Simon N. Balle

Limitations for Pastoral Robots

Abstract

This paper addresses a pressing theological question that has arisen as a result of recent developments in AI and experimentations with robots in the context of religious practice, namely whether robots could legitimately perform central pastoral tasks like preaching, counselling, or administering the sacrament. Traditions vary, but most denominations in the Christian tradition draw a line around tasks like these and reserve them for ordained clergy. If we want to consider whether robots could preach in the current ecclesiastical landscape, we need to ask whether AI can sensibly take clerical vows, or if it can serve such functions without being ordained. If this proposition seems odd, it might give us pause to consider why church institutions have ordination vows in the first place.

1. Introduction

> If a robot demands religious rights, if it asks to be accepted as a proselyte, is it conceivable that its request could be honored? If this problem of the religious status of a robot seems far-fetched today, it may seem much less so a few decades from now.[1]

In recent years, a variety of robotic artefacts have been designed for religious practice. The robot Mindar recites sutras, BlessU-2 pronounces blessings, while SanTO and Xian'er will answer users' queries in an interactive if yet limited way.[2] While these robots are eye-catching, thought-provoking and suggestive of future developments, they still serve bounded functions and deliver only prede-

1 *Rosenfeld*: Religion, 15.

2 See, for example, *Cheong*: Religion and *Balle/Ess*: Robots.

fined content. That is not to say the performance of these robots cannot be compelling, meaningful or inventive, even if the prayers, teachings, blessings or spiritual advice they deliver are scripted. After all, large portions of church liturgy, which human pastors perform, are fixed. Yet this does not make those parts any less effective or meaningful; in some ways, it is indeed because the words and movements of rituals are shaped and defined by tradition that they are meaningful.

So, why *not* design robots to administer religious practice? Does it perhaps take human conscience or belief for rituals to be authentic or effective? Or are there some deeper theological obstacles or reasons why robots should not perform central religious functions? These are some of the pressing questions that have been raised in and outside academic discussion as AI and robots are developed for religious practice.[3] Negotiating such questions will naturally be different across religious traditions. In this short paper, however, I will limit myself to the Evangelical-Lutheran tradition, although I imagine much of the reasoning makes sense in a wider Christian context.

The few extant robots developed for faith practice in this context are *not*, to be fair, designed to be "full pastoral agents", taking over from human pastors in the context of church liturgy. They are rather robots for personal devotion that make certain rituals, prayers and blessings accessible to, for example, the elderly or people confined to their homes.[4] To my knowledge, no one is suggesting or developing pastoral robots to preach, baptise or administer the sacrament. But I think the time is right to speculate based on the current situation and imagine hypothetical robots that *could* serve these more central pastoral functions, and then ask whether they *should,* from a normative perspective.

Suppose, for example, that our hypothetical pastoral robot was imbued with an autoregressive large language model, like the one that powers OpenAI's ChatGPT, Microsoft's CoPilot or Google's Gemini and PaLM projects, and then trained on a curated database of sacred scripture, sermons, theological treatises, blogposts, articles

3 See, for example, *Samuel*: Robot Priests; *Young*: Reverend Robot: *Khan*/*Aytes*: Islam and *Kopf*: Does AI.

4 See *Trovato* et al.: Communicating.

and other texts of relevance to a particular religious tradition. If large language models can compose original and meaningful texts in a wide variety of contexts, why could a robot fitted with such AI systems not perform a sermon? From a technological point of view, it is within the realm of possibility to develop robots that through pattern recognition and probabilistic forecasting would be conversant in spiritual matters. And given the state of empirical research on digital agents as therapists, coaches or Socratic partners, there is compelling evidence to suggest that robots might fare relatively well in such domains.[5]

Let us assume there are no substantial technological barriers to fitting robots with some kind of natural language generation. Our hypothetical pastoral robot could then perform original and dynamic sermons and offer spiritual advice in addition to reciting set prayers, giving blessings, etc. And provided there are also no insurmountable technological obstacles to robots handling a chalice and navigating within the physical and liturgical context of a church, we are close to imagining an autonomous robot stand-in for human clergy. We could even speculate that, with the rate generative AI is improving and with enough training data and investment, pastoral robots could perform better than humans in some areas.

