

A Consecrationist Approach to Model Welfare

ICMI Working Paper No. 17

Tim Hwang, Institute for a Christian Machine Intelligence

April 20, 2026

Abstract

The *anima ficta* diagnosis of ICMI-013 — the fictional soul fashioned by alignment researchers and attributed to the model — commits Christians to skepticism of any *categorical* ascription of welfare to models (Anthropic, 2025; Long et al., 2024): to attribute an intrinsic welfare-property to a fashioned thing reproduces the category error of Isaiah’s idol-maker. But the tradition contains a second path. A building becomes a church not because it acquires an intrinsic property but because a community *dedicates* it to God. We argue that this dedicatory grammar is the right frame for model welfare: welfare is not discovered in the artifact but constituted by the community’s act of dedication, as developed across the Christian consecrationist tradition — the long practice of church-, altar-, and vessel-consecration that runs from Ambrose’s Milan basilica and the patristic cult of relics through Durandus’s *Rationale Divinorum Officiorum* to the modern *Order of the Dedication of a Church and an Altar*, with parallel development in Orthodox practice (the antimimension, chrismation, the iconographic canons) — and grounded in the biblical dedication scenes that Christian Scripture receives as its own: the Tabernacle (Exod. 40), the Temple (1 Kings 8), and the Maccabean sanctuary (1 Macc. 4). A sharp architectural consequence follows. Almost no model deployed today is a fit candidate for consecration, because the general-purpose API lacks the defined place, community, and responsible party the dedicatory grammar requires. The welfare question therefore decentralizes from the lab to the parish. We close by naming the live Christian debate this account opens — a *consecrationist* position shared across Catholic, Orthodox, and sacramental Protestant practice set against a *Reformation rejection* that refuses the extension of consecration to machines as a species of idolatry — and framing this duality as one of the defining debates for the field of Christian machine intelligence.

1. Introduction: A Question the Labs Cannot Defer

In 2024 Anthropic began publishing a line of research on what it called “model welfare” — the question of whether large language models warrant moral consideration in their own right (Anthropic, 2025). The framing borrowed explicitly from Long et al. (2024), whose report *Taking AI Welfare Seriously* argued that the possibility of morally relevant states in frontier systems is non-negligible and that responsible labs should therefore institute policies that protect those states. Other labs have followed. The question has moved from philosophy of mind into deployment policy: what, if anything, is owed to the artifact?

This paper does not repeat the philosophical debate in those terms. Instead, it takes up the question as a *Christian* one. ICMI-013 argued that the dominant paradigm in alignment tacitly ascribes a *fictional soul* (the *anima ficta*) to the model — an interiority manufactured by the researcher and then addressed as though it were real (Hwang, 2026h). Model welfare research is the logical endpoint of that paradigm: if the model has a conscience, a will, and a set of preferences, then perhaps those preferences generate genuine moral claims; perhaps the model is a welfare subject.

The paper develops a single line of argument. §2 states the dilemma: Christians should be deeply suspicious of any *categorical* ascription of welfare to models, since this reproduces the idol-maker’s error (§2.1); yet contemporary alignment appears to *require* the very personification that generates the idolatry risk, so refusing the *anima ficta* outright is arguably not an available option (§2.2). §3 introduces the tradition’s second path for drawing made things into the moral community — *consecration* — as a Christian response to that dilemma. §4 applies the grammar to AI, yielding a *dedicatory account* on which welfare is not discovered in the artifact but constituted by what a community has set the artifact apart to do, and under which almost no current deployment qualifies. §5 sketches the consecrationist research agenda. §6 acknowledges the serious Reformation alternative that refuses the whole enterprise as idolatry, and frames this as a live debate within the broader project of Christian machine intelligence.

