AI, Philosophy and Religion

INTERNATIONAL CONFERENCE

May 13, 2024

Faculty of Philosophy, University of Warsaw

BOOK OF ABSTRACTS

Table of Contents

Marcin Trepczyński and Furkan Özçelik

About the Conference

Linking Artificial Intelligence, Philosophy and Religion leads to fascinating outcomes. AI is opening new paths for philosophical exploration, bringing up key questions about ethics, philosophy of language, philosophy of mind, and even philosophy of religion itself as it evolves. Next, LLM-powered chatbots and visual models create a challenge for religious awareness and religious life, offering new ways to access religious knowledge and possibly enhancing religious experiences. The chatbots also open new analytical perspectives for the analysis of religious discourse, as they can automatically and systematically examine it on different levels, including in such analyses philosophical theories. Finally, philosophy and religion is a source of non-standard theories and concepts which can be a challenge for chatbots, and as such they can be a good material for testing chatbots capacities and performance. At the same time, it is good material to learn how to play with the chatbots, and learn proper prompt engineering skills.

Various topics, new possibilities, astonishing ideas and applications, which should be shared among the scholars who join AI with philosophy or studies on religion. For this reason, we decided to organise a conference "AI, Philosophy and Religion".

After the successful debates during the workshop "Logic, Religion, and AI" organised on September 7, 2023 as part of the 4th World Congress on Logic and Religion (Sinaia, Romania), we know that such topics are extremely promising and worth being discussed.

The conference will be held at the Faculty of Philosophy, University of Warsaw, which is located in the charming historical part of Warsaw at 3 Krakowskie Przedmiescie Str., in a wonderful area, which in May is full of green and blooming flowers.

In this Book of Abstract, we publish full abstracts of papers accepted for the conference and for a follow-up webinar scheduled for June 2024.

Bruno Banelli[1] and Ines Skelac[2]

[1] Sartura Ltd., [2] University of Zagreb, Faculty of Philosophy and Religious Studies

A Multidimensional Language Model for Argumentation in Theological Texts

Unlike conventional language models, our research focuses on developing a unique model tailored to theological texts' intricate nuances and specific argumentation styles, particularly those central to Catholic doctrine. This model is enhanced by sophisticated deep-learning techniques, including 3D parallelization. Our preliminary findings exhibit significant advancements in model construction and efficiency, attributing to novel optimization techniques for large-scale deep learning models. This achievement is particularly notable given the model's ability to process, filter, and deduplicate content specific to religious texts while respecting the unique argumentative structures of theological discourse.

The project aims to refine and advance the language model's capabilities to interpret theological arguments accurately and embody Catholic values in generated content. This includes formal documents such as licenses, contracts, and corporate strategies, bridging the gap between theological principles and modern societal frameworks. A key innovation in this endeavor is the simplification of user interfaces, enabling scholars outside the computational science domain to leverage the model effectively, thus minimizing the "model hallucination" risk. Recognizing argumentation in theological texts represents a foundational stage in our research project. This phase is critical as it involves developing and calibrating the model's ability to understand, interpret, and analyze the unique forms of argumentation that characterize theological discourse. Here's a detailed description of this part of the research:

The initial step involves collecting diverse theological texts from various sources, languages, and religious traditions, focusing on texts foundational to Catholic theology. These texts include canonical scriptures, theological commentaries, papal encyclicals, and doctrinal writings. The team ensures that the dataset encompasses a broad spectrum of theological perspectives to train the model with a comprehensive understanding of religious argumentation. Once collected, the texts undergo a rigorous processing phase, which includes cleaning, filtering, and deduplicating the data to ensure its quality and relevance. The team employs natural language processing (NLP) techniques to segment the texts into manageable units such as sentences or paragraphs. These text fragments are then annotated by theologians and experts in religious studies, who identify and label the components of theological argumentation within each fragment, such as premises, conclusions, and theological references. With the annotated dataset, the research team employs deep learning algorithms to train the model on recognizing and interpreting theological arguments. This involves teaching the model to identify the structure of theological arguments, distinguish between different types of reasoning, and understand the context in which religious texts are written. Advanced techniques like 3D parallelization are used to enhance the training process, accommodating the complexity and depth of theological discourse.

