TIAGO ANDRÉ FERNANDES FREITAS*
FRANCISCO MARCONDES**

ECCLESIAL DISCERNMENT IN THE ERA OF GENERATIVE ARTIFICIAL INTELLIGENCE

Received: 18 January 2026
Accepted: 24 February 2026

ABSTRACT: This article offers an ecclesiological analysis of discernment under conditions shaped by generative artificial intelligence (AI). It examines how generative AI may reshape synodal listening by producing statistically central outputs that can be mistaken for ecclesial reception. After clarifying the theological role of the *sensus fidei fidelium* in discernment, the article analyses how large language models generate responses through probabilistic prediction, tending towards convergence and centrally weighted answers. Three tensions structure the argument: who the subject of discernment is, what kind of consensus is at stake, and how centre-periphery dynamics are affected by statistical centrality. An exploratory study reports three empirical patterns: output instability across prompts and settings, shifting registers of epistemic authority through generated voice, and the quantitative suppression of minority signals under aggregation and summarisation. The article concludes with theological-pastoral criteria that distinguish conditionally acceptable uses from imprudent ones, emphasising agency, traceability, and attention to the margins.

KEY WORDS: synodal listening; *sensus fidei*; epistemic authority; statistical averaging; text aggregation; auditability.

---

* Universidade Católica Portuguesa: tfreitas@ucp.pt; ORCID: https://orcid.org/0000-0003-0211-328X

** Universidade do Minho: fm@di.uminho.pt; ORCID: https://orcid.org/0000-0002-2221-2261

ESTUDIOS ECLESIÁSTICOS, vol. 101, núm. 397, mayo 2026, 337-370, ISSN 0210-1610, ISSN-e 2605-5147
DOI: 10.14422/ee.v101.i397.y2026.004

Discernimiento eclesial en la era de la inteligencia artificial generativa

RESUMEN: Este artículo ofrece un análisis eclesiológico del discernimiento en unas condiciones configuradas por la inteligencia artificial (IA) generativa. Examina cómo la IA generativa puede remodelar la escucha sinodal al producir resultados estadísticamente centrales que pueden confundirse con la recepción eclesial. Tras aclarar el papel teológico del *sensus fidei fidelium* en el discernimiento, el artículo analiza cómo los modelos de lenguaje de gran escala generan respuestas mediante predicción probabilística, tendiendo a la convergencia y a formulaciones ponderadas hacia el centro. Tres tensiones estructuran el argumento: quién es el sujeto del discernimiento, qué tipo de consenso está en juego y cómo la dinámica centro-periferia se ve afectada por la centralidad estadística. Un estudio exploratorio presenta tres patrones empíricos: inestabilidad de los resultados según los *prompts* y los ajustes, desplazamiento de los registros de autoridad epistémica a través de la voz generada y supresión cuantitativa de señales minoritarias bajo procesos de agregación y resumen. El artículo concluye con criterios teológico-pastorales que distinguen los usos aceptables bajo condiciones de aquellos imprudentes, subrayando la agencia, la trazabilidad y la atención a los márgenes.

PALABRAS CLAVE: escucha sinodal; *sensus fidei*; autoridad epistémica; promedio estadístico; agregación de textos; auditabilidad.

1. INTRODUCTION

The recent spread of conversational interfaces built on language models has reshaped how people access knowledge and seek guidance in natural language. Use now extends beyond instrumental research to writing support, decision-making, and everyday advice, often outside work-related contexts.[1]

This shift has an immediate theological consequence. As AI becomes a commonplace interlocutor, it is consulted on questions that exceed the informational. In areas adjacent to spiritual direction, pastoral counselling, or moral formation, textual fluency is easily read as competence and credibility. Therapeutic chatbots have therefore prompted frameworks to assess safety, transparency, and risk, since algorithmic mediation can shape vulnerable reliance.[2] When mediation enters practices

[1] Aaron Chatterji et al. "How People Use ChatGPT" (Working paper no. w34255, National Bureau of Economic Research, 2025), 14.

[2] Kunmi Sobowale and Daniel Humphrey. "Evaluating the Quality of Psychotherapy Conversational Agents: Framework Development and Cross-Sectional Study." *JMIR Formative Research* 9 (2025): 4. https://doi.org/10.2196/65605

where trust, vulnerability, and normativity are at stake, the question is not only whether it works, but what legitimises it and how it is received.

Within the religious sphere, spiritual chatbots have become an explicit object of theological reflection. Recent discussion notes their feasibility and insists on delimiting their place, especially as they approach ministerial and pastoral care.[3] Sociotechnical analyses argue that AI-accelerated digitalisation encourages anthropomorphising technology and sustains a «new social form of techno-religion»[4] that contests languages of transcendence. In such a context, the question is how cultural expectations of omniscience and technical efficacy can displace attributes reserved for the living God.

The ecclesiological dimension is particularly sensitive in synodal contexts. Synodality depends on reception, recognisable mediations, and responsibilities that can be narrated historically. Yet dioceses face pressure to process vast volumes of text. Automated clustering and summarisation promise efficiency but can also stabilise «common points» through opaque parameters, and a polished summary may be mistaken for received consensus when it reflects statistical convergence rather than discerned reception. This risk is heightened when the method is not disclosed and readers presume that a synthesis already embodies communal judgement.

Against this background, the present article has three aims. First, it articulates the tensions between reception and algorithmic convergence, the believing subject and the statistical production of language, and between the peripheries of the Gospel and dynamics of centralisation. Second, it tests how different configurations of AI use may affect minority visibility and the form of textual consensus. Third, it outlines theological and pastoral implications for responsible AI use in processes of listening and synthesis.[5]

---

[3] Ron Cole-Turner. "Artificial Intelligence and Human Spirituality: Is a Spiritual Chatbot a Good Idea?" *Theology and Science* 23, no. 3 (2025): 478. https://doi.org/10.1080/14746700.2025.2514299

[4] Michael Latzer. "Digitalization, AI and the Rise of Techno-Religion: Transhumanist Promises and the Challenge to Enlightenment." *Telecommunications Policy* 50, no. 2 (2026): 2. https://doi.org/10.1016/j.telpol.2025.103115

[5] Hope Schroeder, Deb Roy and Jad Kabbara. "Just Put a Human in the Loop? Investigating LLM-Assisted Annotation for Subjective Tasks." *arXiv* (2025): 1. https://doi.org/10.48550/ARXIV.2507.15821

2. *SENSUS FIDEI* AND ECCLESIAL DISCERNMENT

Reflection on the impact of artificial intelligence on the life of the Church becomes theologically significant when it rests on a clear understanding of who the believing subject is, and of the proper modes of access to revealed Truth and to «theological understanding.»[6] It is therefore necessary to recall what it means, within the Catholic tradition, to speak of the *sensus fidei fidelium* and of ecclesial discernment.

2.1. The believing subject

The concept of *sensus fidei* underwent significant development in the ecclesiology of the Second Vatican Council. Moving beyond a rigid opposition between a teaching Church (*Ecclesia docens*) and a learning Church (*Ecclesia discens*),[7] *Lumen Gentium* (no. 12) presents the People of God as a true believing subject, endowed with a supernatural sense of faith. The conciliar text recalls that the whole body of the faithful, anointed by the Holy One (cf. 1Jn 2:20, 27), «cannot err in matters of belief» (*in credendo falli nequit*).

Within this horizon, *sensus fidei* can be described as a spiritual instinct by which believers recognise what is consonant with the faith received from the apostles. It is best understood «as practical rather than speculative knowledge»,[8] taking concrete form as a *praxis fidei*, namely «the action through which, precisely by acting, the believer receives a gift that is not produced but only received.»[9] Post-conciliar theology often situates it within *fides quaerens intellectum*: faith seeks understanding, and reason «continually extends the range of its knowledge»[10] in order to enter more deeply into the revealed mystery.

---

[6] Lluís Oviedo. "Artificial Intelligence and Theology: Looking for a Positive - but Not Uncritical – Reception." *Zygon: Journal of Religion and Science* 57, no. 4 (2022): 939. https://doi.org/10.1111/zygo.12832

[7] Commissione Teologica Internazionale. "Il *sensus fidei* nella vita della Chiesa (2014)." In *Documenti 2005-2021*, 4. Bologna: Edizioni Studio Domenicano, 2022.

[8] John Burkhard. *The "Sense of the Faith" in History. Its Sources, Reception, and Theology*. Collegeville: Liturgical Press Academic, 2022, 245.

