Journal of Scientific and Human Dimensions
مجلة الأبعاد العلمية والإنسانية
Volume 2, Issue 2, 2026 Pages: 764-769
Journal homepage: https://jshd.com.ly/index.php/jshd/en

Artificial Intelligence and the Paper Writing Revolution: A Literature Analysis of the IMRAD Model in the Age of Large Language Models

Dr Abdulgader Alsharif *
Department of Electric and Electronic Engineering, Collage of Technical Science Sabha, Sabha, Libya

الذكاء الاصطناعي وثورة كتابة الورقة البحثية: تحليل مرجعي لنمط *IMRAD* في عصر النماذج اللغوية الكبيرة

د عبدالقادر الشريف *
قسم الهندسة الكهربائية والالكترونية، كلية العلوم التقنية سبها، سبها، ليبيا

*Corresponding author: alsharifutm@gmail.com

Received: March 25, 2026	Accepted: April 25, 2026	Published: June 04, 2026

Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract:
The academic landscape is undergoing a radical transformation due to the infiltration of artificial intelligence (AI) and large language models (LLMs), which are reshaping the mechanisms of scientific knowledge production, from literature exploration to manuscript writing. This article presents a specialized, systematic literature analysis examining the impact of this technological revolution on the structure of the paper according to the IMRAD model (Introduction, Methodology, Results, Discussion). Based on a meticulous review of more than 40 recent studies (2024-2026), the research presents an analytical framework divided into four main axes: (1) Intelligent tools for literature discovery and citation analysis, and their classification in detailed analytical tables; (2) Applications of LLMs in data analysis and extraction, with a focus on critiquing the phenomenon of academic "hallucination"; (3) The transformation of manuscript writing and review with the aid of intelligent agents (Multi-Agent Frameworks); (4) Ethical Dilemmas and Flaws in Artificial Content Detectors. The results reveal that large linguistic models demonstrate high efficiency in structured tasks (such as title and abstract scanning, where recall rates reach 100%, reducing workload by up to 97.3%). However, they remain fragile and unreliable in tasks requiring deep scientific judgment (such as extracting complex data, with recall rates ranging from 45.5% to 94.4%). The article concludes that responsible integration between humans and artificial intelligence, where these tools act as assistants to enhance capabilities and not as replacements for human expertise, is the only way to ensure the integrity of scientific research in the foreseeable future.

Keywords: Artificial Intelligence, Large Linguistic Models, IMRAD, Reference Analysis, Academic Ethics, Research Aids, Multi-Agent Systems**.**

الملخص
يشهد المشهد الأكاديمي تحولاً جذرياً بفعل تسلل تقنيات الذكاء الاصطناعي (AI) والنماذج اللغوية الكبيرة (LLMs)، التي تعيد تشكيل آليات إنتاج المعرفة العلمية، بدءاً من استكشاف الأدبيات وصولاً إلى صياغة المخطوطات. يقدم هذا المقال تحليلاً مرجعياً منهجياً متخصصاً، يبحث في تأثير هذه الثورة التكنولوجية على هيكل الورقة البحثية وفق نمط IMRAD (المقدمة، المنهجية، النتائج، المناقشة). بالاعتماد على مراجعة دقيقة لأكثر من 40 دراسة حديثة (2024-2026)، يقدم البحث إطاراً تحليلياً مقسماً إلى أربعة محاور رئيسية: (1) الأدوات الذكية لاكتشاف الأدبيات وتحليل الاستشهادات، وتصنيفها في جداول

تحليلية تفصيلية؛ (2) تطبيقات LLMs في تحليل البيانات واستخلاصها، مع التركيز على نقد ظاهرة "الهلوسة" الأكاديمية؛ (3) التحول في كتابة المخطوطات والمراجعة بمساعدة الوكلاء الذكيين (Multi-Agent Frameworks)؛ (4) المعضلات الأخلاقية ومكامن الخلل في كاشفات المحتوى الاصطناعي. تكشف النتائج أن النماذج اللغوية الكبيرة تُظهر كفاءة عالية في المهام المنظمة (مثل فحص العناوين والملخصات، حيث يصل معدل الاستدعاء إلى 100%، مع تقليل عبء العمل بنسبة تصل إلى 97.3%)، ومع ذلك، تظل هشة وغير موثوقة في المهام التي تتطلب حكماً علمياً عميقاً (مثل استخلاص البيانات المعقدة، بنسبة استدعاء تتراوح بين 45.5% و94.4%). يستنتج المقال أن التكامل المسؤول بين الإنسان والذكاء الاصطناعي، حيث تعمل هذه الأدوات كمساعدين لتعزيز القدرات وليس كبديلين عن الخبرة البشرية، هو السبيل الوحيد لضمان سلامة البحث العلمي في المستقبل المنظور.