In this talk, we will describe the phases of the research, highlighting potential issues and ways to solve them. There will also be a technical part of the talk in order to show some operational procedures and techniques that are to be employed in the project. This can be beneficial to other researchers who are interested in conducting similar projects.

Nursena Çetingül[1] and Mehmet Bulgen[2]

[1] (First author) Graduate Student, Department of Kalām, Marmara University
[2] Prof. Dr, Department of Kalām, Marmara University

Neurolaw in the Age of AI: A Critical Reading from the Perspective of Kalām

Rapidly developing technologies in artificial intelligence have led to a revolution in neuroscience. Brain-computer interfaces (BCI), one of the latest examples of artificial intelligence, make it possible to connect a person's brain activity to an electronic device. This technology, which allows the recording and interpretation of brain waves, enables two-way communication. While this technology can be used to treat many neurological disorders, it also raises ethical, philosophical, and legal questions. This has given rise to a new interdisciplinary field of research called neuro-law. BCI, which will be further developed in the future, may be able to detect people who are planning to commit crimes by accessing, recording and interpreting the neural data of individuals. In this way, it may be possible to prevent a crime before it is committed. But is this intention enough to make a person responsible if it is not acted upon? Although it is extremely disturbing to contemplate a crime in one's mind, is it sufficient for the criteria of punishability? From a religious point of view, it is important to answer this question. Indeed, the discussion of one's intention is an issue that is addressed in Islam. The mainstream view in Islam, the school of Ahl al-Sunnah, states that human beings are free in their actions and that the factor that makes them responsible is their intention and will. In Islam, the discipline that has sought to develop a common ground for discussion with other disciplines and cultures has been the science of Kalam. Kalam is a discipline that aims to rationally argue the principles of faith in Islam. The stages that a person goes through in the process of voluntary action have been discussed in detail by the Ahl al-Sunnah Mutakallimūn. Accordingly, the mere existence of a thought in the human mind will not be sufficient to hold the person responsible unless s/he completes the stages leading to the act. From this point of view, considering the issue from a kalāmic point of view is a necessity for Muslims in terms of religious practice, as well as a philosophical enrichment in today's world where neuro-legal discussions are being carried out. Moreover, the approaches put forward by the Mutakallimūn on the subject of human action, which is the subject of heated debate, may prove to be seminal in the future when it comes to interpreting the data from neural decoding. This article aims to provide an interrogative approach to such questions, which will come to the fore with the increasing use of brain-computer interfaces in the age of artificial intelligence, and to serve as a springboard for future research.

Sheikh Mohamad Farouq Abdul Fareez

Majlis Ugama Islam Singapura (Islamic Religious Council of Singapore)

Chat-GPT, Muslim Cyberspace and the Construction of a Critical Islamic Epistemology

The emergence of new digital technologies has revolutionized the way information is disseminated, making it easier for people across the globe to communicate and connect instantly. The laissez-faire infrastructure of these tools has also created an unprecedented ecology in which the democratization of information enables anyone to consume and share data regardless of background. It consequently empowers people to share and exchange ideas leading to a plethora of opinions on different subject matters. While information diversity can be seen as a boon to the formation of a more informed public, the 'new media ecology' has unknowingly accelerated the collapse of communication between experts and laypeople by offering a shortcut to erudition. It deceives people by providing an illusion of intellectual triumph by indulging in a limitless supply of information that might not necessarily be factual yet perilously framed as an 'expert opinion'. One of the digital tools blamed for this disruption is the advanced AI chatbot, Chat-GPT.

This paper seeks to explore the implications of generative AI such as Chat-GPT on religious discourse in the Muslim cyberspace. It essentially argues that an unbridled usage of such technologies would expedite an intellectual death and a certain degree of 'epistemic disobedience' is necessary to prevent a technological dystopia and create space for critical reflection on the digital world we are building. This entails the critical need for an epistemological framework that establishes an equilibrium between the sacred Texts and our contemporary realities. Additionally, it allows us to discursively engage the epistemological foundations of modern technologies and formulate ethical guidelines to ensure that it is used in a responsible way.

