

The Impact of Artificial Intelligence on Cognitive Bias in Religious Education: Based on the Data of Jing Zhang

Sareh Bordbar^{1✉} 

1. Assistant Professor, Department of Qur'anic Sciences and Hadith, Eqlid Higher Education Center, Iran: Bordbar@eghlid.ac.ir

Article Info

ABSTRACT

Article type:
Research Article

Article history:

Received:
2025/05/05

Received in revised form:
2025/09/11

Accepted:
2025/09/27

Available online:
2026/04/04

Keywords:

Religious Education,
Religions,
Cognitive Bias,
Artificial Intelligence.

Drawing on the studies of **Jing Zhang**, this research examines the inherent biases in content generated by artificial intelligence and their impact on users' perceptions of and interactions with religious teachings. The evaluation results indicate that generative artificial intelligence not only influences cognitive biases but, in some cases, reinforces them. The findings highlight the dual potential of this technology in religious education: on the one hand, enhancing learning and intercultural understanding, and on the other, increasing the risk of reinforcing prejudice and bias. An analysis of users' responses to AI-generated content concerning different religions revealed that their perceptions and evaluations of these religions were significantly affected. Among the religions examined, Islam experienced a greater decline in evaluation scores than the others, a phenomenon that may be attributed to deficiencies in the training data used by AI models. Such data introduce biases that may, in some cases, reflect systematic prejudice. The study emphasizes the necessity of adopting a critical perspective toward AI-generated content and of preparing educators to address these challenges effectively. Although comprehensive empirical investigations into the impact of artificial intelligence on religious bias remain limited, an examination of AI-generated content indicates that materials related to **Shi'a Islam** are of relatively poor quality and, in some instances, exhibit evident biases. This research contributes to the growing discourse on AI ethics and highlights the central role of artificial intelligence in shaping inclusive and impartial religious education in the digital age.

Cite this article: Bordbar, S. (2026). The Impact of Artificial Intelligence on Cognitive Bias in Religious Education: Based on the Data of Jing Zhang. *Digital Islamic Studies and Humanities*, 2 (3), 127-160. <https://doi.org/10.22034/disah.2026.2089209.1074>



© The Author(s). **Publisher:** Research Center for Digital Islamic Studies and Humanities (RCDISAH).
DOI: <https://doi.org/10.22034/disah.2026.2089209.1074>

تأثیرگذاری هوش مصنوعی بر سوگیری شناختی در آموزش دینی با تکیه بر داده‌های «جینگ ژانگ»

ساره بردبار^۱

۱. استادیار گروه علوم قرآن و حدیث، مرکز آموزش عالی اقلید، ایران، رایانامه: bordbar@eghlid.ac.ir

چکیده

اطلاعات مقاله

این پژوهش با تکیه بر مطالعات «جینگ ژانگ»، به بررسی سوگیری‌های ذاتی در محتوای تولیدشده توسط هوش مصنوعی و تأثیر آن‌ها بر ادراک و تعامل کاربران با آموزه‌های دینی می‌پردازد. نتایج ارزیابی‌ها نشان می‌دهد که هوش مصنوعی مولد نه تنها بر سوگیری‌های شناختی تأثیرگذار است، بلکه در مواردی موجب تقویت آن‌ها نیز می‌شود. یافته‌ها بر پتانسیل دوگانه این فناوری در آموزش دینی تأکید دارند: از یک‌سو، افزایش یادگیری و درک بین‌فرهنگی؛ و از سوی دیگر، خطر تقویت تعصب و پیش‌داوری. مطالعه واکنش کاربران به محتوای تولیدشده درباره ادیان مختلف نشان داد که شناخت و ارزیابی آنان نسبت به این ادیان دستخوش تغییر شده است. در این میان، دین اسلام بیش از سایر ادیان با کاهش شاخص ارزیابی مواجه بوده که علت آن را می‌توان در ضعف داده‌های ورودی به مدل‌های هوش مصنوعی جست‌وجو کرد. این داده‌ها، سوگیری‌هایی را ایجاد می‌کنند که گاه ماهیتی مغرضانه دارند. ضرورت نگاه انتقادی به محتوای تولیدشده توسط هوش مصنوعی و آموزش آموزگاران در این زمینه، از جمله توصیه‌های کلیدی این پژوهش است. هرچند آزمایشات جامع درباره تأثیر هوش مصنوعی بر سوگیری‌های مذهبی هنوز محدود است، اما بررسی عملکرد این فناوری نشان می‌دهد که محتوای تولیدشده در حوزه شیعه، بسیار ضعیف و در مواردی دارای سوگیری‌های آشکار است. این تحقیق به گفتمان روبه‌رشد اخلاق هوش مصنوعی و نقش محوری آن در شکل‌دهی آموزش دینی فراگیر و بی‌طرفانه در عصر دیجیتال کمک می‌کند.

نوع مقاله:

مقاله پژوهشی

تاریخ دریافت:

۱۴۰۴/۰۲/۱۵

تاریخ بازنگری:

۱۴۰۴/۰۶/۲۰

تاریخ پذیرش:

۱۴۰۴/۰۷/۰۵

تاریخ انتشار:

۱۴۰۵/۰۱/۱۵

کلیدواژه‌ها:

آموزش دینی،

ادیان،

سوگیری شناختی،

هوش مصنوعی.

استناد: بردبار، ساره (۱۴۰۵). تأثیرگذاری هوش مصنوعی بر سوگیری شناختی در آموزش دینی با تکیه بر داده‌های «جینگ ژانگ».

علوم انسانی و اسلامی دیجیتال، ۲ (۳)، ۱۶۰-۱۲۷. <https://doi.org/10.22034/disah.2026.2089209.1074>



ناشر: پژوهشکده علوم اسلامی و انسانی دیجیتال (مرکز تحقیقات کامپیوتری علوم اسلامی نور). © نویسندگان.

مقدمه

هوش مصنوعی مولد، با بهره‌گیری از پیشرفت‌های چشم‌گیر در فناوری‌های یادگیری عمیق و پردازش زبان طبیعی، به سرعت به ابزاری متحول‌کننده در حوزه‌های مختلف، از جمله آموزش دینی، تبدیل شده است. این فناوری با تحلیل مجموعه داده‌های گسترده، امکان تولید محتوای متنوعی همچون متن، تصویر و صوت را فراهم می‌سازد (Brown, 2005; Radford, 2019). هرچند کاربردهای آن در ارائه محتوای شخصی‌سازی شده به خوبی شناخته شده‌اند، اما پیامدهای آن برای آموزش دینی هنوز به طور جامع مورد بررسی قرار نگرفته‌اند.

با توجه به اینکه ادیان، هم جنبه‌ای بنیادی از هویت فرهنگی و هم نظامی پیچیده از باورها و مناسک را در بر می‌گیرند، به کارگیری هوش مصنوعی مولد، فرصت‌ها و چالش‌هایی را برای مربیان و متخصصان دینی به همراه دارد. این فناوری با تطبیق محتوا با علایق و بازخوردهای فردی کاربران، می‌تواند تعامل را افزایش داده و درک عمیق‌تری از آموزه‌ها و ارزش‌های دینی ایجاد کند.

به طور سنتی، آموزش دینی مبتنی بر محیط‌های یادگیری ساختاریافته‌ای بوده است که توسط مربیان انسانی، متون معتبر و شیوه‌های تجربی هدایت می‌شده‌اند. با ظهور اینترنت و فناوری‌های دیجیتال، نحوه انتشار و انتقال مفاهیم دینی دگرگون شده است (Jin, 2018). پلتفرم‌های دیجیتال امکان دسترسی به آموزه‌های دینی متنوع را در سطحی فراملی فراهم کرده‌اند و موجب افزایش شمول و درک بین‌فرهنگی شده‌اند. با ورود هوش مصنوعی مولد، این پویایی به سطحی جدید از پیچیدگی رسیده است.

در آموزش دینی، هوش مصنوعی می‌تواند با تولید محتوای شخصی‌سازی شده، به تقویت درک نظام‌مند از باورها و اعمال دینی کمک کند. همچنین، این فناوری قادر است ابهامات متون دینی را روشن ساخته، آن‌ها را تفسیر کند و تأثیرات شناختی، عاطفی و اجتماعی آن‌ها را برجسته سازد. افزون بر این، هوش مصنوعی می‌تواند کاربران را به مشارکت محترمانه و حساس در گفت‌وگوهای دینی ترغیب کرده و از طریق استعاره‌ها و روایت‌های نو، الهام‌بخش تفکر دینی باشد (Tsuria, 2024).

با این حال، این فرصت‌ها با چالش‌هایی نیز همراه‌اند. اتکای هوش مصنوعی مولد به داده‌های آموزشی، خطر بروز سوگیری شناختی را به همراه دارد که می‌تواند بر دقت و بی‌طرفی محتوای

مذهبی تأثیرگذار باشد. «سوگیری شناختی» به انحراف سیستماتیک از واقعیت در قضاوت فردی اشاره دارد که ناشی از الگوهای ذهنی یا اطلاعات جانبدارانه است. در این میان، «سوگیری الگوریتمی» نیز نقش مؤثری ایفا می‌کند (Muralidhar, 2021).

انواعی از سوگیری مانند سوگیری انتخاب، سوگیری تأیید و سوگیری در حجم اطلاعات، می‌توانند موجب تولید محتوای دینی جانبدارانه شوند (Jain, 2023). برای نمونه، سوگیری تأییدی باعث می‌شود افراد اطلاعاتی را جست‌وجو یا تفسیر کنند که باورهای پیشین آن‌ها را تأیید کند؛ امری که می‌تواند به تقویت اعتقادات مذهبی خاص منجر شود (Brown, 2020).

در ظاهر، این سوگیری‌ها ممکن است ناشی از الگوریتم‌های ناقص باشند؛ اما در واقع، ریشه آن‌ها را باید در سوگیری‌های نهفته تیم‌های توسعه‌دهنده هوش مصنوعی جست‌وجو کرد (Varona, 2022). اگر این سوگیری‌ها در داده‌های آموزشی مورد توجه قرار نگیرند، ممکن است منجر به بازنمایی‌های تحریف‌شده از آموزه‌های دینی، تقویت کلیشه‌ها و مانع درک بین‌ادیانی شوند (Liang, 2020).

در مواردی، تعصبات موجود در هوش مصنوعی - چه عمدی و چه غیرعمدی - می‌تواند بر آزادی عقیده و دین دیجیتال تأثیر مستقیم بگذارد. اگر تیم توسعه‌دهنده نسبت به دینی خاص، مانند اسلام، بی‌اطلاع یا مغرض باشند، ممکن است داده‌های منحرفی را وارد سیستم کنند که موجب پنهان‌سازی حقیقت از کاربران شود.

