Artificial Intelligence and Religion: Recent Advances and Future Directions

with Andrea Vestrucci, *"Introduction: Five Steps Toward a Religion–AI Dialogue"; Lluís Oviedo, "AI and Theology: Looking for a Positive—But Not Uncritical—Reception"; Christoph Benzmüller, "Symbolic AI and Gödel's Ontological Argument"; Sara Lumbreras, "Lessons from the Quest for Artificial Consciousness: The Emergence Criterion, Insight-Oriented AI, and Imago Dei"; Marius Dorobantu, "Artificial Intelligence as a Testing Ground for Key Theological Questions"; and Andrea Vestrucci, "Artificial Intelligence and God's Existence: Connecting Philosophy of Religion and Computation."*

INTRODUCTION: FIVE STEPS TOWARD A RELIGION–AI DIALOGUE

by Andrea Vestrucci

Abstract. This introduction to the thematic section of *Zygon: Journal of Religion and Science* on "Artificial Intelligence and Religion: Recent Advances and Future Directions" outlines the five articles by dividing them into two groups: the three that analyze the impact of recent advances in subsymbolic artificial intelligence (AI) on religion and theology, and the two that explore theological concepts in symbolic AI environments. These five articles are five steps toward a strong, deep, and interdisciplinary dialogue between the research in religion and the research in AI.

Keywords: artificial intelligence; artificial intelligence and religion; artificial intelligence and theology; imago Dei; subsymbolic artificial intelligence; symbolic artificial intelligence; technology and religion

Each of the five articles constituting this thematic section offers to the reader a step of an ideal ladder aimed at strengthening the interdisciplinary dialogue between artificial intelligence (AI) and religion, or at outlining new paths for such a dialogue.

Rather than following the order of the contributions in the tematic section, I organize the outline of the five articles by dividing them into two

Andrea Vestrucci is a Research Professor at Starr King School, Oakland (CA), USA, and he is currently working at the Chair for Artificial Intelligence System Engineering of the Faculty of Information Systems and Applied Computer Sciences, University of Bamberg, Germany; e-mails: avestrucci@sksm.edu, andrea.vestrucci@gmail.com.

groups, following the two notions of symbolic and subsymbolic AI. These notions are explained in Benzmüller's article and summarized here.

By "symbolic AI," we refer to programs able to understand and manipulate symbols representing objects and concepts and apply logical rules to prove propositions (theorems) about these concepts. Clearly, one of such theorems can concern the concept of "God." Sure, for a religious person or a theologian, it might sound bizarre to consider God just as a concept. However, any ontological proof of God's existence begins precisely by positing God as just an abstract concept, and then proving that this position is logically not coherent, or incompatible with a given definition of God. On the other hand, by "subsymbolic AI," we refer to programs based on so-called models learned from a large amount of data, which are used to classify new information and new incoming data. This is the type of AI that appears most frequently on media, in light of the recent improvement of the problem-solving ability of this subfield.

The five articles collected in this thematic section cover both types of AI and analyze their multiple connections with questions, problems, and insights coming from the religious field. The articles by Oviedo, Lumbreras, and Dorobantu focus on subsymbolic AI. The article by Benzmüller, and my own, focus on symbolic AI.

Lluis Oviedo's article explores what has been done, and what is left to do, in the theological approach to the AI phenomenon. Concerning what has been done, Oviedo presents an analysis of the current state-of-the-art advances in subsymbolic AI as interpreted within the theological context of theology, in particular under the ethical, soteriological, and anthropological perspectives. In this way, the article invites us to explore further the ways theology can communicate with AI technology, for theology to rethink and reconfirm its relevance in the modern world. Theology has seemed to progressively lose credibility and interest in the current scientific context, although theological production still seems to be rich and flourishing at least in quantitative terms (Fehige and Vestrucci 2022). Rather than a theological "hermeneutics of suspicion" toward AI achievements, Oviedo stresses the necessity of a collaborative model between the research in technology and theology: a religious approach *can* interact with AI advances to enhance the quality of human life since the two refer to specific aspects of human life. This interaction applies also to the respective fields of research: AI technology might help to clarify the nature of belief and to dealing with big data useful for theological research (texts, statistics, reports on religious phenomena); on the other hand, theology might help to better asses how AI technology can be at the service of human flourishing.

Sara Lumbreras focuses on AI consciousness. Consciousness can be defined in several ways, for example, as awareness, self-awareness, and attention (Russell and Norvig 2022, 1037). Machine performances can be misinterpreted as manifestations of consciousness. Lumbreras discusses the

"emergence criterion": according to this standard, consciousness is to be evaluated not on external outputs (like scoring positively in some tests), but as a result of the machine's underlying structure. This invites us to consider the "black box" notion in AI: for machine learning on a large amount of data, the actual functioning of the algorithm makes for difficult interpretation by programmers and users. This leads to machine biases, and thus to the reinstatement of the fundamental role of human operators to assess the ethics of AI system. Lumbreras connects the quest for artificial consciousness to the Christian notion of *imago Dei*: acknowledging the risks of reducing consciousness to a successful score in some tests invites us to focus the notion of human likeliness to God not on performative intelligence but on interiority, including the appreciation of rich interpersonal relationships, the capacity of contemplation, and artistic creativity. As such, technological challenges allow reinterpretation of theological concepts: the creative innovation of human beings includes the creational power of producing machines that can, at their turn, become co-creator beings.

