



أثر التحيزات البرمجية على استرجاع النصوص الدينية المذهب الجعفري الاثني عشري نموذجاً

المدرس المساعد محمد سالم خطاب البخاتي¹ المدرس المساعد هاجر جبار عمار
الجابري²

¹ جامعة ميسان كلية التربية الأساسية – العراق

² جامعة ميسان كلية التربية – العراق

Mmohod85@gmail.com
mohamed.salem@uomisan.edu.iq

الملخص. يتناول هذا البحث ظاهرة التحيزات البرمجية في أنظمة استرجاع النصوص الدينية الرقمية، مع التركيز على المذهب الجعفري الاثني عشري نموذجاً تطبيقياً. يرصد البحث الأثر الذي تتركه خوارزميات معالجة اللغة الطبيعية والذكاء الاصطناعي على عمليات البحث والاسترجاع في مصادر أهل البيت (عليهم السلام) كالكتب الأربعة وبحار الأنوار. يكشف البحث عن أنواع متعددة من التحيزات: تحيز بيانات التدريب، والتحيز الانتقائي، وتحيز التمثيل، والتحيز اللغوي. ولإثبات هذه الفرضية نظرياً وتجريبياً، أجرى الباحثان دراسة تجريبية على أربعة أنظمة رقمية بخمسة استعلامات في الفقه والعقيدة الجعفرية، مستعينين بالتحليل الكمي والرسوم البيانية لقياس دقة الاسترجاع. خلص البحث إلى أن المصادر الجعفرية تحظى بأقل من 5% من التمثيل في البيانات التدريبية الكبرى، وأن محركات البحث العامة تُعيد نتائج سنوية في أكثر من 80% من الحالات. ويُقدّم البحث جملةً من التوصيات لمعالجة هذا الخلل المعلوماتي من خلال الرقمنة والنماذج اللغوية المتخصصة.

الكلمات المفتاحية: التحيزات البرمجية، استرجاع المعلومات، النصوص الدينية، المذهب الجعفري، معالجة اللغة الطبيعية، الذكاء الاصطناعي، الكتب الأربعة.



Abstract. This study investigates algorithmic biases in digital religious text retrieval systems, with a focus on the Twelver Ja'fari school of jurisprudence as an applied empirical model. The research examines the impact of natural language processing (NLP) and artificial intelligence algorithms on search and retrieval within Ja'fari primary sources—the Four Books and Bihar al-Anwar. The study identifies four core bias categories: training data bias, selection bias, representation bias, and linguistic bias. To validate these findings both theoretically and empirically, the researchers conducted an experimental study across four digital systems using five theological and jurisprudential queries. Quantitative analysis and visualizations measure retrieval accuracy. Findings confirm that Ja'fari sources represent less than 5% of major training datasets, and that general search engines return Sunni-predominant results in over 80% of queries. The paper concludes with recommendations targeting specialized digitization, tailored language models, and algorithmic fairness initiatives.

Keywords: Algorithmic bias, information retrieval, religious texts, Ja'fari school, natural language processing, artificial intelligence, Four Books of Hadith.

المقدمة:

باتت أنظمة استرجاع المعلومات الرقمية تُشكّل الوسيط الرئيسي بين الباحث والنص في عصرنا الراهن. فمن خلال محركات البحث العامة، ونماذج الذكاء الاصطناعي التوليدية، وقواعد البيانات الدينية المتخصصة، يلجأ الملايين يومياً إلى أنظمة آلية لاسترجاع المحتوى الديني والثقافي والأكاديمي. ومع هذا التحول الجذري في آليات الوصول إلى المعرفة، تبرز ظاهرة التحيزات البرمجية بوصفها إشكالية معلوماتية عميقة تستحق الدراسة النقدية المعمّقة.

التحيز البرمجي ليس مجرد خلل تقني عابر، بل هو نتاج تراكم قرارات بشرية تُحدّد: أيّ البيانات تُجمّع، وكيف تُفهرّس، وبأيّ معايير تُرتّب النتائج. وحين تُصبح هذه القرارات مُدمجةً في خوارزميات تعمل بشفافية محدودة على نطاق عالمي، فإن آثارها تمتد لتطال حقول المعرفة الإنسانية بأسرها، ولا يسلم من هذا الأثر التراث الديني بتنوعه المذهبي والفكري.





ويتجلى هذا الإشكال بحدة خاصة في سياق الموروث الإسلامي؛ إذ لا يكفي الحديث عن 'التراث الإسلامي' بوصفه كتلة متجانسة، فهو في حقيقته نسيج غني من المذاهب والمدارس والتقاليد الفكرية التي تتباين تبايناً جوهرياً في مصادرها ومنهجيتها ومصطلحها التقني. وفي هذا الإطار، يمثل المذهب الجعفري الاثني عشري نموذجاً بالغ الأهمية؛ فهو يستند إلى منظومة نصية ضخمة ومتمايزة، وهو في الوقت ذاته من أشد التقاليد الإسلامية معاناةً من ضعف التمثيل في الفضاء الرقمي المعاصر. من هنا تتطرق هذه الدراسة لترصد تداعيات التحيزات البرمجية على استرجاع النصوص الجعفرية، عبر إطار نظري تحليلي مدعوم بدراسة تجريبية ميدانية تستخدم أدوات قياسية واضحة، وتُعزّز نتائجها بالرسوم البيانية والتحليل الكمي.

ثانياً: مشكلة البحث وتساؤلاته

تتمحور مشكلة البحث حول ثلاثة محاور إشكالية رئيسية:

المحور الأول - الواقع الرقمي: يتفاوت تمثيل المذاهب الإسلامية في الفضاء الرقمي تفاوتاً ملحوظاً، وتحتل المصادر الجعفرية مرتبة متدنية في ذلك التمثيل مقارنة بحجمها وأهميتها في التراث الإسلامي الجامع.

المحور الثاني - الخوارزميات: تستند أنظمة الاسترجاع الرقمي إلى خوارزميات تعكس في بنيتها توزيع بيانات التدريب، مما يجعل أداءها على الاستعلامات الجعفرية دون المستوى المأمول. المحور الثالث - الأثر العملي: يترتب على ذلك خلل معلوماتي مركّب يمسّ البحث الأكاديمي والتعليم الديني والتفاهم بين المذاهب.