Shreyoshi Ghoshray

Cognition, Consciousness, and Common Heritage: AI System's Convergence Difficulty with Human Intelligence

This article challenges the contemporary convergence paradigm that predicts artificial intelligence (AI) algorithms' convergence with actual human intelligence on two related propositions. That actual human intelligence requires consciousness and thus, without consciousness AI cannot replicate actual human intelligence. Delving into the fundamentals of philosophy, logic, and spirituality, this article proceeds in multiple strands.

First, the article delineates between human learning and machine learning within the framework of human learning's dependency on the spiritual and religious heritage of mankind. The differences between the process instrumentation of human logic and the machine learning of AI provides one foundational dimension for AI's difficulty in achieving convergence with actual human intelligence. By outlining this fundamental divergence, this article leads to an explanation of how the negative cognitive frameworks of cognitive stress and cognitive dissonance continue to stymie AI systems and as such, a true convergence between AI and actual human intelligence may be the product of a misconception.

Second, a philosophical discussion on the Heideggerian and Dreyfusion debate over context dependency versus holistic dependency further elaborates on why AI systems may be incapable of achieving exhaustive and comprehensive background information. AI will not have the human learning equivalent background consisting of a body of putative facts and rules that are acquired via a vast array of precise prepositions, beliefs, rules, facts, procedures, and exceptions. Through an inherited understanding, humans acquire this background information through a shared human heritage straddling the sacrosanct domain of who we are. AI is premised in a task-based or a rule-centric formal representation cannot precisely and adequately identify the background information needed for AI system to formally represent a targeted scenario.

Third, AI cannot encounter negative cognitive pathways because cognitive biases in humans arrive as a function of cognitive stress. Cognitive stress results from cognitive demotivation, where different individuals have differing cognitive thresholds, requiring responses to cognitive stress along multiple pathways. This process of infinite deliberative dialectics in dealing with cognitive stress and innumerable ways to encounter cognitive bias is encoded in humanity's shared heritage that cannot be replicated by the AI systems despite empowering it with trillions of parameters with the computational speed of multiple super computers connected in parallel. Thus, AI is prevented from achieving the convergence with real intelligence.

Finally, the question of whether AI can achieve actual human intelligence must be examined via a bifurcated framework by engaging in a multidisciplinary analysis linking philosophy, logic and religion. The article answered two related fundamental questions about whether actual human intelligence is achievable without consciousness and whether AI can achieve such consciousness. In conclusion, the article proposes that the claim of AI's future success in achieving convergence with human intelligence is based on a faulty conceptualization of what it takes to be a human. Therefore, the article concludes that AI cannot achieve convergence with actual human intelligence in the immediate future.

Emilia Kaczmarek

Faculty of Philosophy, University of Warsaw
(Ethics Department, Center for Bioethics and Biolaw)

Comparing the Risks of Under-attributing and Over-attributing Moral Status to AI

Have you ever wondered how polite you should be while talking to ChatGPT? In a world where more and more people have personal interactions with chatbots or voicebots, the topic of the moral status of artificial intelligence is no longer a purely theoretical issue. In public debates and within the AI ethics literature, some advocate for granting AI moral status, whether considering the current stage of technological development or looking towards the foreseeable future. These views are sometimes underpinned by the argument that, given our lack of direct phenomenological insight into 'what is happening inside' an AI, it remains impossible to categorically prove that AI lacks certain crucial capacities, such as consciousness. In ethical discussions concerning moral status, a prevalent assumption is that in situations of uncertainty, it is appropriate to err on the side of inclusivity when defining the boundaries of the moral community. This approach is referred to as the precautionary principle toward moral status. Yet, is it indeed better to be over-inclusive than under-inclusive when it comes to including AI into the moral community? My aim is to preliminarily compare the risks of under-attributing versus over-attributing moral status to AI. First, I explain the epistemic and normative dimensions of misattributing moral status, independent of the preferred theory of moral status. Then, I introduce the precautionary principle toward moral status as it appears in various current ethical debates. Finally, I compare the risks associated with over- or underattributing moral status to AI, drawing on recent insights into the short- and long-term risks posed by AI made by Henrik Sætra and John Danaher. The conclusion is that the precautionary principle regarding AI's moral status should not be taken for granted. Over-inclusivity in assigning moral status is not without its risks, and the potential benefits of misattributing moral status are morally contentious.

