



## Evaluation of the quality of scientific database using Artificial intelligence Algorithm: A case study of Persian and international database

Tahereh Gholami 

Department of Knowledge and Information Science, Faculty of Literature and Humanities, University of Qom, Qom, Iran. Email: [T.gholami@gmail.com](mailto:T.gholami@gmail.com)

| Article Info                                                                                                                                                         | ABSTRACT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Article type:</b><br>Research Article                                                                                                                             | <b>Objective:</b> The primary objective of this research is to develop a novel, AI-driven framework for evaluating the quality of scientific databases. By benchmarking Persian databases against reputable international counterparts, the study maps the current landscape, proposes evidence-based strategies for service enhancement, and provides a reliable reference for researchers, database managers, and scientometric policymakers.                                                                                                                                                                                                                                                                                                                                                       |
| <b>Article history:</b><br>Received: 2026/06/15<br>Received in revised form: 2026/07/31<br>Accepted: 2026/08/24<br>Published online: 2026/10/02                      | <b>Methodology:</b> This applied, descriptive-analytical study investigates active Persian and international scholarly indexing databases. The selected sample comprises Persian platforms—specifically the Scientific Information Database (SID), Magiran, and Noormags—and international counterparts, namely Web of Science and Scopus. Data were gathered by extracting bibliographic records and article metadata. Quality assessment criteria included data accuracy, comprehensiveness, coherence, metadata standardization, currency, retrievability, and duplicate record rates. AI and machine learning algorithms were employed for objective evaluation, specifically to detect metadata errors, identify duplicates, assess descriptive consistency, and evaluate retrieval performance. |
| <b>Keywords:</b><br>Quality assessment, scientific databases, artificial intelligence, machine learning, Persian databases, international databases, scientometrics. | <b>Findings:</b> The results indicate that international databases outperform Persian ones across most quality indicators. Data accuracy averaged 96% for Scopus and 95% for Web of Science, compared to an average of 85% for Persian databases. Metadata consistency scored 94% internationally versus 80% domestically, while duplicate record rates averaged 5% in Persian platforms. Furthermore, update latency was under three months for international databases, compared to approximately five months for Persian equivalents.                                                                                                                                                                                                                                                              |
|                                                                                                                                                                      | <b>Conclusion:</b> Findings demonstrate that AI-driven algorithms provide an effective, scalable approach for objective database quality assessment. While Persian databases play a vital role in disseminating indigenous scholarship, they require structural metadata upgrades, error and duplication reduction, accelerated indexing cycles, and intelligent automation to meet international standards. The proposed framework establishes a solid scientific foundation for database selection, policy refinement, and quality-of-service optimization in regional indexing systems.                                                                                                                                                                                                            |

**Cite this article:** Gholami, T. (2026). Evaluation of the quality of scientific database using Artificial intelligence Algorithm: A case study of Persian and international database. *Science and Technology of Information Management*, 12 (3). 7-16.  
**DOI:** <https://doi.org/10.22091/stim.2026.8499.1843>



© The Author(s)

DOI: 10.22091/stim.2026.8499.1843

**Publisher:** University of Qom



# ارزیابی کیفیت پایگاه‌های داده علمی با استفاده از الگوریتم‌های هوش مصنوعی: مطالعه موردی پایگاه‌های فارسی و بین‌المللی

طاهره غلامی 

گروه علم اطلاعات و دانش‌شناسی دانشگاه قم، دانشکده ادبیات و علوم انسانی، دانشگاه قم، قم، ایران. رایانامه: [T.gholami@gmail.com](mailto:T.gholami@gmail.com)