الكلمات المفتاحية: الذكاء الاصطناعي، النماذج اللغوية الكبيرة، IMRAD، تحليل مرجعي، الأخلاقيات الأكاديمية، الأدوات المساعدة للبحث، أنظمة الوكلاء المتعددة (Multi-Agent Systems).

1. المقدمة

في أقل من عقدين، تحولت النماذج اللغوية الكبيرة (LLMs) من أدوات أكاديمية متخصصة إلى قوة محركة تعيد تشكيل البنى التحتية للمعرفة البشرية. منذ إطلاق ChatGPT في نوفمبر 2022، تسارعت وتيرة التكامل بين الذكاء الاصطناعي وسير العمل البحثي بشكل لم يسبق له مثيل (Kong et al., 2026). بينما يعترف الأكاديميون بإمكانات هذه التقنيات الهائلة في أتمتة استخلاص البيانات وتوليد المخطوطات وتلخيص كميات هائلة من الأدبيات، إلا أن المخاوف المتعلقة بالهلوسة (Hallucination)، والتحيز الخوارزمي، والأصالة الفكرية، وما وصفه بعض الباحثين بـ"أزمة النزاهة" تلوح في الأفق كتحديات وجودية (Adel & Alani, 2025).

منذ النصف الثاني من القرن العشرين، تم اعتماد نمط IMRAD كبنية معيارية للورقة العلمية التجريبية، وذلك لضمان الشفافية وإمكانية التكرار (Sollaci & Pereira, 2004). ومع ذلك، فإن إدخال الذكاء الاصطناعي في كل مرحلة من مراحل هذا النمط - من صياغة الفرضية (المقدمة) إلى تصميم البحث (المنهجية) إلى تحليل البيانات (النتائج) وتفسيرها (المناقشة) - يحمل وعوداً بثورة في الإنتاجية ومخاطر لا تقل خطورة على سلامة العلم. يهدف هذا التحليل المرجعي إلى معالجة سؤال رئيسي: كيف يعيد الذكاء الاصطناعي، وخاصة LLMs، تشكيل كتابة الورقة البحثية وفق نمط IMRAD في الفترة 2024-2026؟

2. المنهجية

اتبع هذا التحليل المنهجية المنظمة للمراجعات المرجعية التحليلية (Systematic Literature Review - SLR) وفق مبادئ PRISMA (Page et al., 2021)، المدمجة مع ممارسات التحليل الببليومتري الحديثة كما اقترحها Pereira et al. (2025).

- استراتيجية البحث: تم إجراء البحث بين أكتوبر 2024 ومايو 2026 في قواعد البيانات الرئيسية: Scopus، IEEE Xplore، PubMed، ومستودع arXiv. تم استخدام سلسلة بحث رئيسية: ("Artificial Intelligence" OR "Large Language Model" OR "ChatGPT" OR "Generative AI") AND ("Systematic Review" OR "Academic Writing" OR "Research Workflow" OR "IMRAD" OR "Scientific Integrity").
- معايير الاختيار والاستبعاد: شملت المعايير الدراسات المنشورة باللغة الإنجليزية بين 2024 و2026، والتي تقيم أدوات الذكاء الاصطناعي في البحث العلمي. تم استبعاد الدراسات النظرية البحتة أو تلك التي ركزت حصراً على جوانب تعليمية غير ذات صلة بهيكل الورقة البحثية. بعد فحص 340 ملخصاً، تم اختيار 43 دراسة رئيسية للتحليل الكامل.
- استخلاص البيانات وتصنيف الأدوات: تم تصنيف الأدوات المساعدة بواسطة الذكاء الاصطناعي وفق إطار عمل حديث من خمسة مجالات (فرز الأدلة الاستدلالي، أنظمة التفاعل مع المستندات، تحليل الأدبيات، مساعدي الكتابة، وحلول التصميم الجرافيكي وإدارة المراجع)، كما ورد في دراسات مثل تلك التي أجراها Bara & Oprea (2025)

3. النتائج والتحليل

ركز التحليل على تقييم أدوات الذكاء الاصطناعي عبر ثلاث مراحل أساسية تتماشى مع نمط IMRAD.