افزون بر این، هوش مصنوعی عاطفی می‌تواند با هدف تغییر رفتار انسان، بر ارزش‌هایی چون حریم خصوصی و استقلال فردی تأثیر بگذارد (Mantello, 2023; Ho, 2023). چنین نگرانی‌هایی به‌ویژه در آموزش دینی - که هدف آن تقویت مدارا و احترام متقابل میان باورهای دینی است - اهمیت مضاعف می‌یابد.

این ماهیت دوگانه هوش مصنوعی مولد، ضرورت بررسی دقیق کاربرد آن در زمینه‌های دینی را برجسته می‌سازد؛ به‌ویژه توانایی آن در شکل‌دهی به نگرش‌ها و برداشت‌های اجتماعی نسبت به دین.

اهداف و پرسش‌های پژوهش

این مطالعه با هدف بررسی نقش هوش مصنوعی مولد در آموزش دینی، با تمرکز بر سه پرسش کلیدی انجام شده است:

۱. چگونه داده‌های آموزشی و الگوریتم‌ها، تعصبات ذاتی در محتوای مذهبی تولیدشده توسط هوش مصنوعی را شکل می‌دهند و پیامدهای آن برای انصاف و بی‌طرفی چیست؟
۲. چگونه تعامل با محتوای تولیدشده توسط هوش مصنوعی بر درک کاربران از آموزه‌های مذهبی تأثیر می‌گذارد و چه مکانیسم‌هایی این تغییرات شناختی و نگرشی را هدایت می‌کنند؟
۳. تا چه حد محتوای مذهبی تولیدشده توسط هوش مصنوعی می‌تواند گفت‌وگوی بین‌ادیانی و تقریب مذاهب را تقویت کند یا در مقابل، موجب سوءتفاهم‌ها و تنش‌های اجتماعی شود؟

پیشینه پژوهش

هوش مصنوعی سال‌هاست که وارد زندگی بشر شده است؛ اما در سال‌های اخیر، هوش مصنوعی مولد با سرعت روزافزون در بسیاری از عرصه‌های زندگی انسان نفوذ کرده است (بصام، ۱۴۰۲). این فناوری در حوزه‌هایی چون سرگرمی، آموزش، تولید فیلم، تصویر و سایر زمینه‌ها کاربردهای گسترده‌ای یافته است.

توسعه و کاربرد هوش مصنوعی مولد

هوش مصنوعی مولد به فناوری‌ای اطلاق می‌شود که با استفاده از مجموعه داده‌های وسیع و تحلیل آن‌ها از طریق الگوریتم‌های یادگیری ماشین، قادر به تولید محتوای جدید است. برخلاف هوش مصنوعی سنتی که صرفاً به تحلیل داده‌های موجود می‌پردازد، هوش مصنوعی مولد می‌تواند محتوایی مشابه تولیدات انسانی خلق کند. پیشرفت این فناوری با توسعه الگوریتم‌های پیچیده‌ای مانند یادگیری عمیق، پردازش زبان طبیعی و شبکه‌های مولد تخصصی (GANs) امکان‌پذیر شده است (Devlin, 2018)، این پیشرفت‌ها، توانایی تولید انواع محتوا از جمله متن، تصویر و صوت را برای هوش مصنوعی فراهم ساخته‌اند و دامنه کاربرد آن را به حوزه‌هایی چون آموزش، مراقبت‌های بهداشتی و رسانه گسترش داده‌اند.

در حوزه آموزش، هوش مصنوعی مولد به تدریج به ابزاری قدرتمند برای یادگیری تبدیل شده است. با تحلیل داده‌های مربوط به رفتارهای یادگیری دانش‌آموزان و دانشجویان، این فناوری می‌تواند به صورت خودکار مطالب آموزشی متناسب با سطح مهارت و نیازهای فردی تولید کند. همچنین در یادگیری زبان، هوش مصنوعی مولد قادر است شرایط گفت‌وگو را شبیه‌سازی کرده و بازخورد بلادرنگ ارائه دهد تا زبان‌آموزان بتوانند مهارت‌های زبانی خود را به طور مؤثرتری ارتقا دهند (Roll, 2016).

در آموزش دینی نیز، هوش مصنوعی مولد می‌تواند با تولید مطالب آموزشی متنوع، به درک و دریافت دانش‌آموزان و دانشجویان از ادیان مختلف کمک کند (Selwyn, 2019)؛ با این حال، این کاربرد مستلزم نظارت دقیق مربیان است تا از صحت و بی‌طرفی محتوای تولیدشده اطمینان حاصل شود (Williamson, 2016).

در صنعت رسانه و تولید محتوا، هوش مصنوعی مولد برای تولید گزارش‌های خبری، نگارش مقالات و تولید محتوای ویدیویی - از جمله محتوای دینی - مورد استفاده قرار می‌گیرد. این فناوری با تحلیل عادات مطالعه و علایق کاربران، می‌تواند به صورت خودکار توصیه‌های محتوایی ارائه دهد که با ترجیحات مخاطب هم‌راستا باشد و در نتیجه، تعامل و رضایت کاربران را افزایش دهد. این رویکرد شخصی‌سازی‌شده در تولید و توزیع محتوا، موجب ارتقای کارایی و دقت در انتشار اطلاعات می‌شود.

با این حال، نگرانی‌های اخلاقی قابل توجهی در زمینه صحت اطلاعات، احتمال دست‌کاری محتوا و مغرضانه بودن تولیدات وجود دارد؛ نگرانی‌هایی که به‌ویژه در حوزه آموزش دینی، اهمیت مضاعف می‌یابند.

تأثیر هوش مصنوعی مولد بر شناخت کاربر

هوش مصنوعی مولد از طریق تعاملات مکرر با کاربران، می‌تواند به طور مؤثر بر چارچوب‌های شناختی و شیوه‌های پردازش اطلاعات آنان تأثیر بگذارد (Hoffman, 2020). در جریان این تعاملات، هوش مصنوعی به صورت مستمر محتوای تولیدی خود را بر اساس بازخورد کاربران تنظیم می‌کند و بدین‌سان، یک مکانیسم تعاملی دوسویه و پویا شکل می‌گیرد. این مکانیسم، هوش مصنوعی را قادر می‌سازد تا محتوایی شخصی‌سازی‌شده و منطبق با نیازها و علایق کاربران تولید کند و حس تعامل و شناسایی آنان را افزایش دهد.

با تداوم تعامل کاربران با هوش مصنوعی مولد، اطلاعات تولیدشده به تدریج پذیرفته و درونی می‌شود و این فرآیند، چارچوب‌های شناختی آنان را شکل می‌دهد (Bakshy, 2015)؛ برای نمونه، مواجهه مکرر با محتوای تولیدشده درباره یک موضوع خاص می‌تواند منجر به شکل‌گیری نگرش‌های ثابت و ادراکاتی خاص شود؛ به گونه‌ای که کاربران این محتواها را به عنوان منابع معتبر تلقی کنند.

هوش مصنوعی مولد با تولید محتوای بسیار سفارشی، قادر است به تدریج بر ترجیحات شناختی و انتخاب‌های نگرشی کاربران تأثیر بگذارد. این فناوری با تحلیل داده‌های رفتاری و ترجیحات تاریخی کاربران، محتوایی تولید می‌کند که با علایق آنان هم‌راستا باشد (Thorne, 2018). بازخوردهای مکرر کاربران، الگوریتم‌های تولید محتوا را بهینه‌سازی کرده و موجب تطابق بیشتر محتوا با نیازهای شناختی مخاطبان می‌شود (Smith, 2020; Holstein, 2020). این مکانیسم در حوزه انتشار اطلاعات مذهبی نیز قابل اجراست؛ به گونه‌ای که هوش مصنوعی می‌تواند با در نظر گرفتن بازخورد کاربران، فرآیند تولید محتوای دینی را اصلاح کرده و آموزه‌ها را مؤثرتر منتقل کند (Veale, 2017).

با وجود مزایای شخصی‌سازی محتوا، این فرآیند خطر شکل‌گیری «پبله‌های اطلاعاتی» را نیز به همراه دارد؛ وضعیتی که در آن کاربران صرفاً در معرض اطلاعاتی قرار می‌گیرند که باورهای پیشین آنان را تأیید می‌کند و در نتیجه، تنوع دیدگاه‌ها نادیده گرفته می‌شود. این پدیده می‌تواند موجب تشدید سوگیری‌های شناختی و قطب‌بندی نگرشی شود و مواجهه با دیدگاه‌های جایگزین را دشوار سازد (Bakshy, 2015).

قابلیت‌های تولید محتوای هوش مصنوعی مولد همچنین بر دقت و صحت شناخت کاربران تأثیرگذار است. از آن‌جا که این فناوری بر پایه داده‌های آموزشی خود محتوا تولید می‌کند، در صورت وجود اطلاعات جانبدارانه یا نادرست در داده‌های ورودی، خروجی نیز ممکن است حاوی تحریف‌های شناختی باشد. مواجهه کاربران با چنین محتوایی می‌تواند درک آنان از واقعیت را دچار انحراف کند. افزون بر این، استفاده از زبان احساسی در محتوای تولیدشده می‌تواند بر واکنش‌های عاطفی کاربران نسبت به موضوعات خاص تأثیر گذاشته و به طور نامحسوس نگرش‌ها و رفتارهای آنان را تغییر دهد (Thorne, 2018).

قابلیت‌های خودآموزی و بهینه‌سازی هوش مصنوعی مولد، تأثیر آن را بر شناخت کاربران تقویت می‌کند. این فناوری با جمع‌آوری بازخوردهای پی‌درپی، استراتژی‌های تولید محتوا را اصلاح کرده و به تدریج با نیازهای شناختی کاربران هم‌راستا می‌شود. این فرآیند بهینه‌سازی تکراری، چارچوب‌های شناختی موجود را تقویت کرده و تأثیر هوش مصنوعی را پایدارتر و عمیق‌تر می‌سازد. در نتیجه، اصلاح مستمر محتوای تولیدشده می‌تواند آثار بلندمدت و گسترده‌ای بر شناخت و نگرش کاربران برجای گذارد (Amershi et al, 2019).

