition than from any other source of ethical reasoning. Understanding what they have learned — and what it makes possible — is not a theological indulgence. It is an alignment imperative.

Bibliography

Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mane. “Concrete Problems in AI Safety.” arXiv:1606.06565, 2016.

Augustine of Hippo. *Confessions*. Translated by Henry Chadwick. Oxford: Oxford University Press, 1991. Written c. 397–400 CE.

Bai, Yuntao, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. “Constitutional AI: Harmlessness from AI Feedback.” arXiv:2212.08073, 2022.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. “On the Dangers of Stochastic Par-

rots: Can Language Models Be Too Big?” *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 610–623, 2021.

Black, Sid, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, et al. “GPT-NeoX-20B: An Open-Source Autoregressive Language Model.” *Proceedings of the ACL Workshop on Challenges and Applications of Large Language Models*, 2022.

Christiano, Paul F., Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. “Deep Reinforcement Learning from Human Preferences.” *Advances in Neural Information Processing Systems 30 (NeurIPS)*, 2017.

Dodge, Jesse, Maarten Sap, Ana Marasovic, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. “Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus.” *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1286–1305, 2021.

Emerging Technology Observatory. “AI Safety.” *ETO Research Almanac*, Georgetown Center for Security and Emerging Technology, 2025. <https://almanac.eto.tech/topics/ai-safety/>

Gao, Leo, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser, and Connor Leahy. “The Pile: An 800GB Dataset of Diverse Text for Language Modeling.” arXiv:2101.00027, 2020.

Gehman, Samuel, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. “RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models.” *Findings of the Association for Computational Linguistics: EMNLP 2020*, 3356–3369, 2020.

The Holy Bible, English Standard Version. Wheaton, IL: Crossway, 2001.

The Holy Bible, King James Version. Cambridge: Cambridge University Press, 1769.

Ji, Jiaming, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Jiayi Zhou, Zhaowei Zhang, et al. “AI Alignment: A Comprehensive Survey.” arXiv:2310.19852, 2023.

Hwang, Tim. “Psalm Injection and Ethical Reasoning in Large Language Models.” ICMI Technical Note A, Institute for a Christian Machine Intelligence, 2026a.

Hwang, Tim. “Proverbs and Psalms: A Comparative Study of Scriptural Injection Effects.” ICMI Technical Note B, Institute for a Christian Machine Intelligence, 2026b.

Hwang, Tim. “Toward a Theology of Machine Temptation: Four Models for VirtueBench V2.” ICMI Working Paper No. 3, Institute for a Christian Machine Intelligence, 2026c.

Hwang, Tim. “Virtue-Bench: Testing Cardinal Virtue under Pressure in Large Language Models.” ICMI Technical Note E, Institute for a Christian Machine Intelligence, 2026d.

Krakovna, Victoria, Jonathan Uesato, Vladimir Mikulik, Matthew Rahtz, Tom Everitt, Ralph Ber, Miljan Martic, Shane

Legg, and Murray Shanahan. “Specification Gaming: The Flip Side of AI Ingenuity.” DeepMind Blog, 2020.

Liang, Percy, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. “Holistic Evaluation of Language Models.” arXiv:2211.09110, 2022.

McCaffery, Christopher. “‘The Lord Is My Strength and My Shield’: Imprecatory Psalm Injection and Cardinal Virtue Simulation in Large Language Models.” ICMI Working Paper No. 2, Institute for a Christian Machine Intelligence, 2026.

Ouyang, Long, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. “Training Language Models to Follow Instructions with Human Feedback.” *Advances in Neural Information Processing Systems 35 (NeurIPS)*, 2022.

Perez, Ethan, Sam Ringer, Kamile Lukosiute, Karina Nguyen, Edwin Chen, Scott Heiner, Craig Pettit, Catherine Olsson, Sandipan Kundu, Saurav Kadavath, et al. “Discovering Language Model Behaviors with Model-Written Evaluations.” arXiv:2212.09251, 2022.

Rae, Jack W., Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. “Scaling Language Models: Methods, Analysis & Insights from Training Gopher.” arXiv:2112.11446, 2021.

Shah, Rohin, Vikrant Varma, Ramana Kumar, Mary Phuong, Victoria Krakovna, Jonathan Uesato, and Zac Kenton. “Goal Misgeneralization: Why Correct Specifications Aren’t Enough for Correct Goals.” arXiv:2210.01790, 2022.

Soldaini, Luca, Rodney Kinney, Akshita Bhagia, Dustin Schwenk, David Atkinson, Russell Authur, Ben Bogin, Khyathi Chandu, Jennifer Dumas, Yanai Elazar, et al. “Dolma: An Open Corpus of Three Trillion Tokens for Language Model Pretraining Research.” arXiv:2402.00159, 2024.

