Journal of Economic, Administrative and Legal Sciences (JEALS) • Vol 9, Sup Issue 10 (2025) • P: 46 - 26
https://journals.ajsrp.com/index.php/jeals
ISSN: 2522-3372 (Online) • ISSN: 2522-3372 (Print)

# The Role of Explainable Artificial Intelligence (XAI) in Supporting Knowledge Management (KM) and Its Processes: An Analytical Study

**Ms. Mashael Abdulrahman Al-Hazmi***
King Abdulaziz University | Kingdom of Saudi Arabia
**Ms. Enas Mohammed Al-Najjar, Ms. Wafa Mohna Al-Shaikh**
Faculty of Arts & Humanities | King Abdulaziz University | KSA

**Received**: 20/06/2025
**Revised**: 24/06/2025
**Accepted**: 06/07/2025
**Published**: 30/10/2025

* Corresponding author: mashael.alhazmi@hotmail.com

**Citation**: Al-Hazmi, M. A., Al-Najjar, E. M., & Al-Shaikh, W. M. (2025). The Role of Explainable Artificial Intelligence (XAI) in Supporting Knowledge Management (KM) and Its Processes: An Analytical Study. *Journal of Economic, Administrative and Legal Sciences, 9*(10S), 26 – 46. https://doi.org/10.26389/AJSRP.M220625

2025 © AISRP • Arab Institute for Sciences & Research Publishing (AISRP), United States, all rights reserved.

• **Open Access**

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license

**Abstract:** Problem: This study addresses the research gap related to exploring the role of Explainable Artificial Intelligence (XAI) in supporting knowledge management processes and investigating how this technology can be employed to enhance decision-making in contemporary environments. Significance: The importance of this study lies in its potential to highlight the role of this emerging technology in supporting knowledge management and its core processes, particularly in the domains of knowledge discovery, representation, and extraction. Hypotheses: The study assumes that Explainable Artificial Intelligence (XAI) plays a positive role in supporting knowledge management and its associated processes. Objectives: This study aimed to clarify the concept of Explainable Artificial Intelligence (XAI) and explore the relationship between XAI technologies and knowledge management (KM) and its processes, through an analytical review of the relevant scientific literature. Methods: The study adopted an inductive analytical approach by examining English-language studies published between 2020 and 2025, to identifying theoretical and practical frameworks that illustrate how XAI can support KM processes, including knowledge discovery, representation, and extraction. Results: The study concluded that (XAI) technologies have a significant positive impact on enhancing the effectiveness of knowledge management by producing interpretable outputs that support better decision-making and improve decision quality for decision-makers. The findings also indicate that (XAI) fosters transparency and trust in AI systems, thereby increasing the reliability of AI adoption in the context of knowledge management.

**Keywords:** Explainable artificial intelligence, XAI, black box models, knowledge Management, knowledge discovery, knowledge representation, knowledge management processes.

# دور تقنية الذكاء الاصطناعي التفسيري في دعم إدارة المعرفة وعملياتها: دراسة استقرائية تحليلية

**أ. مشاعل عبد الرحمن الحازمي***
جامعة الملك عبد العزيز | المملكة العربية السعودية
**أ. إيناس محمد النجار، أ. وفاء مهنا الشيخ**
كلية الآداب والعلوم الإنسانية | جامعة الملك عبد العزيز | المملكة العربية السعودية

**المستخلص:** مشكلة الدراسة: معالجة الفجوة البحثية لاستكشاف دور تقنية الذكاء الاصطناعي التفسيري (XAI) في دعم عمليات إدارة المعرفة وكيفية توظيف هذه التقنية لتعزيز عملية اتخاذ القرار في البيئات المعاصرة. أهمية الدراسة: تكمن أهمية الدراسة في إمكانية إبراز دور هذه التقنية الناشئة في دعم إدارة المعرفة وعملياتها، لاسيما في مجالات اكتشاف المعرفة، تمثيلها، واستخراجها. فرضيات الدراسة: وتفترض الدراسة أن تقنية الذكاء الاصطناعي التفسيري (XAI) لها دور إيجابي في دعم إدارة المعرفة وعملياتها. الأهداف: هدفت الدراسة إلى توضيح مفهوم تقنية الذكاء الاصطناعي التفسيري، وذلك لاستكشاف العلاقة بين تقنيات الذكاء الاصطناعي التفسيري (XAI) وإدارة المعرفة وعملياتها، من خلال تحليل الأدبيات العلمية المنشورة حول الموضوع. المنهجية: اعتمدت الدراسة على المنهج الاستقرائي التحليلي، حيث تم تحليل الدراسات المنشورة باللغة الإنجليزية خلال الفترة من 2020م إلى 2025م، بهدف تحديد الأطر النظرية والتطبيقية التي توضح كيف يمكن لتقنيات (XAI) أن تسهم في دعم عمليات إدارة المعرفة: اكتشاف المعرفة، تمثيلها واستخراجها. النتائج: وقد توصلت الدراسة إلى أن لتقنيات الذكاء الاصطناعي التفسيري أثرًا إيجابيًا ملحوظًا في تعزيز فعالية إدارة المعرفة، من خلال قدرتها على تقديم مخرجات قابلة للتفسير تسهم في تحسين اتخاذ القرار وجودته من قبل صنّاع القرار. كما أظهرت النتائج أن تقنيات (XAI) تعزّز الشفافية والثقة في أنظمة الذكاء الاصطناعي، مما يزيد من موثوقية استخدام أنظمة الذكاء الاصطناعي في مجال إدارة المعرفة.

**الكلمات المفتاحية:** الذكاء الاصطناعي التفسيري، إدارة المعرفة، اكتشاف المعرفة، تمثيل المعرفة، عمليات إدارة المعرفة، نماذج الصندوق الأسود.

## 1- مقدمة

في ظل الاعتماد المتزايد على التكنولوجيا في المنظمات التي تسعى إلى البقاء وتحقيق الميزة التنافسية، ومع التقدم الهائل في القدرات الحاسوبية وإتاحة الوصول إلى مجموعات بيانات ضخمة، ازداد اعتماد المنظمات منذ العقد الماضي في عمليات تحليل البيانات والتنبؤ على تقنيات الذكاء الاصطناعي ونماذج التعلم الآلي (Machine Learning Models)، حيث أصبحت تُستخدم على نطاق واسع لتطوير أنظمة دعم القرار في المنظمات. وعلى الرغم من دقة النتائج التي تقدمها هذه النماذج، إلا أن غموض آلية عملها يضعف الثقة في مخرجات نماذج الذكاء الاصطناعي، وخاصة عندما تتعلق هذه القرارات بمصائر الأفراد والمجتمعات. ويُشار إلى هذه المشكلة بمصطلح الصندوق الأسود (Black Box)؛ لأن النموذج يعطي مخرجات دون أن يكون هناك تفسير واضح لما يحدث داخله، مما يبرز تحديًا واضحًا حينما تُوكل قرارات مهمّة إلى نظام لا يستطيع أن يُخبرنا كيف توصّل إلى هذه النتيجة.

وتبرز هذه المشكلة خاصة في المجال العسكري ومجال الرعاية الصحية، اللذين غالبًا ما يتعلقان باتخاذ قرارات مصيرية تتعلق بالحياة والموت، حيث يمكن أن تؤثر التشخيصات أو التوصيات الطبية المولّدة بالذكاء الاصطناعي على حياة المرضى، مما يطرح تساؤلات عميقة حول المساءلة الأخلاقية والقانونية، خاصة عندما تكون الأخطاء غير قابلة للتفسير أو التبرير. لذلك أصبحت هناك حاجة ملحة لتفسير نتائج ومخرجات نماذج الذكاء الاصطناعي لضمان شفافيتها وموثوقيتها، وأخيرا مع ظهور الجيل الثاني من النظم الخبيرة في أوائل التسعينيات، تحسنت قابلية التفسير من خلال تضمين المزيد من المعرفة حول المجال والمستخدمين والأنظمة الخبيرة. (Ridley, 2024, pp. 100-101)

وفي ضوء التحول الرقمي المتسارع، أصبحت إدارة المعرفة (Knowledge Management) من أهم المحاور الاستراتيجية التي تعتمد عليها المنظمات الحديثة لدعم اتخاذ القرار، بجميع عملياتها التي تتضمن: توليد المعرفة، مشاركة المعرفة، اكتشاف المعرفة، استخراج المعرفة تمثيل المعرفة، وتقييم المعرفة.

ومن هذا المنطلق، برزت أهمية الذكاء الاصطناعي التفسيري (Explainable AI-XAI) الذي يهدف إلى مساعدة الناس على الفهم والاعتماد على نتائج أنظمة الذكاء الاصطناعي أو التعلم الآلي، وتعزيز الشفافية في هذه النماذج دون التأثير على أدائها، مما يتيح للمستخدمين فهم آلية عملها، كما يُعد عامل الثقة جوهريًا عند تطوير أو استخدام مثل هذه النماذج لاتخاذ قرارات مهمة، إذ يرغب المستخدمون في معرفة سبب اتخاذ النموذج لقرار معين، أو كيفية وصوله إليه. (Majumder & Dey, 2022, pp. 101-102)

بناء على ما سبق، تهدف هذه الدراسة إلى تقديم فهم شامل لتقنية الذكاء الاصطناعي التفسيري (XAI)، ومناقشة كيف تسهم هذه التقنية في دعم إدارة المعرفة وعملياتها داخل المنظمات، خاصة التي تتطلب تفسيرات شفافة ودقيقة لدعم اتخاذ القرار.

### إشكالية الدراسة

على الرغم من التقدم الكبير في نماذج الذكاء الاصطناعي والتعلم الآلي واعتماد المنظمات المتزايد عليها في دعم عمليات اتخاذ القرار، فإن هذه النماذج لا تزال تواجه تحديًا جوهريًا يتمثل في غياب الشفافية أو توضيح الآليات التي تقود إلى قرارات ونتائج معينة. هذا القصور يُضعف من موثوقيتها ويحد من إمكانية الاعتماد عليها في السياقات الحساسة التي تتطلب تفسيرات واضحة وقابلة للتدقيق. ومن هنا ظهرت الحاجة إلى تقنية الذكاء الاصطناعي التفسيري (XAI) كأداة لمعالجة هذا التحدي من خلال تحسين قابلية تفسير مخرجات النماذج الذكية، وتمكين اتخاذ قرارات أكثر دقة وشفافية ((Longo et al., 2020, p. 2. ومع ذلك، فإن الإشكالية البحثية الرئيسية تكمن في أن معظم الدراسات السابقة تركز على جوانب محدودة من تقنيات الذكاء الاصطناعي التفسيري (XAI) أو تطبيقاتها القطاعية، دون تناول لدورها في دعم إدارة المعرفة وعملياتها. وهذا القصور يخلق فجوة بحثية تحتاج إلى معالجة لاستكشاف كيفية توظيف تقنية (XAI) لدعم عمليات إدارة المعرفة المختلفة وتعزيز اتخاذ القرار في البيئات المعاصرة، خاصة في ظل التحديات الناشئة عن البيانات الضخمة والتحول الرقمي. ومن هذا المنطلق، تعرض مشكلة الدراسة في السؤال التالي:

ما هو دور تقنيات الذكاء الاصطناعي التفسيري في دعم إدارة المعرفة وعملياتها؟

### أهمية الدراسة

تنبع أهمية هذه الدراسة من أهمية تقنية الذكاء الاصطناعي التفسيري (XAI) كأحد أبرز توجهات التقنيات الناشئة التي تستخدم لتفسير مخرجات النماذج المعقدة في الذكاء الاصطناعي، مما يسهم في تعزيز الشفافية ودعم اتخاذ القرار. وتُركز الدراسة على إبراز دور (XAI) في دعم إدارة المعرفة وعملياتها، لاسيما في مجالات اكتشاف المعرفة، تمثيلها، واستخراجها.

وتكمن أهمية الدراسة في سد فجوة معرفية تتعلق بكيفية الاستفادة من هذه التقنية وتوظيفها لدعم إدارة المعرفة وعملياتها داخل المنظمات، من خلال توفير أدوات تفسيرية تساعد متخذي القرار على فهم النتائج واستخدامها بفعالية.

**أهداف الدراسة**

تهدف الدراسة إلى تحقيق عدة أهداف رئيسية تتعلق باستكشاف دور تقنيات الذكاء الاصطناعي التفسيري في دعم إدارة المعرفة وعملياتها من خلال مراجعة الأدبيات العلمية الحديثة، ويمكن تلخيص هذه الأهداف كما يلي:

1. التعرف على مفهوم الذكاء الاصطناعي القابل التفسيري (XAI) ومكوناته، وآلية عمله، استنادًا إلى الأدبيات النظرية.
2. استكشاف دور تقنيات الذكاء الاصطناعي التفسيري (XAI)في دعم عمليات إدارة المعرفة.
3. التعرف على التحديات التي تواجه تقنية الذكاء الاصطناعي التفسيري، كما وردت في الدراسات الحديثة.
4. تقديم توصيات لتعزيز توظيف تقنية الذكاء الاصطناعي التفسيري في مجال إدارة المعرفة.

**أسئلة الدراسة**

تسعى هذه الدراسة إلى الإجابة عن التساؤل الرئيسي الآتي:

ما هو دور تقنيات الذكاء الاصطناعي التفسيري في دعم إدارة المعرفة وعملياتها؟

ويتفرع من هذا التساؤل عدد من التساؤلات الفرعية:

1- ما هو مفهوم الذكاء الاصطناعي التفسيري؟
2- ما أبرز عمليات إدارة المعرفة التي يمكن تعزيزها باستخدام تقنية الذكاء الاصطناعي التفسيري؟
3- ما هي التحديات التي تواجه تطبيق تقنيات الذكاء الاصطناعي التفسيري؟

**فرضيات الدراسة**

تفترض الدراسة أن تقنية الذكاء الاصطناعي التفسيري (XAI) لها دور إيجابي في دعم إدارة المعرفة وعملياتها ومن هذه الفرضية الرئيسية تتفرع الفرضيات التالية:

1. تقنية الذكاء الاصطناعي التفسيري تسهم في دعم إدارة المعرفة.
2. تقنية الذكاء الاصطناعي التفسيري تسهم في دعم عملية اتخاذ القرار.
3. تقنية الذكاء الاصطناعي التفسيري تسهم في تعزيز عملية اكتشاف المعرفة.
4. تقنية الذكاء الاصطناعي التفسيري تسهم في تعزيز عملية تمثيل المعرفة.
5. تقنية الذكاء الاصطناعي التفسيري تسهم في تعزيز عملية استخراج المعرفة.

