

A Framework Based on Crowdsourcing for an Analytical Investigation of Qur'anic Concepts Using Natural Language Processing: Solutions and Limitations

Saeed Abasinia  / Assistant Professor at the Department of Islamic Theology and Teachings, The Qur'an and Hadith Sciences, Islamic Azad University, West Tehran Branch, Tehran, Iran
dr.abasinya@iau.ac.ir

Received: 2025/02/03 - Accepted: 2025/04/19

Abstract

The application of Natural Language Processing (NLP) techniques to illustrate the concepts and meanings of the Qur'anic texts not only helps the researchers of Qur'anic studies to present comprehensive analyses for conveying the authentic messages of this divine text, but also significantly makes a cultural and scientific contribution to a deeper understanding and development of Islamic and spiritual knowledge and values in society. Aiming at connecting the two important fields of "Qur'anic Sciences" and "Natural Language Processing" for an in-depth analysis of the concepts and meanings of the Qur'an, this paper presents a strategic framework based on crowdsourcing. The proposed strategic framework provides solutions for semantic analysis of the Qur'anic text to understand deeper meanings and semantic relationships of its various components using advanced techniques of NLP. It also enables us to examine the differences and similarities in meaning and identify patterns and semantic connections in the Qur'an, which leads to the strategic development of religious theories and interpretation and helps to advance knowledge and education in the fields of philosophy and religion. To measure the efficiency and soundness of machine approaches to the Holy Qur'an, several key evaluation criteria and indicators that facilitate the analysis and comparison of results derived from NLP-based solutions are introduced in this paper.

Keywords: structural analysis, natural language processing, crowdsourcing, pattern identification, semantic analysis.

نوع مقاله: پژوهشی

ارائه چهار چوبی مبتنی بر جمع‌سپاری برای واکاوی تحلیلی مفاهیم قرآنی با استفاده از پردازش زبان طبیعی: راهکار تا محدودیت‌ها

سعید عباسی‌نیا  / استادیار گروه الهیات و معارف اسلامی، علوم قرآن وحديث، دانشگاه آزاد اسلامی، واحد تهران غرب، تهران، ایران
دریافت: ۱۴۰۳/۱۱/۱۵ - پذیرش: ۱۴۰۴/۰۱/۳۰
dr.abasiny@iau.ac.ir

چکیده

استفاده از فنون پردازش زبان طبیعی (NLP) برای تجسم مفاهیم و معانی متن قرآن، نه تنها به پژوهشگران حوزه علوم قرآنی کمک می‌کند تا تحلیل‌های قدرتمندتری برای انتقال مفاهیم اصیل این متن الهی ارائه دهند، بلکه از نظر فرهنگی و علمی به فهم عمیق‌تر و توسعه دانش و ارزش‌های اسلامی و معنوی جامعه کمک قابل توجهی می‌کند. این پژوهش یک چهارچوب راهبردی مبتنی بر جمع‌سپاری به‌منظور ارتباط میان دو حوزه مهم «علوم قرآنی» و «پردازش زبان طبیعی» برای تحلیل عمیق مفاهیم و معانی قرآن ارائه می‌دهد. چهارچوب راهبردی پیشنهادی، راهکار تحلیل معنایی متن قرآن برای استخراج مفاهیم و معانی عمیق‌تر و ارتباطات معنایی بین عناصر مختلف آن را با استفاده از روش‌های پیشرفته NLP ارائه می‌کند و قابلیت را برای بررسی تفاوت‌ها و تطابق‌های معانی و شناسایی الگوها و ارتباطات معنایی در قرآن فراهم می‌آورد که به توسعه راهبردی نظریه‌های دینی و تفسیر منجر می‌شود و به ارتقای دانش و آموزش در زمینه‌های فلسفه و دین کمک می‌کند. برای ارزیابی و سنجش کارایی و صحت عملکرد رویکردهای ماشینی قرآن کریم چند معیار و شاخص کلیدی ارزیابی معرفی می‌کنیم که امکان تحلیل و مقایسه نتایج راهکارهای مبتنی بر NLP را فراهم می‌کنند.

کلیدواژه‌ها: واکاوی ساختاری، پردازش زبان طبیعی، جمع‌سپاری، شناسایی الگوها، تحلیل معنایی.

مقدمه

تحلیل عمیق متون دینی، به‌ویژه قرآن، به دلیل پیچیدگی‌های ساختاری و معنایی آن با چالش‌های متعددی روبه‌روست. تنوع زبانی، ساختار پیچیده متون و حجم عظیم اطلاعات موجود، موانعی بر سر راه درک دقیق و کامل این آثار ارزشمندند. مقایسه و تحلیل جامع تمام متون مذهبی، برای یک انسان کاری تقریباً غیرممکن و برای رایانه‌ها نیز وظیفه‌ای پیچیده به‌شمار می‌رود. استفاده از پردازش زبان طبیعی (NLP) در تحلیل مفاهیم قرآنی و ارائه ابزارهای قدرتمند می‌تواند بخشی از یک راهبرد فرهنگی جامع برای ترویج فهم و درک عمیق‌تر از قرآن کریم در بین مردم باشد.

پردازش زبان طبیعی به‌عنوان یکی از زمینه‌های نوظهور و چشم‌اندازهای شگفت‌انگیز هوش مصنوعی، ابزاری قدرتمند برای تحلیل متون ارائه می‌دهد. NLP با بهره‌گیری از تکنیک‌های پیشرفته محاسباتی و یادگیری ماشین، قادر به درک مفاهیم و اطلاعات موجود در متون زبان طبیعی است (Hupkes et al, 2023; Patil et al, 2023). این قابلیت فرصتی بی‌نظیر برای تحلیل عمیق‌تر و گسترده‌تر متون مذهبی و دستیابی به الگوهای عمیق نهفته در این متون فراهم می‌کند (آذری و دیگران، ۱۳۹۹).

در سال‌های اخیر روش‌های مختلفی با رویکردهای سنتی برای بسیاری از مراحل که شامل شناسایی آیات و واژگان و طبقه‌بندی موضوعی (معجم موضوعی) آنهاست، به کار گرفته شده است (بهجت‌پور، ۲۰۱۴م؛ ایازی، ۱۳۸۹؛ فریادرس، ۲۰۱۵م)؛ اما این روش‌ها کمتر، از توانایی‌های پیشرفته NLP بهره برده‌اند. رویکردهای سنتی معمولاً از تکنیک‌های ساده‌تری مانند تجزیه و تحلیل مبتنی بر واژگان روش‌های دستی برای تفسیر و طبقه‌بندی مفاهیم استفاده می‌کنند. با بهره‌گیری از مدل‌های پیشرفته NLP، مانند بردارواژگان و GloVe، که قادر به نگاشت کلمات به فضای برداری هستند و روابط معنایی بین کلمات را به‌خوبی منعکس می‌کنند، می‌توانیم به تحلیل دقیق‌تر و جامع‌تری از متون قرآنی دست یابیم. این رویکرد به تخصص‌های مختلف، از جمله علوم رایانه، زبان‌شناسی و علوم قرآنی و اسلامی نیاز دارد تا بتواند به‌طور مؤثر به شناسایی و تحلیل مفاهیم و الگوهای پیچیده در متون قرآنی بپردازد. این روش‌ها، اگرچه در زمان خود کارآمد بوده‌اند، توانایی درک و تحلیل عمیق‌تر و دقیق‌تر متون قرآنی را ندارند. به همین دلیل سختی این پژوهش و پیچیدگی‌های موجود در تحلیل متون قرآنی با استفاده از NLP، یک رویکرد جمع‌سپاری میان‌رشته‌ای را می‌طلبد. همچنین تدوین راهبردهای فرهنگی مبتنی بر تحلیل دقیق متون دینی نیازمند رویکردی میان‌رشته‌ای است که از تخصص‌های مختلف علوم رایانه، زبان‌شناسی، معناشناسی و علوم اسلامی بهره‌مند باشد. چهارچوب پیشنهادی، همکاری و تعامل میان متخصصان مختلف را فراهم می‌کند که به درک بهتر و جامع‌تری از مفاهیم قرآنی منجر می‌شود. اهداف اصلی این پژوهش عبارت‌اند از:

- ارائه چهارچوبی نوآورانه برای تحلیل قرآن با استفاده از NLP؛

- راهبردی نوین برای استخراج مفاهیم و معانی عمیق از متن قرآن؛

- راهکاری برای بررسی ارتباطات معنایی بین آیه‌های مختلف قرآن.

با انجام این پژوهش امیدواریم که گامی در جهت درک عمیق‌تر و جامع‌تر قرآن کریم برداریم و به توسعه راهبردی نظریه‌های فلسفی و دینی وابسته به قرآن کمک کنیم.

۱. پیشینه پژوهش

در سال‌های اخیر شاهد پیشرفت‌های چشمگیری در زمینه پردازش زبان طبیعی و یادگیری ماشین بوده‌ایم. این پیشرفت‌ها به توسعه مدل‌های NLP قدرتمندی منجر شده است که می‌توانند برای انجام کارهای مختلفی مانند ترجمه ماشینی، تشخیص گفتار و استخراج اطلاعات از متن استفاده شوند. علاوه بر این، رشد فناوری‌های جدید به‌مانند شبکه‌های عصبی عمیق به افزایش دقت و کارایی مدل‌های NLP کمک کرده است. این پیشرفت‌ها زمینه را برای تحقیقات جدید در زمینه NLP مهیا کرده و امکان انجام کارهایی را که قبلاً غیرممکن به‌نظر می‌رسید، فراهم ساخته است.