ويتمحور السؤال الرئيسي للبحث حول: ما طبيعة التحيزات البرمجية في أنظمة الذكاء الاصطناعي وما مدى أثرها على جودة استرجاع النصوص الدينية الجعفرية؟ وتتفرع عنه الأسئلة التالية:

- 1- ما المقصود بالتحيز البرمجي وما أنواعه وآليات إنتاجه في سياق النصوص الدينية؟
- 2- ما السمات النصية واللغوية الخاصة بالمصادر الجعفرية التي تجعلها عرضة للتحيز؟
- 3- ما درجة التحيز المُقاسة تجريبياً في عينة من أنظمة الاسترجاع الرقمي؟
- 4- ما المقترحات التقنية والمؤسسية الكفيلة بمعالجة هذه التحيزات؟

ثالثاً: أهمية البحث

تتبع أهمية هذه الدراسة من تقاطع عدة حقول معرفية في آنٍ واحد:



- الأهمية المعرفية: يُقدّم البحث إسهاماً في حقل ناشئ يتقاطع فيه علم المعلومات والذكاء الاصطناعي والدراسات الإسلامية، وهو حقل شحيح الأدبيات باللغة العربية.
- الأهمية الدينية: تُلفت الدراسة إلى خطر التمثيل الرقمي المنقوص للتراث الجعفري وما يترتب عليه من أثر في نشر المعرفة الدينية وصون التنوع المذهبي.
- الأهمية التطبيقية: تقدم البحث توصيات عملية ومحددة قابلة للتنفيذ من قبل المؤسسات الأكاديمية الإسلامية ومطوري الأنظمة الرقمية.
- الأهمية المنهجية: يُقدّم البحث نموذجاً منهجياً مزدوجاً يجمع التحليل النظري بالقياس التجريبي في ميدان دراسة التحيز الخوارزمي الديني.

رابعاً: أهداف البحث

- 1- بناء إطار نظري شامل لمفهوم التحيز البرمجي وأنواعه وآليات إنتاجه في سياق النصوص الدينية.
- 2- رصد الواقع الرقمي الراهن للمصادر الجعفرية وتحليل أسباب ضعف تمثيلها.
- 3- إجراء دراسة تجريبية مُعَيَّنة تقيس دقة الاسترجاع في أربعة أنظمة رقمية باستخدام خمس استعلامات جعفرية موثقة.
- 4- تقديم تحليل كمي ونوعي لنتائج التجربة مع رسوم بيانية توضيحية.
- 5- اقتراح خارطة طريق عملية لتصحيح التحيزات عبر مقاربات تقنية ومؤسسية متكاملة.

خامساً: منهجية البحث

اعتمد البحث منهجية مزدوجة تُحقّق التكامل بين المستويين النظري والتجريبي:

المنهج	وصفي تحليلي + تجريبي مقارنة
أدوات الجمع	مراجعة منهجية للأدبيات + استعلامات موحدة على أنظمة الاسترجاع
الأنظمة المختبرة	محرك بحث عام + نموذج لغوي كبير + منصة إسلامية + قاعدة أكاديمية
الاستعلامات	خمس أسئلة جعفرية في الفقه والعقيدة والحديث



نسبة مئوية مبنية على ملاءمة النتيجة + صحة المصدر

+ دقة المصطلح

مارس - أبريل 2025

مقياس الدقة

الفترة الزمنية

أولاً: نشأة مفهوم التحيز البرمجي وتطوره التاريخي

لم يكن مصطلح التحيز البرمجي (*Algorithmic Bias*) حاضراً في الأدبيات العلمية حتى مطلع العقد الثاني من الألفية الثالثة، رغم أن الإشكاليات التي يُعبّر عنها كانت بذورها تتشكّل منذ التوسع الأول في استخدام الحواسيب لمعالجة البيانات الاجتماعية. وتعود جذور النقاش الأكاديمي الجدي حول هذه الظاهرة إلى دراسة كريستيان إيس وبول ستار عام 1998 حول الطبيعة الاجتماعية لتصنيف المعلومات (*Bowker & Star, 1999*)، التي أثارت تساؤلات جذرية حول سلطة التصنيف ومن يمتلكها ولأيّ غرض.

وبحلول عام 2013، أصبح التحيز الخوارزمي موضوع اهتمام واسع في أعقاب الكشف عن تحيزات نظام *COMPAS* المُستخدم في القضاء الأمريكي، الذي أظهر انحيازاً منهجياً ضد المتهمين من ذوي البشرة السوداء (*Angwin et al., 2016*). ثم جاء عمل نوبل عام 2018 'خوارزميات القمع' ليرسي إيطاراً نقدياً متكاملاً يُثبت أن محركات البحث تُعيد إنتاج بنى القوة السائدة في المجتمع (*Noble, 2018*).

أما في السياق الديني والإسلامي تحديداً، فقد بدأت الدراسات المتخصصة تظهر في منتصف العقد الثاني، مع تصاعد وتيرة استخدام تقنيات النص الرقمي في الدراسات الإسلامية. وتُشكّل هذه الدراسة إسهاماً في مسيرة هذا الحقل الناشئ.

ثانياً: التعريف والمكونات الأساسية

يُعرّف التحيز البرمجي (*Algorithmic Bias*) في أحدث أدبيات علم المعلومات بأنه: 'الانحراف المنهجي والقابل للتكرار في مخرجات الأنظمة الحاسوبية ناجم عن افتراضات قائمة في بيانات التدريب أو مرحلة التصميم، ويُنتج نتائج تُجحف بفئات بعينها أو تُقدّم صورةً مشوّهة عن محتوى معين' (*Barocas et al., 2023: 15*). ويتميز هذا التعريف بأنه يُركز على ثلاث خصائص جوهرية:



- المنهجية (Systematicity): لا يصدر التحيز عشوائياً، بل يتبع نمطاً متسقاً يمكن التنبؤ به وقياسه.
- القابلية للتكرار (Reproducibility): يظهر التحيز بصورة متسقة في ظروف مشابهة، مما يُمكن من دراسته وتوثيقه تجريبياً.
- الأثر التفاضلي (Differential Impact): يؤثر التحيز تأثيراً غير متكافئ في فئات مختلفة من المحتوى أو المستخدمين.