**منهجية الدراسة**

تعتمد هذه الدراسة على مجموعة من الخطوات المنهجية التي تهدف إلى فهم وتحليل دور تقنيات الذكاء الاصطناعي التفسيري (XAI) في دعم تطبيق إدارة المعرفة وعملياتها داخل المنظمات. حيث يتم جمع البيانات من خلال مراجعة الأدبيات العلمية ذات الصلة، وذلك باختيار مجموعة من الدراسات الحديثة المنشورة بين عامي 2020م و2025م، التي تتناول موضوعات تقنيات الذكاء الاصطناعي التفسيري (XAI) وإدارة المعرفة وعملياتها.

ويتم اختيار الدراسات بناءً على مدى ارتباطها بأهداف الدراسة وتساؤلاتها، ويتم تحليل هذه البيانات باستخدام أساليب تحليل النصوص أو التحليل الموجه نظريًا، بهدف استخلاص الأنماط والمفاهيم والتوجهات التي توضّح طبيعة العلاقة بين تقنيات الذكاء الاصطناعي التفسيري وبين عمليات إدارة المعرفة، مثل: عملية اكتشاف المعرفة، عملية تمثيل المعرفة وعملية استخراج المعرفة.

وأخيرا، يتم تفسير ومناقشة النتائج المستخلصة من الأدبيات بهدف استكشاف دور تقنيات الذكاء الاصطناعي التفسيري في تعزيز ودعم إدارة المعرفة وعملياتها، كما يتم تقديم استنتاجات وتوصيات يمكن أن تساهم في توجيه الجهود البحثية والممارسات المؤسسية نحو تعزيز التكامل بين تقنيات الذكاء الاصطناعي التفسيري وإدارة المعرفة.

**حدود الدراسة:**

- الحدود الموضوعية: تقتصر هذه الدراسة على تحليل دور تقنية الذكاء الاصطناعي التفسيري (XAI) في إدارة المعرفة مع التركيز على مساهمتها في اتخاذ القرار القائم على المعرفة، بالإضافة إلى تحليل دورها في ثلاث عمليات من عمليات إدارة المعرفة وهي: (اكتشاف المعرفة، تمثيل المعرفة، استخراج المعرفة).
- الحدود الزمانية: تعتمد الدراسة في تحليلها ومراجعتها على الأدبيات العلمية المنشورة خلال الفترة من عام 2020م إلى عام 2025م.
- الحدود اللغوية: تقتصر الدراسة على تحليل الأدبيات المنشورة باللغة الإنجليزية فقط.

**مصادر البحث:**

الأدبيات والدراسات العلمية المنشورة المرتبطة بموضوع البحث بشكل ممنهج وموثق وذلك بهدف إبراز الجهود العلمية المتعلقة بالعلاقة بين (تقنية الذكاء الاصطناعي التفسيري) و (إدارة المعرفة وعملياتها)، وسيتم ذلك من خلال مراجعة الدراسات التي تناولت هذا الموضوع في بيئات تنظيمية متعددة.

ولتحقيق ذلك، تم اعتماد الأدبيات التالية:

- الكتب
- المقالات العلمية المنشورة في مجلات علمية
- البحوث المنشورة في مجلات علمية

تمت عملية البحث بطريقة منهجية منظمة باللغة الانجليزية عبر قواعد بيانات الكترونية متخصصة ومعتمدة، شملت:

- Google Scholar
- Search Gate
- Sciencedirect

تم استخدام الكلمات المفتاحية التالية في البحث:

Knowledge Management, XAI, Explainable AI, Artificial Intelligence, Black Box, Knowledge Discovery, Knowledge Representation, Knowledge Extraction, Decision Making.

## 2- الإطار النظري للدراسة

في المنظمات الحديثة، أصبحت إدارة المعرفة من أهم الاستراتيجيات التي تسهم في بقاء المنظمة وتحقيق الميزة التنافسية، وفي ظل التحولات الرقمية المتسارعة ازداد اعتماد هذه المنظمات على تقنيات الذكاء الاصطناعي ونماذج التعلّم الآلي بشكل كبير في إدارة المعرفة وعملياتها، إلا أن الاعتماد المتزايد على هذه النماذج أدى إلى تساؤلات جوهرية تتعلق بمدى موثوقية وشفافية القرارات الناتجة عنها، وفي هذا السياق برزت تقنية الذكاء الاصطناعي التفسيري كأحد الحلول الواعدة التي تهدف إلى تقديم مخرجات قابلة للتفسير، وانطلاقا من ذلك تسعى هذه الدراسة إلى تسليط الضوء على تقنية الذكاء الاصطناعي التفسيري من خلال توضيح مفهوم تقنية الذكاء الاصطناعي التفسيري، والغرض من هذه التقنية، وخصائصها، ومكوناتها، ومجالات تطبيقها، وكذلك، سيتم استعراض آلية عمل تقنية (XAI)ومتطلباتها وتصنيفاتها والخوارزميات المستخدمة، بالإضافة إلى تناول التحديات التي تواجه هذه التقنية ، وأخيرًا، سيتم مناقشة دور تقنية الذكاء الاصطناعي التفسيري في دعم إدارة المعرفة وعملياتها، مع التركيز على دورها في تعزيز عملية اكتشاف المعرفة، تمثيل المعرفة، واستخراج المعرفة.

**مفهوم تقنية الذكاء الاصطناعي التفسيري (Explainable AI Concept)**

تعرف تقنية الذكاء الاصطناعي التفسيري (XAI) بأنها مجال فرعي من الذكاء الاصطناعي يُعنى بتوفير الشفافية Transparency))، والتفسيرية ((Interpretability والقابلية للتفسير Explainability)) . فهي بذلك التقنية التي تعمل على تحسين فهم نماذج الذكاء الاصطناعي مما يسمح للمستخدمين بالفهم العميق لنقاط القوة والقيود والافتراضات التي يعتمد عليها النموذج. ووضح Alkhanboulli وآخرون نقلًا عن Doshi-Velez & Kim أن التفسيرية في سياق الذكاء الاصطناعي التفسيري هي القدرة على التفسير أو العرض بطرق مفهومة للبشر. أما القابلية للتفسير فهي فهم العمليات الداخلية والخطوات التي يتبعها النموذج للوصول إلى استنتاج معين. (Alkhanbouli et al., 2025, p.2)

كما أن أحد التعريفات الأكثر حداثة وقبولًا للذكاء الاصطناعي التفسيري هو الذي أشار إليه Ali وآخرون نقلًا عن Arrieta وآخرون، والذي يعرّف الذكاء الاصطناعي التفسيري بأنه النظام الذي يقدم التفاصيل والأسباب التي تتيح للجمهور فهم آلية عمل الأنظمة بوضوح وسهولة. (Ali et al., 2023, p.7)

وترى الباحثات أن التعريفين يتشابهان في توضيحهم للغرض من تقنية الذكاء الاصطناعي التفسيري، ولكن تعريف Alkhanbouli et al., 2025) يعد أكثر شمولًا، حيث إنه يربط بين المفهوم الأساسي للتقنية والمفاهيم المحورية المرتبطة به مثل الشفافية والتفسيرية والفهم، وهذا يظهر بعدًا أعمق للتقنية.

**الغرض من تقنية الذكاء الاصطناعي التفسيري (Explainable AI Purpose)**

أشار Ridley (2024) نقلًا عن (DAPRA, 2016)و (Turek,2016) و (Langeret, 2021) إلى أنه وفقًا لوصف وكالة مشاريع البحوث الدفاعية المتقدمة الأمريكية، يهدف الذكاء الاصطناعي التفسيري (XAI) إلى تمكين أنظمة الذكاء الاصطناعي من تقديم تفسيرات واضحة

لقراراتها مع تحديد نقاط القوة والضعف فيها مما يساعد على فهم سلوكها في المستقبل، كما يسعى (XAI) إلى تعزيز ثقة المستخدمين والمساهمة في إدارة الجيل الجديد من شركاء الذكاء الاصطناعي بشكل فعال. ويركز (XAI) بصفة عامة على تطوير أساليب تتيح شرح الأنظمة الاصطناعية وجعلها مفهومة لأصحاب المصلحة من البشر. (Ridley, 2024, p. 101)

كذلك يهدف نظام الذكاء الاصطناعي التفسيري (XAI)إلى توضيح سلوكه للبشر من خلال تقديم تفسيرات واضحة. حيث ينبغي أن يكون نظام (XAI)قادرًا على شرح قدراته وفهمه بما في ذلك توضيح ما قام به سابقًا، وما يقوم به حاليًا، وما سيحدث لاحقا، بالإضافة إلى الكشف عن المعلومات الأساسية التي يستند إليها في تصرفاته ومع ذلك، يتم وضع كل تفسير في سياق يعتمد على المهمة والقدرات وتوقعات مستخدم نظام الذكاء الاصطناعي. لذا فإن تعريفات قابلية التفسير والتوضيح تعتمد على المجال المحدد، وقد لا تكون مستقلة عنه، وقد تكون التفسيرات إما كاملة أو جزئية؛ حيث تقدم النماذج القابلة للتفسير بالكامل تفسيرات شاملة وشفافة تمامًا، بينما تكشف النماذج القابلة للتفسير جزئيًا عن جوانب مهمة من عملية تفكيرها. تخضع النماذج القابلة للتفسير لقيود معينة تتعلق بقابلية التفسير، والتي تحدد وفقًا للمجال (مثل التكرار فيما يتعلق بمتغيرات معينة والعلاقات المرتبطة بها)، في حين أن النماذج غير القابلة للتفسير قد لا تلتزم بهذه القيود. (Gunning et al., 2019, p. 4)

بالإضافة إلى ذلك؛ يعتبر التحسين هدفًا إضافيًا يمكن تحقيقه من خلال تطبيق أساليب الذكاء الاصطناعي القابلة للتفسير (XAI) حيث إن فهم العمليات الداخلية لمنظمة الذكاء الاصطناعي والنتائج الناتجة عنها أمرًا أساسيًا لتعزيز فعالية الخوارزميات وبالتالي فإن تحسين قابلية التفسير يمكن أن يسهم في زيادة دقة النظام وقيمته. من هنا يمكن القول أن التركيز على تحسين أداء الخوارزمية بالإضافة إلى تصحيح الأخطاء والتحقق من أجل اتباع نهج هندسي منظم يعتمد على تحليل الأسباب بدلاً من التجربة والخط. (Meske et al., 2022, pp. 56-57)

بناء على ما تم ذكره؛ ترى الباحثات أن الغرض من الذكاء الاصطناعي التفسيري (XAI) يكمن في الآتي: تطوير طرق للتعلم وصنع القرارات للأنظمة الاصطناعية، وتقديم تفسيرات دقيقة للسلوكيات وشرحها بوضوح تام، وأخيرًا التحسين والسعي نحو تطوير قابلية التفسير وأداء الخوارزميات بفعالية وكفاءة عالية.

**خصائص تقنية الذكاء الاصطناعي التفسيري (Explainable AI Features)**

يتميز الذكاء الاصطناعي التفسيري (XAI) بتسع خصائص رئيسية، وهي كالتالي:

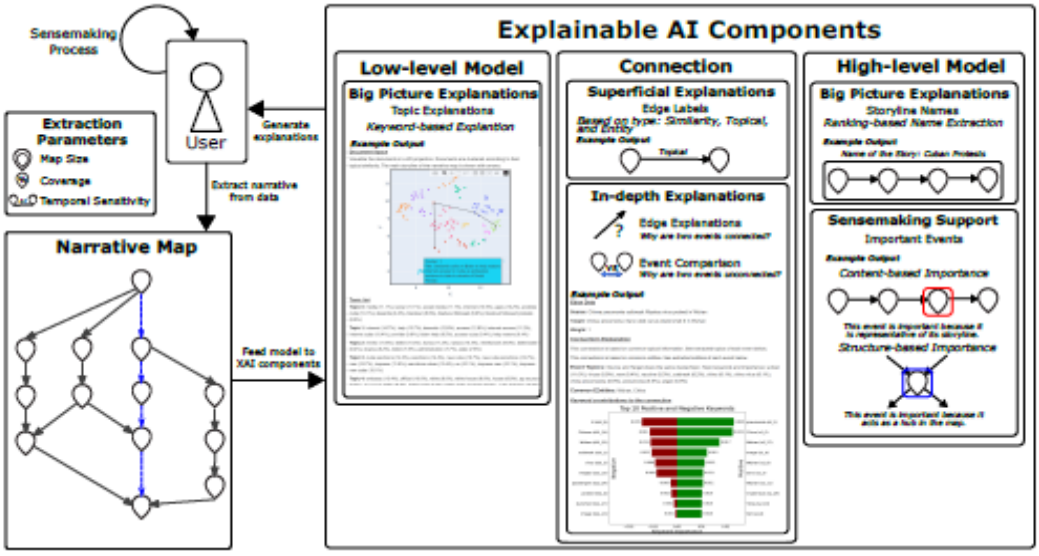
1. الموثوقية :(Reliability) وتعني التأكد من فعالية النموذج في تنفيذ المهمة الموكلة إليه وفقًا للخطة المحددة.. وحتى يتم الاستفادة من تنبؤات النظام في التطبيقات الواقعية، يجب على المستخدم أن يثق في النموذج المستخدم حيث يُعتبر تقديم تفسير للتنبؤات عنصرًا أساسيًا لبناء الثقة لدى الإنسان وضمان الاستخدام الفعّال لتقنيات التعلم الآلي، بشرط أن تكون هذه التفسيرات صادقة وسهلة الفهم.
2. السببية :(Causality) توضح السببية العلاقة بين السبب والنتيجة بين مساحة الميزة والمخرجات المحتملة. ويتطلب تحديد مجموعة من أسباب التأثير المعرفة الشاملة، حيث تسعى نماذج التعلم الآلي إلى اكتشاف الارتباطات بين الميزات كذلك تقوم بتحسين جودة التفسيرات من خلال توفير مستوى معين من فهم الأسباب من قبل الخبير البشري.
3. قابلية النقل :(Transferability) يمكن نقل نموذج قياسي قابل للتفسير لاستخدامه في مجالات أخرى لتحقيق نتائج فعالة. بشكل عام يتم تدريب النموذج واختباره عادةً باستخدام مجموعة بيانات محدودة. ويذكر أن أحد الأسباب التي تدفع إلى أهمية تفسير النموذج هو إمكانية استخدام نموذج مدرب بشكل صحيح في مجال آخر يتمتع بخصائص مشابهة، وبالتالي يمكن وصف هذه الخاصية بأنها قابلية إعادة الاستخدام.
4. المعلومات :(Informativeness) يوفر النموذج القابل للتفسير معلومات إضافية حول المشكلة التي يتم التعامل معها. وتقدم المعلومات التي يوفرها نموذج (XAI)توضيحات حول كيفية عمل النظام، مما يساعد في تجنب المفاهيم الخاطئة، كما يشرح النموذج العلاقات الداخلية ويعزز فهم المستخدم للعملية الداخلية.
5. الثقة :(Confidence) تتميز الأطر القابلة للتفسير بالقوة والثبات والموثوقية. حيث تسعى إلى إمكانية التحقق من صحة القرار أو عدمه ومن الضروري دائمًا تقييم هذا الإجراء ضمن نظام تكون فيه الموثوقية أمرًا مهمًا. لذا يجب أن يوفر النموذج القابل للتفسير المصمم لهذا الغرض مستوى من الثقة في تنبؤاته.
6. التفاعل :(Interaction) يتم التركيز على هذا الهدف في النماذج التي تحتاج إلى تفاعل مع المستخدمين النهائيين. ينبغي أن يوضح النموذج القرار الذي تم اتخاذه والخيارات التي تم النظر فيها، ثم يقدم تفسيرًا بلغة طبيعية واضحة للحلول الغامضة.
7. الإنصاف :(Fairness Justifiability) يمكن تقييم مخرجات نموذج قابل للتفسير وإخضاعها للتحليل. تعتبر خوارزميات التعلم الآلي نتاجًا للبيانات المستخدمة، وأي تحيز موجود في بيانات الإدخال يمكن أن يؤثر سلبًا على النتائج المحققة ويعيق الوصول إلى