بردارهای جمله ابزاری قدرتمند برای نمایش معنای جمله به‌صورت عددی‌اند. این بردارها می‌توانند برای انجام کارهایی مانند ترجمه ماشینی، تجزیه و تحلیل احساسات، و پاسخ به سوالات استفاده شوند. با اینکه تحقیقات زیادی در زمینه بردارهای جمله برای زبان‌های انگلیسی و اروپایی انجام شده است، هنوز نیاز زیادی به تحقیقات بیشتر در زمینه بردارهای جمله برای زبان عربی وجود دارد. این امر به دلیل ساختار پیچیده زبان عربی و کمبود منابع آموزشی برای مدل‌های NLP است. این پژوهش در زمینه بردارهای جمله برای متون عربی می‌تواند به پر کردن این فاصله و بهبود درک ما از زبان عربی کمک کند. در ادامه، برخی از پژوهش‌های مرتبط را بررسی می‌کنیم.

جدول ۱: کارهای مرتبط

مرجع	چکیده
Khadangi et al,) (2022	در این مقاله بر اساس سه روش word2vec، tf-idf و هم‌نشینی ریشه در آیات، تشابه ریشه‌های قرآنی محاسبه شده است؛ سپس میزان تشابه مفاهیم موجود در سوره‌ها با یکدیگر محاسبه و با حالت تصادفی مقایسه می‌شود. نتایج نشان می‌دهند که سوره‌های مورد مطالعه انسجام درونی مفاهیم را به‌گونه‌ای حفظ می‌کنند که در یک موضوع یا چند موضوع مرتبط با یکدیگر شکل گرفته‌اند. علاوه بر این، بررسی شباهت بخش اول و بدنه هر سوره نشان می‌دهد که ساختار مقدمه و تبیین با روش طراحی‌شده برای بسیاری از سوره‌ها صادق است.
Arief Fatchul et) (al, 2021	در این مطالعه، یک روش تگ‌گذاری POS کارآمد برای متن عربی با استفاده از الگوریتم بهینه‌سازی کلونی زنبور بررسی می‌شود. مسئله تگ‌گذاری POS به‌عنوان یک گراف نشان داده می‌شود و روشی جدید برای اختصاص ارزش‌گذاری انتقال به هر برجسب کلاس‌واژه‌ای که ممکن است احتمالی نباشد، پیشنهاد می‌شود؛ سپس زنبورها به‌دنبال بهترین مسیر حل مسئله می‌گردند. مجموعه داده مورد استفاده در این مطالعه، از مجموعه داده ترجمه‌شده قرآن تشکیل شده است که شامل ۱۵۰ جمله کامل ساده به‌عنوان یک دسته داده آسان، پنجاه جمله با بیش از یک S/P/O/K به‌عنوان یک دسته داده متوسط و پنجاه آیه قرآن انتخاب‌شده به‌عنوان یک دسته داده دشوار است.
Auhoud et al,) (2013	اعمال تکنیک‌های پردازش زبان طبیعی بر روی متن قرآن برای استخراج حروف تجویدی، یک مرحله مهم از استخراج اطلاعات (IE) است. این تحقیق روش‌های NLP را در متن قرآنی با استفاده از GATE، یک محیط NLP انعطاف‌پذیر ارائه می‌کند. GATE برای این تحقیق به‌منظور ساخت برنامه استفاده می‌شود که مجموعه داده‌های متنی قرآن است. برنامه ساخته‌شده با استفاده از معیارهای دقت و بازخوانی ارزیابی می‌شود.

در این پژوهش، مازول پردازش زبان طبیعی برای متن به گفتار عربی معرفی شده است که HQB-ATTS نامگذاری شد. سیستم HQB-ATTS با استفاده از برخی از قوانین تجویدی که در عربی استاندارد استفاده می‌شود، پیاده‌سازی شده است.	(Bana et al, 2016)
این مقاله از یادگیری انتقالی استفاده می‌کند تا اطلاعات از متن‌های زبان موجود (متن قرآن) را به مسئله درک خواندن ماشینی انتقال دهد. برای اثبات کارایی روش پیشنهادی ما، آزمایش‌ها روی مجموعه‌داده Quran QA 2022 انجام شده است. در انجام آزمایش‌ها، ما نشان می‌دهیم که روش پیشنهادی ما نسبت به روش‌های موجود در این حوزه بهبودهای معناداری را ایجاد کرده است.	(Damith et al, 2022)
این مقاله روش‌شناسی عینی و طراحی یادگیری ماشینی تفسیر قرآن را با استفاده از پیشرفت‌های فناوری علوم داده پیشنهاد می‌کند که از فناوری هوش مصنوعی شامل یادگیری ماشین و پردازش زبان طبیعی استفاده کرده است.	(Dardya et al, 2021)
در این مقاله با استفاده از روش‌های پردازش زبان طبیعی، موضوع همسانی سوره‌های قرآنی بررسی می‌شود. در این راستا، بر اساس دو روش word2vec و همراهی ریشه در آیات، تشابه ریشه‌های قرآنی محاسبه شده است؛ سپس میزان تشابه عنوان سوره‌ها با مفاهیم درون سوره مورد بررسی قرار می‌گیرد. نتایج نشان می‌دهند که انتخاب عنوان سوره مبتنی بر منطق عقلی است و این امر از سوی عموم مردم عادی اوایل دوران اسلامی امکان پذیر نبوده است.	(Ehsan et al, 2018)
این مقاله یک مدل دسته‌بندی چند برچسبی مجموعه‌ای را ارائه می‌دهد که به‌طور خودکار آیات قرآن را بر اساس مضامین/موضوعات شناسایی و طبقه‌بندی می‌کند. آیات با استفاده از تکنیک‌های فرکانس معکوس فرکانس سند (TF-IDF) و word2vec به بردارهای ویژگی تبدیل می‌شوند. Word2vec برای در نظر گرفتن جنبه معنایی کلمات قرآنی و بهبود عملکرد استفاده می‌شود. همچنین آنها بر روی یک مجموعه کلاسیک عربی جمع‌آوری شده از دویست میلیون کلمه آموزش می‌بینند؛ سپس تکنیک طبقه‌بندی چندبرچسبی مرتبط با اینتری با استفاده از سه طبقه‌بندی کننده مختلف اعمال می‌شود؛ رگرسیون لجستیک، ماشین بردار پشتیبان و جنگل تصادفی، که آیات را به ۳۹۳ موضوع/برچسب دسته‌بندی می‌کند.	(Ensaf Hussein et al, 2021)
این مقاله الگوریتم‌ها و رویکردهایی را ارائه می‌کند که برای استخراج داده‌های انبوه از منابع متنی عظیم عربی، از جمله قرآن کریم و کتاب‌های مرتبط با آن، طراحی شده‌اند. مجموعه‌داده‌ی نهایی در برگیرنده‌ی اکسل و فرمت رکورد‌های پایگاه داده نمایش داده می‌شود. همچنین این مقاله مدلهایی از مجموعه‌داده را در تمام سطوح ارائه می‌کند. مجموعه‌داده‌های قرآنی ارائه‌شده در این مقاله به‌گونه‌ای طراحی شده است که برای پایگاه داده، داده کاوی، متن کاوی و کاربردهای هوش مصنوعی مناسب باشد.	(Mohamed Osman et al, 2015)
در این مقاله تکنیک‌های مدل‌سازی برای تنظیم چهارچوبی برای جست‌وجوی معنایی در قرآن کریم مورد بررسی قرار گرفته است. از آنجاکه قرآن کریم کلام خداوند است، معانی آن نامحدود است. در این مقاله، کلمات سوره یوسف از قرآن کریم بر اساس تکنیک‌های مدل‌سازی موضوعی به‌عنوان مطالعه موردی تحلیل شده است. تکنیک مدل‌سازی موضوعی تخصیص دیریکله پنهان (LDA) در این مقاله در دو ساختار حزب و آیات استفاده شده است.	(Mohammad and Mohammad, 2015)
این مقاله به بررسی تلاش‌های مختلف در زمینه NLP قرآنی می‌پردازد که به‌عنوان خلاصه‌ای از آثار (ابزار)، مجموعه‌داده‌ها، رویکردها) ارائه می‌شود که گستره‌ای از تحلیل ریخت‌شناسی خودکار تا تصحیح تلاوت قرآن از طریق تشخیص گفتار را پوشش می‌دهد. در صورت لزوم، چندین رویکرد برای چندین کار مورد بحث قرار گرفته است. در نهایت، مسیرهای تحقیقاتی آینده در این زمینه را بیان می‌کنیم.	(Muhammad Huzaifa et al, 2023)
این تحقیق در پی شناخت کلمه «التور» انجام شده است («علم و هدایت و هسته و ایمان» و آیه ۳۵ سوره نور). این آیه به این دلیل انتخاب شده است که در مورد سلطنت الله تعالی و نور خدا صحبت می‌کند. در نهایت، کارایی این پژوهش با استفاده الگوریتم «خوشه‌بندی تفریقی و خوشه‌بندی تفریق وزنی» اندازه‌گیری شده و از زبان Matlab (۲۰۱۳) برای دستیابی به جنبه کاربردی تحقیق استفاده کرده است.	(Nima and Nima, 2020)
این مقاله به تشریح ساخت یک روش تشخیص موجودیت نام‌دار مبتنی بر قاعده برای استخراج موجودیت‌های موجود در ترجمه انگلیسی به معنای متن قرآن و ارزیابی عملکرد آن می‌پردازد.	(Shasha Arzila and Saidah, 2022)

