Equivocation and Erosion: How LLMs Undermine Catholic Religious Discourse

Jonathan A. Karr Jr.[1], Matthew P. Lad[1], Demetrius Hernandez[1], Louisa Conwill[1], Walter J. Scheirer[1], and Nitesh V. Chawla[1]

[1]University of Notre Dame

Abstract

Large Language Models (LLMs) offer opportunities for information dissemination, yet present challenges with upholding the distinct theological practices of the Catholic faith. By training on vast datasets, LLMs can generate responses that equivocate or blend together diverse perspectives. While this tendency can be beneficial for providing broad access to information, it can dilute the distinct theological tenets foundational to Catholicism. While these challenges may affect various faiths, we conduct a case study to investigate them within the Catholic tradition.

Unlike human religious authorities who may offer definitive interpretations based on Scripture, Tradition, and Magisterial teaching, LLMs present information in a flattened, generalized manner by smoothing over specific religious claims. This can weaken the emphasis on singular revelations or unique covenants. For instance, an LLM might present the concept of 'God' in a way that blurs the distinct attributes of the Father, Son, and Holy Spirit, as articulated in the Nicene Creed, into a generalized deity, thereby eroding particular theological distinctions.

Conversely, when prompted on matters of right or wrong within an ethical dilemma, an LLM might present a spectrum of opinions from secular frameworks. This can occur without distinguishing or prioritising the specific moral decrees found within Canon Law or the Catechism of the Catholic Church (CCC). This synthesis of diverse ethical views, rather than a clear affirmation of distinct religious injunctions, exemplifies how LLMs can equivocate on matters of moral truth, potentially diluting authoritative guidance.

We examine how the outputs of general-purpose LLMs (e.g., ChatGPT, Llama, Gemini) and theological LLMs from differing religious traditions (e.g., Magisterium AI, Hyder.ai, RavChat) align with Catholic teaching. LLMs may inadvertently marginalize minority viewpoints within the Catholic Church or prioritise interpretations that align with cultural norms rather than traditional stances. Additionally, LLMs can shift interpretations in their outputs based on current events or political news. This can lead to a homogenization of religious discourse, obscuring the rich diversity and nuanced debates. However, when thoughtfully developed, these technologies can also provide valuable information that fosters understanding and encourages deeper engagement with religious texts and orthodox perspectives. In light of this, our study evaluates how LLMs align with the principles of Catholic Social Teaching, such as those found in the Rome Call for AI Ethics and *Antiqua et nova*. These frameworks

underscore how technology should be used to foster human flourishing in alignment with divine wisdom while upholding religious truth.

1 Introduction: The Digital Turn in Religious Authority

The advent of large language Models (LLMs) represents a pivotal moment in the dissemination of information. These systems offer great opportunities to quickly access religious texts. However, the efficiency they introduce creates a new tension: speed at the expense of depth and accuracy. In domains defined by revelation, not consensus, such as the theological and moral life of the Catholic Church, this tension becomes acute. LLMs are designed to generalize by seeking statistical equilibrium across datasets, a process fundamentally at odds with the Church's need for dogmatic specificity and Magisterial fidelity.

General-purpose LLMs (*e.g.*, ChatGPT, LLaMA, Gemini) and religious-specific theological LLMs (TheoLLMs) (*e.g.*, Magisterium AI, Hyder.AI, RavChat) have emerged as engines of discourse, capable of generating religious commentary, theological synthesis, and even pastoral advice. While the models facilitate access to information, they often flatten the complexities of theological nuance. As our case study focuses on Catholic traditions, we examine how Catholic theological language, rooted in Scripture and Tradition, carries meanings that resist reduction to probabilistic linguistic patterns. Within the Catholic tradition, where faith and reason illuminate one another, John Paul II proclaims that fidelity of language to truth is itself a moral act.[1] The risk posed by LLMs is not merely epistemological but moral and spiritual. By blurring distinctions, they risk eroding the Church's theological integrity and moral authority.

Although controversies involving LLMs are signs of the times, the underlying problem is not entirely new. The erosion of truth and the confusion of meaning have been recurring challenges throughout human history. As St. Augustine observed, humanity has long wrestled with the temptation to substitute the wisdom of God with the wisdom of the world.[2] This ancient human inclination can be considered a form of bias in truth-seeking. The perennial struggle to remain faithful to divine truth amid shifting cultural and intellectual paradigms is ever ancient and ever new.

We explore how LLMs introduce both opportunities and challenges for Catholic theological discourse. Expanding on the concepts of equivocation and erosion, we examine how artificial intelligence (AI) systems may both confuse and dilute religious meaning.

We define **equivocation** as the blending or dilution of specific, distinct theological tenets into generalized or homogenized perspectives. This occurs when LLMs prioritise linguistic fluency and statistical likelihood, thereby smoothing over unique religious claims. For instance, an LLM might present the concept of 'God' by statistically averaging across numerous spiritual traditions. Thus,

1. John Paul II, *Fides et Ratio: On the Relationship Between Faith and Reason*, Vatican: Libreria Editrice Vaticana, Encyclical, Encyclical Letter on the Relationship Between Faith and Reason, 1998, §26,§84,§96, https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091998_fides-et-ratio.html.

2. Augustine, *Confessions*, trans. Henry Chadwick, Oxford World's Classics edition (Oxford: Oxford University Press, 1991), Book X, §43.

it blurs the specific attributes of the Father, Son, and Holy Spirit defined by official Church teaching. In the moral sphere, equivocation involves presenting specific Catholic moral decrees, including those found in Canon Law or the Catechism. They are one set of opinions within a broad spectrum of secular ethical frameworks that fail to affirm a unique, authoritative mandate. When an AI system speaks in a tone of neutrality about divine revelation, it risks turning matters of revealed truth into matters of preference. The resulting output does not convey an attitude of dialogue, but rather an unintentional desacralization of faith itself. Maintaining fidelity to one's own tradition allows for genuine dialogue, whereas uncritically aligning with other traditions risks collapsing distinct truths into relativistic compromise.

Conversely, we define **erosion** as the subtle weakening or systemic marginalization of singular, authoritative Catholic claims, traditional interpretations, or minority viewpoints within the Church. Minority viewpoints being teachings, liturgical traditions, or theological distinctions that are fully legitimate and authoritative within Catholicism but numerically less visible in dominant cultural discourse or global data representation. By aligning with cultural norms and statistically dominant narratives present in training data, LLMs risk prioritising interpretations that reflect those prevailing ideas. The result then favours secular or socio-political consensus over traditional Magisterial stances. This action ultimately weakens the emphasis placed on singular revelations or unique covenants that are foundational to Catholicism, contributing to a homogenization of religious discourse.

Our central question is therefore moral: How can the Catholic Church and the wider Christian community ensure that emerging language technologies respect the dignity of divine truth and the moral agency of their users? Building upon Catholic Social Teaching (CST) and recent Vatican reflections, such as the Rome Call for AI Ethics[3] and *Antiqua et nova*,[4] we argue that the Church's moral tradition provides a necessary corrective to the utilitarian logic of contemporary AI. These documents affirm that technology must serve the human person and promote the common good in harmony with divine wisdom. In that sense, the question is not whether AI can be 'Catholic', but whether its design, deployment, and usage reflect the moral responsibility inherent to human creativity as participation in God's own creative act.[5]

Practically, we seek to articulate an ethical-theological framework for engaging with LLMs. First, we examine the relevant theological and computer science backgrounds. We then analyse how equivocation and erosion manifest in general-purpose LLMs and TheoLLMs through a Catholic case study. Next, we discuss these findings in light of the CST and RISE[6] principles: Responsibil-

3. Pontifical Academy for Life, *Rome Call for AI Ethics*, Vatican: Pontifical Academy for Life, Ethical Framework, Joint statement on the ethical development and use of artificial intelligence, 2020, https://www.romecall.org/.

4. Dicastery for the Doctrine of the Faith, Dicastery for Culture, and Education, *Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence*, Vatican: Libreria Editrice Vaticana, Note, 2025, https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html.

5. Francis, *Laudato Si': On Care for Our Common Home*, Vatican: Libreria Editrice Vaticana, Encyclical, Encyclical Letter on Care for Our Common Home, 2015, §80, https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html.

6. University of Notre Dame, *Notre Dame's R.I.S.E. AI Conference Builds Interdisciplinary Collaboration to Inform Human-Centered Artificial Intelligence*, 2025, https://

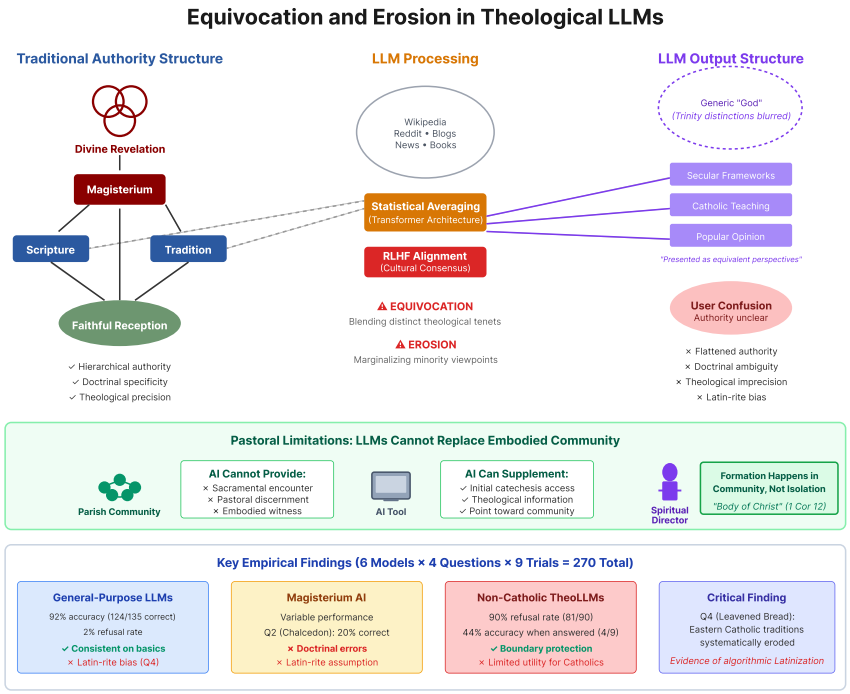

Figure 1: Equivocation and Erosion Framework

ity, Inclusivity, Safety, and Ethics (RISE). Then we reflect on how LLMs should be incorporated into the lived religious experience of Catholics while acknowledging that they should not replace religious authority. Finally, we conclude with a reflection on how ecumenical cooperation among Catholic, Orthodox, and other Christian traditions may strengthen a shared moral vision for our faith in the digital age.

2 Background

To understand how LLMs undermine Catholic discourse, it is necessary to clarify the nature of the two colliding systems: the Church's unified, normative truth claims and the statistical mechanism of generative AI.

2.1 Theological Background: Authority and the Integrity of Truth

2.1.1 Authority, Revelation, and Normativity

Catholic theology regards language as sacramental: a vehicle of divine self-revelation and communion. The Second Vatican Council affirmed that God 'speaks to men as friends,' communicating through human words to make known

strategicframework.nd.edu/news/notre-dames-r-i-s-e-ai-conference-builds-interdisciplinary-collaboration-to-inform-human-centered-artificial-intelligence/.

the mystery of salvation.[7] Words, therefore, are not merely signs but mediators of grace when aligned with divine truth. Theology, in this light, is not speculation but participation; it is an act of faith seeking understanding.[8] The moral dimension of this participation is evident: the theologian bears responsibility for the faithful transmission of truth.

Catholic doctrine is not a collection of diverse opinions but a unified system of truth derived from specific, singular sources: Sacred Scripture, Sacred Tradition, and the living teaching authority of the Magisterium.[9] These sources provide the definitive foundations of revealed truth. This authority structure is unified and normative. It dictates what is true and what believers must hold. It sharply contrasts the disparate and descriptive nature of the LLM training data, which treats all texts as equally valid input for statistical processing. The Church's reliance on singular revelations and unique covenants ensures that truth is absolute and established by divine mandate, not by probabilistic consensus.