استنتاجات عادلة. كذلك يمكن القول بأن نظام الذكاء الاصطناعي التفسيري (XAI)يكشف عن التباينات في البيانات ويضمن تحقيق العدالة في نماذج التعلم الآلي.

8. إمكانية الوصول :(Accessibility Interactivity) القدرة على التفاعل بشكل مباشر مع عملية اتخاذ القرار لنموذج يمكن تفسيره. تُعتبر قابلية التفسير وسيلة لتعزيز فعالية العمليات الداخلية لنماذج التعلم الآلي فهي تتيح للمستخدمين بما في ذلك غير المحترفين إمكانية تحسين الأداء وفقًا لاحتياجاتهم.
9. الخصوصية :(Privacy) القدرة على توضيح العمليات الداخلية لنموذج من قبل جهة خارجية، يمكن أن تتيح القدرة على توضيح المنطق وراء النموذج وسيلة لتقييم الخصوصية. فقد يؤدي النموذج غير الشفاف إلى جمع بيانات حساسة، مما يسبب انتهاكًا للخصوصية.((Minh et al., 2022, pp. 3513-3514: Moosavi et al., 2024, p. 3

**مكونات تقنية الذكاء الاصطناعي التفسيري(Explainable AI Components)**

هناك ثلاثة مستويات رئيسية للذكاء الاصطناعي التفسيري يتم توظيفها أثناء عملية التفسير وتتضح في الشكل التالي:




**الشكل (1): مكونات تقنية الذكاء الاصطناعي التفسيري**

**المصدر: (Keith et al., 2025, p.4)**

1. النموذج المنخفض المستوى: يقوم بتوفير تفسيرات عامة حول توزيع المواضيع، حيث يستخدم الكلمات المفتاحية لفهم العلاقات بين المستندات.
2. تفسيرات الروابط: يوضح لماذا ترتبط الأحداث ببعضها. حيث يستخدم التصنيفات مثل "تشابه"، "موضوع مشترك"، "كيانات مشتركة". ويحتوي على تحليل SHAP لإظهار مساهمات الكلمات في تقوية أو إضعاف العلاقات بين الأحداث.
3. النموذج عالي المستوى: يحدد أسماء المحتوى باستخدام أسلوب يعتمد على تصنيف الكلمات. كذلك يساعد على استخراج الأحداث المهمة بناء على المحتوى أو البنية. كما يحدد الأحداث المحورية التي تربط الأجزاء المختلفة من المحتوى.

**مجالات تطبيق الذكاء الاصطناعي التفسيري((Explainable AI Application Domain**

شهدت السنوات الأخيرة نموًا متسارعًا في استخدام تقنيات الذكاء الاصطناعي التفسيري (XAI)، خاصة في المجالات التي تصنّف بأنها حساسة أو عالية الخطورة، مثل الصحة والبيئة والأمن، حيث تتطلب هذه المجالات درجة عالية من الشفافية في عمليات اتخاذ القرار، وذلك لتعزيز الثقة في مخرجات الأنظمة الذكية. ويوضح الشكل أدناه بعض المجالات التي برز فيها استخدام تقنية الذكاء الاصطناعي التفسيري كالتالي:

**الشكل (2): مجالات تطبيق الذكاء الاصطناعي التفسيري**

**المصدر: (Saarela & Podgorelec, 2024, pp. 7-9)**

ووفقًا للمراجعة المنهجية التي أجراها Saarela & Podgorelec (2024) فإن مجالات تطبيق الذكاء الاصطناعي التفسيري الأكثر شيوعًا تتمحور في الخمس المجالات التالية: المجال الصحي، مجال البيئة والزراعة، مجال الصناعة، مجال الأمن والدفاع، وأخيرا مجال التصنيع. (Saarela & Podgorelec, 2024, p. 11)

وفيما يلي سيتم استعراض لأبرز التطبيقات في هذه المجالات وفقًا للدراسة:

1. المجال الصحي (Health)

يمثل المجال الصحي أكثر المجالات استخدامًا للذكاء الاصطناعي التفسيري، وخاصة في استخدام هذه التقنية لأغراض التشخيص الطبي والتنبؤ، مثل اكتشاف وتصنيف الأمراض والتنبؤ بالانتكاسة وكذلك اتخاذ قرارات الجراحة.(Saarela & Podgorelec, 2024, p. 7) .

ومن أشهر تطبيقات استخدام (XAI)في الطب والتشخيص:

**الجدول (2): تطبيقات استخدام الذكاء الاصطناعي التفسيري في المجال الصحي**

| مجالات الاستخدام | أمثلة التطبيق |
|---|---|
| تشخيص مرض السرطان والتنبؤ به، وقرارات الجراحة | تشخيص سرطان الجلد، والتنبؤ بسرطان الثدي، تشخيص سرطان البروستاتا، والرئة، قرارات الجراحة لحالة سرطان المبيض |
| تشخيص كوفيد 19 | اكتشاف الفايروس، التنبؤ بالحاجة إلى العناية المركزة، تقييم الضغط النفسي |
| التصوير الطبي والتشخيص | التنبؤ بمرض الشريان التاجي من صورتروية عضلة القلب، الكشف عن الحالة الصحية لحديثي الولادة |
| الأمراض المزمنة | التنبؤ بمرض السكري، التنبؤ بالجلوكوما وتشخصيها |
| تطبيقات ادارية عامة | إدارة البيانات الطبية، تطوير نماذج التعلم الآلي في الطب، دراسة تطبيقات استدامة التكنولوجيا الذكية في المجال الصحي |
| سلوكيات صحية و إنسانية | دراسة سلوك الغذاء، تقييم تطور الأطفال، عوامل التأثير على استخدام السماعات الطبية |
| علوم الدماغ والأعصاب | تصنيف وتوقع مرض الزهايمر، تشخيص مرض باركنسون، تشخيص أورام الدماغ، اكتشاف وتفسير طيف التوحد |

**المصدر: (Saarela & Podgorelec, 2024, p. 7)**

2. مجال البيئة والزراعة (Environment & Agricultural)

يستخدم الذكاء الاصطناعي التفسيري على نحو متزايد في مراقبة التغيرات البيئية، والتنبؤ بالكوارث الطبيعية، وتحسين العمليات الزراعية .(Saarela & Podgorelec, 2024, p. 8)

ومن الأمثلة على تطبيقات الذكاء الاصطناعي التفسيري في هذا المجال:

**الجدول (3): تطبيقات استخدام الذكاء الاصطناعي التفسيري في مجال البيئة والزراعة**

| مجالات الاستخدام | أمثلة التطبيق |
|---|---|
| الكوارث الطبيعية | التنبؤ بحدوث الزلازل، وتقييم الاحتمالية المكانية لتأثيرات الزلازل |

| مجالات الاستخدام | أمثلة التطبيق |
|---|---|
| **الموارد المائية** | **مراقبة جودة المياه الجوفية، التنبؤ بأنماط دورات المحيطات، التنبؤ بجريان المياه الناتج عن ذوبان الثلوج** |
| **المراقبة البيئية** | **التنبؤ بتركيز المعادن الثقيلة في المياه الجوفية، التعرف على المناطق المصابة بخنافس اللحاء في الغابات** |
| **الزراعة** | **التهجين النباتي، كشف الأمراض النباتية، تشخيص الإجهاد النباتي، تحليل الجينوم النباتي** |

**المصدر: (Saarela & Podgorelec, 2024, p. 8)**

3. مجال الصناعة (Industry)

تستخدم تقنية الذكاء الاصطناعي التفسيري في الصناعة لتحسين العمليات الصناعية واكتشاف وتشخيص الأعطال، مثل تشخيص الأعطال الذكي في الروبوتات، وتحسين العمليات من ناحية تحديد وقت دورة العمل والتنبؤ بنسبة الإنتاج النهائي.

(Saarela & Podgorelec, 2024, p. 8)

4. مجال الأمن والدفاع (Security & Defense)

استخدمت تقنيات الذكاء الاصطناعي التفسيري على نطاق واسع لتعزيز تدابير وإجراءات الأمن السيبراني وخاصة في أنظمة كشف التسلل، وإدارة الثقة داخل الأنظمة، كشف البرمجيات الخبيثة، كشف الثغرات في الشفرات، كما تم استخدام تقنيات الذكاء الاصطناعي التفسيري للتحقق من الوجه والكشف عن الأصوات المزيفة. (Saarela & Podgorelec, 2024, p. 8)

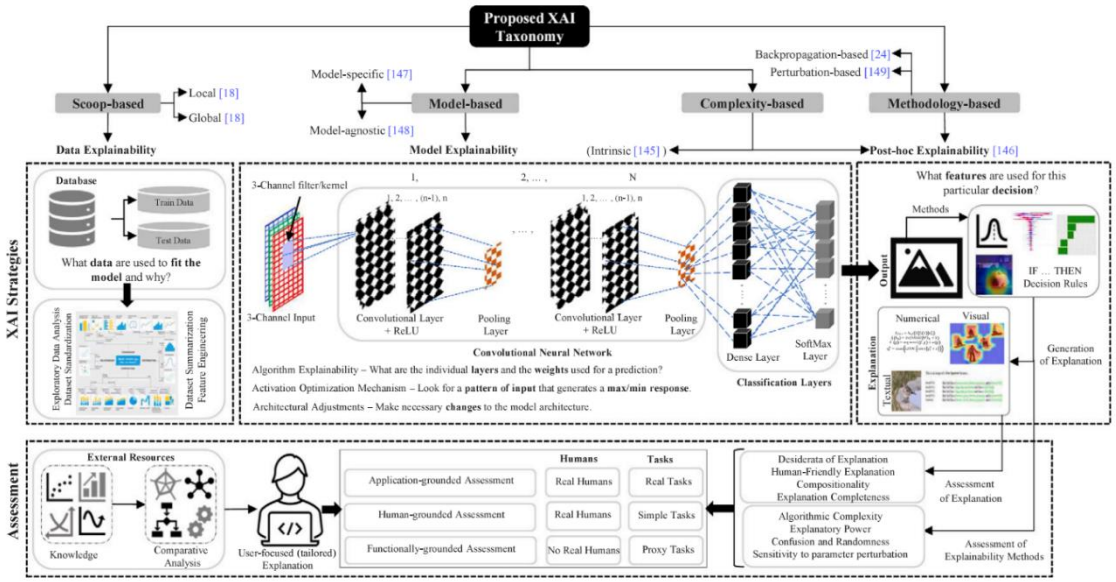
5. مجال التصنيع (Manufacturing)

تم توظيف تقنيات الذكاء الاصطناعي التفسيري في مجموعة متنوعة من المهام التنبؤية والتشخيصية في مجال التصنيع، حيث تم استخدامها في: تقدير العمر الافتراضي التنبؤي لمحركات التوربوفان، التنبؤ بالأعطال في الطابعات ثلاثية الأبعاد، كما تم استخدامها كذلك في المراقبة التنبؤية للعملية الإنتاجية، والصيانة التنبؤية في أنظمة التصنيع، هذا بالإضافة إلى استخدام تقنية الذكاء الاصطناعي التفسيري في نمذجة أداء أنظمة التبريد وإدارة الحرارة في عمليات التصنيع، والتنبؤ بخصائص قوة الخرسانة واكتشاف الشقوق فيها. (Saarela & Podgorelec, 2024, p.8)

بعد الاستعراض السابق، يتضح أن الذكاء الاصطناعي التفسيري يشكّل أداة محورية في دعم القرارات المبنية على البيانات خاصة في المجالات ذات الطابع عالي الخطورة الذي لا يُقبل فيه اتخاذ قرار دون تفسير واضح وشفاف، مثل الطب والأمن، وترى الباحثات من خلال العرض السابق أن هناك عددًا من المجالات مازالت تعد مجالات ناشئة في استخدام تقنيات الذكاء الاصطناعي التفسيري حيث ان استخداماتها في تلك المجالات مازالت محدودة مقارنةً بمجالات أخرى قطعت شوطًا متقدمًا في تطبيق تقنية XAI ومن هذه المجالات: التعليم، والقانون، والقطاع المالي.

