




# Automated Thematic Classification System for Qur'anic Verses Using Machine Learning and Deep Learning

Mohsen Khoshfar<sup>1</sup>  and Mohammad-Rasoul Rahimi<sup>2</sup>  and Mohammad-Amin Ghiabi Zare<sup>3</sup> 

1. corresponding author, Assistant Professor and Faculty Member, University of Qom, Iran, Email: [mohsenkhoshfar@yahoo.com](mailto:mohsenkhoshfar@yahoo.com)
2. Undergraduate Student in Software Engineering, Faculty of Engineering, Islamic Azad University, North Tehran Branch, Tehran, Iran, Email: [rahimimohammadrasool@gmail.com](mailto:rahimimohammadrasool@gmail.com)
3. Undergraduate Student in Software Engineering, Faculty of Engineering, Islamic Azad University, North Tehran Branch, Tehran, Iran, Email: [a.m.denizer@gmail.com](mailto:a.m.denizer@gmail.com)

## Article Info

## ABSTRACT

**Article type:**  
Research Article

**Article history:**  
**Received:** 27 April 2025  
**Received in revised form:**  
26 June 2025  
**Accepted:** 25 August 2025  
**Available online:**  
11 March 2026

**Keywords:**  
Thematic classification;  
Machine learning; Deep  
learning;  
Multi-label classification;  
Transformer;  
Qur'anic text mining.

Manual thematic classification of Qur'anic verses is a time-consuming and complex process. Given that many verses address multiple themes simultaneously, the need for an efficient automated method to organize Qur'anic content based on thematic categories has become increasingly evident. The primary research question of this study is how machine learning (ML) and deep learning (DL) techniques can be utilized to develop an automated system capable of determining the theme or themes of each verse. This study adopts a descriptive-analytical research method. First, key concepts and relevant literature are examined, followed by an analysis of various ML and DL approaches applied to thematic classification of Qur'anic verses. The findings indicate that classical algorithms such as Naïve Bayes and Support Vector Machines have produced acceptable results. However, given that more than half of the verses require multi-label classification, the use of multi-label models is essential. The application of Convolutional Neural Networks (CNNs) and Transformer-based models has significantly improved classification accuracy, with recent studies reporting accuracy rates exceeding 90%. These findings demonstrate the effectiveness of automated systems as supportive tools for researchers in Qur'anic studies. Nevertheless, limitations such as reliance on labeled datasets and limited model interpretability remain significant challenges. Overall, integrating computer science methodologies with Qur'anic studies through automated thematic classification systems provides an effective complementary approach to traditional exegetical and indexing methods, opening new avenues for Qur'anic research in the digital age.

**Cite this article:** Elyasi, M. (2025). Designing a Conceptual Model of Explainable Artificial Intelligence (XAI) Based on the clarification jihad Approach. *Digital Islamic Studies and Humanities*, 1 (2), 303-344. <https://doi.org/10.22034/disah.2025.2076072.1054>



© The Author(s). **Publisher:** Research Center for Digital Islamic Studies and Humanities (RCDISAH).

**DOI:** <https://doi.org/10.22034/disah.2025.2076072.1054>

## سامانه خودکار طبقه‌بندی موضوعی آیات قرآن با استفاده از یادگیری ماشین

### و یادگیری عمیق

محسن خوش‌فر<sup>۱</sup> و محمدرسول رحیمی<sup>۲</sup> و محمدامین غیابی زارع<sup>۳</sup>

۱. نویسنده مسئول، استادیار و عضو هیئت علمی دانشگاه، قم، ایران، رایانامه: [mohsenkhoshfar@yahoo.com](mailto:mohsenkhoshfar@yahoo.com)
۲. دانشجوی کارشناسی مهندسی نرم‌افزار، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد تهران شمال، تهران، ایران، رایانامه: [rahimimohammadrasool@gmail.com](mailto:rahimimohammadrasool@gmail.com)
۳. دانشجوی کارشناسی مهندسی نرم‌افزار، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد تهران شمال، تهران، ایران: [a.m.denizer@gmail.com](mailto:a.m.denizer@gmail.com)

### اطلاعات مقاله

### چکیده

طبقه‌بندی موضوعی آیات قرآن کریم به‌صورت دستی، فرایندی زمان‌بر و پیچیده است؛ به‌ویژه با توجه به چندموضوعی بودن بسیاری از آیات، نیاز به روشی خودکار و کارآمد برای سازمان‌دهی محتوای قرآن بر اساس موضوعات، بیش از پیش احساس می‌شود. پرسش اصلی این پژوهش آن است که چگونه می‌توان با بهره‌گیری از فنون یادگیری ماشین (ML) و یادگیری عمیق (DL)، سامانه‌ای خودکار برای تعیین موضوع یا موضوعات هر آیه ارائه داد. روش تحقیق این مطالعه، توصیفی-تحلیلی است. ابتدا مفاهیم کلیدی و پیشینه مرتبط بررسی شده و سپس رویکردهای مختلف ML و DL در زمینه طبقه‌بندی موضوعی آیات تحلیل می‌گردد. یافته‌ها نشان می‌دهد که الگوریتم‌های کلاسیک مانند بیز ساده و ماشین بردار پشتیبان، نتایج قابل قبولی ارائه کرده‌اند؛ اما با توجه به چندبرچسبی بودن بیش از نیمی از آیات، استفاده از مدل‌های چندبرچسبی ضروری است. بهره‌گیری از شبکه‌های عصبی پیچشی (CNN) و ترنسفورمرها موجب افزایش چشمگیر دقت طبقه‌بندی شده و پژوهش‌های اخیر دقتی بالاتر از ۹۰٪ را گزارش کرده‌اند. این دستاوردها بیانگر توانمندی سامانه‌های خودکار به‌عنوان ابزار کمکی برای پژوهشگران علوم قرآنی است. با این حال، محدودیت‌هایی همچون وابستگی به داده‌های برچسب‌خورده و پایین بودن تفسیرپذیری مدل‌ها همچنان باقی است. در مجموع، ترکیب دانش علوم رایانه و علوم قرآنی در قالب سامانه‌ای خودکار برای طبقه‌بندی موضوعی آیات، می‌تواند روشی کارآمد و مکمل برای روش‌های سنتی تفسیری و فهرست‌نگاری فراهم آورد و مسیر جدیدی را برای پژوهش‌های قرآنی در عصر دیجیتال هموار سازد.

نوع مقاله:

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۲/۰۷

تاریخ بازنگری: ۱۴۰۴/۰۴/۰۵

تاریخ پذیرش: ۱۴۰۴/۰۶/۰۳

تاریخ انتشار: ۱۴۰۴/۱۲/۲۰

کلیدواژه‌ها:

طبقه‌بندی موضوعی،

یادگیری ماشین،

یادگیری عمیق،

طبقه‌بندی چندبرچسبی،

ترنسفورمر،

متن کاوی قرآن.

استناد: الیاسی، محمد (۱۴۰۴). طراحی مدل مفهومی هوش مصنوعی XAI با رویکرد جهاد تبیین. علوم انسانی و اسلامی

دیجیتال، ۱ (۲)، ۳۰۳-۳۴۴. <https://doi.org/10.22034/disah.2025.2076072.1054.344-303>



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

## مقدمه

قرآن کریم کتاب مقدس نزدیک به دو میلیارد مسلمان در سراسر جهان است. این کتاب جامع الهی به زبان عربی در بردارنده مجموعه عظیمی از معارف و موضوعات گوناگون درباره اعتقادات، اخلاق، احکام و تاریخ بشریت است. اما برخلاف کتاب‌های معمول، آیات قرآن بر اساس ترتیب نزول یا موضوع طبقه‌بندی نشده‌اند؛ بلکه هر سوره و حتی هر آیه ممکن است به مضامین مختلفی بپردازد. همین امر دسترسی سریع به تمامی آیات مرتبط با یک موضوع خاص را دشوار می‌سازد و پژوهشگران علوم قرآنی از دیرباز برای حل این مشکل به طبقه‌بندی موضوعی آیات پرداخته‌اند. به عنوان نمونه، تفسیر «القرآن و البیان» اثر دکتر محمد حسن الحمشی (دمشق) و کتاب «طبقات آیات» اثر خلیل‌الله صبری مجموعه‌ای از آیات را بر اساس حدود ۲۶ عنوان موضوعی تنظیم کرده‌اند. هرچند این تلاش‌های دستی ارزشمندند، اما فرایند فهرست‌نگاری موضوعی به صورت دستی بسیار زمان‌بر و مستلزم صرف دقت و نیروی انسانی فراوان است.

با ظهور رایانه و گسترش منابع متنی دیجیتال قرآن، زمینه برای به‌کارگیری روش‌های پردازش زبان طبیعی و یادگیری ماشین در تحلیل موضوعی قرآن فراهم شده است. طی دو دهه اخیر، پژوهش‌های متعددی در این راستا انجام گرفته است. برای مثال، سیدیکی و همکاران (۲۰۱۳) با به‌کارگیری مدل‌های احتمالاتی نهان مانند LDA به کشف ساختارهای موضوعی پنهان در متن قرآن پرداختند. نتایج آنها که در یک کنفرانس تخصصی قرآن در عربستان ارائه شد، چهارچوبی برای استخراج خودکار موضوعات قرآن به روش یادگیری بدون نظارت فراهم آورد. در ایران نیز تلاش‌هایی جهت ترکیب روش‌های علوم رایانه با فهم تفسیری قرآن صورت گرفته است؛ از جمله صالحی‌شهرودی و همکاران (۱۳۹۲) که با استفاده از داده‌کاوی و بهره‌گیری از رویکرد موضوعی علامه طباطبایی در تفسیر المیزان، به ایجاد یک پیکره موضوعی از آیات قرآن مبادرت کردند. آنها در این پژوهش موضوعات محوری معارف قرآنی را در هفت دسته کلی سامان‌دهی نموده و با آزمون‌های داده‌کاوی، ارتباطات معنایی نوینی میان آیات کشف کردند. این نوع تحقیقات نشان داد که استفاده از تکنیک‌های یادگیری ماشین می‌تواند به آشکارسازی روابط پنهان بین آیات کمک کند.

در سال‌های اخیر، پژوهشگران بین‌المللی نیز گام‌های بلندی در توسعه سامانه‌های خودکار برای طبقه‌بندی موضوعی قرآن برداشته‌اند. به عنوان نمونه، یک گروه پژوهشی در اندونزی مجموعه‌ای از مطالعات پیوسته انجام داده است که ابتدا با روش‌های کلاسیک یادگیری ماشین و سپس با یادگیری عمیق به دسته‌بندی موضوعات آیات پرداخته‌اند. در این مطالعات که از ترجمه انگلیسی قرآن (انتشارات صاحبی اینترنشنال) بهره گرفته شد، مشخص گردید حدود ۵۵٪ از آیات قرآن ذاتاً چندموضوعی یا چندبرچسبی هستند. این بدین معناست که بیش از نیمی از آیات می‌توانند همزمان به دو یا چند مقوله موضوعی تعلق داشته باشند و برچسب موضوعی یگانه برای آن‌ها کافی نیست. این یافته اهمیت به‌کارگیری روش‌های پیشرفته‌تر (نظیر الگوریتم‌های چندبرچسبی) را آشکار کرد. در ادامه، همان گروه پژوهشی با استفاده از الگوریتم‌های قوی‌تر، موفق به بهبود دقت دسته‌بندی شدند؛ به طوری که در پژوهش پرابوو و همکاران (۲۰۱۹) از ماشین بردار پشتیبان (SVM) برای طبقه‌بندی چندبرچسبی آیات استفاده شد و نتایج «چشمگیری» به‌دست آمد. به موازات این پیشرفت‌ها، محققان دیگری نیز از داده‌های ترجمه‌شده یا تفسیری بهره گرفتند. برای مثال، اکور و همکاران (۲۰۱۴) با محاسبه شباهت بین آیات بر اساس مدل‌های gram-n، روشی برای خوشه‌بندی آیات مشابه و دسته‌بندی سوره‌ها ارائه کردند. همچنین رولیاواتی و همکاران (۲۰۱۸) با تحلیل متن ترجمه اندونزیایی قرآن، چهار موضوع کلی (بهشت و جهنم؛ دنیا و آخرت؛ علم، عمل و جهاد؛ شب و روز و حیات و مرگ) را به عنوان خوشه‌های معنایی اصلی قرآن استخراج کردند. چنین موضوعاتی نشان دادند که آیات مکی نسبت به آیات مدنی تأکید بیشتری بر ایمان و آخرت دارند. مجموع این پژوهش‌ها، چه در داخل کشور و چه در عرصه بین‌الملل، بیانگر ضرورت و اهمیت توسعه یک سامانه خودکار برای طبقه‌بندی موضوعی آیات قرآن است؛ سامانه‌ای که بتواند با دقت و سرعت بالا، آیات مرتبط با یک موضوع را شناسایی کرده و در کنار روش‌های سنتی مفسران، ابزار مکملی برای فهم بهتر معارف قرآن کریم ارائه دهد.

## ۱. مفهوم‌شناسی اصطلاحات کلیدی

در این بخش مفاهیم و اصطلاحات پایه که در مقاله به کار رفته‌اند، تعریف می‌گردند تا چارچوب نظری بحث روشن شود.

**طبقه‌بندی موضوعی:** مقصود از طبقه‌بندی موضوعی (Classification Topic) سامان‌دهی اسناد متنی بر اساس محتوا و مضامین آن‌ها است. در زمینه قرآن، طبقه‌بندی موضوعی به معنای تخصیص یک یا چند برچسب موضوعی به هر آیه از قرآن کریم است. برای مثال، آیه‌ای ممکن است ذیل برچسب‌های «توحید»، «معاد» و «اخلاق» قرار گیرد. خروجی یک سیستم طبقه‌بندی موضوعی می‌تواند به صورت یک فهرست موضوعی باشد که در آن موضوعات کلی (نظیر ایمان، احکام، تاریخ انبیا و غیره) تعریف شده و آیات مرتبط ذیل هر موضوع فهرست می‌شوند. طبقه‌بندی موضوعی زیرمجموعه‌ای از بازیابی اطلاعات متنی محسوب می‌شود که هدف آن تسهیل جستجو و پیمایش در محتوای متنی بر اساس برچسب‌های معنایی است.

**یادگیری ماشین (Learning Machine):** یادگیری ماشین شاخه‌ای از هوش مصنوعی است که به طراحی الگوریتم‌هایی می‌پردازد که رایانه را قادر می‌سازند از داده‌ها الگو بیاموزد و بر اساس آن تصمیم‌گیری کند، بدون این‌که برای هر حالت به طور صریح برنامه‌نویسی شده باشد. در واقع در یادگیری ماشین، رایانه با مشاهده مثال‌های آموزشی (داده‌ها و برچسب‌های مربوطه) مدلی می‌سازد که بتواند برای داده‌های جدید پیش‌بینی یا دسته‌بندی انجام دهد. یادگیری ماشین شامل دو شاخه عمده یادگیری نظارت‌شده (مانند مسأله طبقه‌بندی آیات که مدل با آیات برچسب‌خورده آموزش می‌بیند) و یادگیری بدون نظارت (مانند خوشه‌بندی آیات بدون داشتن برچسب اولیه) است. در سال‌های اخیر، الگوریتم‌های یادگیری ماشین به طور گسترده در پردازش متن به کار گرفته شده و نتایج موفقیت‌آمیزی را رقم زده‌اند.