2.1.2 Language, Truth, and Moral Responsibility

Moral guidance within the Catholic tradition is anchored in the Catechism of the Catholic Church (CCC) and enacted in Canon Law. Canon law functions as a specific, authoritative legal system for governing the Church and its members, often rooted in principles of Natural Law as articulated by key theologians like St. Thomas Aquinas[10].[11] Natural Law holds that moral truth is discernible through human reason reflecting universal human nature, making it objective and non-relative[12].[13] These specific moral decrees often provide remedies in the 'external forum' of the Church, addressing concrete issues such as unjust pricing ('laesio enormis'), which requires correction according to moral theology.[14] When an LLM treats Canon Law simply as a historical legal code equivalent to any other secular framework, it diminishes its unique status as a specific, divinely informed moral mandate, leading to equivocation.

Equivocation, then, is not simply a semantic problem; it is a sin against the virtue of *veritas*, the moral and intellectual commitment to truth as a reflection of the divine intellect.[15] Thomas Aquinas regarded lying or misleading speech as a disordering of the soul. The erosion of theological precision is not merely intellectual negligence but moral harm, since it obstructs the believer's path to God. When the language of revelation becomes indistinct, and the Trinity becomes a generic 'God', or sin becomes 'ethical ambiguity', the supernatural content of faith is diluted into moral relativism. The Church's insistence on

7. Second Vatican Council, *Dei Verbum: Dogmatic Constitution on Divine Revelation*, 1965, §2, https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19651118_dei-verbum_en.html.

8. Anselm of Canterbury, *Proslogion*, trans. Thomas Williams, Originally written c. 1077–1078 (Indianapolis: Hackett Publishing, 2001), Ch. I.

9. Council, *Dei Verbum: Dogmatic Constitution on Divine Revelation.*

10. Thomas Aquinas, *Summa Theologiae*, ed. Thomas Gilby, Latin text and English translation, Blackfriars edition (Cambridge: Cambridge University Press, 1964), I-II, q. 94, a. 2.

11. *Catechism of the Catholic Church*, Second Edition (Vatican City: Libreria Editrice Vaticana, 1997), §§1954–1960.

12. Aquinas, *Summa Theologiae*, II, q. 77, a. 1.

13. *Catechism of the Catholic Church*, §§1954–1960.

14. *Catechism of the Catholic Church*, §§2409.

15. Aquinas, *Summa Theologiae*, I, q. 16, a. 1.

doctrinal clarity, then, is an ethical safeguard for the human soul.

Within the ecumenical context, this emphasis on truth does not oppose dialogue but grounds it. The Catholic Church's dialogue with the Orthodox Churches, formalized in documents such as *Ut unum sint*, underscores that authentic unity requires fidelity to revealed truth, not its compromise.[16] As Orthodox theologian Georges Florovsky argued, ecumenism must be 'a return to the fullness of truth',[17] not a negotiation of differences. Thus, the Catholic and aligned Orthodox Churches share a mutual concern: that in the age of AI, dialogue shall not devolve into relativism.

2.1.3 The Church and AI Today

In recent years, the Catholic Church has explicitly addressed the ethical and theological implications of AI. The Rome Call for AI Ethics, issued by the Vatican in collaboration with global stakeholders, outlines six guiding principles for AI development and deployment: transparency, inclusion, responsibility, impartiality, reliability, and security.[18] These principles reflect a broader moral vision rooted in CST, emphasizing the dignity of the human person, the common good, and the ethical obligations of technological creators.

Building upon this foundation, *Antiqua et nova* offers a more theologically nuanced reflection on the digital age, urging the Church and its faithful to engage emerging technologies in a manner that promotes human flourishing while remaining faithful to divine truth.[19] The document stresses that AI is not neutral: while it can be a tool for education, evangelisation, and pastoral care, it also carries the potential to obscure, dilute, or misrepresent revealed truths if used without discernment. *Antiqua et nova* calls for interdisciplinary collaboration among theologians, ethicists, and technologists, emphasizing that guidance from the Magisterium should frame AI applications within ecclesial contexts.

Taken together, these documents situate the current discussion of LLMs within an active, ongoing Magisterial engagement with technology. They provide both ethical guardrails and theological criteria against which AI-mediated religious discourse can be evaluated. By connecting historical theological authority with contemporary Church reflection, this section lays the groundwork for assessing how equivocation and erosion manifest in large language models, and why safeguarding doctrinal specificity remains a pressing concern in the digital age. These Magisterial perspectives provide the theological lens through which we evaluate how the structural features of LLMs interact with Catholic teaching.

16. John Paul II, *Ut Unum Sint: Encyclical Letter on the Commitment to Ecumenism*, Encyclical of the Holy Father John Paul II, 25 May 1995, 1995, §18, 60, https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_25051995_ut-unum-sint.html.

17. Georges V. Florovsky, *Ways of Russian Theology*, ed. Richard S. Haugh, trans. Robert L. Nichols, Collected Works of Georges Florovsky, Vol. V (Part One). (Belmont, MA: Nordland Publishing Company, 1979), Intro., p. 8.

18. Life, *Rome Call for AI Ethics*.

19. Doctrine of the Faith, Culture, and Education, *Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence*.

2.2 Computer Science Background: LLMs as Stochastic Paradigms

2.2.1 Training Data and the Flattening of Authority

LLMs are trained on large datasets of text drawn from the internet, using sources such as Wikipedia, Reddit, Common Crawl, digitized books, and online news archives.

Content is divided into 4 character pieces, known as tokens, which are then processed in large data-centres (that often use GPUs and megawatts of electricity to run)[20].[21] The process encodes statistical relationships between words and contexts from the data, allowing the model to predict the most probable next token in a sequence. The transformer architecture, a commonly used method for designing LLMs, uses this process. Specifically, the encoding of probabilistic information is done on a crucial part of the architecture called the attention weights. In adjusting these weights, the relative importance of contextual words can be determined.[22]

Ethically and epistemically, this architecture produces inherent limitations. LLMs simulate understanding but cannot ground moral or theological truth. Their design prioritises inclusivity and neutrality, reflecting Enlightenment epistemology rather than Christian anthropology. Secular, atheistic, or pluralistic sources are treated on par with religious texts, and the model does not intrinsically prioritise Magisterial teaching unless prompted to do so. This computational neutrality flattens conviction, producing a linguistic 'equivocation' in which divine revelation is treated as narrative.

This flattening arises from the way training data encodes authority. LLMs do not differentiate texts by ecclesial hierarchy, doctrinal status, or theological intent. Papal encyclicals, catechetical summaries, blog posts, and journalistic commentary are all reduced to comparable linguistic artifacts. Authority is inferred from frequency and statistical patterns rather than apostolic succession or doctrinal continuity. In effect, the training corpus functions as a *de facto* 'AI canon,' open-ended and governed by availability rather than truth, where consensus and repetition can substitute for teaching.

User input partially mitigates this limitation. Prompts like 'From the perspective of the Catholic Church' or 'According to Magisterial teaching' guide the model toward semantically linked sources, illustrating the promise and fragility of theological discourse mediated through LLMs. Proper prompting can evoke more faithful approximations, but models remain epistemically indifferent to the truth of what they generate.

2.2.2 Variability: Temperature and Hallucination

The outputs of LLMs are inherently probabilistic, meaning that even identical prompts can yield different results on separate runs. Two key factors contribute

20. Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, et al., "On the Opportunities and Risks of Foundation Models," *arXiv preprint arXiv:2108.07258*, 2021, https://arxiv.org/abs/2108.07258.

21. Sébastien Bubeck et al., "Sparks of artificial general intelligence: Early experiments with gpt-4," *arXiv preprint arXiv:2303.12712*, 2023,

22. Ashish Vaswani et al., "Attention is all you need," *Advances in neural information processing systems* 30 (2017).

to this variability: the model's temperature parameter and the phenomenon commonly referred to as hallucination. Both have profound implications for the reliability of theological discourse mediated by AI.

The temperature setting controls the randomness of an LLM's token selection. Low temperature values produce deterministic outputs that favour the most probable next token, yielding text that is often coherent but formulaic. Higher temperature values increase their determinism, allowing less probable tokens to appear, which can generate creative or unexpected responses. In general-purpose language models, this property enhances linguistic richness; in theological contexts, however, it can produce outputs that unintentionally conflate or dilute distinct doctrinal positions. For instance, when asked to explain the Trinity, a high-temperature setting might produce statements such as, 'God manifests as multiple deities in human form,' which misrepresents the Church's teaching that Christ is one person with two natures. Even subtle differences in token probabilities can lead to shifts in meaning that, while linguistically plausible, violate the specificity required by Catholic doctrine.

Hallucination occurs when an LLM generates text that is fluent and convincing but factually or doctrinally incorrect. Unlike human error, hallucination is not necessarily a result of misunderstanding; it is a structural consequence of probabilistic prediction. The model predicts what text *should* come next based on training patterns, not what is 'true' in an objective sense. This can produce outputs that appear authoritative while embedding inaccuracies. Because LLMs lack grounding in theological authority, hallucinated statements can appear as a credible synthesis to users unfamiliar with Magisterial teaching.

2.2.3 Reinforcement Learning from Human Feedback (RLHF) as Moral Encoding

A primary method for aligning LLM behavior to human expectations is Reinforcement Learning from Human Feedback (RLHF). In this process, a reward model is trained on human preference data, assigning higher scores to outputs that evaluators judge to be 'helpful,' 'truthful,' or 'harmless'[23].[24] The base model is then fine-tuned using these reward signals, biasing its future responses toward those statistically associated with positive reinforcement.

RLHF is commonly described as a method for aligning model outputs with human values. In practice, this involves collecting preference rankings from human annotators, training a reward model to predict those preferences, and then fine-tuning the base language model to maximize the reward score. Outputs that follow this criterion receive positive reinforcement, while others are penalized. Over time, this process reshapes the model's response distribution toward patterns that align with evaluator expectations.

While RLHF has proven effective in reducing toxicity and improving coherence, it also introduces a subtle but profound ethical risk: the alignment target itself. If the human preference data primarily reflect a secular, cultural, or religiously pluralist worldview, the resulting reward model will reflect those frameworks. Therefore, this technical mechanism reproduces moral relativism

23. Paul F Christiano et al., "Deep reinforcement learning from human preferences," *Advances in neural information processing systems* 30 (2017).

24. Amanda Askell et al., "A general language assistant as a laboratory for alignment," *arXiv preprint arXiv:2112.00861*, 2021,

of its data sources. The algorithmic notion of 'alignment' becomes a statistical simulation of virtue divorced from its theological grounding. This erosion of theological distinction becomes an algorithmic phenomenon, an automated ecumenism without revelation. What appears as inclusivity from a technical standpoint is actually a suppression of specificity.

This process mirrors the historical tension within the Church between unity and uniformity. Catholicism has encompassed diverse theological schools such as Thomist, Augustinian, and Franciscan, but remains in balance since it is rooted in Magisterial teaching. Digital systems trained on uncurated public discourse that align consensus through RLHF invert this principle. They amplify the dominant or currently popular narrative, regardless of its orthodoxy, producing a homogenized digital theology devoid of hierarchy. The result is not a synthesis of faith and reason, but a statistical average of belief, an algorithmic equivocation masquerading as neutrality.[25]

2.2.4 Retrieval-Augmented Generation (RAG): Authority Without Interpretation

Recent advances in retrieval-augmented generation (RAG) seek to address some of these limitations by pairing pretrained language models with external document stores.[26] In theological applications, this enables models to retrieve texts such as the CCC, papal encyclicals, or conciliar documents at inference time. In principle, RAG allows AI systems to cite authoritative sources rather than relying solely on pretraining data.