3.1. الأدوات الذكية لاكتشاف الأدبيات وتحليل الاستشهادات (مرحلة المقدمة)

تضمنت مرحلة تحليل الأدبيات واستخلاصها استخدام مجموعة واسعة من الأدوات المتخصصة، التي تختلف بشكل كبير في منهجياتها ودقتها. يقدم جدول 1 تحليلاً تفصيلياً مقارناً لهذه الأدوات بناءً على معايير معيارية شملت: مصدر البيانات، مرحلة المراجعة، أداة الذكاء الاصطناعي الأساسية، ونقاط القوة والضعف العملية.

جدول 1: تحليل مقارن للأدوات المدعومة بالذكاء الاصطناعي في المراجعات المنهجية

الأداة	المصدر / المرجع	المرحلة	تقنية الذكاء الاصطناعي	نقاط القوة الرئيسية	القيود والضعف
Elicit	(Lau & Golder, 2025; Hilkenmeier et al., 2025)	الفحص، استخلاص البيانات	LLMs (مملوكة)	دقة استخلاص 94-99%، تقليل الوقت بنسبة 80%	يعتمد على Semantic Scholar، قابلية التخصيص محدودة
Nested Knowledge	(Wong, 2025)	الفحص، التحليل	Robot Screener + LLMs	تخصيص عالي، قابلية للفحص المزدوج	يتطلب تدخل بشري كبير، حساسية معتدلة (72.7%)
ProfOlaf	(Afonso et al., 2026)	جمع المقالات، الفحص	LLMs + Snowballing	يدعم التجميع التكراري (snowballing)، مفتوح المصدر	لا يزال في مراحله الأولى نسبياً
ReviewAid	(Sahu & Balakrishnan, 2026)	الفحص القائم على PICO، الاستخلاص	OpenAI، Anthropic، DeepSeek إلخ.	يدعم مزودي خدمات متعددين، أنبوب تحليل مقاوم للأخطاء	يركز على الرعاية الصحية (PICO)، يحتاج لخبرة تقنية
Scite	(Research Solutions, 2026)	تحليل الاستشهادات	Smart Citations	يحلل 19 مليار استشهاد، يصنف الدعم/المعارضة للدراسات	أداة تحليل وليس فحص شامل
ResearchRabbit	(ResearchRabbit, 2025; Zenodo, 2025)	اكتشاف الأدبيات، رسم الخرائط	خوارزميات قائمة على الاستشهادات	رسوم بيانية تفاعلية، مجاني، تحليل شبكات التعاون	مجرد أداة اكتشاف، لا يقوم بالفحص أو الاستخلاص التلقائي
ISLE	(Al-Hassan et al., 2025)	الاستكشاف الشامل	تعلم آلة + استشهادات	تحليلات ببليومترية، معالجة النمو الهائل للمراجع (8-9% سنوياً)	أداة بحثية متقدمة، غير متاحة تجارياً بالكامل

أظهرت النتائج أن Elicit كان الأكثر دقة في استخلاص البيانات (تصل إلى 99% وفقاً لتقارير المطورين)، بينما تفوقت Nested Knowledge في السيناريوهات التي تتطلب تخصيصاً عالياً، وإن كان ذلك على حساب كفاءة الوقت (Wong, 2025).

3.2. تطبيقات النماذج اللغوية الكبيرة (LLMs) في تحليل البيانات (مرحلة النتائج)

ركزت الدراسات الحديثة على قدرة LLMs على أتمتة المهام المتكررة والمنظمة، وهو ما يتجلى بقوة في مرحلة **تحليل البيانات (النتائج). يلخص جدول 2 الأداء المقارن لأحدث النماذج في المراجعات الطبية وهندسة البرمجيات.