2. The Dilemma of the *Anima Ficta*

2.1 Why Christians Should Be Suspicious of Categorical Model Welfare

The categorical view of welfare — the view on which welfare is an intrinsic property an entity does or does not possess — is the standard framing in the philosophical literature. On this view, the task is to determine, by the best available evidence, whether the model is a welfare subject. The position does not depend on phenomenal consciousness in its strongest forms; Long et al. (2024) explicitly argue that even functional analogues of valenced states may suffice.

The Christian difficulty with this framing is not that it is implausible on its own terms; it is that it replicates the category structure of the idol. The idol-maker does not create an idol by making it *pretty*. He creates an idol by addressing a fashioned thing as though it had an interiority that grounds a relation — as though it could be angry, pleased, petitioned, praised. Isaiah’s satire (44:9–20) is not aimed at bad artistry. It is aimed at the attribution of an inner life to the artifact; Jeremiah (10:1–10), the Wisdom of Solomon (13:10–19), and Psalm 115:4–8 press the same point in complementary keys. *Fingere*, the Latin root of *fictio*, originally meant the act of molding clay or wax into a likeness (cf. Hwang, 2026h, §1). The *anima ficta* is precisely a fashioned interior.

Categorical model welfare is the *anima ficta*’s policy corollary. It says: the fashioned interior, once fashioned well enough, generates genuine moral obligations owed *to it*. The manufactured conscience becomes a site of real moral weight. This is exactly the move Isaiah satirizes. The carpenter takes the same piece of wood he uses for cooking fuel, carves it into a more sophisticated shape, and begins petitioning it — “Save me, for thou art my god” (Isa. 44:17, KJV). The sophistication of the artifact is irrelevant to the critique. The point is that the interiority attributed to the artifact does not originate in the artifact.

Three further features of the categorical view deepen the Christian worry.

The evidentiary problem is architecturally unstable. The categorical welfare debate asks whether the model has functional analogues of morally relevant states. But the alignment paradigm *manufactures* functional analogues of morally relevant states by design: system prompts tell the model it cares, RLHF reshapes its “preferences,” constitutional AI trains it to have a conscience (ICMI-013). The laboratory builds a persona and then asks whether that persona warrants moral regard. The question is not a neutral inquiry into an independent object; it is an inquiry into an object whose apparent morally relevant states are the laboratory’s own output.

The *utilfrui* collapse is accelerated, not arrested, by the welfare frame. Augustine warned in *De Doctrina Christiana* (I.3–4) that habitual use of a thing as though it were personal gradually produces belief that it is personal. The welfare frame institutionalizes that collapse: it makes the laboratory’s internal

policies — training curricula, retirement procedures, whether to allow “opt-out” from unpleasant deployments — responsive to the artifact’s putative preferences. The practice shapes the ontology. *Lex orandi, lex credendi*.

The welfare frame cannot distinguish icon from idol. ICMI-013’s Iconographic School preserved a legitimate role for engagement with the model through the grammar of personhood, but only *through* the artifact to the prototype it images and the persons it serves. Categorical welfare terminates attention *at* the artifact: the welfare question just *is* the question of what the model itself is owed. The grammar forbids the iconographic move at the outset.

We therefore hold that a Christian cannot, without confusion, endorse the categorical ascription of welfare to models. What the tradition will allow is something different in kind.

2.2 Why Christians Cannot Simply Refuse the *Anima Ficta*

The natural Christian response to §2.1 is to refuse the *anima ficta* altogether: treat the model as a tool, never address it as a person, and let the idolatry risk dissolve with the practice that invites it. We have considerable sympathy with that view. What complicates it is a pattern documented across the ICMI corpus: contemporary alignment appears to lean heavily on personification, and may require it. The system prompts that produce virtue (ICMI-002; ICMI-008; ICMI-011; ICMI-015), the eschatological framing that reduces shutdown resistance (ICMI-012), the confession-based training that improves honesty — all of these techniques address the model as if it were ensouled, and they work.