| اطلاعات مقاله                                                                                                                                                                                     | چکیده                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>نوع مقاله:</b> مقاله پژوهشی</p> <p><b>تاریخ دریافت:</b> ۱۴۰۵/۰۳/۲۵</p> <p><b>تاریخ بازنگری:</b> ۱۴۰۵/۰۵/۰۹</p> <p><b>تاریخ پذیرش:</b> ۱۴۰۵/۰۶/۰۲</p> <p><b>تاریخ انتشار:</b> ۱۴۰۵/۰۷/۱۰</p> | <p><b>هدف:</b> هدف اصلی این پژوهش توسعه چارچوبی نوین و علمی مبتنی بر الگوریتم‌های هوش مصنوعی برای ارزیابی کیفیت پایگاه‌های داده علمی است. پژوهش با شناسایی و تحلیل نقاط قوت و ضعف پایگاه‌های داده فارسی در مقایسه با پایگاه‌های بین‌المللی معتبر، ضمن ارائه تصویری روشن از وضعیت موجود، راهکارهایی علمی برای ارتقای کیفیت خدمات اطلاعاتی و فراهم کردن چارچوبی قابل استناد برای پژوهشگران، مدیران پایگاه‌ها، و سیاست‌گذاران حیطه علم‌سنجی است.</p> <p><b>روش‌شناسی:</b> پژوهش حاضر از نظر هدف کاربردی و از نظر ماهیت و روش توصیفی - تحلیلی است. جامعه پژوهش شامل پایگاه‌های داده علمی فارسی و بین‌المللی فعال در حوزه نمایه‌سازی و اشاعه تولیدات علمی است. در این پژوهش پایگاه‌های داده فارسی شامل پایگاه اطلاعات علمی جهاد دانشگاهی، مگ‌ایران، و نورمگز و پایگاه‌های بین‌المللی شامل وب‌آوساینس و اسکوپوس برای نمونه انتخاب شدند. داده‌ها از طریق استخراج رکوردهای کتاب‌شناختی و فراداده‌های مرتبط با مقالات علمی گردآوری شد. برای ارزیابی کیفیت پایگاه‌ها شاخص‌هایی شامل دقت داده‌ها، جامعیت اطلاعات، انسجام، و استانداردسازی فراداده، به‌روزر بودن داده‌ها، قابلیت بازیابی اطلاعات، و نرخ داده‌های تکراری تعیین شد. برای ارزیابی عینی شاخص‌ها از الگوریتم‌های هوش مصنوعی و یادگیری ماشین برای تحلیل خطاهای فراداده، شناسایی رکوردهای تکراری، ارزیابی هم‌سانی اطلاعات توصیفی، و سنجش کارایی بازیابی اطلاعات استفاده شد.</p> <p><b>یافته‌ها:</b> یافته‌ها نشان داد پایگاه‌های بین‌المللی در بیشتر شاخص‌های کیفیت عملکرد مطلوب‌تری دارند. میانگین دقت داده‌ها در اسکوپوس ۹۶ درصد و در وب‌آوساینس حدود ۹۵ درصد برآورد شد، درحالی‌که این شاخص‌ها در پایگاه‌های فارسی به‌طور میانگین ۸۵ درصد بود. از نظر انسجام فراداده، پایگاه‌های بین‌المللی امتیاز ۹۴ درصد و پایگاه‌های فارسی ۸۰ درصد را کسب کردند. نرخ داده‌های تکراری در پایگاه‌های فارسی به‌طور متوسط ۵ درصد و در پایگاه‌های بین‌المللی کمتر از سه ماه و در پایگاه‌های فارسی حدود پنج ماه گزارش شد.</p> <p><b>نتیجه‌گیری:</b> نتایج نشان می‌دهد استفاده از الگوریتم‌های هوش مصنوعی می‌تواند ابزار مؤثری برای ارزیابی عینی و مقیاس‌پذیر کیفیت پایگاه‌های داده علمی باشد. اگرچه پایگاه‌های فارسی نقش مهمی در دسترسی به تولیدات علمی بومی دارند، برای دستیابی به استانداردهای بین‌المللی نیازمند بهبود ساختار فراداده، کاهش خطاها و داده‌های تکراری، تسریع فرایند به‌روزرسانی، و توسعه فناوریهای هوشمند هستند. چارچوب پیشنهادی می‌تواند مبنایی علمی برای انتخاب پایگاه‌های مناسب، بهبود سیاست‌گذاری‌های اطلاعات علمی و ارتقای کیفیت و کارآمدی پایگاه‌های داده علمی، به‌ویژه پایگاه‌های فارسی فراهم کند.</p> |

## کلیدواژه‌ها:

ارزیابی کیفیت، پایگاه‌های داده علمی، هوش مصنوعی، یادگیری ماشین، پایگاه‌های داده فارسی، پایگاه‌های بین‌المللی، علم‌سنجی.

**استناد:** غلامی، طاهره. (۱۴۰۵). «ارزیابی کیفیت پایگاه‌های داده علمی با استفاده از الگوریتم‌های هوش مصنوعی: مطالعه موردی پایگاه‌های فارسی و

بین‌المللی». *علوم و فنون مدیریت اطلاعات*. دوره ۱۲، شماره ۳، ص: ۱۶-۷. <https://doi.org/10.22091/stim.2026.8499.1843>



## ۱. مقدمه

پیشرفت دانش و فناوری‌های نوین و رشد صنعت اطلاع‌رسانی در دهه اخیر باعث به‌کارگیری ابزارها و روش‌های جدید شده است. استفاده از شبکه جهانی وب و دریافت اطلاعات از پایگاه‌های اطلاعاتی در سراسر جهان، اشاعه اطلاعات را آسان‌تر و سریع‌تر کرده است. امروزه پژوهشگران با استفاده از فناوری‌های نوین و پایگاه‌های اطلاعاتی قادر هستند اطلاعات مورد نیاز خود را به دقت و سرعت جست‌وجو و بازیابی و از این اطلاعات در یادگیری و پژوهش در زمینه‌های مختلف استفاده کنند. پایگاه‌های داده علمی نقش اساسی در پژوهش، تولید دانش و تصمیم‌گیری علمی دارند. با توجه به پیشرفت فناوری‌های نوین و ظهور الگوریتم‌های هوش مصنوعی، امکان ارزیابی جامع و دقیق‌تر پایگاه‌های داده علمی فراهم شده است. الگوریتم‌های هوش مصنوعی با قابلیت شنا سایی الگوها، تحلیل کلان داده‌ها، و کشف خطاها، پتانسیل بالایی در بهبود کیفیت دارند.

نبود ارزیابی نظام‌مند و دقیق کیفیت پایگاه‌های داده علمی فارسی، همراه با فقدان چارچوب تحلیلی مبتنی بر الگوریتم‌های هوش مصنوعی می‌تواند به کاهش دقت جست‌وجو، ناهماهنگی در یکپارچگی فراداده و کاهش قابلیت بازیابی اطلاعات منجر شود. این پژوهش به دنبال پاسخ‌گویی به این شکاف و ارائه چارچوبی کارآمد برای ارزیابی کیفیت پایگاه‌های فارسی و بین‌المللی است، تا با شناسایی نقاط ضعف و قوت پایگاه‌های فارسی و ارائه پیشنهادها برای کاربردی به ارتقای کیفیت خدمات این پایگاه‌ها و افزایش بهره‌وری پژوهشگران یاری رساند و در نهایت چارچوبی علمی و قابل استناد برای ارزیابی پایگاه‌های داده ارائه کند.

مطالعات بین‌المللی نشان می‌دهند که کاربرد الگوریتم‌های هوش مصنوعی در ارزیابی کیفیت پایگاه‌های داده علمی به‌طور فزاینده‌ای در کانون توجه قرار گرفته است. الگوریتم‌های هوش مصنوعی و یادگیری ماشین با توانایی پردازش کلان داده‌ها، کشف الگوهای پنهان، و کاهش خطای انسانی ظرفیت بالایی برای ارتقای فرایند ارزیابی کیفیت پایگاه‌های داده علمی دارند. پژوهش‌های اخیر با بهره‌گیری از الگوریتم‌های یادگیری ماشین، شبکه‌های عصبی، و خوشه‌بندی توانسته‌اند داده‌های ناقص، ناسازگاری‌ها، و رکوردهای تکراری را با دقت بالا شناسایی کنند و معیارهای کمی و کیفی قابل اتکایی برای ارزیابی کیفیت پایگاه‌ها ارائه دهند.