این پژوهش با هدف تحلیل کاربرد روش‌های برنامه‌ریزی زبانی NLP/Neuro Linguistic در ارتقای کیفیت حفظ قرآن در مرکز اسلامی انجام شده است. نتایج تحقیق نشان می‌دهد که اجرای روش برنامه‌ریزی زبانی NLP/عصبی باعث بهبود کیفیت حفظ قرآن در مرکز اسلامی می‌شود.	(Syamsu et al, 2023)
این مقاله یک ابزار جست‌وجوی معنایی قرآن را توسعه داده است که آیات مربوط به سؤال یا درخواست کاربر را پیدا می‌کند. برای دستیابی به این هدف، چندین مدل را بر روی مجموعه داده بزرگی از بیش از سی تفسیر آموزش داد که معمولاً هر تفسیر با یک آیه در قرآن مطابقت دارد و با استفاده از تشابه کسینوس، تانسور تفسیری را به دست آورد که شبیه‌ترین تانسور مورد علاقه است؛ سپس برای نمایه‌سازی آیه مربوطه در قرآن استفاده شد.	(Yasser and Maged, 2023)
در این پژوهش، جاسازی‌های کلمه با تکنیک‌های یادگیری عمیق برای شناسایی آیات قرآن در محتوای متنی عربی بررسی شد. کار پیشنهادی با استفاده از دوازده مدل جاسازی کلمه مختلف با دو طبقه‌بندی کننده محبوب برای طبقه‌بندی باینری، یعنی «شبکه عصبی کانولوشن» و «حافظه کوتاه مدت طولانی (LSTM)» ارزیابی می‌شود. نتایج تجربی، برتری رویکرد پیشنهادی را بر روش‌های سنتی در تشخیص آیات قرآن و متن عربی با دقت ۹۸/۳۳ درصد نشان داد.	(Zineb et al, 2022)
این مقاله به بررسی ابزارهای مختلف بازنمایی کلمات مورد استفاده در محتوای متون قرآن کریم به زبان عربی می‌پردازد که شامل دو شکل اصلی نمایش محلی و بازنمایی توزیع شده با هدف استفاده از آنها در زیرمجموعه‌های مختلف هوش مصنوعی است؛ مانند الگوریتم‌های «یادگیری ماشینی» و «یادگیری عمیق» که به NLP نیاز دارند.	(Zineb et al, 2020)
هدف مقاله تعیین مسیری برای تبیین فناوریانه مفهوم‌یابی از آیات قرآن است. مفهوم‌یابی یکی از مراحل اصلی روش تحقیق موضوعی در قرآن کریم است و در عین حال در سایر مطالعات مرتبط با فهم قرآن می‌تواند به کار گرفته شود. استفاده از دانش متن کاوی رایانشی برای پردازش متون، روشی متداول در فناوری دنیای امروز است. قرآن کریم نیز به عنوان یک متن و زبان، مورد توجه این دانش قرار گرفته است.	(آذری و دیگران، ۲۰۲۰م)
در این مقاله با استفاده از تفسیر راهنما، موضوعات مطرح شده در هریک از سوره‌های قرآن کریم شناسایی و میزان شباهت بین سوره‌ها بر اساس موضوعات مشترک آنها با استفاده از معیاری کمی و روش‌های داده کاوی محاسبه شده است. در نهایت، درختواره شباهت موضوعی سوره‌ها بر اساس معیار کمی مورد نظر ترسیم و با استفاده از آن، سوره‌های قرآن به خوشه‌هایی افراز شده‌اند.	(صوفی و دیگران، ۲۰۱۸م)
در این مقاله الگوریتم‌های مختلف داده کاوی و متن کاوی روی قرآن بررسی شده و از جنبه‌های مختلف با یکدیگر مقایسه شده‌اند.	(یزدانپرستی و اسماعیلی، ۱۳۹۴)
در این مقاله با هدف ارائه رویکردی فناوریانه برای مفهوم‌یابی از آیات قرآن، از دانش متن کاوی و یادگیری ماشینی استفاده می‌کنیم. متن کاوی به عنوان شاخه‌ای از هوش مصنوعی، به استخراج دانش از متن‌های حجیم می‌پردازد. یادگیری ماشینی نیز ابزاری قدرتمند برای تحلیل داده‌ها و کشف الگوهای پنهان در آنهاست.	(آذری و دیگران، ۲۰۲۰م)
در این پژوهش روشی نوین برای کشف الگوهای معنایی در قرآن کریم با استفاده از قوانین انجمنی ارائه می‌شود. در این روش، هر آیه قرآن به مثابه یک سبد خرید در نظر گرفته می‌شود و کلمات آن به عنوان اقلام موجود در آن سبد تلقی می‌شوند. با به کارگیری الگوریتم‌های استخراج قوانین انجمنی، ارتباطات آماری بین کلمات در آیات مختلف قرآن تجزیه و تحلیل می‌شود.	(Aslani and Esmaeili, 2018)

۲. چهارچوب نظری روش پیشنهادی

در این پژوهش، روش‌های مختلف NLP برای تحلیل متن قرآن، تعریف کارها برای جمع‌سپاری، چهارچوب تحلیلی و محدودیت‌های موجود بررسی خواهند شد. برخی از این روش‌ها عبارت‌اند از: الف) تجزیه و تحلیل فراوانی کلمات: بررسی تعداد دفعاتی که کلمات خاص در متن ظاهر می‌شوند؛ ب) تحلیل احساسات: طبقه‌بندی متن به مثبت یا منفی بر اساس لحن و اشارات آن؛ ج) استخراج موضوع: تعیین موضوع اصلی مورد بحث در متن؛ د) مدل‌های Word

Embeddings: نمایش کلمات به‌عنوان بردارهای عددی که روابط معنایی بین آنها را نشان می‌دهد؛ (ه) مدل‌های ترجمه ماشینی: ترجمه متن قرآن به زبان‌های مختلف برای مقایسه و تحلیل.

این پژوهش بر جمع‌آوری و تحلیل متون مذهبی با محوریت قرآن کریم تمرکز دارد. این کار با استفاده از روش‌های جمع‌سپاری و پردازش زبان طبیعی انجام می‌شود. در این پروژه، ما از جمع‌سپاری برای جمع‌آوری داده‌های متنی عربی از منابع مختلف، مانند کتاب‌ها، مقالات و وب‌سایت‌ها استفاده می‌کنیم. این داده‌ها سپس برای آموزش مدل‌های NLP مورد استفاده قرار می‌گیرند. علاوه بر این، جمع‌سپاری به ما امکان می‌دهد که از دانش و تخصص افراد مختلف در سراسر جهان برای بهبود کارمان استفاده کنیم. بخش‌های کلیدی روش پیشنهادی مطابق شکل (۱) عبارت‌اند از:

تعریف اهداف: ابتدا باید اهداف و انتظارات را برای این چهارچوب مشخص کنیم؛ برای مثال، آیا هدف بررسی مفاهیم دینی و فلسفی در قرآن و روایات اسلامی به‌شیوه‌ای جدید و دقیق‌تر است یا اینکه می‌خواهیم موضوعات مشترک بین قرآن و روایات اسلامی را شناسایی کنیم؟

تعیین منابع: برای ایجاد چهارچوب، شما به منابع معتبری نیاز دارید که قرآن و روایات اسلامی را پوشش دهند. این منابع می‌توانند کتب دینی، تفاسیر، روایات و متون دینی معتبر دیگر باشند.