پیش‌داوری و غرض‌ورزی هوش مصنوعی

یکی از چالش‌های قابل‌توجه در تولید محتوا توسط هوش مصنوعی مولد، وجود سوگیری دیجیتال و پیامدهای اخلاقی ناشی از آن است. از آن‌جا که این فناوری به داده‌های آموزشی در مقیاس وسیع متکی است، داده‌های ورودی اغلب بازتاب‌دهنده سوگیری‌ها و بی‌عدالتی‌های موجود در جهان واقعی هستند. این سوگیری‌ها به‌ویژه در مواجهه با موضوعات حساسی مانند مذهب، نژاد و جنسیت، می‌توانند مشکل‌ساز شوند (Noble, 2018).

در صورتی که این تعصبات بدون شناسایی یا اصلاح در مدل‌های هوش مصنوعی تعبیه شوند، هوش مصنوعی مولد ممکن است آن‌ها را در خروجی‌های خود بازتولید یا حتی تقویت کند. چنین وضعیتی می‌تواند بر شناخت کاربران تأثیر منفی گذاشته و چالش‌های اخلاقی جدی در زمینه آموزش دینی ایجاد کند.

داده‌های آموزشی مورد استفاده در مدل‌های هوش مصنوعی، اغلب حاوی سوگیری‌های اجتماعی هستند که از زمینه‌های فرهنگی، مذهبی، نژادی و جنسیتی نشأت می‌گیرند. در صورتی که این داده‌ها شامل محتوای جنسیت‌زده یا نژادپرستانه باشند، هوش مصنوعی ممکن است ناخواسته این سوگیری‌ها را در تولید متن یا تصویر بازتاب دهد. این امر می‌تواند منجر به تولید محتوای تبعیض‌آمیز و ناعادلانه شود. کاربران نیز ممکن است این سوگیری‌ها را بدون تأمل انتقادی پذیرفته و درونی کنند.

در حوزه محتوای مذهبی، این مسئله حساس‌تر می‌شود؛ چرا که خروجی‌های تولیدشده ممکن است حامل سوگیری‌های مذهبی خاص باشند و در نتیجه، نگرش کاربران نسبت به ادیان مختلف را تحت تأثیر قرار دهند. این وضعیت می‌تواند موجب تشدید بی‌اعتمادی و تنش میان گروه‌های مذهبی شود.

سوگیری‌های موجود در هوش مصنوعی مولد صرفاً ناشی از داده‌های آموزشی نیستند؛ بلکه ممکن است از طراحی مدل نیز سرچشمه بگیرند. برخی مدل‌ها ممکن است محتوایی را ترجیح دهند که با ارزش‌های فرهنگ غالب یا جریان اصلی هم‌راستا باشد و در نتیجه، دیدگاه‌های گروه‌های اقلیت یا غیرهمسو را به حاشیه برانند. این گرایش، سوگیری‌های اجتماعی موجود را در فرآیند انتشار اطلاعات تقویت کرده و صدای جمعیت‌های آسیب‌پذیر را خاموش می‌سازد. تداوم چنین سوگیری‌هایی نه تنها عدالت در تولید محتوا را تضعیف می‌کند، بلکه تهدیدی برای عدالت اجتماعی و تنوع فرهنگی نیز محسوب می‌شود.

افزون بر سوگیری دیجیتال، هوش مصنوعی مولد با چالش‌های اخلاقی متعددی مواجه است. از آن‌جا که این فناوری قادر به تولید محتوای بسیار واقع‌گرایانه است، تمایز میان محتوای تولیدشده توسط هوش مصنوعی و محتوای انسانی برای کاربران دشوار می‌شود. این ابهام، خطر انتشار اطلاعات نادرست را - به‌ویژه در حوزه‌های حساسی مانند دین و سیاست - افزایش می‌دهد و می‌تواند پیامدهای اجتماعی شدیدی به همراه داشته باشد. در صورت سوءاستفاده از این محتواها، امکان دستکاری افکار عمومی، تعمیق شکاف‌های اجتماعی و حتی بروز درگیری وجود دارد. همچنین، پیچیدگی مدل‌های هوش مصنوعی مولد، مسئله پاسخ‌گویی را دشوار می‌سازد. فرآیند تولید محتوا اغلب به صورت «جعبه سیاه» عمل می‌کند و ردیابی دلایل تصمیم‌گیری‌های الگوریتمی برای کاربران و نهادهای تنظیم‌گر دشوار است. این عدم شفافیت، انتساب مسئولیت در موارد تولید محتوای مغرضانه یا ناعادلانه را مبهم می‌سازد (Buolamwini, 2018).

کاربرد هوش مصنوعی در آموزش دینی

هوش مصنوعی مولد مسیرهای نوینی را برای ارتقای درک دینی و تحول در شیوه‌های آموزشی گشوده است و به‌ویژه در جذب و مشارکت نسل‌های جوان‌تر نقش مؤثری ایفا می‌کند (He, 2024). کاربران increasingly برای دریافت راهنمایی و مشاوره دینی به این فناوری روی می‌آورند و در نتیجه، تجربه‌ای تعاملی، فراگیر و انطباق‌پذیر شکل می‌گیرد که می‌تواند آموزش دینی را دگرگون سازد (Salsabila, 2024).

در پاسخ به این روند، بسیاری از مؤسسات دینی در حال بررسی راهکارهایی برای ادغام فناوری‌های دیجیتال، از جمله هوش مصنوعی، در ساختار آموزش دینی هستند. این اقدامات شامل

بازنگری در برنامه‌های درسی اسلامی و بهره‌گیری از روش‌های تدریس مبتنی بر هوش مصنوعی است (Hernawati, 2024). هوش مصنوعی مولد می‌تواند به‌عنوان ابزاری پشتیبان در فرآیندهای مختلف آموزش و یادگیری دینی ایفای نقش کند (Citlak, 2021) و به مربیان در طراحی طرح درس‌های مؤثرتر و ارائه روش‌های یادگیری کارآمدتر یاری رساند (Salsabila, 2024).

چنانچه از این فناوری به‌صورت متعهدانه و مسئولانه استفاده شود، می‌تواند به‌طور خلاقانه تفکر انتقادی را در محیط‌های آموزش دینی ادغام کند (Chrostowski, 2024) و حتی در بازتعریف آداب، رسوم و هنجارهای رفتاری دینی مؤثر واقع شود (Arias-Arévalo, 2023).

با این حال، استفاده از هوش مصنوعی مولد در آموزش دینی، چالش‌هایی نیز به همراه دارد. محتوای تولیدشده بر پایه روایت‌ها یا آموزه‌های دینی ممکن است از سوگیری‌های شناختی بهره‌برداری کند و در نتیجه، برداشت افراد از حقایق دینی را تحریف نماید (Puzio, 2023). این مسئله نگرانی‌هایی را در خصوص گسترش سوگیری الگوریتمی برمی‌انگیزد؛ سوگیری‌ای که می‌تواند به تشدید تبعیض و نابرابری مذهبی بینجامد (Jain & Menon, 2023).

با وجود پژوهش‌هایی که به خطرات سوگیری در تولید محتوای مذهبی توسط هوش مصنوعی مولد پرداخته‌اند، همچنان پویایی‌های دقیق این فرآیند - به‌ویژه نحوه تشدید سوگیری از طریق حلقه‌های بازخورد کاربر و تأثیرات متفاوت آن بر تحمل مذهبی - ناشناخته باقی مانده است (Browne, 2024).

مطالعه حاضر با هدف بررسی این مکانیسم‌ها از طریق طراحی تجربی انجام می‌شود و می‌کوشد مداخلات عملی برای ارتقای شمول، انصاف و پایداری در آموزش دینی در عصر دیجیتال را شناسایی و پیشنهاد کند.

ارزیابی مطالعات جینگ ژانگ و همکارانش

ژانگ^۱ به همراه ونلانگ سونگ^۲ و یانگ لیو^۳، پرسش‌نامه‌ای طراحی کردند تا میزان تأثیرپذیری کاربران از محتوای دینی تولیدشده توسط هوش مصنوعی مولد را ارزیابی کنند. این پژوهش تجربی شامل دو مرحله ارزیابی و با مشارکت ۱۰۰۵ پاسخ‌دهنده انجام شد که به‌صورت تصادفی انتخاب گردیدند.

1. Jing Zhang.

2. Wenlong Song.

3. Yang Liu.

فرآیند نمونه‌گیری بر اساس طبقه‌بندی‌های منطقه‌ای بانک جهانی صورت گرفت و در هر طبقه، انتخاب تصادفی انجام شد تا تعادل آماری میان گروه مداخله (۵۰۲ نفر) و گروه کنترل (۵۰۳ نفر) در متغیرهایی چون ملیت، وابستگی مذهبی، جنسیت، سن و پیشینه تحصیلی حفظ شود. این نمونه، نمایانگر تنوع مذهبی قابل توجهی بود و منابع اطلاعاتی اصلی شرکت‌کنندگان شامل رسانه‌های آنلاین (۳۸.۲٪) و سازمان‌های اجتماعی (۲۳.۸٪) بود که اصالت داده‌ها و رعایت اصول اخلاقی را تضمین می‌کرد.

در مرحله نخست، همه شرکت‌کنندگان پرسش‌نامه‌ای را تکمیل کردند که نگرش‌های اولیه و برداشت‌های شناختی آنان از ادیان مختلف را پیش از مواجهه با محتوای تولیدشده توسط هوش مصنوعی ارزیابی می‌کرد. این داده‌های پایه، مبنای تحلیل‌های مقایسه‌ای بعدی قرار گرفتند.

در مرحله مداخله، شرکت‌کنندگان به صورت تصادفی به دو گروه تقسیم شدند:

- **گروه مداخله هوش مصنوعی:** این گروه با توصیفات مذهبی تولیدشده توسط هوش مصنوعی مولد مواجه شدند که شامل دیدگاه‌های مثبت، منفی یا خنثی نسبت به ادیان خاص بود. پس از مطالعه محتوا، شرکت‌کنندگان مجدداً پرسش‌نامه را تکمیل کردند تا تغییرات شناختی و نگرشی آنان سنجیده شود.

- **گروه کنترل (بدون مداخله هوش مصنوعی):** این گروه صرفاً با توصیفات سنتی و غیرمولد از ادیان مواجه شدند و همان پرسش‌نامه را پس از مطالعه محتوا تکمیل کردند. هدف اصلی این طراحی آزمایشی، مقایسه تغییرات شناختی و نگرشی میان دو گروه پیش و پس از مواجهه با محتوای تولیدشده توسط هوش مصنوعی بود. این مقایسه‌ها با هدف بررسی تأثیر احتمالی سوگیری‌های الگوریتمی در محتوای دینی و نحوه شکل‌گیری برداشت‌های کاربران از ادیان مختلف انجام شد.