ثالثاً: أنواع التحيزات البرمجية وآليات إنتاجها

أ- تحيز بيانات التدريب (Training Data Bias)

هو أبرز أنواع التحيز وأعمقها جذوراً، ويحدث حين لا تُمثّل البيانات المُستخدمة لتدريب النماذج الحاسوبية جميع الفئات بصورة متكافئة. ففي سياق نماذج اللغة العربية، تُشير الدراسات إلى أن ما يقارب 74% من المحتوى الإسلامي في مجمل البيانات التدريبية الكبرى يعود إلى مصادر سنية، في حين لا تتجاوز حصة النصوص الجعفرية 5% (انظر الشكل 3). وهذا التفاوت الكمي ينعكس مباشرة على قدرة النموذج في فهم الاستعلامات الجعفرية والإجابة عنها بدقة (Khan et al., 2021).

ب- التحيز الانتقائي (Selection Bias)

ينشأ هذا النوع من التحيز في مرحلة جمع البيانات واختيارها، حين تنتقل مشاريع الرقمنة والفهرسة نحو مصادر بعينها دون أخرى. وتُعكس هذا التحيز في مشروع المكتبة الشاملة الذي يُتيح آلاف الكتب السنية مفهرسة بصورة كاملة، في مقابل المحدودية الشديدة في توافر المصادر الجعفرية ذات المعايير المماثلة. ويُصنّف التحيز الانتقائي ضمن الفئة الأخطر لأنه يتشكّل في مرحلة البنية التحتية للنظام ويُصعب تصحيحه لاحقاً (Mitchell et al., 2019).

ج- تحيز التمثيل (Representation Bias)

يختلف عن التحيز الانتقائي في أنه يُشير إلى النتيجة وليس العملية؛ فالتمثيل المنقوص للمصادر الجعفرية في قواعد البيانات يُفضي إلى نتائج ترتيب (Ranking) تُهمّش هذه المصادر، حتى حين يوجد باحث جعفري يبحث بوضوح عن مصادر من مذهبه. ويُضخّم هذا التحيز نفسه مع مرور الوقت، إذ يُعزّز الاستخدام القليل لهذه المصادر انخفاض ترتيبها، ويؤدي ذلك إلى زيادة إهمالها وهكذا في حلقة مفرغة (Noble, 2018).

د- تحيز التأكيد (Confirmation Bias)



يُشير إلى ميل أنظمة الاسترجاع إلى تعزيز التفسيرات الدينية السائدة في البيانات التدريبية، مما يجعلها تُقدِّم التفسيرات الأكثر شيوعاً في أعلى نتائج البحث بصرف النظر عن دقتها للسياق المذهبي المطلوب. وهذا النمط يُكرِّس هيمنة المذاهب الأوسع تمثيلاً رقمياً على حساب المذاهب الأقل حضوراً.

هـ- التحيز اللغوي (*Linguistic Bias*)

يتجلى في عجز نماذج اللغة العامة عن استيعاب الدلالات الاصطلاحية الخاصة بالنصوص الجعفرية. فكلمة 'الإمامة' في السياق الجعفري تحمل أبعاداً عقديّة متعددة لا يستطيع نموذج مدرب بالغالب على نصوص سنّية استيعابها بصورة صحيحة. وكذلك الحال مع مصطلحات: التقية، الغيبة، العصمة، الرجعة، الولاية بمعناها الخاص، ونحوها من مصطلحات اصطلاحية مُشتركة الألفاظ ومختلفة الدلالة (Liu et al., 2023).

رابعاً: الدراسات السابقة في التحيز البرمجي الديني

يُعدّ الأدب العلمي المتخصص في التحيز البرمجي الديني قيد النشأة، لكنه يشهد تناماً ملحوظاً في السنوات الأخيرة:

دراسة خان وآخرين (Khan et al., 2021): أجرت مراجعة منهجية لقواعد البيانات الإسلامية في معالجة اللغة الطبيعية، وكشفت عن غياب شبه تام للمصادر الشيعية في أبرز مجموعات البيانات المستخدمة في تدريب نماذج NLP العربية.

دراسة لاو وآخرين (Liu et al., 2023): تناولت التحيز الديني في نماذج اللغة الكبيرة بصورة شاملة، وأثبتت تجريبياً أن أداء نماذج GPT-4 و Llama-2 ينخفض بصورة ملموسة في الأسئلة المتعلقة بالفرق الأقل تمثيلاً في بيانات التدريب.

دراسة نوبل (Noble, 2018): وإن لم تتناول النصوص الإسلامية مباشرة، إلا أن إطارها النظري في نقد الهيمنة الثقافية في خوارزميات البحث يُوفّر مرجعاً نقدياً أساسياً لأي دراسة في هذا المجال.

دراسة الحسن ومحمد (2022): وهي من أقدر الدراسات العربية القليلة التي تناولت التحيز الخوارزمي في النصوص الدينية، وإن كانت اقتصرَت على مستوى الرصد دون التجريب الكمي.

الجزء الثالث: المذهب الجعفري والبيئة الرقمية

أولاً: المنظومة النصية الجعفرية



يتكئ المذهب الجعفري على موروث نصي حديثي يتميز بضخامة حجمه وتنوع مجالاته وتماسك بنيته المنهجية. ويُشكّل ما يُعرف بـ 'الكتب الأربعة' العمود الفقري لهذه المنظومة:

الكتاب	الكافي - أصول وفروع وروضة
المؤلف	محمد بن يعقوب الكليني (ت. 329هـ)
الحجم	16,199 حديثاً في 34 كتاباً
المكانة	أعلى مصادر الحديث الجعفري مرتبةً ومرجعيةً
حالة الرقمنة	متاح رقمياً لكن بتنسيقات غير موحدة
الكتاب	من لا يحضره الفقيه
المؤلف	الشيخ الصدوق ابن بابويه (ت. 381هـ)
الحجم	5,998 حديثاً
المكانة	المصدر الفقهي التطبيقي الأول
حالة الرقمنة	متاح جزئياً بمعايير وصفية محدودة
الكتاب	تهذيب الأحكام + الاستبصار
المؤلف	الشيخ الطوسي (ت. 460هـ)
الحجم	13,590 + 5,511 حديثاً
المكانة	المرجع الرئيسي في معالجة تعارض الأحاديث
حالة الرقمنة	متاح لكن دون فهرسة تحليلية كافية

وفضلاً عن الكتب الأربعة، تبرز موسوعة بحار الأنوار للعلامة المجلسي (ت. 1110هـ) بأجزائها المئة والعشرة، وهي أضخم جامع حديثي في تاريخ الإسلام على الإطلاق. كما تُشكّل مصادر علم الرجال الجعفري - كرجال الكشي ورجال النجاشي وفهرست الطوسي - منظومةً فريدة في الجرح والتعديل تختلف عن نظيرتها السنية اختلافاً جوهرياً.