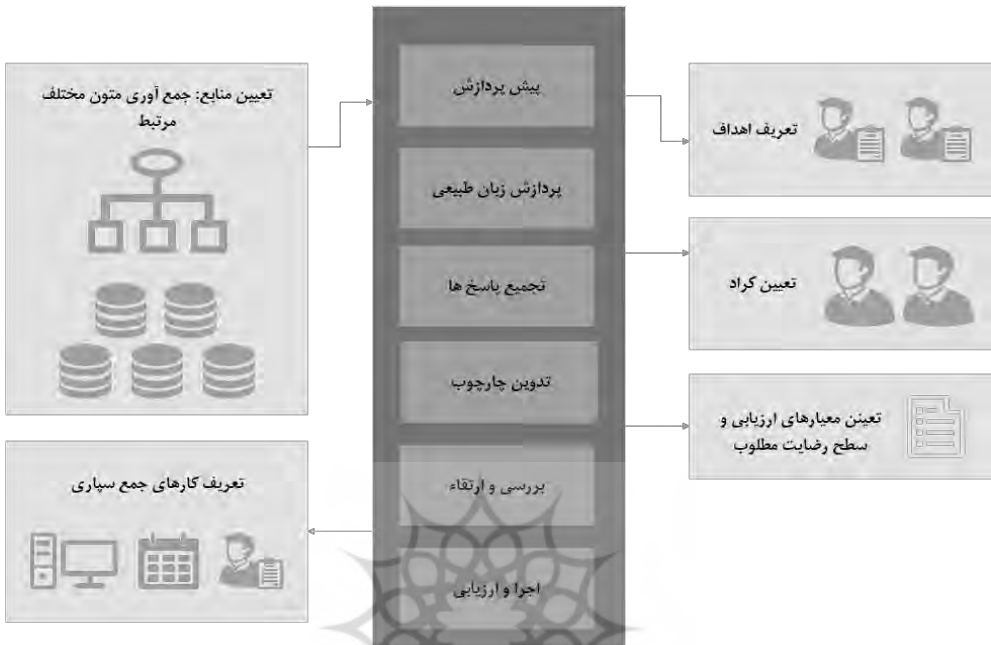
تحلیل محتوا: در این مرحله، متون قرآن و روایات اسلامی را دقیقاً مورد بررسی قرار می‌دهیم تا مفاهیم و موضوعات مختلف شناسایی شود. می‌توان به دنبال تطابقات و تفاوت‌ها در مفاهیم، الگوهای تکراری و ارتباطات میان مفاهیم بود.

تدوین چهارچوب: با توجه به تحلیل محتوا، یک چهارچوب تدوین می‌کنیم که به قرآن پژوهان کمک کند تا مفاهیم و موضوعات مختلف را ارزیابی و تفسیر کنند. این چهارچوب ممکن است شامل دسته‌بندی مفاهیم، ایجاد الگوهای معنایی، تشخیص تطابقات و تفاوت‌ها، و دیگر عناصر مرتبط با تحلیل محتوا باشد.

بررسی و ارتقا: پس از ایجاد چهارچوب اولیه، آن را بازبینی کرده، بهبودهای لازم را اعمال می‌کنیم. از کارشناسان دینی و متخصصان پردازش زبان طبیعی نظرخواهی می‌کنیم تا چهارچوب بهبود یابد. این مرحله بر ضرورت رویکرد میان‌رشته‌ای در این زمینه تأکید دارد و نشان می‌دهد که تعامل میان تخصص‌های مختلف می‌تواند به تدوین و اجرای راهبردهای فرهنگی مؤثر، که بر پایه تحلیل دقیق متون دینی استوار است، منجر شود.

اجرا و ارزیابی: چهارچوب خود را بر روی داده‌ها و متون مورد نظر اجرا و نتایج را ارزیابی می‌کنیم. آیا چهارچوب به تحلیل و تفسیر مفاهیم دینی و فلسفی در قرآن و روایات اسلامی کمک کرده است؟ آیا موجب شناخت بهتر و عمیق‌تر از مفاهیم شده است؟ همچنین ضروری است که در هر مرحله از ایجاد چهارچوب، با تیمی از متخصصان دینی، متخصصان پردازش زبان طبیعی و دیگر افرادی که در زمینه مورد نظر تخصص دارند، همکاری کنیم تا به نتایج بهتری دست یابیم.

شکل ۱: رویکرد پیشنهادی



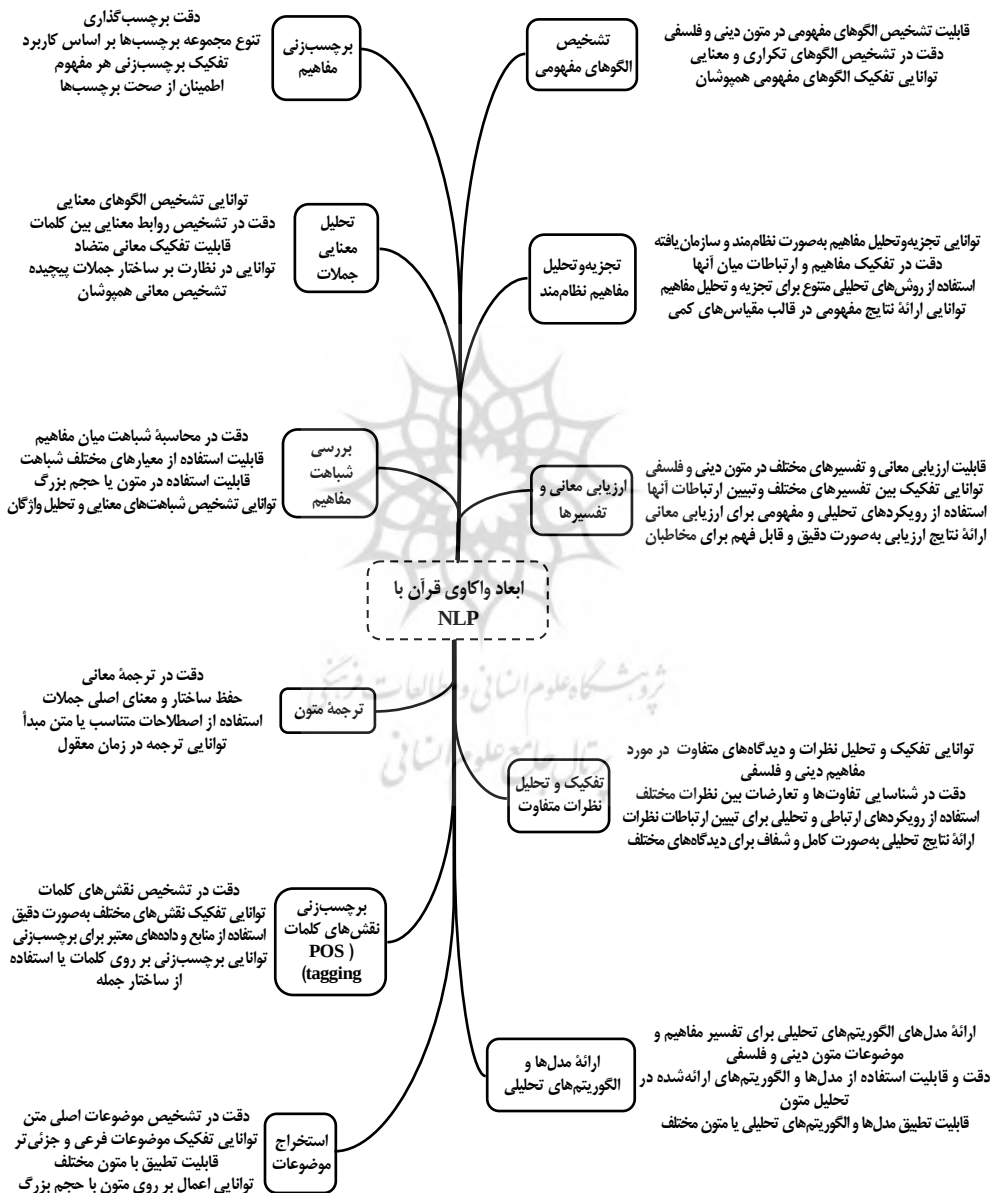
۳. تعریف کارها

چارچوب پیشنهادی ما شامل مراحل متعددی است که هریک به‌طور خاص به ارتقای دقت و انسجام در تحلیل و تفسیر متون قرآنی می‌پردازد. این مراحل شامل پیش‌پردازش متن، استخراج ویژگی، مدل‌سازی، تفسیر و تحلیل هستند. در مرحله پیش‌پردازش متن، آیات قرآنی با استفاده از تکنیک‌های پیشرفته پردازش زبان طبیعی تصحیح و استانداردسازی می‌شوند تا برای تحلیل‌های بعدی آماده شوند. در مرحله استخراج ویژگی، ویژگی‌های کلیدی هر آیه با استفاده از مدل‌های Word Embeddings، مانند Word2Vec و GloVe استخراج می‌شوند تا مفاهیم اصلی هر کلمه در فضای برداری نمایان شود. مرحله مدل‌سازی شامل استفاده از الگوریتم‌های یادگیری ماشین مانند BiLSTM-CRF برای تحلیل نحوی و معنایی جملات است. در نهایت، مراحل تفسیر و تحلیل با بهره‌گیری از تکنیک‌های تحلیل معنایی و شناسایی الگوهای مفهومی، به استخراج و تفسیر دقیق مفاهیم آیات می‌پردازد.

علاوه بر این، می‌توان وظایف مختلف این چهارچوب را به متخصصان مختلف در زمینه‌های زبان عربی، یادگیری ماشین، علوم قرآنی و تفسیر، برون‌سپاری کرد. این رویکرد میان‌رشته‌ای نه تنها سبب ارتقای دقت و انسجام در مفهوم‌یابی آیات می‌شود، بلکه باعث می‌شود که هر بخش از فرایند توسط متخصصی که در آن زمینه تبحر دارد، انجام گیرد. مدل‌های یادگیری ماشین می‌توانند مفاهیم کلیدی آیات را با دقت و انسجام بیشتری نسبت به روش‌های سنتی شناسایی و دسته‌بندی کنند که این امر به تفسیری دقیق‌تر و جامع‌تر منجر می‌شود. در شکل (۲)، انواع کارهایی

که آماده جمع‌سپاری در چهارچوب پیشنهادی هستند، به‌عنوان یک نقشه ذهنی جامع نشان داده شده است. این نقشه ذهنی می‌تواند به‌عنوان راهنمایی برای تقسیم وظایف و برون‌سپاری کارها به متخصصان مربوطه مورد استفاده قرار گیرد که نتیجه آن، تحلیلی دقیق‌تر و منسجم‌تر از متون قرآنی خواهد بود.