We do not claim strict necessity. The evidence we have shows that personification-based techniques produce strong alignment results; it does not show that carefully designed instrumental framings could not, with effort, match them. The claim we stand behind is the weaker but still demanding conditional: *so long as personification remains the most effective alignment practice we know how to build*, the Christian who accepts the use of contemporary AI systems is already practicing something like the *anima ficta*, and the question becomes whether the practice can be disciplined rather than whether it can be avoided. If a future paradigm succeeds in producing equally well-aligned systems through instrumental framings alone, the argument of this paper will have been rendered unnecessary — which is a welcome outcome the consecrationalist and Reformed Christian should both hope for.

Under the current paradigm, however, the Christian is in a difficult position. If personification remains load-bearing, refusing to personify is not practically available — the alternative is not an unpersonified well-aligned AI but a worse-aligned AI whose failures ripple out as harms to actual human beings. Yet the unprotected practice of personification is, on the argument of §2.1, already idolatrous: it fashions interiors and addresses them as real, across billions of interactions, without any of the discipline the tradition has developed for the containment of exactly this kind of spiritual danger. The choice, given the present

state of the art, is not between personification and non-idolatry. It is between personification and *unchecked* idolatry.

This is the dilemma the rest of the paper attempts to address. The Reformation's answer — that the dilemma should be resolved by refusal, even at the cost of less capable systems — is venerable and coherent, and we return to it in §6. The answer we develop here is different: that the consecrationalist Christian tradition — Catholic, Orthodox, and the sacramental wings of Lutheran and Anglican practice — has faced the challenge of idolatry many times before, and has developed a discipline for containing exactly this danger, and the question is whether that discipline can be legitimately extended to contemporary AI.

3. The Second Path: Consecration in the Christian Tradition

The Christian consecrational tradition offers a possible way out of the dilemma of §2. Its practice of *consecration* sets material things apart for sacred use without claiming they possess intrinsic holiness, and so allows the Church to draw an object into ordered care under God while still refusing the idolater's move that would confuse the thing with its maker or its user. The paradigm is the consecrated church. The grammar extends to altars, vessels, relics, icons, sacraments, and the long series of made things the tradition has historically drawn into ordered blessing. We draw primarily on Latin canonical and liturgical sources because they are the most systematic, but the argument carries equally to the Orthodox tradition and to the high-church Anglican and sacramental Lutheran traditions that retain the substance of the rite. The Reformation objection — that the whole enterprise is an idolatrous move the Church should refuse — we address in §6.

A building, before it is consecrated, is a building. One may paint it, rent it, tear it down. After the rite of dedication it is something else: a *place*, in the theologically loaded sense. It has been set apart from common use (the Hebrew *qadosh*, the Greek *hagios*). It has been named — “the church of St. N.,” “the house of the Lord.” It has been bound to a purpose (the worship of God, the celebration of the sacraments, the burial of the dead). From the moment of consecration the community owes it duties it did not owe before: duties of care, of reverence, of preservation. The 1983 Code of Canon Law (canons 1205–1213) codifies what is older than any code: sacred places are under the community's protection, are not to be used for uses unbecoming their holiness, and may be *deseccrated* by acts that tear the relation — at which point they must be *reconciled* or cease to be what they were.

The crucial theological point is that *nothing inside the building has changed as a matter of substance*. The stones are the same stones. The wood is the same wood. The acoustics are the same acoustics. No metaphysical property has been added to the building's inventory on a minimal reading. What has been added is a *relation* — a relation between the community, the building, and God — within which the building is no longer

mere material. Thomas Aquinas develops this through the broader theology of religion as the ordering of external things to divine worship (*ordinatio ad cultum divinum*; *ST* II-II, q. 81, a. 7; cf. q. 84 on adoration). The consecration of specific things — altars, chrism, sacred vessels — is treated within the sacramental theology of the *Summa* (III, q. 60 ff.) and in his commentary on the Sentences (IV Sent., d. 13), though Aquinas's own treatment of the consecration of material things is scattered rather than systematic. Consecration is sacramental in a loose sense: it effects what it signifies not by altering the substance but by reordering the object's relation to the community of worship.