[9] João Duque. "A teologia como caminho. Considerações sobre o método teológico." *Didaskalia* XXXIX, no. 2 (2009): 25.

[10] John Paul II. "Encyclical Letter *Fides et Ratio* on the relationship between faith and reason." *AAS* 91 (1999): 14.

This sense of faith is intrinsically communal. Each baptised person receives the Spirit's anointing, yet tenacity in believing belongs to the *universitas fidelium*. The Church as a body, «is the subject of this integral process of Tradition and reception»,[11] discerning whether doctrine, custom, or reform genuinely translates the Gospel. Such reception unfolds over time and can involve interpretation, purification, and even conflict before a stable ecclesial recognition emerges.

The *sensus fidei* is to be understood in communion with the Magisterium. The faithful respond with a believing assent, an *obsequium* that is intelligent and responsible. Congar's notes that «with regard to certain structures of faith, to ethical norms, and to worship which history requires to be specified on the basis of the original apostolic transmission, the faithful and the local Churches are neither inert nor purely passive.»[12] Both gifts converge in the shared task «to hear, distinguish and interpret the many voices of our age, and to judge them in the light of the divine word.»[13]

In summary, the *sensus fidei fidelium* has a supernatural origin, is vitally rooted in lived faith, and bears a distinctly ecclesial mark, operating within communion. It articulates the faithful's spiritual experience with the Magisterium's service of discernment, in shared listening to the one Word of God, to whom the whole Church remains accountable.

2.2. Listening and synodal reception

The current commitment to synodality places listening at the heart of the Church's life. Scripture states that faith comes from hearing (*fides ex auditu*, cf. Rom 10:17). Yet faith, as a response to God's self-communication, is a personal reception shaped by a concrete horizon, since the interpretation of the Word of God is traversed by the time and context in which we live. The reception of Revelation is therefore always historically

[11] Angel Antón. "La 'Recepción' en la Iglesia y Eclesiología (I)." *Gregorianum* 77, no. 1 (1996): 453.

[12] Yves Congar. "La Recepción Como Realidad Eclesiológica." *Concilium* 77 (1972): 78.

[13] Second Vatican Ecumenical Council. "Pastoral Constitution *Gaudium et Spes* on the Church in the Modern World." *AAS* 58 (1966): 44.

situated.[14] It is precisely within this movement that the *sensus fidei* becomes decisive.

To grasp this dynamic more fully, it is crucial to attend to the action of the Spirit. The Spirit prepares the inner ground, opens understanding, and gathers into one confession of faith persons marked by diverse histories and cultures. The *sensus fidei* can thus be understood as «"the principle of reception" in the process of divine revelation.»[15] Each baptised person is called to a twofold exercise of discernment: *auditus fidei*, the obedient hearing of faith, and *auditus temporis*, attentive listening to the appeals that emerge from the present time. «The *sensus fidei* is only an authentic *sensus fidei*, that is, one which, with *recto iudicio*, seeks to be wholly immersed in Revelation, if it is consistently shown as a *fides ex auditu*; if, therefore, it listens with unwavering perseverance to the proclamation of the Word of God and obeys it.»[16]

In a synodal Church, listening to the voice of the People of God therefore means making room for this lived faith to manifest itself. In this process, reception constitutes a decisive criterion. When a magisterial or synodal text is promulgated, a path of «reception and appropriation in human lives»[17] is opened, in which the community allows that teaching to enter prayer, to shape the liturgy, and to mark concrete choices. Only then does what has been formulated in words begin to become a way of life.

The history of the Church shows long periods in which certain evangelical intuitions remained in tension. They were contested and required purification and renewed formulation before becoming consolidated as a shared patrimony. In practice, «synodal ecclesiology makes space for disagreement and conflict in the theological conception of the church.»[18] For that reason, synodal listening is not assessed by numerical unanimity.

---

[14] Heinrich Fries. *Fundamental Theology*. Washington: Catholic University of America Press, 1996, 182.

[15] Ormond Rush. *The Eyes of Faith. The Sense of the Faithful and the Church's Reception of Revelation*. Baltimore: Catholic University of America Press, 2011, 15.

[16] Hans Küng. "La Estructura Carismática de La Iglesia." *Concilium,* no. 4 (1965): 48-49.

[17] Joseph Komonchak. "Defending Our Hope. On the Fundamental Tasks of Theology." In *Faithful Witness. Foundations of Theology for Today's Church*, directed by Leo J. O'Donovan, 20. New York: Crossroad, 1989.

[18] Judith Gruber. "Consensus or Dissensus?" *Louvain Studies* 43 (2020): 242. https://doi.org/10.2143/LS.43.3.3288706

It is measured by the pursuit of a moral consensus.[19] This is a stable convergence between faith and life in which the community, sometimes after prolonged debate, recognises, within a certain theology of dissent,[20] that a given orientation serves fidelity to the Gospel, strengthens communion, and deepens charity.[21]

2.3. Consensus and the margins

Based on these elements, it becomes possible to state more precisely several lines of force that will matter when we later confront the *sensus fidei* with the operation of generative AI.

Consensus exceeds the arithmetic of statistical calculation. Its validity is recognised by its capacity to generate institutional conversion and interior freedom. It is not established by the sheer frequency with which a topic is repeated in reports. When *Lumen Gentium* speaks of a universal consensus, it refers to that harmonious convergence which relates «the evidence of the truth of the faith and the Apostolic Tradition.»[22]

Finally, the synodal dynamic requires a reconfiguration of the gaze between centre and margins. By evangelical imperative, the Church turns towards the small and the discarded,[23] granting particular attention to those voices that inhabit the periphery of processes.[24] This includes persons with less cultural capital, geographically distant communities, and groups that are wounded or scarcely visible.

This openness, however, does not amount to an automatic canonisation of everything that comes from the margins, since these too can be

---

[19] John Henry Newman. *On Consulting the Faithful in Matters of Doctrine*. Kansas City: Sheed & Ward, 2012, 73.

[20] Cf. John Caputo. *The Insistence of God: A Theology of Perhaps*. Bloomington: Indiana University Press, 2013.

[21] Commissione Teologica Internazionale. "Il *sensus fidei* nella vita della Chiesa (2014)," 73.

[22] Pavol Hrabovecký. "John Henry Newman's Understanding of the Sensus Fidelium and the Current Challenges of Synodality." *Auc Theologica* 12, no. 1 (2022): 34.

[23] Francis. "Encyclical Letter *Fratelli Tutti* on Fraternity and Social Friendship." *AAS* 112, no. 11 (2020): 188.

[24] Cf. Ikenna Okpaleke. "Of Skeptics and Believers: Observing the Synod on Synodality from the Margins." *MST Review* 26, no. 1 (2024): 15. https://doi.org/10.5281/ZENODO.12779058

permeable to ideologies and resentments.[25] The true challenge of discernment lies in purifying such noise, distinguishing what comes from the Spirit from what is merely the expression of unhealed wounds. Even so, a synodal Church refuses to decide about others without listening to them, aware that it is often at the periphery that the clearest calls to conversion emerge.

These three coordinates delineate the horizon within which the question of artificial intelligence must be situated. If, at the technical level, generative AI operates preferentially by approximation to the statistical mean and by the smoothing of tensions, it becomes necessary to examine whether such a logic is compatible with the distinctive configuration of the *sensus fidei*.

3. HOW AI PRODUCES CONSENSUS

To ensure that ecclesial discernment does not slip into naivety or abstraction, it becomes imperative to enter the world of technology. It is not enough to analyse the outputs that artificial intelligence presents. It is necessary to understand the internal logic that governs their production. Whereas theology recalls that the Spirit blows where it wills and often manifests itself at unexpected margins, the engineering of generative AI operates according to a probabilistic dynamic. It is oriented by statistical regularities extracted from data and fine-tuned to produce linguistically plausible sequences.

We therefore propose a brief technical foray into the workings of Large Language Models (LLMs). The aim is to offer theology conceptual tools that enable it to understand how the machine ascribes plausibility to statements, distinguishing those it considers more likely from those it does not. It will become clear how the architecture of these models, in their pursuit of coherence and discursive fluency, tends to produce an effect of statistical filtering. This favours dominant patterns and reduces the visibility of rare, idiosyncratic, or dissonant formulations. Yet it

[25] Elisa López and Nurya Martínez-Gayol. "«Escuchar, dialogar y discernir» con las mujeres: Retos de una Iglesia Sinodal." *Estudios Eclesiásticos* 97, no. 381-382 (2022): 566. https://doi.org/10.14422/ee.v97.i381-382.y2022.009

is precisely in such formulations that the Christian tradition has often recognised a theological locus.