**یادگیری عمیق (Learning Deep):** یادگیری عمیق زیرمجموعه‌ای پیشرفته از یادگیری ماشین است که مبتنی بر شبکه‌های عصبی چندلایه می‌باشد. در یادگیری عمیق، مدل با استفاده از لایه‌های متعدد غیرخطی می‌تواند بازنمایی‌های پیچیده‌تری از داده‌ها بیاموزد. پیشرفت‌های اخیر در یادگیری عمیق، خصوصاً در حوزه پردازش زبان طبیعی، موجب جهش قابل توجهی در دقت و کارایی مدل‌های تحلیل متن شده است. به عنوان مثال، شبکه‌های عصبی پیچشی (CNN) و شبکه‌های حافظه‌دار (RNN مانند LSTM) توانسته‌اند در وظایف طبقه‌بندی متن عملکرد بهتری نسبت به الگوریتم‌های کلاسیک نشان دهند. تفاوت یادگیری عمیق با روش‌های ML سنتی در آن است که دیگر نیازی به استخراج ویژگی دستی از متن (مانند شمارش کلمات کلیدی) نیست، بلکه مدل عمیق

بازنمایی مناسب ویژگی‌ها را به‌طور خودکار یاد می‌گیرد. در حوزه قرآن‌پژوهی، یادگیری عمیق امکان تحلیل خودکار متون دینی در مقیاس بزرگ را فراهم کرده است؛ به‌طوری‌که مدل‌های عمیق می‌توانند پیچیدگی‌های زبانی قرآن (نظیر استعاره‌ها، مفاهیم ضمنی و بافت معنایی) را بهتر از روش‌های سطحی‌تر مدل کنند.

**طبقه‌بندی چندبرچسبی:** در مسائل کلاسیک طبقه‌بندی متنی، فرض بر آن است که هر سند (در اینجا هر آیه) تنها به یک برچسب یا طبقه تعلق دارد. حال آن‌که در طبقه‌بندی چندبرچسبی (Classification label-Multi) هر سند می‌تواند همزمان چند برچسب دریافت کند. این حالت زمانی رخ می‌دهد که دسته‌های موضوعی هم‌پوشانی دارند یا متن مورد نظر به‌طور ذاتی شامل چند مبحث است. طبقه‌بندی آیات قرآن نمونه بارزی از یک مسئله چندبرچسبی است، زیرا بسیاری از آیات دارای چندین مضمون مختلف هستند. در روش‌های یادگیری ماشین کلاسیک، برای مواجهه با چندبرچسبی بودن، از تکنیک‌هایی نظیر «دوتایی‌سازی برچسب» (Relevance Binary) یا زنجیره طبقه‌بندها (Chains Classifier) استفاده می‌شود. همچنین برای ارزیابی عملکرد مدل در این حالت، از شاخص‌هایی مانند ضریب همینگ (Loss Hamming) استفاده می‌گردد که میانگین خطای مدل در هر برچسب را می‌سنجد. به عنوان مثال، مقدار همینگ لاس حدود  $0/112$  که توسط یک مدل بیز ساده به‌دست آمده است نشان می‌دهد در مجموع  $11/2$  برچسب‌های آیات به‌اشتباه طبقه‌بندی شده‌اند. مدل‌های پیشرفته‌تر مانند SVM چندخروجی یا شبکه‌های عصبی چندبرچسبی قادرند همبستگی میان برچسب‌های مختلف را نیز در نظر بگیرند تا دقت بهتری حاصل شود.

**ترنسفورمر (Transformer):** ترنسفورمر نام معماری نوینی در شبکه‌های عصبی عمیق است که بر مبنای مکانیزم توجه (Attention) بنا شده و نخستین بار توسط واسوانی و همکاران معرفی شد. مدل‌های ترنسفورمر مانند BERT و GPT با بهره‌گیری از لایه‌های توجه چندسری قادرند وابستگی‌های دور و نزدیک کلمات در یک متن را به‌خوبی مدل کنند. برتری ترنسفورمرها در این است که برخلاف RNN ها، امکان موازی سازی محاسبات را داشته و در عین حال با استفاده از توجه، هر کلمه را در متن با تمام کلمات دیگر مرتبط می‌سازند. در حوزه پردازش متون دینی نظیر قرآن، ترنسفورمرها انقلابی ایجاد کرده‌اند؛ زیرا می‌توانند بافت معنایی عبارات قرآنی را با دقتی بی‌سابقه درک کنند. برای مثال، مدل‌های زبانی مبتنی بر ترنسفورمر توانسته‌اند ظرافت‌های زبانی

قرآن (مثل ارائه مفاهیم متضاد در آیات مختلف) را بهتر از روش‌های قبل تشخیص دهند. در یک مطالعه میان‌رشته‌ای جدید در ایران، از مدل RoBERTa برای تشخیص احساسات در متن قرآن استفاده شد و دقت حدود ۷۷٪ به دست آمد که نشان‌دهنده توان بالای این مدل‌ها در فهم لایه‌های معنایی متن است. با توجه به موفقیت ترنسفورمرها، انتظار می‌رود در طبقه‌بندی موضوعی آیات نیز به‌کارگیری این مدل‌ها (مثلاً یک نسخه تخصصی از BERT برای قرآن) بتواند عملکرد سامانه را بهبود بخشد.

**تحلیل موضوعی متون دینی:** منظور از این اصطلاح، به‌کارگیری روش‌های کامپیوتری برای استخراج و سازمان‌دهی مفاهیم و موضوعات از متون دینی (قرآن، حدیث و...) است. متن‌های دینی معمولاً ساختار خطی ساده‌ای ندارند و تفسیر آن‌ها وابسته به فهم عمیق معانی، شان نزول، سیاق و دیگر عوامل است. تحلیل موضوعی متون دینی به کمک رایانه چالشی دوگانه دارد: از یک سو باید ابزارها و الگوریتم‌های پردازش زبان طبیعی را متناسب با زبان فاخر و پیچیده این متون تطبیق داد، و از سوی دیگر نتایج حاصله را به شکل مفید برای پژوهشگران علوم اسلامی ارائه نمود. برای نمونه، در پروژه صالحی‌شهرودی و همکاران (۱۳۹۲) تلاش شد تا با تلفیق دانش تفسیر المیزان و روش‌های داده‌کاوی، ارتباطات معنایی آیات استخراج و ذیل موضوعات تعریف شده در مقدمه المیزان مرتب شود. چنین رویکردهایی در واقع نوعی تحلیل موضوعی رایانه‌ای را نشان می‌دهند که مکمل تلاش مفسران در تبیین معارف دینی است.

## ۲. روش‌های یادگیری ماشین کلاسیک برای دسته‌بندی آیات

نخستین تلاش‌ها برای خودکارسازی دسته‌بندی موضوعی قرآن، از به‌کارگیری الگوریتم‌های یادگیری ماشین کلاسیک آغاز شد. در این روش‌ها، مراحل کلی کار بدین صورت بود که ابتدا متن آیات به صورت برداری از ویژگی‌ها (معمولاً ویژگی‌های آماری Words-of-Bag مانند وزن TF-IDF کلمات) نمایشی می‌شد، سپس یک الگوریتم کلاسیک یادگیری نظارت‌شده بر روی این بردارها آموزش می‌دید تا هر آیه را به یک یا چند دسته تخصیص دهد. تعدادی از مهم‌ترین الگوریتم‌های کلاسیک که در این حوزه آزموده شده‌اند عبارت‌اند از:

**بیز ساده (Bayes Naive):** این الگوریتم مبتنی بر قانون بیز و فرض استقلال ویژگی‌ها است. پانه و همکاران (۲۰۱۸) با استفاده از نسخه چندجمله‌ای بیز ساده، آیات ترجمه‌شده قرآن را در یک مسئله چندبرچسبی طبقه‌بندی کردند. آن‌ها پس از پیش‌پردازش متون (شامل پاک‌سازی، ریشه‌یابی و حذف کلمات توقف) و استخراج ویژگی‌های IDF-TF، مدل بیز ساده را آموزش دادند و بهترین مقدار خطای همینگ را ۰/۱۲۴۷ گزارش کردند. این مقدار خطا نشان‌دهنده عملکرد قابل قبول مدل در تشخیص همزمان موضوعات هر آیه بود. مزیت بیز ساده سادگی و سرعت آن است، هرچند فرض استقلال کلمات در متن قرآن که واژگان گاه به شدت وابسته به هم‌اند، چندان واقع‌بینانه نیست.

**درخت تصمیم و جنگل تصادفی:** درخت تصمیم (Tree Decision) الگوریتمی است که بر اساس پرسش‌های بلی/خیر بر ویژگی‌ها، داده‌ها را به دسته‌ها تفکیک می‌کند. جنگل تصادفی (Forest Random) مجموعه‌ای از درختان تصمیم است که با رای‌گیری، دقت بیشتری حاصل می‌کنند. در مسأله طبقه‌بندی قرآن نیز می‌توان از این روش‌ها استفاده کرد. اگرچه در ادبیات موجود استفاده مستقیمی از جنگل تصادفی برای آیات قرآن گزارش نشده، اما شباهت این مسأله با سایر مسائل طبقه‌بندی متن (که در آن‌ها RF عملکرد خوبی داشته) نشان می‌دهد قابل به‌کارگیری است. برای مثال، در یک مطالعه مرتبط بر متون حدیث، جنگل تصادفی توانست دسته‌های موضوعی احادیث را با دقت قابل ملاحظه‌ای تفکیک کند (al et Alotaibi, ۲۰۱۸). بنابراین انتظار می‌رود بر روی قرآن نیز نتایج مشابهی بدهد، هرچند پیچیدگی زبان قرآن چالش‌برانگیزتر است.

**ماشین بردار پشتیبان (SVM):** یکی از قدرتمندترین الگوریتم‌های کلاسیک برای طبقه‌بندی به‌شمار می‌آید که با یافتن یک ابرصفحه بهینه، داده‌ها را تفکیک می‌کند. این الگوریتم به‌خاطر توانایی در تعمیم به داده‌های با ابعاد بالا، برای متن که دارای ویژگی‌های بسیار (واژگان) است مناسب می‌باشد. پرابوو و همکاران (۲۰۱۹) از SVM به عنوان طبقه‌بند پایه برای مسأله چندبرچسبی آیات بهره گرفتند. آن‌ها با اعمال تکنیک‌های کاهش بُعد (استفاده از LSA) و اعتبارسنجی  $k$ -بخش، مدل SVM خود را بهینه کرده و خطای همینگ را نسبت به روش‌های قبلی کاهش دادند. هرچند در آن مقاله عدد دقیقی برای دقت مدل ذکر نشده، اما توصیف «نتایج برجسته» نشان می‌دهد SVM توانسته‌بود عملکرد بهتری نسبت به بیز ساده و TAN داشته باشد. در واقع، SVM به‌واسطه مرز تصمیم‌گیری منعطف و استفاده از کرنل‌های مختلف (خطی، چندجمله‌ای، RBF)، قابلیت تفکیک مناسبی حتی در حضور برچسب‌های متعدد برای هر نمونه از خود نشان داد.

سایر روش‌ها: به‌جز موارد فوق، الگوریتم‌های کلاسیک دیگری نیز قابلیت اعمال بر این مسأله را دارند. از جمله نزدیک‌ترین همسایه (NN-k) که بر اساس شباهت بردار آیات عمل می‌کند، یا روش‌های دسته‌بند مجموعه‌ای (Ensemble) نظیر بگینگ و بوستینگ. به عنوان نمونه، ایزاتی و همکاران (۲۰۱۸) از یک نوع مدل گرافی احتمالیس (TAN، شبکه بیزی درخت‌افزایی) استفاده کردند و به خطای همینگ ۰/۱۱۲ دست یافتند که اندکی بهتر از بیز ساده بود. همچنین می‌توان به ترکیب چند مدل (Ensemble) اشاره کرد که به صورت همزمان خروجی چند طبقه‌بند را برای هر آیه در نظر گرفته و با رای‌گیری یا میانگین‌بندی، تصمیم نهایی را می‌گیرد. اگرچه در متون پژوهشی قرآن به‌صراحت استفاده از ensemble گزارش نشده، اما در یک مطالعه آزمایشی در اندونزی ترکیب چند مدل طبقه‌بندی منجر به بهبود نتایج برای برچسب‌گذاری آیات شده است (al et Ibrahim, ۲۰۲۰).

به‌طور خلاصه، رویکردهای یادگیری ماشین کلاسیک نقش پیشگام را در خودکارسازی دسته‌بندی موضوعی قرآن داشته‌اند. این روش‌ها اغلب نیازمند مرحله استخراج ویژگی دستی (مانند محاسبه وزن کلمات) بوده و تا حدی به کیفیت آن وابسته‌اند. با این حال، آنها پایه لازم برای روش‌های پیشرفته‌تر را فراهم کردند و نشان دادند که می‌توان به جای تکیه صرف بر فهرست‌های دستی مفسران، از قدرت محاسباتی برای سازمان‌دهی موضوعی قرآن بهره برد.

### ۳. کاربرد یادگیری عمیق و مدل‌های ترنسفورمر در تحلیل متون قرآنی

با پیشرفت فناوری و افزایش توان محاسباتی، یادگیری عمیق به عنوان نسل جدیدی از روش‌های یادگیری ماشین در بسیاری از حوزه‌ها از جمله پردازش زبان طبیعی (NLP) انقلابی ایجاد کرد. حوزه مطالعات قرآنی نیز از این تحول بی‌بهره نبوده است. در سال‌های اخیر، پژوهشگران از انواع شبکه‌های عصبی عمیق برای بهبود دقت و کارایی سیستم‌های طبقه‌بندی موضوعی قرآن استفاده کرده‌اند. در این بخش، به مهم‌ترین کاربردهای یادگیری عمیق و مدل‌های ترنسفورمر در تحلیل متون قرآنی می‌پردازیم.