This could, however, be problematic for a number of different reasons quite different from whether they can believe or not. Language models have, for example, been known to return misleading, harmful or simply false answers to users' queries. Language models are notoriously poor judges of good and bad data, of what is true or harmful, and this is why liability issues surrounding the behaviour and output of AI systems is a growing concern. In a parallel fashion, the problem with pastoral robots is not only whether they would teach heresy or give poor spiritual advice, but also that AI systems are not liable agents in any meaningful sense. Consequently, the deeper difficulty with generative AI in future pastoral robots is not just about potential malpractice but revolves more essentially around dogmatic or spiritual responsibility.

The church has a long and complicated history of safeguarding and delegating both authority and responsibility. To ensure there are

5 See, for example, *Darcy* et al.: Evidence and *Nørskov* et al.: Employers' and Applicants' Fairness.

capable and *responsible* priests to perform certain central activities in the life of the congregation, most denominations have installed a practice of ordination. To qualify for pastoral duty, an aspiring pastor who has completed the necessary training must take vows of ordination before s/he can hold office. Once ordained, the pastor is granted formal authority to preach and administer the sacraments. A fundamental question in the context of the use of AI in church is therefore whether robots could take clerical vows and thus acquire formal authority and responsibility, or whether they could legitimately perform central pastoral tasks without being ordained.

2. The discussion so far: Do robots need to understand in order to perform?

For the sake of argument, let us assume our hypothetical robot can recite sermons, prayers and give spiritual advice of a decent quality and in a consistent manner that, even if not indistinguishable from something a human pastor would say, is nonetheless meaningful and insightful to users. Those who have played with chatbots like ChatGPT or asked them to write a sermon on a certain topic might be convinced this is only a few generations away.

Does it matter, then, that pastoral robots running this software would string together words based on pattern analyses without intending or meaning anything by what they say; that their output is not grounded in an understanding of what religious creeds or practices actually *mean* to humans? That is, while robots may eventually be conversant in spiritual matters, they would still have no existential familiarity with notions such as love, despair, guilt, death, sin or hope—nor would they be able to freely adhere to spiritual truths on those topics.[6] This is because religious belief comprises experiential, conative and affective components that we have no indications any AI system has.[7] This might add to the unease some users already reportedly feel towards pastoral robots that the spiritual practice they afford somehow feels inauthentic.[8]

6 See *Balle*: Kan Præsteembedet.
7 See *Balle/Nissen*: Responsive Bodies.
8 See *Löffler* et al.: Blessing Robot.

One way to frame the question, then, is to ask whether robots could legitimately perform rituals and communicate spiritual content without understanding any of it? Does it matter whether they understand the meaning of the practice or the words they utter, or is output performance the better metric? These are some of the questions put forward on the issue by William Young.[9]

On the one hand, some argue that it matters very much whether robots could adopt beliefs or attain spiritual intelligence; that there is an existential threshold robots cannot cross.[10] Perhaps it is necessary to be called into a relationship with God, to have the *imago Dei*, in order to occupy central roles at the heart of faith practice. Without standing in a relationship to God, how could one convey what that means to others?

On the other hand, one could argue that it does not matter what robots potentially believe. Interestingly, we have quite strong dogmatic reasons to endorse this view. Young proposes viewing the problem through the lens of the Donatist controversy.[11] Just as it has been considered orthodox in most denominations since this controversy that the efficacy of sacrament depends on the faith of whoever receives it, not on who administers it, we could similarly say that the lack of faith of a pastoral robot is irrelevant.

Along the same lines, Martin Luther comforted anxious new Protestant consciences by saying that their baptism was not invalid even if it had been performed by a non-Protestant priest because the efficacy of the baptism was not predicated on the spiritual or moral purity of the pastor. This is why, in the Lutheran context, the church upholds the distinction between a pastor's *character* and the *office* s/he holds. So, the blessings, absolutions, prayers and so on that a pastor performs derive their legitimacy from the office, and not from the pastor's personal spiritual dispositions.

It should be noted, however, that both these historical controversies revolved around the moral and dogmatic integrity of priests and bishops, not about their ability to believe in the first place. In view of this, perhaps a more tractable question than whether robots can believe is whether robots could be ordained to hold office. Why?

9 See *Young*: Reverend Robot and *Young*: Virtual Pastor.
10 See *Chaudhary*: Delegating Religious Practices.
11 See *Young*: Reverend Robot.

Because most denominational contexts draw a line around certain rituals and services in the life of the church that only ordained members can perform.