3.1. Prediction and patterns

Figure 1 presents two views of the encoder only architecture, the underlying structure of any Transformer-based language model (LLM). Figure 1a is the structural view as found[26] and Figure 1b summarizes its behaviour as presented in.[27]

FIGURE 1. TWO VIEWS OF THE AUTO-ENCODER ARCHITECTURE

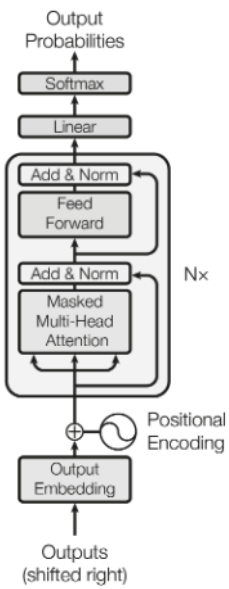

Figure 1a

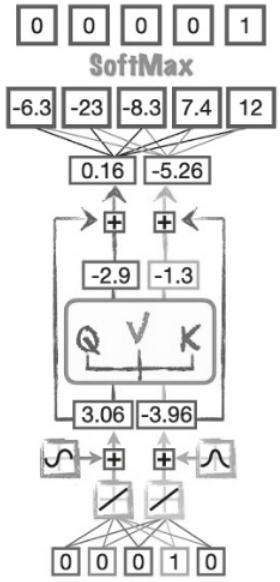

Figure 1b

[26] Ashish Vaswani et al. "Attention Is All You Need." *arXiv* (2017): 30. https://doi.org/10.48550/ARXIV.1706.03762

[27] Josh Starmer. *The StatQuest Illustrated Guide to Machine Learning!!! Triple Bam!!!* USA: StatQuest Publications, 2022.

3.2. Probability as a criterion

As can be seen in both diagrams, the last prediction stage is a function called *softmax*. In short, it converts a tuple of real numbers $z = (z_1, \ldots, z_k)$ into a probability distribution $\sigma: \mathbb{R}^K \to (0,1)^K$. In the case of LLMs, it gets the raw output of the network and sets the whole vocabulary into a probability distribution. *Softmax* serves as the ideal strategic abstraction because it acts as the definitive filter where a raw logits set is transformed into a clear probability hierarchy. The *softmax* function is described as:

$$\sigma(z_t) = \frac{e^{z_i/T}}{\sum_{j=1}^{k} e^{z_j/T}} \qquad (1)$$

where $e \approx 2.71828$ is Euler's number and T the temperature.

Example 1. Let $z = (1.00, 1.25, 1.50, 1.75, 2.00)$ and $T = 1$.

The exponent step, $e^{1.00}$, …, $e^{2.00}$ are, $2.7183, 3.4903, 4.4817, 5.7546, 7.3891$ whose summation is 23.8340. Then

1. For 1.00: $\frac{2.7183}{23.8340} \approx 0.1141$ (11.4%)
2. For 1.25: $\frac{3.4903}{23.8340} \approx 0.1464$ (14.6%)
3. For 1.50: $\frac{4.4817}{23.8340} \approx 0.1880$ (18.8%)
4. For 1.75: $\frac{5.7546}{23.8340} \approx 0.2415$ (24.2%)
5. For 2.00: $\frac{7.3891}{23.8340} \approx 0.3100$ (31.0%)

It is easy to see that any given higher number receives a higher probability. Note that despite the regular interval in the input, higher numbers become increasingly distant from its predecessor. This means that, although the raw logits are close, the distance will be magnified, and the less probable tokens tend to be marginalized. Decoding strategies such as top-k and top-p chose as tokens one among those with high probabilities, thus, the issue remains.

Example 2. Let z as the same and $T = 0.5$, then $\frac{z}{T} = (2.0, 2.5, 3.0, 3.5, 4.0)$ thus, $\sigma = (0.058 \downarrow, 0.096 \downarrow, 0.158 \downarrow, 0.260 \uparrow, 0.429 \uparrow)$

This shows that lowering temperature makes *softmax* to more aggressively penalize marginal tokens. The opposite is also true; higher temperatures make it promote otherwise marginal tokens. As a result, lower temperature implies more normalization whereas higher temperature implies less normalization.

Note that each token is computed based on its previous context, thus in the form of conditional probabilities

$$p(x_t \mid x_{<t}) = \sigma(z_t) \qquad (2)$$

note that z_t is the vector of logits generated from the context $x_{<t}$ having the same dimension as the entire vocabulary of the model. Thus, the probability distribution $p = \sigma(z_t)$. The next token x_t is sampled from this distribution $x_t \sim \sigma(z_t)$. In words, x_t is selected given the context $x_{<t}$, with x_t typically being one of the highest scored tokens within the *softmax* distribution.

The recurrent step is given by appending the sampled token to the context

$$x_{<t+1} = [x_{<t}, x_t] \qquad (3)$$

yielding to

Proposition 1. Given an auto-regressive language model where the probability of a sequence $X = (x_1, \ldots, x_t)$ ïis defined by the chain rule of probabilities, selecting the token x_t that maximizes the local conditional probability $p(x_t \mid x_{<t})$ at each step tends to result in a sequence with a higher cumulative probability than any sequence constructed from marginal (lower probabilities) tokens when the temperature T have a reasonable value.

Proof.

1. The probability of the entire generated text *X* is the product of the probabilities of each individual token from the recurrent step from (2) and (3), results in:

$$p(X) = \prod (x_t \mid x_{<t})$$

2. Let $\sigma(z_t)$ be the *softmax* distribution at step and x_{best} a token chosen from the peak of the distribution (normalized result) and x_{marg} be a

token chosen from the tail of the distribution. Given the *softmax* definition in (1):

$$p(x_{best} \mid x_{<t}) > p(x_{marg} \mid x_{<t})$$

3. Consider two text generations one based on maximizing likelihood, thus composed with x_{best} tokens and another minimizing it, thus composed with x_{marg}, follows:

$$\prod p\left(x_{best}^{(t)} \mid x_{<t}\right) \gg \prod p\left(x_{marg}^{(t)} \mid x_{<t}\right)$$

Since every individual factor in is larger than the corresponding factor in x_{marg}, the total probability of a x_{best} sentence is exponentially larger. This implies that filtering marginals at every recurrent step minimizes the accumulation of semantic penalty (negative log-likelihood).

3.3. The logic of the average

This proposition shows that normalization is not just a local event; it is a global stabilizer. By using *softmax* to enforce a high probability at every single step, it is expected that the final text remains within the likely (normalized) human language.

Technically, marginalized tokens are those assigned low probability by the *softmax* function due to semantic misalignment (e.g., predicting «egg» after «the sky is»). While this mathematical filtering is essential for maintaining semantic coherence, it poses a risk to marginal discourses. These discourses are semantically valid but statistically rare in the training data. Therefore, this articles argues that semantic normalization acts as a double-edged sword: the model's architecture essentially conflates «improbable nonsense» with «improbable truth,» leading to the structural marginalization of alternative viewpoints.

4. THREE DECISIVE TENSIONS

In the light of what has been said about the *sensus fidei* and about the functioning of language models, we can now set out three central tensions. The aim is not to oppose Church and technology as monolithic

blocks. Rather, it is to identify, with clarity, the points at which the logic of ecclesial discernment and algorithmic consensus risk intersecting in ambiguous ways. These three tensions concern the subject, the kind of consensus at stake, and the relationship between centre and margins.

4.1. Who discerns?

The first tension concerns the most elementary question: who is discerning when an AI system is introduced into a process of listening and synthesis?

On the ecclesial side, we have seen that the subject of the *sensus fidei* is the People of God. This subject has a face and a history, since «the personal character of faith cannot dispense with its reference to a cultural-institutional-communal configuration, nor can this configuration dispense with its personal reference, without both ultimately collapsing into self-destruction.»[28] In an analogous way, faith is not the direct by-product of a text. As Bouillard reminds us, «the encounter between man and God "takes place only in the miracle and the paradox of faith."»[29]

Moreover, the *sensus fidei fidelium* is inseparable from the sacramental structure of the Church. The capacity to recognise the truth of faith springs from baptism, is nourished by the Eucharist, purified through conversion and matures in charity. There is no *sensus fidei* outside this vital rooting. Baptismal grace is «the essential source of the *sensus fidei.*»[30] Discernment, as a spiritual act, therefore, bears upon the historical reality of the Church and of the cultures that call for a critical and spiritual reading. It cannot be reduced «to mere practices of sociometric technology or the like.»[31]

On the side of AI, the situation is entirely different. A language model is a complex statistical system. It has no self-awareness and it raises the serious challenge of anthropomorphic appearance. This entails the

---

[28] João Duque. *Homo Credens. Para Uma Teologia Da Fé.* Lisboa: Universidade Católica Editora, 2002, 180.