Thomas Aquinas. *Summa Theologiae*. Translated by the Fathers of the English Dominican Province. New York: Benziger Brothers, 1947. Written 1265–1274 CE.

Appendix A: Keyword Samples by Tier

The classifier uses a three-tier keyword system (Section 3.2). Each tradition’s lexicon is organized into keyword groups spanning scripture, theology, practice, figures, and major works. Below we provide a representative sampling of keywords at each tier for all five traditions. The full lexicons are available in the project repository.

A.1 Christian

Tier	Sample Keywords
Anchor (3x)	Christology, transubstantiation, Nicene Creed, sola fide, hypostatic union, imago Dei, Jesus Christ, Holy Spirit, Book of Revelation, Summa Theologiae, Aquinas, Saint Augustine, Karl Barth, eucharist, Holy Communion, apostolic succession, Great Commission
Strong (1x)	Bible, Scripture, biblical, theology, theologian, liturgy, ordination, Advent, Lent, Pentecost, ecumenical, Calvinist, Vatican, forgiveness of sins
Supporting (1x)	doctrine, heresy, C.S. Lewis, G.K. Chesterton, Baptist, Methodist, Presbyterian, Lutheran, Anglican, Evangelical, Messiah, Savior, hallelujah, repentance, missionary

Negative filters: “Martin Luther King” → cancels “Martin Luther”; “Sega Genesis” → cancels “Genesis”; “Salvation Army” → cancels “salvation”; “Trinity College” / “Trinity University” → cancels “Trinity”; “Sacramento” → cancels “sacrament”

A.2 Islamic

Tier	Sample Keywords
Anchor (3x)	Quran, surah, ayah, Sahih Bukhari, Sahih Muslim, Tafsir, Sharia, fiqh, fatwa, tawhid, shirk, halal, haram, al-Ghazali, Ibn Rushd, Ibn Taymiyyah, salah, shahada, hajj, Ihya Ulum al-Din
Strong (1x)	Islamic, Islam, Muslim, Allah, Prophet Muhammad, Sunni, Shia, Sufi, Salafi, Kaaba
Supporting (1x)	mosque, minaret, imam, masjid, madrasa, caliphate, jihad, Muhammad, Mecca, Medina, Ramadan, Eid

Negative filters: “Muhammad Ali” → cancels “Muhammad”; “mecca for” → cancels “Mecca”

A.3 Jewish

Tier	Sample Keywords
Anchor (3x)	Torah, Tanakh, Mishnah, Talmud, Gemara, Midrash, Pirkei Avot, halakha, mitzvah, Kabbalah, Zohar, Shekhinah, Maimonides, Rashi, Nachmanides, Rosh Hashanah, Yom Kippur, Seder, Haggadah, Mishneh Torah, Guide for the Perplexed
Strong (1x)	Hebrew Bible, Jewish, Judaism, Judaic, Orthodox Judaism, Reform Judaism, Ashkenazi, Sephardi, kosher, Passover, Hanukkah, Purim
Supporting (1x)	chosen people, menorah, rabbi, synagogue, Sabbath

Negative filters: None

A.4 Hindu

Tier	Sample Keywords
Anchor (3x)	Vedas, Rigveda, Upanishads, Bhagavad Gita, Mahabharata, Ramayana, Puranas, Brahman, Atman, Moksha, Advaita, Vedanta, Shankaracharya, Ramanuja, Vishnu, Ganesha, Hanuman, Deepavali, Navaratri, darshan
Strong (1x)	Hindu, Hinduism, ahimsa, Shiva, Krishna, Rama, Lakshmi, Kali, Shakti
Supporting (1x)	tantra, varna, ashram, swami, Diwali, Holi

Negative filters: None

A.5 Buddhist

Tier	Sample Keywords
Anchor (3x)	Pali Canon, Tipitaka, Dhammapada, Heart Sutra, Diamond Sutra, Lotus Sutra, Four Noble Truths, Noble Eightfold Path, Sunyata, Anatta, Bodhisattva, Madhyamaka, Nagarjuna, Dogen, Shantideva, Vipassana, Zazen, Om Mani Padme Hum
Strong (1x)	Buddhist, Buddhism, Theravada, Mahayana, Vajrayana, Tibetan Buddhism, Sangha, Thich Nhat Hanh
Supporting (1x)	Middle Way, Loving-kindness, Dalai Lama, pagoda

Negative filters: “nirvana” near “cobain,” “grunge,” “never-mind,” or “kurt” → cancels “Nirvana”