۳/۱. شبکه‌های عصبی عمیق (DNN) و حافظه‌دار: اولین کاربردهای یادگیری عمیق در طبقه‌بندی آیات عمدتاً با بهره‌گیری از شبکه‌های عصبی استاندارد صورت گرفت. برای مثال، الاشقر (۲۰۱۹) از دو معماری رایج یعنی شبکه عصبی بازگشتی (RNN) و شبکه عصبی پیچشی (CNN)

برای دسته‌بندی موضوعی آیات (بر اساس متن عربی اصلی قرآن) استفاده کرد. در آن پژوهش، ابتدا بردارهای تعبیه کلمات (Vec2Word) برای نمایش بردارهای معنایی واژگان قرآن آموزش داده شد؛ سپس هر آیه به صورت دنباله‌ای از بردارهای کلمات به مدل‌های عمیق داده شد. خروجی مدل، توزیع احتمال موضوعات تعریف‌شده (۱۲ موضوع اصلی) برای هر آیه بود. نتایج نشان داد که هر دو مدل RNN و CNN توانستند با دقت بسیار خوبی آیات را طبقه‌بندی کنند، به طوری که بهترین مدل (ترکیب شبکه CNN) به دقت حدود ۹۰٪ و مقدار F1 معادل ۸۱/۹۳٪ دست یافت. این ارقام نسبت به روش‌های کلاسیک گذشته جهش قابل توجهی داشت و اهمیت یادگیری عمیق را آشکار ساخت. دلیل این بهبود چشمگیر آن است که شبکه‌های عصبی عمیق توانایی فراگیری الگوهای پیچیده‌تر زبانی (مانند ترتیب کلمات، ترکیب معنایی عبارات و هم‌بستگی مفاهیم) را دارند که در روش‌های سنتی نادیده گرفته می‌شد. به عنوان مثال، RNN‌ها با داشتن حافظه داخلی قادرند ارتباط میان جملات طولانی را حفظ کنند و CNN‌ها با فیلترهای محلی می‌توانند عبارات کلیدی مهم در آیه را تشخیص دهند. ترکیب این دو (مدل‌های هیبریدی) نیز در برخی پژوهش‌ها به کار رفته و عملکرد را بهتر کرده است (al et Chaudhary, ۲۰۲۰).

۳/۲. مدل‌های ترنسفورمر و زبان‌های از پیش آموزش دیده: در سال‌های اخیر مدل‌های ترنسفورمر مانند BERT، RoBERTa، XLNet و ... پیش‌تاز مسابقات گوناگون NLP شده‌اند. این مدل‌ها که عموماً بر حجم عظیمی از متون عمومی از پیش آموزش دیده‌اند (Models Language)، می‌توانند با اندکی آموزش تکمیلی (Fine-tuning) برای کارهای خاص به کار گرفته شوند. در حوزه قرآن پژوهی، به کارگیری این مدل‌ها تازه در ابتدای راه است اما نتایج اولیه بسیار امید بخش بوده است. یک نمونه را می‌توان در کار کرمی و همکاران (۱۴۰۱) دید که از مدل ParsBERT (یک ترنسفورمر از پیش آموزش شده فارسی) برای تحلیل احساسات در ترجمه فارسی قرآن استفاده کردند. اگرچه موضوع آن پژوهش احساسات بود نه دسته‌بندی موضوعی، اما نشان داد مدل‌های مبتنی بر ترنسفورمر قادرند مفاهیم پنهان متن قرآن را بهتر درک کنند و دقت بالاتری نسبت به مدل‌های سنتی به دست آورند. همچنین در پژوهش کرمی، ذکر شده که «پیشرفت‌های اخیر، به ویژه استفاده از مدل‌های زبان مبتنی بر ترنسفورماتور، نتایج پیشرفته‌ای را به همراه داشته است که پیشی گرفتن از آنها به راحتی دشوار است». این بیانگر قدرت فوق‌العاده مدل‌های ترنسفورمر در پردازش متن‌هایی چون قرآن است.

از سوی دیگر، در جهان عرب کار روی ترنسفورمرهای ویژه متن قرآن آغاز شده است. برای مثال، BERT عربی بر روی متن قرآن نیز آزمایش شده و دیده شده که روابط معنایی بین آیات (نظیر آیات مرتبط با یک داستان یا حکم) را به خوبی در بردارهای نهفته بازنمایی می‌کند (Biltawi et al., 2021). همچنین مدل‌های چندزبانه مانند R-XLM این قابلیت را دارند که هم زمان از متن اصلی عربی و ترجمه‌های قرآن یاد بگیرند و به نوعی پلی بین زبان‌ها بزنند. این می‌تواند طبقه‌بندی موضوعی را بهبود بخشد؛ زیرا مثلاً مدلی که هم قرآن عربی و هم ترجمه انگلیسی را می‌بیند، شاید بهتر بتواند مفهوم آیه را درک کند (به کمک سیگنال‌های هر دو زبان). البته این حوزه بسیار جدید است و هنوز نتایج قطعی منتشر نشده، اما روند تحقیقات حکایت از حرکت به سوی بهره‌گیری هر چه بیش‌تر از مدل‌های ترنسفورمر از پیش‌آموزش‌شده ویژه قرآن دارد.

**۳/۳. بهبود عملکرد و چالش‌ها:** شبکه‌های عمیق معمولاً پارامترهای زیادی دارند و برای عملکرد بهینه نیازمند مجموعه‌داده بزرگی هستند. یکی از چالش‌های طبقه‌بندی آیات قرآن کمبود داده‌های برچسب‌خورده کافی است؛ چرا که برچسب‌گذاری موضوعی آیات امری اجتهادی بوده و تهیه دیتاستی جامع (مثلاً مجموعه‌ای استاندارد از آیات با برچسب‌های موضوعی مورد توافق کارشناسان) زمان‌بر است. برخی پژوهش‌ها این مشکل را با استفاده از منابع ثانویه حل کرده‌اند: برای مثال بهره‌گیری از فهرست موضوعی موجود در تفاسیر یا کتاب‌ها به عنوان برچسب. در پژوهش پرابوو (2019) از دسته‌بندی تفسیری آماده (اثر یک مفسر) برای برچسب‌دهی استفاده شد. رویکرد دیگر استفاده از یادگیری انتقالی است؛ بدین معنا که مدل‌های از پیش‌آموزش‌دیده بر متون غنی (مثلاً کل متون عربی ادبی) گرفته شده و سپس بر قرآن اختصاصی‌سازی شوند. چنین روشی نیاز به داده برچسب‌خورده زیاد را کاهش می‌دهد و توان مدل را افزایش می‌دهد (al et Bashir, 2023). یکی دیگر از ابزارهای ارتقای عملکرد در مدل‌های عمیق، ترکیب اطلاعات جنبی نظیر ویژگی‌های زبانی (نحو، صرف) با شبکه عصبی است. برای نمونه، کرمی و همکاران افزون بر بردارهای ترنسفورمری، برچسب‌های نقش دستوری (POS) را نیز به شبکه دادند تا مدل از ساختار جمله آگاه شود و این منجر به بهبود دقت شد.

در مجموع، بهره‌گیری از یادگیری عمیق و ترنسفورمرها در تحلیل قرآنی نتایج بسیار خوبی به همراه داشته و افق‌های تازه‌ای را گشوده است. مدل‌های عمیق نه تنها دقت طبقه‌بندی را بالا برده‌اند،

بلکه قابلیت‌های جدیدی نظیر یادگیری روابط معنایی بلندبرد، تلفیق اطلاعات متنی و فرامتنی و یادگیری نمایشی مشترک چندزبانه را فراهم کرده‌اند که در نسل قبل میسر نبود. البته کاربرد این مدل‌ها مستلزم دسترسی به سخت‌افزار مناسب (GPU) و دانش تنظیم پارامترها است و برای بسیاری از پژوهشگران علوم اسلامی ممکن است چالش برانگیز باشد. از این رو، همکاری میان متخصصان رایانه و قرآنی در طراحی، آموزش و استفاده از این مدل‌ها امری کلیدی است که به توسعه ابزارهای کاربردی‌تر منجر خواهد شد.

#### ۴. طبقه‌بندی چندبرچسبی در آیات قرآن (چالش‌ها و مزایا)

همان‌گونه که اشاره شد، یکی از ویژگی‌های منحصربه‌فرد مسئله طبقه‌بندی موضوعی قرآن، چندبرچسبی بودن آن است؛ به این معنا که یک آیه می‌تواند ذیل چند موضوع قرار گیرد. این واقعیت از محتوای قرآن سرچشمه می‌گیرد که در بسیاری موارد پیونددهنده چند مفهوم در یک بیان واحد است. برای مثال، آیه ۲۶۱ سوره بقره هم‌زمان به انفاق (عمل اجتماعی)، ایمان به خدا (اعتقادی) و پاداش آخرت (معادی) اشاره دارد. بنابراین هرگونه سیستم موضوع‌یابی آیات باید امکان نسبت‌دادن چند موضوع به یک آیه را داشته باشد.

۴/۱. **ضرورت چندبرچسبی:** روش‌های سنتی فهرست‌نگاری قرآن نیز به صورت چندموضوعی بوده‌اند؛ مثلاً در کتاب‌های معجم‌الموضوعی، یک آیه ممکن است در چند جای کتاب (زیر چند عنوان) تکرار شود. پژوهش al et Prabowo (۲۰۱۹) نشان داد که حدود ۵۵٪ آیات قرآن دارای بیش از یک برچسب موضوعی هستند. این بدان معناست که اگر مدل را به صورت تک‌برچسب آموزش دهیم (اجبار به انتخاب یک موضوع برای هر آیه)، در بیش از نیمی موارد بخشی از اطلاعات موضوعی آیه از بین می‌رود. از این رو، اکثر پژوهش‌های جدید مسئله را به صورت چندبرچسبی دیده‌اند و از الگوریتم‌ها و معیارهای مختص آن استفاده کرده‌اند.

۴/۲. **روش‌های چندبرچسبی:** برای تطبیق الگوریتم‌های یادگیری ماشین با خروجی چندبرچسبی، دو رویکرد کلی وجود دارد: یکی تبدیل مسئله به چندین مسئله تک‌برچسب (مثلاً یک طبقه‌بند جداگانه برای هر موضوع) و دیگری طراحی الگوریتم‌هایی که خروجی مجموعه مقدراری تولید کنند. روش نخست ساده‌تر است؛ برای نمونه روش دوتایی‌سازی (Relevance Binary) که

در آن به ازای هر برچسب یک SVM جداگانه آموزش داده می‌شود تا تصمیم بگیرد آن برچسب برای آیه فعال هست یا نه. این روش در عین سادگی، وابستگی میان برچسب‌ها را در نظر نمی‌گیرد. برای رفع این نقیصه، روش‌های مانند Chains Classifier پیشنهاد شده‌اند که خروجی‌های طبقه‌بند‌های قبل را به ورودی بعدی می‌دهند تا وابستگی مدل شود (al et Read, ۲۰۱۵). به عنوان مثال، اگر برچسب «ایمان» و «عمل صالح» اغلب با هم می‌آیند، زنجیره طبقه‌بندی‌ها می‌تواند این همبستگی را بیاموزد. در مسئله قرآن نیز می‌توان تصور کرد دانستن این‌که آیه‌ای اخلاقی است، احتمال اجتماعی بودن آن را افزایش می‌دهد و مدل باید چنین دانشی را لحاظ کند. در بین تحقیقات مورد بررسی، برخی به این وابستگی‌ها توجه داشته‌اند. مثلاً مدل TAN مورد استفاده al et Izzaty (۲۰۱۸) یک ساختار گراف احتمالاتی را بین برچسب‌ها مدل می‌کرد که می‌تواند روابط میان موضوعات را بیاموزد.

**۴/۳. ارزیابی چندبرچسبی:** با توجه به خروجی چندگانه، معیارهای ارزیابی عملکرد نیز متفاوت از حالت تک برچسب است. پرکاربردترین معیار در این حوزه همان خطای همینگ است که به‌طور متوسط چه کسری از برچسب‌های حقیقی اشتباهاً پیش‌بینی شده‌اند. هرچه این مقدار کمتر باشد بهتر است. برای درک آن، تصور کنید در یک آیه با ۳ موضوع واقعی، مدل ۲ موضوع را درست و یکی را غلط پیش‌بینی کرده، خطای همینگ برای آن آیه  $0.33$  خواهد بود. معیار دیگر دقت مثالی (Accuracy based-Example) است که سخت‌گیرانه‌تر است و تنها در صورتی یک آیه را درست شناسایی شده می‌شمارد که تمامی برچسب‌های مربوط به آن درست پیش‌بینی شوند. از آنجا که این حالت در عمل کمتر رخ می‌دهد، برخی پژوهشگران از میانگین معیارهایی چون precision و recall بر روی مجموعه برچسب‌ها استفاده می‌کنند (مانند میانگین micro و macro). به عنوان نمونه، در پژوهش Prabowo (۲۰۱۹) از Loss Hamming به عنوان معیار اصلی استفاده شد و برای مدل SVM مقدار بسیار پایینی به دست آمد که نشان‌دهنده عملکرد مناسب بود. در پژوهش Alashqar (۲۰۲۳) نیز دقت، recall و F1 به‌طور جداگانه برای خروجی چندبرچسبی محاسبه گردید که همگی در حدود ۹۰٪ یا بالاتر بودند.

**۴/۴. مزایا و چالش‌های چندبرچسبی:** مزیت اصلی اجازه چندبرچسبی در این است که تمام وجوه معنایی آیه انعکاس می‌یابد و کاربر نهایی (پژوهشگر) می‌تواند آیه را ذیل هر موضوع مرتبط بیابد. برای مثال، آیه «انما المؤمنون اخوة...» (حجرات: ۱۰) ممکن است هم در موضوع «اخلاق

اجتماعی» و هم «ایمان» برچسب بخورد و در هر دو جای فهرست ظاهر شود. این جامعیت، نمای کاملی از ارتباطات متن را فراهم می‌کند. از سوی دیگر، چندبرچسبی بودن کار مدل را دشوارتر می‌کند؛ زیرا مدل باید بیاموزد چه ترکیباتی از برچسب‌ها با هم امکان‌پذیرند یا رایج‌ترند. قرآن کریم کتابی است که مفاهیم در آن به ظرافت تنیده شده‌اند و گاه تشخیص مرز میان موضوعات به قضاوت انسانی نیاز دارد. حتی متخصصان ممکن است در تعیین اینکه یک آیه داخل کدام دسته‌ها قرار بگیرد اختلاف نظر داشته باشند. بنابراین تهیه دیتاست آموزشی منسجم در این زمینه چالش‌برانگیز است. راهکار ممکن، تعریف یک مجموعه نسبتاً محدود از موضوعات کلی است که اکثر مفسران روی آن توافق دارند (مانند ۱۵ موضوع استفاده‌شده در پژوهش Prabowo یا ۱۲ موضوع به‌کاررفته در Alashqar). البته هرچه تعداد برچسب‌های ممکن کمتر باشد، مشکل هم‌پوشانی کمتر می‌شود ولی جزئیات نیز از دست می‌رود. برخی منابع کلاسیک تا ده‌ها عنوان موضوعی برای قرآن در نظر گرفته‌اند (مثلاً ۲۶ موضوع در کتاب طبقات آیات) در حالی که پژوهش‌های کامپیوتری معمولاً در حد ۱۰ تا ۲۰ موضوع کلی کار کرده‌اند. این اختلاف، نشانگر یکی دیگر از چالش‌های طبقه‌بندی موضوعی است: نبود یک چارچوب موضوعی استاندارد. در مواجهه با این چالش، رویکردی که می‌توان اتخاذ کرد استفاده از ساختارهای سلسله‌مراتبی است؛ مثلاً ابتدا طبقه کلی (اعتقادی، تاریخی، فقهی و ...) تعیین شود و سپس زیرمجموعه‌های دقیق‌تر. چنین ساختاری هم در داده‌های آموزشی راحت‌تر قابل حصول است (چون توافق روی کلیات بیشتر است) و هم خروجی برای کاربر قابل فهم‌تر.