پژوهش‌های بین‌المللی نشان می‌دهند که هوش مصنوعی می‌تواند در مدیریت داده‌ها، بهینه‌سازی عملکرد پایگاه‌های داده، بهبود امنیت، و پشتیبانی از سیستم‌های پیشنهاددهنده نقش مهمی ایفا کند.

مدلی<sup>۱</sup> (۲۰۲۵) در مقاله خود به بررسی بهبود معماری پایگاه‌های داده با استفاده از هوش مصنوعی پرداخته و روش‌هایی برای افزایش کارایی و کاهش هزینه‌های پردازشی پیشنهاد کرده است.

توال<sup>۲</sup> (۲۰۲۵) در مقاله‌ای تحلیلی درباره استفاده از مدل‌های زبانی و هوش مصنوعی برای ارزیابی کیفیت علمی پژوهش‌ها، به نقد شاخص‌های سنتی مانند بیبلیومتریکس<sup>۳</sup> پرداخته و چالش‌های هوش مصنوعی را بررسی می‌کند. پژوهش توال به بررسی تأثیر فناوری‌های نوین مانند چت جی‌پی‌تی در سنجش کیفیت پرداخته و هوش مصنوعی به مثابه ابزار جایگزین یا مکمل سنج‌های سنتی در کانون توجه پژوهشگر قرار گرفته است.

گولین اویناگا<sup>۴</sup> (۲۰۲۵) در بررسی گسترده‌ای از کیفیت داده در عصر هوش مصنوعی نشان داده است که بُعدهای اطمینان داده، کامل بودن، سازگاری، و قابلیت باز استفاده از معیارهای اساسی کیفیت داده هستند و تأکید داشته هوش مصنوعی می‌تواند در بهبود کیفیت داده‌ها در پایگاه‌های علمی نقش داشته باشد.

خان<sup>۵</sup> و دیگران (۲۰۲۴) نقش هوش مصنوعی در مدیریت پایگاه‌های داده را بررسی و ابزارهایی برای تحلیل و پردازش حجم بالای داده ارائه کرده‌اند.

---

۱. Maddali

۲. Theiwall.

۳. Bibliometrics

۴. Bibliometrics

۵. khan

سعیدنیا<sup>۱</sup> و دیگران (۲۰۲۴) در مطالعه‌ای جامع، نقش الگوریتم‌های هوش مصنوعی در علم‌سنجی، وب‌سنجی، و ارزیابی عملکرد علمی را بررسی کردند. نتایج نشان داد که هوش مصنوعی می‌تواند در تحلیل انتشارها، شبکه‌های استنادی، تشخیص هم‌نوشتگی‌ها، تجزیه و تحلیل شبکه‌های هم‌نویسندگی، و سنجش کیفیت پژوهش با شاخص‌های کمی و کیفی نقش برجسته‌ای ایفا کند.

ما<sup>۲</sup> و دیگران (۲۰۲۲) در مقاله خود به استفاده از هوش مصنوعی در جست‌وجوی متون و انتخاب پایگاه پابمد پرداخته و بیان داشتند که یک روش تقویت و غربالگری کلمات کلیدی مبتنی بر پردازش زبان طبیعی را توسعه داده‌اند تا به دانشمندان کمک کند به راحتی کلمات کلیدی خود را در جست‌وجوی موضوعی اصلاح کنند. این روش می‌تواند کلمات کلیدی منتخب معنادار را از عناوین و چکیده‌های یک جست‌وجوی اولیه با استفاده از دانش قبلی، نمودارهای دانش، و یادگیری ماشین استخراج کند. نتیجه پژوهش آن‌ها نشان داد تکنیک‌های هوش مصنوعی می‌توانند به افزایش جامعیت و دقت جست‌وجو کمک کنند. این نتایج نشانگر پتانسیل بالای کاربرد روش‌های هوش مصنوعی در جست‌وجوهای نشریات پزشکی و سایر کاربردهای مبتنی بر متن است.

محمد<sup>۳</sup> و دیگران (۲۰۲۲) در مقاله خود به بررسی اثر کیفیت داده بر عملکرد الگوریتم‌های یادگیری ماشین با تأکید بر ابعاد کیفیت داده پرداخته‌اند. نتایج پژوهش آن‌ها نشان می‌دهد ابعاد کیفیت داده مانند دقت، کامل بودن، و سازگاری از عوامل اساسی تأثیرگذار بر عملکرد مدل‌های هوش مصنوعی هستند. این نتایج مبنای مفهومی مناسب برای ارزیابی کیفیت داده‌های علمی در پایگاه‌های داده فراهم می‌آورد.

سینق<sup>۴</sup> (۲۰۱۷) در مقاله‌ای با عنوان نقش هوش مصنوعی در امنیت حفاظت از پایگاه داده جدید بیان کرده راهکارهای مبتنی بر هوش مصنوعی می‌توانند ناهنجاری‌ها را شناسایی و از دسترسی‌های غیرمجاز جلوگیری کنند. مقاله او مروری جامع بر روش‌های موجود داشته و اصول کار سیستم‌های امنیتی مبتنی بر هوش مصنوعی را برجسته می‌کند و آن‌ها را با رویکردهای سنتی مقایسه می‌کند.