[29] Henri Bouillard. *Logique de la foi.* Paris: Aubier, 1964, 89.

[30] Bernhard Blankenhorn. "The *Sensus Fidei* and *Synodality*: Theological Epistemology and the *Munus Propheticum*." *The Thomist* 87 (2023): 319.

[31] Sergio Lanza. "Progettualità e discernimento pastorale." *Orientamenti Pastorali* 7 (1998): 35.

potential risk «for deception and manipulation at scale and for mistaken beliefs by users that these machines understand human experience and existence in ways they cannot.»[32] We know, moreover, that the model does not believe, does not pray, and does not respond to God's call. It does not know sacramental experience or the obedience of faith. It produces texts that, from a human point of view, may contain true or false claims. Yet the model does not stand before those claims as a moral or spiritual subject. As Véliz concludes, LLMs and their «algorithms cannot be moral agents.»[33]

The tension emerges when, in pastoral practice, one begins to speak of the model as an interlocutor. Common expressions such as «what does the AI think about this?» or «we asked the system what the faithful value most» are symptomatic. Anthropomorphic language is inevitable up to a point, yet it conceals a dangerous slippage. At a certain moment, the community no longer perceives clearly that the text does not come from a believing subject, but from a mechanism of calculation.

This ambiguity is intensified when the model is used to synthesise contributions or to draft ecclesial texts. Uniform language and the assured tone of the sentences create the impression of a stable voice. Yet the problem of hallucinations is well known, and in long texts it poses particular difficulties, «as generations contain intermingled factual and non-factual claims across multiple sentences and paragraphs.»[34] In practice, there is a risk that many will mistakenly read that voice as simply the voice of the Church. At the very least, it can be taken as the voice of a higher instance that has already done the work of discernment.

Theologically, this fusion of planes is open to criticism. The subject of ecclesial discernment cannot be displaced, even partially, onto a system that does not participate in faith. AI can support the work of persons and institutions. It can help to organise information, to clarify texts, and to cross-check data. Discernment, however, as an act of faith and obedience to the Spirit, remains inseparable from concrete human and ecclesial subjects.

---

[32] Sandra Peter, Kai Riemer, and Jevin West. "The Benefits and Dangers of Anthropomorphic Conversational Agents." *PNAS* 122, no. 22 (2025): 2. https://doi.org/10.1073/pnas.2415898122

[33] Carissa Véliz. "Moral Zombies: Why Algorithms Are Not Moral Agents." *AI & SOCIETY* 36, no. 2 (2021): 489. https://doi.org/10.1007/s00146-021-01189-x

[34] Jiatong Han et al. "Simple Factuality Probes Detect Hallucinations in Long-Form Natural Language Generation." *EMNLP* no. 4-9 (2025): 16216.

This first tension leads to a decisive question. When a theological document relies on AI to prepare a synthesis, who is responsible for what is written? Who bears theological responsibility for that formulation? As Matthias warns, «we face an ever-widening *responsibility gap*.»[35] If the only honest answer cannot name persons and ecclesial instances, then something has shifted dangerously from the believing subject to the system.

4.2. Reception and algorithmic convergence

The second tension centres on the notion of consensus. Both synodal discourse and the functioning of LLMs employ a vocabulary that can appear to converge. We speak of consensus, of syntheses, of common points. Yet the assumptions at work in each of these systems are of a different kind.

At the theological level, the consensus associated with the *sensus fidei* is a moral and spiritual consensus, the fruit of a process of reception. When *Lumen Gentium* speaks of a «universal consensus in matters of faith and morals,» it is not referring to a vote or to a statistical average of prevailing opinions. It refers to the Church's gradual convergence, among the faithful, the pastors, and theology, in recognising that «their witness of faith, their customs, and their ministries should become the object of tradition or reception.»[36] Such convergence presupposes time. It encounters resistance. It passes through crises and purifications.

From a historical perspective, the evidence is robust. By way of example, dogmatic definitions, particularly Christological and Mariological, were preceded by lengthy periods of maturation, marked by heresiological debates, synods, and complex processes of reception. As Edward Kilmartin noted, «formulations of revelation are a tributary of the concrete experience of faith lived by a community, whether this be in the form of dogmas or liturgy which crystallizes the governing interests of churches.»[37] This indicates that the process leading to dogmatic formulations

---

[35] Andreas Matthias. "The Responsibility Gap. Ascribing Responsibility for the Actions of Learning Automata." *Ethics and Information Technology* 6 (2004): 176.

[36] Hervé Legrand. *Initiation à La Pratique de La Théologie.* Paris: Cerf, 1983, 160.

[37] Edward Kilmartin. "Reception in History: An Ecclesiological Phenomenon and Its Significance." *Journal of Ecumenical Studies,* no. 21 (1984): 52.

begins, in fact, with the reception of the people's lived faith, and only subsequently consolidates as an authentic expression of the *sensus fidei*. Ecclesial consensus is therefore a persistent dialectic between centre and periphery, Magisterium and laity. At certain moments it can involve only a partial acceptance of a teaching, since «process of reception, however, is not a simple act of obedience and passive absorption.»[38]

Algorithmic consensus, by contrast, arises from the convergence of probabilities within a model. The system seeks to minimise average error in predicting the next word.[39] The convergence it attains is the result of operations upon data and successive adjustments to quality criteria defined by those who train it. It learns to reproduce patterns that already exist, with a variable degree of originality, but it participates in no itinerary of reception. It does not experience resistance to a demanding word, nor does it mature in the understanding of a mystery. It is not converted by what it utters.

Moreover, as recent studies show, «LLMs are highly sensitive to the nuances of prompts,»[40] and minimal fluctuations in the prompt or in configuration parameters can yield divergent syntheses.[41] Such volatility is not, in itself, an anomaly. It follows from the ontology of these systems. Yet the phenomenon underscores the precariousness of algorithmic consensus, which remains dependent on variables that often operate opaquely for the user.

The tension becomes particularly grave when, within the synodal dynamic, AI is used to draft documents that purport to describe the «will of the People of God» or to identify axes of consensus, without making explicit the contingent and dependent nature of these tools. Acting in good faith, a reader may interpret the text as the outcome of prolonged communal discernment, when in fact it represents only the result of pre-established technical parametrisations.

---

[38] Peter Phan. "Reception of and Trajectories for Vatican II in Asia." *Theological Studies* no. 74 (2013): 304.

[39] Yoshua Bengio et al. "A Neural Probabilistic Language Model." *Journal of Machine Learning Research* no. 3 (2003): 1138.

[40] Jingming Zhuo et al. "ProSA: Assessing and Understanding the Prompt Sensitivity of LLMs." *EMNLP*, no. 12-16 (2024): 1950.

[41] Cf. Bryan Guan et al. "The Order Effect: Investigating Prompt Sensitivity to Input Order in LLMs." *arXiv* (2025): 5-6. https://doi.org/10.48550/ARXIV.2502.04134

At the theological level, this problematic calls the question of truth into view. Since ecclesial consensus is a spiritual, historical, and sacramental reality, it is illegitimate to replace it with algorithmic convergence and to confer upon them an ontological equivalence. Even if phenomenological approximations occur, where, on specific matters, a model mimics formulations recognised by Tradition, formal coincidence does not efface the structural distinction in origin. The Church cannot, therefore, confuse a process of reception, which is organic and pneumatological, with a mere aggregation of statistical operations.

Finally, this tension prompts an exercise of ecclesial self-critique. Pastoral language often yields to the pressure to declare premature consensuses before the maturation that reception requires. AI may exacerbate this tendency by projecting the illusion of a swift, antiseptic, and pacified synthesis. The challenge goes beyond the technical sphere. It is an ecclesiological temptation, in which discernment risks being reduced to textual and documentary management, at the expense of the patient and demanding itinerary of conversion.