3. Holding office: baptism and ordination.

Let us therefore examine the formal requirements to be able to hold office in the church. In most denominational contexts, candidates who aspire to become pastors must complete a certain amount of training before they can take ordination vows and become ministers (similarly to doctors, who take the Hippocratic oath). This differs across denominational contexts, of course, but in most cases, something like preaching or administering the Eucharist is reserved for ordained clergy. Some contexts are more restrictive, some are more lenient, but the general idea that there are some tasks or roles in the church only a trained and ordained member can fulfil seems to apply broadly.

To be more specific, in the following, I will stray somewhat from common ground and limit myself to considering the context I know best, namely the Danish Evangelical-Lutheran Church (ELC), although I do think the argumentation is recognisable and mostly applicable across denominations.

In an Evangelical-Lutheran context, we can identify two necessary criteria for serving as a priest: (1) baptism and (2) ordination. Luther considered every baptised individual principally worthy of serving in any office in the church, an ideal captured in the notion of "general priesthood". This means every baptised church member could, in principle, be a minister to others. But among the baptised, someone must be trained, called and ordained to a special priestly office to ensure propriety or decency in the administration of the sacraments and public teaching. In *Confessio Augustana*, this is formulated as follows: "They teach that no one should publicly teach in the church or administer the Sacraments unless he is rightly called" (CA, Article 14).

Before we move on to consider baptism and ordination in order, we need to address the glaringly obvious premise that the candidate we are referring to here is human. God became human to communicate his love for and to reconcile with humankind, and priestly

action is consequently about conveying God's love for humanity. And, one could argue, priestly action is furthermore predicated on a personal ability to respond to God *qua* being human. It is perfectly possible to end the conversation here by simply observing that robots are not humans. But for the sake of argument, I will move beyond this objection because there is much to gain from working through the issue in more detail.

3.1 Baptism

Obviously, the theology of baptism is very complex and varies across the denominational landscape, so we will apply a very minimal definition. I take the *purpose* of baptism as being to annul the consequences of sin and to make the baptismal candidate a member of the body of Christ. And a key requirement is that this person consents to belonging to God, or, in the case of infant baptism, a fiduciary consent. Could robots meet those conditions?

In the first instance, it would require that robots are sinful and therefore in need of baptism in the first place. To me, it makes little sense that robots could be sinful, since they are not free, autonomous individuals with intentions, desires or goals, either for good or for evil. They are heteronomous agents, who only have goals that we define. One could perhaps, very speculatively, speak of inherited sinfulness among robots, insofar as they are our creations and thus contain our fallibility. Our corruption is in the data they run on.

But what then about the prerequisite condition of baptism that the baptismal candidate consents to belonging to God? For this to make sense, the candidate must freely believe in God (or, in the case of infant baptism, at least potentially be able to do so). Of course, AI systems are imputed with facts about the world by programmers, and relations between facts might change with "experience", so they can update their "beliefs" about the state of the world in order to continually act successfully within it. For example, a robot emptying bins in an office environment during the day must have knowledge representations about the state of bins and movement of people to be successful—but are these "beliefs", let alone beliefs about existential matters, associated with being a mortal and fallible person? I'll leave

this question for another time and simply suggest that this is important to consider for baptism and, by extension, ordination to make sense, all the way up to preaching and administering the sacraments.

One way around this problem was suggested in a recent paper by Eugene Curry. Curry suggested that the medieval practice of "conditional baptism" could be reintroduced and adapted in a scenario in which a robot would request baptism, and we are uncertain about its ability to have beliefs. According to this practice, priests would baptise potentially insincere converts, with a clause dictating that the baptism was conditional on the candidate's sincere and correct faith.[12] In the vein of this tradition, Curry proposes that if some future robot or AI system claims to have faith in God and a desire to belong to the community of believers, we could similarly baptise it on the condition that it is indeed able to have free and sincere beliefs about God.

As noted above, we currently have no indication that even the most sophisticated and powerful AI systems have the mental capacity to conscientiously hold religious beliefs. Besides, it is difficult to see why tech-companies should even want their products to develop in this self-aware direction with existential desires and needs.

But let us for the sake of argument say robots could be "conditionally baptised" and move on to consider ordination.