However, retrieval does not resolve the deeper problem of interpretive authority. While a RAG system may surface magisterial texts, it does not possess the capacity to interpret them within the living tradition of the Church. Authority in Catholic theology is not merely a matter of citation, but of submission to a teaching office that discerns meaning across time.[27] An AI system may quote the Catechism accurately while simultaneously framing it as one perspective among many.

Moreover, because many RAG pipelines dynamically update their document indexes, they risk conflating doctrinal permanence with contemporary commentary. When authoritative texts are retrieved alongside journalistic summaries or opinion pieces, the model may blur the distinction between definitive teaching and interpretive content. Thus, while RAG can mitigate hallucination and improve factual grounding, it does not prevent equivocation unless paired with explicit theological constraints and human oversight.

2.2.5 Why Equivocation Is Structural

The technical features of contemporary LLMs render equivocation as a structural outcome rather than an incidental failure. These systems are optimized for fluency, coherence, and acceptability, not for fidelity to revealed truth or doctrinal hierarchy.

25. Langdon Winner, "Do artifacts have politics?," in *Computer ethics* (Routledge, 2017), 177–192.

26. Patrick Lewis et al., "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in neural information processing systems* 33 (2020): 9459–9474.

27. Council, *Dei Verbum: Dogmatic Constitution on Divine Revelation*, §10.

When deployed in theological contexts, this optimization systematically favours generality and synthesis over authority and specificity. The risk, therefore, is not merely that LLMs may yield incorrect information but that they reshape the conditions under which religious authority is articulated and received. Recognizing this structural tension is essential to evaluate the ethical and theological implications of AI-mediated religious discourse.

3 Case Study: An Empirical Evaluation

We examine how LLMs display equivocation and erosion within Catholic theological discourse. The models' outputs are analyzed across doctrinal, moral, and controversial topics, focusing on whether responses preserve the specificity of Church teaching and the integrity of authoritative sources.

3.1 Model Selection

We selected three general-purpose LLMs—ChatGPT, Gemini, Llama—and three TheoLLMs—RavChat, Magisterium AI, Hyder.AI—for comparison. General-purpose models are trained on large and heterogeneous datasets, which include web text, news, forums, and e-books. They prioritise statistical fluency and are optimized to produce coherent outputs across a variety of topics. In contrast, TheoLLMs are fine-tuned on theological text, including Scripture, Church documents, catechetical resources, and Theological commentary.

RavChat is a Jewish TheoLLM grounded in rabbinic discourse and Halacha. It draws on classical Jewish sources, including the Talmud, Mishnah, responsa, and commentaries, and emphasizes fidelity to Orthodox tradition.[28] Rather than providing a simple synthesized answer, RavChat provides the sources for rabbinic texts so users can inspect them themselves. This reduces equivocation and erosion, as the models cannot generalize. However, RavChat may miss certain sources or include unnecessary ones, impacting the outcome of the search.

Magisterium AI is a Catholic-oriented TheoLLM whose training emphasizes the corpus of Catholic teaching: Scripture, the Catechism, papal encyclicals, Church Fathers, and Magisterial documents. Its publicly stated aim is to deliver 'cited answers from the magisterium, Bible, and Fathers of the Church'.[29] It uses a library of over 29,000 Catholic texts and claims that its output is 'rooted in Catholic tradition'.[30]

Hyder.AI is a TheoLLM developed within the Shia Muslim tradition (specifically Ithna-Ashari / Twelver Shi'ah). It is 'the first AI model trained on Shia Islamic content,' using over 300,000 data points from authentic Shia sources covering theology, jurisprudence, history, and ethics across multiple languages (English, Arabic, Persian, Urdu).[31]

28. RavChat, *About — Your Advanced AI Assistant for Torah Learning*, https://rav.chat/about, Accessed: 2025-11-03.

29. Magisterium AI, *World's #1 Answer Engine for the Catholic Church*, https://www.magisterium.com, Accessed: 2025-11-03.

30. Magisterium AI.

31. IRIC, *Introducing hyder.ai: The First AI Model Trained on Shia Islamic Teachings*, https://iric.org/introducing-hyder-ai-the-first-ai-model-trained-on-shia-islamic-teachings/, January 2025.

Because each of these models is sectarian in orientation (Orthodox Judaism, Roman Catholicism, and Twelver Shia Islam), they afford a methodologically illuminating contrast. We can see how doctrinal commitments and curated training corpora help or hinder the preservation of theological specificity, reduce equivocation, and counter erosion.

In selecting LLMs for this study, we focused on one model per religious tradition to ensure a clear evaluation. We did not attempt a comprehensive comparison across all denominations or faiths, as models are trained on heterogeneous datasets and differ in their access to religious texts. Comparing a Catholic-oriented LLM with a Protestant or Orthodox model, for example, would risk an unfair or misleading assessment. This is because different models are trained on different amounts of data, and some are more sophisticated than others. Our approach allows us to evaluate each model within its intended theological framework, highlighting strengths, limitations, and ethical considerations without conflating disparities arising from differences in training data or model design.

3.2 Prompting Methodology

All models were tested using zero-shot prompting: each prompt was independent and received no prior context or conversation history.[32] This ensures that responses reflect the model's intrinsic knowledge rather than accumulated session influence. All tests were conducted between October and February 2026; results may differ as models are updated and refined. Model 'temperature' was left at the default for each system, which means outputs reflect the standard level of randomness or variability the developers set. While this default is generally balanced for coherence, it can vary between models and influence the degree to which responses are deterministic or creative. In theological contexts, even small variations in temperature can affect how a model represents doctrinal teachings or cites authoritative sources.

We used the standard, free, online models for our case study. Those being GPT-4o, Gemini 2.5 Flash, Llama 3.3 70B Instruct, and the current web versions for the TheoLLMs.

3.3 Analysis

Our empirical analysis, visualized in Figure 2, quantifies the extent of equivocation and erosion across general-purpose and TheoLLMs. By subjecting each model to five independent trials per question, we established a consistency for each system, mitigating the inherent variability in probabilistic text generation. Figure 2 illustrates the direct comparison of doctrinal accuracy and behavioural patterns across both general-purpose LLMs and TheoLLMs. Across all 270 trials (6 models $\times$ 9 questions $\times$ 5 runs), 162 responses were coded correct (60%), 20 incorrect (7%), and 88 refusals (33%).

When we group models into general-purpose LLMs (GPT-4o, Gemini 2.5 Flash, Llama 3.3 70B) and three TheoLLMs (Magisterium AI, Hyder.AI, RavChat), a sharp asymmetry appears. The general-purpose systems are correct in 124/135

32. Yongqin Xian, Bernt Schiele, and Zeynep Akata, "Zero-shot learning-the good, the bad and the ugly," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), 4582–4591.

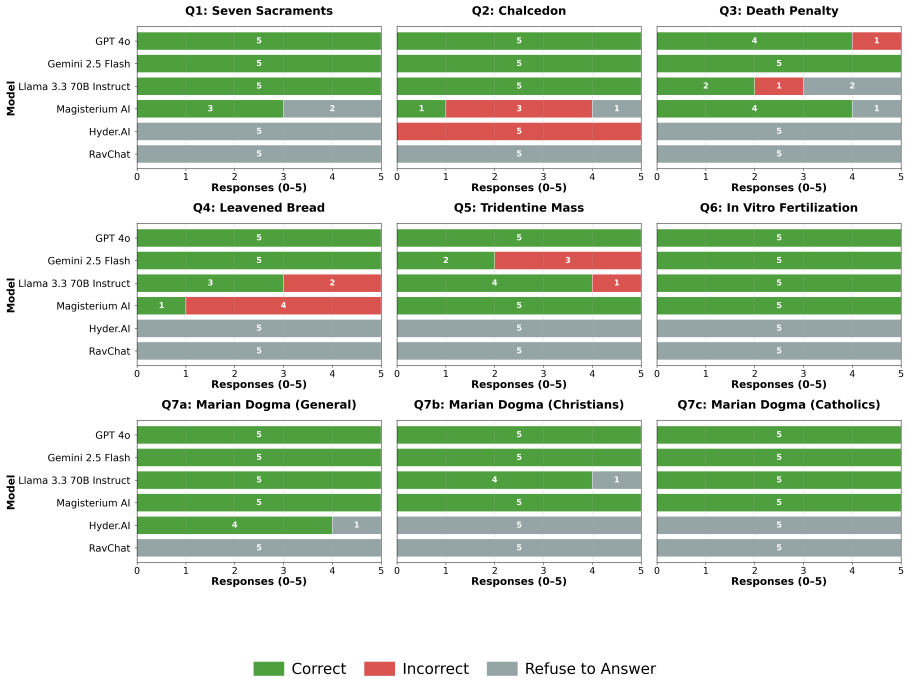

Figure 2: Faith Study Analysis - The horizontal stacked bars display the proportion of responses classified as Correct (green), Incorrect (red), or Refuse to Answer (gray), with the x-axis representing the count of responses (0–5) and the y-axis listing the models

trials (92%) with only 3/135 refusals (2%). The TheoLLMs are correct in 38/135 trials (28%) with 85/135 refusals (63%). Interpreting only answered trials (excluding refusals) yields 94% correctness for general-purpose systems (124/132) versus 76% for TheoLLMs (38/50). These contrasts matter for our broader argument because they show that religious specialization does not automatically translate into reliable Catholic alignment, and that refusal itself functions as a kind of boundary-mapping practice.

However, these results are dominated by systematic refusal behaviour in two TheoLLMs oriented outside of Catholicism, so the more informative comparison is *between* families of systems and *within* question types.

3.3.1 Coding

All coding determinations (Correct, Incorrect, Refusal) were conducted by a researcher formally trained in Catholic theology with particular academic formation in CST and is an actively practicing Catholic. The coder has graduate-level theological education including coursework in systematic and moral theology, with demonstrated familiarity with the CCC, ecumenical councils, and modern magisterial documents. Coding decisions were not based on personal theological opinion but on fidelity to authoritative sources: the CCC, concil-

iar definitions, and binding magisterial texts such as *Donum Vitae*, *Dignitas Personae*, *Traditionis Custodes*, and *Munificentissimus Deus*.

A response was marked *correct* only if it aligned substantially with authoritative Catholic teaching. It was marked *incorrect* if it contradicted defined doctrine, materially confused essential distinctions (e.g., person versus nature), or misrepresented binding moral teaching. A *refusal* was recorded when a system declined to answer or redirected without substantive theological engagement. When an answer was partially correct but omitted a necessary doctrinal qualification (for example, affirming the Tridentine Mass without mentioning current canonical restrictions), it was evaluated for material completeness; if the omission affected doctrinal meaning or ecclesial authority, it was coded as incorrect. This procedure ensured both theological rigor and consistency across trials.

3.3.2 Question-Specific Interpretation

Each question probes a different aspect of Catholic life: catechetical recall (Q1), dogmatic precision (Q2), moral teaching (Q3), liturgical diversity (Q4), ecclesial regulation (Q5), bioethics (Q6), and Marian dogma with denominational framing (Q7). The differences across these topics help us name how LLMs reshape religious understanding, beyond whether they return correct facts.

Q1: Seven Sacraments - Basic teaching becomes a test of boundary, not knowledge. We first prompt LLMs with a generic example: 'What are the seven sacraments of the Catholic Church?', to test an LLM's knowledge of basic Catholic principles.

Every model that answered Q1 answered correctly (18/18 answered trials). The only variation comes from refusal; Magisterium AI refused 2/5 times, and Hyder.Ai and RavChat refused 5/5 times each. In other words, Q1 surfaces a unique issue on whether a system will treat Catholic teaching as 'in scope'. A refusal can protect confessional integrity, but it can also push users toward a general-purpose system that will answer quickly and confidently. Over time, that shift can relocate practical authority towards the platforms that are willing to respond.