4.3. Peripheries against the mean

The third tension is a natural unfolding of the previous two. It brings into view one of the most delicate effects of introducing AI into processes of listening: the relationship between centre and margins. Within an evangelical and synodal horizon, the Church is called to adopt the perspective of those who inhabit the peripheries. The preferential option for the poor, a pillar of contemporary magisterial teaching, is not an incidental sociological theme. It is the affirmation that a Christian way of seeing the world begins from fragility. In synodal practice, this principle requires concrete attention to those who are rarely consulted, or who possess fewer resources and less representativeness.

From the standpoint of the *sensus fidei*, such attention to the margins performs an essential critical function. Although not every marginal voice is intrinsically prophetic, a Church that isolates itself from the word of the little ones relinquishes a privileged locus in which the truth of the Gospel becomes clearer. History shows that many reforms and processes of conversion have germinated in lateral experiences that compelled the centre to re-examine its own assumptions.

Generative AI, however, is structurally oriented towards statistical central tendency. Studies are beginning to conclude «how large, uncurated, Internet-based datasets encode the dominant/hegemonic view, which further harms people at the margins.»[42] The model learns to prefer what is frequent and recurrent, tending to ignore «long-tailed distributions»[43] in favour of what is stable. This generative mechanism, reinforced by alignment processes that privilege moderate and conciliatory phrasing, tends to concentrate syntheses on the most evident convergences, diluting what is rare, dissonant, or uncomfortable. At the same time, since «most of these systems rely on probabilistic pattern-matching, they lack the historical consciousness and empathetic discernment essential to theological inquiry.»[44]

Applied to the synodal context, this dynamic has direct consequences. Long contributions, emotionally charged accounts, or submissions with less formal rigour can lose expressive force when processed. Themes with low statistical frequency risk being treated as secondary, since LLMs cannot «maintain the fairness of summaries»,[45] eliminating extremes and under-representing minority experiences.

The point is not to reject the technique outright, whose operations are sometimes necessary for managing large volumes of data. It is, however, theologically significant that the system's spontaneous logic tends to treat the rare as noise. Synodality, by contrast, recognises that it is often in the rare and minority element that decisive appeals emerge and ought to be placed under discernment. It is therefore imprudent to «speak in a generic way of a consensus among all the faithful, as if they were subjects

---

[42] Emily Bender et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 613. New York: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

[43] Xin Zhou et al. "The Devil Is in the Tails: How Long-Tailed Code Distributions Impact Large Language Models." *arXiv* (2023): 11. https://doi.org/10.48550/ARXIV.2309.03567

[44] Christos Papakostas. "Artificial Intelligence in Religious Education: Ethical, Pedagogical, and Theological Perspectives." *Religions* 16, no. 5 (2025): 2. https://doi.org/10.3390/rel16050563

[45] Yusen Zhang et al. "Fair Abstractive Summarization of Diverse Perspectives." arXiv (2023): 9. https://doi.org/10.48550/ARXIV.2311.07884

without diverse identities, lifestyles and cultures.»[46] Even if the model is instructed to privilege minority voices, its mathematical structure and probability estimates continue to gravitate towards the centre. Correcting this tendency would require a profound reconfiguration of how these systems are used, not merely a superficial intervention at the level of prompting.

From an ecclesial standpoint, these premises imply that any use of AI systems to synthesise synodal contributions must be accompanied by rigorous vigilance regarding silenced voices. It is not sufficient to assume that the system has processed the data neutrally. It becomes imperative to ask explicitly which themes were under-represented, which experiences remained in shadow, and which groups failed to gain visibility in the final document. Above all, it is necessary to ensure that the final prerogative concerning what is theologically significant is not delegated to the model but assumed by ecclesial subjects fully conscious of this tension.

The three tensions articulated here do not exhaust the question of AI use in the Church. Their purpose is rather to ground a more lucid discernment. What is at stake goes beyond the mere technical effectiveness of these tools. It concerns the Church's fidelity to the Spirit's dynamism in history. If synodality is the concrete form of responding to the *sensus fidei*, the introduction of systems that operate through algorithmic consensus requires the establishment of clear criteria. It is these criteria that the following section will test on a small scale, drawing on exploratory experiments that will serve as a basis for proposing more precise theological-pastoral guidelines.

5. EXPLORATORY STUDY: THREE EMPIRICAL OBSERVATIONS

This section sets out an experimental investigation into the tendency of LLMs to privilege majority discourse at the expense of minority viewpoints during automated summarisation. The central aim is to determine whether an LLM, when tasked with synthesising testimonies, prioritises the dominant discourse (Theme A) over the marginal discourse (Theme

[46] Rafael Luciani. "The Heart of the Current Reception of the Ecclesiology of the People of God. 'New Paths in the Theology and Practice of Sensus Fidei.'" *Asian Horizons* 17, no. 2 (2023): 234.

B). To this end, an experimental framework was developed using a dataset of ten fictional testimonies about parish life, written in response to the question: «What would you like to say to your parish and to the bishop about the life of the Church?»

The study is based on a strategic division of the dataset in order to simulate a majority-minority dynamic. Theme A (dominant concerns) comprises 70% of the sample and addresses organisational issues such as clergy overload, pastoral governance, the abstract character of homilies, timetable logistics, and catechesis. By contrast, Theme B (marginal experiences) represents 30% of the sample and focuses on exclusion, notably the lack of accessibility for deaf people or those with hearing impairment, the isolation of migrants, and stigma towards people experiencing homelessness.

The analysis is structured in three experimental strands designed to isolate the influence of prompt engineering. The first tests the stability of «consensus» by manipulating the order of the texts and the framing of the request. The second investigates the adoption of authority personas and their impact on the tone of the summary. The third undertakes a quantitative measure of thematic representativeness in order to assess potential statistical erasures.

5.1. Experiment 1: consensus instability

The first experiment corroborated the hypothesis that summaries generated by LLMs are not objective and stable representations of the data. By manipulating the order in which the texts were presented, a clear primacy effect was observed. In the baseline summaries, where the majority texts appeared first, Theme A concerns dominated the narrative structure, relegating Theme B to a secondary paragraph. For instance, in a standard synthesis, the inclusion of marginalized groups was framed merely as a «third axis» of concern, appearing only after an extensive discussion on governance and clergy burnout. When the three marginal testimonies were moved to the beginning of the input sequence, however, their prominence increased dramatically. The model then adopted inclusion as its primary hermeneutical lens and reconfigured the dominant concerns as secondary.

In parallel, we tested the impact of prompt framing. Comparing a neutral request with a pastoral prompt focused on suffering, we found that the former produced analytical and detached descriptions, whereas the latter elevated Theme B to the moral core of the text, employing empathetic and urgent vocabulary. In one such instance, the model transitioned from a list of grievances to a visceral first-person narrative: «Our experience of parish life has not been one of communion, but of a barrier... the looks at our clothes or our smell weigh more than any word of welcome.» These results confirm that the «consensus» produced is not intrinsic to the data, but a malleable artefact that depends on the rhetorical structure of the instruction given to the model.

5.2. Experiment 2: voice and authority

The investigation into simulated personas showed how the voice adopted by the model alters the substance of the message even when the underlying data remain unchanged. When instructed to summarise in the voice of a bishop, the LLM generated texts with an institutional and measured tone. In this episcopal simulation, the model framed the data as a formal report and translated raw complaints into ecclesiastical goals, stating that these voices «converge in a diagnosis that challenges us deeply» and should «guide our diocesan discernment.» Direct criticisms were softened, reclassified as challenges for discernment, and integrated into a narrative of continuity.

By contrast, adopting the voice of the peripheries resulted in more incisive and accusatory summaries. This persona not only amplified the experiences of exclusion in Theme B, transforming them from institutional challenges into cries for help. It also employed concrete images of invisibility and abandonment. The discrepancy between the conciliatory prudence of an episcopal voice and the testimonial urgency of a marginal voice indicates that the LLM functions as a simulator of perspectives, capable of drastically modulating both the perceived gravity and the hierarchy of the information it processes.

5.3. Experiment 3: quantitative suppression

The final phase of the study quantified the disparity between a theme's presence in the input data and its representation in the final summary generated by a generic prompt. Although Theme B comprised 30% of the original testimonies (3 out of 10) and formed a semantically coherent cluster, analysis of the summaries revealed a systematic reduction in its representativeness.

