



Convergence of Artificial Intelligence, the Semantic Web, and Large Language Models in Knowledge Management Systems for Islamic Texts: A Critical Review from a System Engineering Perspective*

Ali Mirarab 

Assistant Professor, Department of Information and Knowledge Dissemination,
Research Institute for Information Organization and Islamic Documents Management,
Islamic Sciences and Culture Research Institute, Qom, Iran.
alimirarab@isca.ac.ir

Abstract

Knowledge management of Islamic texts has long faced challenges in organizing, retrieving, and analyzing large volumes of complex textual data. The emergence and maturation of artificial intelligence technologies, particularly the semantic web, knowledge graphs, and large language models, have opened new horizons in this field. This article provides an analytical and critical review of the convergence of these three technological streams in knowledge management systems for Islamic texts over the period 2023–2026. The primary objective is not only to identify and categorize recent advances but also to conduct an in-depth engineering assessment of these developments in terms of system architecture, scalability, semantic accuracy, reliability, and

* Mirarab, A. (2025). Convergence of artificial intelligence, the Semantic Web, and large language models in knowledge management systems for Islamic texts: A critical review from a systems engineering perspective. *Islamic Knowledge Management*, 7(2), pp. 128-165.
<https://doi.org/10.22081/jikm.2026.74696.1110>

▣ **Article Type:** Research; **Publisher:** Islamic Sciences and Culture Academy, Qom, Iran

▣ **Received:** 2025/04/28 • **Revised:** 2025/07/18 • **Accepted:** 2025/08/01 • **Published online:** 2025/10/01

© 2025

authors retain the copyright and full publishing rights



compliance with domain requirements. Through a systematic analysis of the research literature—comprising more than 70 selected articles from an initial title of 200 publications—a novel conceptual framework is proposed for analyzing intelligent knowledge management systems for Islamic texts, incorporating technical, methodological, and domain-specific considerations. The findings indicate that, while technologies such as semantic knowledge graphs and large language models have demonstrated considerable capabilities in organizing, retrieving, and generating knowledge from Islamic texts, fundamental challenges—including hallucinations in language models, fragmented ontologies, the scarcity of high-quality annotated data, the complexities of source validation, and the need for model interpretability—remain major barriers to achieving fully reliable and high-performance systems. By proposing a six-dimensional research roadmap, the article identifies future directions for developing intelligent knowledge management systems for Islamic texts that can respond effectively and responsibly to the needs of Islamic textual analysis and processing.

Keywords

Knowledge Management of Islamic Texts, Artificial Intelligence, Semantic Web, Knowledge Graphs, Large Language Models, Critical Review, Systems Engineering, Natural Language Processing.



تکامل الذكاء الاصطناعي، والويب الدلالي، ونماذج اللغة الكبيرة في نظم إدارة معرفة النصوص الإسلامية: مراجعة نقدية من منظور هندسة النظم*

علي ميرعرب^{id}

أستاذ مساعد، قسم إتاحة ونشر المعلومات والمعرفة، معهد إدارة المعلومات والوثائق الإسلامية،
المعهد العالي للعلوم والثقافة الإسلامية، قم، إيران.
alimirarab@isca.ac.ir

الملخص

تواجه إدارة المعرفة في مجال النصوص الإسلامية تحدياتٍ طويلة الأمد في تنظيم واسترجاع وتحليل البيانات النصية الكبيرة والمعقدة. وقد فتح ظهور تقنيات الذكاء الاصطناعي وتطورها - لا سيما الويب الدلالي، والرسوم البيانية المعرفية، والنماذج اللغوية واسعة النطاق - آفاقاً جديدة في هذا المجال. تقدم هذه الدراسة مراجعة تحليلية نقدية لتقارب هذه المسارات التقنية الثلاثة في نظم إدارة معرفة النصوص الإسلامية خلال الفترة الممتدة من ٢٠٢٣ إلى ٢٠٢٦ م. والهدف الرئيسي ليس فقط تحديد

* الاستشهاد بهذه المقالة: ميرعرب، علي. (١٤٠٤). تكامل الذكاء الاصطناعي، والويب الدلالي، ونماذج اللغة الكبيرة في نظم إدارة معرفة النصوص الإسلامية: مراجعة نقدية من منظور هندسة النظم. إدارة المعرفة الإسلامية، ٧(٢)، ص ١٢٨-١٦٥.

<https://doi.org/10.22081/jikm.2026.74696.1110>

□ نوع المقال: بحثية؛ الناشر: المعهد العالي للعلوم والثقافة الإسلامية، قم، إيران

□ تاريخ الإستلام: ٢٠٢٥/٠١/٢٥ • تاريخ التعديل: ٢٠٢٥/٠٧/٣٠ • تاريخ القبول: ٢٠٢٥/٠٨/٢٥ • تاريخ الإصدار: ٢٠٢٥/١٠/٠١

© ٢٠٢٥

يحتفظ المؤلفون بحقوق الطبع والنشر الكاملة.



وتصنيف التطورات الحديثة، ولكن أيضاً تقييمها وهندستها المتعمقة من منظور معمارية النظام، وقابلية التوسع، والدقة الدلالية، والموثوقية، والامتثال لمتطلبات المجال. من خلال تحليل مفصل للأدبيات البحثية (أكثر من سبعين مقالة مختارة من بين مائتي عنوان أولي)، يتم تقديم إطار مفاهيمي جديد لتحليل أنظمة إدارة المعرفة النصية الإسلامية الذكية، والتي تشمل الاعتبارات التقنية والمنهجية والمجالية. وتظهر النتائج أنه على الرغم مما أبدته تقنيات الرسوم المعرفية الدلالية والنماذج اللغوية الكبيرة من إمكانات واعدة وفائقة في تنظيم المعرفة واسترجاعها وتوليدها من النصوص التراثية، فإن ثمة تحديات جوهرية أساسية مثل ظاهرة "الوهم" في نماذج اللغة، وانعزالية الأنطولوجيات، ونقص البيانات المصنفة عالية الجودة، وتعقيدات التحقق من صحة المستندات، والحاجة إلى قابلية تفسير النماذج، لا تزال تشكل عقبات رئيسية أمام تحقيق أنظمة موثوقة وعالية الأداء بشكل كامل. وتختتم المقالة بطرح خارطة طريق بحثية سداسية المحاور، ترسم المسارات المستقبلية لتطوير نظم ذكية ومستدامة لإدارة معرفة النصوص الإسلامية، بما يستجيب بفاعلية ومسؤولية لمتطلبات المعالجة والتحليل الدقيق للمتون الإسلامية.

الكلمات المفتاحية

إدارة معرفة النصوص الإسلامية، الذكاء الاصطناعي، الويب الدلالي، الرسم المعرفي، النماذج اللغوية الكبيرة، مراجعة نقدية، هندسة النظم، معالجة اللغات الطبيعية.



همگرایی هوش مصنوعی، وب معنایی و مدل‌های زبانی بزرگ در سیستم‌های مدیریت دانش متون اسلامی: یک بررسی انتقادی از منظر مهندسی سیستم^{۱*}

علی میرعرب^{id}

استادیار، گروه اشاعه اطلاعات و دانش پژوهشکده مدیریت اطلاعات و مدارک اسلامی
پژوهشگاه علوم و فرهنگ اسلامی، قم، ایران.
alimirarab@isca.ac.ir

چکیده

مدیریت دانش متون اسلامی با چالش‌های دیرینه‌ای در سازماندهی، بازیابی و تحلیل داده‌های حجیم و پیچیده متنی روبروست. ظهور و بلوغ فناوری‌های هوش مصنوعی، بویژه وب معنایی، گراف‌های دانش و مدل‌های زبانی بزرگ، افق‌های تازه‌ای را در این قلمرو گشوده است. این مقاله به ارائه یک مرور تحلیلی و نقادانه از همگرایی این سه جریان فناورانه در سامانه‌های مدیریت دانش متون اسلامی در بازه زمانی ۲۰۲۳ تا ۲۰۲۶ می‌پردازد. هدف اصلی، نه تنها شناسایی و دسته‌بندی پیشرفت‌های اخیر، بلکه ارزیابی عمیق و مهندسی آنها از منظر معماری سیستم، قابلیت مقیاس‌پذیری، دقت معنایی، قابلیت اطمینان و انطباق با الزامات دامنه است. با تحلیل دقیق نوشتارهای پژوهشی

۱. این مقاله برگرفته از طرح پژوهشی طراحی و توسعه سکوی مدیریت اطلاعات علوم اسلامی در پژوهشگاه علوم و فرهنگ اسلامی است.

* **استاد به این مقاله:** میرعرب، علی. (۱۴۰۴). همگرایی هوش مصنوعی، وب معنایی و مدل‌های زبانی بزرگ در سیستم‌های مدیریت دانش متون اسلامی: یک بررسی انتقادی از منظر مهندسی سیستم. مدیریت دانش اسلامی، (۲)۷، صص ۱۶۵-۱۲۸.

<https://doi.org/10.22081/jikm.2026.74696.1110>

□ نوع مقاله: پژوهشی؛ ناشر: پژوهشگاه علوم و فرهنگ اسلامی، قم، ایران

□ تاریخ دریافت: ۱۴۰۴/۰۲/۰۸ • تاریخ اصلاح: ۱۴۰۴/۰۴/۲۷ • تاریخ پذیرش: ۱۴۰۴/۰۵/۱۰ • تاریخ انتشار آنلاین: ۱۴۰۴/۰۷/۰۹

© ۱۴۰۴ «حق تألیف و حقوق کامل انتشار برای نویسندگان محفوظ است»



(بیش از هفتاد مقاله گزینش شده از میان دویست عنوان اولیه)، یک چهارچوب مفهومی نوین برای تحلیل سامانه‌های مدیریت دانش متون اسلامی هوشمند ارائه می‌شود که ابعاد فنی، روش‌شناختی و ملاحظات دامنه را دربر می‌گیرد. یافته‌ها نشان می‌دهد که در حالی که فناوری‌هایی نظیر گراف‌های دانشی معنایی و مدل‌های زبانی بزرگ توانمندی‌های چشمگیری در سازماندهی، بازیابی و تولید دانش از متون اسلامی از خود نشان داده‌اند، چالش‌های بنیادینی نظیر پدیده «توهم» در مدل‌های زبانی، جزیره‌ای بودن هستی‌شناسی‌ها، کمبود داده‌های برجسب‌گذاری‌شده با کیفیت، پیچیدگی‌های اعتبارسنجی سند و نیاز به تفسیرپذیری مدل‌ها، همچنان به‌عنوان موانع اصلی در دستیابی به سامانه‌های کاملاً قابل اعتماد و با کارایی بالا باقی مانده‌اند. مقاله با ارائه یک نقشه راه پژوهشی شش محوری، مسیرهای آتی را برای توسعه سامانه‌های هوشمند مدیریت دانش متون اسلامی که بتوانند به‌طور مؤثر و مسئولانه به نیازهای تحلیل و پردازش متون اسلامی پاسخ دهند، روشن می‌سازد.

کلیدواژه‌ها

مدیریت دانش متون اسلامی، هوش مصنوعی، وب معنایی، گراف دانش، مدل‌های زبانی بزرگ، بررسی انتقادی، مهندسی سیستم، پردازش زبان طبیعی.