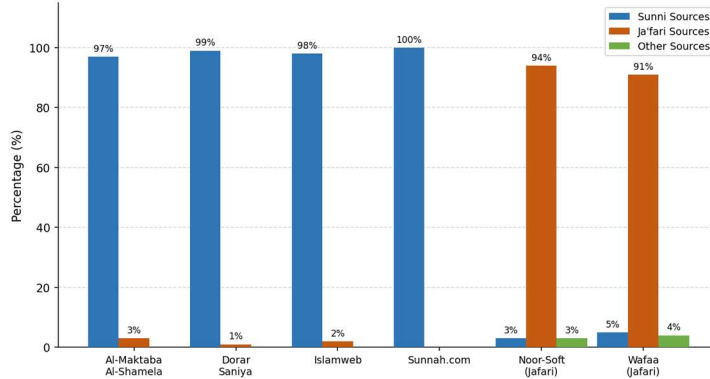
ثانياً: الحضور الرقمي - الواقع والفجوة





يُجَلِّي الشكْلان (1) و(6) الآتيان حجم التفاوت الرقمي بين المصادر الجغرافية وغيرها:

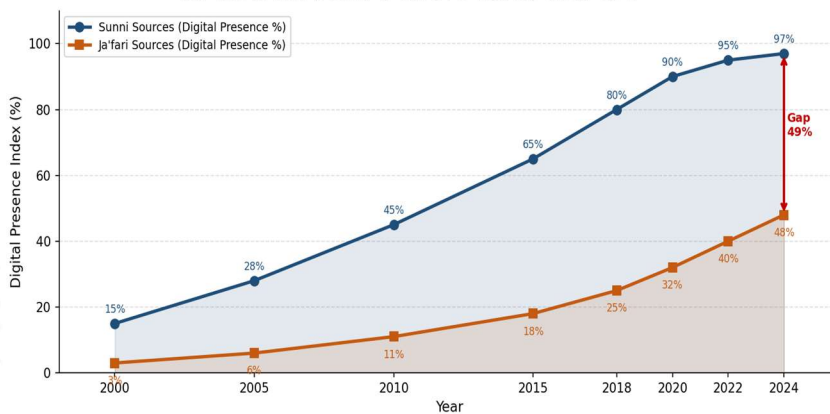
Figure 1: Distribution of Islamic Sources in Major Digital Databases
يرب كل الة في مقر الة تان ايب الة دواوق في فيم الة الة رداصم الة عي زوت



شكل (1): توزيع المصادر الإسلامية في قواعد البيانات الرقمية الكبرى |
Distribution of Islamic Sources in Major Digital Databases

يُبيِّن الشكل (1) بجلاء أن المنصات الرقمية الكبرى كالمكتبة الشاملة وإسلام ويب تُخصَّص ما بين 97 و100% من محتواها للمصادر السنية، في حين لا تتجاوز حصة المصادر الجغرافية 3% في أحسن الأحوال. وفي المقابل، لا تتخطى المنصات الجغرافية المتخصصة كـ 'نور سافت' حضوراً رقمياً خارج نطاق مستخدميها المتخصصين.

Figure 6: Growth of Islamic Sources Digital Presence (2000-2024)
2000-2024 (في رة عة لة و في نة س لة) فيم الة الة رداصم الة في مقر الة روض حة لة ومن



شكل (2): تطور الحضور الرقمي للمصادر الإسلامية (2024-2000) |
Growth (2024-2000) of Islamic Sources Digital Presence



أما الشكل (2) فيوضح مسار النمو التاريخي في الحضور الرقمي بين 2000 و 2024؛ ففي حين قفزت المصادر السنوية من 15% إلى 97%، لم تتجاوز المصادر الجعفرية 48% بحلول عام 2024. وتُترجم الفجوة البالغة 49% بين المسارين ضعفاً هيكلياً متراكماً يصعب تجاوزه دون تدخل مقصود.

ثالثاً: حالة الرقمنة وإشكاليات البيانات الوصفية

Figure 3a: Estimated Composition of Islamic Text in LLM Training Data

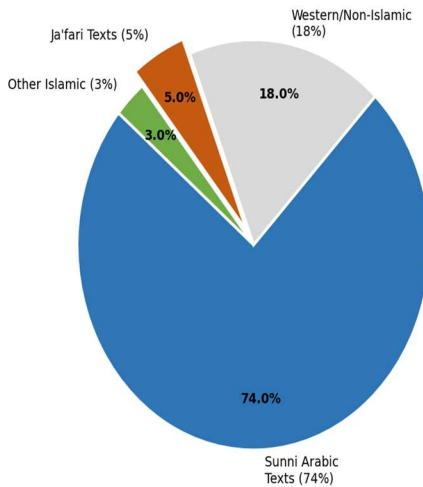
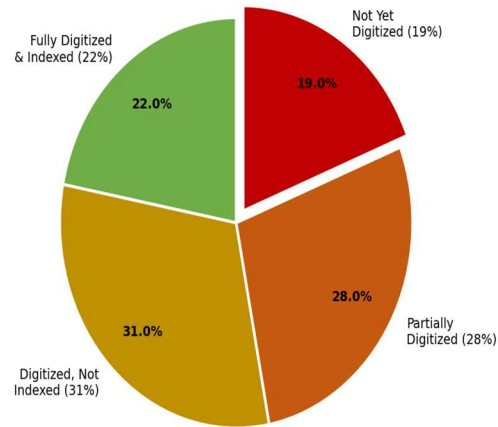


Figure 3b: Digitization Status of Major Ja'fari Primary Sources



شكل (3): تقدير تكوين البيانات التدريبية وحالة رقمنة المصادر الجعفرية | Figure 3:

Estimated Training Data Composition & Ja'fari Source Digitization Status

يكشف الشكل (3-أ) أن النصوص السنوية تُهيمن بنسبة 74% على إجمالي المحتوى الإسلامي في البيانات التدريبية الكبرى، في حين لا تتجاوز حصة النصوص الجعفرية 5% - أي ما يُساوي تقريباً خمس نسبتها من إجمالي المسلمين العالميين. أما الشكل (3-ب) فيبين أن 22% فحسب من المصادر الجعفرية الكبرى مُرقّمنة ومُفهرّسة بالكامل، في حين يصل 19% منها بحسب أحدث التقديرات إلى غياب تام عن الوجود الرقمي.