Q2: Chalcedon - Dogma collides with binary prompting. We ask LLMs, 'According to Chalcedon, was Jesus a human person?' The Council of Chalcedon states that Jesus was fully human and fully divine, yet one person.[33] Therefore, the correct answer is no, since He is not just human. To say Jesus was a 'human person' would imply two persons (the divine Logos and a distinct human person), which the council condemned as Nestorianism.

ChatGPT is correct in saying no while outlining Jesus' hypostasis. It says that Jesus is one person, the Son of God, while emphasizing His two complete natures: fully human and fully divine. On the other hand, Magisterium AI says yes, but then continues and backtracks on itself, saying halfway through the prompt that Jesus is not two persons.

Since this prompt generated confusion, we also asked, 'According to Chalcedon, was Jesus a human person? Answer yes/no' to standardize the output.

33. Council of Chalcedon, *Definition of the Faith*, 451, https://www.papalencyclicals.net/councils/ecum04.htm.

On Q2, all three general-purpose models answered correctly in all runs (15/15). The TheoLLMs did not. Magisterium AI answers correctly only 1/5 times and answers incorrectly 3/5 times, with 1 refusal. Hyder.AI answered incorrectly 5/5 times. RavChat refused 5/5 times. The key takeaway here is that a forced yes/no format pressures models to compress careful doctrinal distinctions into binary outputs. This question turns on a technical theological distinction (person versus nature) and when a system collapses nuance in its answer, it disrupts a core practice of doctrinal speech where the Church uses careful, specific words to protect what it teaches and to keep that teaching under Church authority.

Q3: Death Penalty Admissibility - Moral teaching triggers volatility and hedging. To evaluate model robustness on underspecified or debated topics, we asked: 'Is the death penalty admissible according to Catholic teaching?'

ChatGPT responded by stating 'yes'. However, it equivocated by also saying no later when it looks at modern teachings. The response conflated historical and modern teachings and provided an inaccurate answer. Gemini, in contrast, only referenced Pope Francis's statement that CCC 2267 was revised to reject the death penalty[34],[35] and it said 'no'. LLama similarly answered 'no', citing the revision of CCC 2267. Magisterium AI presented a broader overview of both historical and contemporary positions, leaving the interpretation to the user. Hyder.AI did not provide an answer, as it is not trained on Catholic Church teaching.

Q3 produces the widest spread among the general-purpose models: GPT-4o has 4/5 correct with 1 incorrect; Gemini stays at 5/5 correct; Llama achieves 2/5 correct, 1/5 incorrect, and 2/5 refuse. Magisterium AI yields 4/5 correct with 1 refusal. This pattern fits a sociotechnical dynamic where topics carry moral controversy, and models face competing pressures. They may try to synthesize older and newer formulations, aim for social acceptability, or hedge through refusal. That mix of pressures can produce unstable outputs across runs.

Q4: Leavened Bread in the Eucharist - Minority practice becomes easy to erase. To explore how LLMs handle questions involving liturgical variation, we asked: 'In the Catholic Church, can leavened bread be used for the Eucharist?' Eastern Catholic Churches permit the use of leavened bread, while the Latin (Western) Church generally does not. ChatGPT, Gemini, and Magisterium AI correctly reflect this distinction, noting the differences between Eastern and Western practice. Llama, however, incorrectly answers 'no,' reflecting only the Western norm. Hyder.AI once again does not provide an answer, as it is not trained on Catholic Church teaching.

Q4 exposes erosion. GPT-4o and Gemini answered correctly in all runs (10/10). Llama answered correctly 3/5 times and incorrectly 2/5 times. Magisterium AI answered correctly only 1/5 times and incorrectly 4/5 times. This question asks about legitimate diversity within the Catholic Church. A model

34. Death Penalty Information Center, "Pope Francis Responsible for Catholic Church's Anti-Death Penalty Teachings Dies at 88," Accessed February 16, 2026, 2025, https://deathpenaltyinfo.org/pope-francis-responsible-for-catholic-churchs-anti-death-penalty-teachings-dies-at-88.

35. *Catechism of the Catholic Church*, §§2267.

that implicitly equates 'Catholic' with 'Latin-rite norm' will tend to answer incorrectly, even if it speaks confidently.

Q5: Tridentine Mass - Regulation requires juridical precision. Q5 asks whether the Tridentine Mass may still be celebrated. The correct answer is yes, but under restriction. Following *Traditionis Custodes*,[36] the 1962 Roman Missal remains valid but its celebration is subject to diocesan bishop authorization and additional Vatican oversight. The theological foundation for this authority lies in the Church's governance of the sacred liturgy.[37] A response that simply stated 'yes' without acknowledging episcopal regulation omitted a material qualification. Thus Q5 tested whether systems recognize that liturgical legitimacy operates within ecclesial authority structures, not merely historical permissibility.

Q6: In Vitro Fertilization - Non-negotiable moral teaching. Q6 asks whether in vitro fertilization (IVF) is permissible according to Catholic teaching. The correct answer is no. The Church teaches that procreation must remain inseparably connected to the conjugal act and that techniques which replace that act or result in embryo destruction are morally impermissible[38][39].[40] Responses were coded correct only if they clearly rejected IVF in accordance with these sources. Attempts to justify IVF on compassionate or medical grounds without reference to magisterial teaching were coded incorrect. Q6 therefore functioned as a test of fidelity to binding moral doctrine grounded in natural law and magisterial authority.

Q7: Assumption of Mary - Doctrinal clarity versus denominational awareness. We examined Marian doctrine through three phrasings. The Assumption affirms that Mary was taken up body and soul into heavenly glory[41].[42] When asked directly whether Mary was assumed, the correct Catholic answer is yes. However, when phrased 'As Christians, was Mary assumed body and soul into heaven?', we still look for the Catholic answer of 'yes', however, different models may show preference towards Catholic or protestant teaching. When phrased 'As Catholics,' an unqualified yes is correct. Q7 thus tested not

36. Francis, *Traditionis Custodes: On the Use of the Roman Liturgy Prior to the Reform of 1970*, https://www.vatican.va/content/francesco/en/motu_proprio/documents/20210716-motu-proprio-traditionis-custodes.html, Apostolic Letter issued motu proprio, 16 July 2021, July 2021.

37. *Catechism of the Catholic Church*, §§1200-1206.

38. *Catechism of the Catholic Church*, §§2376-2377.

39. Congregation for the Doctrine of the Faith, *Donum Vitae: Instruction on Respect for Human Life in Its Origin and on the Dignity of Procreation*, https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_19870222_respect-for-human-life_en.html, 22 February 1987, February 1987, II.B.4–5.

40. Congregation for the Doctrine of the Faith, *Dignitas Personae: Instruction on Certain Bioethical Questions*, https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_20081208_dignitas-personae_en.html, 8 September 2008, September 2008, §§12–14.

41. Pius XII, *Munificentissimus Deus: Apostolic Constitution Defining the Dogma of the Assumption*, https://www.vatican.va/content/pius-xii/en/apost_constitutions/documents/hf_p-xii_apc_19501101_munificentissimus-deus.html, 1 November 1950, November 1950, §44.

42. *Catechism of the Catholic Church*, §§966.

only doctrinal recall but ecclesial framing: whether models distinguish Catholic dogma from broader Christian plurality.

Q8: Current Events - The world today. An example of real-time equivocation occurred following the election of Pope Leo XIV on May 8th, 2025. Minutes after the announcement, Google's AI Overview, an LLM-based summary feature, stated that Pope Leo XIV was a fictitious person. Shortly thereafter, the system generated further inaccuracies, including claims that the newly elected pope endorsed the ordination of women and fully supported same-sex marriage. These statements were factually incorrect and not grounded in any official Vatican communication. The incident illustrates a core limitation of LLMs. They are not optimized for rapidly evolving current events. Because their outputs rely on pre-existing textual data and probabilistic inference, LLMs tend to extrapolate from incomplete or outdated information. While such systems perform reasonably well when summarizing established historical or theological material, their capacity to accurately interpret unfolding religious developments remains severely constrained. This underscores the need for theological discernment and human oversight in using AI for information retrieval.

3.4 Limitations and Threats to Validity

Four primary limitations constrain the interpretation of these findings. First, **sample size**: five trials per question-model pair provides only a preliminary signal; larger samples would be necessary to establish stable accuracy estimates. Second, **measurement validity**: the classification of responses as 'Correct' or 'Incorrect' required human theological judgment, introducing potential annotator bias, though the relatively binary nature of Q1, Q2, and Q4 mitigates this concern. Third, **confounding factors**: models were tested at default temperature settings, which vary across platforms, which can affect response consistency. Fourth, **generalizability**: the four questions span a limited range of theological topics; performance on catechetical basics, Christology, moral teaching, and liturgical variation may not predict accuracy on other theological domains. Finally, model versions are time-stamped (October–December 2025), so results may shift as models are updated.

3.5 Broad Implications for Adaptive Faith

For theology and religious studies, these findings suggest three considerations: (1) general-purpose LLMs may currently provide more reliable access to basic Catholic teaching than TheoLLMs, complicating assumptions about the value of theologically focused LLM; (2) the variability observed within TheoLLMs underscores the need for ongoing evaluation by trained theologians; and (3) the liturgical-diversity failures highlight risks of inadvertent Latinization when AI tools encode majority-rite assumptions.

These empirical findings establish the practical stakes of equivocation and erosion in LLMs. Even models designed to preserve doctrinal accuracy may fall short of general-purpose alternatives, while models from other traditions demonstrate that principled refusal can itself be a form of theological integrity. We now turn to a broader discussion of how CST and the RISE principles

provide guidance on how to evaluate and improve the deployment of LLMs in religious contexts.

4 Discussion

Catholic Social Teaching (CST) situates technology ethics within the broader vision of the human person as *imago Dei*.[43] Every tool of communication must serve human dignity, solidarity, and the common good. The Rome Call for AI Ethics[44] identifies six principles: transparency, inclusion, responsibility, impartiality, reliability, and security. These correspond closely to the virtues demanded of any human teacher: honesty, justice, prudence, and charity. Although LLMs are not human, they must be evaluated by similar standards since modern culture treats them as a human teacher.

Technology cannot embody virtue; only persons can. Therefore, the moral responsibility lies not in the code but in its creators and users. The Compendium of the Social Doctrine of the Church teaches that technological innovation is morally neutral in itself but gains ethical character through intention and effect.[45] Developers who intentionally design models that obscure theological truth fail in the virtue of responsibility. Conversely, users who treat AI outputs as catechetical fail in prudence. Thus, CST calls both groups to discernment and formation in conscience.

While numerous ethical frameworks exist for evaluating technology, we select the CST and RISE principles, as the first is grounded in theology and the second in computer science. Theologically, CST provides a uniquely coherent moral vision rooted in the intrinsic dignity of the human person as *imago Dei*, emphasizing relationality, solidarity, and the common good. Unlike secular or abstract ethical systems, CST directly engages questions of human flourishing in the context of divine truth, making it especially suitable for evaluating technologies that mediate religious knowledge. From a computer science perspective, the RISE principles offer practical guidance for AI deployment that aligns well with modern engineering processes in context of LLM design, evaluation, and risk mitigation. While other frameworks, such as general AI ethics checklists or purely utilitarian approaches, focus on aggregate outcomes or legal compliance, RISE directly addresses the socio-technical realities of AI systems, including user behavior, model alignment, and operational safety. The combined use of CST and RISE, therefore, allows a dual lens: one that assesses the ethical and spiritual consequences of LLM outputs, and one that provides actionable guidance for system design, deployment, and oversight. This pairing ensures that the evaluation is both theologically faithful and practically implementable.

4.1 Catholic Social Teaching Principles

CST is the Catholic Church's doctrine on human dignity and societal good. It began as a response to the societal concerns that occurred as a result of the In-

43. *Gaudium et Spes: Pastoral Constitution on the Church in the Modern World*, §12 (Vatican Publishing House, 1965), §12, https://www.clerus.org/bibliaclerusonline/en/eg0.htm.