شکل ۲: نقشه ذهنی ابعاد واکاوی قرآن با NLP



۴. ابعاد چهارچوب واکاوی تحلیلی مفاهیم قرآنی با استفاده از پردازش زبان طبیعی

چهارچوب واکاوی تحلیلی مفاهیم قرآنی با استفاده از پردازش زبان طبیعی، رویکردی مبتنی بر جمع‌سپاری است که به تحلیل و درک عمیق‌تر مفاهیم قرآنی می‌پردازد. این چهارچوب از چندین مؤلفه کلیدی تشکیل شده است که در زیر به آنها اشاره می‌شود.

۴-۱. رویکرد جمع‌سپاری

رویکرد جمع‌سپاری در این چهارچوب نقش اساسی دارد. این رویکرد بر اساس همکاری چندین فرد یا سیستم برای رسیدن به هدف مشترک، یعنی تحلیل مفاهیم قرآنی، استوار است.

۴-۲. پردازش زبان طبیعی

پردازش زبان طبیعی زیرمجموعه‌ای از هوش مصنوعی است که با تعامل بین رایانه‌ها و زبان طبیعی انسان سروکار دارد. در چهارچوب تحلیل مفاهیم قرآنی، NLP برای پردازش و تحلیل متن قرآن به کار می‌رود تا الگوها، روابط و معانی را در متن شناسایی کند.

۴-۳. تحلیل معنایی

تحلیل محتوا در چهارچوب به کار می‌رود تا الگوها، موضوعات و معانی را در متن قرآن شناسایی کند. این تحلیل می‌تواند برای شناسایی، بافت و روابط بین مفاهیم قرآنی استفاده شود.

۴-۴. تحلیل ساختاری

تحلیل ساختاری در چهارچوب به کار می‌رود تا ساختار زیربنایی متن قرآن را بررسی کند. این تحلیل می‌تواند برای درک بهتر سازمان‌دهی، روابط و معانی مفاهیم قرآنی استفاده شود.

جدول ۲: ابعاد کلیدی چهارچوب پیشنهادی و معیارهای ارزیابی

توضیحات	معیار ارزیابی	بعد دوم: کاربردهای جمع‌سپاری	بعد اول: تکنیک‌های NLP
این شاخص درصد کلمات به‌درستی تفسیرشده در مجموعه داده را اندازه‌گیری می‌کند.	WSE: ارزیابی رفع ابهام معنایی کلمات	رفع ابهام معنایی کلمات: می‌توان از جمع‌سپاری برای پالایش جاسازی کلمات و بهبود توانایی آنها در درک ظرافت‌های واژگان قرآنی استفاده کرد.	جاسازی کلمات
این شاخص کیفیت خلاصه‌سازی متن را با مقایسه آن با خلاصه‌های مرجع ارزیابی می‌کند.	ارزیابی خلاصه‌سازی متن	خلاصه‌سازی متن: می‌توان از جمع‌سپاری برای تولید خلاصه‌ای از آیات و سوره‌های قرآنی استفاده کرد که فهم و دسترسی به آنها را آسان‌تر می‌کند.	تقسیم‌بندی جمله
این شاخص دقت و توانایی سیستم را در تشخیص درست احساسات آیات قرآنی اندازه‌گیری می‌کند.	FI-score: دقت و بازخوانی برای تجزیه و تحلیل احساسات	تجزیه و تحلیل احساسات: می‌توان از جمع‌سپاری برای تجزیه و تحلیل احساسات آیات قرآنی استفاده کرد و بینشی در مورد لحن احساسی و پیام‌های نهفته ارائه کرد.	برچسب‌گذاری اجزای گفتار

شناسایی موجودیت نام‌دار	مدل‌سازی موضوع: می‌توان از جمع‌سپاری برای شناسایی و دسته‌بندی موجودیت‌های نام‌دار در قرآن، مانند افراد، مکان‌ها و رویدادها استفاده کرد.	دقت، بازخوانی و F1-score برای شناسایی موجودیت نام‌دار	این شاخص توانایی سیستم را در شناسایی و دسته‌بندی موجودیت‌های نام‌دار مانند افراد، مکان‌ها و رویدادها در قرآن اندازه‌گیری می‌کند.
مدل‌سازی موضوع	پاسخ به سؤال: می‌توان از جمع‌سپاری برای توسعه سیستم‌هایی استفاده کرد که می‌توانند به سؤالات مربوط به قرآن پاسخ دهند و اطلاعات سریع و دقیقی را به کاربران ارائه دهند.	رتبه متقابل میانگین برای پاسخ به سؤال	این شاخص توانایی سیستم را در پاسخگویی دقیق و مرتبط به سؤالات مربوط به قرآن اندازه‌گیری می‌کند.
ترجمه ماشینی	تجزیه و تحلیل چندزبانه قرآن: می‌توان از جمع‌سپاری برای تسهیل ترجمه قرآن به زبان‌های مختلف استفاده کرد و دامنه دسترسی آن را به مخاطبان گسترده‌تر گسترش داد.	ارزیابی ترجمه ماشینی	این شاخص کیفیت ترجمه ماشینی قرآن را با مقایسه آن با ترجمه‌های انسانی مرجع ارزیابی می‌کند.
طبقه‌بندی متن	طبقه‌بندی آیات قرآنی: می‌توان از جمع‌سپاری برای طبقه‌بندی آیات قرآنی بر اساس مضامین، موضوعات یا اهداف آنها استفاده کرد.	دقت: درصد آیات به درستی طبقه‌بندی شده	این شاخص توانایی سیستم را در طبقه‌بندی صحیح آیات قرآنی بر اساس موضوع، مضمون یا هدف آنها اندازه‌گیری می‌کند.
استخراج رابطه	ساخت نمودار دانش قرآنی: می‌توان از جمع‌سپاری برای ساخت نمودار دانشی استفاده کرد که آیات، مفاهیم و موجودات قرآنی را به هم مرتبط می‌کند.	کیفیت لینک: انسجام و قابلیت اتکای متوسط لینک‌ها	این شاخص انسجام و قابلیت اتکای لینک‌ها بین مفاهیم، آیات و موجودیت‌های مختلف در نمودار دانش قرآنی را اندازه‌گیری می‌کند.

علاوه بر ابعاد اصلی چهارچوب پیشنهادی، ابعاد دیگری لازم است به عنوان ابعاد تکمیلی در واکاوی تحلیلی مفاهیم قرآنی با استفاده از پردازش زبان طبیعی در نظر گرفته شوند که در جدول (۳) جمع‌بندی و سازمان‌دهی شده‌اند.

۴-۵. روش‌شناسی تحقیق کیفی

چهارچوب از روش‌شناسی تحقیق کیفی استفاده می‌کند که بر جمع‌آوری و تحلیل داده‌های غیرعددی متمرکز دارد. این روش برای درک عمیق‌تر مفاهیم قرآنی و روابط بین آنها مناسب است.

۴-۶. فرایند چرخه‌ای تحلیل

فرایند چرخه‌ای تحلیل در چهارچوب به کار می‌رود تا داده‌ها را به صورت مداوم جمع‌آوری، تحلیل و تفسیر کند. این فرایند به درک عمیق‌تر مفاهیم قرآنی و شناسایی الگوها و روابط جدید در متن کمک می‌کند.

۴-۷. دیدگاه اسلامی

چهارچوب بر اساس دیدگاه اسلامی استوار است؛ بدین معنا که تحلیل مفاهیم قرآنی از دیدگاه اسلامی و با توجه به بافت فرهنگی و تاریخی قرآن انجام می‌شود.

۴-۸. مدل‌سازی پویا

مدل‌سازی پویا در چهارچوب به کار می‌رود تا سیستم‌های پیچیده و رفتار آنها را در طول زمان نمایش دهد. در تحلیل مفاهیم قرآنی، از مدل‌سازی پویا می‌توان برای نمایش روابط بین مفاهیم و تکامل آنها در طول قرآن استفاده کرد.

۹-۴. رویکرد میان‌رشته‌ای

چهارچوب از رویکرد میان‌رشته‌ای استفاده می‌کند که بر تلفیق دیدگاه‌ها و روش‌شناسی‌های مختلف از رشته‌های گوناگون، مانند علوم رایانه، زبان‌شناسی، الهیات، فلسفه و مطالعات تاریخی و فرهنگی استوار است. این رویکرد به درک عمیق‌تر مفاهیم قرآنی و شناسایی روابط بین آنها کمک می‌کند.