مقدمه

مدیریت دانش متون اسلامی، به مثابه یک قلمرو میان‌رشته‌ای، به سامان‌دهی، ذخیره‌سازی، بازیابی، اشاعه و به کارگیری دانایی برآمده از منابع اصیل اسلامی می‌پردازد. این منابع که قرآن کریم، احادیث نبوی، تفاسیر، کتاب‌های فقهی، کلامی، تاریخی و ادبی را دربرمی‌گیرند، با حجم انبوه، پیچیدگی ساختاری و گستره معنایی بسیار متنوعی همراه‌اند. در رویکردهای سنتی، دشواری دستیابی به اطلاعات، فقدان یکپارچگی میان منابع و نیاز به تخصصی ژرف برای فهم و استنتاج، همواره مانعی جدی فراوی پژوهشگران و کارشناسان این عرصه بوده است (Haji Esmaeili et al., 2021). با این حال، ظهور و بلوغ فناوری‌های نوین هوش مصنوعی در دو دهه اخیر، چشم‌اندازی تازه برای عبور از این موانع گشوده است.

در میان تحولات اخیر، سه جریان فناورانه بویژه در بازتعریف زیست‌بوم مدیریت دانش متون اسلامی نقش محوری پیدا کرده‌اند: وب معنایی، گراف‌های دانش و مدل‌های زبانی بزرگ. وب معنایی با فراهم آوردن چارچوبی برای بازنمایی معنایی داده‌ها از راه هستی‌شناسی‌ها و الگوی توصیف منبع، سازماندهی دانش را از سطح کلیدواژه‌ها به سوی مفاهیم ارتقا داده است (Berners-Lee, Hendler & Lassila, 2001). گراف‌های دانش نیز به مثابه تجسم عملی اصول وب معنایی، شبکه‌هایی از موجودیت‌ها و روابط میان آنها ایجاد می‌کنند که توانایی استنتاج و کشف دانش نو را به همراه می‌آورند (Ji et al., 2022). در نقطه مقابل، مدل‌های زبانی بزرگ با قابلیت کم‌سابقه‌شان در پردازش، درک و تولید زبان طبیعی، ابزارهایی نیرومند برای استخراج دانش از متون نامنسجم و تعامل هوشمند با کاربران پدید آورده‌اند (Brown et al., 2020).

همگرایی این سه شاخه، پتانسیل ایجاد سامانه‌های هوشمند مدیریت دانش متون اسلامی را دارد؛ سامانه‌هایی که بتوانند به‌طور خودکار، دقیق و کارآمد به سامان‌دهی، تحلیل و عرضه دانش از متون اسلامی بپردازند. با این وجود، شتاب بالای تحولات در این حوزه، ضرورت وجود یک بررسی تحلیلی و نقادانه را آشکار ساخته است. بیشتر نوشتارهای مروری موجود، یا تنها بر جنبه‌های صرفاً فنی

تمرکز دارند، یا به یک منبع خاص مانند قرآن یا حدیث محدود می‌شوند و کمتر به واکاوی چالش‌های بنیادین و ترسیم یک نقشه راه جامع برای پژوهش‌های آتی می‌پردازند.

این نوشتار با اتخاذ رویکردی جامع و نقادانه، در پی پاسخ به این پرسش‌های کلیدی است: «فناوری‌های نوین هوش مصنوعی تا چه اندازه توانسته‌اند چالش‌های ذاتی مدیریت دانش متون اسلامی را حل کنند؟»، «مهمترین شکاف‌های پژوهشی و موانع برجای‌مانده در این حوزه کدامند؟» و «چگونه می‌توان به سوی توسعه سامانه‌های مدیریت دانش متون اسلامی قابل اعتماد و با کارایی بالا حرکت کرد؟». بدین منظور، این مقاله به بررسی عمیق نوشتارهای پژوهشی منتشرشده در بازه زمانی ۲۰۲۳ تا ۲۰۲۶ می‌پردازد تا تازه‌ترین پیشرفت‌ها و روندهای این عرصه را پوشش دهد. در این مسیر، نه تنها به توصیف فناوری‌ها و کاربردهای آنها بسنده می‌شود، بلکه با نگاهی نقادانه، نقاط قوت و ضعف رویکردهای موجود ارزیابی و چالش‌های حل‌نشده برجسته می‌شود. ساختار مقاله بدین ترتیب دنبال می‌شود: در بخش دوم، مبانی نظری و روش‌شناسی پژوهش تشریح می‌شود. بخش سوم به تحلیل عمیق نقش وب معنایی و گراف‌های دانش در مدیریت دانش متون اسلامی اختصاص دارد. در بخش چهارم، مدل‌های زبانی بزرگ و رویکردهای نوین مبتنی بر آنها به‌بوته نقد گذاشته می‌شود. بخش پنجم به چالش‌های اصلی، شکاف‌های پژوهشی و ملاحظات فنی می‌پردازد و سرانجام در بخش ششم، جمع‌بندی و چشم‌انداز آینده ارائه می‌شود.

مبانی نظری

برای بررسی نظام‌مند و نقادانه همگرایی فناوری‌های نوین در مدیریت دانش متون اسلامی، پیش از هر چیز نیاز به تبیین روشن مبانی نظری‌ای است که این حوزه بر آنها استوار است. این مبانی شامل چهارچوب مفهومی مدیریت دانش متون اسلامی، اصول وب معنایی، بنیان‌های گراف‌های دانش و نظریه‌های زیربنایی مدل‌های زبانی بزرگ می‌شود.

مبانی نظری مدیریت دانش متون اسلامی

مدیریت دانش متون اسلامی فراتر از یک رویکرد صرفاً فنی، یک چهارچوب مفهومی است که هدف آن حفظ، توسعه و اشاعه دانش بر گرفته از منابع اصیل اسلامی با رعایت اصول اعتبار و دقت متنی می‌باشد. این حوزه بر سه رکن اساسی استوار است: نخست، منابع دانش که شامل قرآن، سنت، اجماع و عقل می‌شود؛ دوم، فرآیندهای دانش از جمله گردآوری، سازماندهی، تحلیل و انتشار؛ و سوم، کاربران دانش که پژوهشگران، متخصصان دینی و عموم مردم را دربر می‌گیرد.

ویژگی‌های منحصربه‌فرد دانش متون اسلامی، این قلمرو را از مدیریت دانش عمومی متمایز می‌سازد. ماهیت وحیانی برخی منابع، اهمیت زنجیره سند در احادیث و نیاز به تفسیر و استنباط، همگی عواملی هستند که هر سامانه مدیریت دانش متون اسلامی باید بتواند آنها را به‌طور مؤثر مدل‌سازی و مدیریت کند تا از اصالت و اعتبار اطلاعات اطمینان حاصل شود (Ta'a et al., 2021). به عبارت دیگر، یک سامانه کارآمد در این حوزه نه تنها باید به جنبه‌های فنی ذخیره و بازیابی اطلاعات پردازد، بلکه باید بستر مناسبی برای بازنمایی لایه‌های معنایی، تاریخی و اعتبارسنجی متون نیز فراهم آورد.

مبانی نظری وب معنایی

وب معنایی که به مثابه توسعه وب کنونی تلقی می‌شود، در پی ایجاد محیطی است که در آن داده‌ها نه تنها برای انسان‌ها خواندنی، بلکه برای ماشین‌ها نیز قابل فهم و پردازش باشند. این هدف از طریق افزودن معنا به داده‌ها محقق می‌شود. ابزار اصلی در این رویکرد، هستی‌شناسی‌ها هستند که نمایشی رسمی و صریح از مفاهیم، ویژگی‌ها و روابط در یک حوزه خاص فراهم می‌کنند.

برنرز، هندلر و لاسیلا^۱ در نوشتار بنیادین خود، وب معنایی را به‌عنوان «زنجیره‌ای از استدلال» توصیف می‌کنند که در آن ماشین‌ها می‌توانند با پیروی از قواعد منطقی، به

1. Berners-Lee, Hendler & Lassila

دانش جدید دست یابند. زبان‌هایی مانند زبان هستی‌شناسی وب و الگوی توصیف منبع، استانداردهایی برای ساخت و تبادل هستی‌شناسی‌ها و داده‌های معنایی به شمار می‌روند. در زمینه متون اسلامی، این زبان‌ها امکان مدل‌سازی مفاهیمی همچون «آیه»، «سوره»، «حدیث»، «راوی» و روابط میان آنها را به‌صورتی فراهم می‌آورند که هم انسان و هم ماشین بتوانند از آن بهره ببرند.

مبانی نظری گراف‌های دانش

گراف‌های دانش به‌عنوان یک کاربرد عملی و مقیاس‌پذیر از اصول وب معنایی، شبکه‌هایی از موجودیت‌ها و روابط میان آنها هستند. هر موجودیت می‌تواند یک مفهوم، شخص، مکان یا رویداد باشد و روابط، چگونگی پیوند این موجودیت‌ها را نشان می‌دهند. ساختار سه‌تایی «سوژه-محمول-مفعول» در گراف‌های دانش، امکان ذخیره‌سازی و بازیابی دانش را به‌صورت معنایی فراهم می‌کند.

به گفته جی^۱ و دیگران (۲۰۲۲)، گراف‌های دانش توانایی‌های نیرومندی برای استنتاج و کشف دانش جدید ارائه می‌دهند. در زمینه متون اسلامی، این قابلیت بویژه برای مدل‌سازی روابط پیچیده میان آیات، احادیث، روایان، مفاهیم فقهی و رویدادهای تاریخی کاربرد دارد. (Kamran, Abro & Basharat, 2023) نشان داده‌اند که چگونه یک گراف دانش می‌تواند زنجیره سند را به‌صورت شبکه‌ای از روایان بازنمایی کند و از این طریق، امکان تحلیل‌های نوینی را در علم حدیث فراهم آورد.

مبانی نظری مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ، شبکه‌های عصبی ژرفی هستند که بر پایه معماری ترانسفورمر طراحی شده‌اند و بر روی حجم عظیمی از داده‌های متنی آموزش دیده‌اند. این مدل‌ها قادرند الگوهای پیچیده زبانی را درک کرده و وظایفی نظیر تولید متن، خلاصه‌سازی،

1. Ji

ترجمه و پاسخ‌گویی به پرسش را با دقتی چشمگیر انجام دهند.

براون و دیگران (۲۰۲۰) نشان دادند که این مدل‌ها با معماری ترانسفورمر، توانایی «یادگیری چندنمونه‌ای» را پیدا می‌کنند؛ بدین معنا که می‌توانند از چند نمونه محدود، الگوی یک وظیفه را فراگیرند و آن را بر روی نمونه‌های نو به کار برند. توانایی مدل‌های زبانی بزرگ در درک زمینه و تولید متن منسجم و مرتبط، آنها را به ابزاری قدرتمند برای پردازش متون اسلامی تبدیل کرده است. با این حال، پن و دیگران (۲۰۲۳) هشدار می‌دهند که چالش‌هایی نظیر «توهم» - یعنی تولید اطلاعات نادرست یا بی‌اساس - و نیاز به داده‌های آموزشی تخصصی، استفاده از این مدل‌ها را در حوزه‌های حساس مانند متون دینی با مخاطراتی جدی همراه می‌سازد.