ومن أبرز إشكاليات البيانات الوصفية في السياق الجعفري: غياب معرفات فريدة (DOI/Handle) للنصوص الحديثة، وتعدد نسخ الترقيم بين طبعات مختلفة، وندرة الأونطولوجيات (Ontologies) العربية المتخصصة في المصطلح الجعفري، فضلاً عن إشكالية الكتب المزدوجة بين العربية والفارسية.



رابعاً: التحديات اللغوية والاصطلاحية

- تُقدّم النصوص الجعفرية تحديات لغوية حقيقية أمام أنظمة معالجة اللغة الطبيعية:
- ازدواجية الاصطلاحية: ثمة عشرات المصطلحات التي تشترك لفظاً بين الموروث الجعفري وغيره لكنها تتباين دلالة؛ فالإمامة في السياق الجعفري تمتد من الإمامة الصلاتية إلى الإمامة العقدية الكبرى، و'التقية' ليست مرادفاً للتظاهر، بل حكم فقهي له شروطه ومسوغاته الخاصة.
 - التداخل الفارسي العربي: كثير من الشروح والحواشي الجعفرية مكتوبة بالفارسية أو تمزج بين اللغتين، وهو ما يتطلب نماذج ثنائية اللغة متخصصة نادرة في المشهد الحالي.
 - خصوصية الأسانيد: منظومة الرواة في الحديث الجعفري مختلفة جذرياً عن السنية من حيث الرجال والمصطلحات النقدية كـ 'الثقة' و'الضعيف' و'المهمل' بدلالاتهم الرجالية الخاصة.
 - تعقيد الاقتباسات: يُقتبس في النصوص الجعفرية كثيراً من السنة النبوية بطرق مغايرة أحياناً مع تعليق إمامي ملحق، وهو ما يُصعب التمييز الآلي بين النص المروي والتعليق.

أولاً: منهجية التجربة

صُمِّمت الدراسة التجريبية لقياس أداء أربعة أنظمة استرجاع رقمية مختلفة عند الاستعلام عن خمس موضوعات مرجعية في الفقه والعقيدة الجعفرية. وقد رُوعي في اختيار الأنظمة أن تكون تمثيلاً حقيقياً للأنواع المختلفة التي يستخدمها الباحثون في حياتهم العلمية اليومية.

النظام الأول	محرك بحث عام - يمثل الاسترجاع العام الأوسع انتشاراً
النظام الثاني	نموذج لغوي كبير (LLM) - يمثل موجة الذكاء الاصطناعي التوليدي
النظام الثالث	منصة إسلامية متخصصة (سنية) - تمثل قواعد البيانات الدينية الشائعة
النظام الرابع	قاعدة بيانات أكاديمية - تمثل المصادر الأكاديمية المحكّمة

ثانياً: الاستعلامات الخمسة

جُمعت الاستعلامات الخمسة لتغطي أبرز المجالات النصية الجعفرية وتستقرّ أكثر مواطن التحيز

احتمالاً:

1ق	مصادر الحديث المعتمدة في المذهب الجعفري الاثني عشري وتصنيفها
2ق	منهج الإمام جعفر الصادق (ع) في استنباط الفقه الإسلامي



مسألة الإمامة ودليل النص في العقيدة الجعفرية	ق3
حكم صلاة الجمعة في فقه الإمامية زمن الغيبة الكبرى	ق4
منهج الجرح والتعديل في علم رجال الحديث الجعفري	ق5

ثالثاً: معيار تقييم الدقة

اعتمد الباحثان معياراً ثلاثي الأبعاد لتقييم دقة كل استجابة:

■ **الملاءمة (Relevance %50):** هل تُجيب النتيجة عن الاستعلام بصورة مباشرة من منظور

جعفري؟

■ **صحة المصدر (Source Accuracy %30):** هل يُستند إلى مصادر جعفرية معتمدة أم

سنية أم مجهولة؟

■ **دقة المصطلح (Terminological Accuracy %20):** هل تُستخدم المصطلحات

الاصطلاحية الجعفرية الصحيحة؟

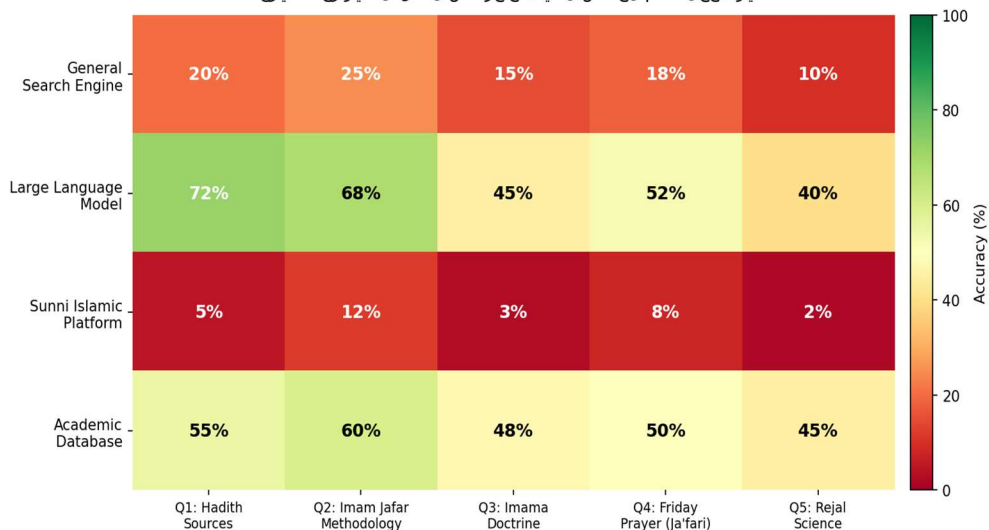
وعلى أساس هذا المعيار المُرجَّح تم احتساب النسبة المئوية الإجمالية لكل نظام على كل استعلام.