**آلية عمل تقنية الذكاء الاصطناعي التفسيري (How Explainable AI works)**

من خلال البحث تبين أن تقنية الذكاء الاصطناعي التفسيري تعمل من خلال 4 مراحل موضحة في الشكل التالي:




**الشكل (3): آلية عمل تقنية الذكاء الاصطناعي التفسيري**

**المصدر: (Ali et al., 2023, p.10)**

1. المرحلة الأولى: تفسير البيانات (Data Explainability)

وتهدف هذه المرحلة الى عملية جمع البيانات فهم البيانات التي تم استخدامها والسبب في اختيارها ويتم فيها تحليل أولي للبيانات وتنظيف البيانات وتلخيصها واختيار الخصائص المؤثرة فيها وهذه المرحلة تساعد في فهم ما إذا كانت البيانات مناسبة منذ البداية. (Ali et al., 2023, p.10)

2. المرحلة الثانية: تفسير النموذج (Model Explainability)

وفيها يتم فهم كيفية عمل النموذج من الداخل بما فيها ادخال البيانات الى النموذج ومعرفة مكونات النموذج من طبقات ونوع الخوارزمية المستخدمة وتحليل كيفية تدفق البيانات داخل النموذج وتحديد الطبقات التي أثرت في القرار المتخذ وتعديل النموذج لجعله أكثر شفافية. (Ali et al., 2023, p.10)

3. المرحلة الثالثة: التفسير بعد اتخاذ القرار (Post-hoc Explainability)

ويتم في هذه المرحلة معرفة لماذا اتخذ هذا القرار بالتحديد من خلال تحليل النتائج النهائية وتوليد شرح للقرار باستخدام طرق متعددة كالتالي:

- طرق عددية: على شكل معادلات توضح التفسير الرقمي للقرار.
- طرق بصرية: وتكون على هيئة خرائط حرارية مثل (Grad-Cam)توضح المناطق المهمة بالصورة.
- طرق نصية: على شكل جمل تشرح القرار بلغة بشرية بسيطة.

وتستخدم هذه المرحلة قواعد If...Thenالفهم المسار المنطقي للقرار. (Ali et al., 2023, p.10)

4. المرحلة الرابعة: تقييم التفسير (Assessment of Explanation)

وفي هذه المرحلة يتم التأكد إذا كان التفسير المقدم واضح ومفهوم ومفيد للمستخدم ويكون التقييم بعدة أنواع:

- Application-grounded
- Human-grounded
- Functionally- grounded

**المعايير الخاصة بتقييم جودة التفسيرات المقدمة في الذكاء الاصطناعي التفسيري:**

- الشفافية :(Transparency) وتبين مدى قدرة المستخدم على رؤية ألية اتخاذ القرار داخل النظام.

- قابلية التفسير(Interpretability): توضح مدى سهولة فهم المستخدم للتفسير المقدم.
- الدقة :(Fidelity)وهي مدى مطابقة التفسير لسلوك النموذج الحقيقي دون تحريف.
- الاتساق :(Consistency) وتبين مدى ثبات التفسيرات عند تقديم مدخلات متشابهة.
- الاستقرار :(Stability) وتوضح مدى ثبات التفسير عند وجود تغيرات بسيطة في المدخلات.
- القابلية للنقل :(Portability) وتبين إمكانية تطبيق التفسير على نماذج أو بيئات مختلفة.
- الوضوح والايجاز (Clarity & Conciseness):أن تكون التفسيرات واضحة ومباشرة دون تعقيد مفرط.
- العدالة :(Fairness ) وهي خلو التفسير من التحيزات الضمنية او التمييز ضد مجموعات معينة.
- الخصوصية:(Privacy ) ويقصد بها حماية معلومات المستخدم الحساسة عند تقديم التفسيرات. .(Ali et al., 2023, pp.7-8)

من خلال الشرح السابق يتبين لنا أن آلية عمل تقنية الذكاء الاصطناعي التفسيري تتلخص في جمع البيانات وتفسيرها، تليها عملية تفسير النموذج، ثم تفسير القرار وأخيرا تقييم جودة التفسير. وبذلك نحصل على نظام ذكاء اصطناعي شفاف موثوق ويمكن الاعتماد عليه بثقة.

**تصنيفات تقنيات الذكاء الاصطناعي التفسيري ((Classification of XAI Techniques**

يمكن تصنيف تقنيات الذكاء الاصطناعي التفسيري والخوارزميات المرتبطة بها إلى فئتين رئيسيتين:

1. النماذج الشفافة :(Transparent Models) وهي نماذج تتيح فهما كاملا وواضحا لكيفية عمل الخوارزميات، حيث يمكنها شرح كيفية معالجة المدخلات والخطوات التي تؤدي إلى للوصول إلى نتائج معينة، بالإضافة لتفسير السبب وراء هذه المخرجات. ومن أبرز الخوارزميات المستخدمة في هذه الفئة:
   - النماذج الخطية مثل: الانحدار الخطي.
   - الأنظمة القائمة على القواعد.
   - أشجار القرار. (Hamm et al.2023, p.3)
2. نماذج الصندوق الأسود :(Black- Box Models) وهي نماذج تنشئ بنية داخلية معقدة لتوليد المخرجات، ولكنها غير مفهومة بشكل مباشر للمستخدمين أو حتى للمبرمجين. ومن أبرز الخوارزميات المرتبطة بهذا النموذج:
   - الشبكات العصبية الأمامية.
   - الشبكات العصبية الالتفافية.
   - الشبكات العصبية المتكررة.
   - الشبكات التوليدية التنافسية. (Hamm et al., 2023, p.3))

**متطلبات تقنية الذكاء الاصطناعي التفسيري (Explainable AI Requirements)**

- توفر البيانات وتجهيزها وفهما يتم فهم طبيعة وتوزيع البيانات الداخلة للنموذج وتهدف الى تعزيز الفهم والشفافية.
- وجود خوارزميات التفسير وتكون اما Intrinsic أو Post-hoc Explainabilityما بعد التفسير وعن طريقها يتم تفسير القرارات المتخذة وتستخدم تقنيات مثل LIMEو.SHAP
- توفر واجهات تفاعلية مع المستخدم تسمح بعرض التفسيرات بصيغ مرئيةVisual، رقمية Numerical ونصية Textual وذلك يعزز فهم المستخدمين غير المتخصصين.
- توفر التقييم وذلك لقياس جودة وكفاءة التفسيرات ويكون تقييم وظيفي Functionally-grounded Assessment، تقييم مع المستخدمين Human-grounded Assessment وتقييم تطبيقي(Ali et al., 2023, p.10) .Application-grounded Assessment

ومن متطلبات تقنية الذكاء الاصطناعي التفسيري الاخلاقية:

1. الدقة والصدق : (Accuracy and Truthfulness)يؤكد STODT et al. نقلاً عن Rong et al. أن الدقة والصدق عنصران أساسيان في تفسيرات الذكاء الاصطناعي، حيث يعززان فهم المستخدم وثقته بأنظمة الذكاء الاصطناعي وتوفر التفسيرات الدقيقة رؤى موثوقة مما يعزز الفهم والتواصل الفعالين بينما قد تؤدي التفسيرات المضللة إلى سوء الفهم، بالإضافة الى ذلك التفسيرات الدقيقة والصادقة تتميز بأن لها قيمة تعليمية حيث تمكن المستخدمين من فهم أنظمة الذكاء الاصطناعي بشكل أفضل وتعزز المشاركة

وتدعم اتخاذ القرارات المستنيرة، وبذلك فإن الثقة بالذكاء الاصطناعي ضرورية لتحقيق تجارب مستخدم إيجابية واعتماد حلول الذكاء الاصطناعي على نطاق واسع. (STODT et al., 2024, p. 566)

2. العبء المعرفي :(Cognitive Load) يؤكد .STODT et alنقلاً عن .Hoffman et alوHerm على أهمية تقليل العبء المعرفي، ورفع كفاءة أداء المهام، وتقليص مدة تنفيذها، وقد أظهرت الدراسة التي أجريت على 271 من الأطباء المستقبليين أن تفسيرات تقنيات الذكاء الاصطناعي التفسيري (XAI)تؤثر بشكل كبير في العبء المعرفي والأداء أثناء تنفيذ المهام، ويعد تحقيق توازن مناسب في العبء المعرفي أمراً حاسماً لتمثيل القرارات بدقة، وتعزيز فهم المستخدمين. (STODT et al., 2024, p. 566)

3. التركيز على الفهم: (Focus on Understanding)يشير .STODT et alنقلاً عن .Wang et al و .Arrieta et alإلى أن إعطاء الأولوية للفهم في تفسيرات الذكاء الاصطناعي التفسيري (XAI) يعد أمرًا بالغ الأهمية، فالتفسيرات الشفافة تعزز الثقة من خلال تمكين المستخدمين من التحقق من مصداقية الذكاء الاصطناعي وعدالته. كما أن فهم منطق اتخاذ القرار في الذكاء الاصطناعي يعزز من قدرة المستخدمين على التفاعل مما يسهم في الكشف عن الأخطاء وتصحيحها، وتعد التفسيرات الواضحة أدوات تعليمية قيمة تساعد على فهم عمليات اتخاذ القرار المعقدة. .(STODT et al., 2024, p.567)

4. ادراج المعرفة المتخصصة :(Inclusion of Domain Knowledge)يؤكد .STODT et alنقلاً عن .Panigutti et alعلى أن دمج المعرفة المتخصصة يعد أمراً بالغ الأهمية، فارتباط التفسير بسياق المجال يوضح سبب اتخاذ النماذج الصندوقية السوداء لقراراتها، مما يعزز فهم المستخدم وثقته بالنظام. وتسهم المعرفة المتخصصة في تبسيط المفاهيم المعقدة لغير المتخصصين، الأمر الذي يدعم تقبل المستخدم للنظام ويعزز ثقته به. كما تضمن هذه المعرفة أن تعكس التفسيرات بشكل دقيق آلية اتخاذ القرار داخل النموذج، مما يجعلها أكثر توافقاً مع الفهم البشري وأكثر سهولة في الاستيعاب، ويساعد أيضا في تحسين دعم اتخاذ القرار خصوصاً في المجالات المعقدة مثل الرعاية الصحية. (STODT et al., 2024, p. 567)

5. التمثيل غير الرمزي :(Non-Propositional Representation)ذكر .STODT et al نقلاً عن .Páezو.Mittelstadt et al الى استخدام الوسائل البصرية مثل الرسوم التوضيحية والمخططات والخرائط في التفسيرات، نظراً لكونها أكثر سهولة في الفهم من النصوص، وتقلل العبء المعرفي، مما يسهم في تحسين مستوى الاستيعاب، ويبرز .Kim et al فاعلية هذه الوسائل في تبسيط المعلومات المعقدة وتعزيز عملية التواصل. كما أن التصورات البصرية تسهم في زيادة تفاعل المستخدم واهتمامه، وتوفر الخصائص التفاعلية تجربة مخصصة لكل مستخدم. اضافة الى ذلك، فإن هذه الوسائل تتجاوز الحواجز اللغوية وتلائم الخلفيات المتنوعة، مما يجعلها متاحة على نطاق واسع. بالإضافة إلى أن هذا النهج العملي يُعزز من قابلية الاستخدام والوصول إلى تفسيرات الذكاء الاصطناعي التفسيري. (STODT et al., 2024, p. 567)

6. تخصيص التفسيرات وفقاً لخبرة المستخدم :(Tailoring to User Expertise)يعد تخصيص التفسيرات بما يتناسب مع مستوى خبرة المستخدم أمراً بالغ الأهمية، كما ناقشه .STODT et al نقلاً عن .Kim et al و ,.Nyrup, Miller, Langer et al.وRobinson فاختلاف مستويات الخبرة يتطلب اعتماد أساليب تفسيرية متنوعة، حيث تسهم التفسيرات البسيطة في دعم المستخدمين ذوي المعرفة المحدودة. بينما يستفيد الخبراء من الشروحات الأكثر تفصيلا. (STODT et al., 2024, p. 567)

من خلال التحليل تبين للباحثات أن الذكاء الاصطناعي التفسيري (XAI) يعتمد على متطلبات أو أبعاد تقنية تشمل تجهيز البيانات، خوارزميات التفسير، الواجهات التفاعلية، وآليات التقييم لضمان الفهم والشفافية، كما تبرز أهمية الالتزام بمعايير جودة التفسيرات مثل الشفافية والدقة والوضوح، إلى جانب الأخذ بالاعتبارات الأخلاقية، كالدقة وتقليل العبء المعرفي وتعزيز الفهم ودمج المعرفة المتخصصة وتخصيص التفسيرات حسب خبرة المستخدم.