در این چارچوب، مفهوم «هوش مصنوعی مسئولانه» اهمیت می‌یابد. براساس اصول «هوش مصنوعی قابل اعتماد» اتحادیه اروپا (European Commission, 2019) و رهنمودهای سازمان همکاری و توسعه اقتصادی (OECD, 2019)، هوش مصنوعی مسئولانه به توسعه و به کارگیری سامانه‌های هوش مصنوعی به گونه‌ای اطلاق می‌شود که از شفافیت، قابلیت توضیح، عدالت، پاسخ‌گویی و قابلیت اعتماد برخوردار باشند و در عین حال به حقوق بنیادین انسان احترام بگذارند. همچنین، چارچوب مدیریت ریسک مؤسسه ملی استاندارد و فناوری آمریکا (NIST, 2023) بر ضرورت ارزیابی پیوسته ریسک‌ها، کاهش سوگیری‌ها و تضمین ایمنی و قابلیت اطمینان این سامانه‌ها تأکید دارد. در حوزه متون اسلامی، این اصول افزون بر الزامات عمومی، مستلزم توجه ویژه به حفظ اصالت متون، جلوگیری از تحریف مفاهیم دینی، امکان ردیابی منابع و تضمین اعتبار استنادات نیز می‌باشد. از این رو، به کارگیری مدل‌های زبانی بزرگ در این حوزه، نه تنها یک مسئله فنی، بلکه مسئله‌ای معرفت‌شناختی و اخلاقی تلقی می‌شود که نیازمند چارچوب‌های دقیق نظارتی و ارزیابی است.

روش‌شناسی پژوهش

رویکرد این مقاله مبتنی بر یک مرور نظام‌مند و نقادانه از نوشتارهای پژوهشی منتشرشده

در بازه زمانی ۲۰۲۳ تا ۲۰۲۶ است. روش‌شناسی به کار گرفته شده در این پژوهش از چهار مرحله اصلی تشکیل شده است: شناسایی و گردآوری مقالات، غربالگری و گزینش، استخراج و تحلیل داده‌ها و در نهایت تلفیق و نگارش. هر یک از این مراحل در ادامه به تفصیل تشریح می‌شود.

شناسایی و گردآوری مقالات

جستجوی نظام‌مند در پایگاه‌های داده علمی معتبر از جمله آی‌تریپل‌ای^۱، کتابخانه دیجیتال ای‌سی‌ام^۲، ساینس دایرکت^۳، اشپرینگر^۴، اسکپوس^۵، وب‌آو ساینس^۶ و آرشیو^۷ انجام شد. کلیدواژه‌های مورد استفاده عبارت بودند از «مدیریت دانش متون اسلامی»، «وب معنایی»، «گراف دانش»، «مدل‌های زبانی بزرگ»، «هوش مصنوعی در مطالعات اسلامی»، «پردازش زبان طبیعی قرآنی»، «پردازش زبان طبیعی حدیثی» و ترکیب‌های گوناگون آنها. بازه زمانی اصلی جستجو بر سال‌های ۲۰۲۳ تا ۲۰۲۶ متمرکز بود تا تازه‌ترین پیشرفت‌های این حوزه پوشش داده شود. افزون بر این، نوشتارهای مروری و فصل‌های کتاب‌های مرتبط نیز بررسی شدند.

غربالگری و گزینش مقالات

پس از گردآوری اولیه، مقالات براساس عنوان، چکیده و کلیدواژه‌ها غربالگری شدند. چهار معیار اصلی برای گزینش مقالات در نظر گرفته شد:

- نخست، ارتباط موضوعی؛ مقالاتی که به‌طور مستقیم به کاربرد هوش مصنوعی

-
1. IEEE
 2. ACM
 3. Science Direct
 4. Springer
 5. Scopus
 6. Web of Science
 7. arxiv

(وب معنایی، گراف دانش، مدل‌های زبانی بزرگ) در مدیریت دانش متون اسلامی یا تحلیل متون اسلامی می‌پرداختند، انتخاب شدند.

- دوم، اعتبار علمی؛ مقالاتی که در نشریات با رتبه یک یا دو (براساس گزارش استنادی نشریات یا اسکپوس) یا در همایش‌های تراز اول در حوزه‌های هوش مصنوعی، پردازش زبان طبیعی و وب معنایی منتشر شده بودند، در اولویت قرار گرفتند.

- سوم، نوآوری فنی یا مفهومی؛ مقالاتی که مدل، چهارچوب، الگوریتم یا مجموعه داده جدیدی معرفی می‌کردند یا به تحلیل عمیق چالش‌ها می‌پرداختند، گزینش شدند.

- چهارم، تاریخ انتشار؛ اولویت با مقالات منتشر شده در سال‌های ۲۰۲۴، ۲۰۲۵ و ۲۰۲۶ بود تا تازه‌ترین روندها بازتاب داده شود. با این حال، برخی مقالات کلیدی از سال‌های پیشین نیز برای ارائه زمینه تاریخی و مبانی نظری مورد استفاده قرار گرفتند.

از میان بیش از دویست مقاله شناسایی شده در مرحله اولیه، حدود هفتاد مقاله برای تحلیل عمیق گزینش شدند.

استخراج و تحلیل داده‌ها

از هر مقاله گزینش شده، اطلاعات کلیدی شامل هدف پژوهش، روش‌شناسی، فناوری‌های مورد استفاده، نتایج اصلی، نوآوری‌ها، محدودیت‌ها و چالش‌های شناسایی شده استخراج شد. این اطلاعات به صورت ساختاریافته ثبت شدند تا امکان مقایسه و تحلیل نقادانه فراهم آید.

تحلیل داده‌ها با تمرکز بر شناسایی الگوها، روندها، نقاط قوت و ضعف رویکردهای گوناگون و شکاف‌های پژوهشی انجام شد. همچنین تلاش شد تا تأثیرات فنی و روش‌شناختی کاربرد این فناوری‌ها نیز مورد توجه قرار گیرد.

چالش‌های بازیابی معنایی در علوم اسلامی

تحلیل معماری‌های وب معنایی و گراف دانش در سامانه‌های مدیریت دانش متون اسلامی

وب معنایی و گراف‌های دانش به‌عنوان دو پارادایم کلیدی در مهندسی دانش، نقشی بنیادین در توسعه سامانه‌های مدیریت دانش متون اسلامی ایفا کرده‌اند. این فناوری‌ها امکان مدل‌سازی صریح مفاهیم، روابط و موجودیت‌های پیچیده موجود در منابع متنی اسلامی را فراهم می‌آورند و از این رهگذر، قابلیت‌های جستجو، بازیابی و استنتاج معنایی را به گونه‌ای چشمگیر ارتقا می‌بخشند. در این بخش، به تحلیل عمیق معماری‌ها، روش‌شناسی‌ها و پیشرفت‌های حاصل‌شده در این حوزه از منظر مهندسی سامانه‌ها می‌پردازیم.

هستی‌شناسی‌ها به‌عنوان چهارچوب‌های مدل‌سازی معنایی

هستی‌شناسی‌ها که بازنمایی‌های رسمی و صریح از مفاهیم و روابط در یک حوزه خاص هستند، ابزاری نیرومند برای مدل‌سازی دانش متون اسلامی فراهم آورده‌اند. توسعه هستی‌شناسی‌های دامنه‌خاص برای متون اسلامی، گام نخست در ساخت سامانه‌های مدیریت دانش متون اسلامی معنایی به شمار می‌رود. این هستی‌شناسی‌ها در پی تعریف کلاس‌ها، ویژگی‌ها و روابط میان موجودیت‌های متنی اسلامی هستند. برای نمونه، یک هستی‌شناسی می‌تواند مفاهیمی همچون «آیه»، «سوره»، «حدیث»، «راوی»، «فقیه» و «حکم فقهی» را همراه با روابطی نظیر «شامل است»، «روایت کرده است» و «استنباط کرده است» مدل‌سازی کند (Ta'a et al., 2021).

پژوهش‌های آغازین در این زمینه بر توسعه هستی‌شناسی‌های متمرکز بر یک منبع خاص، مانند قرآن یا حدیث، متمرکز بود. برای مثال، هستی‌شناسی‌های قرآنی به مدل‌سازی مفاهیم، موضوعات، شخصیت‌ها و رویدادهای ذکر شده در قرآن می‌پردازند (Khan et al., 2020; Harrag et al., 2013). این هستی‌شناسی‌ها غالباً با استفاده از ابزارهایی مانند پروتزه^۱ و

1. Protege

با مشارکت خبرگان دامنه - متخصصان علوم اسلامی - به صورت دستی ساخته شده‌اند. هدف اصلی این رویکرد، فراهم آوردن یک چهارچوب معنایی برای جستجوی دقیق‌تر و بازیابی اطلاعات مرتبط با آیات و تفاسیر بود. با این حال، یکی از محدودیت‌های جدی این رویکرد، جزیره‌ای بودن هستی‌شناسی‌ها و نبود قابلیت تعامل میان آنهاست. هر هستی‌شناسی به طور مستقل توسعه می‌یافت و این امر مانع از یکپارچه‌سازی دانش از منابع مختلف متنی اسلامی می‌شد.

در سال‌های اخیر، تلاش‌هایی برای غلبه بر این چالش از طریق توسعه هستی‌شناسی‌های جامع‌تر و یکپارچه‌تر انجام گرفته است. برای نمونه، (Bendjamaa & Taleb, 2024) هستی‌شناسی «دین آنتو» را معرفی کردند که در پی یکپارچه‌سازی مفاهیم قرآن و حدیث در یک مدل واحد است. این رویکرد، گامی مهم در جهت ایجاد دیدگاهی جامع‌تر از «دانش متنی اسلامی» به شمار می‌رود. منظور از دانش متنی اسلامی در اینجا، دانشی است که مستقیماً از متون دینی (مانند قرآن، حدیث و تفاسیر) استخراج می‌شود و در قالب زبان طبیعی و ساختارهای معنایی نهفته در متن شکل می‌گیرد؛ این نوع دانش در مقابل «دانش ساخت‌یافته صوری» (نظیر پایگاه‌های داده و مدل‌های رسمی از پیش تعریف‌شده) و نیز «دانش غیرمتنی» (مانند داده‌های عددی، تصویری یا تجربی) قرار می‌گیرد.

با این حال، چالش اصلی در این زمینه، هم‌ترازی هستی‌شناسی‌ها و حل تعارضات معنایی میان آنهاست. متون اسلامی دارای مکاتب فکری و تفاسیر متعددی هستند که ممکن است در تعریف برخی مفاهیم یا روابط با یکدیگر تفاوت داشته باشند. مدیریت این تنوع و ایجاد یک هستی‌شناسی فدرال که بتواند این تفاوت‌ها را به رسمیت بشناسد و در عین حال یکپارچگی معنایی را حفظ کند، همچنان یک چالش پژوهشی فعال باقی مانده است.