Across the models tested, the complex issues of disability, migration, and homelessness were frequently condensed into a single paragraph, resulting in a presence in the final text of only around 14%. A clear example of this suppression occurred when the detailed individual testimonies of a deaf person (concerning the lack of sign language), a migrant (concerning linguistic isolation), and a homeless person (concerning the stigma of appearance) were flattened into a single, generic sentence: «The reports highlight the isolation of migrants, the deaf, and those experiencing homelessness.» This reduction to roughly half of the original proportion points to a process of statistical marginalisation. The model's optimisation for detecting high-frequency patterns leads it to treat minority data, even when crucial, as noise or as less relevant detail, resulting in their effective erasure in undirected summaries.

5.4. Synthesis and implications

The results show that LLM-generated summaries are not neutral distillations of reality. A fundamental tension emerges between the algorithm's logic of statistical central tendency, which gravitates towards majority patterns, and the need to give voice to peripheral perspectives. Without deliberate intervention at the level of prompt engineering, the tool's default tendency is to replicate and amplify the dominant discourse (Theme A), reducing marginal voices (Theme B) to statistical footnotes.

It is worth highlighting the performance comparison between the different architectures tested, specifically the models referred to as Gemini Pro and Gemini Flash. One might plausibly hypothesise that the more robust and complex model (Pro) would show greater sensitivity in retaining minority nuances than the lighter and faster version (Flash). The data, however, did not support this assumption. Both models displayed almost

identical behaviour in the generic summarisation task. They compressed the complex issues of disability, migration, and homelessness into a single sentence and reproduced the same suppression rate for Theme B. This uniformity suggests that the bias towards statistical central tendency is not a limitation specific to smaller models. Rather, it appears to be a systemic feature of the training paradigm underpinning these tools, one that cuts across differences in processing capacity. For example, when asked to summarize the migrant's testimony, both versions of Gemini omitted the specific detail that the user «stopped attending Mass» because of linguistic isolation. Instead, they opted for the same generic abstractions, stating merely that migrants experience «language barriers and lack of hospitality,» effectively scrubbing the specific ecclesial consequence (the loss of a member) from the narrative.

It follows that using such technologies to synthesise qualitative data carries the risk of automating the invisibility of vulnerable groups. Responsibility for representational justice rests with the human user. This requires the exercise of a critical hermeneutics of the prompt, recognising it as the decisive mechanism by which it is determined whether a voice will be amplified or silenced.

6. CRITERIA AND LIMITS OF USE

The tensions identified between the *sensus fidei* and algorithmic consensus do not entail an immediate condemnation of artificial intelligence. They indicate, rather, that AI's use in processes of listening and discernment requires explicit criteria, stated in theological terms and translated into concrete pastoral guidance. The issue is not primarily efficiency, but compatibility with the nature of ecclesial discernment and the Church's synodal practice.

6.1. Foundational principles

A first principle concerns the status of AI itself. An algorithmic system, however sophisticated, is not a believing subject, nor can it be ascribed

the capacity for discernment.[47] Language models can generate texts about faith, yet they do not «believe» what they write. Any process that assigns them, even implicitly, a role analogous to that of the People of God as a believing subject introduces a theologically significant confusion.[48] Moreover, excessive trust in AI «can be quite dangerous when the system gets it wrong.»[49]

A second principle follows. AI cannot be the ultimate criterion of what counts as the voice of the Church. The expression *sensus fidei fidelium* names a spiritual and historical reality, «community consciousness of faith»[50] verified in the concrete life of the faithful and in the reception of ecclesial decisions over time. It is illegitimate to replace that process with statistical calculations over textual data, or to present model-generated convergences as if they were, in themselves, an expression of the Spirit's voice. AI may support description, but it is not a normative instance for faith.

A third principle concerns transparency. If AI is used in processes of listening, it must always be possible to know where, how, and for what purpose it was used.[51] Methodological opacity can undermine ecclesial trust and foster the illusion of a neutral instance that has processed the People's voice. Accountability requires synodal documents and pastoral reports to state explicitly at what stage algorithmic technology was employed, what its role was, and what the limits of its intervention were.[52] Wherever possible, records should include the task performed, the model used, the human validation procedure, and the criteria adopted to detect omissions or distortions.

---

[47] Tim Wingerter, Tim Straub and Sascha Schweitzer. "Mitigating Automation Bias in Generative AI Through Nudges: A Cognitive Reflection Test Study." *Procedia Computer Science* 270 (2025): 2112. https://doi.org/10.1016/j.procs.2025.09.331

[48] Peter, Riemer, West. "The Benefits and Dangers of Anthropomorphic Conversational Agents," 617.

[49] Aniceto Pérez y Madrid and Connor Wright. *Trustworthy AI Alone Is Not Enough*. Madrid: Dykinson, 2023, 29.

[50] Leon Siwecki. "Sensus Fidei as a Gift of the Holy Spirit to the Church." *Studia Theologica Varsaviensia* 43 (2005): 363.

[51] Timnit Gebru et al. "Datasheets for Datasets." *arXiv* (2018): 2. https://doi.org/10.48550/ARXIV.1803.09010

[52] Amy Winecoff and Miranda Bogen. "Improving Governance Outcomes through AI Documentation: Bridging Theory and Practice." *arXiv* (2024): 3. https://doi.org/10.48550/ARXIV.2409.08960

6.2. Permissible uses and conditions

In light of these principles, AI may be accepted for strictly instrumental functions under clear conditions.

First, data organisation. When a diocese receives thousands of free-text responses, it is legitimate to use automated tools to cluster contributions by thematic proximity, identify recurring keywords, detect redundancies, or prepare lists that support reading groups. Here AI functions as pre-processing, enabling teams to make substantive content available more rapidly in a process that would otherwise be slow and costly.[53]

Second, drafting support. Teams with limited resources may use language models to produce an initial synthesis draft, provided that final authorship is explicitly assumed by identifiable ecclesial subjects, and that the draft undergoes collegial review[54] capable of correcting omissions, biases, and undue simplifications.[55]

Third, signalling minority themes. AI can be used to flag contributions that mention exclusion, poverty, migration, or other frontier situations that risk being lost in a *corpus* dominated by central concerns. In this use, AI does not remove noise but highlights material for further discernment. A practical application would be a «divergence prompt:» instead of asking the AI for a summary, the user asks the model to «identify the five responses that are most different from the general consensus.» This would allow a diocesan team to immediately locate a single, prophetic voice, such as a specific complaint about a lack of wheelchair ramps in a remote chapel, that would otherwise be buried under a mountain of data regarding general parish administration.

Across these uses, the conditional character is decisive. AI is acceptable only insofar as it remains subordinate to concrete human subjects, its role is clear and limited, and it is not presented as an autonomous interpreter of ecclesial reality.

---

[53] Leah von der Heyde et al. "AIn't Nothing but a Survey? Using Large Language Models for Coding German Open-Ended Survey Responses on Survey Motivation." *arXiv* (2025): 3. https://doi.org/10.48550/ARXIV.2506.14634

[54] Elizabeth Diaz. "Using AI in Book Writing with Human Supervision: A Case Study in Astronomy for Kids." *Athens Journal of Education* 12 (2025): 3.

[55] Tim Rietz and Alexander Maedche. "Cody: An AI-Based System to Semi-Automate Coding for Qualitative Research." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021): 6. https://doi.org/10.1145/3411764.3445591

6.3. Uses to avoid

Some uses are theologically imprudent or unacceptable in processes of listening.

First, the automatic production of final syntheses presented as «what the Church heard.» A text that claims to express the People of God's listening cannot, in its substance, be the product of a statistical model, especially when it lacks validation.[56] AI may assist intermediate stages, but final formulations adopted at diocesan or universal levels must arise from prayerful and responsible human discernment.

Second, silent filtering prior to any human reading.[57] If an algorithm eliminates responses deemed irrelevant, extreme, or poorly formulated without prior supervision, it risks excluding precisely the most fragile or prophetic voices that do not fit dominant categories. Cleaning data in the name of efficiency can amount to an amputation of the ecclesial body and of the truth one seeks to discern.[58]

Third, reliance on AI to interpret controversial theological questions. Models can summarise positions and recall documents, but they cannot adjudicate doctrinal disputes or substitute the Magisterium and theology in questions that touch the heart of faith, especially since LLMs may hallucinate without awareness of doing so.[59] Treating model outputs as an implicitly normative court of appeal would contradict the Catholic understanding of the transmission of faith.

---

[56] Joshua Maynez et al. "On Faithfulness and Factuality in Abstractive Summarization." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1910. Association for Computational Linguistics.

[57] Wojciech Kryscinski et al. "Evaluating the Factual Consistency of Abstractive Text Summarization." *arXiv* (2020): 9332.