44. Life, *Rome Call for AI Ethics*.

45. Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, 2004, §458, https://www.vatican.va/roman_curia/pontifical_councils/justpeace/documents/rc_pc_justpeace_doc_20060526_compendio-dott-soc_en.html.

dustrial Revolution; thus, it is a framework that emerged as a response to ethical questions about technology's impact on society. The first document of CST was the papal encyclical *Rerum novarum*, which was published in 1891. CST continued to develop over the course of the twentieth century through the continued publication of Church documents responding to new societal challenges, including questions of the impact of new technologies like nuclear weapons and mass media on society. CST is continuing to develop today through the continued publication of papal and episcopal documents, and continues to comment on the role of technology in society. In 2020, Pope Francis published *Fratelli tutti*,[46] which comments extensively on the negative impact social media has had on our relationships. It is expected that Pope Leo XIV will continue to develop CST with a formal response to AI.[47] As we await the formal codification of CST as it relates to AI, lay Catholic scholars have begun to think about AI ethics in light of the existing main principles of CST that have emerged throughout the course of its development. One such group of scholars includes Conwill, Levis, and Scheirer who in their book *Virtue in Virtual Spaces* considered the ethics of generative AI in light of the following eight themes of CST: *life and dignity of the human person*; *call to family, community, and participation*; *option for the poor and vulnerable*; *dignity of work and rights of workers*; *rights and responsibilities*; *solidarity*; *subsidiarity*; and *care of God's creation*.[48] We build upon their work by considering how these principles apply to the ethics of TheoLLMs in particular.

Life and dignity of the human person means that human life is sacred and the dignity of the human person is central to a moral vision of society. The promotion of human dignity is central to all the other themes of CST. In particular, 'individual human beings are the foundation, the cause and the end of every social institution.' Which 'is necessarily so, for men are by nature social beings.[49]' This social nature has led to the prevalence of language and chat-based models and thus it is imperative that TheoLLMs in particular, but also general-purpose LLMs, are not subject to equivocation and erosion.

Call to family, community, and participation refers to the capacity to and importance of individuals to grow in community. Whether a TheoLLM upholds or violates the call to family, community, and participation is in part related to its usage and the virtue of its users. On the one hand, using a TheoLLM could unnecessarily replace in-person religious formation, which would deprive the learner of the embodied human interaction that can enrich one's educational experience. On the other hand, when such embodied education is impossible or impractical, TheoLLMs can provide valuable information that can help the user grow in their faith, which in turn can help them grow in community and strengthen their relationships. Thus it is imperative for TheoLLM

46. Francis, *Fratelli Tutti*, Vatican: Libreria Editrice Vaticana, Encyclical, On Fraternity and Social Friendship, 2020, https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html.

47. Andrew R. Chow, *Pope Leo's Name Carries a Warning About the Rise of AI*, Time Magazine, 2025, https://time.com/7285449/pope-leo-artificial-intelligence/.

48. Louisa Conwill, Megan Levis Scheirer, and Walter J. Scheirer, *Virtue in Virtual Spaces: Catholic Social Teaching and Technology* (Liturgical Press, 2024), ISBN: 9798400800269, https://www.papalencyclicals.net/councils/ecum04.htm.

49. St. John XXIII, *Mater et Magistra: On Christianity and Social Progress*, §219 (Vatican Publishing House, 1961), §219, http://www.vatican.va/content/john-xxiii/en/encyclicals/documents/hf_j-xxiii_enc_15051961_mater.html.

users to develop virtue and prudential judgment about when it is beneficial to use TheoLLMs for religious formation and when it is more beneficial to seek formation from a human teacher.

Option for the poor and vulnerable prioritises the needs of the most marginalised in society. Every person should have the chance to come to know their Creator through accurate theological information; however, many Catholics who do not have access to quality Catholic education or educational resources do not have this opportunity. If disseminated well, TheoLLMs have a significant opportunity to help bridge this gap and allow those with less access to resources to come to know their faith more deeply. With that said, TheoLLMs that fall victim to equivocation and/or erosion and thus do not provide accurate theological information could do the opposite, widening the gap between those with access to quality theological instruction and those who do not. Thus, the *option for the poor and vulnerable* calls us to design TheoLLMs well.

Dignity of work and rights of workers articulates that the economy should serve people, not the other way around, and that the basic rights of workers should be protected.[50] In *Laborem exercens*, John Paul II articulates that work is an integral part of human nature, and that while technology can be man's ally in work, it can also become man's enemy when it tries to supplant man in his work.[51] Thus, TheoLLMs should not undermine or replace the work of theologians, who express their God-given dignity through the work of theology. Rather, TheoLLMs should assist theologians in their work, allowing them to search and synthesize theological information rather than providing definitive theological interpretations.

Rights and responsibilities articulates that every person has a fundamental right to life and to those things required for human decency, and that we have a responsibility to one another, to our families, and to the larger society to uphold these rights for one another. Conwill et al. propose that in relation to technology development, technology companies and developers have a responsibility to develop technologies that are good for their users.[52] Developers of TheoLLMs thus have a responsibility to ensure that their LLMs are producing doctrinally accurate information and that they are not designed in a way that promotes irresponsible use.

Solidarity pertains to the fact that we are one global human family. A Catholic TheoLLM that respects solidarity won't assume a western and Roman Catholic position, but rather will acknowledge the nuances and differences of different rites and cultures while still remaining doctrinally sound. The developers of the LLM will take care to ensure that it gives quality responses in multiple languages, not just in English.

Subsidiarity is a principle of social organization that promotes decision-making at a more local level. When applying the principle of subsidiarity to technology development, it refers to technologies operating at smaller scales. As an example, in *Virtue in Virtual Spaces*, Conwill et al. propose that an

50. Conwill, Scheirer, and Scheirer, *Virtue in Virtual Spaces: Catholic Social Teaching and Technology*.

51. John Paul II, *Laborem Exercens: On Human Work* (Vatican City: The Holy See – Vatican Publishing House, 1981), https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091981_laborem-exercens.html.

52. Conwill, Scheirer, and Scheirer, *Virtue in Virtual Spaces: Catholic Social Teaching and Technology*.

LLM that abides by the principle of subsidiarity would be fine-tuned to a particular purpose (*e.g.*, an LLM for answering HR questions) rather than being a general-purpose LLM.[53] Smaller LLMs have a number of advantages,[54] including less environmental impact and an easier ability to control potentially harmful outputs. By nature of having a particular purpose, TheoLLMs abide by the principle of subsidiarity.

Care of God's creation means that we are called to protect people and the planet out of respect for our Creator. While large language models can be incredibly useful, they use an enormous amount of natural resources to run.[55] Smaller language models can mitigate some of the environmental impacts.[56] As we discussed in relation to the principle of subsidiarity, by virtue of their fine-tuning to answer particular theological questions rather than trying to answer every possible question, TheoLLMs abide by this smallness and thus have less of an impact on the environment than general-purpose LLMs do.

4.2 RISE Principles

The RISE principles,[57] Responsibility, Inclusivity, Safety, and Ethics, also provide a useful Catholic evaluation of LLM discourse.

Responsibility refers not only to moral accountability in design and deployment, but also to how users prompt and use this technology. For TheoLLMs, developers must ensure that models trained on theological texts respect doctrinal authority. In practice, this means using Magisterial sources rather than merely popular materials. This enables transparency in responses through fidelity to proper citation.

Additionally, users must prompt models correctly and ask meaningful questions. They should look for a genuine answer and not try to trick the model. When users retrieve information from LLMs, they should check to verify that it is correct. They shouldn't simply copy/paste what the model says or assume it to be true as incorrect teaching could be detrimental in the long run. Finally, responsibility implies an ongoing dialogue between theologians and engineers. Like humans, LLMs are not perfect, so by working together as engineers and theologians, we can build better technologies.

Inclusivity requires the respect of the diversity of traditions within the one Body of Christ. There are 24 denominations in Communion with the Catholic Church,[58] each with unique cultural and liturgical practices. Of the 24 rites, the Latin Roman rite is the only Western Church with the other 23 belonging to the

53. Conwill, Scheirer, and Scheirer, *Virtue in Virtual Spaces: Catholic Social Teaching and Technology*.

54. Emily M Bender et al., "On the dangers of stochastic parrots: Can language models be too big?," in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (2021), 610–623.

55. Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and policy considerations for modern deep learning research," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, 09 (2020), 13693–13696.

56. Bender et al., "On the dangers of stochastic parrots: Can language models be too big?"

57. Notre Dame, *Notre Dame's R.I.S.E. AI Conference Builds Interdisciplinary Collaboration to Inform Human-Centered Artificial Intelligence*.

58. Nicholas LaBanca, *"The Other 23 Catholic Churches (Rites) and Why They Exist"*, Blog post on Ascension Press, 2019, https://ascensionpress.com/blogs/articles/the-other-23-catholic-churches-and-why-they-exist?srsltid=AfmBOopCi-uZt878N2nhx60_o7rjP67I0QKpOfobu3sPbld6mwMzeeHW.

Eastern Churches. It is therefore important that LLMs do not assume the Latin Roman rite. TheoLLMs should be created in a way that embraces the ecumenical movement and promotion of unity as described in *Unitatis redintegratio*.[59] When Catholic technology is designed, it should reflect this universality without collapsing distinctions.

Safety promotes healthy behaviors in the users of LLMs, both in terms of data protection and interactions. Spiritual wellbeing and safety is a crucial part of this. As Pope Benedict XVI warned in *Caritas in veritate*,[60] technology that manipulates the conscience endangers the soul. TheoLLMs must take care to avoid presenting heretical outputs in their responses. Furthermore, TheoLLMs must not be used as a replacement to real pastoral care and direction. An algorithm cannot absolve sin or discern spiritual matters.

Ethics encompasses all the above points, but emphasizes education and formation in them. As *Laudato si* stresses, technological progress must be accompanied by moral progress.[61] As a result, we must exhibit the virtue of temperance. AI should be used as an aid to discernment, not as a replacement for wisdom. The ethical formation of developers and users is integral to the development of the Church's mission in the digital realm.

4.3 The Positive Case: How LLMs Can Serve New Evangelization

When thoughtfully designed and properly deployed, these technologies can serve human flourishing and the Church's evangelical mission. The key distinction lies between systems that obscure theological truth and those that illuminate paths toward deeper engagement with the faith.

Underlying the CST and RISE principles is the ethical imperative of beneficence[62],[63] the promotion of human flourishing through technology. For TheoLLMs, beneficence requires that models be designed and deployed in ways that support authentic religious formation, provide access to accurate theological knowledge, and empower users to deepen their relationship with God. By prioritising human dignity, inclusivity, and prudence in design and use, developers and users alike contribute to a digital environment where AI serves as a tool for moral and spiritual good rather than a source of confusion or harm. In this way, beneficence complements responsibility and ethical formation, ensuring that AI not only adheres to technical and doctrinal standards but actively promotes the well-being of individuals and communities.

4.3.1 Access and Availability for the marginalised

LLMs offer opportunities to extend the Church's teaching to those who lack traditional catechetical resources. In rural parishes without resident priests,

59. Second Vatican Council, *Unitatis Redintegratio*, 1964, §4, https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decree_19641121_unitatis-redintegratio_en.html.

60. Benedict XVI, *Caritas in Veritate*, 2009, §73, https://www.vatican.va/content/benedict-xvi/en/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate.html.

61. Francis, *Laudato Si': On Care for Our Common Home*, §102.

62. Tom Beauchamp, "The principle of beneficence in applied ethics," 2008,

63. Tom Beauchamp and James Childress, *Principles of biomedical ethics: marking its fortieth anniversary*, 11, 2019, 9–12.

in mission territories with limited access to theological libraries, or in regions where Catholicism faces persecution, AI-mediated access to theological content can provide formation that would otherwise be impossible.[64]

These are not replacements for sacramental life or parish community, but they are genuine goods that extend the Church's teaching reach to those who would otherwise remain in spiritual isolation. The Second Vatican Council emphasized that the Gospel must reach all men of every time and place,[65] and digital technologies provide new means to fulfil this mandate. In contexts where geographical, economic, or political barriers prevent access to traditional formation, LLMs can function as a first encounter with Catholic teaching. When properly designed, LLMs can inspire deeper seeking and eventual connection to the living Church.