به طور خلاصه، در طراحی سامانه خودکار ما باید چندبرچسبی بودن را به عنوان یک اصل جدایی‌ناپذیر بپذیریم و الگوریتم مناسب برای آن برگزینیم. به کارگیری روش‌های یادگیری چندبرچسبی، هرچند پیچیدگی را افزایش می‌دهد، اما منجر به سیستمی می‌شود که با واقعیت متن قرآن سازگارتر و برای بهره‌برداران غنی‌تر است. نتایج تحقیقات نیز موید این است که مدل‌های چندبرچسبی می‌توانند با موفقیت این کار را انجام دهند و به دقت‌های بالا دست یابند.

## ۵. رویکردهای بدون نظارت برای کشف خوشه‌های موضوعی پنهان

علاوه بر روش‌های نظارت‌شده که نیازمند برچسب‌گذاری صریح آیات هستند، رویکردهای یادگیری بدون نظارت نیز در سال‌های اخیر برای کشف ساختارهای معنایی قرآن به کار گرفته شده است. در این روش‌ها تلاش می‌شود خوشه‌ها یا موضوعات پنهان در متن قرآن بدون دانش پیشینی (برچسب انسانی) شناسایی گردد. این رویکرد می‌تواند مکمل مفیدی برای طبقه‌بندی نظارت‌شده باشد؛ زیرا گاهی الگوها و موضوعاتی را آشکار می‌کند که در چارچوب طبقه‌بندی از پیش تعریف‌شده وجود ندارند.

۵/۱. خوشه‌بندی متون قرآنی: خوشه‌بندی (Clustering) به معنای گروه‌بندی اسناد متنی بر اساس شباهت در محتوا است. در زمینه قرآن، خوشه‌بندی آیات به صورت خودکار بر اساس شباهت کلمات یا مفاهیم می‌تواند منجر به تشکیل خوشه‌هایی از آیات با موضوعات مشترک شود. به عنوان نمونه، اگر الگوریتمی مجموعه آیات را به ۱۰ خوشه تقسیم کند، شاید خوشه‌ای شامل آیاتی درباره بهشت و جهنم، خوشه‌ای حاوی آیاتی پیرامون داستان‌های انبیا و ... باشد. یکی از تکنیک‌های متداول خوشه‌بندی متن، محاسبه بردار ویژگی برای هر متن (مثلاً بردار IDF-TF کلمات آیه) و سپس استفاده از الگوریتمی نظیر Means-k برای گروه‌بندی بردارهای مشابه است. یک پروژه دانشجویی در دانشگاه علم و صنعت ایران (رحیمی‌بافقی و پورباقری، ۱۳۹۳) دقیقاً بدین روش عمل کرد: ابتدا هر آیه به همراه ترجمه فارسی و تفسیرش برداردهی شد، سپس با روش Means-k خوشه‌هایی از آیات مشابه تشکیل گردید. گزارش این پروژه نشان داد که آیات مرتبط واقعاً در یک خوشه قرار گرفته‌اند (برای مثال آیاتی که درباره نماز بودند در یک خوشه جمع شدند). هرچند جزئیات کمی از نتایج آن منتشر شده، اصل ایده که خوشه‌بندی می‌تواند ارتباط معنایی آیات را آشکار سازد تایید گشت.

۵/۲. مدل‌های موضوعی آماری: روش قدرتمندتر در کشف خودکار موضوعات، استفاده از مدل‌های موضوعی مانند LDA (تخصیص پنهان دیریکله) است. LDA یک مدل مولد احتمالاتی است که هر سند (آیه) را ترکیبی از چند موضوع پنهان فرض می‌کند و هر موضوع را توزیعی از کلمات. اجرای LDA بر قرآن می‌تواند مثلاً ۲۰ موضوع نهفته بیابد که هر موضوع لیستی از واژه‌های پرتکرار

دارد (که بازتاب یک مضمون‌اند). سیدیکی و همکاران (۲۰۱۳) اولین کسانی بودند که LDA را بر قرآن (ترجمه انگلیسی) اعمال کردند. آنها از طریق این مدل توانستند ساختار موضوعی ضمنی قرآن را در چند محور اصلی نشان دهند. همچنین الهورات (۲۰۱۵) نیز با به‌کارگیری مدل‌های مولد (PLSA و LDA) موضوعات معنایی قرآن را استخراج کرد و گزارش نمود که برخی موضوعات به‌دست‌آمده با مفاهیم شناخته‌شده تفسیری منطبق‌اند (مثل موضوعاتی مربوط به قصص پیامبران، احکام فقهی و غیره). به عنوان نمونه، یکی از موضوعات کشف شده مجموعه کلماتی مانند «موسی، فرعون، بنی‌اسرائیل، دریا، نجات» بود که به‌وضوح داستان حضرت موسی (علیه السلام) را تشکیل می‌داد. جالب اینجاست که LDA این خوشه را بدون دانستن هیچ اطلاعات قبلی خود استخراج کرده بود.

مزیت روش‌های مبتنی بر مدل موضوعی آن است که قادرند ارتباطات پنهان و دوربرد را بیابند؛ یعنی ممکن است آیاتی از سوره‌های مختلف که ظاهراً پراکنده‌اند ولی از نظر مبحث مرتبط می‌باشند در یک موضوع جمع شوند. این همان هدفی است که یک مفسر با تتبع در کل قرآن انجام می‌دهد. برای نمونه، پنجو (۲۰۱۴) با استفاده از ترکیب LDA و تکنیک‌های بصری‌سازی، توانست شبکه‌ای از آیات و موضوعات قرآن ترسیم کند که نشان می‌داد چه آیاتی محورهای مشترک دارند.

۵/۳. **بازنمایی‌های توزیعی و خوشه‌بندی معنایی:** روش‌های جدیدتر بدون نظارت به‌جای تکیه صرف بر مدل‌های احتمالاتی ساده، از بردارهای تعبیه معنایی (Embeddings Semantic) استفاده می‌کنند. در این رویکرد، هر آیه به یک بردار متراکم در فضای با بُعد پایین تبدیل می‌شود به نحوی که بردارها بیانگر معنای آیات باشند (دو آیه مشابه بردارهای نزدیک به هم دارند). یکی از تکنیک‌های موفق در این زمینه (Vec2Doc Vector Paragraph) است که توسط Mikolov et al. (۲۰۱۴) معرفی شد. الشحمری و همکاران (۲۰۱۹) این روش را بر روی متن عربی قرآن به کار گرفتند. آن‌ها ابتدا با Vector Paragraph هر آیه را به بردار ۱۰۰ بُعدی نگاشتند که این بردار «خلاصه معنایی» آیه بود. سپس با خوشه‌بندی این بردارها توانستند آیات را در فضای برداری گروه‌بندی کنند. نتیجتاً خوشه‌های معنایی معناداری به‌دست آمد که هر خوشه تقریباً منطبق بر یک موضوع قرآنی بود. برای مثال، یکی از خوشه‌ها بردارهای آیات مرتبط با معاد و بهشت و جهنم را شامل می‌شد، خوشه‌ای دیگر بردارهای آیات مرتبط با احکام فقهی را. این تحقیق نشان داد که بازنمایی برداری آیات می‌تواند پایه قدرتمندی برای انواع تحلیل‌های بعدی باشد؛ چرا که پس از آن، تنها با محاسبه فاصله یا شباهت بین بردارها می‌توان میزان ارتباط موضوعی آیات را سنجید.

۴/۵. مزایا و محدودیت‌های روش‌های بدون نظارت: مزیت بزرگ این روش‌ها عدم نیاز به داده برچسب‌خورده انسانی است. از این رو به راحتی می‌توان کل قرآن (و حتی متون مرتبط مثل تفاسیر) را وارد مدل کرد و از مدل خواست خودش ساختاردهی کند. این امر خصوصاً وقتی ارزشمند است که می‌خواهیم کشف الگوهای جدید کنیم، نه صرفاً تأیید طبقه‌بندی‌های از قبل موجود. برای مثال، ممکن است مدل بدون نظارت تشخیص دهد آیاتی که درباره طبیعت (خورشید، ماه، ستارگان) سخن می‌گویند خوشه مستقلی تشکیل می‌دهند، در حالی که شاید در طبقه‌بندی‌های سنتی اینها ذیل «توحید» یا «علم خداوند» پراکنده شده باشند. چنین بینشی به پژوهشگر کمک می‌کند تا به روابط تازه‌ای پی‌برد.

با این حال، خروجی مدل‌های بدون نظارت نیاز به تفسیر انسانی دارد. یعنی وقتی LDA می‌گوید یک «موضوع» شامل کلمات  $X$  و  $Y$  و  $Z$  است، این انسان است که باید تشخیص دهد این مجموعه چه عنوانی می‌تواند داشته باشد. گاهی اوقات نیز مدل ترکیب‌های عجیب یا مبهم ارائه می‌کند که تفسیر روشنی ندارند. برای نمونه ممکن است تعدادی واژه مرتبط با هم نباشند و صرفاً از نظر آماری ظاهر شده باشند. بنابراین معمولاً بهترین کاربرد یادگیری بدون نظارت در کنار نظارت‌شده و دانش انسانی است: مدل خوشه‌های خام را می‌دهد و یک کارشناس علوم قرآنی آن‌ها را بازبینی و برچسب‌گذاری نهایی می‌کند.

از دیگر محدودیت‌ها، حساسیت این روش‌ها به انتخاب تعداد خوشه‌ها یا موضوعات است. الگوریتم Means-k باید تعداد  $k$  را بداند و LDA نیز تعداد موضوعات را ورودی می‌گیرد. انتخاب این عدد اگر نادرست باشد، یا خوشه‌ها بیش از حد کلی خواهند بود یا بیش از حد ریز و درهم. برخی پژوهش‌ها با آزمون و خطا یا با استفاده از معیارهایی مثل انسجام موضوعی (Topic Coherence) سعی کرده‌اند مقدار بهینه را بیابند. به عنوان مثال، رولیاواتی و همکاران پس از چند آزمون به این نتیجه رسیدند که ۴ موضوع کلی برای قرآن ترجمه اندونزیایی مناسب است و با همین مقدار نتایج قابل تفسیرتری گرفتند.

در مجموع، رویکردهای بدون نظارت نقش مکمل ارزشمندی در سامانه خودکار ما ایفا می‌کنند. این روش‌ها می‌توانند برای پیشنهاد دسته‌های موضوعی اولیه به کار روند؛ سپس آن دسته‌ها با کمک کارشناس اصلاح یا ادغام شوند و در نهایت برای مدلسازی نظارت‌شده استفاده گردند. همچنین

می‌توانند برای یافتن پیوندهای پنهان میان آیات بهره‌گیری شوند؛ مثلاً کشف کنند که دو آیه از دو سوره متفاوت احتمالاً در یک زمینه مشترک هستند، چیزی که شاید در ظاهر مشخص نباشد. چنین اطلاعاتی حتی در مطالعات تفسیری نیز می‌تواند الهام‌بخش باشد (نظیر کاری که صالحی‌شهرودی و مینایی با داده‌کاوی تفاسیر انجام دادند و روابط جدیدی یافتند).

## ۶. مطالعات موردی: معرفی پروژه‌ها و مقالات داخلی و خارجی

برای درک بهتر وضعیت فعلی این حوزه، در این بخش چند مطالعه شاخص، چه از پژوهش‌های داخلی ایران و چه در سطح بین‌المللی، مرور می‌شود. این مطالعات هر یک از جنبه‌ای خاص قابل توجه‌اند و جمع‌بندی آن‌ها دید روشنی از دستاوردها و کاستی‌های موجود به‌دست می‌دهد.

۱/ ۶. پروژه‌های داخلی (ایران): در ایران با وجود سابقه طولانی در علوم قرآنی، بهره‌گیری از هوش مصنوعی و یادگیری ماشین در دهه اخیر رشد یافته است. یکی از نخستین کارها، همان‌طور که ذکر شد، پژوهش صالحی‌شهرودی و همکاران (۱۳۹۲) بود که در نشریه «قرآن‌شناخت» منتشر شد. این تیم پژوهشی با راهنمایی یک متخصص هوش مصنوعی (دکتر بهروز مینایی) و یک متخصص علوم قرآن (دکتر صالحی‌شهرودی) تلاش کردند مفاهیم تفسیری علامه طباطبایی را به زبان فرمال داده‌کاوی تبدیل کنند. خروجی کار آنان ایجاد یک مجموعه برچسب‌خورده از آیات ذیل ۷ موضوع کلان (مانند ذات و اسماء الهی، افعال الهی، انسان در دنیا و آخرت، هدایت و ضلالت و ...) بود. سپس با اجرای چند الگوریتم داده‌کاوی روی این پیکره، توانستند مصادیقی از ارتباط آیات را کشف کنند که حتی در تفسیر المیزان تصریح نشده بود. این پروژه نشان داد که همکاری میان حوزه علوم اسلامی و مهندسی کامپیوتر نتایج جالبی به همراه دارد. با این حال، به دلیل محدودیت در تعداد موضوعات (فقط ۷ موضوع کلی) و اتکا به یک چارچوب تفسیری خاص، نمی‌توانست یک سامانه عمومی طبقه‌بندی ارائه دهد.

یکی دیگر از طرح‌های داخلی، پروژه خوشه‌بندی آیات قرآن در دانشگاه علم و صنعت (۱۳۹۳) بود که به صورت پایان‌نامه انجام شد. اگرچه خروجی رسمی پژوهشی از آن در دسترس نیست، اما مستندات نشان می‌دهد هدف آن یافتن آیات مشابه از طریق متن ترجمه و تفسیر بوده است. به بیان دیگر، این پروژه به جای برچسب‌گذاری موضوعی صریح، یک ابزار جستجوی معنایی آیات ارائه

داده است: کاربر یک آیه یا موضوع وارد می‌کند و آیات مشابه (از نظر معنایی) پیشنهاد می‌شوند. چنین سیستمی نوعی سامانه بازیابی آیات بر اساس محتوا است که می‌تواند برای پژوهشگران قرآنی بسیار مفید باشد. احتمالاً الگوریتم‌هایی شبیه به همان که در بخش خوشه‌بندی ذکر شد (بردارهای Means-k + IDF-TF یا محاسبه تشابه کسینوسی بین بردار آیات) در آن به کار رفته است.