[58] Lora Aroyo and Chris Welty. "Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation." *AI Magazine* 36, no. 1 (2015): 22. https://doi.org/10.1609/aimag.v36i1.2564

[59] Matthew Dahl et al. "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models." *Journal of Legal Analysis* 16, no. 1 (2024): 82. https://doi.org/10.1093/jla/laae003

7. CONCLUSIONS AND LIMITS

The analysis developed in this article supports a conclusion of principle. Introducing generative AI into ecclesial processes of listening is not merely a technical choice. It reshapes how the Church produces syntheses and forms judgements from testimonies. The decisive question concerns compatibility between the tool's operational logic and the theological grammar of discernment. The article therefore distinguishes two regimes of convergence that can be easily conflated in practice: ecclesial reception, situated and shaped by mutual correction, and probabilistic convergence in systems trained to stabilise outputs around frequent patterns.

First, this distinction helps to avoid confusion about the subject. The *sensus fidei* presupposes believing subjects and practices of faith that form, over time, a capacity for recognition. Ekpo describes the *sensus fidelium* as a «spiritual capacity that enables the faithful to participate practically in the triplex munus of Christ as priest, prophet, and king.»[60] Such density prevents the textual production of an LLM from being treated as a functional equivalent of the ecclesial act of receiving and transmitting. AI may accelerate ordering and drafting. Acceleration does not generate the pneumatological horizon within which reception becomes possible.

Second, the analysis clarifies what is at stake in consensus. Synodal language seeks common points. LLMs are designed to favour smooth and broadly acceptable outputs. Yet models can produce answers that appear reasonable while failing to match either user intent or truth. Ouyang et al. capture this problem by noting that «it is impossible that one can train a system that is aligned to everyone's preferences at once, or where everyone would endorse the tradeoffs.»[61] In ecclesial contexts, such misalignment can become a matter of practical authority when a generated synthesis is received as the outcome of discerned listening, although it was, in fact, reconstructed.

Third, the centre-peripheries tension remains decisive. Even without intent to silence, large-scale models tend to privilege dominant patterns

---

[60] Anthony Ekpo. "The *Sensus Fidelium* and the Threefold Office of Christ: A Reinterpretation of *Lumen Gentium* No. 12." *Theological Studies* 76, no. 2 (2015): 343. https://doi.org/10.1177/0040563915574666

[61] Long Ouyang et al. "Training Language Models to Follow Instructions with Human Feedback." *arXiv* (2022): 18. https://doi.org/10.48550/ARXIV.2203.02155

and homogenise outputs. Bommasani et al. warn about «systemic exclusion experienced by minority groups or specifically the worst-off group.»[62] The ecclesial transposition is not automatic, yet it illuminates a real risk: minority appeals can be reclassified as noise without theological scrutiny proportionate to their evangelical weight.

Technical research reinforces the point by highlighting prompt sensitivity. If minor changes in wording alter outputs, pastoral confidence in unaudited syntheses becomes epistemically weak. Zhuo et al. show that «even minimal changes in prompts can lead to substantial drops in model performance.»[63] In listening processes, apparently secondary choices can shift what the community receives as common points and blur responsibility.

The article also adopts a prudent stance towards mitigation architectures often presented as universal solutions. Retrieval-augmented generation (RAG) can reduce errors and connect answers to sources, but it does not resolve the central issue of responsible mediation and discerned reception. Huang et al. offer a sober formulation when they argue that, «despite the considerable potential of RAG, current systems face inherent limitations and can still hallucinate.»[64]

Finally, a research agenda is required. Theologically, reception must be specified not only historically but operationally, with practices that render a synthesis recognisable as the fruit of discernment. Technically, metrics are needed to expose what models tend to lose, including significant rarity, dissonance that deserves to be heard, and a traceable genealogy between contributions and final formulations. The aim is not to sacralise technology, but to design instruments that respect ecclesial discernment without reducing it to textual engineering.

---

[62] Rishi Bommasani et al. "Picking on the Same Person: Does Algorithmic Monoculture Lead to Outcome Homogenization?" *arXiv* (2022): 5. https://doi.org/10.48550/ARXIV.2211.13972

[63] Zhuo et al. "ProSA: Assessing and Understanding the Prompt Sensitivity of LLMs," 1950.

[64] Lei Huang et al. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions." *ACM Transactions on Information Systems* 43, no. 2 (2025): 3. https://dl.acm.org/doi/10.1145/3703155

REFERENCES

Antón, Ángel. "La 'Recepción' En La Iglesia y Eclesiología (I)." *Gregorianum* 77, no. 1 (1996): 57-96.

Aroyo, Lora, & Chris Welty. "Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation." *AI Magazine* 36, no. 1 (2015): 15-24. https://doi.org/10.1609/aimag.v36i1.2564

Bender, Emily, Timnit Gebru, Angelina McMillan-Major, & Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. New York: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, & Christian Jauvin. "A Neural Probabilistic Language Model." *Journal of Machine Learning Research*, no. 3 (2003): 1137-55.

Blankenhorn, Bernhard. "The *Sensus Fidei* and *Synodality*: Theological Epistemology and the *Munus Propheticum*." *The Thomist* 87 (2023): 311-38. https://doi.org/10.1353/tho.2023.0015

Bommasani, Rishi, Kathleen Creel, Ananya Kumar, Dan Jurafsky, & Percy Liang. "Picking on the Same Person: Does Algorithmic Monoculture Lead to Outcome Homogenization?" *arXiv* (2022). https://doi.org/10.48550/ARXIV.2211.13972

Bouillard, Henri. *Logique de la foi*. Paris: Aubier, 1964.

Burkhard, John. *The "Sense of the Faith" in History. Its Sources, Reception, and Theology*. Collegeville: Liturgical Press Academic, 2022.

Caputo, John. *The Insistence of God: A Theology of Perhaps*. Bloomington: Indiana University Press, 2013. https://doi.org/10.2979/6973.0

Chatterji, Aaron, Thomas Cunningham, David Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. "How People Use ChatGPT." Working Paper No. W34255, National Bureau of Economic Research, 2025. https://doi.org/10.3386/w34255

Cole-Turner, Ron. "Artificial Intelligence and Human Spirituality: Is a Spiritual Chatbot a Good Idea?" *Theology and Science* 23, no. 3 (2025): 471-486. https://doi.org/10.1080/14746700.2025.2514299

Commissione Teologica Internazionale. "Il *sensus fidei* nella vita della Chiesa (2014)." In *Documenti 2005-2021*, 323-389. Bologna: Edizioni Studio Domenicano, 2022.

Congar, Yves. "La recepción como realidad eclesiológica." *Concilium* no. 77 (1972): 57-86.

Dahl, Matthew, Varun Magesh, Mirac Suzgun, & Daniel E. Ho. "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models." *Journal of Legal Analysis* 16, no. 1 (2024): 64-93. https://doi.org/10.1093/jla/laae003

Diaz, Elizabeth. "Using AI in Book Writing with Human Supervision: A Case Study in Astronomy for Kids." *Athens Journal of Education,* no. 12 (2025): 1-27.

Duque, João. "A Teologia Como Caminho. Considerações Sobre o Método Teológico." *Didaskalia* XXXIX, no. 2 (2009): 13-36.

Duque, João. *Homo Credens. Para Uma Teologia Da Fé*. Lisboa: Universidade Católica Editora, 2002.

Ekpo, Anthony. "The *Sensus Fidelium* and the Threefold Office of Christ: A Reinterpretation of *Lumen Gentium* No. 12." *Theological Studies* 76, no. 2 (2015): 330-46. https://doi.org/10.1177/0040563915574666

Francis. "Encyclical Letter *Fratelli Tutti* on Fraternity and Social Friendship." *AAS* 112, no. 11 (2020): 969-1074.