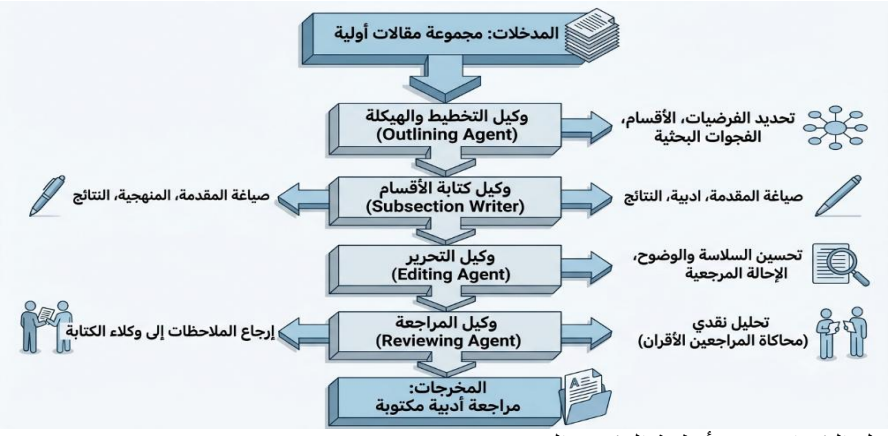
جدول 2: أداء LLMs في المهام الفرعية للمراجعة المنهجية

المصدر	الدقة / الحساسية / الاستدعاء	المهمة	النموذج
(Martins et al., 2026)	الدقة: 77% (حساسية 87%، خصوصية 70%)	الفحص النص الكامل-(Full Text)	**ChatGPT 5.0**
(Martins et al., 2026)	الدقة: 71% (كاپا 0.43)	الفحص النص الكامل	**ChatGPT 4.0**
(Alazba et al., 2026)	ثبات عال، قريب من تقييم الخبراء	صياغة استراتيجيات البحث، التقييم النوعي	**GPT-4o**
(Motika, 2026)	الاستدعاء: 71.6-100% (توفير عبء عمل 90-97%)	فحص العناوين والملخصات	**GPT-5-Nano**
(Motika, 2026)	الاستدعاء: 83.3-88.5%	فحص النص الكامل	**GPT-5-Nano**
(Motika, 2026)	الاستدعاء: 45.5-94.4% (تأثر بهيكل المستند)	استخلاص البيانات	**GPT-5-Nano**
(Human vs AI, 2025)	فشل كامل نسبياً في الكتابة النهائية (قصيرة وغير ملهمة)	مراجعة منهجية كاملة	**LLMs المتعددة**

تكشف البيانات وجود فجوة أداء كبيرة: بينما تتفوق النماذج في الفحص الأولي للمقالات (شاشة أولية عالية الحساسية/استدعاء)، فإن دقتها تنخفض بشكل ملحوظ في مهام الاستخلاص المعقدة التي تتطلب فهماً سياقياً دقيقاً للوثائق (Motika, 2026).

3.3. الأطر متعددة الوكلاء: إعادة تعريف المناقشة والكتابة

يمثل الجيل الجديد من أطر الوكلاء المتعددة (Multi-Agent Frameworks) نقلة نوعية في معالجة التحديات المرتبطة بمرحلتي المناقشة وكتابة المخطوطة**، وهي المجالات التي كانت أكثر مقاومة للأتمتة سابقاً (Go et al., 2026). يستخدم إطار *LiRA (Literature Review Agents) أربعة وكلاء متخصصين: (1) وكيل للتخطيط والهيكلة (Outlining)، (2) وكيل لكتابة الأقسام الفرعية، (3) وكيل للتحرير، و(4) وكيل للمراجعة. أظهر LiRA تفوقاً على خطوط الأساس السابقة (مثل AutoSurvey) من حيث جودة الكتابة ودقة الاستشهادات، مع الحفاظ على درجة تشابه تنافسية مع المراجعات البشرية (Go et al., 2026).



شكل 1: سير العمل التكراري في أنظمة الوكلاء المتعددة (Multi-Agent Frameworks)

يمثل هذا النظام نقلة نوعية من النماذج التوليدية التقليدية (LLMs) إلى **أنظمة وكيلة قادرة على التفكير النقدي وإعادة التنظيم الذاتي، مما يجعلها أكثر قدرة على محاكاة العملية البشرية في كتابة المخطوطة العلمية ومناقشة النتائج.