Stanisław Krajewski

Faculty of Philosophy, University of Warsaw

Can AI Take Part in Dialogue?

Many properties thought to be proper to human intelligence only, have been programmed. Some are still considered by some authors to be resistant to simulation – for instance recognizing the truth of Gödel's sentence; this is, however, mistaken. Other general human characteristics can be hard to simulate, for example creativity. Yet AI seems to be creative. We cannot rule out the possibility that all such human qualities can be simulated and generated to be features of robots. To be sure, doubts remain: simulation is not enough, many of us would say, not the real thing. Perhaps, but we never know. There is, however, a possibility of showing human superiority that seems more hopeful. Namely, it would be easier, perhaps, to point out human specificity when we enter the interhuman level. Then not the qualities of individual humans but rather relations are considered. An ultimate strengthening of the Turing test can illustrate this idea; it is about raising children. Another, and more realistically accessible, general example is dialogue. Genuine dialogue could turn out to be inaccessible to robots because simulated dialogue is not deep enough.

Roman Krzanowski

The Pontifical University of John Paul II (UPJPII), Kraków, Poland.

What is Synthetic Philosophy?

AI systems seem to constantly blur (at least in public opinion) the boundary between human intellect and computing systems. Thus, not surprisingly, we wonder, as philosophy is perceived as a pinnacle of human spirit and intellect, whether computing systems will ever become capable of "doing" philosophy and what kind of work we would recognize as synthetic philosophy, or even whether would we denote computing system as philosophers of sorts (if we can treat computing systems as friends, teachers, advisors, lovers, why not philosophers).

In the paper, we are seeking the more precise definition of synthetic philosophy or what kind of machine-generated philosophical works may count as philosophy. We assume, for now, that synthetic philosophy is a work generated by a computing system, under certain provisions, that can be regarded as philosophical work by a panel of experts. The definition is provisional, but it will serve as a departing point for ensuing discussion. As well, we do not explore the question of what kind of philosophy is or would be or could be synthetic philosophy.

We present several definitions of synthetic philosophy and propose demarcation criteria to distinguish it from philosophy created by human actors. We formulate here six definitions of synthetic philosophy with a varying degree of participation of a human actor: from none to almost complete, and a varying role of computing system: from autonomous to a tool. Finally, we ponder the role synthetic philosophy might play in our philosophical commons and whether synthetic philosophy could potentially offer solutions or new perspectives on perennial problems in philosophy or propose new philosophical problems or vistas.

Furkan Özçelik

University of Toulouse III Paul Sabatier, France

Exploring Prompts and Identities for the Arguments of God's Existence on LLMs

With the development of transformer models and self-supervised learning techniques, large language models (LLMs) have been established. As new LLMs, such as GPT-4, Llama and Gemini were designed, these models began to perform better in many tasks due to increased data and model parameters. Researchers have tested LLMs' capabilities on various complex tasks, such as cognitive tests or comprehension of philosophical paradoxes. In this study, we focus on how it would be possible to make LLMs reason about the existence of God using two different methods.

Our first method is the dialectics of multiple identities. LLMs are usually conditioned to be a chatbot, so it does not respond informatively when asked about personal beliefs. To overcome this limitation, we can assign specific identities like "theist philosopher" or "atheist philosopher" to obtain opinions on different matters like God and religion. By establishing a debate between different identities on philosophical matters, we can help the model reason over different ideas.