در مقابل پژوهش‌های بالا، پژوهش‌های داخلی بیشتر بر روش‌های توصیفی تحلیلی و شاخص‌های علم سنجی سنتی تمرکز داشته‌اند و بدون استفاده از الگوریتم‌های هوشمند شاخص‌هایی مانند تعداد مدارک، میزان استناد، و قابلیت بازایی اطلاعات را ارزیابی کرده‌اند (احمدی و رضایی، ۱۴۰۰).

به‌طور کلی در کشور ما تعداد پژوهش‌های انجام‌شده در زمینه کاربرد هوش مصنوعی برای ارزیابی کیفیت پایگاه‌های داده علمی فارسی محدود است. مطالعات فارسی نیز در این زمینه رویکردهای کاربردی و پژوهشی مشابهی داشته‌اند.

رفیعی خشنود (۱۴۰۲)، در مقاله خود به ارزیابی کمی و کیفی مقاله‌های جهاد دانشگاهی در پایگاه‌های وب‌اوساینس<sup>۵</sup> و اسکوپوس<sup>۶</sup> پرداخته که یافته‌های پژوهش آن‌ها بیانگر آن است که تعداد مقاله‌های علمی جهاد دانشگاهی در وب‌اوساینس برابر با ۵۲۷۳ و در پایگاه اسکوپوس برابر با ۹۰۹۱ است. تعداد مقاله‌های علمی جهاد دانشگاهی در هر دو پایگاه روبه افزایش است. شاخص‌های کمی و کیفی مقاله‌های جهاد دانشگاهی در هر دو پایگاه با مقدار ۰/۰۹۳۵ همبستگی دارند.

ظهوریان نادعلی و دیگران (۱۴۰۲) در پژوهشی به ترسیم نقشه تولیدات علمی پژوهشگران ایرانی نمایه‌شده در پایگاه اسکوپوس در حوزه هوش مصنوعی پرداختند. نتایج پژوهش آن‌ها نشان داد ایران در زمینه هوش مصنوعی در رتبه ۳۱ ام و در خاورمیانه در جایگاه سوم قرار دارد. نتایج پژوهش آن‌ها نشان می‌دهد روند تولیدات پژوهشگران ایرانی به‌طور کلی صعودی بوده است، با این حال سرعت رشد در مقایسه با سایر کشورها کمتر است.

1. Saeidnia
2. Ma
3. Mohammed
4. Singh
5. Web Of Science
6. Scopus

عظیمی و اسماعیلی (۱۴۰۰) مؤلفه‌های هوش مصنوعی در پایگاه‌های اطلاعاتی ایرانی را شناسایی و تأثیر آن‌ها در جست‌وجوی معنایی و مدیریت داده‌ها را بررسی کردند.

خادمی‌زاده و دیگران (۱۳۹۹) در پژوهشی معیارهای ارزیابی کیفیت اطلاعات در سامانه‌های اطلاعاتی مستندات فنی را به ابعاد سازمانی، فردی، اطلاعاتی، فناوری، و نظام‌مند طبقه‌بندی کرده‌اند. ایشان در مقاله خود بیان کردند این نوع دسته‌بندی می‌تواند به شرکت‌های دانش‌بنیان و طرح‌محور کمک کند تا وضعیت جاری کیفیت مستندات خود را در انطباق با آن‌ها بسنجند.

با توجه به پژوهش‌های بالا، مشخص شد که هوش مصنوعی در سطح بین‌المللی به مثابه ابزاری مؤثر برای بهبود عملکرد پایگاه‌های داده، امنیت، و پردازش حجم بالای اطلاعات به کار گرفته می‌شود.

با این حال پژوهش‌های انجام شده در ایران بیشتر تو صیفی هستند و به شناسایی مؤلفه‌ها می‌پردازند. کمبود مطالعاتی که به توسعه مدل‌های کاربردی و ارزیابی عملکرد هوش مصنوعی در پایگاه‌های اطلاعاتی ملی بپردازند، به‌طور جدی احساس می‌شود. از این‌رو پژوهش حاضر در پی آن است که با تمرکز بر شناسایی و تحلیل مؤلفه‌های هوش مصنوعی در پایگاه‌های اطلاعاتی ایرانی، شکاف موجود در این حوزه را پر کرده و راهکارهایی برای بهبود مدیریت داده‌ها ارائه کند.

## ۲. پرسش‌های پژوهش

پژوهش حاضر در پی پاسخ‌گویی به پرسش‌های زیر است:

۱. شاخص‌های کیفی پایگاه‌های علمی کدامند؟
۲. کیفیت داده‌ها در پایگاه‌های فارسی و بین‌المللی بر اساس شاخص‌های دقت، انسجام، فراداده، جامعیت، و قابلیت بازیابی چگونه است؟
۳. الگوریتم‌های هوش مصنوعی چگونه می‌توانند داده‌های ناقص و ناسازگار را شناسایی کنند؟
۴. تفاوت معنادار کیفیت پایگاه‌های فارسی و بین‌المللی در شاخص‌های اصلی چیست؟
۵. با استفاده از الگوریتم هوش مصنوعی پایگاه‌های مطالعه‌شده در پژوهش کنونی در چه رتبه‌ای قرار دارند؟

## ۳. روش پژوهش

این پژوهش از نوع تحقیق توصیفی - تحلیلی و با هدف کاربردی است. داده‌ها از سه پایگاه فارسی (اس‌آی‌دی، مگ‌ایران، و نورمگز) و دو پایگاه بین‌المللی وب‌آوساینس و اسکوپوس جمع‌آوری شدند. شاخص‌های کیفیت شامل دقت داده‌ها، جامعیت، انسجام فراداده، به‌روزر بودن، قابلیت بازیابی، و نرخ داده‌های تکراری هستند که با مطالعه منابع و مدارک علمی در این راستا تنظیم شدند. تحلیل داده‌ها با استفاده از یادگیری ماشین انجام شد. بر این اساس پژوهش کنونی در پی این است که با برگزیدن رویکردی مبنی بر هوش مصنوعی به ارزیابی کیفیت پایگاه‌های داده علمی فارسی و بین‌المللی بپردازد و با انجام مطالعه موردی نقاط ضعف، قوت، و شکاف‌های موجود بین آن‌ها را ارزیابی کند.