جدول ۳: ابعاد تحلیلی روش پیشنهادی

معیار / زیرمعیار	توضیحات ویژگی‌ها	متغیرهای مؤثر در ارزیابی
رویکرد جمع‌سپاری	- همکاری چندین فرد یا سیستم برای تحلیل مفاهیم قرآنی	تعداد شرکت‌کنندگان
	- پایه‌گذاری بر اعتماد به همکاران و سیستم‌ها	اعتماد به دقت برچسب‌ها
	- انعطاف‌پذیری در استفاده از اطلاعات و تجربیات شرکت‌کنندگان	تنوع برچسب‌ها
	- استفاده از مکانیسم‌های تشویقی برای حفظ و تشویق همکاری	رضایت شرکت‌کنندگان
	- بررسی صحت و اعتماد به دقت برچسب‌گذاری و تحلیل مشارکت‌کنندگان	
پردازش زبان طبیعی	- توانایی شناسایی الگوها و روابط معانی در متن قرآن	دقت در تحلیل مفاهیم
	- قابلیت استفاده از الگوریتم‌های مختلف برای تحلیل متون	پوششی گسترده مفاهیم
	- انعطاف‌پذیری در تطبیق با متون با ساختار متنوع و زبان‌های مختلف	سرعت پردازش متن
	- دقت و قابلیت اعتماد در تشخیص معانی و روابط متنی	
تحلیل محتوا	- شناسایی الگوها، موضوعات و معانی در متن قرآن	دقت در تحلیل الگوها و موضوعات
	- درک بهتر سازمان‌دهی و روابط بین مفاهیم قرآنی	انعطاف‌پذیری در شناسایی مفاهیم
تحلیل ساختاری	- بررسی ساختار زیربنایی متن قرآن	دقت در تحلیل ساختار
	- شناسایی روابط و معانی مفاهیم در ساختار متن	
روش‌شناسی	- تمرکز بر جمع‌آوری و تحلیل داده‌های غیر عددی	دقت در تحلیل داده‌ها
تحقیق کیفی	- توانایی در درک عمیق‌تر مفاهیم قرآنی و روابط بین آنها	انعطاف‌پذیری در تحلیل
فرایند چرخه‌ای تحلیل	- جمع‌آوری، تحلیل و تفسیر داده‌ها به صورت مداوم	پیگیری و انعطاف‌پذیری فرایند
	- شناسایی الگوها و روابط جدید در طول زمان	
ابزارهای ساخته‌شده	- استفاده از نرم‌افزارها، الگوریتم‌ها و ابزارهای دیجیتال طراحی‌شده برای تحلیل مفاهیم قرآنی	کارایی و کاربردی بودن ابزار
دیدگاه اسلامی	- تحلیل مفاهیم قرآنی از دیدگاه اسلامی و با توجه به بافت فرهنگی و تاریخی قرآن	تطابق با مفاهیم اسلامی
مدل‌سازی پویا	- نمایش سیستم‌های پیچیده و رفتار آنها در طول زمان	پوششی زمانی مفاهیم
	- نمایش تکامل مفاهیم قرآنی و روابط بین آنها در طول قرآن	
رویکرد میان‌رشته‌ای	- تلفیق دیدگاه‌ها و روش‌شناسی‌های مختلف از رشته‌های گوناگون	تنوع و کارایی در تحلیل
	- بهبود درک عمیق‌تر مفاهیم قرآنی و شناسایی روابط بین آنها	

به‌طور کلی چهارچوب واکاوی تحلیلی مفاهیم قرآنی با استفاده از پردازش زبان طبیعی، رویکردی مبتنی بر جمع‌سپاری است که برای تحلیل و درک عمیق‌تر مفاهیم قرآنی طراحی شده است. این چهارچوب با تلفیق مؤلفه‌های مختلف، ابزار قدرتمندی برای تحلیل و درک مفاهیم قرآنی فراهم می‌کند.

۵. معیارهای تحلیل و ارزیابی

ارزیابی و سنجش کارایی و صحت رویکردهای ماشینی قرآن کریم و کتب مهم اهمیت بسیاری دارد و همواره از دغدغه‌های اصلی پژوهشگران و مفسران است. در این راستا، چهار معیار و شاخص کلیدی برای ارزیابی این روش‌ها معرفی می‌کنیم: ۱. دقت: برای بررسی میزان صحت و درستی مفاهیم و معانی استخراج‌شده از متون قرآنی، که شاخص‌های سنجش آن عبارت‌اند از: نرخ تناقض (میزان مغایرت با تفاسیر معتبر)، نرخ ابهام (میزان عدم وضوح مفاهیم استخراج‌شده) و نرخ صحت معنایی (میزان انطباق با معنای مورد نظر آیه)؛ ۲. انعطاف‌پذیری: به توانایی روش در انطباق با طیف وسیعی از متون قرآنی، با موضوعات و ساختارهای مختلف می‌پردازد و شاخص‌های سنجش آن عبارت‌اند از: میزان پوشش موضوعات (تعداد و تنوع موضوعات قابل تحلیل)، میزان تطبیق با ساختارها (توانایی تحلیل آیات با ساختارهای متفاوت) و سهولت به‌کارگیری؛ ۳. پوشش: به جامعیت روش در تحلیل و استخراج مفاهیم و معانی متنوع موجود در متون قرآنی اشاره دارد و شاخص‌های سنجش آن عبارت‌اند از: تعداد مفاهیم استخراج‌شده، عمق مفاهیم استخراج‌شده (از کلیات تا جزئیات) و تطابق با مفاهیم کلیدی (انطباق با مفاهیم بنیادی در علوم قرآنی)؛ ۴. صحت اطلاعات: به میزان قابل اعتماد بودن و اتکا به اطلاعات و مفاهیم استخراج‌شده از متون قرآنی می‌پردازد. شاخص‌های سنجش صحت اطلاعات عبارت‌اند از: اعتبار منابع و مدارک، تطابق روش‌شناسی با اصول علمی و وجود فرایندهای بازبینی و نظارت توسط متخصصان. خلاصه معیارها و شاخص‌هایی که برای ارزیابی و تحلیل مفاهیم و معانی در متون قرآنی مورد استفاده قرار می‌گیرد، در جدول (۴) نشان داده شده است.

جدول ۴: معیارهای ارزیابی و تحلیل چهارچوب پیشنهادی

معیار/شاخص	توضیح
دقت	میزان صحت و دقت در تحلیل و استخراج مفاهیم و معانی از متون قرآنی
انعطاف‌پذیری	توانایی چهارچوب برای انطباق با متون متنوع و پیچیده قرآنی به‌صورت انعطاف‌پذیر و منطبق
پوشش	میزان پوشش چهارچوب در تحلیل و استخراج مفاهیم و معانی متنوعی از متون قرآنی و شامل متن‌های مختلف
صحت اطلاعات	صحت و قابل اطمینان بودن اطلاعات و مفاهیم استخراج‌شده از متون قرآنی

این معیارها و شاخص‌ها به‌عنوان ابزارهایی برای ارزیابی کیفیت و کارایی چهارچوب تحلیلی واکاوی فلسفی و تحلیلی از مفاهیم با پردازش زبان طبیعی در رویکرد جمع‌سپاری پیشنهاد می‌شوند و به این وسیله می‌توان نتایج کارهای واگذارشده به افراد مختلف با تخصص‌های گوناگون را بررسی و بر اساس این معیارها و شاخص‌ها امتیازدهی کرد. در شکل (۳) این شاخص‌ها با زیرمعیارهای آن نشان داده شده‌اند.