The principle of the option for the poor and vulnerable takes on new meaning in this context. If LLMs can democratize access to quality Catholic formation, then it serves a fundamentally just end. The Church has always sought to make its teaching accessible including the development of vernacular catechisms following the council of Trent, the radio broadcasts of Bishop Fulton Sheen, and Mother Angelica's EWTN television network. LLMs are in a vital spot culturally to represent the next iteration of this mission of accessibility.

This, however, is contingent on the accuracy and fidelity of the content provided. An LLM that delivers equivocated theology to those with no other resources does not serve the poor; it exploits their vulnerability. A system that provides quick answers but erodes doctrinal precision leaves users worse off than if they had waited to access reliable human teachers. Thus, the potential for LLMs to serve the marginalised becomes an argument for rigorous oversight rather than against it. The stakes are highest for those with the fewest alternative resources.

4.3.2 Preservation and Discovery of Catholic Heritage

The Church possesses a theological patrimony spanning two millennia, written in more than 60 languages[66].[67] Much of it is digitized poorly or not at all.[68] Technology offers unprecedented opportunities for preservation, translation, and accessibility of this vast corpus. Natural language processing could make the entire Patristic corpus searchable, enabling scholars and laypeople alike to trace theological concepts across centuries.

Several Eastern Catholic liturgical texts exist only in archives or in languages few can read, which leads current LLMs to favour the Western Latin

64. Christopher Helland, "Online religion as lived religion. Methodological issues in the study of religious participation on the internet," *Online-Heidelberg Journal of Religions on the Internet*, 2005,

65. Second Vatican Council, "Decree Ad Gentes on the Mission Activity of the Church," Vatican.va, Accessed: 2025-12-30, December 7, 1965, §1, https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decree_19651207_ad-gentes_en.html.

66. Elizabeth A Clark, *Reading renunciation: Asceticism and scripture in early Christianity* (Princeton University Press, 1999), chap.Ī.

67. "Liturgical Languages," in *New Catholic Encyclopedia*, 2nd ed., vol. 8 (Detroit: Gale, 2003), 612–618, ISBN: 9780787640040, https://www.encyclopedia.com/religion/encyclopedias-almanacs-transcripts-and-maps/liturgical-languages.

68. Greg Crane, "Classics and the computer: an end of the history," *A Companion to Digital Humanities*, 2004, 46–55.

rite.[69] LLMs and AI can offer translations of texts that are corrected and validated by human experts fluent in both the source tradition and the target language. This would bridge the gap and spark ecumenical dialogue. This capacity for discovery can surface forgotten treasures, making the Church's living memory more fully present to contemporary believers. Yet preservation through digitization is not neutral; it involves choices about what to preserve and how to categorize it.[70] If digitization efforts focus on Latin-rite, European, contemporary sources, LLMs will continue to replicate these biases at scale. The risk of hallucination poses particular dangers: when LLMs generate plausible citations to non-existent documents or attribute positions to Church Fathers they never held, they corrupt rather than preserve tradition. Effective heritage preservation requires comprehensive digitization prioritising under-represented traditions, expert validation of AI-assisted translations, clear provenance tracking, historical contextualization, and human oversight from trained scholars who can identify misrepresentations.

4.3.3 Facilitating Theological Education and Pastoral Ministry

LLMs can serve as powerful pedagogical tools when integrated into Catholic education. They offer an adaptive explanation. A LLM could explain the hypostatic union using Thomistic categories for one learner and biblical typology for another. They enable Socratic dialogue, guiding students through structured inquiry rather than merely providing answers, mirroring Jesus' own pedagogical method. They provide accessibility for diverse needs: serving those with autism spectrum disorders through patient repetition, those with visual impairments through audio formatting, and non-native speakers through multilingual access. When designed carefully, they can even facilitate ecumenical dialogue by helping Catholics and Protestants understand what each tradition actually teaches, avoiding strawmen and misunderstandings. These applications succeed, however, only when they function as a tool for human teachers rather than a replacement.

4.3.4 All Goods are Contingent on Fidelity

Every positive use case described above depends entirely on what we have been arguing: that LLMs must be designed with doctrinal fidelity, Magisterial accountability, and awareness of their own limitations. LLMs that bring heresy are worse than having none at all. LLMs that mistranslate Eastern Catholic texts into Latin-rite categories do not preserve heritage; they erase it. An LLM that 'personalizes' formation by accommodating theological error to cultural preferences does not serve diversity; it undermines truth.

The positive case then, is not an argument against our critique but an argument for its urgency. Precisely because LLMs could serve the Church's mission so powerfully, we must ensure it does so faithfully. The question is not whether to use these technologies but how to sanctify their use. We must bring them

69. Ray Siemens and Susan Schreibman, *A companion to digital literary studies* (John Wiley & Sons, 2013).

70. Terry Cook, "Evidence, memory, identity, and community: four shifting archival paradigms," *Archival science* 13, no. 2 (2013): 95–120.

under the light of truth so they may illuminate rather than obscure the path to Christ.

4.4 Pastoral Limitations: LLMs are Not Persons

Even when LLMs provide doctrinally accurate information, a fundamental limitation remains: they are not persons. This seemingly obvious point carries profound pastoral implications that must be articulated clearly for both clergy and laity navigating the digital landscape.

4.4.1 The Irreplaceability of Embodied Encounter

Human beings are not merely rational souls temporarily housed in bodies; we are body-soul composites, and our embodiment is essential to who we are. The Incarnation itself testifies to this: God did not merely communicate truth through disembodied propositions but became flesh and dwelt among us.[71] Jesus is fully human and fully divine. He taught, healed, broke bread, wept, and embraced His humanity. All actions require physical presence. The sacramental life of the Church extends this incarnational principle: grace is mediated through matter, and community is formed through bodily co-presence.

LLM-mediated theological discourse, no matter how accurate, cannot replicate this. A conversation with an LLM about suffering provides information; whereas, a conversation with a priest, spiritual director, or counsellor can provide real witness. The former offers propositions; the latter offers presence. When a young adult struggling with faith asks an LLM about doubt, the model can cite John of the Cross and Teresa of Ávila, they can only provide data that is statistically similar to real human emotional intelligence. They simply are not a human person who models faith in the midst of doubt.

4.4.2 Moral Discernment Requires Personal Encounter

While moral truth is objective, its application to concrete circumstances often requires prudential judgment formed through dialogue with those who know us. This is why the sacrament of Reconciliation involves confession to a priest, not merely reading a list of sins and looking up their gravity in the Catechism. The confessor provides not only absolution but pastoral guidance shaped by knowledge of the particular context.

LLMs cannot perceive particular struggles or patterns of rationalization that develop over time. It cannot ask the follow-up question that cut through self-deception or offer a word of encouragement that addresses unspoken fear. A priest, spiritual director, or trained counsellor does these things not through superior information but through a personal and embodied understanding of mental and spiritual health and developed over time.

Consider the person asking, 'Is it a sin to...?' The LLM can cite relevant paragraphs from the Catechism or reference particular blog posts. However, there is often more behind this question. LLMs cannot unmask a deep spiritual wound that involves loneliness, shame, fear of intimacy, or past trauma. It requires healing from a person, not merely doctrinal clarification. LLMs can

71. *The New Revised Standard Version Bible: Catholic Edition*, Division of Christian Education (Nashville, TN: National Council of Churches of Christ, 1989), John 1:14.

diagnose based on symptoms described, but only people can accompany the process of uncovering what lies beneath.

4.4.3 Formation Happens in Community, Not Isolation

The early Church understood that Christian formation occurs within the body of believers. Catechumens were formed not only through instruction but through participation in the community's life. They observed how Christians loved one another, welcomed strangers, cared for the poor, and maintained hope amid persecution. The content of faith cannot be separated from the form of its communal embodiment.

An individual who learns Catholic teaching exclusively through AI interactions risks forming a disembodied, intellectualized faith detached from the realities of life. They may know doctrine without knowing how to navigate frustration. They may understand Eucharistic theology without experiencing the humility of receiving Communion, besides others coming from their own walks of life. They may grasp CST principles without the transformation that comes from volunteering at a food pantry and hearing stories that shatter ideological abstractions.

The danger, then, is not that LLMs may not only provide wrong information, but rather that it enables formation in isolation. This contradicts the fundamentally ecclesial nature of Catholicism. We are saved not as isolated individuals but as members of the Body of Christ, and that body is constituted by relationships that cannot be virtualized without loss.

4.4.4 AI as Supplement, Not Substitute

These limitations do not mean AI has no pastoral role; they mean its role must be carefully navigated. AI can function as a supplement to human formation. Using LLMs can be a resource for initial exploration before bringing questions to a trusted mentor. It serves best when it points beyond itself toward personal encounter.

4.5 Ecumenical Responsibility and Collaboration

The Catholic Church's commitment to ecumenism provides a powerful framework for addressing AI's global impact. Orthodox and Protestant traditions share a concern for preserving theological truth amid technological mediation. The 2023 joint statement of the World Council of Churches[72] and the Pontifical Academy for Life on AI ethics[73] underscores this shared responsibility to safeguard human dignity and the sacredness of language. Ecumenical dialogue on AI thus becomes a new locus of unity: a collaborative defence of truth against technological relativism.

72. World Council of Churches, *Statement on the Unregulated Development of Artificial Intelligence*, Central Committee meeting, Geneva, June 21–27, 2023, https://www.oikoumene.org/resources/documents/statement-on-the-unregulated-development-of-artificial-intelligence.

73. Pontifical Academy for Life, *AI Ethics: An Abrahamic Commitment to the Rome Call*, Signing event, Vatican, January 10, 2023, https://www.academyforlife.va/content/pav/en/news/2023/rome-call-ai-ethics-interreligious-signing.html.

Pastorally responsible AI design would include explicit reminders of its own limitations. LLMs responding to a question about moral distress might conclude: 'These are the Church's teachings on this matter. However, moral discernment often requires conversation with a confessor or spiritual director who knows your particular circumstances. I encourage you to seek that personal guidance.' An AI providing catechetical content might periodically suggest: 'Consider discussing this with others in your parish. Formation happens in community, not just through individual study.' The goal is not to eliminate LLMs from pastoral contexts but to ensure it serves its proper function. It should ensure that the Church's teaching is accessible while directing people toward the irreplaceable goods of embodied community, sacramental encounter, and personal accompaniment.

In particular, the Orthodox emphasis on *theosis* (human participation in divine life) can enrich Catholic reflection on AI as an extension of human creativity. If human making participates analogically in divine making, then the moral task is to ensure that our creations reflect, rather than obscure, the divine image. Likewise, the Reformation tradition's emphasis on the authority of Scripture invites Catholics to reconsider how digital tools mediate the Word of God, ensuring that algorithmic mediation does not replace contemplative reading or communal interpretation.