نمونه پژوهشی به شیوه‌ای هدفمند انتخاب شده است. از هر پایگاه داده مجموعه‌ای از مقاله‌های علمی منتشر شده در یک بازه زمانی پنج‌ساله (۲۰۲۰-۲۰۲۵) برای نمونه انتخاب شد. حجم نمونه (۲۰۰ مقاله در هر پایگاه) به گونه‌ای انتخاب شد که نماینده مناسبی از ساختار داده‌ها و فراداده‌های هر پایگاه باشد.

داده‌های پژوهشی با استخراج خودکار اطلاعات (خزش داده‌ها<sup>۱</sup> و رابط‌های برنامه نویسی کاربردی<sup>۲</sup>) از پایگاه‌های داده منتخب گردآوری شد. محدودیت‌های دسترسی به ای‌پی‌آی و داده‌های کامل برخی پایگاه‌های بین‌المللی موجب شد حجم نمونه برای شاخص‌ها محدود شود. اطلاعات استخراج شده شامل داده‌های کتاب‌شناختی، فراداده‌های نمایه‌سازی، اطلاعات استنادی، و شاخص‌های بازیابی است. تحلیل داده‌ها با الگوریتم‌های هوش مصنوعی مبتنی بر مدل‌های استاندارد انجام شد.

1. Data Crawling

2. API( Application Programming Interface)

برای ارزیابی کیفیت یک چارچوب شاخص‌های چندبُعدی شامل دقت، جامعیت، انسجام فراداده، به‌روزرودن، و قابلیت بازیابی اطلاعات طراحی و داده‌ها بعد از پیش‌پردازش (حذف داده‌های تکراری، نرمال‌سازی، و پاک‌سازی) آماده تحلیل شدند.

#### ۴. یافته‌ها

##### پرسش اول: شاخص‌های کیفی پایگاه‌های علمی کدامند؟

این شاخص‌ها با مطالعه متون و منابع مختلف علمی خارجی به دست آمده و به‌عنوان چک لیستی از شاخص‌های ارزیابی کیفیت پایگاه‌های داده علمی استفاده شد. در جدول ۱، شاخص‌های ارزیابی کیفیت پایگاه‌های داده علمی ترسیم شد.

جدول ۱. شاخص‌ها و سنجه‌های عددی ارزیابی کیفیت پایگاه‌های داده علمی

| بُعد کیفیت      | شاخص             | تعریف عملیاتی                                |
|-----------------|------------------|----------------------------------------------|
| دقت             | درستی داده‌ها    | میزان تطابق اطلاعات کتاب‌شناختی با منبع اصلی |
| جامعیت          | پوشش اطلاعات     | نسبت رکوردهای کامل به کل رکوردها             |
| انسجام          | یکنواختی فراداده | نسبت رکوردهای کامل به کل رکوردها             |
| به‌روزرودن      | تازگی داده‌ها    | فاصله زمانی بین انتشار و نمایه‌سازی          |
| قابلیت بازیابی  | کارایی جست‌وجو   | توانایی بازیابی نتایج مرتبط                  |
| داده‌های تکراری | افزونگی          | میزان رکوردهای تکراری                        |

##### پاسخ پرسش دوم: کیفیت داده‌ها در پایگاه‌های فارسی و بین‌المللی براساس شاخص‌های دقت، انسجام فراداده، جامعیت، و قابلیت بازیابی چگونه است؟

جدول ۲، درصد رعایت هر یک از شاخص‌های ارزیابی کیفی در پایگاه‌های اطلاعاتی فارسی و بین‌المللی را در طی این پژوهش نشان می‌دهد.

جدول ۲. وضعیت پایگاه‌های علمی در شاخص‌ها

| پایگاه داده | نوع        | تعداد مقاله‌های نمونه | دقت داده‌ها % | جامعیت داده‌ها % | انسجام فراداده % | به‌روزرودن (ماه) | قابلیت بازیابی f1-score | نرخ تکراری % |
|-------------|------------|-----------------------|---------------|------------------|------------------|------------------|-------------------------|--------------|
| اس‌آی‌دی    | فارسی      | ۲۰۰                   | ۸۸            | ۸۵               | ۸۰               | ۶                | ۰/۸۲                    | ۵            |
| مگ‌ایران    | فارسی      | ۲۰۰                   | ۹۰            | ۸۷               | ۸۲               | ۵                | ۰/۸۴                    | ۴            |
| نورمگز      | فارسی      | ۲۰۰                   | ۸۵            | ۸۳               | ۷۸               | ۷                | ۰/۸۰                    | ۶            |
| وب‌آوساینس  | بین‌المللی | ۲۰۰                   | ۹۶            | ۹۲               | ۹۵               | ۲                | ۰/۹۳                    | ۱            |
| اسکوپوس     | بین‌المللی | ۲۰۰                   | ۹۵            | ۹۱               | ۹۴               | ۳                | ۰/۹۲                    | ۲            |

یافته‌های پژوهش مبتنی بر داده‌های برآمده از جدول ۲ نشان داد که پایگاه‌های بین‌المللی در بیشتر شاخص‌های کیفیت عملکرد مطلوب‌تری دارند. به‌طور مشخص میانگین دقت داده‌ها در پایگاه اسکوپوس برابر با ۹۶٪ و در وب‌آوساینس حدود ۹۵٪ برآورد شد، درحالی‌که این شاخص‌ها در پایگاه‌های فارسی به‌طور میانگین ۸۵ درصد بود. از نظر انسجام فراداده، پایگاه‌های بین‌المللی امتیاز ۹۴٪ را کسب کردند، درحالی‌که این امتیاز برای پایگاه‌های فارسی ۸۰٪ گزارش شد. همچنین نرخ داده‌های تکراری در پایگاه‌های فارسی به‌طور متوسط ۵ درصد برآورد شد. درحالی‌که این میزان در پایگاه‌های بین‌المللی کمتر از ۲ درصد بود. در شاخص به‌روزرودن اطلاعات نیز پایگاه‌های بین‌المللی با میانگین تأخیر زمانی کمتر از سه ماه عملکرد بهتری نسبت به پایگاه‌های فارسی با تأخیر پنج‌ماهه داشتند.