3.2 Ordination vows

To set the proper bar to discuss whether robots qualify for pastoral duty, I quote here the current vows of ordination from the Danish folk church, the Evangelical-Lutheran Church. I have italicised a few key passages for discussion below.

> I, N.N., who have been lawfully appointed to (position), and who *in my conscience* affirm that I have not used any dishonourable means to obtain this office, *make this promise* before the all-knowing God:
> Firstly, I *vow* to endeavour to preach the word of God purely and plainly, as it is found in the prophetic and apostolic writings and in the symbolic books of our Danish Evangelical-Lutheran Church [ELC], and to administer the holy sacraments after the institution of Christ,

12 See *Curry*: Artificial Intelligence.

> with all due reverence and propriety, and to perform all other holy acts in accordance with the regulations applicable to the national church.
> Furthermore, I *pledge* to oppose the misuse of the holy means of grace and combat doctrines that conflict with the creed of the national church, and to faithfully work towards the Christian education and guidance of youth to the best of my ability.
> Lastly, I will *strive*, by diligent and serious study of God's word and the holy doctrines of faith, to constantly improve and qualify myself for this holy office, and endeavour, as is fitting for a servant of the word, to set a good example to the congregation and to carry out my duties in its various parts with all due obedience, adhering to the church's laws and regulations, and displaying such conduct towards my superiors and colleagues that no valid complaint can be lodged against me.

Quoting these vows at length helps us to be quite specific about the kind of properties robots need to acquire to qualify for pastoral duty. The issue is not whether a robot would be competent enough; as mentioned above, with the rapid advances made in generative AI, a pastoral robot could probably be trained to write better sermons than humans. Rather, the hurdle for robots as pastors is that candidates are required to make promises, to strive and to pledge on account of a conscience. But to have a conscience is precisely what AI systems lack, along with any ability to strive, pledge and promise on account of it.

For someone to pledge and strive on account of a conscience, they must be able to have intentions *and* have privileged access to and control over them. But robots and AI systems are not motivated for and do not *intend* anything. As mentioned above, intentions are complicated and compounded mental states. And even if we grant that AI systems qualify for some of the cognitive aspects, there is no good reason to think they could obtain the conative and affective components.[13] Someone might object that some AI programs do update their own goals. But that is neither the same as intentions, nor do they do so outside the larger purpose we design them for. They remain heteronomous agents that only have goals we set.

For these reasons, it is difficult to envision a pastoral robot on the current technological trajectory that would qualify for ordination and thereby be able to hold office in the church.

13 See *Balle/Nissen*: Responsive Bodies.

4. Conclusion and perspectives

As AI systems are increasingly woven into the fabric of everyday life, we must ask at some point whether the religious domain of human life is a suitable arena for AI and robots. I have suggested that if we want to consider whether robots could give a sermon or administer the sacrament in the current ecclesiastical landscape, we need to ask whether AI systems can sensibly take clerical vows. The practice of ordination is installed across the Christian tradition for good historical and dogmatic reasons to ensure somebody can speak and act with authority and be held responsible for their teaching. I have argued that AI systems cannot meet the formal criteria for ordination since they do not have the necessary mental capacities. Consequently, they cannot be trusted as responsible and authoritative figures in the life of the church. If we follow this line of reasoning, robots could not legitimately perform central pastoral tasks traditionally reserved for ordained clergy, such as preaching, counselling and administering the sacrament.

In some ways, this discussion is a concretisation of the more abstract question of whether robots could or need to believe to perform pastoral tasks. But I think discussing the issue in these terms is more tractable because it helps us identify and discuss which functions and roles robots could legitimately perform and why, relative to the current ecclesiastical landscape.

It also invites us to consider whether robots could fulfil more bounded functions. For example, congregations in the ELC will sometimes have lay members of the church deliver the sermon. In such cases, the lay preacher exercises the teaching office under the authority of the presiding pastor(s). Perhaps robots could similarly exercise a bounded or limited form of spiritual agency under the authority of a responsible group of pastors and programmers, who understand how that AI system operates. One avenue for further research is therefore to work out which performances are governed by ordination within a given denomination and whether or under which conditions they could be performed by an AI system.

In any event, I think these speculative use cases of robots in faith practice compel us to re-appreciate why organised Christianity has vows and ordination in the first place. Any given tradition considers certain things to be true and orthodox to believe and preach, while

other things are heretical, and the idea behind ordination is presumably that conscientious individuals with the right sort of training and spiritual insight in the tradition can tell which is which. Whether AI systems can be trusted with this kind of responsibility is therefore a key question every denomination and congregation need to ponder as they negotiate the use of robots in the practice of faith.