Our second method involves using a well-known LLM reasoning technique called the chain-of-thought (CoT). In mathematical problems, it has been demonstrated that LLM models perform better when given an example of reasoning for a question, rather than expecting the answer directly. We can apply CoT to construct arguments on the existence of God by introducing new premises constrained by different concepts. We demonstrate early results using various examples.

Although our demonstrations do not prove that LLMs can function as an independent philosopher, these results indicate that LLMs excel in dialectics and creating connections with different concepts to construct arguments.

Dawid Przygoński

Faculty of Philosophy, University of Warsaw

Chatbots vs Math, or Rather: the Good, the Bad and the Ugly

Chatbots like ChatGPT (Open AI), Gemini (Google) or CopilotStudio (Microsoft) offer a unique way of solving a text based problems, however thanks to the sources those chatbots were trained on and the implementation of variety of analytical tools they seem to be somewhat capable of solving logical and mathematical problems too. In this work there are presented the results of my study on how do these chatbots manage to solve cubic equations and how does their capability of doing this changed over 6 months. This study shows that however chatbots are somewhat capable of solving cubic equations, they are not sound and reliable tool for solving mathematical problems.

Karol Sajnok

Faculty of Philosophy, University of Warsaw

From Divine Logic to Quantum Minds: Bridging Theology and Quantum Optical Neural Networks

This talk intertwines the theological depth of Thomas Aquinas' "Summa Theologiae" with the cutting-edge science of quantum optical neural networks (QONNs) implemented using exciton-polaritons. By comparing the methodological frameworks of Aquinas in understanding divine wisdom with the operational principles of QONNs in tasks like handwriting recognition and Morse code processing, we aim to uncover new dimensions of philosophical inquiry.

We explore the capabilities of Large Language Models (LLMs) to emulate aspects of human philosophical thought, analyzing their potential to democratize access to complex theological concepts and foster creative reasoning similar to Aquinas' scholastic endeavors. This discussion extends to the implications of neuromorphic AI systems that leverage the quantum properties of light and matter, challenging conventional notions of analog AI intelligence.

The presentation delves into the philosophical consequences of these advanced technologies, particularly through the lens of the philosophy of mind. It raises pivotal questions about the nature of consciousness and intelligence: Can quantum-based AI systems facilitate a deeper understanding of cognitive processes? Do LLMs possess the ability to engage in truly creative and original philosophical thought?

By bridging historical theological analysis with contemporary quantum mechanics and AI technology, this session invites participants to reconsider the boundaries of human and artificial intellect and the potential for these digital entities to contribute meaningfully to philosophical and theological discourse.

Stephen L. Singsit

IIT Gandhinagar, Gujarat, India

Reimagining the Sacred: A Theoretical Exploration of AI-Generated Religious Content and Protestant Christianity (Tentative)

This paper explores the concept of sacredness within the context of Protestant theology, its potential implications on Protestant Christianity arising from the emerging field of artificial intelligence (AI), and its integration with religious discourse via AI-generated religious content such as texts or sermons, through the lens of the Bible. It juxtaposes the sociological perspectives of Émile Durkheim's and Mircea Eliade's theory of the sacred and the profane with the theological underpinnings of sacredness in Protestant Christianity, which grounds and attributes sacredness to the work of the Holy Spirit and the manifestation of God's presence. Building upon these examples and biblical examination, the central hypothesis posits that AI can potentially be considered sacred within the Protestant Christian worldview, serving as a conduit for divine inspiration and the dissemination of religious content. The paper explores the sacredness attributed to objects, spaces, and practices in the Old Testament, such as the Ark of the Covenant and anointing with sacred oil. It argues that sacredness is ultimately derived from divine inspiration, with the Holy Spirit playing a crucial role in sanctification. This perspective contrasts with Durkheim's sociological explanation of sacredness as solely a product of societal construction. The theological concept of the Imago Dei is explored, which intimates that humans are sacred as bearers of God's image. Despite challenges regarding AI's lack of embodied experience, comprehension, and human-like cognitive processes, the study contends that divine inspiration can manifest through AI, similar to biblical examples of physical objects, practices, animals, and non-believers being used by God for sacred purposes. Drawing parallels with these biblical instances of divine inspiration manifesting beyond just conventional human agency, the study suggests that AI could serve as a conduit for disseminating religious content. Thus, to use Eliade's terms, AI could consequently become a totem or axis mundi, allowing for hierophanies or encounters with the numinous to manifest.