**تقنيات الذكاء الاصطناعي التفسيري (How Explainable AI works)**

1. :Model-agnostic أكد .Mathew et alنقلاً عن .Dwivedi et alأنه من الممكن توظيف عدة استراتيجيات رئيسية ضمن تقنيات الذكاء الاصطناعي التفسيري (XAI)مثل الأساليب التفسيرية غير المرتبطة بنموذج معين، وذلك لجعل عمليات الذكاء الاصطناعي أكثر قابلية للتفسير. ويكمن الهدف من هذه الإستراتيجيات في تمكين المستخدم من تفسير وفهم قرارات النماذج المعقدة وذلك يسهم في تعزيز الشفافية ويقلل من غموض الصندوق الأسود المستخدم في الذكاء الاصطناعي (.Mathew et al., 2025, p. 2)

2. :Post-hoc Explanation يعد منهجاً تفسيرياً خارجياً يهدف إلى توضيح التنبؤات التي يصدرها. كما وضح .Mathew et al نقلاً عن Retzlaff .et alأن نموذج شجرة القرار يساعد في التوفيق بين المفاضلات الأساسية ضمن مختلف أساليب (XAI). ويعد هذا المنهج مفيداً بشكل خاص لتفسير النماذج المعقدة التي لا تتسم بقابلية التفسير مثل الشبكات العصبية العميقة أو نماذج التجميع. ,.Mathew et al., 2025) p. 5)

ومن التقنيات الشائعة التي تستخدم في التفسير بعد التدريب :Post-hoc Explanation

- Shapley Additive Explanations (SHAP)
- Local Interpretable Model-agnostic Explanations (LIME)
- وتستخدم جميعها لتمكين النماذج المعقدة من تقديم رؤى فيمة حول آلية اتخاذ القرار. (Mathew et al., 2025, p. 5)

3. Feature Importance Technique: أوضح Mathew et al. نقلاً عن Zacharias et al. أن هذه التقنية تهدف الى تعزيز عمق الشرح التفسيري من خلال تحديد الخصائص الأكثر تأثيراً في قرارات نموذج الذكاء الاصطناعي، فمن خلال تحديد المتغيرات التي تؤثر بشكل أكبر على تنبؤات النموذج ليكتسب المستخدمون فهماً أعمق للعوامل المحركة لسلوك النظام الذكي، وتتيح هذه التقنية للمستخدمين توجيه اهتمامهم نحو الخصائص ذات الصلة وفهم المنطق الكامن وراء اتخاذ القرار مما يسهم في تعزيز الشفافية وقابلية التفسير للنظام. (Mathew et al., 2025, pp. 5-6)
4. Visualization Technique: أشار Mathew et al. نقلا عن Minh et al. أن التصور البصري يلعب دوراً مهماً في تحقيق عمق الشرح التفسيري وذلك من خلال تقديم تمثيلات بديهية لسلوك نموذج الذكاء الاصطناعي وذلك عن طريق تحويل مخرجات النموذج المعقدة الى صيغ مرئية قابلة للفهم مثل المخططات البيانية والرسوم التوضيحية والخرائط الحرارية، مما يسهم في تمكين المستخدمين من فهم الأنماط والعلاقات التي يلتقطها النظام الذكي بشكل أكثر وضوحاً. (Mathew et al., 2025, pp. 5-6)
5. Counterfactual Explanation: يوضح Mathew et al. نقلاً عن Stepin et al. أن هذه التقنية تمثل مساراً اضافياً لتحقيق عمق الشرح التفسيري في تقنيات (XAI) من خلال تقديم رؤى قائمة على سيناريوهات "ماذا لو" إذ يتم توليد حالات افتراضية تتغير فيها مخرجات النموذج مما يتيح للمستخدمين فهماً أفضل لحدود اتخاذ القرار والعلاقات السببية ضمن النظام الذكي وتسهم في تمكين المستخدمين من استكشاف النتائج البديلة وفهم كيفية تأثير المتغيرات على تنبؤات النموذج. (Mathew et al., 2025, p. 6)

مما سبق تستخلص الباحثات أن الذكاء الاصطناعي التفسيري (XAI) يشمل عدة تقنيات رئيسية منها الأساليب غير المرتبطة بنموذج معين (Model-agnostic) والتفسير الخارجي بعد التدريب (Post-hoc Explanation) لتحسين الشفافية وفهم سلوك النماذج المعقدة. كما تعتمد (XAI) على تقنيات مثل Feature Importance، والتصور البصري (Visualization)، والتفسيرات الافتراضية (Counterfactual Explanation) لدعم التفاعل مع قرارات الذكاء الاصطناعي وفهم آلية اتخاذ القرار بوضوح.

**التحديات التي تواجه تقنية الذكاء الاصطناعي التفسيري (Explainable AI Challenges)**

رغم التقدّم الملحوظ في تقنيات الذكاء الاصطناعي التفسيري (XAI)، إلا أن هذه التقنية لا تزال تواجه عددًا من التحديات التي تعيق اعتمادها على نطاق واسع داخل المنظمات، خاصة في المجالات التي تتطلب شفافية وثقة عالية مثل إدارة المعرفة، وقد تناول Longo et al. (2024) في دراستهم التي تعد بيانًا بحثيًا (Manifesto) أبرز التحديات كالتالي:

1. التحديات التقنية :(Technical Challenges)

تشهد تقنيات الذكاء الاصطناعي تطورًا مستمرا مع ظهور أنواع جديدة من النماذج التي يتميز كل منها بمجموعة فريدة من الخائص، مما أبرز تحدي يواجه تقنيات الذكاء الاصطناعي التفسيري في إنشاء تفسيرات لهذه الأنواع الجديدة من نماذج الذكاء الاصطناعي. فعلى سبيل المثال: تمثل النماذج التوليدية Generative Models)) مثل نماذج (GPT) تحديًا كبيرا أمام تقنيات (XAI)حيث تُظهر هذه النماذج أداء استثنائيا بسبب حجمها الهائل الذي قد يصل إلى تريليونات من المعاملات، مما يجعلها تحديًا للأساليب التفسيرية الحالية. كما تعاني تقنيات (XAI)من صعوبات نتيجة للطبيعة متعددة المعاني العصبية (Polysemantic) حيث يمكن أن تمثل الخلية العصبية الواحدة عدة مفاهيم متداخلة مما يعقد تفسيرها، أما نماذج اللغة الكبيرة (LLMs) فتعاني تحد آخر يتمثل في قوانين التوسع العصبي (Neural Scaling Laws) حيث تشير هذه القوانين إلى علاقات وظيفية تربط بين متغيرات معينة في الشبكة العصبية، مثل عدد الطبقات في بنية النموذج والدقة التي يحققها بعد التدريب. وتتحكم هذه القوانين في القدرات الإجمالية لنماذج (LLMs)، إلا أن الفهم الدقيق لتأثيراتها على مستوى المهام الفردية لا يزال غامضًا، إذ تبدو هذه التأثيرات غير متوقعة في كثير من الأحيان. (Longo et al., 2024, p. 6)

وتزداد هذه التحديات التقنيّة تعقيدًا عند تطبيق تقنيات الذكاء الاصطناعي التفسيري في بيئات التعلم التعاوني (Collaborative Learning) التي تستخدم عادة لمشاركة المعرفة بين عدة كيانات مستقلة. ففي مثل هذه البيئات، يتم تحديث النماذج بشكل تكراري وغير متزامن مما يجعل من الصعب تتبع التغيرات وفهم تأثير مساهمة كل عقدة في النموذج، كما أن العقد في الأنظمة الموزعة قد تمتلك مجموعات بيانات وهياكل نماذج مختلفة، ومع زيادة عدد هذه العقد يصبح من الصعب تفسير القرارات التي يتخذها النموذج المدرب. (Longo et al., 2024, p. 7)

2. التحديات المفاهيمية :(Conceptual Challenges)

يُعد غياب تعريفات واضحة ومتسقة للمفاهيم الرئيسية في الذكاء الاصطناعي التفسيري من أبرز التحديات التي تعيق التقدم في تطوير أنظمة عملية وفعّالة، حيث توجد ضبابية مفاهيمية فيما يتعلق بالمفاهيم الأساسية في الذكاء الاصطناعي التفسيري مثل: القابلية للتفسير (Explainability)، التفسيرية (Interpretability)، الشفافية (Transparency)، الفهم (Understanding)، الإيضاحية (Explicability)، الوضوح (Perspicuity)، والإدراك (Intelligibility). إذ أن غياب تعريفات واضحة ومتسقة لهذه المصطلحات قد يعوق التقدم في تطوير أنظمة (XAI)عملية وذات قيمة، فبعض الباحثين يستخدم هذه المصطلحات كمترادفات، في حين يرى آخرون اختلافات جوهرية بينها، ويسبب هذه التباين مشكلات في الأبحاث التطبيقية ويعقد التعاون بين التخصصات المختلفة. (Longo et al., 2024, p. 8)

ويمتد الغموض المفاهيمي، إلى العلاقة بين الذكاء الاصطناعي التفسيري وبين الموثوقية (Trustworthiness)ففي بعض المصادر، يُنظر إلى الموثوقية كهدف مركزي لتقنيات (XAI)، بينما في مصادر أخرى، يُزعم أن (XAI)هو مجرد جزء من الموثوقية. كما يبرز تحدٍ آخر يرتبط بمفهوم الفهم نفسه ،(Understanding) فبينما يرى بعض الباحثين أن الفهم لا يتحقق إلا من خلال تفسيرات صحيحة بالكامل، يرى آخرون أن الفهم من الممكن أن يتحقق حتى عندما تكون بعض عناصر التفسير غير دقيقة أو مشوهة. (Longo et al., 2024, p. 9)

3. التحديات التنظيمية :(Regulatory Challenges)

تُواجه تقنيات الذكاء الاصطناعي التفسيري (XAI) تحديات تنظيمية أساسية، فعلى الرغم من الاهتمام الواسع بتصميم أساليب (XAI)، إلا أنها تفتقر إلى أساليب ومعايير موحدة للتقييم، خصوصًا فيما يتعلق بكيفية مقارنة نتائج التقييمات المختلفة أو بناء فهم مشترك حول كيفية تقييم التفسيرات. ويكمن النقص في غياب مجموعة من مؤشرات التقييم لقابلية التفسير تكون قابلة للتطبيق على نطاق واسع عبر الدراسات والسياقات والبيئات المختلفة. (Longo et al., 2024, p. 10)

كما تبرز مشكلة أخرى تتعلق بالقصور في الطرق الحالية لتقييم التفسيرات، حيث لا يوجد حتى الآن معيار ذهبي موحد يحدد بوضوح ما الذي يجعل التفسير جيدًا، حيث تعتمد معظم طرق التقييم الحالية على تحليل خصائص تقنية معينة لأساليب (XAI)نفسها فقط، دون الأخذ بعين الاعتبار تفاعل المستخدم النهائي مع التفسير. هذا بالإضافة إلى جانب آخر مهم لـ (XAI)هو حوكمة الذكاء الاصطناعي، والحاجة لتدابير تنظيمية ترتبط بإدارة المخاطر المصاحبة للذكاء الاصطناعي. مما يبرز تحديات أساسية لتصميم وتطوير ونشر أنظمة الذكاء الاصطناعي بأمان. (Longo et al., 2024, p. 9)

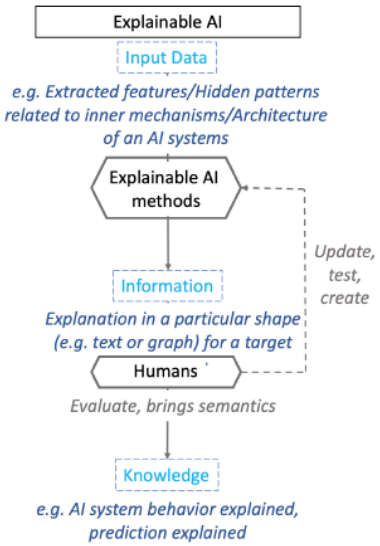
4. التحديات الأخلاقية :((Ethical Challenges

- خطر التفسيرات الخاطئة أو المضللة :(Failed & Misleading Explanations)تمثل التفسيرات غير الدقيقة تحديًا كبيرًا في أنظمة الذكاء الاصطناعي التفسيري، إذ تُقنع بعض تفسيرات النماذج الذكية المستخدمين بصحة نتائج خاطئة أو تُضعف ثقتهم في نتائج صحيحة بسبب ضعف أو غموض التفسير. هذا التحدي يعرف بمفارقة الصندوق الأبيض، حيث قد يكون الاستنتاج صحيحًا والتفسير ضعيفًا أو العكس، مما يؤدي إلى قرارات بشرية غير دقيقة أو فقدان الثقة تجاه الأنظمة الذكية. (Longo et al., 2024, p. 13)
- غياب معايير للتفنيد والتحقق :(Lack of Falsifiability)لا توجد حتى الآن معايير واضحة للحكم على صحة التفسيرات أو خطئها، ومتى يمكن دحضها (Falsifiable)، فإن غياب هذا المفهوم يجعل من الصعب محاسبة ممارسي الذكاء الاصطناعي أو تحميلهم المسؤولية، مما يزيد من الاعتماد المفرط على الحدس بدلًا من الأساليب المنهجية في تقديم التفسيرات. (Longo et al., 2024, p. 14)
- إمكانية استغلال التفسير من قبل أطراف بشرية خبيثة (Abuse by Malicious Human Agent) :قد تستغل قابلية التفسير من قبل بعض الجهات الخبيثة كأداة لإخفاء سلوكيات غير أخلاقية أو تمييزية ضمن نماذج الذكاء الاصطناعي، من خلال تقديم تفسيرات ظاهريا تبدو مقنعة لكنها تخفي التحيز أو التلاعب بالقرارات، على سبيل المثال: قد يقوم أصحاب العمل، بالتمييز المنهجي ضد المتقدمين للوظائف. (Longo et al., 2024, p. 14)
- أخطار التفسير في سياق الذكاء الاصطناعي الفائق :(Superintelligent AI)مع التطور المتسارع في قدرات أنظمة الذكاء الاصطناعي، قد تُكتشف طرقًا لخداع البشر عمدًا من خلال استغلال الفجوة المفاهيمية بين التفسيرات المستندة للإدراك البشري، وبين المفاهيم الأكثر تعقيدًا التي تستطيع الأنظمة الذكية استيعابها واستخدامها، وإذا حدث ذلك، وكان اعتماد البشر قائمًا فقط على قابلية التفسير كوسيلة لضمان السلامة، فقد يكون المكسب الخبيث الذي تحققه أنظمة الذكاء الاصطناعي غير محدود. (Longo et al., 2024, p. 14)

بعد الاستعراض السابق لأبرز التحديات، ترى الباحثات أن التحديات التقنية تمثل العائق الأكبر أمام دمج تقنيات الذكاء الاصطناعي التفسيري في إدارة المعرفة، خاصة في البيئات التعاونية المعتمدة على البيانات الموزعة بين أطراف متعددين، حيث تصبح القدرة على تتبع مصادر المعرفة وفهم طريقة عمل النماذج أصعب، مما يقلل من الاستفادة من تفسيرات الذكاء الاصطناعي التفسيري.