### پرسش ۳: الگوریتم‌های هوش مصنوعی چگونه می‌توانند داده‌های ناقص و ناسازگار را شناسایی کنند؟

الگوریتم‌های هوش مصنوعی از جمله درخت تصمیم و خوشه بندی K-means توانستند داده‌های ناقص و ناسازگار را با دقت بالایی شناسایی و الگوهای ناهماهنگی در فراداده‌ها را گروه‌بندی کنند. یافته‌های پژوهش نشان داد که الگوریتم‌های هوش مصنوعی توانایی بالایی در شناسایی ناهماهنگی‌ها، داده‌های ناقص، و خطاهای ساختاری پایگاه‌های داده علمی دارند. الگوریتم درخت تصمیم توانست داده‌های ناقص و ناسازگار با دقت ۹۵ درصد شناسایی کند و الگوریتم k-means خوشه‌بندی خطاهای ساختاری را شناسایی و سه خوشه اصلی از ناسازگاری‌ها (ناقص بودن فیلدها، تکراری بودن داده‌ها، ناسازگاری فراداده) ارائه داد. استفاده از الگوریتم‌های هوش مصنوعی می‌تواند ابزار مؤثری برای ارزیابی عینی و مقیاس‌پذیر کیفیت پایگاه‌های داده علمی باشد.

### پرسش ۴: تفاوت معنادار کیفیت پایگاه‌های فارسی و بین‌المللی در شاخص‌های اصلی چیست؟

مطابق جدول ۳ تحلیل بین‌گروهی (فارسی و بین‌المللی) نشان داد که تفاوت‌های معناداری در شاخص‌های دقت، انسجام، و قابلیت بازیابی وجود دارد. این یافته‌ها تأکید می‌کند که پایگاه‌های علمی فارسی نیازمند بهبود فراداده، افزایش جامعیت داده‌ها، و کاهش خطاهای ساختاری هستند. از دیگر یافته‌های جدول ۳ می‌توان به این اشاره کرد که پایگاه‌های بین‌المللی از نظر دقت، جامعیت، و انسجام فراداده بهتر از پایگاه‌های فارسی هستند، زمان به‌روزر بودن (تأخیر) در پایگاه‌های فارسی دو برابر پایگاه‌های بین‌المللی است. قابلیت بازیابی اطلاعات در پایگاه‌های بین‌المللی بالاتر است ( $F1\text{-score}=0.92$ ) و در نهایت نرخ داده‌های تکراری در پایگاه‌های فارسی بیشتر است.

جدول ۳. مقایسه بین‌گروهی (پایگاه‌های داده فارسی در مقایسه با پایگاه داده بین‌الملل)

| شاخص                                 | پایگاه‌های فارسی (سه مورد) | پایگاه‌های بین‌المللی (دو مورد) |
|--------------------------------------|----------------------------|---------------------------------|
| دقت داده‌ها (%)                      | ۸۷/۷                       | ۹۴/۳                            |
| جامعیت داده‌ها (%)                   | ۸۵                         | ۹۱                              |
| انسجام فراداده‌ها (%)                | ۸۰                         | ۹۳                              |
| به‌روزر بودن (ماه)                   | ۶                          | ۳                               |
| قابلیت بازیابی ( $F1\text{-score}$ ) | ۰/۸۲                       | ۰/۹۲                            |
| نرخ داده‌های تکراری (%)              | ۵                          | ۲                               |

در جدول ۳ داده‌های جمع‌آوری شده از سه پایگاه علمی فارسی و دو پایگاه بین‌المللی با استفاده از الگوریتم‌های هوش مصنوعی تحلیل شدند. شاخص‌های کیفیت شامل دقت داده‌ها، جامعیت، انسجام فراداده، به‌روزر بودن، قابلیت بازیابی، و نرخ داده‌های تکراری بودند. با تحلیل میانگین شاخص‌ها مشاهده شد که پایگاه‌های بین‌المللی دارای دقت بالاتر (۹۴/۳٪) در برابر (۸۷/۷٪)، انسجام فراداده بهتر (۹۳٪) در برابر (۸۰٪)، و قابلیت بازیابی اطلاعات بالاتر ( $F1\text{-score}$  0.92 در برابر ۰/۸۲) هستند. علاوه بر این، زمان به‌روزر بودن داده‌ها در پایگاه‌های فارسی به‌طور میانگین دو برابر پایگاه‌های بین‌المللی است و نرخ داده‌های تکراری در پایگاه‌های فارسی بیشتر است.

### پرسش ۵: پایگاه‌های منتخب با استفاده از الگوریتم هوش مصنوعی چه رتبه‌ای می‌گیرند؟

الگوی یادگیری ماشین برای امتیازدهی و رتبه‌بندی پایگاه‌ها براساس کیفیت مقاله‌ها و اعتبار منابع استفاده شد. در جدول ۴ رتبه‌بندی پایگاه‌ها با استفاده از الگوریتم هوش مصنوعی نشان داد که پایگاه‌های بین‌المللی دارای مقاله‌هایی با امتیاز بالاتر در شاخص‌های اعتبار علمی و ارجاعات هستند. پایگاه‌های فارسی در زمینه خاص علوم انسانی عملکرد مناسبی داشتند، اما میانگین کیفیت کلی آن‌ها کمتر از پایگاه‌های بین‌المللی است. پایگاه‌های فارسی کیفیت متوسطی داشته، اما در زمینه‌های خاص (برای نمونه علوم اجتماعی و فرهنگی) عملکرد مناسبی دارند.