شکل ۳: معیارهای ارزیابی به‌همراه زیرمعیارها

موارد ارزیابی و ارزیابی			دقت شامل صحت یادآوری و امتیاز F1	
تعدد موارد ارزیابی	تنوع موارد ارزیابی	ارزیابی توسط متخصصان	امتیاز F1: معیاری ترکیبی است که دقت و یادآوری را به‌طور هم‌زمان در نظر می‌گیرد. این معیار، میانگین وزنی دقت و یادآوری را نشان می‌دهد و به‌عنوان یک شاخص واحد برای ارزیابی عملکرد چهارچوب NLP در استخراج مفاهیم از متن قرآنی به‌کار می‌رود.	صحت نسبت تعداد مفاهیم صحیح استخراج‌شده به تعداد کل مفاهیم استخراج‌شده را نشان می‌دهد. به عبارت دیگر، دقت بیانگر این است که از بین تمام مفاهیمی که چهارچوب NLP از متن استخراج کرده، چه درصدی از آنها با مفاهیم واقعی موجود در متن مطابقت دارند.
نشان می‌دهد که چهارچوب تا چه حد در برابر طیف وسیعی از متون قرآنی با موضوعات و سبک‌های مختلف، انعطاف‌پذیری خود را حفظ می‌کند.	تنوع انواع متن نشان می‌دهد که چهارچوب تا چه حد قادر به تحلیل و استخراج مفاهیم از انواع مختلف متون قرآنی مانند آیات، احادیث و تفاسیر است.	تعدد مفاهیم نشان می‌دهد که چهارچوب تا چه حد قادر به استخراج طیف وسیعی از مفاهیم و معانی از متون قرآنی است.	یادآوری: نسبت تعداد مفاهیم صحیح استخراج‌شده به تعداد کل مفاهیم موجود در متن را نشان می‌دهد. به عبارت دیگر، یادآوری بیانگر این است که از بین تمام مفاهیم واقعی موجود در متن، چه درصدی از آنها توسط چهارچوب NLP استخراج شده‌اند.	صحت نسبت تعداد مفاهیم صحیح استخراج‌شده به تعداد کل مفاهیم استخراج‌شده را نشان می‌دهد. به عبارت دیگر، دقت بیانگر این است که از بین تمام مفاهیمی که چهارچوب NLP از متن استخراج کرده، چه درصدی از آنها با مفاهیم واقعی موجود در متن مطابقت دارند.
پوشش شامل تنوع مفاهیم و تنوع متن			صحت اطلاعات شامل نرخ خطا و ارزیابی توسط متخصصان	
تعدد مفاهیم نشان می‌دهد که چهارچوب تا چه حد قادر به استخراج طیف وسیعی از مفاهیم و معانی از متون قرآنی است.	تنوع انواع متن نشان می‌دهد که چهارچوب تا چه حد قادر به تحلیل و استخراج مفاهیم از انواع مختلف متون قرآنی مانند آیات، احادیث و تفاسیر است.	ارزیابی توسط متخصصان شامل نظرخواهی از متخصصان در حوزه تحلیل متن قرآنی و سنجش صحت اطلاعات استخراج‌شده از دیدگاه آنهاست. نمره صحت اطلاعات بر اساس میانگین نظرات متخصصان محاسبه می‌شود.	نرخ خطا نشان می‌دهد که چه درصدی از مفاهیم و معانی استخراج‌شده نادرست هستند.	صحت نسبت تعداد مفاهیم صحیح استخراج‌شده به تعداد کل مفاهیم استخراج‌شده را نشان می‌دهد. به عبارت دیگر، دقت بیانگر این است که از بین تمام مفاهیمی که چهارچوب NLP از متن استخراج کرده، چه درصدی از آنها با مفاهیم واقعی موجود در متن مطابقت دارند.

به‌منظور ایجاد الگویی برای استفاده از معیارها، دقت را با مثالی بررسی می‌کنیم. فرض کنید چهارچوب NLP ما وظیفه دارد تا مفاهیم کلیدی را از سوره مبارکه بقره استخراج کند. این سوره شامل ۲۸۶ آیه است که هر کدام حاوی مفاهیم و معانی عمیق و پیچیده‌ای هستند. برای سنجش دقت چهارچوب NLP در این وظیفه، از سه زیرمعیار صحت، یادآوری و نمره F1 استفاده می‌کنیم:

صحت: چهارچوب NLP در مجموع ۱۲۰۰ مفهوم از ۲۸۶ آیه سوره بقره استخراج می‌کند. پس از بررسی توسط کارشناسان علوم قرآنی، مشخص می‌شود که ۹۶۰ مورد از این مفاهیم با مفاهیم واقعی موجود در متن مطابقت دارند.

$$\text{صحت} = (960 / 1200) * 100 = 80\%$$

این نتیجه نشان می‌دهد که چهارچوب NLP در ۸۰ درصد موارد، مفاهیم صحیح را از متن استخراج کرده است. **یادآوری:** کارشناسان علوم قرآنی پس از بررسی متن سوره بقره، ۱۵۰۰ مفهوم کلیدی را در آن شناسایی می‌کنند.

$$\text{یادآوری} = (960 / 1500) * 100 = 64\%$$

این نتیجه نشان می‌دهد که چهارچوب NLP توانسته است ۶۴ درصد از مفاهیم کلیدی موجود در متن را به‌درستی استخراج کند.

امتیاز F1: با توجه به مقادیر دقت و یادآوری، نمره F1 چهارچوب NLP را محاسبه می‌کنیم:

$$\text{امتیاز } F1 = (80 * 64) / (80 + 64) * 2 = 71.11\%$$

این نمره نشان می‌دهد که عملکرد چهارچوب NLP در استخراج مفاهیم از سورة بقره، «متوسط» بوده است و جای ارتقا دارد.

در این مثال، چهارچوب NLP توانسته است ۸۰ درصد از مفاهیم را به‌درستی استخراج کند. با این حال، ۲۰ درصد از مفاهیم استخراج‌شده نادرست بوده‌اند و نیاز به اصلاح دارند. برای ارتقای دقت چهارچوب NLP، می‌توان از تکنیک‌های پیشرفته‌تر NLP، مانند یادگیری عمیق و شبکه‌های عصبی استفاده کرد. همچنین می‌توان با استفاده از پایگاه‌های دانش تخصصی در حوزه علوم قرآنی، به چهارچوب NLP در درک بهتر مفاهیم و معانی متن کمک کرد.

۶. محدودیت‌ها

پردازش زبان طبیعی برای تحلیل متن‌های قرآن با چالش‌ها و سختی‌های خاصی روبه‌رو می‌شود:

۱. **ساختار زبانی پیچیده:** قرآن به زبان عربی نوشته شده است که دارای یک ساختار زبانی پیچیده و قواعد دقیقی است. این ساختار شامل استفاده از اصطلاحات، قواعد خاص و تنوع زبانی بسیاری است که برای الگوریتم‌های پردازش زبان طبیعی سختی می‌آفریند.

۲. **استفاده از مفاهیم فلسفی و دینی:** قرآن به‌عنوان متن دینی دارای مفاهیم فلسفی و دینی عمیقی است که برای درک آنها به شناخت عمیق و تخصصی از این مفاهیم نیاز است. تحلیل و استخراج این مفاهیم به‌وسیله الگوریتم‌های NLP چالش بزرگی است.

۳. **ترجمه و تفسیر معانی چندگانه:** قرآن به زبان عربی نوشته شده است و ترجمه آن به زبان‌های دیگر، معانی چندگانه‌ای دارد. این موضوع باعث می‌شود که الگوریتم‌های NLP برای ترجمه و تفسیر متون قرآنی به زبان‌های دیگر نیازمند دقت و دانش عمیقی باشد.

۴. **حجم بزرگ متن:** قرآن به‌عنوان یکی از کتب مقدس و مهم دینی، دارای حجم بسیار زیادی از متن است. پردازش و تحلیل کلیه متن‌های قرآنی به‌صورت دستی یا با استفاده از الگوریتم‌های NLP نیازمند منابع محاسباتی بالا و زمان بسیاری است.

۵. **تفسیر متون با رویکردهای مختلف:** قرآن معمولاً به تفسیرهای مختلفی تحت عناوین مختلف مورد بررسی قرار می‌گیرد. این باعث می‌شود که تحلیل و استخراج معانی قرآن با استفاده از الگوریتم‌های NLP دقت و کنترل بیشتری نیاز داشته باشد.

اساساً پردازش متون قرآنی با استفاده از NLP نیازمند الگوریتم‌ها و روش‌های پیشرفته‌ای است که بتوانند با چالش‌های یادشده برخورد کنند و به نتایج قابل قبولی برسند. استفاده از مدل‌های زبانی بزرگ، به‌ویژه در زمینه پردازش زبان طبیعی، می‌تواند به‌طور قابل توجهی به پیشرفت و توسعه فناوری‌های مربوط به زبان فارسی کمک کند. این مدل‌ها،

که اکثراً با استفاده از شبکه‌های عصبی عمیق آموزش داده می‌شوند، قادرند به تولید متون، ترجمه خودکار، تحلیل معنایی، تولید متن تصویری، پرسش و پاسخ، و بسیاری دیگر از وظایف مرتبط با زبان فارسی و زبان‌های دیگر کمک کنند. دولت می‌تواند با ارائه حمایت‌های لازم از پژوهش‌ها و پروژه‌های مرتبط با پردازش زبان طبیعی در زبان فارسی، اقدام به توسعه و آموزش مدل‌های زبانی متنوع و پیشرفته کند. این اقدامات می‌توانند شامل اعطای منابع مالی، ایجاد محیط‌های همکاری بین دانشگاه‌ها، صنعت و دولت، تسهیل در دسترسی به داده‌های متنوع و بزرگ، و تشویق به پژوهش‌های مبتنی بر این فناوری باشد.

نتیجه‌گیری

این مقاله به بررسی استفاده از روش‌های نوین پردازش زبان طبیعی در تحلیل متون قرآنی پرداخت و نشان داد که چگونه این رویکرد می‌تواند در مقایسه با روش‌های سنتی، تحلیل دقیق‌تر و جامع‌تری از مفاهیم و الگوهای معنایی فراهم آورد. بهره‌گیری از روش‌های نوین پردازش زبان طبیعی در تحلیل متون قرآنی نقشی کلیدی در تدوین و اجرای راهبردهای فرهنگی مؤثر ایفا می‌کند. روش‌های سنتی، اگرچه در زمان خود کارآمد بوده‌اند، با محدودیت‌های جدی در تحلیل دقیق و جامع متون روبه‌رویند. استفاده از مدل‌های پیشرفته NLP امکان شناسایی و تحلیل عمیق‌تر مفاهیم و الگوهای معنایی را فراهم می‌آورد. این رویکرد نوین می‌تواند به تفسیر دقیق‌تر آیات و واژگان قرآنی کمک کند و از این طریق زمینه‌ساز انتقال صحیح و جذاب پیام‌های فرهنگی شود.