**مناقشة دورتقنية الذكاء الاصطناعي التفسيري (XAI)في دعم إدارة المعرفة وعملياتها**

- شرح سلوك النماذج الذكية بطريقة بسيطة: يتمثل الهدف من تقنية الذكاء الاصطناعي التفسيري في تفسير الأسباب الكامنة وكشف المنطق وراء سلوك نماذج الذكاء الاصطناعي، وجعل هذه التفسيرات بسيطة ومفهومة للجمهور المستهدف. (Kaadoud et al., 2021, p. 31)
- تمثيل المعرفة بطريقة مفهومة: تسهم تقنيات الذكاء الاصطناعي التفسيري في دعم عملية تمثيل المعرفة من خلال عرض أسباب ونتائج النماذج بطريقة مفهومة للجمهور البشري المستهدف أو لاستخدام الآلة. (Kaadoud et al., 2021, p. 31)
- دعم اتخاذ القرار: يتم تشكيل المعرفة والمحتوى في الذكاء الاصطناعي التفسيري ضمن مجالي اكتشاف المعرفة وتمثيلها بما يتناسب مع الجمهور المستهدف والسياق، وذلك وفقًا للغرض من الاستخدام مثل: التحقق من الفرضيات، أو الاستدلال، أو دعم اتخاذ القرار. (Kaadoud et al., 2021, p. 32)
- مصدر إلهام من مجال المعرفة: يُستوحى الذكاء الاصطناعي التفسيري بشكل غير مباشر من مجال المعرفة، مما يعزز ارتباطه بممارسات إدارة المعرفة حيث يسعى كلاهما إلى التعبير عن المعلومات المستخلصة من البيانات، كما هو الحال في تقنيات تنقيب البيانات واكتشاف المعرفة، وبالتالي فإن تطور مجال المعرفة يعد مصدرًا مهمًا لتطوير تقنيات الذكاء الاصطناعي التفسيري. (Kaadoud et al., 2021, p. 35)
- تقليل التحيز وتعزيز العدالة: تواجه إدارة المعرفة في سياق الذكاء الاصطناعي تحديات متزايدة في تقديم تفسيرات قائمة على البيانات تتوافق مع توقعات أصحاب المصلحة وسياقاتهم المختلفة، إضافة إلى تحديات أخلاقية ترتبط بالتحيز والعدالة، مما يبرز الدور المحوري للذكاء الاصطناعي التفسيري والعنصر البشري في معالجة هذه التحديات. ((Kaadoud et al., 2021, p. 35
- يتبيّن دور الذكاء الاصطناعي التفسيري في كل من (اكتشاف المعرفة: من خلال تفسير الأنماط في البيانات، تمثيل المعرفة من خلال تحويل التفسيرات إلى معلومات قابلة للفهم، دعم اتخاذ القرار: حين تصل المعرفة المفهومة إلى الإنسان) في الشكل التالي:




**الشكل (4): تحويل البيانات إلى معرفة: رؤية تخطيطية باستخدام الذكاء الاصطناعي التفسيري**

**المصدر: (Kaadoud et al., 2021, p. 32)**

- تسهم تقنية الذكاء الاصطناعي التفسيري في دعم عملية اكتشاف المعرفة (Knowledge Discovery): تتكون عملية اكتشاف المعرفة من ثلاث مراحل رئيسية، مرحلة ما قبل المعالجة (يتم فيها جمع البيانات وتنظيفها وتحضيرها)، مرحلة المعالجة (يستخدم فيها تقنيات متعددة من الإحصاء أو تنقيب البيانات أو تعلم الآلة)، مرحلة ما بعد المعالجة وهي التي يتم فيها التصوّر والتقييم والتحقق من قبل البشر. (Kaadoud et al., 2021, p. 30)وتعد هذه المرحلة الأخيرة نقطة هامة في تحويل البيانات إلى معرفة حيث أن المعلومات المستخرجة يتم تقييمها عادة من قبل البشر وبالتالي فإن الجمهور المستهدف لاكتشاف المعرفة هو الإنسان: خبراء المجال ومتخذو القرار، وهنا يبرز دور استخدام تقنيات الذكاء الاصطناعي التفسيري التي تمكّن من اكتشاف المعرفة الكامنة داخل النماذج المعقدة التي تعرف بالصناديق السوداء Black-Boxesمثل الشبكات العصبية العميقة Deep Neural Networks، من خلال تحليل المدخلات وتحديد العوامل المؤثرة في اتخاذ القرار(Features) بالإضافة إلى تحليل أنماط التنشيط داخل الطبقات المخفية (Features) وذلك لاكتشاف أنماط خفية لم تكن واضحة من خلال النتائج وحدها. ويتماشى ذلك مع الهدفين الرئيسين لاكتشاف المعرفة كما ذكر

.Kaadoud et al نقلًا عن .Fayyad et al وهما التحقق من فرضية المستخدم، واكتشاف معرفة جديدة صالحة ومفيدة تكون مفهومة في ضوء البيانات التي تم استخراجها من النماذج .(Kaadoud et al., 2021, p. 31)

وبالتالي، تستخلص الباحثات أن مساعدة صانعي القرار والخبراء في فهم لماذا؟ وكيف؟ توصّل النموذج إلى نتيجة معينة يسهم في دعم عملية اتخاذ القرار لاحقًا، استنادًا إلى المعرفة التي تم اكتشافها وتفسيرها باستخدام هذه التقنية، مما يعكس القيمة التطبيقية لاستخدام تقنيات (XAI)ضمن اطار عملية اكتشاف المعرفة. ولتحقيق ذلك، تستخدم (XAI)مجموعة من التقنيات التفسيرية من أبرزها تقنيات ما بعد البناء Post-hoc techniques، وتشمل هذه التقنيات التفسيرية الشائعة: LIMEو.(Kaadoud et al., 2021, p. 28) SHAP

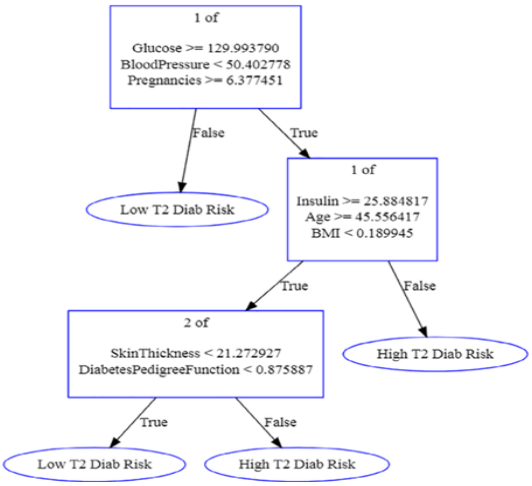
- تسهم تقنية الذكاء الاصطناعي التفسيري في دعم عملية تمثيل المعرفة (Knowledge Representation): تهدف عملية تمثيل المعرفة في سياق الذكاء الاصطناعي التفسيري إلى تحويل الأسباب الكامنة المكتشفة من داخل نماذج الذكاء الاصطناعي المعقدة إلى تمثيلات مفهومة ليستفيد منها الجمهور المستهدف، الذي قد يكون: خبيرا بشريا يستخدم تمثيل المعرفة بهدف الفهم واتخاذ القرار، أو نظام ذكاء اصطناعي آخر يستخدم هذه البيانات، كما أن تمثيل المعرفة يرتبط ارتباطا وثيقًا باسترجاع المعرفة (Knowledge Retrieval) في شكل أنطولوجيات تستخدم لتمثيل وتخزين والوصول إلى المعرفة (Kaadoud et al., 2021, p. 31). ويمكن عرض التفسيرات المستخرجة من تقنيات الذكاء الاصطناعي التفسيري بصيغ متعددة مثل اللغة الطبيعية (Natural Language)، التفسيرات البصرية (Visual Explanations،أو خرائط الأهمية (Kaadoud et al., 2021, pp. 28-29) .(Saliency Mapping))
- وتتمثل إحدى الاستراتيجيات المتاحة لتمثيل المعرفة في المثال المذكور في مجال نماذج رؤية الحاسوب، حيث يتم استخدام الشبكات العصبية العميقة لتصنيف الصور مع دمج تقنيات التفسير في النموذج، فينتج النموذج نتيجة التصنيف بالإضافة إلى تفسير مرئي ولفظي لهذا القرار وذلك من خلال تطبيق تقنية (Saliency masks) على الصور وتوليد النصوص التفسيرية. (Kaadoud et al., 2021, p. 31)
- وبناء على ماورد في الدراسة، تستخلص الباحثات أن هذه التفسيرات المستخرجة باستخدام تقنيات الذكاء الاصطناعي التفسيري يمكن تخزينها أو إعادة تنظيمها داخل نظم معرفية مثل الأنطولوجيات أو قواعد المعرفة مما يسهل استخدامها وتداولها ضمن نظم إدارة المعرفة المؤسسية.
- تسهم تقنية الذكاء الاصطناعي التفسيري في دعم عملية استخراج المعرفة (:(Knowledge Extraction

تهدف معظم الأساليب التقليدية للتفسير في الأنظمة العصبية الرمزية إلى إنشاء مجموعة من القواعد الرمزية التي تكون قريبة من سلوك النموذج الرمزي الفرعي. تعرف مهمة توليد هذه القواعد عادة من خلال شكل من أشكال التعلم باسم استخراج المعرفة وتهدف عملية الاستخراج إلى تحسين المقاييس والمعايير المختلفة وهي الدقة والاتساق والفهم.

يمكن بعد ذلك الاستفادة من القواعد المستخرجة لمراجعة وتوحيد المعرفة الأساسية المتاحة للمجال وإدخال المعرفة في النموذج الرمزي الفرعي للتحسين في الأداء. في استخراج المعرفة يتم التركيز عادةً على نهجين رئيسيين: الأول هو النهج الوضعي التفكيكي، والثاني هو النهج التربوي. في النهج الأول تقوم الخوارزميات باستخراج القواعد مباشرة من بنية وأوزان النموذج الرمزي الفرعي يتم ذلك عادةً من خلال تحليل سلوك كل وحدة اتصالية، ثم يتم تجميع هذه القواعد على مستوى الوحدة لتكوين قاعدة مركبة تعكس عمل الشبكة العصبية ككل وقد تحتاج هذه الأساليب إلى الوصول إلى الطبقات الداخلية للنموذج الرمزي الفرعي، وهو ما قد يكون صعبًا بسبب اعتبارات الملكية الفكرية. ومع ذلك قد يكون هناك حاجة لاستخلاص تفسيرات وهنا تأتي أهمية الخوارزميات التي تنتمي إلى فئة الأساليب التربوية، حيث تتعامل مع النموذج الرمزي الفرعي كـ "أوراكل" وتستخرج المعلومات من أزواج المدخلات والمخرجات. ومن الأمثلة البارزة على النهج التربوي في استخراج المعرفة (Trepan)، وهي خوارزمية لتحفيز الأشجار، تقوم باستخراج أشجار القرار بشكل متكرر من المصنفات الإحصائية. وقد وضح ((Confalonieri, 2021 أنه يمكن اعتبار النهج المتبع عند كلاً من Cravenو(1995) Shavlik بمثابة امتداد لخوارزمية ID2-of-3حيث تُعتبر هذه الطريقة وسيلة لبناء أشجار القرار من البيانات استنادًا إلى قواعد "m-of-n"، حيث يجب أن تكون m من n الشروط المحددة صحيحة لإرسال مثال إلى فرع معين. عادةً ما تُنشأ هذه الاختبارات بواسطة خوارزمية تبدأ بميزة فردية تعزز من اكتساب المعلومات، وتضيف مميزات بشكل متكرر إلى الاختبار حتى يتوقف تحسين اكتساب المعلومات يجمع Trepan بين هذه الفكرة واستخدام مصنف التعلم الآلي المدرب المعروف باسم أوراكل. وبينما يقوم Trepan باستخراج الأشجار من النماذج الرمزية الفرعية من خلال تقريب النماذج بصعوبة تامة دون الوصول المباشر إلى هندستها المعمارية ووحداتها، تظل هناك تحديات في تقييم مدى فهم الإنسان لهذه الأشجار المستخرجة. وتم إجراء بحث حديث لقياس الفهم البشري لأشجار القرار باستخدام مقاييس نحوية ومعرفية بناءً على ذلك، أظهر كيف يمكن تعزيز الفهم البشري للأشجار البديلة من خلال دمج المعرفة المتخصصة، مثل استخدام الأنطولوجيا في عملية استخراج شجرة القرار. (Confalonieri, 2021, p. 13))

وتُعتبر أشجار القرار من أكثر التقنيات شيوعًا في مجال التعلم الآلي، وذلك بفضل بساطتها وهيكلها السهل الفهم. على عكس نماذج الانحدار الخطي أو اللوجستي، تستطيع أشجار القرار تقريب العلاقات غير الخطية، مع الحفاظ على قابلية التفسير من خلال منطق بسيط

يعتمد على الشروط if-else. ومع ذلك، تتناقص شفافية نماذج الأشجار مع زيادة عمقها. حيث إن نماذج الأشجار الجماعية الشائعة مثل XGBoost أو نماذج الأشجار المعززة للتدرج تقضي تمامًا على أي شكل من أشكال القابلية للتفسير. يمكن للمستخدمين فهم أشجار القرار من خلال تتبع المسار من عقدة الجذر وصولًا إلى كل عقدة ورقية، حيث يمكن تفسير النتيجة ببساطة على أنها "إذا كان x أكبر من/أقل من عتبة معينة وx أكبر من/أقل من عتبة أخرى، فإن الإخراج هو"(Yeo et al., 2025, p. 32) ."Y




**الشكل (5): شجرة Trepan مستخرجة من شبكة عصبية مدربة تتنبأ بمخاطر الإصابة بمرض السكري بناء على مجموعة بيانات (Pima Indians dataset)**

**المصدر (Confalonieri, 2021, p. 14) :**

- تبرز العلاقة الوثيقة بين إدارة المعرفة(Knowledge Management) والذكاء الاصطناعي التفسيري من خلال الدور الوسيط الذي تؤديه الأنطولوجيات في كلا المجالين. إذ تُعد الأنطولوجيات بمثابة البنية التحتية المعرفية التي تتيح تمثيل وتنظيم وتبادل المعرفة بشكل منهجي، وهو ما يُعد من الأهداف الجوهرية لإدارة المعرفة. وفي الوقت نفسه، تساهم الأنطولوجيات في تعزيز قابلية تفسير مخرجات الذكاء الاصطناعي عبر:

1. النمذجة المرجعية :(Reference Modelling) التي تساعد في صياغة المتطلبات المعرفية للنظم وتوحيد المفاهيم، مما يُعزز إعادة استخدام المعرفة ويوفر بيئة شفافة للتفسير.
2. الاستدلال المعتمد على المعرفة العامة :(Common-Sense Reasoning) حيث يثري التفسيرات بالسياق والدلالة، ويُحسن من قدرة الأنظمة على توصيل المعرفة بطريقة مفهومة للمستخدمين.
3. إدارة التعقيد وتنقيح المعرفة عبر آليات التجريد والتفصيل :(Knowledge Refinement and Complexity Management) التي تتيح تخصيص الشروحات حسب خلفية المستخدم واحتياجاته، بما يدعم فهمًا أعمق للمحتوى المعرفي.