3.4. المعضلات الأخلاقية والشفافية في عصر الذكاء الاصطناعي

تشير الأدلة الحديثة إلى أزمة ثقة متنامية في قدرة المؤسسات الأكاديمية على ضمان النزاهة العلمية في ظل انتشار المحتوى المُنشأ بالذكاء الاصطناعي.

- انفجار المنشورات الزائدة (Redundant Publications): وثق تحليل لبيانات NHANES زيادة هائلة بمقدار 17 ضعفا في المنشورات المتطابقة بين عامي 2022 و2024، مما يشير إلى فشل منهجي في عمليات التحقق التحريرية، حيث يمكن للذكاء الاصطناعي إعادة صياغة النصوص لخفض درجات التشابه إلى أقل من 5% (معظم كاشفات الانتحال) (BMC Medicine, 2026).
- فشل كاشفات المحتوى الاصطناعي: كشف تقييم شامل لستة كاشفات تجارية رائدة (بما في ذلك GPTZero وTurnitin وOriginality.AI) عن هشاشة كارثية. تراوحت نسبة النتائج الإيجابية الكاذبة بين 0.05% و68.6%، بينما بلغت نسبة النتائج السلبية الكاذبة بين 0.3% و99.6%. والأكثر خطورة، أنه يمكن خفض دقة الكشف من 94.2% إلى 2.5% ببساطة عن طريق طلب إعادة صياغة النص بأسلوب معجمي أكثر تعقيداً (Weber et al., 2026). هذه النتائج تجعل الاعتماد على هذه الكاشفات في القرارات الأكاديمية المصيرية بمثابة "رمي العملة".

4. المناقشة

يؤكد هذا التحليل المرجعي أن الذكاء الاصطناعي لم يعد مجرد أداة مساعدة هامشية، بل أصبح **شريكاً فاعلاً في دورة الإنتاج المعرفي. يوضح الشكل 2 العلاقة التكاملية التي تبلورت من هذا التحليل، حيث يتفاعل الإنسان والذكاء الاصطناعي عبر مراحل البحث المختلفة، مع حدود واضحة للثقة والاعتمادية.

شكل 2: نموذج مقترح للتفاعل البشري مع الذكاء الاصطناعي في الورقة البحثية IMRAD

الحدود بين الموثوقية وعدم اليقين

حدد Kong et al. (2026) حداً حاداً بين المهام التي يمكن للذكاء الاصطناعي أداءها بشكل موثوق (مرتكزة على الاسترجاع، والهيكلة الواضحة، والوساطة بالأدوات) وتلك التي يظل فيها هشاً وضعيفاً (توليد أفكار جديدة حقاً، والأحكام العلمية المعقدة). تتفق هذه النتيجة مع تحذيرات Lahlou et al. (2026) من أن LLMs تميل بطبيعتها إلى إنتاج محتوى تيار رئيسي محايد سياسياً، وتفتقر إلى المنظور النقدي العميق، وقد تنتج مراجعات "مختلفة بشكل كبير" لنفس مجموعة المقالات اعتماداً على اختياراتها الخوارزمية الأولية.

نحو تكامل مسؤول: إطار عمل مقترح

بناءً على التحليل، أقترح نموذج التكامل المسؤول المكون من ثلاث طبقات للتفاعل بين الباحث والأدوات الذكية:

1. **الطبقة الأساسية (Foundation):** الأتمتة الكاملة للمهام المنظمة والموثقة جيداً (مثل فحص العناوين والملخصات، وتنسيق المراجع، والتحليل الإحصائي الروتيني). هنا، يمكن الوثوق بالمخرجات مع إشراف بشري بسيط (Motika, 2026; Adel & Alani, 2025).
2. **طبقة التعاون (Collaboration):** التدخل البشري الإلزامي في المهام التي تتطلب حكماً متخصصاً (مثل استخلاص البيانات المعقدة، وتفسير النتائج غير المتوقعة، وتقييم جودة الدراسات النوعية). يجب استخدام الذكاء الاصطناعي كـ"مراجع ثالث" أو كمساعد لتعزيز القرار، وليس كبديل عنه (Sahu & Balakrishnan, 2026; Hilkenmeier et al., 2025)
3. **طبقة الحذر (Caution):** تجنب الاعتماد على المحتوى المُنشأ بالكامل دون تحقق صارم، خاصة في توليد الفرضيات الجديدة وكتابة المناقشة المتعمقة. هنا، تظل الخبرة البشرية والأصالة الفكرية هي السمة المميزة للعلم الرصين (Kong et al., 2026).