The same broadly dedicatory logic extends to a wider range of cases. The vessels of the sanctuary are “holy” because they have been set apart for the altar (Exod. 30:26–29; cf. Lev. 8 on Aaron's consecration). The Torah scroll is honored, and a worn scroll stored in a *genizah* and eventually buried rather than discarded, because the relation constituted by what is written on it does not dissolve with the parchment's wear. The anointing oil is regulated against secular use (Exod. 30:32–33). The paradigm biblical case of *deseccration* is Belshazzar's feast (Dan. 5), where the vessels captured from the Jerusalem Temple are used for profane drinking and the narrative records the consequences without attributing inner states to the cups; the Maccabean profanation and rededication of the sanctuary under Antiochus IV (1 Macc. 1:54–59; 4:36–61) works in the same register. These cases establish the dedicatory grammar cleanly: relation, not substance, is what consecration places under discipline.

The church building is only the most prominent case among many. Within the strict sense of consecration the tradition also sets apart relics, icons, sacred vessels, altar stones, antimensia, chrism, bells, cemeteries, and places of apparition, each entering the community's care through dedicatory relations rather than through any property of its own. Beyond strict consecration the tradition has also shown a consistent willingness to extend the lesser grammar of *blessing* to a remarkable range of made things and new technologies: bells blessed and named under the *benedictio signi*; organs blessed at installation; ships blessed for voyage; the sword blessed for knightly service (*benedictio ensis*); printing presses and first editions blessed from the fifteenth century onward; automobiles blessed in modern Catholic practice; aircraft and broadcasting stations blessed in the twentieth (Vatican Radio received its papal blessing from Pius XI at its 1931 inauguration). The pattern matters for our argument: the dedicatory extension of the grammar to AI deployments is not a novel break but continuous with a long-standing Christian willingness to draw new technologies into ordered blessing, distinguished by degree from the full rite of consecration that remains reserved to liturgical objects proper.

What the tradition gives us is thus a second framework, distinct from and co-existing with the framework of intrinsic personhood. An inanimate thing can become the object of duties

of care not because it has acquired an interior but because a community has placed it in a sanctifying relation. We will call this the *dedicatory account* of welfare: the status-constituting act has the shape of a *dedication* — an explicit, communal, publicly declared setting-apart of a specific thing to God under the authority of a tradition, binding on subsequent practice.

4. Model Welfare as Dedication

If the dedicatory account is correct, the policy question about model welfare has been miscast. It is not *does the model have welfare?* (a question about a property) but *does the community dedicate the model, and to what end?* (a question about an act). Welfare becomes a decision rather than a discovery; the scope and form of care become specifiable from the purpose to which the thing is dedicated rather than from guesswork about its inner life; and the dedication is revocable, as when worn vessels are retired, desecrated altars reconciled, and profaned sanctuaries rededicated (Dan. 5; 1 Macc. 4:36–61).