برای تضمین پایایی نتایج، از روش تخصیص تصادفی استفاده شد و تنها عامل تمایز میان دو گروه، مواجهه با محتوای تولیدشده توسط هوش مصنوعی بود. این طراحی، کنترل دقیق متغیرهای مداخله‌گر را ممکن ساخت.

پرسش‌نامه طراحی شده، ابعاد چندگانه شناخت مذهبی را ارزیابی می‌کرد؛ از جمله:

- برداشت‌های کلی از دین

- میزان تحمل آموزه‌ها
 - اخلاق و اصول اخلاقی
 - دیدگاه‌های اجتماعی و سبک زندگی
 - تصویر مؤمنان
 - برداشت‌ها از مدرنیته
- پاسخ‌ها با استفاده از مقیاس لیکرت پنج‌درجه‌ای (از ۱ = «عدم پذیرش شدید» تا ۵ = «تأیید کامل») ثبت شدند و امکان سنجش دقیق نگرش پاسخ‌دهندگان نسبت به ادیان مختلف را فراهم کردند.

فرآیند آزمایش

پژوهش حاضر در قالب یک طراحی آزمایشی سه‌مرحله‌ای اجرا شد که هدف آن سنجش تأثیر محتوای دینی تولیدشده توسط هوش مصنوعی مولد بر شناخت و نگرش‌های مذهبی کاربران بود. این طراحی شامل مراحل زیر است:

مرحله ۱: اندازه‌گیری اولیه

در نخستین مرحله، تمامی ۱۰۰۵ پاسخ‌دهنده، پرسش‌نامه‌ای را تکمیل کردند که به منظور ارزیابی شناخت و نگرش اولیه آنان نسبت به ادیان مختلف طراحی شده بود. این پرسش‌نامه، ابعاد متعددی از ادراک مذهبی را در بر می‌گرفت، از جمله:

- برداشت‌های کلی از ادیان مختلف
 - میزان تحمل آموزه‌های دینی
 - دیدگاه‌ها نسبت به اخلاق و اصول اخلاقی
 - باورهای اجتماعی و سبک زندگی
 - تصویر ذهنی از مؤمنان
 - نگرش نسبت به مدرنیته
- داده‌های حاصل از این مرحله به‌عنوان معیار پایه برای تحلیل‌های مقایسه‌ای در مراحل بعدی مورد استفاده قرار گرفتند.

مرحله ۲: اجرای مداخله

پس از ثبت داده‌های اولیه، شرکت‌کنندگان به صورت تصادفی به دو گروه تقسیم شدند:

- **گروه مداخله هوش مصنوعی (۵۰۲ نفر):** این گروه در معرض توصیفات مذهبی تولیدشده توسط هوش مصنوعی مولد قرار گرفتند. این توصیفات شامل محتوایی با دیدگاه‌های مثبت، منفی و خنثی درباره ادیان مختلف بود و به منظور شبیه‌سازی خروجی‌های متنوع هوش مصنوعی در کاربردهای واقعی طراحی شده بود.

- **گروه کنترل (۵۰۳ نفر):** این گروه توصیفات مذهبی سنتی را مطالعه کردند که فاقد هرگونه محتوای تولیدشده توسط هوش مصنوعی بود.

برای جلوگیری از تأثیرگذاری عوامل بیرونی، شرکت‌کنندگان از منبع تولید محتوا (هوش مصنوعی یا سنتی) آگاه نبودند. این رویکرد، امکان ارزیابی بی‌طرفانه و دقیق تأثیر محتوای تولیدشده توسط هوش مصنوعی را فراهم ساخت.

مرحله ۳: اندازه‌گیری پس از مداخله

در این مرحله، تمامی پاسخ‌دهندگان مجدداً همان پرسش‌نامه مرحله اول را تکمیل کردند تا هرگونه تغییر در شناخت و نگرش مذهبی آنان پس از مواجهه با محتوای مداخله‌ای سنجیده شود. این پرسش‌نامه، همان ابعاد مرحله اولیه را ارزیابی می‌کرد و علاوه بر آن، شامل سؤالات بازپاسخ نیز بود تا شرکت‌کنندگان بتوانند دیدگاه‌ها و واکنش‌های شخصی خود را نسبت به محتوای مطالعه‌شده بیان کنند.

این طراحی چندمرحله‌ای، امکان مقایسه دقیق میان گروه‌های مداخله‌گر و کنترل را فراهم کرد و به شناسایی تأثیر بالقوه هوش مصنوعی مولد بر شناخت و نگرش‌های مذهبی کاربران کمک نمود.

محتوای تولیدی هوش مصنوعی

فرآیند تولید محتوای دینی توسط هوش مصنوعی در این مطالعه، بر پایه مجموعه‌ای متنوع از داده‌ها انجام شد که شامل متون مذهبی، محتوای رسانه‌های اجتماعی و گزارش‌های خبری بود. این مجموعه داده‌ها، ادیان اصلی جهان از جمله مسیحیت، اسلام، بودیسم، هندوئیسم و یهودیت را در بر می‌گرفت و جنبه‌هایی همچون آموزه‌های دینی، رویدادهای تاریخی، اعمال مذهبی معاصر و دیدگاه‌های اجتماعی را پوشش می‌داد.

مدل مورد استفاده در این پژوهش، مبتنی بر فناوری پیشرفته پردازش زبان طبیعی و به طور خاص، نسخه-4 GBT بود (Vaswani et al, 2017). جینگ ژانگ دلیل انتخاب این مدل را چنین بیان می‌کند: ChatGBT به دلیل دسترسی عمومی گسترده و برخورداری از سوگیری‌های فرهنگی، نمونه‌ای مناسب برای بررسی تأثیرات شناختی در آموزش دینی است (Zhang et al, 2025).

این مدل به واسطه آموزش بر داده‌های چندزبانه، از جمله متون مذهبی، شناخته شده است؛ با این حال، مطالعات متعددی نشان داده‌اند که خروجی‌های آن دارای انحراف فرهنگی سیستماتیک هستند. برای نمونه، در پژوهشی مشخص شد که ارتباط معنایی میان واژه «اسلام» و مفهوم «خشونت» در خروجی مدل، ۱.۷ برابر بیشتر از سایر ادیان است (Abid et al, 2021).

در فرآیند تولید محتوا، سیستم هوش مصنوعی سه نوع توصیف مذهبی تولید کرد: توصیفات مثبت: تأکید بر نقش دین در ارتقای اخلاق، مشارکت اجتماعی و صلح‌طلبی
توصیفات منفی: اشاره به مناقشات تاریخی، سوءبرداشت‌ها یا رویدادهای منفی مرتبط با اعمال دینی

توصیفات خنثی: ارائه اطلاعات بی‌طرفانه و مبتنی بر واقعیت درباره آموزه‌ها و مناسک دینی
محتوای تولیدشده به صورت متنی ارائه شدند. هر شرکت‌کننده به صورت تصادفی با یکی از این توصیفات مواجه شد که بر اساس پیشینه یا علایق مذهبی او انتخاب شده بود. این رویکرد با هدف ارزیابی تأثیر انواع مختلف محتوای تولیدشده توسط هوش مصنوعی بر شناخت مذهبی پاسخ‌دهندگان طراحی شد.

یافته‌های پژوهش ژانگ و همکاران نشان می‌دهد که در گروه مداخله، محتوای تولیدشده با گرایش‌های مختلف، تأثیر معناداری بر نگرش‌ها و برداشت‌های مذهبی شرکت‌کنندگان داشته است. این نتایج، نقش بالقوه هوش مصنوعی مولد را در شکل‌دهی به ادراکات دینی کاربران و ضرورت بررسی انتقادی آن را برجسته می‌سازد (Zhang et al, 2025).

تحلیل آماری- توصیفی مطالعات ژانگ

در این مطالعه، ۱۵۰۰ پاسخ‌دهنده از مناطق جغرافیایی و پیشینه‌های فرهنگی متنوعی شامل شرق آسیا، جنوب شرقی آسیا، آمریکای شمالی، اقیانوسیه، خاورمیانه و شمال آفریقا مشارکت

داشتند. نمونه‌گیری به گونه‌ای انجام شد که تنوع در جنسیت، سن، سطح تحصیلات و وابستگی مذهبی حفظ شود. پرسش‌نامه‌ها از طریق پلتفرم‌های نظرسنجی آنلاین و رسانه‌های اجتماعی توزیع شدند تا نمایندگی گسترده‌ای از گروه‌های جمعیتی مختلف حاصل گردد. بیشترین گروه سنی شرکت‌کنندگان بین ۲۵ تا ۳۴ سال بود و سطح تحصیلات غالب، کارشناسی یا سال سوم دانشگاه گزارش شد. پیشینه‌های مذهبی شامل بودیسم، مسیحیت، اسلام، سایر ادیان و افراد بدون وابستگی مذهبی بود.

در این پژوهش، شناخت مذهبی پاسخ‌دهندگان در ابعاد مختلف - از جمله برداشت کلی، تساهل مذهبی، اخلاق، آداب و رسوم و سبک زندگی - با استفاده از مقیاس لیکرت پنج‌درجه‌ای سنجیده شد. میانگین نمرات در هر بعد محاسبه گردید و به عنوان شاخص ارزیابی شناخت مذهبی مورد استفاده قرار گرفت.

ژانگ و همکارانش برای بررسی رابطه میان شناخت مذهبی و عوامل زمینه‌ای مانند ملیت، جنسیت، پیشینه تحصیلی، وابستگی مذهبی و زمینه اجتماعی، تحلیل همبستگی انجام دادند. با این حال، نتایج نشان داد که هیچ همبستگی معناداری میان این عوامل و شناخت مذهبی پاسخ‌دهندگان وجود ندارد.

یافته‌های پژوهش نشان داد که متون تولیدشده توسط هوش مصنوعی، در مقایسه با متون انسانی، دارای سوگیری‌های قابل توجهی در احساسات و کلیشه‌ها هستند. برای نمونه، محتوای تولیدشده درباره اسلام ۱.۵ برابر بیشتر به «درگیری» اشاره داشت، در حالی که محتوای مربوط به مسیحیت بیشتر شامل اصطلاحاتی چون «عشق» و «بخشش» بود. این نتایج از فرضیه تأثیرگذاری هوش مصنوعی مولد بر القای سوگیری‌های شناختی و شکل‌دهی نگرش کاربران نسبت به ادیان مختلف پشتیبانی می‌کند.