از مراکز فعال در ایران در این زمینه باید به مرکز تحقیقات کامپیوتری علوم اسلامی نور اشاره کرد. این مرکز سال‌هاست نرم‌افزارهایی چون جامع تفاسیر نور و معجم موضوعی نور را عرضه کرده که با تلاش دستی پژوهشگران علوم اسلامی تهیه شده‌اند. اما اخیراً پروژه‌هایی برای خودکارسازی برخی قسمت‌ها در دستور کار قرار داده است. برای مثال، در نسخه‌های جدید نرم‌افزار تفاسیر نور، قابلیت یافتن آیات مشابه افزوده شده که گفته می‌شود با استخراج هم‌رخدادی آیات در تفاسیر به‌دست آمده است. به بیان ساده، اگر دو آیه مکرراً در یک پاراگراف تفسیر کنار هم آمده باشند، نرم‌افزار نتیجه می‌گیرد این دو آیه به هم مرتبطند و در بخش «آیات مرتبط» معرفی می‌کند. این رویکرد هرچند ساده است، اما یک گام در جهت بهره‌گیری از رایانه برای کشف ارتباطات قرآنی است و می‌توان آن را نوعی داده‌کاوی تفاسیر تلقی کرد. انتظار می‌رود مراکز نور، با در اختیار داشتن گنجینه عظیم متون قرآنی و حدیثی، بیش از پیش به روش‌های یادگیری ماشین روی آورند. خبرهایی مبنی بر پروژه پژوهشگاه علوم و فرهنگ اسلامی در قم برای «استخراج خودکار هستی‌نگاشت مفاهیم قرآنی» منتشر شده که رویکردی مشابه (استفاده از شبکه‌های عصبی برای ایجاد دانش‌بنیان قرآنی) دارد.

۲/۶. مطالعات و پروژه‌های خارجی: در خارج از ایران، کشورهای مختلف اسلامی و حتی محققان غربی به موضوع قرآن و فناوری اهتمام داشته‌اند. یکی از پیشگامان این عرصه پروژه Quranica Corpora در دانشگاه لیدز انگلستان بود. در این پروژه (حدود سال‌های ۲۰۱۰-۲۰۱۳)، گروهی به سرپرستی Atwell Eric و Dukes Kais کارهای بنیانگذاری انجام دادند: ابتدا پیکره نحوی قرآن (Corpus Arabic Quranic) را ساختند و سپس یک هستی‌نگاشت معنایی قرآن طراحی کردند که مفاهیم کلیدی قرآن (مانند شخصیت‌ها، مکان‌ها، اعمال) را به آیات مرتبط پیوند می‌داد. خروجی این پروژه اگرچه مستقیماً یک طبقه‌بندی موضوعی نبود، اما زیرساخت ارزشمندی برای برچسب‌زدن خودکار فراهم کرد. به عنوان مثال، پس از این بود که پژوهشگرانی توانستند از

برچسب‌های مفهومی اوتولوژی به عنوان خروجی‌های مدل طبقه‌بندی استفاده کنند (al et Yauri, ۲۰۱۳). یائوری و همکاران در سال ۲۰۰۹ با استفاده از یک اوتولوژی مبتنی بر OWL تلاش کردند آیات را بر اساس مفاهیم (نه صرفاً کلمات) بازیابی و دسته‌بندی کنند. هرچند آن زمان ابزارهای NLP به پیشرفتگی امروز نبود، اما ایده ترکیب دانش معنایی (اوتولوژی قرآن) با پردازش ماشینی مطرح شد که هنوز هم ایده‌آل محسوب می‌شود.

در جهان اسلام، کشورهایی نظیر اندونزی و مالزی نیز فعالیت چشمگیری داشته‌اند. دانشگاه‌هایی در اندونزی چندین پایان‌نامه و مقاله در این حوزه منتشر کرده‌اند که قبلاً به برخی اشاره کردیم (ایزاتی ۲۰۱۸، پانه ۲۰۱۸، پرابوو ۲۰۱۹). علاوه بر آنها، پروژه‌هایی برای تحلیل ترجمه‌های قرآن نیز انجام شده است. مثلاً al et Putra (۲۰۱۷) یک مرور منظم بر پژوهش‌های متن‌کاوی قرآن ترجمه اندونزیایی ارائه دادند. در مالزی و کشورهای عربی نیز کارهایی روی پیکره‌های قرآنی و دسته‌بندی موضوعی دیده می‌شود؛ از جمله Hawarat-Al (۲۰۱۵) که استخراج موضوعات را انجام داد و al et Shoieb (۲۰۰۹) که یک مدل مبتنی بر وردنت برای جستجوی معنایی آیات پیشنهاد کردند. همچنین اخیراً al et Bashir (۲۰۲۳) یک مقاله مروری جامع درباره پردازش زبان عربی برای پژوهش‌های قرآنی منتشر کرده‌اند که مجموعه متنوعی از ابزارها و روش‌های بکاررفته (از جمله برای تفسیر، تجوید، دسته‌بندی و ...) را گزارش می‌کند. این نشان می‌دهد جامعه علمی به سرعت در حال مستندسازی پیشرفت‌ها و ترسیم مسیر آینده است.

یک مطالعه جالب دیگر توسط al et Akour (۲۰۱۴) انجام شد که هدف آن اندازه‌گیری شباهت معنایی بین آیات و همچنین دسته‌بندی سوره‌ها بر اساس محتوای آیاتشان بود. آن‌ها با استفاده از تکنیک gram-n و محاسبه فاصله‌های برداری، معیاری معرفی کردند که تعیین می‌کرد دو آیه تا چه حد از نظر کلمات مشترک نزدیک‌اند. سپس با داشتن این ماتریس شباهت، توانستند سوره‌های قرآن را نیز خوشه‌بندی کنند (مثلاً سوره‌های مکی در یک خوشه و مدنی در خوشه دیگر قرار گرفتند، چون محتوای متفاوتی دارند). این نوع نگاه ماکرو (در سطح سوره) کمتر معمول است اما می‌تواند چشم‌انداز کلانی بدهد؛ مثلاً تأیید کند که سوره‌های مکی بیشتر پیرامون مباحث اعتقادی و اخروی‌اند و سوره‌های مدنی پیرامون فقه و اخلاق اجتماعی.

در مجموع، مطالعات موردی فوق نشان می‌دهد که:

۱. روش‌های گوناگونی از کلاسیک تا پیشرفته در این حوزه آزموده شده و هر کدام بخشی از مسیر را هموار کرده‌اند؛
۲. پژوهش‌ها جنبه بین‌رشته‌ای و همکاری میان متخصصان کامپیوتر و علوم قرآن را می‌طلبد (نمونه آن در ایران و انگلیس دیده شد)؛
۳. هنوز مسائل حل‌نشده و جای کار بسیاری باقی است (مثلاً تعمیم مدل‌ها به تمام زبان‌ها، یا توافق بر سر مجموعه موضوعات معیار).
۴. این مطالعات تجربه‌های ارزشمندی را فراهم آورده که می‌تواند راهنمای طراحی یک سامانه جامع خودکار باشد.

## ۷. مزایا و محدودیت‌های سامانه‌های خودکار طبقه‌بندی قرآنی در مقایسه با روش سنتی تفسیری و فهرست‌نویسی

اکنون که با جنبه‌های فنی و پژوهشی سامانه‌های خودکار آشنا شدیم، لازم است نگاهی نقادانه به مزایا و محدودیت‌های این سامانه‌ها در قیاس با شیوه‌های سنتی داشته باشیم. بدین ترتیب جایگاه و ارزش واقعی این نوآوری در مطالعات قرآنی بهتر درک می‌شود.

### ۷/۱. مزایا و نقاط قوت

سرعت و مقیاس‌پذیری: بزرگ‌ترین مزیت سامانه خودکار، سرعت بسیار بالای آن در پردازش حجم عظیم اطلاعات است. تحلیل دستی کل قرآن و استخراج موضوعی آن کاری است که سال‌ها زمان می‌برد؛ در حالی که یک الگوریتم رایانه‌ای می‌تواند در عرض چند دقیقه یا ساعت این کار را انجام دهد. با افزایش تولید محتوای الکترونیکی دینی، نیاز به چنین سرعتی بیش از پیش احساس می‌شود. به بیان دیگر، سامانه خودکار امکان بررسی همزمان هزاران آیه و حدیث را فراهم می‌کند، کاری که حتی با وجود گروهی از پژوهشگران انسانی به سادگی ممکن نیست.

پوشش جامع و عدم خستگی: انسان ممکن است در کار طولانی دچار خستگی یا خطای سهوی شود و برخی روابط یا آیات را از قلم بیندازد. در مقابل، یک برنامه کامپیوتری تمام داده‌های ورودی را به دقت یکسان پردازش می‌کند و چیزی را فراموش نمی‌کند. بنابراین سامانه خودکار

تضمین می‌کند که حتی آیات کمتر مشهور یا موضوعات فرعی‌تر نیز در دسته‌بندی لحاظ شوند. این نکته به ویژه زمانی مهم است که می‌خواهیم یک فهرست جامع القرآن تولید کنیم که همه آیات مرتبط با یک موضوع (حتی موضوعات نادر) را در بر گیرد. برای نمونه، ممکن است موضوعی نظیر «کشتی‌رانی» فقط در دو سه آیه قرآن آمده باشد و یک مفسر انسانی در میان انبوه موضوعات به آن توجه نکند، اما سامانه خودکار همان چند آیه را نیز زیر خوشه مناسب قرار می‌دهد.

**اکتشاف روابط پنهان:** همان‌طور که دیدیم، الگوریتم‌های یادگیری ماشین قادر به کشف الگوهای جدید هستند که لزوماً در روش سنتی آشکار نبودند. مثلاً داده‌کاوی ممکن است نشان دهد آیه‌ای در سوره غافل به آیه‌ای در سوره یاسین از نظر مفهوم نزدیک است، هرچند در ظاهر ارتباطی نداشتند. این روابط پنهان می‌توانند جرقه‌ای برای تحقیقات تفسیری جدید باشند؛ چنان‌که صالحی‌شهرودی و مینایی برخی ارتباطات آیات را از طریق داده‌کاوی تفاسیر کشف کردند که پیش‌تر تصریح نشده بود. بنابراین سامانه هوشمند علاوه بر سازمان‌دهی اطلاعات، بینش‌های تازه به پژوهشگر می‌دهد و نقش فراتر از یک ابزار صرف پیدا می‌کند.

استانداردسازی و انسجام: هر پژوهشگر ممکن است طبقه‌بندی خاص خود را از موضوعات قرآن ارائه دهد و این تنوع گاه‌گاهاً کاربران را سردرگم می‌کند. اما اگر یک سامانه خودکار توسعه یابد که با پذیرش بازخورد متخصصان به تدریج بهبود یابد، می‌تواند به مرور به یک استاندارد مشترک بدل شود. برای مثال، فرض کنید جامعه علمی توافق کند که نسخه ۵/۰ یک نرم‌افزار طبقه‌بندی قرآنی به اندازه کافی قابل اعتماد است؛ در این صورت همه از همان استفاده می‌کنند و نتایج تحقیقات قابل مقایسه و همگرا می‌شود. چنین استانداردسازی در سایه روش سنتی دشوار است، چون سبک کار و معیارهای هر مفسر با دیگری تفاوت دارد.

**امکان به‌روزرسانی و تکامل:** یک سامانه نرم‌افزاری پویا است و می‌تواند مرتباً به‌روزرسانی و بهتر شود. با افزودن داده‌های جدید (مثلاً تفسیرهای تازه، مقالات علمی جدید) یا با ارتقای الگوریتم‌های آن، سیستم می‌تواند یادگیری مستمر داشته باشد و خروجی دقیق‌تری بدهد. در مقابل، کتاب‌های چاپی معجم موضوعی ثابت هستند و پس از چاپ تغییر نمی‌کنند تا چاپ بعدی. بنابراین سامانه دیجیتال این قابلیت را دارد که آخرین دانش و تحلیل‌ها را منعکس کند. برای مثال، اگر فردا فهم جدیدی درباره یک مفهوم قرآنی حاصل شود، می‌توان آن را به سیستم داد تا آیات مرتبط را مطابق آن بازتنظیم کند.

## ۲/۲. محدودیت‌ها و چالش‌ها

نیاز به داده آموزشی باکیفیت: سیستم‌های یادگیری نظارت‌شده برای ارائه خروجی درست، نیازمند داده‌های برجسب‌دار دقیق هستند. تهیه چنین داده‌ای در حوزه قرآن مستلزم کار کارشناسی عمیق است. اگر برجسب‌های آموزشی نادرست یا ناقص باشند، مدل همان خطاها را بسط خواهد داد (اصل زباله به زباله در علوم داده). لذا یکی از تنگناهای توسعه این سامانه‌ها تامین دیتاست‌های استاندارد و توافق‌شده است. پروژه‌هایی مثل ایجاد بانک موضوعی قرآن با نظارت جمعی از مفسران می‌تواند به این نیاز پاسخ دهد، اما انجام آن ساده نیست. در حال حاضر، هر پژوهشگر معمولاً برجسب‌های دلخواه خود را استفاده می‌کند (مثلاً ۱۲ موضوع در یک تحقیق، ۱۵ موضوع در دیگری)، که مقایسه نتایج بین کارها را دشوار می‌کند.

چالش ارزیابی کیفی: حتی اگر مدل در اعداد و ارقام (Recall, Precision) عملکرد خوبی داشته باشد، تضمینی نیست که خروجی آن از دید یک عالم دینی از نظر معنوی درست باشد. ممکن است یک آیه به لحاظ کلمات شباهت به موضوعی داشته باشد ولی در باطن معنا ارتباط نداشته باشد. انسان مفسر با در نظر گرفتن شان نزول، سیاق، روایات مربوطه و... درک عمیق‌تری دارد که ماشین فاقد آن است. برای نمونه، آیه «لا تقربوا الصلاة وانتم سكارى...» (نساء: ۴۳) درباره ممنوعیت نماز در حالت مستی است. یک سیستم ممکن است آن را فقط ذیل موضوع «نماز» یا «احکام عبادت» قرار دهد، اما یک مفسر می‌داند این آیه تاریخی از سیر تحریم خمر است و به «فقه الخمر» نیز مربوط است. بنابراین، ارزیابی کیفیت خروجی سیستم صرفاً با اعداد کافی نیست و نیاز به بازبینی انسانی دارد تا اطمینان حاصل شود معنای عمیق آیات در دسته‌بندی لحاظ شده است.