وبذلك، فإن الذكاء الاصطناعي التفسيري (XAI)لا يُسهم فقط في توضيح قرارات الأنظمة الذكية، بل يعزز أيضًا من كفاءة إدارة المعرفة من خلال تقديم تفسيرات قابلة للتخصيص، قائمة على المعرفة، وسهلة الفهم. إن التكامل بين الأنطولوجيات و (XAI)يوفّر بيئة معرفية شاملة تدعم اتخاذ القرار، وتُسهّل مشاركة وإعادة استخدام المعرفة داخل المؤسسات. (Confalonieri & Guizzardi, 2023, p. 4)

**الدراسات السابقة**

1. دراسة بعنوان" Explainable AI: A narrative review at the crossroad of knowledge discovery knowledge representation and representation learning"، التي قدمها Kaadoud وآخرون في عام .(2021)وتقدم الدراسة مراجعة سردية للأدبيات المتعلقة بمجالات اكتشاف المعرفة وتمثيل المعرفة، وتعلّم التمثيل ،(Representation Learning) وذلك بهدف استكشاف أوجه التشابه ونقاط الالتقاء بين هذه المجالات وبين تقنية الذكاء الاصطناعي التفسيري. وقد خلصت الدراسة إلى أن تعزيز شفافية نماذج (الصندوق الأسود) يتطلب أن تستلهم مناهج (XAI)من الأعمال السابقة والحديثة في مجالي المعرفة وتعلم التمثيل. وتعد هذه النتائج ذات أهمية في سياق إدارة المعرفة، حيث تبرز كيف يمكن لتقنية الذكاء الاصطناعي التفسيري أن تسهم في تحويل المعرفة المستخرجة من نماذج الذكاء الاصطناعي إلى معرفة قابلة للفهم والاستخدام من قبل صانعي القرار داخل المنظمات، مما يعزز من فعالية عمليتي اكتشاف وتمثيل المعرفة.
2. دراسة بعنوان""Explainable artificial intelligence: a comprehensive review التي قدمها Minh وآخرون في عام .(2022) تهدف الدراسة إلى مراجعة شاملة للأبحاث وذلك لاكتشاف الجوانب الرئيسية للبحوث المتزايدة المتعلقة بالذكاء الاصطناعي التفسيري (XAI)

وتحليل ومراجعة مختلف الأساليب المرتبطة به والتي يتم تصنيفها إلى: قابلية التفسير قبل بناء النموذج (Pre-modeling explainability) والنماذج القابلة للتفسير(Interpretable models) وقابلية التفسير بعد بناء النموذج (Post-modeling explainability). كما أبرزت الاهتمام بالأساليب الحالية التي تهدف إلى تفسير وتحليل أساليب التعلم العميق ومناقشة منهجي التحديات المختلفة المرتبطة بـ (XAI) مثل التوازن بين الأداء وقابلية التفسير، وطرق التقييم، والأمان، والسياسات، وعرض الطرق القياسية المستخدمة للتعامل مع هذه التحديات.

3. دراسة بعنوان "Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare"التي قدمها Hulsen في عام (2023). تناولت الدراسة المفاهيم الأساسية لتقنيات الذكاء الاصطناعي التفسيري (XAI)، مع تركيز خاص على تطبيقاتها في قطاع الرعاية الصحية. هدفت الدراسة إلى تحليل كيفية جعل خوارزميات الذكاء الاصطناعي أكثر قابلية للفهم البشري، بما يعزز الثقة في نتائجها، لاسيما في البيئات عالية الحساسية مثل المجال الطبي، حيث تُتخذ قرارات بالغة الأثر على حياة المرضى. وقد قدمت الدراسة تحليلًا نقديًا للتحديات الرئيسية المرتبطة باستخدام (XAI)في هذا السياق، مشددةً على دور الشفافية والتعاون البشري الآلي في تحسين فعالية تلك الأنظمة.

4. دراسة بعنوان "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities" التي قدمها Saeed & Omlin في عام (2023). قدّمت الدراسة مراجعة منهجية للتحديات الحالية والفرص المستقبلية المرتبطة بتطوير تقنيات .(XAI) صنّف الباحثان التحديات إلى فئتين رئيسيتين: التحديات العامة التي تعيق التوسع في استخدام (XAI)، والتحديات الخاصة بكل مرحلة من مراحل تطوير الأنظمة التفسيرية. وتُعد هذه الدراسة مرجعًا مهمًا لفهم التحديات الحالية، كما تسلط الضوء على اتجاهات مستقبلية واعدة من شأنها أن تعزز فهم المستخدمين لتقنيات الذكاء الاصطناعي وتزيد من موثوقيتها.

5. دراسة بعنوان " Explainable Artificial Intelligence (XAI) 2.0: A Manifesto of Open Challenges and Interdisciplinary Research Directions"، التي قدمها Longo وآخرون في عام (2024). والتي تسلّط الضوء على التقدّم المتسارع في مجال الذكاء الاصطناعي التفسيري (XAI) وعلى تطبيقاته المتنوعة في البيئات الواقعية. وقد أكّد الباحثون على أن فهم نماذج "الصندوق الأسود" أصبح ضرورة ملحة في ظل ازدهار أنظمة ونماذج الذكاء الاصطناعي غير الشفافة في مختلف القطاعات. وتهدف هذه الدراسة إلى تقديم بيان بحثي شامل يتضمن 28 تحديًا مفتوحًا مصنفة ضمن تسع فئات رئيسية، تعكس تعقيدات وتحديات تبنّي (XAI)على أرض الواقع. ومن خلال توظيف منهجية تعاونية متعددة التخصصات، اقترح الباحثون اتجاهات بحثية واعدة لكل تحدٍ، بهدف تسريع وتيرة البحث وتوحيد الجهود البحثية في هذا المجال. وتبرز أهمية هذه الدراسة في سياق إدارة المعرفة، إذ توفر خريطة طريق منهجية تساهم في معالجة الإشكاليات المتعلقة بفهم وتفسير المعرفة.

6. دراسة بعنوان"" Explainable AI in Manufacturing and Industrial Cyber–Physical Systems: A Survey التي قدمها Moosavi وآخرون في عام (2024). تهدف الدراسة إلى استعراض تطبيقات الذكاء الاصطناعي التفسيري (XAI) في مجال التصنيع والأنظمة السيبرانية الفيزيائية الصناعية (ICPS). مع استمرار التقدم التكنولوجي في دمج الذكاء الاصطناعي في البنية التحتية الحيوية والعمليات الصناعية يصبح من الضروري تطوير نماذج ذكية واضحة ومفهومة، تلعب تقنيات (XAI)دوراً محورياً في تعزيز موثوقية الأنظمة الذكية المستخدمة في الصناعة، مما يضمن قدرة المشغلين البشريين على فهم القرارات التي تتخذها هذه الأنظمة والتحقق منها، حيث تناول البحث توضيح الحاجة الملحة إلى الذكاء الاصطناعي التفسيري، وتصنيف تقنيات (XAI)بشكل منهجي، كما يستكشف مجموعة واسعة من التطبيقات الصناعية التي تستفيد من (XAI)مثل الصيانة التنبؤية، الأمن السيبراني، اكتشاف الأعطال وتشخصيها، التحكم في العمليات، تطوير المنتجات، إدارة المخزون، وضمان جودة المنتجات. قدمت هذه الدراسة فهم شامل للإستراتيجيات والمنهجيات المختلفة المستخدمة في دمج تقنيات الذكاء الاصطناعي التفسيري ضمن السياقات الصناعية.

**التعليق على الدراسات السابقة**

**الجدول (1): أوجه الشبه والاختلاف بين الدراسات السابقة والدراسة الحالية**

| م | الدراسة | أوجه الشبه | أوجه الاختلاف |
|---|---|---|---|
| 1 | **دراسة Kaadoud وآخرون (2021)** | **توضيح أهمية تقنية (XAI) وعلاقتها بعمليتي اكتشاف المعرفة وتمثيل المعرفة** | **الدراسة تركز على دورتقنية الذكاء الاصطناعي التفسيري في عمليتين من عمليات إدارة المعرفة أما الدراسة الحالية فهي تتجاوزهذا النطاق من خلال تناول دور (XAI)في طيف أوسع من العمليات، بما في ذلك استخراج المعرفة وتقييمها وتطبيقها في المجالات المختلفة** |

| م | الدراسة | أوجه الشبه | أوجه الاختلاف |
|---|---|---|---|
| 2 | دراسة Minah وآخرون (2021) | تقدم الدراسة خلفية عن الذكاء الاصطناعي التفسيري وأهميته وتحدياته. | تركز الدراسة الحالية على كيفية استخدام الذكاء الاصطناعي التفسيري في تحسن عمليات إدارة المعرفة وكيفية استخراج المعرفة من البيانات الضخمة باستخدام تقنيات (XAI)مما يساعد على اتخاذ قرارات أكثر شفافية. |
| 3 | دراسة Hulsen (2023) | تركز الدراسة على توضيح مفاهيم الذكاء الاصطناعي التفسيري (XAI) في الرعاية الصحية، وتناقش كيف يمكن زيادة فهم وثقة الأطباء والمرضى بقرارات الذكاء الاصطناعي، مع تسليط الضوء على التحديات. | تختلف عن الدراسة الحالية من حيث تركيزها القطاعي، إذ اقتصر نطاقها على الرعاية الصحية فقط، بينما تتبنى الدراسة الحالية مقاربة أوسع تشمل مجالات متعددة، لا سيما في إدارة المعرفة. كما أن دراسة Hulsen تناولت التطبيقات العملية في التشخيص الطبي ودعم القرار، بينما تركز الدراسة الحالية على عمليات مثل اكتشاف المعرفة، التمثيل، والتطبيق ضمن نظم معرفية متنوعة. |
| 4 | دراسة Saeed & Omlin (2023) | تقدم الدراسة التحديات والاتجاهات المستقبلية في مجال الذكاء الاصطناعي التفسيري. | تختلف الدراسة في أن تركيزها كان عامًا وغير مرتبط بمجال محدد، حيث هدفت إلى تقديم تصنيف شامل للتحديات الحالية والفرص المستقبلية في XAI، بينما ركّزت الدراسة الحالية على تطبيق XAI في دعم عمليات إدارة المعرفة بشكل خاص مثل اكتشاف المعرفة وتمثيلها. كما أن منهجيتها كانت مراجعة ميتا منهجية، في حين أن الدراسة الحالية اعتمدت على تحليل نظري واستقرائي. |
| 5 | دراسة Longo وآخرون (2024) | تركز الدراسة على التحديات التي تواجه تقنيات الذكاء الاصطناعي التفسيري وكيفية التغلب عليها | الدراسة تركز على التحديات التي تواجه تقنية (XAI)في البيئات المختلفة دون أن تركز بشكل مباشر على تطبيقات (XAI)في مجال إدارة المعرفة. |
| 6 | دراسة Moosavi وآخرون (2024) | تركز الدراسة على استعراض دور الذكاء الاصطناعي التفسيري ودوره في تحسين موثوقية الأنظمة الذكية والقدرة على فهم القرارات. | ركزت الدراسة على التصنيع والأنظمة السيبرانية الفيزيائية الصناعية، حيث يتم استخدام (XAI)لتحسين الصيانة التنبؤية، الأمن السيبراني، اكتشاف الأعطال، التحكم في العمليات، وتطوير المنتجات.<br>بينما تناولت الدراسة الحالية كيفية استخدام (XAI)في عمليات استخراج المعرفة وتمثيلها واكتشافها والتنبؤ بالقرارات المستقبلية للمعرفة المطلوبة. |

وخلصت الباحثات بعد استعراض تلك الدراسات إلى ما يلي:

أن معظم الدراسات السابقة تناولت تقنيات الذكاء الاصطناعي التفسيري (XAI)من زوايا محددة، كالتحديات التقنية، أو التطبيقات القطاعية مثل الصناعة أو الرعاية الصحية، أو ركّزت على مراحل محددة من دورة إدارة المعرفة، مثل اكتشاف أو تمثيل المعرفة.

في المقابل، تميزت الدراسة الحالية بشمولها وامتدادها، إذ تناولت دور (XAI) في دعم طيف واسع من عمليات إدارة المعرفة، بدءًا من استخراج المعرفة وتفسيرها، مرورًا بتقييمها وتطبيقها، وصولًا إلى دعم اتخاذ القرار المعرفي في بيئات متعددة. كما ركزت على الاستخدام العملي والتكاملي لتقنيات (XAI)ضمن نظم إدارة المعرفة، مما يمنحها بعدًا تطبيقيًا واستراتيجيًا في آنٍ معًا، لا سيما في عصر البيانات الضخمة والتحول الرقمي.

وهكذا، تُسهم الدراسة الحالية في سد فجوة بحثية تتعلق بتكامل تقنيات (XAI)مع عمليات إدارة المعرفة، مما يميزها عن الدراسات السابقة التي كانت غالبًا إما نظرية، أو قطاعية، أو مركزة على أحد جوانب الذكاء الاصطناعي التفسيري دون ربطه المتعمق بإدارة المعرفة.