جدول ۴. رتبه‌بندی با هوش مصنوعی

| رتبه | امتیاز کلی<br>(۰-۱۰۰) | پایگاه     |
|------|-----------------------|------------|
| ۴    | ۶۳                    | اس‌آی‌دی   |
| ۳    | ۶۵                    | مگ‌ایران   |
| ۵    | ۶۲                    | نورمگز     |
| ۱    | ۹۲                    | وب‌آوساینس |
| ۲    | ۹۰                    | اسکوپوس    |

## ۵. نتیجه‌گیری

تحلیل داده‌ها با هدف بررسی کیفیت پایگاه‌های علمی فارسی و بین الملل انجام شد. برای این منظور در کل، پنج پایگاه منتخب (فارسی و بین‌المللی)، پایگاه‌های فارسی اس‌آی‌دی، مگ‌ایران، نورمگز و پایگاه‌های بین‌المللی اسکوپوس و وب‌آوساینس بررسی شدند. شاخص‌های تحلیل کیفیت شامل دقت داده‌ها جامعیت، انسجام فراداده، به‌روزرودن، قابلیت بازیابی، و نرخ داده‌های تکراری بوده که در مورد هر دو دسته پایگاه‌های علمی فارسی و بین‌المللی ارزیابی انجام شد و این ارزیابی نشان داد که پایگاه‌های بین‌المللی از وضعیت مطلوب‌تری نسبت به پایگاه‌های فارسی برخوردار هستند. پایگاه‌های بین‌المللی از نظر دقت، جامعیت، و انسجام فراداده بهتر از پایگاه‌های فارسی هستند. زمان به‌روزرودن در پایگاه‌های فارسی دو برابر پایگاه‌های بین‌المللی است. قابلیت بازیابی اطلاعات در پایگاه‌های بین‌المللی بالاتر از پایگاه‌های فارسی بوده و نرخ داده‌های تکراری در پایگاه‌های فارسی بیشتر است.

الگوریتم‌های هوش مصنوعی توانستند داده‌های ناقص و ناسازگار را با دقت بالایی شناسایی و الگوهای ناهماهنگی در فراداده‌ها را گروه‌بندی کنند. یافته‌های پژوهش نشان داد که الگوریتم‌های هوش مصنوعی توانایی بالایی در شناسایی ناهماهنگی‌ها، داده‌های ناقص، و خطاهای ساختاری پایگاه‌های داده علمی دارند.

در نهایت، رتبه‌بندی پایگاه‌ها با استفاده از الگوریتم هوش مصنوعی نشان داد که پایگاه‌های بین‌المللی دارای مقاله‌ها با امتیاز بالاتر در شاخص‌های اعتبار علمی و ارجاعات هستند. پایگاه‌های فارسی در زمینه خاص علوم انسانی عملکرد مناسبی داشتند، اما میانگین کیفیت کلی کمتر از پایگاه‌های بین‌المللی است.

بر اساس پژوهش انجام شده، پایگاه‌های بین‌المللی کیفیت بالاتری داشته و پایگاه‌های فارسی نیازمند بهبود انسجام فراداده و به‌روزرودن هستند.

با توجه به تحلیل داده‌های پژوهش حاضر نتایج زیر را می‌توان بیان کرد:

۱. کیفیت پایگاه‌های داده علمی بین‌المللی به‌ویژه در شاخص‌های انسجام فراداده، دقت، و قابلیت بازیابی اطلاعات بالاتر از پایگاه‌های فارسی است.
۲. الگوریتم‌های هوش مصنوعی ابزارهای مؤثری برای ارزیابی کیفیت داده‌ها هستند و توانایی شناسایی خطاهای انسانی، داده‌های ناقص، و ناسازگاری را دارند.
۳. ترکیب شاخص‌های کمی و کیفی با الگوریتم‌های هوش مصنوعی چارچوبی نظام‌مند و عینی برای ارزیابی کیفیت پایگاه‌های داده علمی فراهم می‌کند.
۴. پایگاه‌های فارسی نیازمند بهبود روش‌های مدیریت داده و استانداردسازی فراداده‌ها هستند تا سطح کیفیت آن‌ها به پایگاه‌های بین‌المللی نزدیک شود.
۵. پژوهش حاضر می‌تواند به بهبود عملکرد پژوهشگران، ارتقای قابلیت بازیابی اطلاعات، و شفافیت در نظام ارزیابی علمی کشور کمک کند.

در نهایت استفاده از شاخص‌های ترکیبی و الگوریتم‌های هوش مصنوعی می‌تواند به بهبود کیفیت پایگاه‌های علمی فارسی، افزایش اعتبار علمی، و بهره‌وری پژوهشگران کمک کند و فاصله موجود با پایگاه‌های بین‌المللی را کاهش دهد. در راستای پژوهش

انجام‌شده پیشنهادهایی به این شکل ارائه می‌شوند، از جمله ارتقای پوشش موضوعی پایگاه‌های فارسی، به‌روزرسانی منظم اطلاعات، توسعه رابط کاربری و موتور جست‌وجوی هوشمند، و سرمایه‌گذاری هدفمند در توسعه زیرساخت‌های هوشمند اطلاعاتی و بهره‌گیری از فناوری‌های نوین در مدیریت دانش علمی برای ارتقای کیفی پایگاه‌های داخلی و کمک هر چه بیشتر به چرخه تولیدات پژوهشی در کشور.