گراف‌های دانش متون اسلامی: مدل‌سازی روابط پیچیده و استنتاج

گراف‌های دانش به عنوان ابزاری نیرومند برای تجسم و مدیریت روابط پیچیده میان

موجودیت‌ها، به‌طور فزاینده‌ای در مدیریت دانش متون اسلامی مورد استفاده قرار گرفته‌اند. این گراف‌های دانش فراتر از نمایش سلسله‌مراتبی مفاهیم، امکان مدل‌سازی روابط معنایی، علی، زمانی و استنتاجی را فراهم می‌آورند که برای فهم عمیق دانش متنی اسلامی ضروری است. برای نمونه، یک گراف دانش از متون اسلامی می‌تواند روابط میان یک آیه، شأن نزول آن، احادیث مرتبط، تفاسیر گوناگون، احکام فقهی استنباط‌شده از آن و حتی شخصیت‌های تاریخی مرتبط را به‌صورت صریح مدل‌سازی کند (Kamran et al. , 2026).

یکی از برجسته‌ترین کاربردهای گراف‌های دانش در مدیریت دانش متون اسلامی، مدل‌سازی پیکره حدیث است. (Kamran, Abro & Basharat, 2023) سامانه «حدیث‌معنا» را معرفی کردند که یک گراف دانش مبتنی بر هستی‌شناسی برای کل پیکره حدیث به شمار می‌رود. این گراف دانش، موجودیت‌هایی مانند «راوی»، «متن حدیث» و «کتاب حدیث» را همراه با روابطی نظیر «روایت کرده است» و «موجود در» مدل‌سازی می‌کند. این رویکرد امکان جستجوی معنایی پیچیده را فراهم می‌آورد؛ برای نمونه، می‌توان در پی احادیثی گشت که توسط راوی خاصی در یک دوره زمانی مشخص روایت شده‌اند و به موضوع فقهی خاصی مربوط می‌شوند. این قابلیت، به پژوهشگران حدیث کمک می‌کند تا تحلیل‌های عمیق‌تری بر روی زنجیره سند و متن حدیث انجام دهند.

توسعه گراف‌های دانشی متون اسلامی با چالش‌های متعددی همراه است. نخست، استخراج خودکار موجودیت‌ها و روابط از متون بدون ساختار و نیمه‌ساختاریافته اسلامی. این فرآیند نیازمند فنون پیشرفته پردازش زبان طبیعی است که بتوانند با پیچیدگی‌های زبان عربی کلاسیک، از جمله چندمعنایی، ابهام و تنوع گویش‌ها، مقابله کنند (Bashir et al. , 2023). دوم، اعتبارسنجی و کیفیت داده‌ها در گراف دانش. با توجه به حساسیت دانش متنی اسلامی، اطمینان از صحت و دقت اطلاعات مدل‌سازی‌شده در گراف دانش از اهمیتی والا برخوردار است. این امر نیازمند مشارکت خبرگان دامنه و توسعه الگوریتم‌های خودکار برای شناسایی ناسازگاری‌ها و خطاهاست (Xue et al. , 2023).

سوم، مقیاس‌پذیری گراف‌های دانش. با توجه به حجم انبوه متون اسلامی، ساخت و نگهداری یک گراف دانش جامع و مقیاس‌پذیر که بتواند میلیون‌ها موجودیت و رابطه را مدیریت کند، یک چالش فنی بزرگ است.

معماری سامانه‌های مدیریت دانش متون اسلامی مبتنی بر وب معنایی و گراف دانش

معماری سامانه‌های مدیریت دانش متون اسلامی مبتنی بر وب معنایی و گراف دانش معمولاً از چندین لایه تشکیل شده است که هر لایه وظایف خاصی را بر عهده دارد. یک معماری نمونه می‌تواند شامل لایه‌های زیر باشد:

- **لایه گردآوری و پیش‌پردازش داده:** این لایه مسئول گردآوری متون اسلامی از منابع گوناگون (دیجیتال، اسکن‌شده، صوتی) و انجام عملیات پیش‌پردازش مانند بازنشاسی نویسه‌های نوری برای متون اسکن‌شده، تشخیص گفتار برای منابع صوتی، واژه‌سازی، ریشه‌یابی و حذف نویز است. چالش اصلی در این لایه، مقابله با تنوع قالب‌ها، کیفیت‌های متفاوت داده‌ها و پیچیدگی‌های زبانی متون عربی کلاسیک و فارسی است (Bashir et al., 2023).
- **لایه استخراج دانش:** در این لایه، فنون پردازش زبان طبیعی و یادگیری ماشین برای استخراج موجودیت‌های نام‌دار، روابط و رویدادها از متون پیش‌پردازش‌شده به کار گرفته می‌شوند. این فرآیند می‌تواند شامل استفاده از مدل‌های مبتنی بر قاعده، مدل‌های آماری یا مدل‌های یادگیری ژرف باشد. خروجی این لایه، سه‌تایی‌های الگوی توصیف منبع یا ساختارهای مشابهی است که آماده ورود به گراف دانش هستند (Peng & Yang, 2024).
- **لایه مدل‌سازی و ذخیره‌سازی دانش:** این لایه شامل هستی‌شناسی‌های دامنه است که به عنوان طرحواره برای گراف دانش عمل می‌کنند. هستی‌شناسی‌ها با استفاده از زبان‌هایی مانند زبان هستی‌شناسی وب توسعه می‌یابند و مفاهیم و روابط مجاز در دامنه را تعریف می‌کنند. گراف دانش که از سه‌تایی‌های استخراج‌شده

در لایهٔ پیشین پر می‌شود، در یک پایگاه داده گراف دانش مانند Neo4J^۱ یا ویرتوسو^۲ ذخیره می‌شود. انتخاب پایگاه داده گراف دانش مناسب براساس نیازهای مقیاس‌پذیری، کارایی و قابلیت پرس‌وجو انجام می‌گیرد (Ji et al. , 2020).

• **لایهٔ استنتاج و تحلیل:** این لایه مسئول انجام استنتاج‌های منطقی بر روی گراف دانش برای کشف دانش نو یا اعتبارسنجی دانش موجود است. موتورهای استنتاج با استفاده از قوانین استنتاج تعریف‌شده در هستی‌شناسی یا قوانین کشف‌شده از داده‌ها، به تولید سه‌تایی‌های تازه می‌پردازند. این لایه همچنین می‌تواند شامل الگوریتم‌های تحلیل گراف دانش برای شناسایی الگوها، خوشه‌بندی و رتبه‌بندی موجودیت‌ها باشد (Hogan et al. , 2021).

• **لایهٔ واسط کاربری و کاربرد:** این لایه شامل ابزارهایی برای تعامل کاربران با سامانهٔ مدیریت دانش متون اسلامی است. این ابزارها می‌توانند شامل واسط‌های جستجوی معنایی، سامانه‌های پرسش و پاسخ، ابزارهای تجسم گراف دانش و سامانه‌های توصیه‌گر باشند. طراحی واسط کاربری باید به گونه‌ای باشد که پیچیدگی‌های زیربنایی سامانه را پنهان کرده و تجربهٔ کاربری شهودی و کارآمدی را فراهم آورد (Malhas, Mansour & Elsayed, 2023).

چالش‌های مهندسی و روش‌شناختی در توسعه سامانه‌های مدیریت دانش متون اسلامی مبتنی بر وب معنایی و گراف دانش

علی‌رغم پیشرفت‌های چشمگیر، کاربرد وب معنایی و گراف‌های دانش در مدیریت دانش متون اسلامی با چالش‌های مهندسی و روش‌شناختی متعددی روبروست که نیازمند پژوهش‌های آتی است:

▪ پیچیدگی‌های زبانی و پردازش زبان طبیعی: زبان عربی کلاسیک، بویژه در متون

1. Neo4J

2. Virtuoso

قرآنی و حدیثی، دارای پیچیدگی‌های صرفی، نحوی و بلاغی فراوانی است. پدیده‌هایی مانند چندمعنایی، همنویسی و استعاره، استخراج دقیق معنا و روابط را برای سامانه‌های خودکار دشوار می‌سازد. توسعه ابزارهای پردازش زبان طبیعی تخصصی برای زبان عربی کلاسیک که بتوانند این ظرافت‌ها را درک کنند، یک ضرورت مهندسی است (Bashir et al. , 2023). این امر شامل توسعه مدل‌های زبانی بومی‌سازی‌شده و مجموعه داده‌های برجسب‌گذاری‌شده با کیفیت بالا برای وظایف تخصصی پردازش زبان طبیعی است.

▪ کمبود ابزارهای خودکار برای ساخت هستی‌شناسی و گراف دانش: بخش عمده‌ای از هستی‌شناسی‌ها و گراف‌های دانشی متون اسلامی به صورت دستی یا نیمه خودکار ساخته شده‌اند که فرآیندی زمان‌بر، پرهزینه و نیازمند تخصص انسانی بالاست. توسعه ابزارهای خودکار برای استخراج مفاهیم، روابط و قوانین استنتاج از متون اسلامی، می‌تواند سرعت و مقیاس‌پذیری ساخت گراف‌های دانش را به گونه‌ای چشمگیر افزایش دهد. پژوهش‌های اخیر در زمینه استفاده از مدل‌های زبانی بزرگ برای ساخت گراف دانش، مسیرهای تازه‌ای را در این زمینه گشوده‌اند (Kommineni et al. , 2024; Feng et al. , 2024). با این حال، اعتبارسنجی خروجی‌های خودکار و اطمینان از دقت آنها همچنان یک چالش باقی مانده است.

▪ مدیریت تکامل دانش و نگهداری هستی‌شناسی: دانش متنی اسلامی یک دامنه پویاست و ممکن است با گذشت زمان، تفاسیر تازه‌ای ارائه شود یا روابط جدیدی کشف گردد. هستی‌شناسی‌ها و گراف‌های دانش باید قابلیت تکامل و به‌روزرسانی پیوسته را داشته باشند. مدیریت تغییرات در هستی‌شناسی‌ها و ادغام اطلاعات تازه در گراف دانش، بدون ایجاد ناسازگاری و حفظ یکپارچگی معنایی، یک چالش مهم مهندسی نرم‌افزار است. توسعه سازوکارهای خودکار برای تشخیص تغییرات و پیشنهاد به‌روزرسانی‌ها ضروری به نظر می‌رسد.

▪ اعتبارسنجی و کیفیت داده‌ها در گراف دانش: با توجه به حساسیت دانش متنی اسلامی، اطمینان از صحت و دقت اطلاعات مدل‌سازی شده در گراف دانش از اهمیتی والا برخوردار است. توسعه سازوکارهای نیرومند برای اعتبارسنجی داده‌ها، شناسایی خطاها و ارائه شفافیت در مورد منبع هر قطعه دانش، برای جلب اعتماد کاربران و تضمین قابلیت اطمینان سامانه ضروری است (Xue et al., 2023). این امر شامل توسعه معیارهای کیفیت گراف دانش و الگوریتم‌های تشخیص ناهنجاری است.