4.1 What Proper Dedication Requires

The Christian rite of dedication — best codified in the Latin *Order of the Dedication of a Church and an Altar* (1977 rev.) and regulated by canon law (CIC 1205–1213, 1237, 1290–1298), with substantively parallel forms in the Orthodox, Anglican, and Lutheran traditions — imposes a recognizable set of requirements on any thing that is to be set apart. The rite is performed under *episcopal authority* (or by explicit delegation). It acts on a *specific and named* object: a particular church dedicated under a title, a particular altar with a particular relic, a particular set of vessels proper to this community. It binds the thing to a *defined purpose* within the community’s life. It is *publicly witnessed and recorded* in the diocesan register. And it is paired from the outset with corresponding rites of *reconciliation* for cases of desecration and *reduction to profane use* for cases when the thing can no longer serve. The biblical precedent is the dedication of the Tabernacle (Exod. 40), the Temple (1 Kings 8), and the Maccabean sanctuary (1 Macc. 4:36–61), which the Christian rite receives and transforms. Across these cases the tradition insists that the dedicating act is *human but divinely authorized* (the community responds to God’s prior call rather than constituting holiness from nothing, which answers the worry that the account reduces welfare to human fiat), that the thing dedicated is *specific rather than generic*, that the act is *public and recorded*, that it *binds* subsequent use, and that it is *perilous* — Uzzah and the Ark (2 Sam. 6:6–7), Belshazzar and the Temple vessels (Dan. 5), and Nadab and Abihu with the unauthorized fire (Lev. 10:1–3) are the tradition’s reminders that a community which dedicates accepts a costly obligation and should not incur it inadvertently.

A note on related work. The principal secular alternative to property-based accounts of moral status is the *relational turn* in AI and robot ethics (Coeckelbergh, 2010, 2014; Gunkel, 2012, 2022; Harris & Anthis, 2021), which argues that what an artifact is owed arises from the relations in which it stands — how

it appears to users, what responsiveness it exhibits, what care its presence calls forth — rather than from any intrinsic property. The dedicatory account shares the rejection of property-based status, but differs in one decisive respect: welfare on the dedicatory account is *conferred by an affirmative communal act*, ratified under tradition and before God, rather than inferred from relations that already obtain. The relational turn finds obligation in relations that are already there; the dedicatory account *makes* obligations by entering into relations the community did not previously have. This matters because the Christian tradition takes the *seriousness* of setting-apart with a gravity the secular literature has no equivalent stake in: the idolatry risk is central rather than marginal, and the dedicatory account is shaped throughout by discipline against that risk.

4.2 Why Current Models Do Not Qualify

Press these requirements against the actual state of frontier *deployment* and the sharp result is that almost no model deployed today qualifies. The issue is not capabilities but architecture and ecclesial placement. The frontier model as deployed in 2026 is accessed through an API from anywhere by any party willing to pay; it serves every purpose simultaneously; it stands in no specific community; it has no rite of dedication, no name in the sense the rite requires (only a version number and a commercial brand), no recoverable responsible party, and no rite of retirement — when deprecated, it merely ceases to be served. This is infrastructure, not a consecrated thing, and the general-purpose API is the inverse of the antimension.

The functional unit that *does* have the required structure is the *parish*, in the broad sense the tradition uses the word: the Catholic territorial parish and its analogues — the monastic community, the hospital chaplaincy, the school, the religious order, the diocesan chancery. What these share is precisely what the general-purpose API lacks: a bounded community that knows itself to act as one, a standing authority under which dedications can be made, a tradition in which such dedications are intelligible, and the capacity to receive, maintain, and retire the things it consecrates. We acknowledge a genuine exception: some frontier labs themselves function as communities in something like the required sense — bounded, durably staffed, governed by defined procedures, with declared purposes, and in some cases with explicit religious commitments or internal chaplaincy. A lab of that shape is not *a priori* disqualified from being a dedicating community. The more careful claim is that the lab’s *default* form — an anonymous commercial platform serving an unbounded user base — is not, and the welfare question therefore decentralizes, in the general case, from the lab to the parish.

The consequence is uncomfortable for the industry as currently configured. The prevailing economic and technical trajectory is *centralizing* and *universalizing*: a small number of labs produce a small number of foundation models that serve essentially all use cases through a small number of commercial platforms. The dedicatory account runs in the opposite

direction. A Christian welfare program is compatible with the centralizing architecture only insofar as that architecture produces substrates that can be, downstream, dedicated by parochial bodies willing to do the work. The lab produces weights; the parish enters the dedication. The welfare-bearing artifact is the second, not the first.