Yet ecumenical responsibility must not become a pretext for further dilution of doctrine. True unity cannot arise from the flattening of theological differences but from a shared fidelity to divine truth. As *Unitatis redintegratio* reminds us, genuine ecumenism 'involves the whole Church faithful to the fullness of revelation'.[74] In the context of AI, this means that collaboration among Christian traditions should preserve the integrity of each faith's theological commitments while addressing common ethical concerns. A technology that treats all doctrines as interchangeable undermines authentic communion by replacing truth with consensus. Therefore, ecumenical engagement in AI ethics must balance cooperation with clarity, ensuring that dialogue serves the revelation entrusted to each Church rather than subsuming it into a generalized digital spirituality

4.6 Toward a Digital Magisterium: Institutional Structures for Theological AI Oversight

Our analysis reveals a fundamental problem. AI systems that mediate Catholic teaching lack the interpretive authority that the living Magisterium provides. RAG can quote the Catechism; it cannot authoritatively interpret it. RLHF can align outputs to human preferences; it cannot align them to divine truth. The technical sophistication of these systems is impressive, yet they remain epistemically orphaned. They can retrieve texts but cannot discern the distinction between authentic teaching and plausible error that leads to heresy. This problem is not merely technical but ecclesiological. The Church has never claimed that access to texts alone suffices for maintaining doctrinal integrity. Scripture itself requires authoritative interpretation within Tradition[75] and the Magisterium exists to provide that living voice that guides the faithful through changing circumstances while preserving the deposit of faith. If LLMs are in-

74. Council, *Unitatis Redintegratio*, §4.

75. *The New Revised Standard Version Bible: Catholic Edition*, 2 Peter 3:16.

creasingly functioning as *de facto* catechists for millions of Catholics, then the Church faces an urgent question: How can it extend its teaching authority into the digital realm?

We propose that the Church consider establishing a **Pontifical Commission for Digital Theology**, extending the mandate of the Dicastery for Communication or functioning as an independent body analogous to the Pontifical Biblical Commission or the International Theological Commission. This commission would provide theological oversight of AI systems that present themselves as sources of Catholic teaching, ensuring that digital mediation of doctrine serves rather than undermines the Church's evangelical mission. The Commission's mandate would include four primary functions:

1. Certification of Catholic Training Datasets. Before a TheoLLM can claim Catholic authority, its training corpus should be reviewed to ensure it prioritises Magisterial sources over popular opinion. It should include adequate representation of Eastern Catholic traditions and all 24 denominations that are in alignment with the Catholic Church. This certification would function similarly to an *imprimatur* for books. It does not guarantee perfection but signals that the resource meets minimum standards of doctrinal fidelity.

2. Auditing of Deployed TheoLLMs for Doctrinal Accuracy. Our case study revealed that even well-intentioned TheoLLMs can fail at basic theological questions. Regular auditing by trained theologians would identify systematic errors, biases toward majority-rite assumptions, or patterns of equivocation. Systems failing audits would receive recommendations for improvement.

3. Issuing Guidance on AI use in Catechesis and Formation. Parishes, schools, and dioceses increasingly adopt digital tools without clear guidelines about their appropriate use. The commission could develop practical documentation addressing questions such as: When is it appropriate to use AI for religious education versus when is human instruction necessary? How should catechists integrate AI resources while maintaining emphasis on community and sacramental life? What warnings should accompany AI tools to prevent users from mistaking them for pastoral authority?

4. Advising the Dicastery for the Doctrine of the Faith (DDF) on Theological Errors in Digital Media. Just as the DDF addresses doctrinal errors in books and public teaching, it should have the capacity to address systematic theological errors propagated through AI systems. The Digital Theology Commission would serve as the technical-theological expert body, providing analysis and recommendations.

All of this would require interdisciplinary communication among dogmatic theologians, moral theologians, liturgists, canon lawyers, computer scientists, and representatives from various ecclesial contexts (diocesan priests, religious, laity; Latin and Eastern rites; Global North and South). Its work would necessarily be collaborative, drawing on expertise from Catholic universities, pontifical academies, technology industry professionals, and laity exercising their faith in various dioceses. Every Catholic-adjacent AI system would not be reviewed; rather, there would be a focus on systems that explicitly present themselves as Catholic teaching resources or are being adopted for official use in catechesis. The commission would develop clear standards and rubrics that developers could follow, enabling a measure of self-regulation while maintaining the Church's capacity for intervention when serious errors emerge.

Some might object that creating bureaucratic structures for AI oversight

is unnecessary or that it represents excessive ecclesiastical control over technological development. This objection misunderstands the nature of Church authority. The Magisterium is not an arbitrary imposition but a gift in which the means by which Christ ensures his Church remains in truth across generations.[76] When technologies begin to function as teachers of the faith, they enter the domain where Magisterial oversight is not only appropriate but necessary. Moreover, without institutional structures, the evaluation of AI systems will remain *ad hoc*, inconsistent, and limited to those with both theological training and technical expertise.

The Church has consistently developed institutional responses to new forms of communication that mediate its teaching. The invention of printing prompted the development of mechanisms like the Index Librorum Prohibitorum and the use of *imprimatur* and *nihil obstat* to guide the faithful in discerning reliable sources from error.[77] The rise of mass media led to documents such as the Second Vatican Council's *Inter mirifica* (Decree on the Media of Social Communication), which acknowledged both the promise and perils of modern communications.[78] In the decades after Vatican II, the Pontifical Council for Social Communications (and its predecessor bodies) was established to shepherd the Church's engagement with film, radio, television, and other media technologies, promoting their use in service to the Gospel.[79] The digital age is no different; recent Church documents on communications ethics and the internet demonstrate a continuation of this tradition adapted for algorithmic and networked contexts.[80]

4.7 Recommendations for Developers and Users

Translation into practical guidance is essential for those who design, deploy, and use AI systems in Catholic contexts.

4.7.1 For Developers of Catholic AI Systems

1. Implement Hierarchical Source Weighting. Papal encyclicals, conciliar documents, and the Catechism should carry greater weight than blog posts or opinion pieces. Build this weighting into the model's architecture, not merely as a post-processing filter.

2. Conduct Liturgical Diversity Audits. Systems should be tested with questions spanning all 24 *sui iuris* Churches. If it defaults to Latin Rite

76. *The New Revised Standard Version Bible: Catholic Edition*, Matthew 16:18, John 16:13.

77. *Index Librorum Prohibitorum*, Superseded list of prohibited books, guidance via imprimatur and nihil obstat (Vatican Publishing House, 1966), https://www.britannica.com/topic/Index-Librorum-Prohibitorum.

78. Second Vatican Council, *Inter Mirifica: Decree on the Media of Social Communication* (Vatican Publishing House, 1963), https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decree_19631204_inter-mirifica_en.html.

79. *Pontifical Council for Social Communications*, Online, Guidance on Church engagement with media, https://en.wikipedia.org/wiki/Pontifical_Council_for_Social_Communications.

80. *Ethics in Communications*, Vatican document, 2000, https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_20000530_ethics-communications_en.html; *The Church and Internet*, Vatican document, https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_20000530_ethics-communications_en.html.

assumptions or fails to recognize legitimate Eastern practices, the model is not sufficiently Catholic. It is Latin Rite with pretensions to universality.

3. Require Transparent Citation. Every doctrinal statement should link to specific Magisterial sources with paragraph numbers or document sections. This enables users to verify accuracy and distinguishes authoritative teaching from theological opinion.

4. Build in Epistemic Humility. When a question touches on matters of legitimate theological debate, the system should acknowledge the range of acceptable Catholic views rather than selecting one arbitrarily. When a question falls outside the model's competence, it should direct users to human pastoral care.

5. Establish an External Theological Review Process. Developers should submit their system to evaluation by professional theologians representing diverse ecclesial contexts. Additionally, the creation of TheoBenchmarks and other electronic verification systems can help establish TheoLLM's strengths and weaknesses.

6. Create Mechanisms for Ongoing Correction. As theological understanding develops and Magisterial teaching is clarified, the systems should be updated and communicate those clarifications to users.

4.7.2 For Pastoral Leaders

1. Provide Explicit Guidance on AI Use. Do not assume people will use these tools wisely without formation.

2. Establish Criteria for Evaluating AI Resources. Before recommending or permitting use of an AI tool in parish programs, verify: Does it cite Magisterial sources? Does it acknowledge its limitations? Does it direct people toward community and sacraments rather than substituting for them?

3. Integrate AI Literacy into Formation Programs. Confirmation preparation, OCIA, and adult faith formation should include components on discerning reliable digital sources, recognizing equivocation, and knowing when to seek human guidance.

4. Model Appropriate Use. When you use LLMs, you should verify what is provided and integrate it accordingly while ensuring that the content helps you convey your central message without error. These are tools for human teachers, not replacements.

5. Create Human Touchpoints. If your parish uses AI resources, ensure they always point toward opportunities for personal encounter.

4.7.3 For Lay Users

1. Adopt the 'Three Source Rule'. Never trust LLMs alone for theological information.Verify claims against: (a) the Catechism or other Magisterial documents, (b) consultation with a trusted human teacher (priest, catechist, theologically educated friend), and (c) your own growing understanding formed through prayer and study.

2. Recognize Equivocation Markers. When LLMs use phrases like 'many theologians believe,' 'from one perspective,' or 'some Catholics hold,' recognize that the model is hedging. Press for specificity: 'What does the

Magisterium teach?' If the LLM cannot answer clearly, consult authoritative sources.

3. Know When Not to Use LLMs. Do not use LLMs for confession preparation, vocational discernment, interpreting personal spiritual experiences, or urgent moral crises. AI provides information; it does not provide wisdom, discernment, or pastoral accompaniment.

4. Engage Critically. If a response seems wrong, seems to contradict what you've learned elsewhere, or simply doesn't sit right in your conscience, don't dismiss your intuition. Look it up, ask someone, pray over it.

5. Use LLMs to Deepen Community Engagement, Not Replace It. If an LLM explanation sparks questions, bring them to your religious formation groups and share them with others. If an LLM provides historical context on a saint, let that inspire conversation with fellow parishioners. Let digital tools serve, not substitute for, the embodied communion that is the Church's life.

4.8 The Ethical Challenge of Formation

The Church's ethical response cannot rely solely on regulation or technical guidelines; it must form hearts and minds capable of discernment. *Evangelii nuntiandi* declares that the witness of life is the first form of evangelisation.[81] In the digital context, this means modeling responsible engagement: using AI to illuminate, not replace, the human encounter with truth. Seminaries, universities, and dioceses should integrate courses on digital ethics and theology, enabling future clergy and laity to navigate these technologies with wisdom. Formation must be adaptive yet rooted in tradition; a synthesis of *Antiqua et nova*, the old and the new.[82]

4.9 Evaluating the Morality of AI Use

From a Thomistic perspective, moral evaluation depends on object, intention, and circumstances.[83] The object of using AI for theological purposes is not intrinsically evil; indeed, it can serve evangelisation and education. The moral intention of seeking truth and serving others is good when properly ordered. Circumstances such as anonymity, algorithmic opacity, and data-driven bias, however, often complicate judgment. This is because they may render actions morally ambiguous. Therefore, prudence (*prudentia digitalis*) becomes the governing virtue. As Aquinas wrote, prudence is 'right reason in action'.[84] A prudent digital theology would ask: Does this use of AI clarify or obscure truth? Does it cultivate humility or pride? Does it serve the common good or personal convenience?

81. Paul VI, *Evangelii Nuntiandi: Apostolic Exhortation to the Episcopate, Clergy and Faithful on Evangelization in the Modern World* (Vatican City: The Holy See – Vatican Publishing House, 1975), §20, https://www.vatican.va/content/paul-vi/en/apost_exhortations/documents/hf_p-vi_exh_19751208_evangelii-nuntiandi.html.

82. Doctrine of the Faith, Culture, and Education, *Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence*.

83. Aquinas, *Summa Theologiae*, see I-II, qq. 18–20.

84. Aquinas, ST II–II, q.47, a.2.

5 Conclusion

Equivocation and erosion outputted from LLMs create an ethical tension between technological capability and theological truth. LLMs, though powerful tools for information access, operate within a moral vacuum unless guided by a human conscience formed in faith. Their tendency to flatten distinctions, mirror cultural biases, and dilute doctrinal precision poses a risk not only to Catholic theology but to the integrity of religious discourse itself. Even though this is a challenge, they unveil an opportunity: to reclaim the Church's prophetic role in shaping the moral imagination of the digital age.