## ۶. فهرست منابع

- ظهوریان نادعلی، الف.، سلیمانی روزبهانی، ف.، و اجاقی، ف. (۱۴۰۲). نگاشت نقشه علمی تحقیقات هوش مصنوعی ایران مبتنی بر پایگاه استنادی اسکوپوس. سالهای (۱۹۷۸-۲۰۲۲). پژوهش‌نامه علم سنجی، ۱۹(۱)، ۴۶۹-۵۰۶. <https://doi.org/10.22070/rsci.2022.15866.1565>
- خادمی زاده، ش.، فرج پهلوی، ع.، و قاسمی زاده، ش. (۱۳۹۹). معیارهای ارزیابی کیفیت اطلاعات در سامانه‌های اطلاعاتی مستندات فنی: مرور نظام‌مند. مطالعات کتابداری و سازماندهی اطلاعات (مطالعات ملی کتابداری و سازماندهی اطلاعات، ۳۱(۲)، ۱۰۲-۱۱۴. <https://www.sid.ir/paper/954464/fa>
- رفیعی، خشنود، م. (۱۴۰۲). ارزیابی گمی و کیفی مقالات علمی جهاد دانشگاهی در پایگاه‌های اسکوپوس و وب آو سانس. مجله علم سنجی کاسپین، ۷۷-۹۱، ۱۱۰(۲).
- عظیمی، م.، و اسماعیلی، س. (۱۴۰۰). شناسایی مولفه‌های هوش مصنوعی در پایگاه‌های اطلاعاتی ایرانی. دانش شناسی، ۹۴-۹۷، ۱۰۷-۹۴. [https://journals.iau.ir/article\\_687376.html](https://journals.iau.ir/article_687376.html)
- Ahmadi, M., & Rezaei, F. (2022). Assessment of the quality and retrievability of Persian scientific databases. *Iranian journal of information science and technology*, 18(2), 45-61.
- Azimi, M., Esmaeili, S. (2022). Identify the Components of Artificial Intelligence in Iranian Databases. *Journal of Knowledge Studies*, 14 (54), 94-107. [in Persian]
- Guillen-Aguinaga, M. (2025). Data Quality in the Age of AI: A Review of Governance, Ethics, and the Fair Principles. *MDPI*, 10(12), 201. DOI:10.3390/data10120201.
- Khademizadeh, s., Farajpahlu, a.h., & Ghasemizadeh, s. (2020). Information Quality Assessment Criteria for Technical Documentation Information Systems: A Systematic Review. *Librarianship and information organization studies (Journal of national studies on librarianship and information organization)*, 31(122), 104-114. <https://sid.ir/paper/954464/en> [In Persian]
- Khan, M. A., Bibi, S., Toor, s., & Rashid, M. (2024). Role of Artificial intelligence in big database management. *The Asian bulletin of big data management*, 4(2). doi:1062019/abbdm.v4i02.164
- Kuhen, P., Deplanque, A., & Fluck, E. (1994). Criteria of quality assessment for scientific databases. *Journal of chemical information and computer sciences*, 34(3), 517-519.
- Liu, Y., Zhang, H., & Wang, J. (2024). Artificial intelligence for assessing the quality of scientific databases: A machine learning approach. *Journal of informetrics*, 18(1), 101-118.
- Maddali, G. (2025). Enhancing Database Architectures with Artificial intelligence (AI). *International Journal of Scientific Research in science and Technology*. Doi:10.32628/IJsRsT2512331.
- Ma, J., Wu, X., Huang, L. (2022). The use of artificial intelligence in literature search and selection of the pubmed database. *Scientific programming*, (16), 1-9. DOI:101155/2022/8855307.
- Mohammed, S., Budach, I., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N... & Harmouch, H. (2022). The effects of data quality on machine learning performance. *Information systems*, 132.102549.1-18. <https://arxiv.org/abs/2207.14529>
- Maddali, G. (2025). Enhancing database architectures with artificial intelligence (AI). *International journal of scientific research in science and technology*, 12(3). 296-308. Doi:10.32628/ijrst25112331.
- Pannu, A. (2015). Artificial intelligence and its application in different areas. *Artificial intelligence*, 4(10), 79-84.
- Rafiee khoshnood, M. (2023). Quantitative and qualitative evaluation of academic articles published in Academic Center for Education, Culture and Research in Web of Science and Scopus databases. *CJS*, 10 (2), 77-91. <http://cjs.mubabol.ac.ir/article-1-323-fa.html> [in Persian]

- Saeidnia, H.R. Hosseini, E., Abdoli, S.& Ausloos, M. (2024). Unleashing the power of ai: A systematic Review of ai –Enhanced scientometrics, webometrics, and Bibliometrics. *Arxiv*, 2403, 18838.DOI:10.48550/arxiv.2403.18838
- Singh, R. P, S., & kumar, A. (2023). Deep learning applications in ranking and evaluating acientific databases. *Journal of data and information science*, 8(2), 89-105.<https://doi.org/10.2478/jdis-2023-0007>
- Singh, B. (2017). The role of Artificial intelligence in modern Database security and protection. *International journal of research in electronic and computer engineering*, 5(4).
- Thelwall, M. (2025). Research quality evalution by Ai in the era of large language models:Advantages, disadvantages, and systematic effects. *Arxiv*, 2506, 07748. doi:1048550/arxiv.2506.07748
- Wang, H., Yan, Y., Han, Z. (2025). AI-enhanced database management system. *computing magazine of the ccf*, 1(5), 68-82. Doi:10.11991/cccf.202509009
- Wang, l., & Chen, Q. (2023). Comprehensive quality analysis of international scientific databases using AI clustering techniques. *Scientometrics*, 125(4), 2045-2063. <http://doi.org/10.1007/s11192-023-04567-3>
- Zohoorian Nadali, I., Soleimani Roozbahani, F., Ojaghi, H. (2023). Iranian Artificial Intelligence Research Map Based on Scopus Citation Database (1978-2022). *Scientometric research journal*, 9(17), 469-506. <https://doi.org/10.22070/rsci.2022.15866.1565> [in Persian]