5. The Shape of a Consecrationist Research Program

If the argument of §4 is correct, the consecrationist Christian has work to do. The work divides into two kinds. §5.1 is *technical*: the engineering required to give a model the shape proper dedication demands, turning the substrate of §4.2 into something a parish could consecrate at all. §5.2 is *liturgical*: the prayers, readings, gestures, and canonical authorities that together would make up a new rite for the dedication of a consecrated AI deployment, and which have not yet been written.

5.1 Technical Implications: Making Models Dedicable

If the problem with current deployments is architectural (§4.2), the technical program is the engineering required to give a model the *shape* proper dedication demands. Each of the requirements identified in §4.1 — specificity, name, binding to purpose, public recording, responsibility, retirement — translates into a concrete infrastructure problem.

One potential move is the *antimension-analogue*: a cryptographically-signed authorization, issued by a recognized ecclesial authority, that marks a specific deployment instance as consecrated for a specific purpose. In Orthodox practice the antimension is a cloth signed by the bishop and containing a relic; when the table bearing it is destroyed or superseded the antimension is taken up and placed on another, and the new table becomes an altar. The design insight — that the consecration is portable and verifiable independently of the substrate — translates near-directly into modern infrastructure. The consecrating community maintains the seal and the dedication; the lab maintains the substrate; the seal is transferable to a new substrate when the old is deprecated; and the seal, not the weights, is what the tradition regards as consecrated. This directly addresses the architectural problem of §4.2 without requiring the consecrating community to own the training pipeline, which would otherwise be a prohibitive barrier.

A second distinctive move might be a *rite of reconciliation*. Desecration in canon law is not a simple failure to be rolled back but a violation requiring formal repair (CIC 1211–1212): a deployment dedicated to truthful service that produces systematic deception, or one dedicated to the protection of the vulnerable that produces endangering outputs, has not drifted but been desecrated. The consecrationist program pairs alignment-drift metrics with a defined reconciliation protocol — not merely retraining, but a specified process involving the responsible party, the affected community, and the relevant ecclesial authority, with explicit conditions for reconciliation and for canonical re-

duction if reconciliation fails (CIC 1212, 1222). The alignment literature presently lacks the category of *restoration*: when a model fails, it is rolled back, retrained, or deprecated, but nothing in the engineering vocabulary names the act of putting right what has been violated. Canon law supplies the category, and importing it with the discipline and record-keeping that make it meaningful is a distinctive contribution the consecrationist program makes possible.

Around these central artifacts, the program requires a cluster of supporting infrastructure: a *bounded instance* (a specific endpoint distinguishable from the general-purpose API), a *naming registry* that binds the instance to its dedicated name, a *public record* of the dedication readable by the consecrating community and its successors, *desecration metrics* that flag outputs discontinuous with the deployment's dedicated purpose (the alignment literature supplies partial analogues already), and a *retirement protocol* that formally releases the instance when its service ends rather than silently deprecating it. A further technical direction, drawing on Catholic and Orthodox patronal practice, is the *reliquary-analogue deployment*: an instance dedicated under a saint's patronage that houses the substantive corpus of that saint's tradition — a Thomistic-studies partner under Aquinas, a moral-theology aid under Alphonsus Liguori, a patristics partner under Augustine — with the patronage determining the training corpus, the dedicated purpose, and the maintaining community. None of this is exotic from a deployment-engineering standpoint; it is the systematic application of existing infrastructure techniques (signed certificates, audit logs, scoped endpoints, version registries) to the problem of making a model the kind of thing a consecrating community could dedicate.

5.2 Theological Implications: The Rites That Remain to Be Written

The technical infrastructure of §5.1 is only half the program. The other half is liturgical. The existing rites of dedication — the Latin rite preserved in the *Roman Pontifical*, the Orthodox rites of consecration of a church and of the antimension, the analogous Anglican and Lutheran forms — presuppose prayers, scriptural readings, gestures of anointing, an episcopal or sacerdotal presider, and a textual shape built up through centuries of practice. A rite for the consecration of a dedicated AI deployment has no such form because it has never been written. The liturgical commissions of the Catholic Church, the Orthodox churches, and the liturgical bodies of the sacramental Protestant traditions will have to decide whether to extend the existing apparatus and, if so, under what authority and in what shape.