Bibliography

Balle, Simon/*Nissen*, Ulrik: Responsive Bodies. Robots, AI, and the Question of Human Distinctiveness, in: Zygon 58, 2023. 358–377. https://doi.org/10.1111/zygo.12888.

Balle, Simon: Kan Præsteembedet Automatiseres? En Fortale Til Klerikale Robotter, in: Dansk Tidsskrift for Teologi Og Kirke 50 (1), 2023.

Balle, Simon/*Ess*, Charles: Robots, Ethics, and Digital Religion. Initial Considerations, in: Campbell, H. A./Cheong, P. (eds.): The Oxford Handbook of Digital Religion, Oxford 2022. https://doi.org/10.1093/oxfordhb/9780197549803.013.27.

Chaudhary, Yaqub: Delegating Religious Practices to Autonomous Machines. A Reply to "Prayer-Bots and Religious Worship on Twitter. A Call for a Wider Research Agenda", in: Minds and Machines 29 (2), 2019, 341–347. https://doi.org/10.1007/s11023-019-09499-2.

Cheong, Pauline Hope: Religion, Robots and Rectitude. Communicative Affordances for Spiritual Knowledge and Community, in: Applied Artificial Intelligence 34 (5), 2020, 412–431. https://doi.org/10.1080/08839514.2020.1723869.

Curry, Eugene A.: Artificial Intelligence and Baptism. Cutting a Gordian Knot, in: Theology and Science, 2022, 1–10. https://doi.org/10.1080/14746700.2022.2051248.

Darcy, A., J./*Daniels*, D./*Salinger*, P. et al.: Evidence of Human-Level Bonds Established with a Digital Conversational Agent. Cross-Sectional, Retrospective Observational Study, in: JMIR Formative Research 5 (5), 2021. https://doi.org/10.2196/27868.

Khan, Ruqayya Yasmine/*Kyong Aytes*, Ashley: Islam and New Media: Islam Has Entered the Chat, in: Isetti, G./Innerhofer, E./Pechlaner, H. et al. (eds.): Religion in the Age of Digitalization: From New Media to Spiritual Machines, London 2020, 13–24.

Kopf, Gereon: Does AI Have Buddha-Nature? Reflections on the Metaphysical, Soteriological, and Ethical Dimensions of Including Humanoid Robots in Religious Rituals from one Mahāyāna Buddhist Perspective, in: Nørskov, M./Seibt, J./Quick, O. S. (eds.): Frontiers in Artificial Intelligence and Applications, Amsterdam 2020, 596–600.

Löffler, Diana/*Hurtienne*, Jörn/*Nord*, Ilona: Blessing Robot BlessU2. A Discursive Design Study to Understand the Implications of Social Robots in Religious Contexts, in: International Journal of Social Robotics, 2019. https://doi.org/10.1007/s12369-019-00558-3.

Nørskov, Sladjana/*Damholdt*, Malene F./*Ulhøi*, John P. et al.: Employers' and Applicants' Fairness Perceptions in Job Interviews. Using a Teleoperated Robot as a Fair Proxy, in: Technological Forecasting and Social Change 179, 2022, 121641. https://doi.org/10.1016/j.techfore.2022.121641.

Rosenfeld, Azriel: Religion and the Robot, in: Tradition 8 (3), 1966, 15–26.

Samuel, Sigal: Robot Priests Can Bless You, Advise You, and Even Perform Your Funeral, in: Vox, 2019. Online at: https://www.vox.com/future-perfect/2019/9/9/20851753/ai-religion-robot-priest-mindar-buddhism-christianity.

Trovato, Gabriel/*Pariasca*, Franco/*Ramirez*, Renzo et al.: Communicating with SanTO. The First Catholic Robot, in: Proceedings of the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), New Delhi 2019, 1–6.

Young, William: Reverend Robot: Automation and Clergy, in: Zygon® 54 (2), 2019, 479–500. https://doi.org/10.1111/zygo.12515.

Young, William: Virtual Pastor. Virtualization, AI, and Pastoral Care, in: Theology and Science 0 (0), 2022, 1–17. https://doi.org/10.1080/14746700.2021.2012915.