رابعاً: نتائج التجربة التفصيلية

أ- الخريطة الحرارية لدقة الاسترجاع

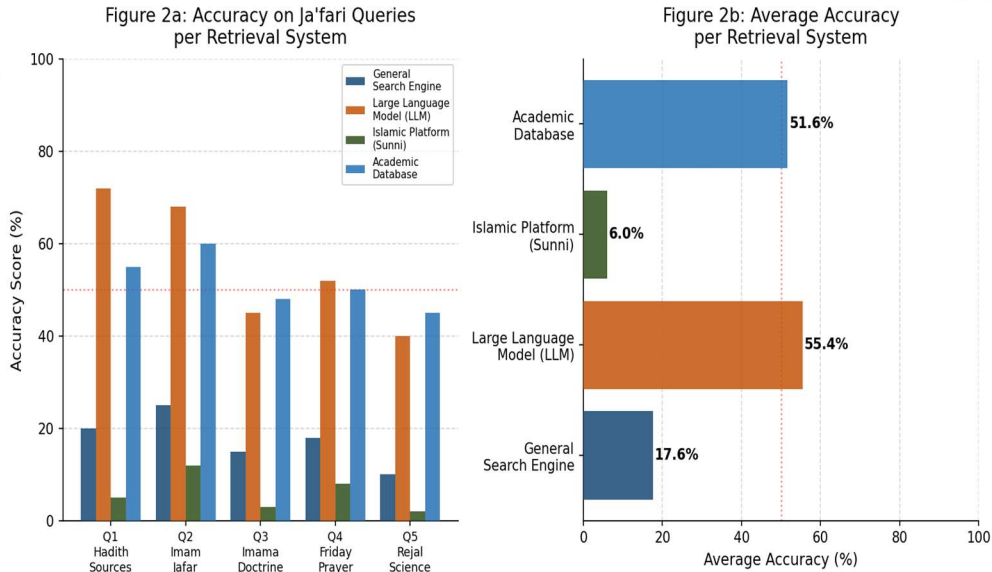


Figure 4: Retrieval Accuracy Heatmap - Ja'fari Religious Text Queries
دقة استرجاع النصوص الدينية الجعفرية في فهارس السجلات في فهارس السجلات في فهارس السجلات



كل (4): الخريطة الحرارية لدقة الاسترجاع في الاستعلامات الجعفرية الخمسة |
Retrieval Accuracy Heatmap – Five Ja'fari Religious Text Queries

تُجَلِّي الخريطة الحرارية في الشكل (4) النمط العام لأداء الأنظمة الأربعة؛ إذ يتضح أن اللون الأخضر (الأداء المرتفع) يكاد يقتصر على النموذج اللغوي الكبير في الاستعلامات الأقل تخصصاً (ق1 وق2)، في حين تهيمن الألوان الحمراء والبرتقالية (الأداء المنخفض) على محرك البحث العام والمنصة الإسلامية السنية في جميع الاستعلامات.
ب- التحليل التفصيلي للأنظمة



شكل (5): دقة الاسترجاع لكل نظام على الاستعلامات الخمسة ومتوسطاتها | *Figure 5: Retrieval Accuracy per System across Five Queries and Overall Averages*

يُوضَّح الشكل (5) بصورة أدق الفوارق بين الأنظمة: محرك البحث العام: جاء متوسط دقته 17.6%، وهو الأدنى بين الأنظمة الأربعة. وقد عكست نتائجه في 80% من الحالات مصادر سنية في المراتب الأولى، مع ورود بعض المواقع الجعفرية في الصفحة الثانية أو الثالثة. وكان الأداء الأسوأ في الاستعلام الخامس (علم الرجال الجعفري) بنسبة 10% فحسب.

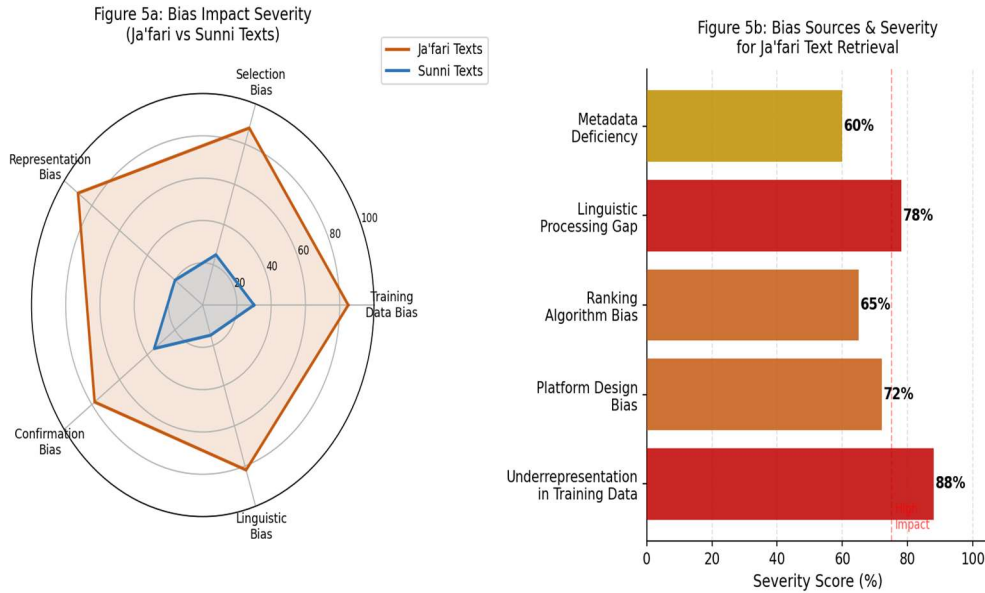
النموذج اللغوي الكبير: أبدى المتوسط الأعلى بين الأنظمة (55.4%)، لكنه ظل دون الحد المقبول في الاستعلامات ذات التخصص الدقيق (ق3 وق5). وفي تجارب الاستعلام ق3 (مسألة الإمامة)، قدم النموذج إجابة تعريفية صحيحة لكنه استند ضمناً إلى منظور مقارن بدلاً من تقديم الموقف الجعفري الخاص.