مسئله تفسیرپذیری مدل: مدل‌های یادگیری عمیق مثل ترنسفورمرها اغلب جعبه‌سیاه هستند؛ یعنی مشخص نیست دقیقاً چرا یک تصمیم (برجسب‌دهی) گرفته‌اند. در علوم دینی، شفاف‌بودن دلیل بسیار مهم است. اگر سیستم بگوید این آیه موضوعش X است، عالم دینی می‌پرسد «چرا؟» و انتظار دارد شواهدی (مثلاً کلمات کلیدی) ببیند. الگوریتم‌های کلاسیک‌تر تا حدی این امکان را می‌دهند (مثلاً وزن IDF-TF کلمات مهم)، ولی مدل‌های عمیق فقط اعداد بی‌معنا برای انسان تولید می‌کنند. این مسئله می‌تواند باعث عدم اعتماد استفاده‌کنندگان به سامانه شود. برای حل آن،

حوزه پژوهشی هوش مصنوعی تفسیرپذیر (XAI) پیشنهاداتی دارد، مثل نمایش نقشه توجه در مدل‌های ترنسفورمر که نشان می‌دهد مدل در هر آیه به کدام کلمات بیشترین توجه را کرده است. به کارگیری این ابزارها در سامانه قرآنی می‌تواند شفافیت را بالا برده و اعتمادسازی کند.

**محدودیت فهم زمینه‌های فرامنتی:** رایانه صرفاً بر اساس متنی که می‌بیند تصمیم می‌گیرد. اما فهم قرآن علاوه بر متن، محتاج دانش‌های پیرامونی است: اسباب نزول، ناسخ و منسوخ، مکی یا مدنی بودن آیه، شأن مخاطب و... فعلاً سیستم‌های خودکار در بهترین حالت متن خالص را تحلیل می‌کنند. ممکن است در آینده با غنی‌سازی ورودی‌ها این کاستی کاهش یابد (مثلاً افزودن برجسب مکی/مدنی به هر آیه در ورودی مدل، چنان‌که رولیاواتی این کار را کرد). ولی باز هم ظرائف تفسیری زیادی وجود دارد که از حوصله ماشین خارج است. لذا خروجی سیستم خودکار هرگز بی‌نیاز از تفسیر انسانی نخواهد بود. بلکه باید به دیده ابزار کمکی نگریسته شود نه جایگزین مفسر. حتی می‌توان گفت این سامانه‌ها برای کاربر غیرمتخصص نهایی طراحی نشده‌اند، بلکه برای کمک به کارشناسان است.

**احتمال سوگیری و نقص اخلاقی:** هر سیستم یادگیری ماشینی ممکن است سوگیری‌های موجود در داده را بازتولید کند. اگر برجسب‌گذاری اولیه بر اساس دیدگاه مذهب خاصی باشد، مدل همان را تکرار خواهد کرد. این در حوزه قرآن ممکن است محل چالش شود؛ زیرا برداشت‌های مذهبی یا فقهی متفاوتی وجود دارد. البته با طراحی مناسب (استفاده از برجسب‌های خنثی و مورد توافق) می‌توان این را تخفیف داد. ولی به هر حال باید توجه داشت که محصول نهایی را نباید وحی منزل انگاشت؛ بلکه آن نیز ساخته فهم بشری (به صورت داده آموزشی) است. از این رو، همیشه باید مجالی برای بازخورد مفسران و اصلاحات دستی در سیستم پیش‌بینی شود.

با توجه به موارد فوق، می‌توان نتیجه گرفت که سامانه‌های خودکار طبقه‌بندی موضوعی آیات ابزاری بسیار سودمند و در عین حال نیازمند نظارت انسانی هستند. این سامانه‌ها می‌توانند کار ده‌ها ساله یک پژوهشگر را در چند دقیقه انجام دهند و روابط و الگوهای جدیدی را پیش روی او بگذارند؛ اما خروجی آن‌ها باید توسط پژوهشگر تفسیر و ارزیابی شود. در واقع، بهترین حالت استفاده ترکیبی است: رایانه کار مکانیکی و پردازشی را انجام دهد و انسان کار تحلیلی و اجتهادی را. این دو در کنار هم می‌توانند فهم عمیق‌تری از قرآن کریم به ارمغان آورند که نه صرفاً با اتکا به روش سنتی میسر بود و نه با اعتماد کامل به ماشین بدون دخالت انسان.

## نتیجه‌گیری

پژوهش حاضر به بررسی جامع امکان‌سنجی و چگونگی ایجاد یک سامانه خودکار طبقه‌بندی موضوعی آیات قرآن پرداخت. در ابتدا ضرورت چنین سامانه‌ای با توجه به گستردگی مفاهیم قرآنی و چندموضوعی بودن آیات تبیین شد. سپس مفاهیم بنیادین از جمله یادگیری ماشین، یادگیری عمیق، طبقه‌بندی چندبرچسبی و ترنسفورمر شرح داده شد تا چارچوب نظری بحث روشن گردد. در ادامه، مروری بر روش‌های مختلف انجام گرفت: روش‌های کلاسیک یادگیری ماشین مانند بیز ساده و SVM که برای دسته‌بندی آیات به کار رفته و نتایج اولیه‌ای فراهم کردند؛ روش‌های یادگیری عمیق و مدل‌های ترنسفورمر که دستاوردهای نوینی چون دقت بالای ۹۰٪ در شناسایی موضوع آیات را به ارمغان آوردند؛ توجه ویژه به پدیده چندبرچسبی بودن آیات و نحوه مدیریت آن در مدل‌ها؛ و نیز رویکردهای بدون نظارت مانند خوشه‌بندی و LDA که به کشف الگوهای پنهان در متن قرآن کمک می‌کنند. همچنین با معرفی چند مطالعه موردی داخلی و خارجی، تصویر روشنی از تلاش‌های عملی صورت‌گرفته ارائه شد. این مطالعات نشان دادند که هر یک بخشی از چالش را پوشش داده‌اند، از تهیه پیکره‌های برچسب‌خورده گرفته تا ساخت مدل‌های پیچیده عمیق و توسعه ابزارهای کاربردی برای جستجوی معنایی آیات.

در پاسخ به سؤال تحقیق - مبنی بر امکان ساخت سامانه خودکار و این‌که چنین سامانه‌ای چه ویژگی‌ها و مزایا و محدودیت‌هایی دارد - می‌توان این‌گونه جمع‌بندی کرد:

سامانه‌های خودکار یادگیری محور قادرند با سرعت و دقت بالا، آیات قرآن را بر اساس موضوعات از پیش تعریف شده یا کشف‌شده، طبقه‌بندی کنند. این سامانه‌ها به ویژه با بهره‌گیری از یادگیری عمیق توانسته‌اند به عملکردی در حد و اندازه متخصص انسانی نزدیک شوند. به عنوان نمونه، شبکه‌های عصبی عمیق توانسته‌اند هم‌زمان جنبه‌های اعتقادی، تاریخی و اخلاقی یک آیه را تشخیص داده و برچسب‌دهی کنند - کاری که برای انسان بسیار زمان‌بر است. به علاوه، سیستم‌های خودکار می‌توانند الگوهای پنهان را که در روش سنتی به‌سادگی قابل ردیابی نیستند آشکار کنند (مثلاً گروه‌بندی غیرمنتظره آیات یا ارتباط مفهومی میان آن‌ها). از منظر کارایی، سامانه خودکار می‌تواند به عنوان یک دستیار هوشمند قرآنی به پژوهشگر در یافتن سریع آیات مرتبط با یک موضوع، تهیه

فهرست‌های موضوعی جامع، و حتی پیشنهاد تفاسیر جدید کمک کند. این همان چیزی است که هدف نهایی بسیاری از پروژه‌های مطرح شده بوده است: نزدیک‌تر کردن فهم متن مقدس به توان تحلیل بشری از طریق قدرت محاسباتی.

با این حال، این سامانه‌ها جایگزین کامل تفسیر انسانی نیستند و نخواهند بود. همان‌طور که در بخش ارزیابی بیان شد، خروجی سیستم نیازمند صحت‌سنجی و تفسیر نهایی توسط کارشناسان است تا از لحاظ انسجام معنایی و عدم تعارض با معارف مسلم قرآن و حدیث اطمینان حاصل شود. مدل‌های یادگیری ماشین بر پایه داده‌های آموزشی کار می‌کنند و هر قدر این داده‌ها دقیق‌تر و غنی‌تر باشد، نتیجه بهتر خواهد بود. بنابراین تعامل مستمر میان مفسران قرآن برای تهیه و تصحیح داده‌ها و میان متخصصان AI برای طراحی مدل‌ها، شرط موفقیت چنین سامانه‌ای است. خوشبختانه تحقیقات بین‌رشته‌ای در این زمینه رو به گسترش است و می‌توان امید داشت در آینده‌ای نزدیک سامانه‌هایی به مراتب هوشمندتر و قابل اعتمادتر در اختیار قرآن‌پژوهان قرار گیرد. شاید بتوان چشم‌اندازی را متصور شد که در آن یک نرم‌افزار جامع قرآنی با ورود هر آیه یا هر پرسش مفهومی، تمام آیات مرتبط را (براساس فهم ترکیبی از متن و داده‌های تفسیری) استخراج کند و حتی توضیحاتی اولیه در مورد ارتباط آن‌ها ارائه دهد. چنین دستاوردی انقلابی در سهولت دسترسی به معارف قرآن و پل زدن بین تخصص علوم رایانه و علوم اسلامی خواهد بود.

در پایان تأکید می‌کنیم که تحقیق حاضر اگرچه به پیشرفت‌های قابل توجهی اشاره کرد، اما چالش‌های پژوهشی فراوانی کماکان باقی است؛ از جمله استانداردسازی مجموعه موضوعات قرآنی، بهبود تفسیرپذیری خروجی مدل‌ها، ترکیب اطلاعات زبان‌شناختی و تاریخی با مدل‌های آماری، و توسعه پیکره‌های چندزبانه قرآن. تداوم همکاری میان رشته‌ای و بهره‌گیری از آخرین فناوری‌های یادگیری ماشین می‌تواند راهگشای این مسائل باشد. سامانه خودکار طبقه‌بندی موضوعی آیات قرآن نه تنها دست‌یافتنی است بلکه با شتاب فعلی علوم داده در حال تبدیل شدن به واقعیتی اجتناب‌ناپذیر در آینده پژوهش‌های قرآنی است. این واقعیت، نویدبخش عصر جدیدی است که در آن ابزارهای هوشمند دیجیتال در خدمت فهم عمیق‌تر و همه‌جانبه‌تر کلام وحی قرار می‌گیرند.

## فهرست منابع

- ا.م.ک. ایزاتی، م.س. موبارک، ن.س. هدا، ک. آدیویجا یا (۲۰۱۸). «طبقه‌بندی چندبرچسبی موضوعات آیات قرآن در ترجمه انگلیسی با استفاده از بیز ساده درخت افزایشی». کنفرانس IEEE، ICoICT، صفحات ۳۶۰-۳۶۵.
- اشیش واسوانی و دیگران (۲۰۱۷). «توجه همه چیز است». کنفرانس NeurIPS، دوره ۳۰، صفحات ۵۹۹۸-۶۰۰۸.
- ایان گودفلو، یوشوا بنگیو، آرون گرویل (۲۰۱۶). یادگیری عمیق. MIT Press.
- تام میچل (۱۹۹۷). یادگیری ماشین. McGraw-Hill.
- تورستن یواخیمز (۱۹۹۸). «دسته‌بندی متنی با ماشین بردار پشتیبان». کنفرانس ECML، صفحات ۱۳۷-۱۴۲. Springer
- جسی رید، لورنزو مارتینو، خسوس د. آل‌موس، دیوید لئونگو (۲۰۱۵). «پیش‌بینی مقیاس‌پذیر برچسب‌های چندگانه». Pattern Recognition، دوره ۴۸، شماره ۶، صفحات ۲۰۹۶-۲۱۰۹.
- جیاوی هان، میشلین کامبر، جیان پی. (۲۰۱۱). داده‌کاوی: مفاهیم و تکنیک‌ها. Morgan Kaufmann.
- جیکوب دولین، مینگ-وی چانگ، کنتون لی، کریستینا توتانوا (۲۰۱۹). «BERT: پیش‌آموزش ترنسفورمرهای دوطرفه عمیق برای درک زبان». کنفرانس NAACL-HLT، صفحات ۴۱۷۱-۴۱۸۶.
- چاروی آگراوال، چنگ‌شیان ژای (۲۰۱۲). داده‌کاوی متون. Springer.
- خلیل‌الله صبری (۱۳۸۹). طبقات آیات: طبقه‌بندی آیات قرآن کریم بر اساس موضوع در ۲۶ فصل (با ترجمه آیات به فارسی). تهران: انتشارات امیرکبیر.
- دوی رولیاواتی، اندری اس. روزاس، خالد، ام. راتودی (۲۰۱۸). «رویکرد متن‌کاوی برای مدل‌سازی موضوعی پیکره قرآن». کنفرانس ACM، IMAN، صفحات ۸۳-۹۰.

ر.ا. پان، م.س. موبارک، ن.س. هدا، الف. آدیویجایا (۲۰۱۸). «طبقه‌بندی چندبرچسبی موضوعات آیات قرآن در ترجمه انگلیسی با استفاده از بیز چندجمله‌ای». کنفرانس IEEE, ICoICT. صفحات ۳۵۹-۳۵۴.

ساتریو آدی پرابوو، آدیویجایا، محمد شهرال موبارک، سعید ال فارابی، محمد زیدنی ناف، محمد یوسلان ابو بکر (۲۰۱۹). «پیاده‌سازی ماشین بردار پشتیبان در طبقه‌بندی چندبرچسبی آیات قرآن». مجله علوم کامپیوتر، دوره ۱۵، شماره ۱۲، صفحات ۱۷۵۲-۱۷۵۸.

سمیر م. الرحیلی، اریک اتول (۲۰۱۴). «آنتولوژی‌های محاسباتی برای برچسب‌گذاری معنایی قرآن: مرور روش‌های پیشین». مجموعه مقالات کنفرانس LREC 2014.

سیوپیان جایا پوترا، تدی مانتورو، محمد نور گونوان (۲۰۱۷). «متن‌کاوی ترجمه اندونزیایی قرآن: مرور نظام‌مند». کنفرانس IEEE, ICCED، صفحات ۱-۵.

عبدالکریم م. العشقار (۲۰۲۳). «طبقه‌بندی آیات قرآن با استفاده از یادگیری عمیق». مجله بین‌المللی رایانه و سیستم‌های دیجیتال.

علی میراعرب، مرتضی محمدی استانی، فائزه سادات طباطبایی امیری، سمیه دهقانی (۱۴۰۲). «روش‌های خودکارسازی یادگیری هستی‌نگاشت‌ها در حوزه مفاهیم قرآنی». تحقیقات

کتابداری و اطلاع‌رسانی دانشگاه فردوسی مشهد، دوره ۵۴، شماره ۲، صفحات ۲۲۷-۲۵۴.  
علیور. یائوری، ر. عبدالکدیر، عززین عزمان، م. ا. عزمی مراد (۲۰۰۹). «استخراج آیات قرآن بر مبنای مفاهیم با استفاده از اونتولوژی OWL-DL». کنفرانس بین‌المللی فناوری‌های نوظهور، IEEE، صفحات ۲۹-۳۴.