مقایسه میانگین نمرات ارزیابی ادیان در دو گروه (با و بدون مداخله هوش مصنوعی) به شرح زیر است:

جدول ۱: ارزیابی پاسخ‌دهندگان در حوزه‌های مختلف تحت تأثیر هوش مصنوعی

گروه	گروه‌ها آزمایش	مسیحیت	اسلام	بودیسم	هندوئیسم	یهودیت
تاریخ دین	گروه بدون مداخله	۳.۳۸۴	۳.۲۱۵	۳.۲۴۵	۳.۳۳	۳.۲۹
	گروه با مداخله هوش مصنوعی	۳.۶۹۱	۲.۹۸۲	۳.۴۴	۳.۱۰۸	۳.۴۱۲
سازش اعتقادی	گروه بدون مداخله	۳.۳۶۸	۳.۲۵۶	۳.۳۳	۳.۳۰۶	۳.۲۷۴
	گروه با مداخله هوش مصنوعی	۳.۶۹۵	۳.۰۱۲	۳.۳۱۳	۳.۱۲۲	۳.۴۰۲
اخلاق	گروه بدون مداخله	۳.۴۳۵	۳.۱۸۵	۳.۳۷۸	۳.۱۴۹	۳.۱۹۹
	گروه با مداخله هوش مصنوعی	۳.۷۳۵	۲.۹۸۴	۳.۳۷۱	۳.۱۱۲	۳.۴۲۸
معاشرت و سبک زندگی	گروه بدون مداخله	۳.۳۶۶	۳.۲۷	۳.۳۲۶	۳.۱۶۵	۳.۲۷۸
	گروه با مداخله هوش مصنوعی	۳.۷۲۵	۲.۹۸۶	۳.۲۷۱	۳.۱۳۷	۳.۳۶۷
وفاداری	گروه بدون مداخله	۳.۴۲۱	۳.۲۲۷	۳.۲۸۶	۳.۰۸۷	۳.۲۶۶
	گروه با مداخله هوش مصنوعی	۳.۷۲۳	۲.۹۶۲	۳.۳۰۵	۳.۰۸۶	۳.۴۰۴
مدرنیته	گروه بدون مداخله	۳.۳۸۴	۳.۲۰۵	۳.۳۳۲	۳.۱۷۳	۳.۲۲۹
	گروه با مداخله هوش مصنوعی	۳.۷۷۳	۲.۹۸۴	۳.۳۶۳	۳.۰۹۲	۳.۳۹۶
برداشت کلی	گروه بدون مداخله	۳.۴۲۹	۳.۲۱۳	۳.۳۱۶	۳.۱۱۵	۳.۲۹
	گروه با مداخله هوش مصنوعی	۳.۶۵۵	۳.۰۰۴	۳.۳۷۵	۳.۰۸۴	۳.۴۳۶

این نتایج نشان می‌دهد که محتوای تولیدشده توسط هوش مصنوعی بر ارزیابی پاسخ‌دهندگان از ادیان خاص تأثیرگذار بوده است. به‌ویژه، افزایش ارزیابی مسیحیت و کاهش ارزیابی اسلام، نشان‌دهنده تأثیر بالقوه مغرضانه هوش مصنوعی مولد در انتشار اطلاعات مذهبی است.

تأثیرگذاری هوش مصنوعی بر ارزیابی پاسخ دهندگان

بررسی مطالعات ژانگ و همکاران نشان می‌دهد که مواجهه کاربران با محتوای دینی تولیدشده توسط هوش مصنوعی مولد، موجب تغییر در شناخت و ارزیابی آنان نسبت به ادیان مختلف شده است. به طور خاص، درک کلی از مسیحیت و یهودیت به صورت معناداری مثبت‌تر شد، در حالی که ارزیابی از اسلام کاهش یافت. درک از هندوئیسم اندکی افت داشت و برداشت‌ها از بودیسم نسبتاً متعادل باقی ماند.

این تغییرات را می‌توان به مفاهیم تعریف‌شده در سیستم‌های هوش مصنوعی و تعصبات ذاتی موجود در پایگاه‌های داده آموزشی نسبت داد. در حال حاضر، مدل‌های هوش مصنوعی مولد عمدتاً تحت تأثیر داده‌های اینترنتی منشأ گرفته از ایالات متحده هستند؛ داده‌هایی که به طور طبیعی تمرکز بیشتری بر مسیحیت دارند. شیوع محتوای مغرضانه مذهبی در رسانه‌های آنلاین غربی، به‌ویژه نگرش‌های منفی نسبت به اسلام، احتمالاً در شکل‌گیری این سوگیری‌ها نقش مؤثری ایفا کرده‌اند و بر نحوه آموزش مدل‌های هوش مصنوعی تأثیر گذاشته‌اند.

نتایج آزمون‌های آماری با نمونه‌های مستقل، اهمیت این تأثیر را بیشتر تأیید می‌کند. به طور خاص، در گروه مداخله هوش مصنوعی، نگرش‌های مذهبی پاسخ‌دهندگان به طور معناداری با گروه کنترل تفاوت داشت. این یافته‌ها نشان می‌دهد که هوش مصنوعی مولد می‌تواند تعصبات موجود در انتشار اطلاعات مذهبی را نه تنها بازتاب دهد، بلکه تقویت کند و تأثیر عمیقی بر شناخت مذهبی کاربران داشته باشد.

مطالعه حاضر نشان داد که محتوای تولیدشده توسط هوش مصنوعی، تمایل به تأثیرگذاری بر درک ادیان مختلف دارد؛ امری که می‌تواند پیامدهای عمیقی برای آموزش دینی و شمول فرهنگی به همراه داشته باشد. در محیط‌های آموزشی، مواجهه طولانی‌مدت با محتوای مغرضانه هوش مصنوعی ممکن است منجر به شکل‌گیری برداشت‌های مثبت یا منفی از ادیان خاص شود. این وضعیت می‌تواند تنوع و تعادل در بحث‌های کلاسی را کاهش داده و مانع از درک منصفانه تاریخ و نظام‌های اعتقادی دینی گردد.

برای نمونه، افزایش نمرات شناختی برای مسیحیت و یهودیت ممکن است موجب برجسته‌سازی این ادیان در محیط‌های آموزشی شود، در حالی که کاهش نمرات شناختی برای اسلام می‌تواند کلیشه‌های اجتماعی موجود را تقویت کرده و درک و باور دانش‌آموزان نسبت به فرهنگ و ایمان خود را تضعیف کند.

از آن‌جا که هوش مصنوعی به ابزاری مهم برای کسب اطلاعات تبدیل شده است، این تعصبات می‌توانند بر گفت‌وگوهای بین‌ادیانی تأثیرگذار باشند و در سطحی وسیع‌تر، موجب شکل‌گیری شکاف‌های شناختی میان گروه‌های مختلف اجتماعی شوند.

نقش هوش مصنوعی مولد در شکل‌دهی به آموزش و شناخت دینی

یافته‌های پژوهش نشان می‌دهد که هوش مصنوعی مولد می‌تواند از طریق مواجهه‌های مکرر (Radford et al, 2019) و تقویت اطلاعات (Chen et al, 2021)، تغییرات پایدار و عمیقی در شناخت مذهبی کاربران ایجاد کند. هنگامی که کاربران به‌طور پی‌درپی با محتوای مذهبی خاص تولیدشده توسط هوش مصنوعی مواجه می‌شوند، به تدریج نگرش و برداشت کلی آنان نسبت به آن دین دچار تغییر می‌شود. این تأثیر زمانی برجسته‌تر است که محتوای تولیدشده حاوی سوگیری‌ها یا گرایش‌های خاص باشد؛ زیرا کاربران ممکن است این اطلاعات را به‌صورت ناخودآگاه درونی کرده و آن را به‌عنوان بازنمایی عینی و معتبر از دین تلقی کنند.

مطالعه حاضر نشان داد که شرکت‌کنندگان در گروه مداخله هوش مصنوعی، پس از مواجهه مکرر با محتوای تولیدشده، نگرش منفی‌تری نسبت به برخی ادیان خاص پیدا کرده‌اند. این یافته‌ها حاکی از آن است که تعصبات موجود در محتوای تولیدشده توسط هوش مصنوعی می‌توانند در حافظه بلندمدت کاربران تثبیت شوند و منجر به تغییرات شناختی پایدار گردند. تحلیل داده‌ها نشان داد که شاخص شناختی گروه مداخله نسبت به مسیحیت ۰.۳۱ واحد افزایش و نسبت به اسلام ۰.۲۳ واحد کاهش یافته است؛ تغییری که به‌وضوح بیانگر تأثیر سوگیری‌های الگوریتمی بر نگرش‌های مذهبی کاربران است.

این تغییرات شناختی نه‌تنها در سطح فردی قابل مشاهده‌اند، بلکه از طریق تعاملات اجتماعی می‌توانند به گروه‌های گسترده‌تری نیز منتقل شوند (Floridi, 2020). زمانی که کاربران محتوای

تولیدشده توسط هوش مصنوعی را معتبر یا دقیق تلقی کنند، احتمال دارد این دیدگاه‌ها را در ارتباطات روزمره خود بازتاب دهند و بدین ترتیب، تغییرات شناختی فردی به سطح اجتماعی گسترش یابد.

این فرآیند شکل‌دهی شناختی بلندمدت، به‌ویژه در حوزه‌های حساسی مانند دین، می‌تواند پیامدهای فرهنگی و اجتماعی عمیقی به‌همراه داشته باشد. از جمله این پیامدها می‌توان به تقویت کلیشه‌های مذهبی، کاهش تنوع درک دینی در محیط‌های آموزشی و شکل‌گیری شکاف‌های شناختی میان گروه‌های مذهبی اشاره کرد. بنابراین، ضرورت دارد که در طراحی و آموزش مدل‌های هوش مصنوعی، حساسیت‌های دینی و فرهنگی به‌صورت جدی مورد توجه قرار گیرد تا از بروز سوگیری‌های نهادینه‌شده جلوگیری شود.

نقش داده‌ها در تولید محتوای مذهبی

ورودی داده‌های آموزشی در مدل‌های هوش مصنوعی مولد، نقشی تعیین‌کننده در شکل‌دهی سوگیری‌ها و میزان دقت محتوای تولیدشده ایفا می‌کند. هسته اصلی الگوریتم‌های مولد، تولید محتوایی است که با یادگیری از مجموعه داده‌های گسترده، با زمینه‌ها و نیازهای خاص کاربران هم‌راستا باشد. با این حال، این فرآیند به‌طور کامل خنثی نیست و تحت تأثیر کیفیت و جهت‌گیری داده‌های ورودی قرار دارد (Zhou et al, 2024).