تدوین راهبردهای فرهنگی مبتنی بر تحلیل دقیق متون دینی نیازمند رویکردی میان‌رشته‌ای است که از تخصص‌های مختلف علوم رایانه، زبان‌شناسی و علوم اسلامی بهره‌مند باشد. این تعامل میان‌رشته‌ای امکان بهره‌گیری از تخصص‌ها و دانش‌های مختلف را فراهم می‌کند و در نهایت به درک بهتر و جامع‌تری از مفاهیم قرآنی منجر می‌شود. در این راستا، این مقاله بر ضرورت رویکرد میان‌رشته‌ای در این زمینه تأکید کرد و نشان داد که تعامل میان تخصص‌های مختلف می‌تواند به تدوین و اجرای راهبردهای فرهنگی مؤثر، که بر پایه تحلیل دقیق متون دینی استوار است، بینجامد. در نهایت، بهره‌گیری از این روش‌ها می‌تواند به ارتقای آگاهی فرهنگی و دینی در جامعه کمک کند. این خود بستر مناسبی برای برنامه‌ریزی و اجرای راهبردهای فرهنگی مؤثر فراهم می‌سازد که می‌تواند به ارتقای آگاهی فرهنگی و دینی در جامعه منجر شود. در نتیجه، پژوهش‌های مبتنی بر NLP نه تنها برای درک بهتر متون دینی مفیدند، بلکه نقش بسزایی در شکل‌دهی به راهبردهای فرهنگی دارند که به ترویج و نهادینه‌سازی ارزش‌های دینی و فرهنگی در جامعه کمک می‌کند. در آخر باید اشاره کرد که نیاز به پلتفرم‌های هوش مصنوعی قرآنی که بتوانند به‌طور خودکار تحلیل‌های دقیق و جامع را ارائه دهند، بیش از پیش احساس می‌شود. این پلتفرم‌ها می‌توانند ابزارهای قدرتمندی برای پژوهشگران و علاقه‌مندان به مطالعات قرآنی فراهم کنند و راهبردهای فرهنگی را با دقت و کارآمدی بیشتری به اجرا درآورند.

منابع

- آذری، سلمان، شکرالهی فر، محمود و فشارکی، محمدعلی لسانی (۱۳۹۹). الگودارزی مفهوم‌یابی از آیات قرآن کریم با استفاده از دانش متن‌کاوی رایانشی. ذهن، ۲۱(۸۲)، ۱۰۸-۸۱.
- ایازی، محمدعلی (۱۳۸۹). نگاهی به معجم‌های موضوعی قرآن کریم. بینات، ۲، ۳۳-۱۲۷.
- بهجت‌پور، عبدالکریم (۲۰۱۴). روش کشف الگوی اسلامی مهندسی فرهنگی. فصلنامه علمی پژوهشی راهبرد فرهنگ، ۷(۲۷)، ۱۱۸-۹۱.
- صوفی، محسن، علی‌احمدی، علیرضا، احمدی، حسین‌علی و مینایی، بهروز (۱۳۹۴). خوشه‌بندی سوره‌های قرآن با تکنیک‌های داده‌کاوی. علوم قرآن و حدیث، ۲۶(۱۵۳)، ۵۶-۶۲.
- فریادرس، زینب (۲۰۱۵). گزارشی توصیفی از چهار معجم موضوعی قرآن. آینه پژوهش، ۲۶، ۴۶-۶۲.
- یزدان‌پرستی، عالیه و اسماعیلی، مهدی (۱۳۹۴). مروری بر انواع روش‌های داده‌کاوی و متن‌کاوی روی قرآن مقدس. در: دومین کنفرانس ملی مهندسی برق و کامپیوتر سیستم‌های توزیع‌شده و شبکه‌های هوشمند.
- Arief Fatchul, Huda, Huda Arief Fatchul, Fauziah None, Fauziah, Ratnawulan Elis, Ratnawulan Elis, Gumelar Dian Rachmat, and Gumelar Dian Rachmat (2021). *Arabic Part Of Speech (POS) Tagging Analysis using Bee Colony Optimization (BCO) Algorithm on Quran Corpus*. 7th International Conference on Wireless and Telematics (ICWT).
- Aslani, Akram and Mahdi Esmaeili (2018). Finding Frequent Patterns in Holy Quran Using Text Mining. *Signal and Data Processing*, (15), 89-100.
- Auhood, Alfaries, Alfaries Auhood, Al-Bahlal Manal, Al-Bahlal Manal, Almazruea Manal, Almazruea Manal, Almazruea Amal, and Almazruea Amal (2013). *A Rule-Based Annotation System to Extract Tajweed Rules from Quran*. Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences.
- Bana, Alsharif, R. Tahboub, and L. Arafah (2016). Arabic text to speech synthesis using Quran-based natural language processing module. *Journal of Theoretical and Applied Information Technology*, 83(1), 148.
- Damith, Premasiri, Ranasinghe Tharindu, W. Zaghouni, and R. Mitkov (2022). *DTW at Quran QA 2022: Utilising Transfer Learning with Transformers for Question Answering in a Low-resource Domain*, OSACT.
- Dardya, Putra, D. I. A. Putra, Yusuf Muhammad, Yusuf Mohd Ali Mohd, M. Yusuf, and M. Yusuf (2021). *Proposing machine learning of Tafsir al-Quran: In search of objectivity with semantic analysis and Natural Language Processing*, IOP Conference Series: Materials Science and Engineering.
- Ehsan, Khadangi, Khadangi Ehsan, Khadangi Ehsan, Fazeli Mohammad Moein, Fazeli Mohammad Moein, Shahmohammadi Amin, and Shahmohammadi Amin (2018). *The Study on Quranic Surahs Topic Sameness Using NLP Techniques*, International Conference on Computer and Knowledge Engineering.
- Ensaf Hussein, Mohamed, Mohamed Ensaf Hussein, H. El-Behaidy Wessam, and H. El-Behaidy Wessam (2021). *An Ensemble Multi-label Themes-Based Classification for Holy Quran Verses Using Word2Vec Embedding*, Arabian Journal for Science and Engineering.

- Hupkes, Dieuwke, Mario Giulianelli, Verna Dankers, Mikel Artetxe, Yanai Elazar, Tiago Pimentel, Christos Christodoulopoulos, Karim Lasri, Naomi Saphra, and Arabella Sinclair (2023). A taxonomy and review of generalization research in NLP. *Nature Machine Intelligence*, (5), 1161-74.
- Khadangi, Ehsan, Mohammad Moein Fazeli, and Mehdi Naghavi (2022). *The Study on Quranic surahs Structured-ness and their Order Organization Using NLP Techniques*, Journal of Interdisciplinary Quranic Studies (JIQS), 1.
- Mohamed Osman, Hegazi, Hegazi Mohamed Osman, Hilal Anwer Mustafa, Hilal Anwer, Alhawarat Mohammad, and Alhawarat Mohammad (2015). *Fine-Grained Quran Dataset*, International Journal of Advanced Computer Science and Applications.
- Mohammad, Alhawarat, and Alhawarat Mohammad (2015). *Extracting Topics from the Holy Quran Using Generative Models*, International Journal of Advanced Computer Science and Applications.
- Muhammad Huzaifa, Bashir, M. Azmi Aqil, Nawaz Haq, Zaghouni Wajdi, Diab Mona, Al-Fuqaha Ala, and Qadir Junaid (2023). Author Correction: Arabic natural language processing for Quranic research: a systematic review, *Artificial Intelligence Review*.
- Nima, A .Al-Fakhry, and Al-Fakhry Nima (2020). *The Linguistic Connotations of the Word Light in the Holy Quran (An analytical study of Quranic verses using Artificial intelligent techniques)*, Al-Rafidain Journal of Computer Sciences and Mathematics.
- Patil, Rajvardhan, Sorio Boit, Venkat Gudivada, and Jagadeesh Nandigam (2023). *A survey of text representation and embedding techniques in nlp*, IEEE Access.
- Shasha Arzila, Tarmizi, and Saad Saidah (2022). *NAMED ENTITY RECOGNITION FOR QURANIC TEXT USING RULE BASED APPROACHES*, Asia-Pacific Journal of Information Technology and Multimedia.
- Syamsu, Nahar, Daulay Nurul Sakinah, and M. Nazri (2023). *Optimalisasi Menghafal Al-Quran: Penerapan Metode Neuro Linguistic Programming (NLP) di Pesantren Islamic Centre Sumut, Fikrotuna: Jurnal Pendidikan dan Manajemen Islam*.
- Yasser, Shohoud and Shoman Maged (2023). *Quranic Conversations: Developing a Semantic Search tool for the Quran using Arabic NLP Techniques*, arXiv.org.
- Zineb, Touati-Hamad, Touati-Hamad Zineb, Laouar Mohamed Ridda, Laouar Mohamed Ridda, Bendib Issam, and Bendib Issam (2020). *Quran content representation in NLP*, International Conference on Information and Software Technologies.
- Zineb, Touati-Hamad, Laouar Mohamed Ridda, Bendib Issam, and Hakak Saqib (2022). *Arabic Quran Verses Authentication Using Deep Learning and Word Embeddings*, ~The cinternational Arab journal of information technology.