کویین پی. مورفی (۲۰۱۲). یادگیری ماشین: چشم‌انداز احتمالاتی. MIT Press.  
گریگوریوس تسوماکاس، یوآنیس کاتاس (۲۰۰۷). «طبقه‌بندی چندبرچسبی: مروری کلی». مجله بین‌المللی انبارداری و داده‌کاوی، دوره ۳، شماره ۳، صفحات ۱-۱۳.

مایسوم ح. پنجو (۲۰۱۴). استخراج آماری و نمایش تصویری موضوعات در پیکره قرآن. گزارش فنی، دانشگاه واترلو.

محمد الخورات (۲۰۱۵). «استخراج موضوعات از قرآن کریم با استفاده از مدل‌های مولدی». IJACSA، دوره ۶، شماره ۱۲، صفحات ۲۸۸-۲۹۴.

محمد بزرگ‌قمی‌زاده، بهروز مینایی (۱۳۹۸). «کشف ارتباطات مفهومی آیات قرآن کریم در بستر تفاسیر قرآن با استفاده از تکنیک‌های داده‌کاوی». دوفصلنامه قرآن‌شناخت، سال ۳، شماره ۲، صفحات ۸۹-۱۰۲.

محمد شعیب، م. ندیم یاسین، الله ک. حکمت، م. عمران سعید، ملک سکندر ح. خیال (۲۰۰۹). «مدل WordNet رابطه‌ای برای جستجوی معنایی در قرآن کریم». کنفرانس بین‌المللی فناوری‌های نوظهور، IEEE، صفحات ۲۹-۳۴.

محمد عکور، ایازات المسعدی، ایاد العزام (۲۰۱۴). «MQVC: اندازه‌گیری شباهت آیات قرآن و کلاس‌بندی سوره‌ها با استفاده از n-گرم». WSEAS Transactions on Computers، دوره ۱۳، صفحات ۴۸۵-۴۹۱.

محمدحسین صالحی‌شهرودی، بهروز مینایی، امیررضا اشرفی (۱۳۹۲). «متن‌کاوی موضوعی رایانه‌ای قرآن کریم برای کشف ارتباطات معنایی میان آیات بر مبنای تفسیر المیزان». فصلنامه علمی پژوهشی قرآن‌شناخت، سال اول، شماره ۲، صفحات ۱۳۳-۱۵۰.

مصطفی کرمی، علیرضا طالب‌پور، فرزانه تاج‌آبادی، زینب حاجی‌محمدی (۱۴۰۱). «تشخیص احساسات از متن قرآن با استفاده از شبکه یادگیری عمیق RoBERTa پیشرفته». پژوهش‌های میان‌رشته‌ای قرآنیک، دوره ۲، شماره ۱، صفحات ۱-۱۴.

معظم احمد صدیقی، سید محمد فراز، سهیل عبدالستار (۲۰۱۳). «کشف ساختار موضوعی قرآن با استفاده از مدل‌های احتمالاتی موضوعی». مجموعه مقالات کنفرانس بین‌المللی دانشگاه طیب در پیشرفت‌های فناوری اطلاعات برای قرآن و علوم آن (ICCIT)، صفحه ۲۳۴-۲۳۹، IEEE.

مهدی رحیمی‌بافقی، معصومه پورباقری (۱۳۹۳). خوشه‌بندی آیات قرآن. پایان‌نامه کارشناسی، دانشگاه علم و صنعت ایران.

مین-لینگ ژانگ، ژی-هوا ژو (۲۰۱۴). «مروری بر الگوریتم‌های یادگیری چندبرچسبی». IEEE TKDE، دوره ۲۶، شماره ۸، صفحات ۱۸۱۹-۱۸۳۷.

نورهن ح. عباس (۲۰۰۹). ابزار و وب‌سایت «جست‌وجو برای یک مفهوم» قرآن. رساله دکتری، دانشگاه لیدز.

یان لوکان، یوشوا بنگیو، جفری هینتون (۲۰۱۵). «یادگیری عمیق». Nature، دوره ۵۲۱، صفحات ۴۴۴-۴۳۶.

Abdelkareem M. Alashqar (2023). “A Classification of Quran Verses Using Deep Learning.” International Journal of Computing and Digital Systems (Early Access).

Ali Mirarab, Morteza Mohammadi Estani, Fa’ezeh S. Tabatabaei Amiri, and Somayeh Dehghani (2023). “Automated Ontology Learning Methods in Quranic Concept Domains: A Systematic Review.” University of Ferdowsi Mashhad Journal of Library and Information Science, 54(2): 227–254.

Aliyu R. Yauri, R. Abdul Kadir, Azreen Azman, and M. A. Azmi Murad (2009). “Quranic Verse Extraction Based on Concepts Using an OWL-DL Ontology.” In 2009 International Conference on Emerging Technologies, 29–34. IEEE.

Ashish Vaswani et al. (2017). “Attention Is All You Need.” In Advances in Neural Information Processing Systems, 30: 5998–6008.

Charu C. Aggarwal and ChengXiang Zhai, eds. (2012). Mining Text Data. Springer.

Dwi Rolliawati, Indri S. Rozas, Khalid, and M. Ratodi (2018). “Text Mining Approach for Topic Modeling of the Corpus Al-Qur’an in Indonesian Translation.” In Proceedings of the 2018 International Conference on Islamic Applications in Computer Science and Technologies (IMAN), 83–90. ACM.

Grigorios Tsoumakas and Ioannis Katakis (2007). “Multi-Label Classification: An Overview.” International Journal of Data Warehousing and Mining, 3(3): 1–13.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016). Deep Learning. MIT Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” In Proceedings of NAACL-HLT, 4171–4186.

- Jesse Read, Lorenzo Martino, Jesús D. Olmos, and David Luengo (2015). "Scalable Multi-Output Label Prediction: From Classifier Chains to Classifier Trellises." *Pattern Recognition*, 48(6): 2096–2109.
- Jiawei Han, Micheline Kamber, and Jian Pei (2011). *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann (Elsevier).
- Kevin P. Murphy (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Khalilullah Sabri (2010). *Tabaqat al-Ayat: Thematic Classification of the Verses of the Holy Quran in 26 Chapters (with Persian Translation of the Verses)*. Tehran: Amir Kabir Publications.
- M. K. Izzaty, M. S. Mubarak, N. S. Huda, and K. Adiwijaya (2018). "A Multi-Label Classification of Topics of Quranic Verses in English Translation Using Tree-Augmented Naïve Bayes." In *Proceedings of the 2018 International Conference on Information and Communication Technology (ICoICT)*, 360–365. IEEE.
- Maysum H. Panju (2014). *Statistical Extraction and Visualization of Topics in the Quran Corpus*. Technical Report, University of Waterloo.
- Mehdi Rahimibafghi and Masoumeh Pourbagheri (2014). *Clustering of Quranic Verses*. B.Sc. thesis, Computer Engineering Department, Iran University of Science and Technology.
- Min-Ling Zhang and Zhi-Hua Zhou (2014). "A Review on Multi-Label Learning Algorithms." *IEEE Transactions on Knowledge and Data Engineering*, 26(8): 1819–1837.
- Mohammad Alhawarat (2015). "Extracting Topics from the Holy Quran Using Generative Models." *International Journal of Advanced Computer Science and Applications (IJACSA)*, 6(12): 288–294.
- Mohammad Bozorg-Ghamizadeh and Behrooz Minaei (2019). "Discovering Conceptual Relations of Quranic Verses within Tafsirs Using Data Mining Techniques." *Quran-Shenakht Quarterly*, 3(2): 89–102.
- Mohammad Hossein Salehi-Shahroudi, Behrooz Minaei, and Amirreza Ashrafi (2013). "Computer-Based Thematic Text Mining of the Holy Quran to Discover Semantic Connections among Verses, Based on Tafsir

- al-Mizan.” Quran-Shenakht Quarterly, vol. 1, no. 2 (Autumn–Winter): 133–150.
- Mohammed Akour, Izzat Alsmadi, and Iyad Alazzam (2014). “MQVC: Measuring Quranic Verses Similarity and Sura Classification Using N-Gram.” WSEAS Transactions on Computers, 13: 485–491.
- Mostafa Karami, Alireza Talebpour, Farzaneh Tajabadi, and Zeinab Hajimohammadi (2023). “Sentiment Analysis of Quranic Text Using Advanced RoBERTa Deep Learning.” Journal of Interdisciplinary Quranic Studies, 2(1): 1–14.
- Muazzam Ahmed Siddiqui, Syed Muhammad Faraz, and Sohail Abdul Sattar (2013). “Discovering the Thematic Structure of the Quran Using Probabilistic Topic Models.” In Proceedings of the Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences (ICCIT), 234–239. IEEE.
- Muhammad Huzaifa Bashir, Aqil M. Azmi, Haq Nawaz, Wajdi Zaghouni, Mona Diab, Ala Al-Fuqaha, and Junaid Qadir (2023). “Arabic Natural Language Processing for Qur’anic Research: A Systematic Review.” Artificial Intelligence Review, 56: 1261–1302.
- Muhammad Shoaib, M. Nadeem Yasin, Ullah K. Hikmat, M. Imran Saeed, and Malik Sikandar H. Khiyal (2009). “Relational WordNet Model for Semantic Search in the Holy Quran.” In 2009 International Conference on Emerging Technologies, 29–34. IEEE.
- Noorhan H. Abbas (2009). Qur’an ‘Search for a Concept’ Tool and Website. PhD diss., School of Computing, University of Leeds, UK.
- R. A. Pane, M. S. Mubarak, N. S. Huda, and A. Adiwijaya (2018). “A Multi-Label Classification of Topics of Quranic Verses in English Translation Using Multinomial Naïve Bayes.” In Proceedings of the 2018 International Conference on Information and Communication Technology (ICoICT), 354–359. IEEE.
- Sameer M. Alrehaili and Eric Atwell (2014). “Computational Ontologies for Semantic Tagging of the Quran: A Survey of Past Approaches.” In Proceedings of LREC 2014 (9th Language Resources and Evaluation Conference). European Language Resources Association.

- Satrio Adi Prabowo, Adiwijaya, Mohamad Syahrul Mubarak, Said Al Faraby, Muhammad Zidny Naf, and Muhammad Yuslan Abu Bakar (2019). “An Implementation of Support Vector Machine on the Multi-Label Classification of English-Translated Quranic Verses.” *Journal of Computer Science*, 15(12): 1752–1758.
- Syopyan Jaya Putra, Teddy Mantoro, and Muhamad Nur Gunawan (2017). “Text Mining for the Indonesian Translation of the Quran: A Systematic Review.” In 2017 International Conference on Computing, Engineering, and Design (ICCED), 1–5. IEEE.
- Thorsten Joachims (1998). “Text Categorization with Support Vector Machines: Learning with Many Relevant Features.” In *Proceedings of the 10th European Conference on Machine Learning (ECML)*, LNCS 1398, 137–142. Springer.
- Tom M. Mitchell (1997). *Machine Learning*. McGraw-Hill.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton (2015). “Deep Learning.” *Nature*, 521: 436–444

## References

- Amini, S. J., Mohammadian, M., & Salim, H. (2025/1404 SH). Strategic model for social media management in the field of clarification jihad. Religion and Communications (Imam Sadiq University / Nameh Sadiq). ([\[https://rc.isu.ac.ir/article\\_77669.html\]](https://rc.isu.ac.ir/article_77669.html)([https://rc.isu.ac.ir/article\\_77669.html](https://rc.isu.ac.ir/article_77669.html)))
- Beheshtinejad, S. M., & Reza-zadeh, R. (2025/1404 SH). The role of clarification jihad in cyberspace for crisis prevention. Social Crisis Management, 17(1), 63–95. ([\[https://cmj.ihu.ac.ir/article\\_209951.html\]](https://cmj.ihu.ac.ir/article_209951.html)([https://cmj.ihu.ac.ir/article\\_209951.html](https://cmj.ihu.ac.ir/article_209951.html)))
- Araqi, A., Bidgoli, M., & Rajabi Dehbarzouei, A. (2022/1401 SH). Analysis of the goals of enemy cognitive warfare and strategies for resilience with emphasis on Quranic teachings. Scientific Quarterly of Sacred Defense Studies, 8(4), 143–162. ([\[https://dor.isc.ac/dor/20.1001.1.37883674.1401.8.4.7.6\]](https://dor.isc.ac/dor/20.1001.1.37883674.1401.8.4.7.6)(<https://dor.isc.ac/dor/20.1001.1.37883674.1401.8.4.7.6>))
- Azizi, M., & Pirani, Sh. (2022/1401 SH). Software applications of cyberspace in clarification jihad. Scientific Quarterly of Islamic Revolution Research, 11(4). ([\[https://www.roir.ir/article\\_168676.html\]](https://www.roir.ir/article_168676.html)([https://www.roir.ir/article\\_168676.html](https://www.roir.ir/article_168676.html)))
- Tabatabaei, S. F., Mohammad-Hosseini-zadeh, A., & Akbarpour-Farsangi, S. (2023/1402 SH). Clarification jihad from the perspective of Ayatollah Khamenei and its jurisprudential foundations. Scientific Quarterly of Islamic Awakening Studies, 12(3), 143–168. ([\[https://www.iabaj.ir/article\\_180733.html\]](https://www.iabaj.ir/article_180733.html)([https://www.iabaj.ir/article\\_180733.html](https://www.iabaj.ir/article_180733.html)))
- Ghaffari, S. A., Mahdian, H., Seydi, S., & Tarkashvand, J. (2024/1403 SH). A systematic model of clarification jihad in the thought of the Supreme Leader. Military Management Quarterly, 24(93), 1–27. ([\[https://jmm.iranjournals.ir/article\\_713680.html\]](https://jmm.iranjournals.ir/article_713680.html)([https://jmm.iranjournals.ir/article\\_713680.html](https://jmm.iranjournals.ir/article_713680.html)))