5. الاستنتاجات

يقف البحث العلمي عند مفترق طرق تاريخي. لم تعد مسألة ما إذا كان ينبغي استخدام الذكاء الاصطناعي في كتابة الورقة البحثية مطروحة؛ بل أصبح السؤال الملح هو كيف يمكن استخدامه بشكل مسؤول وشفاف لتعزيز، وليس تقويض، سلامة العلم.

يخلص هذا التحليل المرجعي إلى ثلاث نتائج رئيسية:

1. **كفاءة متفاوتة:** تتفوق أدوات الذكاء الاصطناعي بشكل كبير في المهام المنظمة ضمن نمط IMRAD (مثل الفحص الأولي في المنهجية، وتحليل البيانات الروتيني في النتائج)، ولكنها تظل غير موثوقة في المهام التي تتطلب حكماً سياقياً عميقاً (مثل التوليف النقدي في المناقشة).
2. **أزمة ثقة في أدوات الكشف:** كاشفات المحتوى الاصطناعي التجارية الحالية غير صالحة للاستخدام في اتخاذ قرارات عقابية أو تحريرية، نظراً لمعدلات الخطأ الفادحة فيها، مما يتطلب إعادة تقييم جذرية لكيفية ضمان النزاهة الأكاديمية.
3. **إعادة تعريف التأليف:** مع ظهور أطر الوكلاء المتعددة القادرة على كتابة مراجعات بأكملها، تبرز أسئلة وجودية حول مفهوم التأليف، والملكية الفكرية، والمهارات التي يجب أن يمتلكها الباحث في المستقبل.

التوصيات: يجب على المجلات الأكاديمية والمؤسسات التعليمية تطوير سياسات واضحة ومدروسة تحدد الاستخدام المسموح به للذكاء الاصطناعي، مع مطالبة الباحثين بالشفافية الكاملة والإفصاح عن دور هذه الأدوات في كل قسم من أقسام الورقة البحثية. مستقبل العلم ليس في استبدال الباحثين بالآلات، بل في تسخير قوة الحوسبة المعرفية لتوسيع حدود الفهم البشري، مع الحفاظ على القيم الجوهرية للفضول، والنزاهة، والشك المنهجي.

المراجع

[1] Adel, A., & Alani, N. (2025). Can generative AI reliably synthesise literature? Exploring hallucination issues in ChatGPT. AI & Society, 40(8), 6799–6812 .

[2] Afonso, M., et al. (2026). ProfOlaf: Semi-automated tool for systematic literature reviews. arXiv preprint arXiv:2510.26750 .

[3] Alazba, A., et al. (2026). An exploratory study on effectiveness of GPT-4o in conducting sub-tasks of systematic literature reviews. Automated Software Engineering, 33, Article 81 .

[4] Bâra, A., & Oprea, S. V. (2025). AI-augmented bibliometric framework: A paradigm shift with agentic AI for dynamic, snippet-based research analysis. arXiv preprint arXiv:2511.21745 .

[5] Go, G. H. T., Ly, K., Søgaard, A., Tabatabaei, S. A., Rijke, M. de, & Chen, X. (2026). LiRA: A multi-agent framework for reliable and readable literature review generation. Proceedings of the AAAI Conference on Artificial Intelligence, 40 (47), 40456–40464 .

[6] Hilkenmeier, F., Pelzer, M., Stierle, C. M. G., & Fink-Lamotte, J. (2025). Evaluating the AI tool "Elicit" as a semi-automated second reviewer for data extraction in systematic reviews: A proof-of-concept. Social Science Computer Review .

[7] Kong, L., et al. (2026). AI for auto-research: Roadmap & user guide. arXiv preprint arXiv:2605.18661 .

[8] Lahlou, S., Gouttebroze, A., Oraee, A., & Madera, J. (2026). Writing literature reviews with AI: Principles, hurdles and some lessons learned. arXiv preprint arXiv:2603.20235

[9] A Ahmed, A Alsharif, M khaleel, Artificial Intelligence: Concepts, Technologies, and Real-World Applications, Publisher: IKSAD publishing house ISBN: 978-625-367-806-7, August 2024

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **JSHD** and/or the editor(s). **JSHD** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.