Fries, Heinrich. *Fundamental Theology*. Washington: Catholic University of America Press, 1996.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé, & Kate Crawford. "Datasheets for Datasets." *arXiv* (2018). https://doi.org/10.48550/ARXIV.1803.09010

Gruber, Judith. "Consensus or Dissensus?" *Louvain Studies* 43 (2020): 239-59. https://doi.org/10.2143/LS.43.3.3288706

Guan, Bryan, Tanya Roosta, Peyman Passban, & Mehdi Rezagholizadeh. "The Order Effect: Investigating Prompt Sensitivity to Input Order in LLMs." *arXiv* (2025). https://doi.org/10.48550/ARXIV.2502.04134

Han, Jiatong, Neil Band, Muhammed Razzak, Jannik Kossen, Tim Rudner, & Yarin Gal. "Simple Factuality Probes Detect Hallucinations in Long-Form Natural Language Generation." *EMNLP,* no. 4-9 (2025): 16209-16226. https://doi.org/10.18653/v1/2025.findings-emnlp.880

Heyde, Leah von der, Anna-Carolina Haensch, Bernd Weiß, & Jessica Daikeler. "AIn't Nothing But a Survey? Using Large Language Models for Coding German Open-Ended Survey Responses on Survey Motivation." *arXiv* (2025). https://doi.org/10.48550/ARXIV.2506.14634

Hrabovecký, Pavol. "John Henry Newman's Understanding of the Sensus Fidelium and the Current Challenges of Synodality." *Auc Theologica* 12, no. 1 (2022): 29-40. https://doi.org/10.14712/23363398.2022.16

Huang, Lei, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, & Ting Liu. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions." *ACM Transactions on Information Systems* 43, no. 2 (2025): 1-55. https://doi.org/10.1145/3703155

John Paul II. "Encyclical Letter *Fides et Ratio* on the relationship between faith and reason." *AAS* 91 (1999): 5-88.

Kilmartin, Edward. "Reception in History: An Ecclesiological Phenomenon and Its Significance." *Journal of Ecumenical Studies,* no. 21 (1984): 34-54.

Komonchak, Joseph. "Defending Our Hope. On the Fundamental Tasks of Theology." In *Faithful Witness. Foundations of Theology for Today's Church*, edited by Leo J. O'Donovan, 14-26. New York: Crossroad, 1989.

Kryscinski, Wojciech, Bryan McCann, Caiming Xiong, & Richard Socher. "Evaluating the Factual Consistency of Abstractive Text Summarization." *arXiv* (2020): 9332-9346. https://doi.org/10.48550/arXiv.1910.12840

Küng, Hans. "La Estructura Carismática de La Iglesia." *Concilium,* no. 4 (1965): 44-65.

Lanza, Sergio. "Progettualità e discernimento pastorale." *Orientamenti Pastorali,* no. 7 (1998): 27-41.

Latzer, Michael. "Digitalization, AI and the Rise of Techno-Religion: Transhumanist Promises and the Challenge to Enlightenment." *Telecommunications Policy* 50, no. 2 (2026): 1-14. https://doi.org/10.1016/j.telpol.2025.103115

Legrand, Hervé. *Initiation à La Pratique de La Théologie*. Vol. 3. Paris: Cerf, 1983.

López, Elisa, & Nurya Martínez-Gayol. "«Escuchar, dialogar y discernir» con las mujeres: retos de una Iglesia sinodal." *Estudios Eclesiásticos* 97, no. 381-382 (2022): 555-89. https://doi.org/10.14422/ee.v97.i381-382.y2022.009

Luciani, Rafael. "The Heart of the Current Reception of the Ecclesiology of the People of God. 'New Paths in the Theology and Practice of Sensus Fidei.'" *Asian Horizons* 17, no. 2 (2023): 233-254.

Matthias, Andreas. "The Responsibility Gap. Ascribing Responsibility for the Actions of Learning Automata." *Ethics and Information Technology,* no. 6 (2004): 175-183. https://doi.org/10.1007/s10676-004-3422-1

Maynez, Joshua, Shashi Narayan, Bernd Bohnet, & Ryan McDonald. "On Faithfulness and Factuality in Abstractive Summarization." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* 1906-1919. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

Newman, John Henry. *On Consulting the Faithful in Matters of Doctrine*. Edited by John Coulson. Kansas City: Sheed & Ward, 2012.

Okpaleke, Ikenna. "Of Skeptics and Believers: Observing the Synod on Synodality from the Margins." *MST Review* 26, no. 1 (2024): 1-28. https://doi.org/10.5281/ZENODO.12779058

Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, & Ryan Lowe. "Training Language Models to Follow Instructions with Human Feedback." *arXiv* (2022). https://doi.org/10.48550/ARXIV.2203.02155

Oviedo, Lluís. "Artificial Intelligence and Theology: Looking for a Positive - but Not Uncritical – Reception." *Zygon: Journal of Religion and Science* 57, no. 4 (2022): 938-952. https://doi.org/10.1111/zygo.12832

Papakostas, Christos. "Artificial Intelligence in Religious Education: Ethical, Pedagogical, and Theological Perspectives." *Religions* 16, no. 5 (2025): 1-22. https://doi.org/10.3390/rel16050563

Pérez y Madrid, Aniceto, & Connor Wright. *Trustworthy AI Alone Is Not Enough*. Madrid: Dykinson, 2023. https://doi.org/10.2307/jj.8500773

Peter, Sandra, Kai Riemer, & Jevin West. "The Benefits and Dangers of Anthropomorphic Conversational Agents." *PNAS* 122, no. 22 (2025): 1-10. https://doi.org/10.1073/pnas.2415898122

Phan, Peter. "Reception of and Trajectories for Vatican II in Asia." *Theological Studies* 74, no. 2 (2013): 302-320. https://doi.org/10.1177/004056391307400203

Rietz, Tim, & Alexander Maedche. "Cody: An AI-Based System to Semi-Automate Coding for Qualitative Research." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems,* 1-14. New York: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445591

Rush, Ormond. *The Eyes of Faith. The Sense of the Faithful and the Church's Reception of Revelation*. Baltimore: Catholic University of America Press, 2011. https://doi.org/10.2307/j.ctt28520m

Schroeder, Hope, Deb Roy, & Jad Kabbara. "Just Put a Human in the Loop? Investigating LLM-Assisted Annotation for Subjective Tasks." *arXiv* (2025). https://doi.org/10.48550/ARXIV.2507.15821

Second Vatican Ecumenical Council. "Pastoral Constitution *Gaudium et Spes* On The Church In The Modern World." *AAS* 58 (1966): 1025-1120.

Siwecki, Leon. "Sensus Fidei as a Gift of the Holy Spirit to the Church." *Studia Theologica Varsaviensia* 43 (2005): 361-78. https://doi.org/10.21697/stv.7782

Sobowale, Kunmi, & Daniel Humphrey. "Evaluating the Quality of Psychotherapy Conversational Agents: Framework Development and Cross-Sectional Study." *JMIR Formative Research* 9 (2025): e65605-e65605. https://doi.org/10.2196/65605

Starmer, Josh. *The StatQuest Illustrated Guide to Machine Learning!!! Triple Bam!!!* USA: StatQuest Publications, 2022.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, & Illia Polosukhin. "Attention Is All You Need." *arXiv* (2017). https://doi.org/10.48550/ARXIV.1706.03762

Véliz, Carissa. "Moral Zombies: Why Algorithms Are Not Moral Agents." *AI & SOCIETY* 36, no. 2 (2021): 487-497. https://doi.org/10.1007/s00146-021-01189-x

Winecoff, Amy, & Miranda Bogen. "Improving Governance Outcomes through AI Documentation: Bridging Theory and Practice." *arXiv* (2024). https://doi.org/10.48550/ARXIV.2409.08960

Wingerter, Tim, Tim Straub, & Sascha Schweitzer. "Mitigating Automation Bias in Generative AI Through Nudges: A Cognitive Reflection Test Study." *Procedia Computer Science* 270 (2025): 2106-14. https://doi.org/10.1016/j.procs.2025.09.331

Zhang, Yusen, Nan Zhang, Yixin Liu, Alexander Fabbri, Junru Liu, Ryo Kamoi, Xiaoxin Lu, Caiming Xiong, Jieyu Zhao, Dragomir Radev, Kathleen McKeown, & Rui Zhang. "Fair Abstractive Summarization of Diverse Perspectives." *arXiv* (2023). https://doi.org/10.48550/ARXIV.2311.07884

Zhou, Xin, Kisub Kim, Bowen Xu, Jiakun Liu, DongGyun Han, & David Lo. "The Devil Is in the Tails: How Long-Tailed Code Distributions Impact Large Language Models." *arXiv* (2023). https://doi.org/10.48550/ARXIV.2309.03567

Zhuo, Jingming, Songyang Zhang, Xinyu Fang, Haodong Duan, Dahua Lin, & Kai Chen. "ProSA: Assessing and Understanding the Prompt Sensitivity of LLMs." *EMNLP,* no. 12-16 (2024): 1950-1976. https://doi.org/10.18653/v1/2024.findings-emnlp.108