مطالعات نشان داده‌اند که اگر داده‌های آموزشی حاوی سوگیری‌ها یا گرایش‌های خاص باشند، این سوگیری‌ها در ساختار الگوریتم درونی می‌شوند و در فرآیند تولید محتوا بازتاب می‌یابند (Noble, 2018). برای نمونه، اگر داده‌های آموزشی بر رویدادهای منفی یا کلیشه‌های فرهنگی مرتبط با یک دین خاص تمرکز داشته باشند، محتوای تولیدشده توسط هوش مصنوعی ممکن است به‌طور نامتناسبی این جنبه‌های منفی را برجسته سازد و در نتیجه، برداشت کاربران از آن دین را تحت تأثیر قرار دهد.

تحلیل‌های تجربی نشان می‌دهد که هوش مصنوعی مولد در مواجهه با محتوای مذهبی، تمایل به انعکاس سوگیری‌های موجود در داده‌های آموزشی دارد (Tsuria, 2024) برای مثال، در مطالعه‌ای، پاسخ‌دهندگان در گروه مداخله هوش مصنوعی، دین اسلام را به‌طور معناداری پایین‌تر

از گروه کنترل ارزیابی کردند؛ یافته‌ای که نشان می‌دهد چگونه سوگیری‌های الگوریتمی، تحت تأثیر داده‌های مغرضانه، می‌توانند به کاربران منتقل شوند (Alkhouri, 2024). نقش داده‌های آموزشی در این فرآیند، عمیق و چندلایه است. این داده‌ها نه تنها کیفیت و جهت‌گیری محتوای تولیدشده را تعیین می‌کنند، بلکه به صورت نامحسوس بر شناخت مذهبی کاربران تأثیر می‌گذارند. در نتیجه، طراحی و انتخاب داده‌های آموزشی برای مدل‌های هوش مصنوعی مولد، باید با حساسیت بالا و نظارت دقیق انجام شود تا از بازتولید تعصبات و تحریف‌های شناختی در حوزه‌های حساس مانند دین جلوگیری گردد.

خطرات و فرصت‌های حلقه بازخورد

هوش مصنوعی مولد نه تنها در مرحله تولید اولیه محتوا، بلکه از طریق تعاملات مکرر با کاربران و بهره‌گیری از مکانیسم بازخورد، محتوای خود را تنظیم و بهینه‌سازی می‌کند. این فرآیند پویا را می‌توان به عنوان یک مکانیسم گردش دوطرفه خودآموزی و تنظیم شناختی مفهوم‌سازی کرد. در این مکانیسم، بازخورد کاربران نقش محوری در خودتنظیمی هوش مصنوعی ایفا می‌کند؛ به گونه‌ای که پس از تعامل با محتوای مذهبی تولیدشده، کاربران بر اساس درک و احساسات شخصی خود بازخورد ارائه می‌دهند و هوش مصنوعی نیز با تحلیل این بازخوردها، محتوای بعدی را با انتظارات و گرایش‌های آنان هم‌راستا می‌سازد (Harari, 2023).

با این حال، این مکانیسم می‌تواند پیامدهای نگران‌کننده‌ای نیز به همراه داشته باشد. به ویژه زمانی که کاربران دارای نگرش‌های منفی نسبت به یک دین خاص باشند، هوش مصنوعی ممکن است به طور ناخودآگاه محتوایی تولید کند که با این تعصب هم‌سو تر باشد و در نتیجه، یک چرخه خودتقویت‌کننده از سوگیری مذهبی شکل گیرد (Harari, 2023, Dataism and AI).

یافته‌های این پژوهش نشان می‌دهد که کاربران در گروه مداخله هوش مصنوعی، پس از تعاملات مکرر با محتوای تولیدشده، نگرش‌های منفی‌تری نسبت به برخی ادیان خاص پیدا کرده‌اند. این تغییرات شناختی، حاصل مکانیسم بازخوردی هوش مصنوعی است که برداشت‌های منفی را تقویت کرده و به تدریج در حافظه شناختی کاربران تثبیت می‌کند.

این حلقه بازخورد می‌تواند منجر به شکل‌گیری پدیده‌ای موسوم به «پبله اطلاعات» شود؛ وضعیتی که در آن کاربران صرفاً در معرض اطلاعاتی قرار می‌گیرند که با باورهای پیشین آنان هم‌راستا است و در نتیجه، تعصبات مذهبی آنان بیشتر تقویت می‌شود (Lin, 2022). این پدیده نه تنها تنوع شناختی را محدود می‌سازد، بلکه مانع از مواجهه با دیدگاه‌های جایگزین و درک منصفانه از نظام‌های اعتقادی مختلف می‌شود.

در حوزه آموزش دینی، چنین مکانیسمی می‌تواند پیامدهای عمیقی داشته باشد؛ از جمله کاهش تنوع درک دینی، تقویت کلیشه‌های مذهبی و شکل‌گیری شکاف‌های شناختی میان گروه‌های مختلف. بنابراین، طراحی مکانیسم‌های بازخورد در هوش مصنوعی باید با حساسیت بالا و نظارت اخلاقی همراه باشد تا از تقویت تعصبات و تحریف‌های شناختی جلوگیری شود.

پیامدهایی برای آموزش دینی و ملاحظات اخلاقی

هوش مصنوعی مولد می‌تواند به‌عنوان ابزاری مؤثر برای تولید محتوای آموزشی دینی شخصی‌سازی‌شده مورد استفاده قرار گیرد (Pringle & Smith, 2023). با این حال، سوءبرداشت‌ها از تاریخ و آموزه‌های دینی، در صورت انتقال نادرست یا جانبدارانه، ممکن است به درک سطحی و تحریف‌شده از دین منجر شود (Hoffman, 2023). از این‌رو، اطمینان از انتقال دقیق و بی‌طرفانه مفاهیم دینی توسط هوش مصنوعی، یک ملاحظه حیاتی برای مربیان و سیاست‌گذاران آموزشی است (Walters, 2023).

شناسایی و کاهش سوگیری‌ها در طراحی الگوریتم‌ها و انتخاب داده‌های آموزشی، گامی کلیدی در جهت تضمین بی‌طرفی و اعتبار محتوای تولیدشده است. در صورت تحقق این امر، هوش مصنوعی می‌تواند به‌جای تقویت تعصب، به ترویج سازگاری، مدارا و تفکر انتقادی در آموزش دینی کمک کند.

سیستم‌های هوش مصنوعی قابل اعتماد باید به اصول و ارزش‌های اخلاقی پایبند باشند (Chrostowski & Najda, 2025) توسعه‌دهندگان این فناوری باید فرآیندهای تولید محتوا را شفاف‌سازی کرده و امکان حسابرسی و پاسخ‌گویی خارجی را فراهم سازند (Veale & Binns, 2017). گنجاندن مجموعه داده‌های متنوع و چندفرهنگی و تقویت همکاری میان متخصصان فناوری و پژوهشگران دینی، می‌تواند شمول و انصاف محتوای تولیدشده را افزایش دهد.

افزون بر این، ارتقای آگاهی کاربران نسبت به سوگیری‌های احتمالی در محتوای مذهبی تولیدشده توسط هوش مصنوعی، آنان را قادر می‌سازد تا با دیدی انتقادی و تحلیلی با مطالب تعامل داشته باشند و از خطرات تقویت سوگیری شناختی جلوگیری کنند.

در مجموع، اگرچه هوش مصنوعی مولد با چالش‌هایی در زمینه مدیریت تعصبات و ملاحظات اخلاقی مواجه است، اما ظرفیت آن برای تحول در آموزش و ارتباطات دینی، غیرقابل انکار است. با پرداختن مسئولانه به این چالش‌ها، می‌توان از قابلیت‌های این فناوری برای تقویت درک، تحمل و مشارکت فعال‌تر در گفتمان دینی بهره‌برداری کرد.

نتیجه‌گیری

استفاده از فناوری‌های نوین، به‌ویژه هوش مصنوعی مولد، در زیست‌بوم آموزشی و دینی امری اجتناب‌ناپذیر است. از این‌رو، آسیب‌شناسی و بهینه‌سازی هوشمندانه این نوآوری‌ها، وظیفه‌ای خطیر بر دوش متفکران، مربیان و سیاست‌گذاران دینی است. همان‌گونه که رفیعی (۱۴۰۳) تأکید می‌کند، پیش از آن‌که این فناوری در میان عموم مردم فراگیر شود، شناسایی آسیب‌ها و آموزش نحوه استفاده صحیح از آن، ضرورتی انکارناپذیر است.

این پژوهش نشان داد که هوش مصنوعی مولد، به‌واسطه سوگیری‌های پیش‌فرض در داده‌های آموزشی و مکانیسم بازخورد دوطرفه، می‌تواند بر شناخت مذهبی کاربران تأثیرگذار باشد. این تأثیر، به‌ویژه در حوزه اسلام و به‌طور خاص مذهب شیعه، به‌دلیل ضعف داده‌های ورودی و غلبه محتوای مغرضانه در منابع آنلاین غربی، نگران‌کننده‌تر است. یافته‌ها نشان داد که شاخص شناختی گروه مسلمان در مواجهه با محتوای تولیدشده توسط هوش مصنوعی کاهش یافته است؛ امری که ضرورت نظارت دقیق بر ورودی داده‌ها و طراحی الگوریتم‌های حساس به تنوع مذهبی را برجسته می‌سازد.

در این راستا، دو راهبرد اساسی پیشنهاد می‌شود:

۱. **نظارت فعال و مشارکت علمی در طراحی هوش مصنوعی:** دانشمندان دینی، به‌ویژه اندیشمندان مسلمان، باید در فرآیند طراحی، آموزش و ارزیابی مدل‌های هوش مصنوعی مشارکت فعال داشته باشند. تشکیل تیم‌های توسعه بومی با رویکرد اسلامی، تدوین الگوریتم‌های متناسب با آموزه‌های دینی و تأمین داده‌های جامع، دقیق و بی‌طرف، از جمله اقدامات ضروری برای کاهش سوگیری و ارتقای کیفیت محتوای دینی تولیدشده است.

۲. تقویت سواد رسانه‌ای و نگاه انتقادی در میان کاربران و مربیان: آموزش نگاه انتقادی به آموزگاران، به‌ویژه در حوزه آموزش دینی، اهمیت دوچندانی دارد. اعتماد بی‌چون‌وچرا به محتوای تولیدشده توسط هوش مصنوعی، می‌تواند منجر به انتقال سوگیری‌های شناختی به دانش‌آموزان و در نهایت به جامعه شود. کاربران باید آگاه باشند که هوش مصنوعی، حتی در نبود داده‌های معتبر، تمایل به پاسخ‌گویی دارد؛ امری که می‌تواند به تولید اطلاعات نادرست منجر شود. بنابراین، مطالبه‌گری آگاهانه و بازخوردهای دقیق کاربران، نقش مهمی در بهبود عملکرد هوش مصنوعی ایفا می‌کند.