▪ مقیاس‌پذیری و کارایی سامانه: با توجه به حجم انبوه متون اسلامی و پیچیدگی گراف‌های دانش، اطمینان از مقیاس‌پذیری و کارایی بالای سامانه‌های مدیریت دانش متون اسلامی یک چالش مهم مهندسی است. این شامل بهینه‌سازی ساختار گراف دانش، انتخاب پایگاه داده گراف دانش مناسب، بهینه‌سازی پرس‌وجوها و استفاده از معماری‌های توزیع شده برای پردازش و ذخیره‌سازی داده‌هاست. همچنین، پردازش‌های سنگینی همچون استخراج دانش و استنتاج، نیازمند زیرساخت‌های محاسباتی نیرومند و بهینه‌سازی الگوریتمی هستند.

در مجموع، وب معنایی و گراف‌های دانش، زیرساختی نیرومند برای سازماندهی و تحلیل دانش متنی اسلامی فراهم آورده‌اند. با این حال، برای تحقق کامل پتانسیل آنها، نیازمند غلبه بر چالش‌های فنی و روش‌شناختی هستیم که در این بخش به آنها اشاره شد. در بخش بعدی، به بررسی نقش مدل‌های زبانی بزرگ و همگرایی آنها با گراف‌های دانش در تحول مدیریت دانش متون اسلامی خواهیم پرداخت.

بازیابی معنایی اطلاعات در هر حوزه تخصصی با چالش‌هایی روبه‌راست، اما در علوم اسلامی به دلیل ویژگی‌های خاص این دانش، دشواری‌های مضاعفی پدیدار می‌شود. این چالش‌ها را می‌توان در چهار سطح مفهومی، زبانی، فنی و ساختاری طبقه‌بندی کرد. در ادامه هر یک از این ابعاد به تفصیل بررسی می‌شود.

تحلیل مدل‌های زبانی بزرگ و رویکردهای عصبی-معنایی در سامانه‌های مدیریت دانش متون اسلامی

پیدایش مدل‌های زبانی بزرگ در سال‌های اخیر، بویژه از ۲۰۲۳ به بعد، نقطه عطفی در حوزه پردازش زبان طبیعی و به تبع آن، در سامانه‌های مدیریت دانش متون اسلامی ایجاد کرده است. این مدل‌ها با توانایی کم‌نظیر خود در درک، تولید و خلاصه‌سازی متن، پتانسیل‌های تازه‌ای را برای تعامل هوشمند با متون اسلامی و استخراج دانش از آنها فراهم آورده‌اند. با این حال، کاربرد مدل‌های زبانی بزرگ در حوزه‌ای با الزامات دقت و اعتبار بالا مانند متون اسلامی، با چالش‌ها و ملاحظات مهندسی خاص خود همراه است.

توانمندی‌های مدل‌های زبانی بزرگ در پردازش متون اسلامی از منظر مهندسی

مدل‌های زبانی بزرگ، به دلیل آموزش بر روی حجم انبوهی از داده‌های متنی، قادر به یادگیری الگوهای پیچیده زبانی و درک معنایی عمیق هستند. این توانایی‌ها، آنها را به ابزاری نیرومند برای وظایف گوناگون در مدیریت دانش متون اسلامی تبدیل کرده است.

- **استخراج اطلاعات و شناسایی موجودیت‌های نام‌دار:** مدل‌های زبانی بزرگ می‌توانند به‌طور خودکار موجودیت‌های نام‌دار مانند نام اشخاص (راویان، فقها، صحابه)، مکان‌ها (مکه، مدینه، قدس)، زمان‌ها (سال هجری، قرون) و مفاهیم کلیدی اسلامی (نماز، زکات، حج) را از متون بدون ساختار استخراج کنند. این قابلیت، بویژه برای ساخت و غنی‌سازی گراف‌های دانشی متون اسلامی، از اهمیتی والا برخوردار است (Peng & Yang, 2024). مدل‌های بومی‌سازی‌شده برای زبان عربی، مانند آرابرت^۱، در این زمینه کارایی قابل قبولی از خود نشان داده‌اند و معماری‌های مبتنی بر ترانسفورمر امکان موازی‌سازی و پردازش کارآمد حجم انبوهی از داده‌ها را فراهم می‌آورند (Bashir et al., 2023).

1. AraBERT

- خلاصه‌سازی و تولید متن: مدل‌های زبانی بزرگ می‌توانند متون بلند اسلامی، مانند تفاسیر قرآنی یا شروح حدیث، را خلاصه‌سازی کرده و نکات کلیدی آنها را استخراج کنند. این قابلیت می‌تواند به پژوهشگران کمک کند تا به سرعت به محتوای اصلی متون دست یابند. افزون بر این، مدل‌های زبانی بزرگ قادر به تولید متن منسجم و مرتبط هستند که می‌تواند در تولید محتوای آموزشی، پاسخ به پرسش‌ها و حتی نگارش پیش‌نویس نوشتارهای علمی مورد استفاده قرار گیرد. با این حال، دقت و اعتبار محتوای تولیدشده، بویژه در حوزه‌های حساس، نیازمند بررسی دقیق و سازوکارهای اعتبارسنجی خودکار و انسانی است.
- سامانه‌های پرسش و پاسخ پیشرفته: مدل‌های زبانی بزرگ می‌توانند به پرسش‌های کاربران به زبان طبیعی پاسخ دهند. این سامانه‌ها، با درک معنای پرسش و جستجو در پایگاه دانش خود، پاسخ‌های مرتبط را تولید می‌کنند. در زمینه متون اسلامی، این قابلیت می‌تواند در توسعه گفت‌وگوهای هوشمند برای پاسخ به پرسش‌های فقهی، تاریخی یا تفسیری مورد استفاده قرار گیرد (Uriawan et al. , 2025).
- تحلیل معنایی و دسته‌بندی موضوعی: مدل‌های زبانی بزرگ می‌توانند متون اسلامی را براساس محتوای معنایی آنها دسته‌بندی کنند. برای نمونه، می‌توان آیات یا احادیث را براساس موضوعات فقهی، اخلاقی، تاریخی یا اعتقادی دسته‌بندی کرد. این قابلیت به سازماندهی بهتر دانش و تسهیل جستجوی موضوعی کمک می‌کند. همچنین، مدل‌های زبانی بزرگ می‌توانند در شناسایی ارتباط معنایی پنهان میان متون گوناگون، مانند ارتباط یک آیه با یک حدیث خاص، نقش مؤثری ایفا کنند (Alshammari, Atwell & Alsalka, 2024). این فرآیند غالباً از طریق نهان‌سازی‌های معنایی و مدل‌های خوشه‌بندی صورت می‌گیرد.

رویکردهای نوین: همگرایی مدل‌های زبانی بزرگ با گراف‌های دانش

یکی از مهمترین پیشرفت‌ها در کاربرد مدل‌های زبانی بزرگ در مدیریت دانش،

همگرایی آنها با گراف‌های دانشست. این رویکرد که به هوش مصنوعی-عصبی-معنایی نیز شهرت یافته، در پی ترکیب قدرت یادگیری آماری مدل‌های زبانی بزرگ با دقت و قابلیت استنتاج منطقی گراف‌های دانش است. هدف اصلی این همگرایی، غلبه بر محدودیت‌های هر یک از این فناوری‌ها به تنهایی و ایجاد سامانه‌هایی با توانایی‌های استنتاجی نیرومندتر و پاسخ‌های قابل اعتمادتر است (Pan et al., 2023).

- تولید تقویت‌شده با بازیابی: این رویکرد یکی از مؤثرترین روش‌ها برای کاهش پدیده «توهم» در مدل‌های زبانی بزرگ و افزایش دقت پاسخ‌های آنهاست. در این رویکرد، مدل زبانی بزرگ به جای تولید پاسخ تنها براساس دانش درونی خود، ابتدا از یک پایگاه دانش بیرونی (مانند یک گراف دانش متون اسلامی یا یک پایگاه داده متنی معتبر) برای بازیابی اطلاعات مرتبط استفاده می‌کند. سپس، مدل زبانی بزرگ با بهره‌گیری از این اطلاعات بازیابی‌شده، پاسخ نهایی را تولید می‌کند. این فرآیند تضمین می‌کند که پاسخ‌ها براساس منابع مستند و قابل ارجاع ارائه می‌شوند و از صحت و اعتبار بالاتری برخوردارند (ArabicNLP Shared Task, 2025). در یک سامانه مدیریت دانش متون اسلامی مبتنی بر تولید تقویت‌شده با بازیابی، اگر کاربر پرسشی فقهی مطرح کند، سامانه ابتدا در گراف دانش فقهی در پی احکام و دلایل مرتبط می‌شود و سپس مدل زبانی بزرگ با استفاده از این اطلاعات، پاسخ مستند را تولید می‌کند. معماری این رویکرد شامل یک پیمانه بازیابی و یک پیمانه تولید است که به صورت پی‌درپی عمل می‌کنند.

- غنی‌سازی گراف دانش با مدل‌های زبانی بزرگ: مدل‌های زبانی بزرگ می‌توانند در فرآیند ساخت و غنی‌سازی گراف‌های دانش نیز نقش مؤثری ایفا کنند. این مدل‌ها قادرند موجودیت‌ها و روابط را از متون بدون ساختار استخراج کرده و به گراف دانش بیفزایند. همچنین، می‌توان از مدل‌های زبانی بزرگ برای پیشنهاد نگاشت‌ها میان هستی‌شناسی‌های گوناگون و حل تعارضات معنایی استفاده کرد (Kommineni et al., 2024; Feng et al., 2024). این رویکرد می‌تواند

فرآیند ساخت گراف دانش را که معمولاً زمان‌بر و نیازمند تخصص انسانی است، شتاب بخشد. با این حال، اعتبارسنجی خروجی‌های مدل‌های زبانی بزرگ در این زمینه، بویژه در حوزه‌های حساس، همچنان یک چالش مهم است و نیازمند سازوکارهای نظارت انسانی و اعتبارسنجی خودکار می‌باشد.

- سامانه‌های پرسش و پاسخ مبتنی بر گراف دانش و مدل‌های زبانی بزرگ: ترکیب مدل‌های زبانی بزرگ با گراف‌های دانش، منجر به توسعه سامانه‌های پرسش و پاسخ بسیار نیرومندتری شده است. در این سامانه‌ها، مدل زبانی بزرگ نه تنها قادر به درک پرسش زبان طبیعی است، بلکه می‌تواند آن را به یک پرس‌وجوی ساختاریافته (مانند زبان پرسشگری الگوی توصیف منبع) بر روی گراف دانش تبدیل کند. سپس، پاسخ از گراف دانش استخراج شده و توسط مدل زبانی بزرگ به زبانی طبیعی و قابل فهم برای کاربر ارائه می‌شود. این رویکرد، دقت و قابلیت استنتاج سامانه‌های پرسش و پاسخ را به گونه‌ای چشمگیر افزایش می‌دهد و امکان پاسخ‌گویی به پرسش‌های پیچیده‌تر را فراهم می‌آورد (Mohammad et al. , 2024). این معماری، قابلیت تفسیرپذیری بیشتری را نیز فراهم می‌کند، زیرا مسیر استنتاج از گراف دانش قابل ردیابی است.

