Equivocation and Erosion: How LLMs Undermine Catholic Religious Discourse

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

Abstract

Large Language Models (LLMs) offer opportunities for information dissemination, yet present challenges that uphold 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.

LLMs often smooth over specific religious claims. Unlike human religious authorities who may offer definitive interpretations based on scripture, tradition, and Magisterial teaching, LLMs present information in a flattened, generalized manner. 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 prioritizing the specific moral decrees found within Canon Law or the Catechism of the Catholic Church. 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., Chat-GPT, Llama, Claude, Gemini) and theological LLMs from differing religious traditions (e.g., Magisterium AI, Hyder.ai, RavGPT) align with Catholic teaching. LLMs may inadvertently marginalize minority viewpoints within the Catholic Church or prioritize 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.