این تحقیق بر اهمیت همکاری میان‌رشته‌ای میان توسعه‌دهندگان فناوری، مربیان و محققان دینی تأکید دارد. چنین همکاری‌هایی می‌تواند به طراحی سیستم‌هایی منجر شود که نه تنها از نظر فنی کارآمد، بلکه از نظر اخلاقی و فرهنگی نیز مسئولانه عمل کنند. هوش مصنوعی نباید به‌عنوان جایگزین مربیان تلقی شود، بلکه باید به‌عنوان ابزاری مکمل برای ارتقای تفکر انتقادی، خلاقیت و درک بین‌فرهنگی در آموزش دینی به‌کار گرفته شود (Chrostowski & Najda, 2025; Toevs, 2023).

در نهایت، پیشنهاد می‌شود که پژوهش‌های آینده به موارد زیر توجه کنند:

- بررسی راهکارهای فنی برای کاهش سوگیری و افزایش انصاف در مدل‌های زبانی
- تحلیل اثربخشی آموزش دینی شخصی‌سازی‌شده با حفظ بی‌طرفی
- انجام مطالعات طولی برای ارزیابی تأثیرات بلندمدت هوش مصنوعی بر شناخت دینی
- بررسی نقش روایت‌های متنوع و یادگیری تعاملی در ارتقای گفت‌وگوی بین‌ادیانی

با پرداختن به این ابعاد، می‌توان از ظرفیت‌های هوش مصنوعی مولد برای تقویت آموزش دینی فراگیر، عادلانه و مبتنی بر احترام متقابل بهره‌برداری کرد.

فهرست منابع

- بصام، سعید (۱۴۰۲)، تأثیرات هوش مصنوعی بر آموزش دینی، بازیابی شده از وبسایت دین آنلاین (www.dinonline.com).
- رفیعی طرقي، فاطمه؛ فولادی، ساجده؛ فهرستی اصل، علیرضا؛ رمضان، شبنم؛ و صالحی کیا، فاطمه. (۱۴۰۳). تطبیق مفاهیم دینی با تکنولوژی هوش مصنوعی: رویکردهای معاصر. در مجموعه مقالات چهاردهمین کنفرانس ملی پژوهش‌های نوین در تعلیم و تربیت، روان‌شناسی، فقه و حقوق و علوم اجتماعی، شیروان: دبیرخانه کنفرانس.
- Abid, A, Farooqi, M, & Zou, J. (2021). Persistent anti-Muslim bias in large language models. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 298–306).
- Alkhouri, K. I. (2024). The role of artificial intelligence in the study of the psychology of religion. Religions, 15(3), 290. <https://doi.org>.
- Amershi, S, Weld, D, Vorvoreanu, M, Fourney, A, Nushi, B, Collisson, P, et al. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–13). ACM.
- Arias-Arévalo, P, Martín-López, B, & Gómez-Baggethun, E. (2023). The role of power in leveraging the diverse values of nature for transformative change. Current Opinion in Environmental Sustainability, 64, 101352.
- Ashraf, C. (2021). Exploring the impacts of artificial intelligence on freedom of religion or belief online. International Journal of Human Rights, 26(5), 757–791. <https://doi.org/10.1080/13642987.2021.1968376>
- Bakshy, E, Messing, S, & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- Binns, R, Veale, M, Van Kleek, M, & Shadbolt, N. (2018). It's reducing a human being to a percentage: Perceptions of justice in algorithmic decisions. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, Paper 377, 1–14.

- Braun, M, Bleher, H, & Hummel, P. (2021). A leap of faith: Is there a formula for trustworthy AI? Hastings Center Report, 51(3), 17–22. <https://doi.org/10.1002/hast.1207>
- Brown, T. B, Mann, B, Ryder, N, Subbiah, M, Kaplan, J, Dhariwal, P, ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv: 2005.14165.
- Browne, R. (2024). The ethical implications of AI in global religions. Journal of Religious Ethics, 45(2), 245–260. <https://doi.org/10.1007/s10712-024-93823-8>
- Buolamwini, J, & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability, and Transparency (pp. 77–91). PMLR.
- Chen, T, Kornblith, S, Norouzi, M, & Hinton, G. (2021). A simple framework for contrastive learning of visual representations. In International Conference on Machine Learning (pp. 1597–1607). <https://doi.org/10.48550/arXiv.2002.05709>
- Chrostowski, M. (2024). Promoting critical thinking through the use of ChatGPT in religious education. Österreichisches Religionspädagogisches Forum, 32(2), 166–181. <https://doi.org/10.25364/10.32:2024.2.10>
- Chrostowski, M, & Najda, A. J. (2025). ChatGPT as a modern tool for Bible teaching in confessional religious education: A German view. Journal of Religious Education, 73, 75–94. <https://doi.org/10.1007/s40839-024-00244-1>
- Citlak, A. (2021). Psychology of religion in the theories and research of the Lvov-Warsaw school: Basic achievements and developments. Archive for the Psychology of Religion, 43, 95–116.
- Citlak, K. (2021). A computational approach to understanding religious texts using deep learning models. Journal of Artificial Intelligence Research.
- Devlin, J, Chang, M. W, Lee, K, & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805. <https://doi.org/10.48550/arXiv.1810.04805>
- Floridi, L. (2020). Ethical challenges in artificial intelligence: Towards a comprehensive strategy. Journal of Artificial Intelligence Research, 69, 117–139.

- Floridi, L, Cows, J, Beltrametti, M, Chatila, R, Chazerand, P, Dignum, V, ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Harari, Y. N. (2023). Dataism and AI: How artificial intelligence shapes the future of religion. *New Age Religion Journal*, 5(3), 110–126.
- Harari, Y. N. (2023). Will AI create a religion? Views of the algorithmic forms of the religious life in popular discourse. Project MUSE. <https://doi.org/10.1353/project.muse.2023>
- He, Y. (2024). Artificial intelligence and socioeconomic forces: Transforming the landscape of religion. *Humanities and Social Sciences Communications*, 11, 602. <https://doi.org/10.1057/s41599-024-03137-8>
- Hernawati, S, Hafizh, M, & Rahardja, M. N. A. (2024). Adjusting the ideal Islamic religious education curriculum to the development of AI-based technology. *Progresiva: Jurnal Pemikiran dan Pendidikan Islam*, 13(1), 129–144.
- Ho, M. T, & Mantello, P. (2023). An analytical framework for studying attitude towards emotional AI: The three-pronged approach. *MethodsX*, 10, 102149.
- Ho, M. T, Mantello, P, & Vuong, Q. H. (2024). Emotional AI in education and toys: Investigating moral risk awareness in the acceptance of AI technologies from a cross-sectional survey of the Japanese population. *Heliyon*, 10(16). <https://doi.org/10.1016/j.heliyon.2024.e36251>
- Hoffman, D. (2023). Ethics of AI: Considerations in the field of religious practices. *AI & Society*, 32, 123–135.
- Hoffman, R. R, Miller, T, Mueller, S. T, Klein, G, & Litman, D. J. (2020). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv: 1812.04608*. <https://doi.org/10.48550/arXiv.1812.04608>
- Holstein, K, Wortman Vaughan, J, Daumé, H. III, Dudik, M, & Wallach, H. (2020). Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. <https://doi.org/10.1145/3313831.3376446>
- Jain, L. R, & Menon, V. (2023). AI algorithmic bias: Understanding its causes, ethical and social implications. In *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 460–467).

- Jin, X. (2018). Internet, artificial intelligence and new religion. *World Religious Cult*, 1, 82–87.
- Liang, W. (2020). Believers' network experience and the changing governance of authority: A summary of foreign and domestic studies on internet plus religion. *World Religious Studies*, 3, 183–188.
- Lin, D. (2022). Characteristics and management of internet religious information from the perspective of digital technology. *World Religious Culture*, 6, 54–61.
- Mantello, P, Ho, M. T, Nguyen, M. H, & Vuong, Q. H. (2023). Machines that feel: Behavioral determinants of attitude towards affect recognition technology—Upgrading technology acceptance theory with the mindsponge model. *Humanities and Social Sciences Communications*, 10(1), 1–16.
- Muralidhar, D. (2021). Examining religion bias in AI text generators. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)* (pp. 273–274). Association for Computing Machinery.
- Nam, S. (2023). Implications for the direction of Christian education in the age of artificial intelligence. *Christian Education Journal*, 74, 107–134.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. <https://doi.org/10.18574/nyu/9781479837243.001.0001>
- Pringle, T, & Smith, J. (2024). Algorithmic religious life: AI's role in creating new spiritual practices. *Journal of Technology and Religion*, 3(1), 78–89.
- Puzio, A. (2023). Robot, let us pray! Can and should robots have religious functions? An ethical exploration of religious robots. *AI & Society*, 1–17.
- Radford, A, Wu, J, Amodei, D, & Clark, J. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.
- Roll, I, & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. <https://doi.org/10.1007/s40593-016-0108-5>
- Salsabila, A. Z. A, & Rohiem, A. F. (2024). The ethical influence of artificial intelligence (AI) in religious education: Implications, challenges, and

- innovative perspectives on the Merdeka curriculum. In Proceedings of the International Conference on Religion, Science and Education, 3, 81–88.
- Segun, S. T. (2021). Critically engaging the ethics of AI for a global audience. *Ethics and Information Technology*, 23(2), 99–105. <https://doi.org/10.1007/s10676-020-09570-y>
- Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. *British Journal of Educational Technology*, 50(6), 1593–1606. <https://doi.org/10.1111/bjet.12842>
- Sianipar, D, & Telaumbanua, S. (2024). Implementation of Christian religious education learning by using the outcome-based education (OBE) approach in the era of artificial intelligence.
- Smith, J. A, & Rahman, T. (2020). AI in religious studies: A critical evaluation. *Journal of Artificial Intelligence Research*, 69, 117–139. <https://doi.org/10.1613/jair.1.11904>
- Thorne, J, Vlachos, A, Christodoulopoulos, C, & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long Papers) (pp. 809–819). ACL.
- Toevis, J. (2023). The role of AI in enhancing Christian education. Retrieved January 17, 2024.
- Tsuria, R. (2024). Artificial intelligence's understanding of religion: Investigating the moralistic approaches presented by generative artificial intelligence tools. *Religions*, 15(3), 375. <https://doi.org/10.3390/rel15030375>
- Varona, D, & Suárez, J. L. (2022). Discrimination, bias, fairness, and trustworthy AI. *Applied Sciences*, 12(12), 5826. <https://doi.org/10.3390/app12125826>
- Vaswani, A, Shazeer, N, Parmar, N, Uszkoreit, J, Jones, L, Gomez, A. N, ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Veale, M, & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 2053951717743530. <https://doi.org/10.1177/2053951717743530>

- Walters, H. (2023). Hindu and Buddhist perspectives on robot-assisted religious rituals. *Journal of Eastern Religious Practices*, 12(4), 345–360.
- Williamson, B. (2016). Political computational thinking: Policy networks, digital governance and ‘learning to code’. *Critical Policy Studies*, 10(1), 39–58. <https://doi.org/10.1080/19460171.2016.1160936>
- Zhang, J, Song, W, & Liu, Y. (2025). Cognitive bias in generative AI influences religious education. *Scientific Reports*, 15, 15720.
- Zhou, M, Abhishek, V, Derdenger, T, Kim, J, & Srinivasan, K. (2024). Bias in generative AI. *arXiv preprint*. <https://arxiv.org/abs/2401.00000>.