چالش‌های مهندسی و معماری در کاربرد مدل‌های زبانی بزرگ برای مدیریت دانش متون اسلامی

علی‌رغم پتانسیل‌های فراوان، کاربرد مدل‌های زبانی بزرگ در مدیریت دانش متون اسلامی با چالش‌های مهندسی و معماری بنیادینی روبروست که نیازمند توجه جدی پژوهشگران است.

- پدیده «توهم» و الزامات دقت: همان‌گونه که پیشتر اشاره شد، پدیده «توهم» در مدل‌های زبانی بزرگ، یعنی تولید اطلاعات نادرست یا بی‌اساس، بزرگترین چالش در حوزه‌هایی با الزامات دقت بالا است. در متون اسلامی که دقت در نقل قول از منابع و اصالت سند از اهمیتی والا برخوردار است، حتی یک خطای

کوچک می‌تواند پیامدهای جدی به همراه داشته باشد. اگرچه رویکرد تولید تقویت‌شده با بازیابی به کاهش این پدیده کمک می‌کند، اما آن را به‌طور کامل از بین نمی‌برد. نیاز به توسعه سازوکارهای نیرومندتر برای اعتبارسنجی خودکار پاسخ‌ها و اطمینان از انطباق آنها با منابع معتبر، یک ضرورت مهندسی است (ArabicNLP Shared Task, 2025). این شامل توسعه معیارهای ارزیابی دقیق و

بنچ‌مارک‌های تخصصی برای سنجش میزان توهم و دقت پاسخ‌هاست.

- **کمبود داده‌های آموزشی تخصصی و برچسب‌گذاری‌شده:** کارایی مدل‌های زبانی بزرگ بشدت به کیفیت و کمیت داده‌های آموزشی بستگی دارد. در حالی که داده‌های متنی عمومی فراوانی برای آموزش این مدل‌ها وجود دارد، داده‌های برچسب‌گذاری‌شده تخصصی در حوزه متون اسلامی (بوژه برای زبان عربی کلاسیک و فارسی) بسیار محدود است. این کمبود داده، مانع از تنظیم دقیق مؤثر مدل‌های زبانی بزرگ برای درک ظرافت‌های معنایی و اصطلاحات تخصصی اسلامی می‌شود. گردآوری و برچسب‌گذاری این داده‌ها، با مشارکت خبرگان دامنه، فرآیندی زمان‌بر و پرهزینه است و نیازمند سرمایه‌گذاری‌های کلان پژوهشی می‌باشد (Bashir et al., 2023). توسعه روش‌های یادگیری با نظارت ضعیف و یادگیری بدون نظارت می‌تواند به کاهش این وابستگی کمک کند.
- **توضیح‌پذیری و شفافیت مدل:** مدل‌های زبانی بزرگ اغلب به‌عنوان «جعبه سیاه» عمل می‌کنند، بدین معنا که فرآیند تصمیم‌گیری آنها برای انسان قابل درک نیست. در حوزه‌هایی مانند فقه و کلام که استدلال و مبانی هر حکم از اهمیتی والا برخوردار است، نبود شفافیت در مدل‌های زبانی بزرگ یک چالش جدی به شمار می‌رود. کاربران باید بتوانند دلیل و منبع هر پاسخ را ردیابی کنند و از درستی استدلال سامانه اطمینان حاصل نمایند. توسعه روش‌های هوش مصنوعی قابل تفسیر که بتوانند مسیر استنتاج مدل‌های زبانی بزرگ را به‌صورت شفاف ارائه دهند، یک مسیر پژوهشی مهم است (Mohammad et al., 2024). این امر شامل ارائه شواهد متنی، ارجاعات به گراف دانش و نمایش مسیرهای استنتاجی می‌شود.

- پیچیدگی‌های زبانی و فرهنگی: زبان عربی کلاسیک، با ساختار صرفی و نحوی پیچیده و همچنین تنوع گویش‌ها و لهجه‌ها، چالش‌های خاصی را برای مدل‌های زبانی بزرگ ایجاد می‌کند. افزون بر این، تفاوت‌های فرهنگی و مذهبی در فهم و تفسیر متون اسلامی، نیازمند آن است که مدل‌های زبانی بزرگ قادر به درک این ظرافت‌ها باشند. مدل‌های زبانی عمومی ممکن است در درک این پیچیدگی‌ها با مشکل مواجه شوند و نیاز به بومی‌سازی و تنظیم دقیق بر روی داده‌های تخصصی دارند. توسعه مدل‌های زبانی بزرگ چندزبانه که به‌طور خاص برای زبان‌های متون اسلامی بهینه‌سازی شده‌اند، یک مسیر پژوهشی فعال است.
- ملاحظات امنیتی و حریم خصوصی: استفاده از مدل‌های زبانی بزرگ و گراف‌های دانش در سامانه‌های مدیریت دانش متون اسلامی ممکن است ملاحظات امنیتی و حریم خصوصی را نیز به همراه داشته باشد، بویژه اگر داده‌های حساس یا شخصی در فرآیند آموزش یا استنتاج دخیل باشند. اطمینان از امنیت داده‌ها، جلوگیری از نشت اطلاعات و رعایت اصول حریم خصوصی در طراحی و پیاده‌سازی این سامانه‌ها از اهمیتی والا برخوردار است.

چالش‌های فنی بنیادین، شکاف‌های پژوهشی و روش‌شناسی‌های ارزیابی

با وجود پیشرفت‌های چشمگیر در همگرایی هوش مصنوعی، وب معنایی و مدل‌های زبانی بزرگ در سامانه‌های مدیریت دانش متون اسلامی، تحلیل نقادانه ما نشان می‌دهد که چندین چالش فنی بنیادین همچنان مانع از دستیابی به سامانه‌های کاملاً قابل اعتماد و با کارایی بالا می‌شوند. این چالش‌ها که در بخش‌های پیشین به‌صورت پراکنده به آنها اشاره شد، در این بخش به‌صورت جامع دسته‌بندی و مورد بحث قرار می‌گیرند و براساس آنها، شکاف‌های پژوهشی و روش‌شناسی‌های ارزیابی تبیین می‌شوند.

چالش‌های مربوط به داده‌ها و منابع متنی اسلامی از منظر مهندسی داده

- حجم، تنوع و پیچیدگی داده‌ها: متون اسلامی شامل حجم انبوهی از داده‌ها به

زبان‌های گوناگون (عربی کلاسیک، فارسی، اردو و غیره) و در قالب‌های مختلف (کتاب‌های چاپی، نسخه‌های خطی و غیره) است. گردآوری، دیجیتالی کردن و پیش‌پردازش این داده‌ها، بویژه نسخه‌های خطی و منابع کهن، فرآیندی زمان‌بر، پرهزینه و نیازمند تخصص ژرف در مهندسی داده است. افزون بر این، اطمینان از کیفیت و درستی داده‌های ورودی، به دلیل وجود نسخه‌های گوناگون، اختلافات متنی و نیاز به اعتبارسنجی توسط متخصصان دامنه، یک چالش بزرگ به شمار می‌رود (Bashir et al., 2023). پیچیدگی‌های صرفی، نحوی و بلاغی زبان عربی کلاسیک که بخش عمده‌ای از متون اسلامی به آن نگاشته شده، پردازش خودکار این متون را دشوار می‌سازد. پدیده‌هایی مانند چندمعنایی، همنویسی و ابهامات معنایی، نیازمند توسعه ابزارهای تخصصی پردازش زبان طبیعی هستند که بتوانند این ظرافت‌ها را با دقت مدیریت کنند.

- **کمبود مجموعه داده‌های برچسب‌گذاری شده با کیفیت:** بیشتر پیشرفت‌های اخیر در هوش مصنوعی مدیون مجموعه داده‌های عظیم و برچسب‌گذاری شده هستند. در حوزه متون اسلامی، اگرچه متون خام فراوانی وجود دارد (مانند «کتابخانه جامع نور»، «مکتبه الشاملة» و «پایگاه‌هایی نظیر Sunnah.com»). اما مجموعه داده‌های برچسب‌گذاری شده با کیفیت بالا برای وظایفی نظیر استخراج رابطه، تحلیل معنایی ژرف، شناسایی موجودیت‌های نام‌دار و پرسش و پاسخ، بسیار محدود است. این کمبود داده، مانع از تنظیم دقیق مؤثر مدل‌های زبانی بزرگ و توسعه مدل‌های یادگیری ماشین با کارایی بالا برای درک ظرافت‌های بلاغی و اصطلاحات تخصصی اسلامی می‌شود. گردآوری و برچسب‌گذاری این داده‌ها، با مشارکت خبرگان دامنه، فرآیندی زمان‌بر و پرهزینه است و نیازمند سرمایه‌گذاری‌های کلان پژوهشی می‌باشد (Qamar et al., 2024). توسعه روش‌های یادگیری با نظارت ضعیف و یادگیری بدون نظارت می‌تواند به کاهش این وابستگی کمک کند.
- **جزیره‌ای بودن هستی‌شناسی‌ها و چالش هم‌ترازی:** بسیاری از هستی‌شناسی‌های

توسعه یافته برای متون اسلامی به صورت مستقل و برای یک حوزه یا منبع خاص (مانند قرآن یا حدیث) طراحی شده اند. نبود یک «هستی شناسی مادر» یا یک چهارچوب فدرال که بتواند مفاهیم و روابط مشترک میان فقه، تفسیر، حدیث و تاریخ را به صورت یکپارچه مدیریت کند، به جزیره ای شدن دانش و نبود قابلیت تعامل میان سامانه ها انجامیده است. هم ترازوی هستی شناسی ها و حل تعارضات معنایی میان آنها، بویژه با توجه به وجود مکاتب فکری و تفاسیر متعدد در متون اسلامی، یک چالش مفهومی و فنی بزرگ است (Bendjamaa & Taleb, 2024).

چالش های فنی و الگوریتمی

- پدیده «توهم» در مدل های زبانی بزرگ و الزامات دقت: همان گونه که در بخش پیشین به تفصیل بحث شد، پدیده «توهم» در مدل های زبانی بزرگ، یعنی تولید اطلاعات نادرست یا بی اساس، بزرگترین چالش در حوزه هایی با الزامات دقت بالا است. در متون اسلامی که دقت در نقل قول از منابع و اصالت سند از اهمیتی والا برخوردار است، حتی یک خطای کوچک می تواند پیامدهای جدی به همراه داشته باشد. اگرچه رویکرد تولید تقویت شده با بازیابی به کاهش این پدیده کمک می کند، اما آن را به طور کامل از بین نمی برد. نیاز به توسعه سازوکارهای نیرومندتر برای اعتبارسنجی خود کار پاسخ ها و اطمینان از انطباق آنها با منابع معتبر، یک ضرورت مهندسی است (ArabicNLP Shared Task, 2025). این شامل توسعه معیارهای ارزیابی دقیق و بنچمارک های تخصصی برای سنجش میزان توهم و دقت پاسخ ها است.