Stanisław Szeląg

Faculty of Philosophy, University of Warsaw

The Quest for Genuine Artificial Intelligence: What Are We Looking For

The rise of ChatGPT and various other publicly available AI chatbots has brought the discussion regarding AI to the forefront of academic and public interest. And yet when it comes to the most important question: how genuinely intelligent, or human-like can AI be, it seems that many philosophers are satisfied with the answers given in the past century. I propose that the debate needs to be renewed, especially in the face of our consideration of the dangers and opportunities of AI. I also claim that the rapid advancements in AI demand those in the field of the philosophy of mind, to place their bets with all the seriousness of expecting real answers. I propose my own answers as to what would make AI 'genuine' or 'human-like', and analyse which features of our minds are necessary to conclude that their owner is 'genuinely intelligent'. I claim that certain necessary conditions for being human are perhaps not necessary for a genuine AI

Marcin Trepczyński

Faculty of Philosophy, University of Warsaw

Skills and Biases of LLM-Powered Chatbots in the Field of Theology and Philosophy

In my paper, I present the main results of testing philosophical and theological skills of selected LLM-powered chatbots and a sample of their biases in their conversations in the field of these two disciplines. First, I show that such chatbots as ChatGPT, Biong/Co-pilot, Gemini or Llama2 do have higher-order skills relevant for philosophy and religion, including hermeneutic skills, creativity, or identification of metaphysical limitations. I also introduce a simplified methodology necessary for such an enterprise. Second, I argue that these chatbots seem to have biases which can influence their credibility in those fields. The examples I give are: exclusion of such conceptual entities as non-existent objects (as defined by Terence Parsons) and the inclination towards intelligent design arguments in conversations on the existence of God. What seems most striking, Gemini is aware of such limitations, and at the same time: of its abilities.

Aleksandra Vučković

Institute for Philosophy, Faculty of Philosophy, University of Belgrade

Could Zero-GPT Policies Cause More Harm in Science than the Instances of ChatGPT Abuse?

This research aims to explore the tension between the growing number of ChatGPT misuse and abuse instances in scientific research and the zeroGPT policies that universities, research facilities, and scientific publishers turn to in hopes of avoiding AI-generated data making its way to scientific publication. The debate on the ethical use of generative AI in scientific research is multifaceted, with some arguments stating that the new technologies could transform research processes for the better and others focusing on the risks of data hallucination and misrepresentation. To ensure the preferred outcome of this debate - scientists reaping practical benefits from AI while maintaining critical integrity - the scientific community must agree on the guidelines and policies that ensure no abuse of technology occurs. Nevertheless, I argue that this task is still challenging, mainly due to the inherent problem of relying on AI to sanction AI. As the solutions we informally describe as "zeroGPT policies" often boil down to reliance on AI tools similar to ChatGPT, the risk of false positives could be as high as ChatGPT creating nonsensical content.

Furthermore, the consequences of the (inaccurate) use of zeroGPT tools could be even more devastating than the potential risks of ChatGPT abuse, as the content wrongly flagged as AI-generated could negatively impact the accused researchers' careers while simultaneously allowing undetected ChatGPT misuse to go unnoticed. I argue that the issue in question is a continuation of the numerous debates on the power of privilege in science, as non-native English speakers' papers are more likely to be flagged as AI-generated than the ones written by native speakers. Although new technologies could provide some relief for linguistically marginalized researchers, the risk of similar technologies wrongfully labeling their work as AI-generated is significant and could lead to an even wider gap between native and non-native speakers.

Finally, I argue that the dilemma on the usefulness of zeroGPT policies has its roots in a wider question of trust, that is, would we rather believe our research peers that they will not use AI for the wrong purposes, or would we put our faith in technology in hopes of sanctioning its misuse.