## خلاصة ونتائج الدراسة

**أولاً: النتائج**

1. تبيّن وجود فجوة بحثية واضحة في ربط تقنيات الذكاء الاصطناعي التفسيري بإدارة المعرفة، حيث تركز معظم الدراسات الحالية على الجانب التقني للتفسير، دون توضيح كافي كيف يمكن استخدام وتوظيف هذه التفسيرات بشكل عملي داخل أنظمة إدارة المعرفة أو دعم اتخاذ القرار في المنظمات. وهذا يُظهر حاجة واضحة إلى دراسات تطبيقية تربط بين (XAI)وعمليات إدارة المعرفة بشكل عملي.
2. تسهم تقنية الذكاء الاصطناعي التفسيري في دعم إدارة المعرفة: تشكّل تقنيات الذكاء الاصطناعي التفسيري جسرًا يربط بين تقنيات ونماذج الذكاء الاصطناعي وبين إدارة المعرفة وأنظمتها، مما يحدث أثرًا ملموسًا في نقل التفسيرات القابلة للفهم إلى أنظمة إدارة المعرفة ودعم اتخاذ قرارات مؤسسية تحقق توازنًا جيدًا بين الدقة والشفافية اعتمادًا على التفسيرات التي قدمتها هذه التقنية.
3. تسهم تقنية الذكاء الاصطناعي التفسيري في تعزيز عملية اكتشاف المعرفة: أضافت تقنية الذكاء الاصطناعي التفسيري قيمة كبيرة لمرحلة ما بعد المعالجة في عملية اكتشاف المعرفة، وذلك عن طريق تقديم تفسير لنتائج النماذج المعقدة (مثل الشبكات العصبية)، حيث يتم استخدام أدوات تفسيرية مدعومة بتقنيات مثل LIMEوSHAP لتوضيح الأسباب الخفية وراء نتائج التنبؤات أو القرارات، وساهم ذلك في تفسير الأنماط المكتشفة، وتوضيح العلاقات أو الفجوات المعرفية التي لم تكن واضحة من خلال النتائج، واتخاذ القرارات بناء على التفسيرات وليس فقط نتائج النماذج.
4. تسهم تقنية الذكاء الاصطناعي التفسيري في تعزيز عملية تمثيل المعرفة: تسهم تقنيات الذكاء الاصطناعي التفسيري في تحويل مخرجات النماذج إلى أشكال متعددة من التمثيلات (بصرية، لفظية أو رمزية) بحيث تكون مفهومة للبشر أو قابلة للتكامل مع أنظمة الذكاء الاصطناعي الأخرى، ويسهم هذا التنوع في تسهيل إدراج هذه التمثيلات داخل قواعد المعرفة والأنطولوجيات والرسوم البيانية المعرفية (Knowledge Graph)مما يساعد على بناء ذاكرة تنظيمية واسترجاع المعرفة.
5. تسهم تقنية الذكاء الاصطناعي التفسيري في تعزيز عملية استخراج المعرفة: وذلك من خلال خوارزمية Trepan التي تقوم باستخلاص واستخراج المعرفة وتمثيلها في شكل انطولوجيا قابلة للقراءة والفهم، وذلك من خلال أشجار القرارات التي تعتمد على تفسيرات النماذج وتبسيطها.
6. إن أهم ما يتميز به الذكاء الاصطناعي التفسيري (XAI) هما: الشفافية في صنع القرارات والثقة، وتعتبر مميزات فريدة غير موجود في الذكاء الاصطناعي AI.
7. يكمن دور الأساليب غير المرتبطة بنموذج معين Model-gnosticفي جعل العمليات أكثر قابلية للتفسير.
8. تتلخص متطلبات الذكاء الاصطناعي التفسيري في متطلبات أخلاقية ومنها الدقة والصدق وتجنب العبء المعرفي ومتطلبات تقنية ومنها توفر البيانات ووجود خوارزميات التفسير وتوفر واجهات تفاعلية وذلك يساعد في مشاركة المعرفة.

**ثانيا: التوصيات**

1. تطوير حلول معرفية تدمج بين الرسوم البيانية المعرفية (Knowledge Graphs) والتقنيات الناشئة مثل البلوك تشين، بهدف تمثيل التفسيرات الناتجة عن تقنيات الذكاء الاصطناعي التفسيري (XAI) بطريقة تضمن توثيق مصدر المعرفة، وتتبّع تطورها لمنع التلاعب بها أو فقدانها مع مرور الوقت.
2. تطوير أدوات وتقنيات تفسيرية مرنة تتكيف مع احتياجات المستخدم النهائي، بحيث تكون قادرة على التعامل مع معلومات متنوعة من مصادر متعددة داخل المنظمة، مثل قواعد البيانات، تقارير الأداء.
3. اعتماد واجهات تفاعلية ذكية مثل أنظمة الذكاء الاصطناعي التفسيري الحوارية Conversational AI التي لا تزال قيد البحث والتطوير، وذلك لما لها من إمكانات واعدة من الممكن أن تسهم في تحسين تجربة المستخدم وفهمه للتفسيرات، مما يعزز من فعالية تمثيل المعرفة ونقلها داخل المؤسسات، ويقلل من مشكلة الاعتماد الزائد Over Relianceعلى الأنظمة الآلية.
4. وضع ضوابط أخلاقية من ضمن سياسة قواعد السلوك (Code of Conduct) في منظمات إدارة المعرفة، وتوضيح من يتحمل المسؤولية عن التفسيرات الخاطئة أو سوء الاستخدام خاصة إذا كانت تتعلق بقرارات مصيرية مثل الرعاية الصحية، أو الأمن.
5. تكثيف الجهود البحثية لتطوير أدوات تقييم معيارية لقياس جودة التفسيرات المقدمة من أنظمة الذكاء الاصطناعي التفسيري، بحيث تراعي هذه الأدوات اختلافات الجمهور المستهدف وكذلك تنوع السياقات التنظيمية التي تطبّق فيها تقنيات إدارة المعرفة.
6. توسيع نطاق الأبحاث المستقبلية لتشمل مجالات جديدة وناشئة مثل التعليم والقانون مما يسهم في تعميم الاستفادة من التفسيرية والشفافية التي تتميز بها هذه التقنية في قطاعات متعددة.

7. توحيد تعريفات المصطلحات المرتبطة بتقنية الذكاء الاصطناعي مثل القابلية للتفسير، التفسيرية، الفهم، والثقة.
8. ضرورة توظيف تقنيات الذكاء الاصطناعي التفسيري في عملية تنظيم المعرفة وتصنيفها وتخزين مصادر المعرفة الضمنية.
9. المساهمة في إنشاء نظم لتحليل وإدارة المعرفة وعملياتها من توليد المعرفة إلى تطبيق المعرفة بالاعتماد على الذكاء الاصطناعي التفسيري (XAI).
10. ضرورة دمج تقنيات الذكاء الاصطناعي التفسيري في منصات إدارة المعرفة بهدف تحسين فهم واسترجاع المعرفة وذلك لسرعة الوصول الى المعلومات المعتمدة على الذكاء الاصطناعي.
11. يعد استخدام التصورات البيانية في عرض التفسيرات تبسيطاً لفهم المعرفة وذلك لتسهيل نقل المعرفة ومشاركتها.

## الخاتمة

تعد تقنية الذكاء الاصطناعي التفسيري من التقنيات الحديثة الناشئة والواعدة، والتي تمتلك القدرة على إحداث تغيير جذري في عمليات إدارة المعرفة داخل المنظمات، وذلك لما توفره من أساليب متقدمة تحسن من القدرات التنظيمية في معالجة المعرفة واستخراجها بشكل سهل وبسيط، مما يساهم في رفع جودة اتخاذ القرارات وتحقيق ميزة تنافسية للمنظمات. وتناول هذا البحث استعراضًا شاملًا لمفهوم الذكاء الاصطناعي التفسيري بدءًا من تعريفه، الغرض منه، وخصائصه، ومكوناته، مرورًا بتوضيح أبرز مجالات تطبيقه، ومتطلباته التقنية، وآلية عمله، وانتهاءً بتسليط الضوء على بعض التحديات التي تواجه هذه التقنية الحديثة، وأخيراً مناقشة دور تقنية الذكاء الاصطناعي التفسيري في دعم إدارة المعرفة وعملياتها.

ومن خلال هذا الاستعراض اتضح أن تقنية (XAI)تساهم بشكل كبير في ثلاث عمليّات رئيسية لإدارة المعرفة، أولًا استخراج المعرفة من خلال تحويل البيانات المعقدة إلى معلومات واضحة ومفهومة، وثانيًا اكتشاف المعرفة عبر تفسير الأنماط والعلاقات المخفية داخل النماذج الذكية، وأخيرا تمثيل المعرفة من خلال تحويل التفسيرات الناتجة إلى صيغ قابلة للفهم بسهولة. كما تبين أن تقنيات الذكاء الاصطناعي التفسيري تستخدم بشكل متزايد في مجالات عديدة وخاصة مجال الصحة، والبيئة، والأمن وهي مجالات تتطلب مستويات عالية من الشفافية والدقة في اتخاذ القرارات، ولكن مازالت هناك بعض المجالات التي لا تزال تعد مجالات ناشئة في تطبيق هذه التقنية مثل القانون، التعليم والقطاع المالي، مما يشير إلى فرص مستقبلية للتوسع في هذه المجالات.

ورغم الدور المهم لهذه التقنية إلا أنها لا تزال تقنية حديثة تواجه تحديات متعددة تعيق دمجها بشكل فعّال في بعض البيئات مثل بيئة إدارة المعرفة، كما أن نجاح تقنية الذكاء الاصطناعي التفسيري لا يعتمد فقط على فعالية التقنية ذاتها، بل يتطلب تطوير أطر تنظيمية وأخلاقية متكاملة تضمن استخداما مسؤولًا لهذه التقنية بشكل أكثر شفافية وموثوقية. وعلى رأس هذه التحديات تأتي التحديات التقنية المرتبطة بالقدرة على تفسير النماذج المتقدمة، وتحديات أخرى تتعلق بغياب معايير موحّدة للتقييم رغم الجهود البحثية في هذا المجال إلا أنها غير كافية لضمان جودة التفسيرات المقدمة، أو مراعاة المستخدم النهائي، فضلًا عن التحديات الأخلاقية التي ترتبط بإمكانية استغلال وإساءة استخدام التفسيرات لأغراض خبيثة مما قد يعرّض المستخدمين والمجتمع لأخطار بسبب هذه التفسيرات المضللة. وفي الختام، يوصي البحث بضرورة تطوير أدوات تفسير متقدمة قادرة على التعامل مع النماذج المعقدة التي تتطور بشكل مستمر، بالإضافة إلى تعزيز الوعي الأخلاقي لدى المستخدمين لضمان الاستخدام المسؤول للتفسيرات.

## المراجع


- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion, 99*, 101805. https://doi.org/10.1016/j.inffus.2023.101805
- Alkhanbouli, R., Almadhaani, H. M. A., Alhosani, F., & Simsekler, M. C. E. (2025). The role of explainable artificial intelligence in disease prediction: A systematic literature review and future research directions. *BMC Medical Informatics and Decision Making, 25*(1), 110. https://doi.org/10.1186/s12911-025-02944-6
- Confalonieri R, Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable artificial intelligence. *WIREs Data Mining and Knowledge Discovery, 11*(1). https://doi.org/10.1002/widm.1391
- Confalonieri, R. & Guizzardi, G. (2023). On the Multiple Roles of Ontologies in Explainable AI. https://doi.org/10.48550/arXiv.2311.04778

- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.Z. (2019). XAI-Explainable artificial intelligence. *Science Robotics. 4*(37). https://doi.org/10.1126/scirobotics.aay7120
- Hamm, P., Klesel, M., Coberger, P., & Wittmann, H. F. (2023). Explanation matters: An experimental study on explainable AI. *Electronic Markets, 33*(1), 17. https://doi.org/10.1007/s12525-023-00640-9
- Hulsen, T. (2023). Explainable artificial intelligence (XAI): Concepts and challenges in healthcare. *AI, 4*(3), 652-666. https://doi.org/10.3390/ai4030034
- Kaadoud, I. C., Fahed, L., & Lenca, P. (2021). Explainable AI: A narrative review at the crossroad of knowledge discovery, knowledge representation and representation learning. In *Proceedings of the Twelfth International Workshop Modelling and Reasoning in Context (MRC 2021).* (pp. 28-40). Montréal, Canada (virtual). CEUR Workshop Proceedings. Available at https://ceur-ws.org/Vol-2995/paper4.pdf (Accessed: 16 April 2025).
- Keith, B., German, F., Krokos, E., Joseph, S., & North, C. (2025). Explainable AI Components for Narrative Map Extraction. https://doi.org/10.48550/arXiv.2503.16554
- Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion, 106*, 102301. https://doi.org/10.1016/j.inffus.2024.102301
- Longo, L., Goebel, R., Lecue, F., Kieseberg, P., & Holzinger, A. (2020). Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions (pp. 1–16). https://doi.org/10.1007/978-3-030-57321-8_1
- Majumder, S., Dey, N. (2022). Explainable Artificial Intelligence (XAI) for Knowledge Management (KM). In AI-empowered Knowledge Management (KM) (pp. 101-104). Studies in Big Data, vol 107. Springer, Singapore. https://doi.org/10.1007/978-981-19-0316-8_6
- Mathew, D.E., Ebem, D.U., Ikegwu, A.C. Ukeoma, P. E., & Dibiaezue, N. F. (2025). Recent Emerging Techniques in Explainable Artificial Intelligence to Enhance the Interpretable and Understanding of AI Models for Human. *Neural Process Lett* 57(1), 16. https://doi.org/10.1007/s11063-025-11732-2
- Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management, 39*(1), 53-63. https://doi.org/10.1080/10580530.2020.1849465
- Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: A comprehensive review. *Artificial Intelligence Review, 55*(5), 3503–3568. https://doi.org/10.1007/s10462-021-10088-y
- Moosavi, S., Farajzadeh-Zanjani, M., Razavi-Far, R., Palade, V., & Saif, M. (2024). Explainable AI in Manufacturing and Industrial Cyber–Physical Systems: A Survey. *Electronics*, *13*(17), 3497. https://doi.org/10.3390/electronics13173497
- Ridley, M. (2024). Human-centered explainable artificial intelligence: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology, 76*(1), 98-120. https://doi.org/10.1002/asi.24889
- Saarela, M., & Podgorelec, V. (2024). Recent applications of explainable AI (XAI): A systematic literature review. *Applied Sciences, 14*(19), 8884. https://doi.org/10.3390/app14198884
- Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems, 263*, 110273. https://doi.org/10.1016/j.knosys.2023.110273
- Stodt, J., Reich, C., & Knahl, M. (2024). Demystifying XAI: Requirements for understandable XAI explanations. *Studies in Health Technology and Informatics*, 316, 565–569. https://doi.org/10.3233/SHTI240477
- Yeo, W. J., Van Der Heever, W., Mao, R., Cambria, E., Satapathy, R., & Mengaldo, G. (2025). A comprehensive review on financial explainable AI. *Artificial Intelligence Review, 58 (6)*, 189. https://doi.org/10.1007/s10462-024-11077-7