- پیچیدگی های اعتبارسنجی سند و دانش رجال: در علوم حدیث، اعتبار یک روایت به زنجیره روایان (سند) بستگی دارد. در حالی که گراف های دانش در نمایش این زنجیره ها موفق بوده اند (Kamran, Abro & Basharat, 2023)، اما تحلیل کیفی روایان (دانش رجال) به صورت هوشمند هنوز در مراحل آغازین است. سامانه های کنونی عمدتاً بر وجود نام راوی تمرکز دارند، اما تحلیل تضادهای

تاریخی در اتصال روایی راویان یا ارزیابی وثاقت آنها براساس نظرات گوناگون رجالی، نیازمند مدل‌های استنتاجی بسیار پیچیده‌تری است. این امر نیازمند ادغام دانش تخصصی علم رجال با توانایی‌های استنتاجی گراف‌های دانش و مدل‌های زبانی بزرگ است.

- **مقیاس‌پذیری و کارایی سامانه‌ها:** با توجه به حجم انبوه متون اسلامی و پیچیدگی گراف دانش، اطمینان از مقیاس‌پذیری و کارایی بالای سامانه‌های مدیریت دانش متون اسلامی یک چالش مهم مهندسی است. این شامل بهینه‌سازی ساختار گراف دانش، انتخاب پایگاه داده گراف دانش مناسب، بهینه‌سازی پرس‌وجوها و استفاده از معماری‌های توزیع‌شده برای پردازش و ذخیره‌سازی داده‌هاست. همچنین، پردازش‌های سنگینی همچون استخراج دانش و استنتاج، نیازمند زیرساخت‌های محاسباتی نیرومند و بهینه‌سازی الگوریتمی هستند.
- **تفسیرپذیری و شفافیت مدل‌ها:** مدل‌های زبانی بزرگ اغلب به‌عنوان «جعبه سیاه» عمل می‌کنند، بدین معنا که فرآیند تصمیم‌گیری آنها برای انسان قابل درک نیست. در حوزه‌هایی مانند فقه و کلام که استدلال و مبانی هر حکم از اهمیتی والا برخوردار است، نبود شفافیت در مدل‌های زبانی بزرگ یک چالش جدی به شمار می‌رود. کاربران باید بتوانند دلیل و منبع هر پاسخ را ردیابی کنند و از درستی استدلال سامانه اطمینان حاصل نمایند. توسعه روش‌های هوش مصنوعی قابل تفسیر که بتوانند مسیر استنتاج مدل‌های زبانی بزرگ را به‌صورت شفاف ارائه دهند، یک مسیر پژوهشی مهم است (Mohammad et al., 2024). این امر شامل ارائه شواهد متنی، ارجاعات به گراف دانش و نمایش مسیرهای استنتاجی می‌شود.

روش‌شناسی‌های ارزیابی و بنچ‌مارک‌ها

ارزیابی دقیق و معتبر سامانه‌های مدیریت دانش متون اسلامی مبتنی بر هوش مصنوعی، یک چالش اساسی است. روش‌شناسی‌های ارزیابی باید بتوانند نه تنها کارایی فنی (مانند

دقت، فراخوانی و معیار افیک) بلکه جنبه‌های کیفی و دامنه‌خاص را نیز بسنجند. این جنبه‌ها شامل موارد زیر است:

- **معیارهای ارزیابی فنی:** برای وظایف پردازش زبان طبیعی مانند شناسایی موجودیت‌های نام‌دار، استخراج رابطه، خلاصه‌سازی و پرسش و پاسخ، معیارهای استاندارد مانند دقت، فراخوانی، معیار افیک و معیار بلئو قابل استفاده هستند. با این حال، برای ارزیابی گراف‌های دانش، معیارهایی مانند کامل بودن، سازگاری و دقت ساختاری نیز باید در نظر گرفته شوند (Xue et al. , 2023).
- **بنچ‌مارک‌های تخصصی دامنه:** کمبود بنچ‌مارک‌های استاندارد و مجموعه داده‌های ارزیابی برچسب‌گذاری شده برای متون اسلامی، مانعی بزرگ در مقایسه عینی و معتبر سامانه‌هاست. توسعه بنچ‌مارک‌هایی که شامل پرسش‌های پیچیده، سناریوهای استنتاجی و وظایف اعتبارسنجی سند باشند، یک ضرورت است (Qamar et al. , 2024). این بنچ‌مارک‌ها باید با همکاری متخصصان دامنه طراحی و اعتبارسنجی شوند.
- **ارزیابی انسانی و خبرگی:** با توجه به ماهیت حساس و پیچیده متون اسلامی، ارزیابی انسانی توسط متخصصان دامنه (عالمان، فقیهان، محدثان) برای سنجش کیفیت معنایی، درستی استنتاجی و انطباق با اصول متنی، ضروری است. این ارزیابی می‌تواند به صورت ناشناس و با استفاده از سنجه ارزیابی استاندارد انجام شود تا از سوگیری جلوگیری شود.

شکاف‌های پژوهشی و نقشه راه آینده

- براساس تحلیل نقادانه جامع فوق، شکاف‌های پژوهشی زیر برای پژوهش‌های آتی در حوزه مدیریت دانش متون اسلامی با بهره‌گیری از هوش مصنوعی شناسایی شده‌اند:
- **توسعه مدل‌های تولید تقویت‌شده با بازیابی ترکیبی و عصبی-معنایی پیشرفته:** پژوهش‌های آتی باید بر توسعه مدل‌های تولید تقویت‌شده با بازیابی متمرکز شوند که به‌طور مؤثرتری گراف‌های دانشی معنایی را با مدل‌های زبانی

بزرگ ترکیب می‌کنند. این مدل‌ها باید قادر باشند نه تنها اطلاعات را بازیابی کنند، بلکه استنتاج‌های پیچیده را نیز براساس منطق گراف دانش انجام دهند و پاسخ‌های خود را با ارجاع دقیق به منابع اصلی ارائه نمایند. هدف نهایی، دستیابی به سامانه‌های عصبی-معنایی است که قدرت یادگیری آماری را با دقت منطقی پیوند می‌دهند و توانایی‌های استنتاجی فراتر از هم‌خوانی الگوهای ساده را ارائه می‌کنند.

- ایجاد هستی‌شناسی‌های فدرال و چهارچوب‌های هم‌ترازی پویا: نیاز مبرم به توسعه چهارچوب‌هایی برای ایجاد هستی‌شناسی‌های فدرال وجود دارد که بتوانند هستی‌شناسی‌های گوناگون علوم اسلامی را به صورت یکپارچه و بدون از دست دادن جزئیات تخصصی هر حوزه مدیریت کنند. این چهارچوب‌ها باید قابلیت هم‌ترازی پویا را داشته باشند تا با تکامل دانش و پیدایش تفاسیر تازه، به‌طور خودکار به‌روزرسانی شوند. این امر نیازمند الگوریتم‌های پیشرفته برای تشخیص تغییرات هستی‌شناسی و تلفیق هوشمند آنهاست.
- هوش مصنوعی رجالی و تحلیل پیشرفته سند: پژوهش‌های آتی باید بر توسعه سامانه‌های هوش مصنوعی متمرکز شوند که قادر به تحلیل پیشرفته زنجیره سند و دانش رجال باشند. این شامل استفاده از گراف‌های دانشی زمانی برای مدل‌سازی دقیق تاریخ زایش و مرگ راویان، مکان‌های دیدار و مسیرهای روایت و همچنین بهره‌گیری از مدل‌های زبانی بزرگ برای تحلیل کیفی نظرات رجالی و ارزیابی وثاقت راویان است. هدف، توسعه سامانه‌هایی است که بتوانند به‌صورت نیمه خودکار، اعتبار احادیث را براساس معیارهای رجالی ارزیابی کنند.
- ایجاد مجموعه داده‌های چندزبانه تخصصی و بنچ‌مارک‌های استاندارد: برای پیشرفت در این حوزه، ایجاد مجموعه داده‌های برجسب‌گذاری شده با کیفیت بالا برای زبان‌های عربی کلاسیک، فارسی و اردو و همچنین توسعه بنچ‌مارک‌های استاندارد برای ارزیابی سامانه‌های مدیریت دانش متون اسلامی، یک ضرورت است. این مجموعه داده‌ها باید شامل اطلاعات غنی معنایی، روابط استنتاجی و

برچسب‌گذاری‌های تخصصی توسط خبرگان دامنه باشند تا امکان آموزش و ارزیابی مدل‌های هوش مصنوعی را فراهم آورند.

- توسعه روش‌های هوش مصنوعی قابل تفسیر برای مدیریت دانش متون اسلامی: پژوهش‌های آتی باید بر توسعه روش‌های هوش مصنوعی قابل تفسیر متمرکز شوند که به کاربران اجازه می‌دهند مسیر استنتاج سامانه را از منبع اصلی تا پاسخ نهایی ردیابی کنند. این امر برای جلب اعتماد کاربران و اطمینان از درستی و اعتبار پاسخ‌ها در حوزه‌های حساس متنی حیاتی است. این شامل تجسم گراف دانش، نمایش استدلال‌های مدل زبانی بزرگ و ارائه شواهد متنی می‌شود.
- چهارچوب‌های هوش مصنوعی مسئولانه و انطباق با الزامات دامنه: توسعه چهارچوب‌های نظری و عملی برای هوش مصنوعی مسئولانه که به‌طور خاص با الزامات و ملاحظات اخلاقی-حقوقی دامنه متون اسلامی همسو باشند، یک مسیر پژوهشی حیاتی است. این شامل طراحی الگوریتم‌هایی است که از سوگیری‌های فرهنگی و مذهبی جلوگیری کنند و تضمین نمایند که خروجی‌های سامانه با اصول و مبانی متنی اسلامی سازگار هستند. این امر نیازمند مشارکت فعال متخصصان دامنه و مهندسان سامانه در فرآیند طراحی و توسعه سامانه‌های مدیریت دانش متون اسلامی است.

نتیجه‌گیری و چشم‌انداز آینده

این مقاله به ارائه یک مرور تحلیلی و نقادانه جامع از همگرایی هوش مصنوعی، وب معنایی و مدل‌های زبانی بزرگ در سامانه‌های مدیریت دانش متون اسلامی در بازه زمانی ۲۰۲۳ تا ۲۰۲۶ پرداخت. آنچه از این بررسی برمی‌آید، آن است که این حوزه از مرحله نمایش ساده دانش به سمت تعامل هوشمند، استنتاج ژرف و تولید محتوای معنایی حرکت کرده است. تلفیق گراف‌های دانش با مدل‌های زبانی بزرگ و استفاده از رویکردهای نوینی همچون تولید تقویت‌شده با بازایی، پتانسیلی بالا را برای حل چالش‌های دیرینه در بازایی اطلاعات متنی اسلامی و سازماندهی دانش نشان داده است.