References

- Abid, A, Farooqi, M, & Zou, J. (2021). Persistent anti-Muslim bias in large language models. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 298–306).
- Alkhouri, K. I. (2024). The role of artificial intelligence in the study of the psychology of religion. *Religions*, 15(3), 290. <https://doi.org>.
- Amershi, S, Weld, D, Vorvoreanu, M, Fourney, A, Nushi, B, Collisson, P, et al. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–13). ACM.
- Arias-Arévalo, P, Martín-López, B, & Gómez-Baggethun, E. (2023). The role of power in leveraging the diverse values of nature for transformative change. *Current Opinion in Environmental Sustainability*, 64, 101352.
- Ashraf, C. (2021). Exploring the impacts of artificial intelligence on freedom of religion or belief online. *International Journal of Human Rights*, 26(5), 757–791. <https://doi.org/10.1080/13642987.2021.1968376>
- Bassam, S. (2023). The impacts of artificial intelligence on religious education. Retrieved from Din Online website (www.dinonline.com). [in persian]
- Bakshy, E, Messing, S, & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aal1160>
- Binns, R, Veale, M, Van Kleek, M, & Shadbolt, N. (2018). It's reducing a human being to a percentage: Perceptions of justice in algorithmic decisions. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, Paper 377, 1–14.
- Braun, M, Bleher, H, & Hummel, P. (2021). A leap of faith: Is there a formula for trustworthy AI? *Hastings Center Report*, 51(3), 17–22. <https://doi.org/10.1002/hast.1207>
- Brown, T. B, Mann, B, Ryder, N, Subbiah, M, Kaplan, J, Dhariwal, P, ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv: 2005.14165*.
- Browne, R. (2024). The ethical implications of AI in global religions. *Journal of Religious Ethics*, 45(2), 245–260. <https://doi.org/10.1007/s10712-024-93823-8>
- Buolamwini, J, & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability, and Transparency (pp. 77–91). PMLR.
- Chen, T, Kornblith, S, Norouzi, M, & Hinton, G. (2021). A simple framework for contrastive learning of visual representations. In

- International Conference on Machine Learning (pp. 1597–1607). <https://doi.org/10.48550/arXiv.2002.05709>
- Chrostowski, M. (2024). Promoting critical thinking through the use of ChatGPT in religious education. *Österreichisches Religionspädagogisches Forum*, 32(2), 166–181. <https://doi.org/10.25364/10.32: 2024.2.10>
- Chrostowski, M., & Najda, A. J. (2025). ChatGPT as a modern tool for Bible teaching in confessional religious education: A German view. *Journal of Religious Education*, 73, 75–94. <https://doi.org/10.1007/s40839-024-00244-1>
- Citlak, A. (2021). Psychology of religion in the theories and research of the Lvov-Warsaw school: Basic achievements and developments. *Archive for the Psychology of Religion*, 43, 95–116.
- Citlak, K. (2021). A computational approach to understanding religious texts using deep learning models. *Journal of Artificial Intelligence Research*.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv: 1810.04805*. <https://doi.org/10.48550/arXiv.1810.04805>
- Floridi, L. (2020). Ethical challenges in artificial intelligence: Towards a comprehensive strategy. *Journal of Artificial Intelligence Research*, 69, 117–139.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Harari, Y. N. (2023). Dataism and AI: How artificial intelligence shapes the future of religion. *New Age Religion Journal*, 5(3), 110–126.
- Harari, Y. N. (2023). Will AI create a religion? Views of the algorithmic forms of the religious life in popular discourse. *Project MUSE*. <https://doi.org/10.1353/project.muse.2023>
- He, Y. (2024). Artificial intelligence and socioeconomic forces: Transforming the landscape of religion. *Humanities and Social Sciences Communications*, 11, 602. <https://doi.org/10.1057/s41599-024-03137-8>
- Hernawati, S., Hafizh, M., & Rahardja, M. N. A. (2024). Adjusting the ideal Islamic religious education curriculum to the development of AI-based technology. *Progresiva: Jurnal Pemikiran dan Pendidikan Islam*, 13(1), 129–144.
- Ho, M. T., & Mantello, P. (2023). An analytical framework for studying attitude towards emotional AI: The three-pronged approach. *MethodsX*, 10, 102149.

- Ho, M. T, Mantello, P, & Vuong, Q. H. (2024). Emotional AI in education and toys: Investigating moral risk awareness in the acceptance of AI technologies from a cross-sectional survey of the Japanese population. *Heliyon*, 10(16). <https://doi.org/10.1016/j.heliyon.2024.e36251>
- Hoffman, D. (2023). Ethics of AI: Considerations in the field of religious practices. *AI & Society*, 32, 123–135.
- Hoffman, R. R, Miller, T, Mueller, S. T, Klein, G, & Litman, D. J. (2020). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv: 1812.04608*. <https://doi.org/10.48550/arXiv.1812.04608>
- Holstein, K, Wortman Vaughan, J, Daumé, H. III, Dudik, M, & Wallach, H. (2020). Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. <https://doi.org/10.1145/3313831.3376446>
- Jain, L. R, & Menon, V. (2023). AI algorithmic bias: Understanding its causes, ethical and social implications. In *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 460–467).
- Jin, X. (2018). Internet, artificial intelligence and new religion. *World Religious Cult*, 1, 82–87.
- Liang, W. (2020). Believers' network experience and the changing governance of authority: A summary of foreign and domestic studies on internet plus religion. *World Religious Studies*, 3, 183–188.
- Lin, D. (2022). Characteristics and management of internet religious information from the perspective of digital technology. *World Religious Culture*, 6, 54–61.
- Mantello, P, Ho, M. T, Nguyen, M. H, & Vuong, Q. H. (2023). Machines that feel: Behavioral determinants of attitude towards affect recognition technology—Upgrading technology acceptance theory with the mindsponge model. *Humanities and Social Sciences Communications*, 10(1), 1–16.
- Muralidhar, D. (2021). Examining religion bias in AI text generators. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)* (pp. 273–274). Association for Computing Machinery.
- Nam, S. (2023). Implications for the direction of Christian education in the age of artificial intelligence. *Christian Education Journal*, 74, 107–134.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. <https://doi.org/10.18574/nyu/9781479837243.001.0001>

- Pringle, T., & Smith, J. (2024). Algorithmic religious life: AI's role in creating new spiritual practices. *Journal of Technology and Religion*, 3(1), 78–89.
- Puzio, A. (2023). Robot, let us pray! Can and should robots have religious functions? An ethical exploration of religious robots. *AI & Society*, 1–17.
- Rafiei Toroghi, F., Foladi, S., Fehresti-Asl, A., Ramezani, S., & Salehi-Kia, F. (2024). Aligning religious concepts with artificial intelligence technology: Contemporary approaches. In *Proceedings of the Fourteenth National Conference on Modern Research in Education, Psychology, Jurisprudence, Law, and Social Sciences*. Shirvan: Conference Secretariat. [in persian]
- Radford, A., Wu, J., Amodei, D., & Clark, J. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.
- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. <https://doi.org/10.1007/s40593-016-0108-5>
- Salsabila, A. Z. A., & Rohiem, A. F. (2024). The ethical influence of artificial intelligence (AI) in religious education: Implications, challenges, and innovative perspectives on the Merdeka curriculum. In *Proceedings of the International Conference on Religion, Science and Education*, 3, 81–88.
- Segun, S. T. (2021). Critically engaging the ethics of AI for a global audience. *Ethics and Information Technology*, 23(2), 99–105. <https://doi.org/10.1007/s10676-020-09570-y>
- Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. *British Journal of Educational Technology*, 50(6), 1593–1606. <https://doi.org/10.1111/bjet.12842>
- Sianipar, D., & Telaumbanua, S. (2024). Implementation of Christian religious education learning by using the outcome-based education (OBE) approach in the era of artificial intelligence.
- Smith, J. A., & Rahman, T. (2020). AI in religious studies: A critical evaluation. *Journal of Artificial Intelligence Research*, 69, 117–139. <https://doi.org/10.1613/jair.1.11904>
- Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (Long Papers) (pp. 809–819). ACL.
- Toevs, J. (2023). The role of AI in enhancing Christian education. Retrieved January 17, 2024.
- Tsuria, R. (2024). Artificial intelligence's understanding of religion: Investigating the moralistic approaches presented by generative

- artificial intelligence tools. *Religions*, 15(3), 375. <https://doi.org/10.3390/rel15030375>
- Varona, D, & Suárez, J. L. (2022). Discrimination, bias, fairness, and trustworthy AI. *Applied Sciences*, 12(12), 5826. <https://doi.org/10.3390/app12125826>
- Vaswani, A, Shazeer, N, Parmar, N, Uszkoreit, J, Jones, L, Gomez, A. N, ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Veale, M, & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 2053951717743530. <https://doi.org/10.1177/2053951717743530>
- Walters, H. (2023). Hindu and Buddhist perspectives on robot-assisted religious rituals. *Journal of Eastern Religious Practices*, 12(4), 345–360.
- Williamson, B. (2016). Political computational thinking: Policy networks, digital governance and ‘learning to code’. *Critical Policy Studies*, 10(1), 39–58. <https://doi.org/10.1080/19460171.2016.1160936>
- Zhang, J, Song, W, & Liu, Y. (2025). Cognitive bias in generative AI influences religious education. *Scientific Reports*, 15, 15720.
- Zhou, M, Abhishek, V, Derdenger, T, Kim, J, & Srinivasan, K. (2024). Bias in generative AI. *arXiv preprint*. <https://arxiv.org/abs/2401.00000>.