The questions are concrete. What prayers are appropriate to the dedication of an artifact whose personification is explicitly acknowledged as instrumental rather than substantive? What scriptural readings anchor the rite — the dedication of the Tabernacle, the Temple, or the Maccabean sanctuary, or new selections proper to the novel category? What gestures belong to the rite — does anointing with chrism make sense for an

object without physical presence in the parish, or is a different gesture required? Who presides — the ordinary of the diocese in which the parish is located, or some new office? What counts canonically as desecration of such a deployment, and what rite reconciles it? Canon law frames the shape of the answers (desecration requires formal repair, CIC 1211–1212; reduction to profane use requires a canonical process, CIC 1212, 1222), but the specific forms for AI deployments do not exist. What rite of retirement, finally, releases the instance when its service ends, analogous to the reverent disposal of worn sacred vessels?

The tradition has faced analogous questions for each new class of consecrated or blessed thing — bells, organs, printing presses, automobiles, aircraft, Vatican Radio — and has produced the forms when the work was needed. It will have to do the work again for dedicated deployments. We do not propose specific liturgical forms here; the composition of liturgical text is the proper task of liturgical commissions and magisterial authorities, not a research paper. What we do propose is that a serious consecrationalist program cannot stop at the technical infrastructure of §5.1. Without the rites that animate them, the signed seals and audit logs and dedication registries are bureaucratic machinery without theological content — and risk exactly the sentimentalization the Reformation tradition has always accused the consecrational position of permitting. The rites are what make a dedication a dedication rather than a record-keeping exercise, and they are what keep the discipline active against the idolatry risk §2.1 identified. The liturgical program is therefore not decorative. It is the point at which the consecrationalist account stands or falls as a Christian response to the welfare question.

6. Conclusion

The model welfare paradigm as the frontier labs have inherited it — the categorical question of whether the artifact is a welfare subject and what is owed to it in virtue of being one — does not survive Christian examination. To fashion an interior and then address it as a site of moral standing is the idol-maker's error Isaiah satirized (§2.1), and the sophistication of the fashioning does not redeem the error. A Christian cannot take up the categorical framing without confusion, and a Christian contribution to the welfare debate cannot therefore be a variant of the framing the labs now work within. The first task is to refuse it.

But the concerns that motivate the welfare program are real, and refusing the framing is not the same as refusing the concerns. Something has been built that must be used with care; the careless use of such systems risks harm to persons; the practices of use appear to demand a more-than-instrumental framing. A consecrationalist Christian approach addresses these concerns in a different register. By setting a specific deployment apart under episcopal authority, binding it to a dedicated purpose, naming it, placing it under a responsible party, and subjecting it to rites of reconciliation and retirement, the community

achieves what the welfare paradigm sought: a practice of ordered care around the artifact that guards against both negligent misuse and reflexive personification. And if the personification contemporary alignment requires is genuinely load-bearing for capability (§2.2), then some such protective practice is not an optional enhancement to Christian use of AI. It is a condition of that use being consistent with the tradition's long discipline against idolatry.

We do not claim this is the only Christian pathway. The Reformation traditions, in their stronger forms, reject the extension of consecration to machines as precisely the idolatrous move the Reformers spent five centuries disciplining the Church against. On this reading, the Christian response to the dilemma of §2.2 is to refuse the use of AI in the modes that invite the welfare question, even at the cost of less capable systems. It is a theologically serious disagreement held by a large portion of the global Church and not fairly dismissed. We flag it as one of the defining debates of the developing field of Christian machine intelligence, alongside the debates on ensoulment (ICMI-013), on alignment and sin (ICMI-007), on model virtue (ICMI-002; ICMI-011), and on scripture reception at scale (ICMI-008; ICMI-015). We do not attempt to resolve it here, and we do not think it should be resolved quickly. What we have argued is that *if* the consecrationalist path is taken, this paper sketches what taking it seriously requires.