Marius Dorobantu presents a rich spectrum of ways theology engages AI advances and technologies. He discusses the limits of the hypothesis of subsymbolic machine learning algorithms being applied to theological-educational purposes, for example, the formulation of the computational model of a saint. Dorobantu also harks back to the notion of *imago Dei*, in his case formulating alternative questions and analyses. The hypothetical emergence of human-level AI seems to challenge the distinctiveness of human cognitive capacities. However, as in Darwinian evolutionary theory, this challenge is the chance for theology to improve the clarity of its notions and to keep up to its "transformative powers." On the other hand, considering the relational approach to *imago Dei*, if some interpret the aim to build intelligent AI as a twisted replacement of our missed relation with God, the *imago* could be a work in progress, a constant tension toward an ideal. Following this scenario, machines could be included too in this performative approach, by considering the possibility of machines being aware of their own finitude. This might give new directions to inclusive pluralist theology. Finally, Dorobantu proposes speculative ways in which theology could engage with AI and computer science to tease out solutions to classical questions such as divine infinity, theodicy, or the nature of demonic intelligence.

I pass now to symbolic AI. Research on AI is also, simultaneously, research on the notion of intelligence, including human intelligence, its powers, and its limits. Symbolic AI is strongly anchored in logic, which might be interpreted as the formal study of intelligence and the plural forms of reasoning. Thus, the implementation and use of symbolic AI systems make it possible for us to gain new access to our reasonings and to the outputs of our intelligence, in line with logic's double task to analyze/test (*ars iudicandi*) and to discover/experiment with (*ars inveniendi*)

propositions, certainties, and theories. This is the idea underlying the two articles on symbolic AI and religion, Benzmüller's and Vestrucci's.

Christoph Benzmüller's article provides the reader with two important analyses. First, a summary of the evolution of the research in AI, and the different ways the advances in symbolic and subsymbolic AI interacted until today. Second, Benzmüller summarizes almost ten years of research by him and his team on the use of symbolic AI to explore Gödel's ontological argument of God's existence. The article reconstructs the vicissitudes that Gödel's argument underwent throughout the years since 1970, outlining the changes and criticisms that logicians and mathematicians formulated to the argument. But more importantly, Benzmüller addresses the different steps of the implementation of AI symbolic systems to the argument: this required the translation of the (varieties) of Gödel's argument into the specific syntaxes of two automated reasoning assistants (or theorem provers): Leo-II and Isabelle-HOL. The AI program helps to simplify the argument by applying new mathematical concepts, confirms the inconsistency of some versions of Gödel's argument, and clarifies how to solve some logical problems resulting from the argument such as "modal collapse," that is, the affirmation of determinism. The use of theorem provers not only opens up unexpected perspectives on metaphysical concepts, but also allows manipulation of these concepts as in an empirical experimental setting.

My contribution (Vestrucci) to the thematic section is also constituted by two parts. The first part analyses another exploration of metaphysical and theological arguments via the use of automated reasoning programs: this time it is Anselm's version of the ontological argument, which Oppenheimer and Zalta formalized and translated in the syntax of first-order theorem prover Prover9. The article reconstructs the steps that led to the discovery that the AI program needs fewer principles to prove the argument, compared to the original formalization. This computationally discovered simplification raised unexpected questions about the structure of the ontological argument. Then, the article's second part presents an assessment of one of these questions: whether, and with which results, the mathematical method called diagonal method (or diagonalization argument) can be applied to the ontological argument. Assessing the diagonalization of the ontological argument means to evaluate the consequences that this mathematical method has on our understanding of metaphysical and theological concepts, such as the conceivability of God and the property "thing than which none greater can be conceived." Assessing the applications of symbolic AI programs to metaphysical/theological arguments means applying the outputs of such applications to current debates in philosophy of religion and theology, for example, the limits of natural theology, the relationship between theology and mathematics, and the scientificity of theology.

These five articles are five steps toward a strong, deep, and interdisciplinary dialogue between the research in religion and the research in AI. Each of them, from its specific perspective, shows the importance of theology and philosophy of religion to familiarize with the outputs and the techniques in AI, and for AI research to refer to philosophy to test and, possibly, better conceive its tasks and ends.

At the same time, these five articles show that much is still to be done, and that the interaction between religion and AI is just beginning. Some of the key questions for future research may include: How can we conciliate the problem-solving approach in AI developments with a religious tendency to overcome a performative attitude? What are the specific ways in which classical topics and recent advances in theology can be fruitful and useful for symbolic AI research? What are the future, concrete outcomes and performances in subsymbolic AI that will raise questions to theology?

This thematic section of *Zygon: Journal of Religion and Science* expresses the hope that these five articles will invite and possibly lead readers to engage with these, and other important questions to be explored.

References

Fehige, Yiftach, and Andrea Vestrucci. 2022. "On Thought Experiments, Theology, and Mathematical Platonism." *Axiomathes*.

Russell, Stuart, and Peter Norvig. 2022. *Artificial Intelligence: A Modern Approach, Global Edition*. London: Pearson Education.