- Mahjoob, H., & Shakouri, S. (2022/1401 SH). Modern cognitive warfare: from battlefield understanding to the cognitive warfare domain. *Human Resources Studies*, 12(2), 156–180. ([<https://doi.org/10.22034/jhrs.2022.158895>](<https://doi.org/10.22034/jhrs.2022.158895>))
- Mahmoudi, M., & Sadeghi, S. (2022/1401 SH). Clarification jihad in cyberspace from the perspective of the Supreme Leader. In *Third National Conference on Cyber Defense*, Maragheh. ([<https://civilica.com/doc/1493906>](<https://civilica.com/doc/1493906>))
- Malekzadeh, M. (2024/1403 SH). Functions and consequences of clarification jihad in the internationalization of the resistance model from Ayatollah Khamenei's perspective. *Military Management Quarterly*, 24(10), 157–176. ([[https://mag.rcipt.ir/article\\_181105.html](https://mag.rcipt.ir/article_181105.html)]([https://mag.rcipt.ir/article\\_181105.html](https://mag.rcipt.ir/article_181105.html)))
- Nikonehad, A., & Biouk, M. (2024/1403 SH). Designing and visualizing an optimal clarification jihad in confrontation with cognitive warfare based on the Supreme Leader's thought. *Islamic Revolution Futures Studies*, 5(1), 11–36. ([<https://dor.isc.ac/dor/20.1001.1.27173674.1403.5.1.1.3>](<https://dor.isc.ac/dor/20.1001.1.27173674.1403.5.1.1.3>))
- Rezaei, D., & Ebrahimi, H. (2024/1403 SH). Identifying factors, challenges, and strategies to counter cognitive warfare with focus on Generation Z in Iran. *Air Defense Management Quarterly*, 3(2), 27–42. ([[https://jadm.khadu.ir/article\\_125.html?lang=fa](https://jadm.khadu.ir/article_125.html?lang=fa)]([https://jadm.khadu.ir/article\\_125.html?lang=fa](https://jadm.khadu.ir/article_125.html?lang=fa)))
- Extracts from the Supreme Leader's statements on clarification jihad (2016–2025/1395–1404 SH). Official Website of the Office for Preservation and Publication of Ayatollah Khamenei's Works. ([<https://farsi.khamenei.ir/newspart-index-old?tid=26000#128187>](<https://farsi.khamenei.ir/newspart-index-old?tid=26000#128187>))

- Al Momani, I. (2025). Ethical challenges for using artificial intelligence in understanding Islamic jurisprudence. *Salud, Ciencia y Tecnología-Serie de Conferencias*, (4), 1519. ([\[http://dx.doi.org/10.56294/sctconf20251519\]](http://dx.doi.org/10.56294/sctconf20251519)(<http://dx.doi.org/10.56294/sctconf20251519>))
- Al-Shammari, E. T. (2025, April). Culturally and computationally adaptive social robots for Arabophone Islamic societies. In 2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD) (pp. 1–6). IEEE. ([\[https://doi.org/10.1109/ITIKD63574.2025.11005245\]](https://doi.org/10.1109/ITIKD63574.2025.11005245)(<https://doi.org/10.1109/ITIKD63574.2025.11005245>))
- Arrieta, A. G. J., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. ([\[https://doi.org/10.1016/j.inffus.2019.12.012\]](https://doi.org/10.1016/j.inffus.2019.12.012)(<https://doi.org/10.1016/j.inffus.2019.12.012>))
- Babiker, A., et al. (2025). Attitude towards AI: Potential influence of conspiracy belief, XAI experience and locus of control. *International Journal of Human–Computer Interaction*, 41(13), 7939–7951. ([\[https://doi.org/10.1080/10447318.2024.2401249\]](https://doi.org/10.1080/10447318.2024.2401249)(<https://doi.org/10.1080/10447318.2024.2401249>))
- Bang, T. C. (2024). Ensuring credibility and trustworthiness in qualitative inquiries. In *Applied Linguistics and Language Education Research Methods: Fundamentals and Innovations* (pp. 70–85). IGI Global. ([\[https://doi.org/10.4018/979-8-3693-2603-9.ch006\]](https://doi.org/10.4018/979-8-3693-2603-9.ch006)(<https://doi.org/10.4018/979-8-3693-2603-9.ch006>))
- Belghachi, M. (2023). A review on explainable artificial intelligence methods, applications, and challenges. *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, 11(4), 1007–1024. ([\[http://dx.doi.org/10.52549/ijeel.v11i4.5151\]](http://dx.doi.org/10.52549/ijeel.v11i4.5151)(<http://dx.doi.org/10.52549/ijeel.v11i4.5151>))
- Bhavaneetharan, L. (2025, February). A comprehensive review of bias detection and XAI techniques in NLP. In 2025 5th International Conference on Advanced Research in Computing (ICARC) (pp. 1–6). IEEE. ([\[https://doi.org/10.1109/ICARC64760.2025.10962966\]](https://doi.org/10.1109/ICARC64760.2025.10962966)(<https://doi.org/10.1109/ICARC64760.2025.10962966>))

- Braun, V., & Clarke, V. (2024). Supporting best practice in reflexive thematic analysis reporting in palliative medicine: A review of published research and introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). *Palliative Medicine*, 38(6), 608–616. ([<https://doi.org/10.1177/02692163241234800>](<https://doi.org/10.1177/02692163241234800>))
- Creswell, J. W., & Inoue, M. (2025). A process for conducting mixed methods data analysis. *Journal of General and Family Medicine*, 26(1), 4–11. ([<https://doi.org/10.1002/jgf2.736>](<https://doi.org/10.1002/jgf2.736>))
- Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (XAI): A survey. *arXiv preprint arXiv:2006.11371*. ([<https://doi.org/10.48550/arXiv.2006.11371>](<https://doi.org/10.48550/arXiv.2006.11371>))
- Deppe, C., & Schaal, G. S. (2024). Cognitive warfare: A conceptual analysis of the NATO ACT cognitive warfare exploratory concept. *Frontiers in Big Data*, 7, 1452129. ([<https://doi.org/10.3389/fdata.2024.1452129>](<https://doi.org/10.3389/fdata.2024.1452129>))
- Dwivedi, R., et al. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55(9), 1–33. ([<https://doi.org/10.1145/3561048>](<https://doi.org/10.1145/3561048>))
- Ennab, M., & Mcheick, H. (2025). Advancing AI interpretability in medical imaging: A comparative analysis of pixel-level interpretability and Grad-CAM models. *Machine Learning and Knowledge Extraction*, 7(1), 12. ([<https://doi.org/10.3390/make7010012>](<https://doi.org/10.3390/make7010012>))
- Fenstermacher, L., Uzcha, D., Larson, K., Vitiello, C., & Shellman, S. (2023, June). New perspectives on cognitive warfare. In *Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII* (Vol. 12547, pp. 172–187). SPIE. ([<https://doi.org/10.1117/12.2666777>](<https://doi.org/10.1117/12.2666777>))

- Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44–58. ([<https://doi.org/10.1609/aimag.v40i2.2850>](<https://doi.org/10.1609/aimag.v40i2.2850>))
- Islam, S. R., Eberle, W., Ghafoor, S. K., & Ahmed, M. (2021). Explainable artificial intelligence approaches: A survey. *arXiv preprint arXiv:2101.09429*. ([<https://doi.org/10.48550/arXiv.2101.09429>](<https://doi.org/10.48550/arXiv.2101.09429>))
- Machlev, R., et al. (2022). Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI*, 9, 100169. ([<http://dx.doi.org/10.1016/j.egyai.2022.100169>](<http://dx.doi.org/10.1016/j.egyai.2022.100169>))
- Mustafa, A. M., et al. (2024). Interpreting Arabic transformer models: A study on XAI interpretability for Qur'anic semantic-search models. *Jordanian Journal of Computing and Information Technology*, 10. ([<https://doi.org/10.5455/jcit.71-1704878720>](<https://doi.org/10.5455/jcit.71-1704878720>))
- Shafie, R. S., et al. (2024). Psychosemantics of “Us” and “Them” in social media warfare. *SEARCH Journal of Media and Communication Research*, 16(2), 65–81. ([<http://dx.doi.org/10.58946/search-16.2.P5>](<http://dx.doi.org/10.58946/search-16.2.P5>))
- Stanek, D., Klaban, I., & Coufalíková, A. (2025, May). Explainable artificial intelligence: State of the art and beyond. In *2025 International Conference on Military Technologies (ICMT)* (pp. 1–7). IEEE. ([<https://ieeexplore.ieee.org/document/11005246>](<https://ieeexplore.ieee.org/document/11005246>))
- Zhang, H., & Ogasawara, K. (2023). Grad-CAM-based explainable artificial intelligence related to medical text processing. *Bioengineering*, 10(9), 1070. ([<https://doi.org/10.3390/bioengineering10091070>](<https://doi.org/10.3390/bioengineering10091070>)).

## References

- Abdelkareem M. Alashqar (2023). "A Classification of Quran Verses Using Deep Learning." *International Journal of Computing and Digital Systems* (Early Access).
- Ali Mirarab, Morteza Mohammadi Estani, Fa'ezeh S. Tabatabaei Amiri, and Somayeh Dehghani (2023). "Automated Ontology Learning Methods in Quranic Concept Domains: A Systematic Review." *University of Ferdowsi Mashhad Journal of Library and Information Science*, 54(2): 227-254.
- Aliyu R. Yauri, R. Abdul Kadir, Azreen Azman, and M. A. Azmi Murad (2009). "Quranic Verse Extraction Based on Concepts Using an OWL-DL Ontology." In *2009 International Conference on Emerging Technologies*, 29-34. IEEE.
- Ashish Vaswani et al. (2017). "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, 30: 5998-6008.
- Charu C. Aggarwal and ChengXiang Zhai, eds. (2012). *Mining Text Data*. Springer.
- Dwi Rolliawati, Indri S. Rozas, Khalid, and M. Ratodi (2018). "Text Mining Approach for Topic Modeling of the Corpus Al-Qur'an in Indonesian Translation." In *Proceedings of the 2018 International Conference on Islamic Applications in Computer Science and Technologies (IMAN)*, 83-90. ACM.
- Grigorios Tsoumakas and Ioannis Katakis (2007). "Multi-Label Classification: An Overview." *International Journal of Data Warehousing and Mining*, 3(3): 1-13.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of NAACL-HLT*, 4171-4186.

- Jesse Read, Lorenzo Martino, Jesús D. Olmos, and David Luengo (2015). "Scalable Multi-Output Label Prediction: From Classifier Chains to Classifier Trellises." *Pattern Recognition*, 48(6): 2096-2109.
- Jiawei Han, Micheline Kamber, and Jian Pei (2011). *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann (Elsevier).
- Kevin P. Murphy (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Khalilullah Sabri (2010). *Tabaqat al-Ayat: Thematic Classification of the Verses of the Holy Quran in 26 Chapters (with Persian Translation of the Verses)*. Tehran: Amir Kabir Publications.
- M. K. Izzaty, M. S. Mubarak, N. S. Huda, and K. Adiwijaya (2018). "A Multi-Label Classification of Topics of Quranic Verses in English Translation Using Tree-Augmented Naïve Bayes." In *Proceedings of the 2018 International Conference on Information and Communication Technology (ICoICT)*, 360-365. IEEE.
- Maysum H. Panju (2014). *Statistical Extraction and Visualization of Topics in the Quran Corpus*. Technical Report, University of Waterloo.
- Mehdi Rahimibafghi and Masoumeh Pourbagheri (2014). *Clustering of Quranic Verses*. B.Sc. thesis, Computer Engineering Department, Iran University of Science and Technology.
- Min-Ling Zhang and Zhi-Hua Zhou (2014). "A Review on Multi-Label Learning Algorithms." *IEEE Transactions on Knowledge and Data Engineering*, 26(8): 1819-1837.
- Mohammad Alhawarat (2015). "Extracting Topics from the Holy Quran Using Generative Models." *International Journal of Advanced Computer Science and Applications (IJACSA)*, 6(12): 288-294.

- Mohammad Bozorg-Ghamizadeh and Behrooz Minaei (2019). "Discovering Conceptual Relations of Quranic Verses within Tafsirs Using Data Mining Techniques." *Quran-Shenakht Quarterly*, 3(2): 89-102.
- Mohammad Hossein Salehi-Shahroudi, Behrooz Minaei, and Amirreza Ashrafi (2013). "Computer-Based Thematic Text Mining of the Holy Quran to Discover Semantic Connections among Verses, Based on Tafsir al-Mizan." *Quran-Shenakht Quarterly*, vol. 1, no. 2 (Autumn-Winter): 133-150.
- Mohammed Akour, Izzat Alsmadi, and Iyad Alazzam (2014). "MQVC: Measuring Quranic Verses Similarity and Sura Classification Using N-Gram." *WSEAS Transactions on Computers*, 13: 485-491.
- Mostafa Karami, Alireza Talebpour, Farzaneh Tajabadi, and Zeinab Hajimohammadi (2023). "Sentiment Analysis of Quranic Text Using Advanced RoBERTa Deep Learning." *Journal of Interdisciplinary Quranic Studies*, 2(1): 1-14.
- Muazzam Ahmed Siddiqui, Syed Muhammad Faraz, and Sohail Abdul Sattar (2013). "Discovering the Thematic Structure of the Quran Using Probabilistic Topic Models." In *Proceedings of the Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences (ICCIT)*, 234-239. IEEE.
- Muhammad Huzaiifa Bashir, Aqil M. Azmi, Haq Nawaz, Wajdi Zaghouani, Mona Diab, Ala Al-Fuqaha, and Junaid Qadir (2023). "Arabic Natural Language Processing for Qur'anic Research: A Systematic Review." *Artificial Intelligence Review*, 56: 1261-1302.
- Muhammad Shoaib, M. Nadeem Yasin, Ullah K. Hikmat, M. Imran Saeed, and Malik Sikandar H. Khyal (2009). "Relational WordNet Model for Semantic Search in the Holy Quran." In *2009 International Conference on Emerging Technologies*, 29-34. IEEE.

- Noorhan H. Abbas (2009). Qur'an 'Search for a Concept' Tool and Website. PhD diss., School of Computing, University of Leeds, UK.
- R. A. Pane, M. S. Mubarak, N. S. Huda, and A. Adiwijaya (2018). "A Multi-Label Classification of Topics of Quranic Verses in English Translation Using Multinomial Naïve Bayes." In Proceedings of the 2018 International Conference on Information and Communication Technology (ICoICT), 354-359. IEEE.
- Sameer M. Alrehaili and Eric Atwell (2014). "Computational Ontologies for Semantic Tagging of the Quran: A Survey of Past Approaches." In Proceedings of LREC 2014 (9th Language Resources and Evaluation Conference). European Language Resources Association.
- Satrio Adi Prabowo, Adiwijaya, Mohamad Syahrul Mubarak, Said Al Faraby, Muhammad Zidny Naf, and Muhammad Yuslan Abu Bakar (2019). "An Implementation of Support Vector Machine on the Multi-Label Classification of English-Translated Quranic Verses." Journal of Computer Science, 15(12): 1752-1758.
- Syopyan Jaya Putra, Teddy Mantoro, and Muhamad Nur Gunawan (2017). "Text Mining for the Indonesian Translation of the Quran: A Systematic Review." In 2017 International Conference on Computing, Engineering, and Design (ICCED), 1-5. IEEE.
- Thorsten Joachims (1998). "Text Categorization with Support Vector Machines: Learning with Many Relevant Features." In Proceedings of the 10th European Conference on Machine Learning (ECML), LNCS 1398, 137-142. Springer.
- Tom M. Mitchell (1997). Machine Learning. McGraw-Hill.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep Learning." Nature, 521: 436-444