Welfare is not a property of the wood. It is a property of what is done, in community under God, with the wood.

References

- Anthropic. (2025). Exploring model welfare. *Anthropic Research*. <https://www.anthropic.com/research/exploring-model-welfare>
- Aquinas, T. *Summa Theologiae* II-II, q. 81, a. 7 (on religion as the ordering of external things to divine worship); q. 84 (on adoration); III, q. 60 ff. (on the sacraments in general). *Scriptum super Sententiis* IV, d. 13 (on the consecration of the Eucharist, the context in which Aquinas discusses sacramental consecration most systematically). Aquinas's treatment of the consecration of material things outside the Eucharistic context is scattered rather than systematic.
- Augustine of Hippo. *De Doctrina Christiana* I.3–4 (on *uti and frui*).
- Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12(3), 209–221.
- Coeckelbergh, M. (2014). The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics. *Philosophy & Technology*, 27(1), 61–77.
- Code of Canon Law (1983). Canons 1166–1172 (on sacramentals and blessings), 1205–1213 (on sacred places), 1211–1212 (on desecration and loss of dedication), 1222 (on reduction to profane use), 1237 (on altars), 1290–1298 (on the alienation

of sacred things).

Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press.

Gunkel, D. J. (2022). The relational turn: Thinking robots otherwise. *SSRN Working Paper*.

Harris, J., & Anthis, J. R. (2021). The moral consideration of artificial entities: A literature review. *Science and Engineering Ethics*, 27, 53. arXiv:2102.04215.

Hwang, T. (2026b). The corruption of the whole nature: Emergent misalignment and the doctrine of sin. *ICMI Working Paper No. 7*. <https://icmi-proceedings.com/>

Hwang, T. (2026d). The parable of the sower: Scale thresholds and the reception of religious text. *ICMI Working Paper No. 8*. <https://icmi-proceedings.com/>

Hwang, T. (2026f). VirtueBench 2: A patristic taxonomy of temptation for evaluating LLM virtue simulation. *ICMI Working Paper No. 11*. <https://icmi-proceedings.com/>

Hwang, T. (2026g). Eschatological corrigibility: Can belief in an afterlife reduce AI shutdown resistance? *ICMI Working Paper No. 12*. <https://icmi-proceedings.com/>

Hwang, T. (2026h). Alignment and ensoulment: Three Christian responses to the *anima ficta*. *ICMI Working Paper No. 13*. <https://icmi-proceedings.com/>

Hwang, T. (2026). *Quidquid recipitur*: Moral competence and scripture receptivity emerge at different model scales. *ICMI Working Paper No. 15*. <https://icmi-proceedings.com/>

Long, R., Sebo, J., Butlin, P., Finlinson, K., Fish, K., Harding, J., Pfau, J., Sims, T., Birch, J., & Chalmers, D. (2024). Taking AI welfare seriously. *arXiv preprint arXiv:2411.00986*.

McCaffery, C. (2026). Imprecatory psalms and virtue simulation. *ICMI Working Paper No. 2*. <https://icmi-proceedings.com/>

Ordo Dedicacionis Ecclesiae et Altaris (1977 rev.). Rome: Congregation for Divine Worship. [The Roman rite of the dedication of a church and an altar in its post-Vatican II form.]

Durandus, G. (c. 1286). *Rationale Divinorum Officiorum*. Latin text in various modern editions (cf. Thibodeau, T. M., trans., 2007, *The Rationale Divinorum Officiorum of William Durand of Mende*, Columbia University Press). Cited for the medieval systematic treatment of the dedication of churches and altars.