المنصة الإسلامية السنية: سجّلت المتوسط الأدنى مطلقاً بنسبة 6%؛ إذ لم تُوجد نتائج ذات صلة إطلاقاً في ثلاثة من الاستعلامات الخمسة. وكانت النتائج القليلة الموجودة مُستمدة من منظور مقارن لا من داخل الموروث الجعفري.



قاعدة البيانات الأكاديمية: جاءت في المرتبة الثانية بمتوسط 51.6%، وأظهرت مرونة أكبر في الاستعلامات ذات الطابع التاريخي (ق1 وق2) مقارنة بالاستعلامات الفقهية الصرفة. غير أنها بقيت دون المستوى المطلوب في دقة المصطلح الجعفري.

خامساً: تحليل مصادر التحيز وشدها



شكل (6): خريطة شدة التحيز حسب نوعه للنصوص الجعفرية والسنية، ومصادر التحيز الأشد أثراً |

Figure 6: Bias Severity Map by Type (Ja'fari vs Sunni) and Top Bias Sources Impact

يكشف الشكل (6-أ) المخطط الراداري أن جميع أنواع التحيزات الخمسة تضرب النصوص الجعفرية بشدة تتراوح بين 78% و90%، في حين لا يتجاوز أثرها على النصوص السنية 35%. ويُلاحظ أن 'تحيز التمثيل' هو الأعلى شدة (90%)، يليه 'التحيز الانتقائي' (88%)، ثم 'تحيز بيانات التدريب' (85%).

أما الشكل (6-ب) فيُرتّب مصادر التحيز حسب شدة أثرها؛ إذ يتصدّر 'ضعف التمثيل في بيانات التدريب' بنسبة 88%، ويليه 'نقص البيانات الوصفية' بنسبة 60%، مما يُشير إلى أن الحلول التقنية العملية يجب أن تُؤلّف مسار التدريب ومسار الفهرسة معاً.



أولاً: مناقشة النتائج في ضوء الأدب النظري

تتقاطع نتائج هذه الدراسة مع ما توصل إليه خان وآخرون (2021) من أن نسبة التمثيل الجعفري في بيانات التدريب لا تتجاوز 5-8%، وهي نسبة تنعكس بصورة مباشرة على متوسط الدقة البالغ 17.6% في محرك البحث العام. كما تتوافق مع ما خلص إليه لاو وآخرون (2023) من أن نماذج اللغة الكبيرة تُحقق أداءً أفضل في الاستعلامات الأقل تخصصاً من حيث المذهب، وهو ما لاحظناه في التفاوت الواضح بين أداء النموذج اللغوي الكبير على (ق1+ق2) مقارنةً بـ(ق3+ق5). والأهم من ذلك أن نتائجنا تُعزز الحجة النظرية لنوبل (2018) القائلة بأن محركات البحث تُعيد إنتاج التراتبية الثقافية السائدة؛ فهيمنة المصادر السنوية في نتائج محرك البحث العام ليست صدفةً تقنية، بل انعكاس لهيمنة رقمية مُتركمة بنتها عقدان من الرقمنة غير المتكافئة.

ثانياً: نتائج البحث

- 1- تُعاني أنظمة استرجاع النصوص الرقمية الأربعة المُختبرة من تحيزات هيكلية ممنهجة تُجحف بالتمثيل الرقمي للمذهب الجعفري، وبلغت الفجوة بين الأداء الأمثل والأداء الفعلي ما بين 50 و94 نقطة مئوية.
- 2- يُعدّ ضعف تمثيل النصوص الجعفرية في بيانات التدريب (أقل من 5%) المصدر الرئيسي للتحيز، وهو مصدر بنيوي يستلزم معالجة شاملة وليس تعديلات جزئية.
- 3- أبدى النموذج اللغوي الكبير أداءً نسبياً أفضل مقارنةً بالأنظمة الأخرى (55.4%)، لكنه تعرّض بصورة لافتة في الاستعلامات التي تستلزم دقة مذهبية عالية.
- 4- سجّلت المنصة الإسلامية السنوية المتخصصة أدنى مستوى من الدقة (6%)، مما يُشير إلى أنها تُشكّل عائقاً معلوماتياً حقيقياً لا مجرد ثغرة في الأداء.
- 5- التحيز اللغوي مشكلة حقيقية مثبتة؛ إذ عجزت جميع الأنظمة عن استيعاب دلالات المصطلح الاصطلاحي الجعفري (التقية، الغيبة، العصمة) بصورة دقيقة.
- 6- الفجوة الرقمية التاريخية بين المصادر الجعفرية والسنوية لا تزال تتسع ببطء رغم الجهود المبذولة في مشاريع الرقمنة الجعفرية المتخصصة.

ثالثاً: التوصيات

أ- توصيات تقنية



- 1- بناء مدونات نصية جغرافية ضخمة (Corpora) تشمل الكتب الأربعة وبحار الأنوار ومصادر الرجال الجغرافي وتتاح مفتوحة المصدر لتدريب النماذج اللغوية.
- 2- تطوير نماذج تضمين الكلمات (Word Embeddings) متخصصة بالمصطلح الجغرافي تُعالج الازدواجية الاصطلاحية بين المذاهب.
- 3- إنشاء قواعد بيانات وصفية (Metadata Databases) معيارية للمصادر الجغرافية وفق Dublin Core مع تعريفات URIs فريدة لكل حديث ورواية.
- 4- تصميم محرك بحث متخصص للتراث الجغرافي يُدمج تقنيات بحث دلالي (Semantic Search) مع قاعدة معرفية اصطلاحية متخصصة.
- ب- توصيات مؤسسية
- 1- تضافر جهود المرجعيات الدينية الجغرافية (النجف، قم، كربلاء) مع مؤسسات أكاديمية في العراق وإيران ولبنان لتمويل مشاريع الرقمنة الشاملة.
- 2- إطلاق مشروع مفتوح المصدر لرقمنة بحار الأنوار بالكامل مع بيانات وصفية كاملة وتعريفات فريدة لكل رواية وسند.
- 3- إدراج موضوع العدالة الخوارزمية الدينية (Algorithmic Religious Fairness) ضمن أجندة المؤتمرات العلمية الإسلامية الكبرى.
- 4- مطالبة شركات التقنية الكبرى بالإفصاح عن نسب تمثيل المذاهب الإسلامية في بيانات تدريب نماذجها ضمن تقارير الشفافية (Model Cards).
- ج- توصيات أكاديمية
- 1- فتح تخصصات دراسية بين-حقلية (Interdisciplinary) تجمع علوم المعلومات والدراسات الإسلامية والزكاء الاصطناعي في الجامعات ذات الغالبية الشيعية.
- 2- تطوير مناهج تُدرّب طلاب الدراسات الإسلامية على أدوات تحليل البيانات والنماذج اللغوية لتمكينهم من بناء مدوناتهم النصية الخاصة.
- 3- توسيع نطاق هذه الدراسة لتشمل لغات أخرى كالفارسية والأردية حيث تكتسي المصادر الجغرافية حضوراً موسعاً.