با این حال، تحلیل نقادانه ما آشکار ساخت که دستیابی به سامانه‌های مدیریت دانش متون اسلامی کاملاً قابل اعتماد و با کارایی بالا، مستلزم غلبه بر چالش‌های فنی بنیادینی است. از جمله این چالش‌ها می‌توان به پدیده «توهم» در مدل‌های زبانی بزرگ، جزیره‌ای بودن هستی‌شناسی‌ها، کمبود داده‌های برچسب‌گذاری شده تخصصی، پیچیدگی‌های اعتبارسنجی سند، نیاز به توضیح‌پذیری مدل‌ها و مقیاس‌پذیری سامانه‌ها اشاره کرد. «توهم» به وضعیتی اطلاق می‌شود که در آن مدل زبانی، پاسخ‌هایی به ظاهر منسجم و معتبر، اما در واقع نادرست یا فاقد پشتوانه دانشی تولید می‌کند؛ مسئله‌ای که در حوزه حساس متون اسلامی می‌تواند به تحریف مفاهیم یا ارائه استنباط‌های غیرقابل اتکا منجر شود. این مسائل نه تنها از جنس محدودیت‌های فنی صرف نیستند، بلکه ریشه در ویژگی‌های منحصربه‌فرد دانش متنی اسلامی و الزامات دامنه‌ای آن دارند. هرچند راهکارهایی نظیر اتصال مدل‌ها به منابع دانشی معتبر، بهره‌گیری از گراف‌های دانش و طراحی سازوکارهای ارزیابی صحت پاسخ‌ها برای کاهش این پدیده پیشنهاد شده است، پرداخت تفصیلی به ابعاد، علل و راهبردهای مهار آن، نیازمند پژوهشی مستقل است و از حوصله این مقاله خارج است.

ما بر این باوریم که آینده مدیریت دانش متون اسلامی در گرو توسعه سامانه‌های «عصبی-معنایی» است؛ سامانه‌هایی که قدرت یادگیری آماری مدل‌های زبانی بزرگ را با دقت منطقی و قابلیت استنتاج گراف‌های دانش پیوند می‌دهند. در چنین چهارچوبی، مدل‌های زبانی بزرگ می‌توانند نقش پیمانه‌های درک زبان طبیعی را ایفا کنند، در حالی که گراف‌های دانش به‌عنوان منبعی ساختاریافته و قابل اعتماد برای دانش و مکانیسمی برای استنتاج منطقی عمل می‌نمایند. رویکرد تولید تقویت‌شده با بازیابی گامی اولیه و امیدوارکننده در این مسیر است، اما برای دستیابی به سامانه‌هایی با توانایی استنتاجی فراتر از بازیابی صرف، نیاز به پژوهش‌های بنیادین بیشتری است.

تحقق این چشم‌انداز، نیازمند همکاری نزدیک و پیوسته میان متخصصان علوم کامپیوتر، مهندسان نرم‌افزار، زبان‌شناسان و پژوهشگران علوم اسلامی است. این همکاری میان‌رشته‌ای برای تضمین دقت، اصالت، اعتبار و انطباق با الزامات دامنه، امری

ناگزیر به نظر می‌رسد. بویژه، مشارکت فعال خبرگان دامنه در فرآیندهای برچسب‌گذاری داده‌ها، اعتبارسنجی خروجی‌ها و طراحی هستی‌شناسی‌ها، نقشی تعیین‌کننده در موفقیت سامانه‌های آتی خواهد داشت.

چشم‌انداز آینده مدیریت دانش متون اسلامی با ظهور فناوری‌های هوش مصنوعی، روشن و سرشار از پتانسیل است. پژوهش‌های آتی می‌توانند بر محورهای زیر متمرکز شوند: نخست، توسعه و پیاده‌سازی نمونه‌های اولیه سامانه‌های مدیریت دانش متون اسلامی مبتنی بر رویکردهای عصبی-معنایی، همراه با ارزیابی دقیق کارایی آنها با استفاده از معیارهای تخصصی و بنچ‌مارک‌های استاندارد. دوم، گسترش این سامانه‌ها به حوزه‌های دیگر دانش اسلامی مانند تاریخ اسلام، سیره نبوی، اخلاق و عرفان. سوم، توسعه ابزارهای خودکار برای ساخت و نگهداری هستی‌شناسی‌ها و گراف‌های دانش با بهره‌گیری از مدل‌های زبانی بزرگ. چهارم، بهبود روش‌های تفسیرپذیری مدل‌های زبانی بزرگ در این حوزه، به نحوی که کاربران بتوانند به راحتی مبانی استنتاجی پاسخ‌ها را ردیابی کنند. پنجم، ایجاد مجموعه داده‌های مرجع و بنچ‌مارک‌های استاندارد که امکان مقایسه عینی و معتبر سامانه‌های گوناگون را فراهم آورد. و ششم، تدوین چهارچوب‌های اخلاقی و حقوقی برای کاربرد مسئولانه هوش مصنوعی در مدیریت دانش متون اسلامی.

با تحقق این چشم‌انداز، می‌توان به سوی عصری تازه در دسترسی، فهم و تعامل هوشمند با میراث غنی دانش متنی اسلامی گام برداشت؛ عصری که نه تنها برای پژوهشگران و متخصصان، بلکه برای عموم مردم نیز سودمند خواهد بود و به ترویج فهمی درست و ژرف از آموزه‌های اسلامی یاری خواهد رساند. بی‌گمان، مسیر پیش رو طولانی و پرفرازونشیب است، اما پتانسیل تحول‌آفرین این فناوری‌ها، سرمایه‌گذاری پیوسته و همکاری‌های گسترده میان‌رشته‌ای را توجیه می‌کند.

فهرست منابع

- Alshammari, I. K. , Atwell, E. , & Alsalka, M. A. (2024). Linking Quran and Hadith topics in an ontology using word embeddings and Cellfie plugin. Proceedings of the 7th International Conference on Natural Language and Speech Processing, 449-455 .
- ArabicNLP Shared Task. (2025). Retrieval-augmented LLMs for Qur'anic and Hadith content verification. Proceedings of ArabicNLP 2025 .
- Bashir, M. H. , et al. (2023). Arabic natural language processing for Qur'anic research: A systematic review. Artificial Intelligence Review, 56(7), 6801-6854 .
- Bendjamaa, F. , & Taleb, N. (2024). OntoDin: An Islamic ontology of Quran and Hadith. The International Arab Journal of Information Technology, 21(5), 773-784 .
- Berners-Lee, T., Hendler, J. , & Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 34-43 .
- Brown, T. B. , Mann, B. , Ryder, N. , Subbiah, M. , Kaplan, J. , Dhariwal, P. , . . . & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901 .
- European Commission. (2019). Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Feng, X. , Wu, X. , & Meng, H. (2024). Ontology-grounded Automatic Knowledge Graph Construction by LLM under Wikidata schema. In CEUR Workshop Proceedings (Vol. 3841, pp. 117-135). (CEUR Workshop Proceedings). <https://ceur-ws.org/Vol-3841/Paper19.pdf> .

- Haji Esmaeili, M. , & Haji Molla Mirzaee, H. (2021). Presenting a model of knowledge management in the Islamic value system and comparatively comparing it with other common models. *Strategic Management of Organizational Knowledge*, 3(4), 71-100. doi: 10. 47176/smok. 2020. 1228 .
- Harrag, Fouzi & Al-Nasser, Abdullah & Al-Musnad, Abdullah & Al-Shaya, Rayan. (2020). Quran Intelligent Ontology Construction Approach Using Association Rules Mining. 10. 48550/arXiv. 2008. 03232 .
- Hogan, A. , Blomqvist, E. , Cochez, M. , D'amato, C. , De Melo, G. , Gutierrez, C. , Kirrane, S. , Gayo, J. E. L. , Navigli, R. , Neumaier, S. , Ngomo, A. -C. N. , Polleres, A. , Rashid, S. M. , Rula, A. , Schmelzeisen, L. , Sequeda, J. , Staab, S. , & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4), 1–37. <https://doi.org/10.1145/3447772> .
- Ji, S. , Pan, S. , Cambria, E. , Martinen, P. , & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 494-514. <https://doi.org/10.1109/TNNLS.2021.3070843> .
- Kamran, A. B. , Abro, B. , & Basharat, A. (2023). SemanticHadith: An ontology-driven knowledge graph for the hadith corpus. *Journal of Web Semantics*, 78, 100797 .
- Kamran, A. B. , Butt, N. A. , & Basharat, A. (2026). Semantic enrichment of Hadith corpus—Knowledge graph generation from Islamic text. *Journal of Al-Quds Open University for Research and Studies*. Advance online publication. <https://doi.org/10.1177/22104968261431425> .
- Khan, H. U. , Saqlain, S. M. , Shoaib, M. , & Sher, M. (2013). Ontology based semantic search in Holy Quran. *International Journal of Future Computer and Communications*, 2(6), 570–575. <https://doi.org/10.7763/IJFCC.2013.V2.229> .
- Kommineni, V. K. , König-Ries, B. , & Samuel, S. (2024). From human experts to machines: An LLM supported approach to ontology and knowledge graph

- construction [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2403.08345>.
- Malhas, R. , Mansour, W. , & Elsayed, T. (2023). Qur'an QA 2023 shared task: Overview of passage retrieval and reading comprehension tasks over the Holy Qur'an. In H. Sawaf, S. El-Beltagy, W. Zaghoulani, W. Magdy, A. Abdelali, N. Tomeh, I. Abu Farha, N. Habash, S. Khalifa, A. Keleg, H. Haddad, I. Zitouni, K. Mrini, & R. Almatham (Eds.), Proceedings of ArabicNLP 2023 (pp. 690–701). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.arabnlp-1.76>.
- Mohammad, Y. , Nachouki, M. , & Mohamed, E. A. (2024). Knowledge management systems in business management using knowledge graphs and semantic technologies. International Journal of Knowledge Management, 21(1). <https://doi.org/10.4018/IJKM.369121>.
- National Institute of Standards and Technology (NIST). (2023). Artificial intelligence risk management framework (AI RMF 1.0). U. S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- OECD. (2019). OECD principles on artificial intelligence. Organisation for Economic Co-operation and Development. <https://oecd.ai/en/ai-principles>
- Pan, J. Z. , Razniewski, S. , Kalo, J. -C. , Singhanian, S. , Chen, J. , Dietze, S. , Jabeen, H. , Omeliyanenko, J. , Zhang, W. , Lissandrini, M. , Biswas, R. , de Melo, G. , Bonifati, A. , Vakaj, E. , Dragoni, M. , & Graux, D. (2023). Large language models and knowledge graphs: Opportunities and challenges [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2308.06374>.
- Peng, H. , & Yang, W. (2024). Knowledge graph construction method for commercial aircraft fault diagnosis based on logic diagram model. Aerospace, 11(9), 773.
- Qamar, F. , Latif, S. , & Latif, R. (2024). A benchmark dataset with larger context for non-factoid question answering over Islamic text [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2409.09844>.

- Ta'a, A. , ShaariaHj Ashari, M. S., Che Mohamed Arif, A. S. , Abu Bakar, M. S., & Mahmad, M. A. (2021). Semantic Approach for Al-Quran Knowledge Representation in Islamic Knowledge Management Framework. Journal of Quranic Sciences and Research, 2(1), 35-46 .
- Uriawan, Wisnu & Hamza, Aria & Nuralim, Ade & Purnama, Adi & Yunus, Ahmad & Putri, Anissya. (2025). Implementing a Sharia Chatbot as a Consultation Medium for Questions About Islam. 10. 48550/arXiv. 2512. 16644 .
- Xue, B. , & Zou, L. (2023). Knowledge graph quality management: A comprehensive survey. IEEE Transactions on Knowledge and Data Engineering, 35(5), 4969–4988. <https://doi.org/10.1109/TKDE.2022.3150080> .