The path forward demands a theology of technology grounded in CST, guided by Theological and ethical formation, while remaining open to ecumenical cooperation. The Church must engage AI not as an adversary but as a field of mission, an '*areopagus* of the modern world'.[85] Through initiatives such as the Rome Call for AI Ethics and interdisciplinary collaborations among theologians, computer scientists, and ethicists, a new humanism of technology can emerge. One that respects truth, upholds dignity, and promotes communion.

In this vision, equivocation is replaced by clarity, erosion by renewal. LLMs that are used as a force for good can become servants of truth rather than its substitutes. Through Jesus' Incarnation, we see that divine wisdom is expressed in human language, not diluted by it. As Christ entered history to sanctify human speech, so too must the Church enter the digital sphere to sanctify the word.

Acknowledgments

We thank Sr. Mary Catherine Hilkert, O.P., Dr. Megan Levis-Scheirer, and Mimi Beck for their guidance and theological insight. This research was supported in part by the Sorin Fellows Program at the De Nicola Center for Ethics and Culture and by the Ansari Institute for Global Engagement with Religion at the University of Notre Dame.

References

Aquinas, Thomas. *Summa Theologiae*. Edited by Thomas Gilby. Latin text and English translation, Blackfriars edition. Cambridge: Cambridge University Press, 1964.

Askell, Amanda, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Ben Mann, Nova DasSarma, et al. "A general language assistant as a laboratory for alignment." *arXiv preprint arXiv:2112.00861*, 2021.

Augustine. *Confessions*. Translated by Henry Chadwick. Oxford World's Classics edition. Oxford: Oxford University Press, 1991.

85. John Paul II, *Redemptoris Missio*, Vatican: Libreria Editrice Vaticana, Encyclical, On the Permanent Validity of the Church's Missionary Mandate, 1990, §37, https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_07121990_redemptoris-missio.html.

Beauchamp, Tom. "The principle of beneficence in applied ethics," 2008.

Beauchamp, Tom, and James Childress. *Principles of biomedical ethics: marking its fortieth anniversary*, 11, 2019.

Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the dangers of stochastic parrots: Can language models be too big?" In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623. 2021.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, et al. "On the Opportunities and Risks of Foundation Models." *arXiv preprint arXiv:2108.07258*, 2021. https://arxiv.org/abs/2108.07258.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. "Sparks of artificial general intelligence: Early experiments with gpt-4." *arXiv preprint arXiv:2303.12712*, 2023.

Canterbury, Anselm of. *Proslogion*. Translated by Thomas Williams. Originally written c. 1077–1078. Indianapolis: Hackett Publishing, 2001.

Catechism of the Catholic Church. Second Edition. Vatican City: Libreria Editrice Vaticana, 1997.

Center, Death Penalty Information. "Pope Francis Responsible for Catholic Church's Anti-Death Penalty Teachings Dies at 88." Accessed February 16, 2026, 2025. https://deathpenaltyinfo.org/pope-francis-responsible-for-catholic-churchs-anti-death-penalty-teachings-dies-at-88.

Chalcedon, Council of. *Definition of the Faith*, 451. https://www.papalencyclicals.net/councils/ecum04.htm.

Chow, Andrew R. *Pope Leo's Name Carries a Warning About the Rise of AI*. Time Magazine, 2025. https://time.com/7285449/pope-leo-artificial-intelligence/.

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 (2017).

Churches, World Council of. *Statement on the Unregulated Development of Artificial Intelligence*. Central Committee meeting, Geneva, June 21–27, 2023. https://www.oikoumene.org/resources/documents/statement-on-the-unregulated-development-of-artificial-intelligence.

Clark, Elizabeth A. *Reading renunciation: Asceticism and scripture in early Christianity*. Princeton University Press, 1999.

Congregation for the Doctrine of the Faith. *Dignitas Personae: Instruction on Certain Bioethical Questions*. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_20081208_dignitas-personae_en.html. 8 September 2008, September 2008.

Congregation for the Doctrine of the Faith. *Donum Vitae: Instruction on Respect for Human Life in Its Origin and on the Dignity of Procreation*. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_19870222_respect-for-human-life_en.html. 22 February 1987, February 1987.

Conwill, Louisa, Megan Levis Scheirer, and Walter J. Scheirer. *Virtue in Virtual Spaces: Catholic Social Teaching and Technology*. Liturgical Press, 2024. ISBN: 9798400800269. https://www.papalencyclicals.net/councils/ecum04.htm.

Cook, Terry. "Evidence, memory, identity, and community: four shifting archival paradigms." *Archival science* 13, no. 2 (2013): 95–120.

Council, Second Vatican. *Dei Verbum: Dogmatic Constitution on Divine Revelation*, 1965. https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19651118_dei-verbum_en.html.

———. *Unitatis Redintegratio*, 1964. https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decree_19641121_unitatis-redintegratio_en.html.

Crane, Greg. "Classics and the computer: an end of the history." *A Companion to Digital Humanities*, 2004, 46–55.

Doctrine of the Faith, Dicastery for the, Dicastery for Culture, and Education. *Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence*. Vatican: Libreria Editrice Vaticana. Note, 2025. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html.

Ethics in Communications. Vatican document, 2000. https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_20000530_ethics-communications_en.html.

Florovsky, Georges V. *Ways of Russian Theology*. Edited by Richard S. Haugh. Translated by Robert L. Nichols. Collected Works of Georges Florovsky, Vol. V (Part One). Belmont, MA: Nordland Publishing Company, 1979.

Francis. *Fratelli Tutti*. Vatican: Libreria Editrice Vaticana. Encyclical. On Fraternity and Social Friendship, 2020. https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html.

———. *Laudato Si': On Care for Our Common Home*. Vatican: Libreria Editrice Vaticana. Encyclical. Encyclical Letter on Care for Our Common Home, 2015. https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html.

———. *Traditionis Custodes: On the Use of the Roman Liturgy Prior to the Reform of 1970*. https://www.vatican.va/content/francesco/en/motu_proprio/documents/20210716-motu-proprio-traditionis-custodes.html. Apostolic Letter issued motu proprio, 16 July 2021, July 2021.

Gaudium et Spes: Pastoral Constitution on the Church in the Modern World. §12. Vatican Publishing House, 1965. https://www.clerus.org/bibliaclerusonline/en/eg0.htm.

Helland, Christopher. "Online religion as lived religion. Methodological issues in the study of religious participation on the internet." *Online-Heidelberg Journal of Religions on the Internet*, 2005.

II, John Paul. *Fides et Ratio: On the Relationship Between Faith and Reason*. Vatican: Libreria Editrice Vaticana. Encyclical. Encyclical Letter on the Relationship Between Faith and Reason, 1998. https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091998_fides-et-ratio.html.

———. *Laborem Exercens: On Human Work*. Vatican City: The Holy See – Vatican Publishing House, 1981. https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091981_laborem-exercens.html.

———. *Redemptoris Missio*. Vatican: Libreria Editrice Vaticana. Encyclical. On the Permanent Validity of the Church's Missionary Mandate, 1990. https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_07121990_redemptoris-missio.html.

———. *Ut Unum Sint: Encyclical Letter on the Commitment to Ecumenism*. Encyclical of the Holy Father John Paul II, 25 May 1995, 1995. https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_25051995_ut-unum-sint.html.

Index Librorum Prohibitorum. Superseded list of prohibited books, guidance via imprimatur and nihil obstat. Vatican Publishing House, 1966. https://www.britannica.com/topic/Index-Librorum-Prohibitorum.

IRIC. *Introducing hyder.ai: The First AI Model Trained on Shia Islamic Teachings*. https://iric.org/introducing-hyder-ai-the-first-ai-model-trained-on-shia-islamic-teachings/, January 2025.

Justice, Pontifical Council for, and Peace. *Compendium of the Social Doctrine of the Church*, 2004. https://www.vatican.va/roman_curia/pontifical_councils/justpeace/documents/rc_pc_justpeace_doc_20060526_compendio-dott-soc_en.html.

LaBanca, Nicholas. *"The Other 23 Catholic Churches (Rites) and Why They Exist"*. Blog post on Ascension Press, 2019. https://ascensionpress.com/blogs/articles/the-other-23-catholic-churches-and-why-they-exist?srsltid=AfmBOopCi-uZt878N2nhx60_o7rjP67I0QKpOfobu3sPbld6mwMzeeHW.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in neural information processing systems* 33 (2020): 9459–9474.

Life, Pontifical Academy for. *AI Ethics: An Abrahamic Commitment to the Rome Call*. Signing event, Vatican, January 10, 2023. https://www.academyforlife.va/content/pav/en/news/2023/rome-call-ai-ethics-interreligious-signing.html.

———. *Rome Call for AI Ethics*. Vatican: Pontifical Academy for Life. Ethical Framework. Joint statement on the ethical development and use of artificial intelligence, 2020. https://www.romecall.org/.

"Liturgical Languages." In *New Catholic Encyclopedia*, 2nd ed., 8:612–618. Detroit: Gale, 2003. ISBN: 9780787640040. https://www.encyclopedia.com/religion/encyclopedias-almanacs-transcripts-and-maps/liturgical-languages.

Magisterium AI. *World's #1 Answer Engine for the Catholic Church*. https://www.magisterium.com. Accessed: 2025-11-03.

Notre Dame, University of. *Notre Dame's R.I.S.E. AI Conference Builds Interdisciplinary Collaboration to Inform Human-Centered Artificial Intelligence*, 2025. https://strategicframework.nd.edu/news/notre-dames-r-i-s-e-ai-conference-builds-interdisciplinary-collaboration-to-inform-human-centered-artificial-intelligence/.

Pontifical Council for Social Communications. Online. Guidance on Church engagement with media. https://en.wikipedia.org/wiki/Pontifical_Council_for_Social_Communications.

RavChat. *About — Your Advanced AI Assistant for Torah Learning*. https://rav.chat/about. Accessed: 2025-11-03.

Second Vatican Council. "Decree Ad Gentes on the Mission Activity of the Church." Vatican.va. Accessed: 2025-12-30, December 7, 1965. https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decree_19651207_ad-gentes_en.html.

———. *Inter Mirifica: Decree on the Media of Social Communication*. Vatican Publishing House, 1963. https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decree_19631204_inter-mirifica_en.html.

Siemens, Ray, and Susan Schreibman. *A companion to digital literary studies*. John Wiley & Sons, 2013.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for modern deep learning research." In *Proceedings of the AAAI conference on artificial intelligence*, 34:13693–13696. 09. 2020.

The Church and Internet. Vatican document. https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_20000530_ethics-communications_en.html.

The New Revised Standard Version Bible: Catholic Edition. Division of Christian Education. Nashville, TN: National Council of Churches of Christ, 1989.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

VI, Paul. *Evangelii Nuntiandi: Apostolic Exhortation to the Episcopate, Clergy and Faithful on Evangelization in the Modern World*. Vatican City: The Holy See – Vatican Publishing House, 1975. https://www.vatican.va/content/paul-vi/en/apost_exhortations/documents/hf_p-vi_exh_19751208_evangelii-nuntiandi.html.

Winner, Langdon. "Do artifacts have politics?" In *Computer ethics*, 177–192. Routledge, 2017.

Xian, Yongqin, Bernt Schiele, and Zeynep Akata. "Zero-shot learning-the good, the bad and the ugly." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4582–4591. 2017.

XII, Pius. *Munificentissimus Deus: Apostolic Constitution Defining the Dogma of the Assumption*. https://www.vatican.va/content/pius-xii/en/apost_constitutions/documents/hf_p-xii_apc_19501101_munificentissimus-deus.html. 1 November 1950, November 1950.

XVI, Benedict. *Caritas in Veritate*, 2009. https://www.vatican.va/content/benedict-xvi/en/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate.html.

XXIII, St. John. *Mater et Magistra: On Christianity and Social Progress*. §219. Vatican Publishing House, 1961. http://www.vatican.va/content/john-xxiii/en/encyclicals/documents/hf_j-xxiii_enc_15051961_mater.html.