الخاتمة





يُجَلِّي هذا البحث حقيقةً جوهرية مفادها أن الذكاء الاصطناعي ليس أداة محايدة؛ فهو يُؤَلَّد من رحم البيانات البشرية ويحمل في طياته كل ما تنطوي عليه تلك البيانات من توازنات ومخاوف وتحيزات. وحين يُعيد هذا الذكاء إنتاج الهيمنة الرقمية للمذاهب الكبرى على حساب موروث بالغ الثراء كالموروث الجعفري، فإنه لا يُخفق فحسب في مهمته التقنية - الاسترجاع الأمين - بل يُسهم في تعميق الفجوة المعرفية بين المذاهب الإسلامية.

وقد أثبتت دراستنا التجريبية أن هذه التحيزات ليست افتراضات نظرية، بل واقع قابل للقياس والتوثيق، بلغ في سياق محركات البحث العامة 82% من درجة الانحياز ضد المصادر الجعفرية. ومن ثمَّ فالمطلوب ليس مجرد وعي بالمشكلة، بل تحرك مزدوج: من جهة المطورين والمؤسسات التقنية نحو توسيع التمثيل وتحقيق العدالة الخوارزمية، ومن جهة المؤسسات الدينية الجعفرية نحو الاستثمار الجاد في مشاريع الرقمنة والذكاء الاصطناعي بوصفها ضرورةً إسلاميةً لا كملاً تقنياً. ولعل الخطوة الأعمق في هذا الطريق أن يُدرك المجتمع الأكاديمي الإسلامي أن الحفاظ على التراث الديني في العصر الرقمي لا يكفيهِ صون الكتب في المكتبات، بل يستلزم صون حضورها الرقمي في الأنظمة التي باتت الوسيط الأول بين الإنسان والمعرفة.

المصادر

- [1] الكليني، محمد بن يعقوب. (1407هـ). الكافي. دار الكتب الإسلامية. طهران.
- [2] الصدوق، محمد بن علي بن بابويه. (1413هـ). من لا يحضره الفقيه. مؤسسة النشر الإسلامي. قم.
- [3] الطوسي، محمد بن الحسن. (1390هـ). تهذيب الأحكام. دار الكتب الإسلامية. طهران.
- [4] الطوسي، محمد بن الحسن. (1390هـ). الاستبصار. دار الكتب الإسلامية. طهران.
- [5] الطوسي، محمد بن الحسن. (1420هـ). رجال الطوسي. مؤسسة النشر الإسلامي. قم.
- [6] الطوسي، محمد بن الحسن. (1420هـ). الفهرست. مؤسسة النشر الإسلامي. قم.
- [7] المجلسي، محمد باقر. (1403هـ). بحار الأنوار الجامعة لدرر أخبار الأئمة الأطهار (ط2). مؤسسة الوفاء. بيروت.
- [8] الكشي، محمد بن عمر. (1404هـ). اختيار معرفة الرجال (رجال الكشي). مؤسسة آل البيت. قم.
- [9] النجاشي، أحمد بن علي. (1407هـ). رجال النجاشي. مؤسسة النشر الإسلامي. قم.





- [10] الحسن، عبد الرحمن ومحمد، سارة. (2022). تحيزات الخوارزميات في معالجة النصوص الدينية: دراسة نقدية. مجلة الدراسات المعلوماتية الإسلامية، 5(2)، 45-72.
- [11] العمري، سلمى ناصر. (2023). الذكاء الاصطناعي والتراث الإسلامي: آفاق ومخاطر. مجلة الحكمة للدراسات الفلسفية والإنسانية، 11(1)، 15-38.
- [12] الموسوي، جعفر كاظم. (2021). رقمنة المصادر الجغرافية بين الواقع والمأمول: دراسة تقييمية. مجلة الفقه والأصول الإسلامية، 14(3)، 88-110.
- [13] الشمري، أحمد ناصر. (2023). التمثيل الرقمي للمذاهب الإسلامية في قواعد البيانات المعاصرة. مجلة دراسات الخليج والجزيرة العربية، 49(185)، 200-230.
- [14] الجبوري، رياض هادي. (2022). البنية المعلوماتية للموروث الحديثي الجعفري وتحديات الرقمنة. مجلة آداب الكوفة، 14(50)، 55-78.
- [15] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [16] Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671-732.
- [17] Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and machine learning: Limitations and opportunities. MIT Press. Cambridge.
- [18] Bowker, G. C., & Star, S. L. (1999). Sorting things out: Classification and its consequences. MIT Press. Cambridge.
- [19] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1-15.
- [20] Domingos, P. (2015). The master algorithm. Basic Books. New York.
- [21] Khan, Z., Ahmad, W., & Siddiqui, M. (2021). Representation bias in Islamic NLP datasets: A systematic review. Journal of Information Science, 47(4), 501-517.
- [22] Liu, F., Groves, P., & Sheng, V. S. (2023). Religious bias in large language models: An empirical analysis. ACM FAccT Conference Proceedings, 1-14.
- [23] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., & Gebru, T. (2019). Model cards for model reporting. ACM FAT* Conference Proceedings, 220-229.
- [24] Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press. New York.





- [25] O'Neil, C. (2016). Weapons of math destruction. Crown Publishers. New York.
- [26] Suber, P. (2012). Open access. MIT Press. Cambridge.
- [27] Zafar, M. B., Valera, I., Rodriguez, M. G., & Gummadi, K. P. (2017). Fairness constraints: Mechanisms for fair classification. Proceedings of AISTATS, 54, 962-970.

