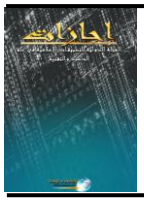
مؤسسة النهضة العلمية



توظيف البلوك–تشين في خدمة اللغة العربية: إطار مفاهيمي وتطبيقي لتعزيز التوثيق الرقمي

جميل اطميزي[1]، محمد طميزه[2]

[1] و [2] كلية الهندسة وتكنولوجيا المعلومات، جامعة فلسطين الاهلية، بيت لحم، فلسطين

[1] j.itmazi@gmail.com ، [2] mohammedt@paluniv.edu.ps

الخلاصة: شهدت اللغة العربية خلال العقدين الأخيرين توسعاً رقمياً ملحوظاً انعكس في ازدهار المحتوى التعليمي والأدبي والبحثي والإعلامي، غير أن هذا النمو السريع أفرز تحديات محورية تتعلق بموثوقية المحتوى، وحماية الملكية الفكرية، واعتماد معاجم موحّدة، وتوثيق النصوص، وضمان جودة البيانات اللغوية المستخدمة في تطبيقات الذكاء الاصطناعي. وفي هذا الإطار، تبرز تقنية البلوك–تشين بوصفها نموذجاً ثورياً قائماً على اللامركزية وعدم قابلية البيانات للتعديل، بما يتيح تعزيز الشفافية وبناء الثقة في المنظومات الرقمية. وتهدف الدراسة إلى استقصاء إمكانات توظيف هذه التقنية في خدمة اللغة العربية عبر بناء إطار نظري وتطبيقي شامل يتضمن نماذج عملية قابلة للتطبيق، ويناقش التحديات والعوامل المؤسسية والتقنية المؤثرة في تبنيها. وتخلص النتائج إلى أن دمج البلوك–تشين في إدارة الموارد اللغوية العربية قد يحدث تحولاً نوعياً عبر رفع جودة المحتوى، وحماية حقوق المؤلفين، ودعم تطوير حلول ذكاء اصطناعي أكثر دقة وموثوقية. وقد اعتمدت الدراسة منهجية البحث في تصميم الأنظمة (DSR)، وقدمت إطاراً مفاهيمياً يربط بين المدخلات والعمليات والتطبيقات والمخرجات، مستندة إلى مراجعة أدبية واسعة وتحليل معمق للتحديات، ومختتمة بجملة من التوصيات التي تعزز الاستدامة الرقمية للغة العربية.

الكلمات الجوهرية: تقنية البلوك–تشين، المحتوى العربي الرقمي، التوثيق الرقمي، سلامة البيانات اللغوية، الملكية الفكرية.

1. المقدمة

وتشير التقارير الدولية والإقليمية إلى تنامٍ واضح في إنتاج ونشر المحتوى العربي الرقمي خلال السنوات الأخيرة، مدفوعاً بالانتشار الواسع لاستخدام الإنترنت والمنصّات الرقمية في المنطقة العربية، سواء في التواصل الاجتماعي أو المجالات التعليمية والبحثية (ESCWA, 2023). كما تؤكد الجهات الأممية—مثل لجنة الأمم المتحدة الاقتصادية والاجتماعية لغربي آسيا (الإسكوا) واليونسكو—أن هناك اهتماماً متزايداً بتطوير المحتوى العربي وتوسيعه، إلى جانب استمرار الفجوة بين الحجم الفعلي للمحتوى المتاح على الشبكة وعدد الناطقين بالعربية (ESCWA, 2013; UNESCO, 2023). وتشير دراسات حديثة إلى أن المحتوى الرقمي التعليمي العربي ينمو لكنه ما يزال غير متوازن بين الدول العربية، مما يعكس الحاجة إلى بنى تحتية أقوى لإنتاج ونشر الموارد الرقمية (Zawacki-Richter et al., 2020). وتوصي هذه المؤسسات بضرورة تعزيز المبادرات الرقمية العربية وتحسين بنية إنتاج المحتوى لضمان حضور أكبر وأكثر تأثيراً للغة العربية في الفضاء الرقمي العالمي (UNESCO, 2023). ويواكب هذا التوسع تطور كبير في أدوات معالجة اللغة الطبيعية (NLP) والتعلم العميق، واتساع رقعة المستخدمين والتطبيقات الرقمية التي تعتمد على المصادر اللغوية العربية.

ومع هذا التوسع والتطور، ظهرت تحديات جوهرية تتعلق بالحوكمة الرقمية وجودة المحتوى، وهي تحديات وثقتها تقارير ودراسات متعددة حول حالة المحتوى العربي الرقمي. فقد أشارت تقارير الإسكوا إلى أن صناعة المحتوى العربي ما تزال تعاني من ضعف في البنى التنظيمية ونماذج الأعمال، إضافة إلى قصور في سياسات دعم المحتوى وضمان جودته (ESCWA, 2012; ESCWA, 2013). وتتمثل أبرز هذه التحديات في ضعف التوثيق الرقمي للنصوص العربية وانتشار السرقات الأدبية والانتحال الرقمي، وهو ما أكدته دراسات متخصصة بيّنت أن البيئة الإلكترونية العربية تشهد نسباً مرتفعة من النسخ غير الموثّق والتلاعب بالنصوص وصعوبة التحقق من أصالتها (Farahat & Hamouda, 2015; Al-Thwaib et al., 2020). كما تواجه المنظومة اللغوية تحديًا يتمثل في غياب معايير موحدة لاعتماد النصوص والمعاجم الإلكترونية، حيث تُظهر أبحاث المعالجة الحاسوبية والمعاجم الرقمية العربية وجود تباين كبير في الهياكل المعجمية وغياب نماذج معيارية للتوثيق والضبط الإملائي (Baccar Ben Amar et al., 2008; Habash et al., 2018). ويُضاف إلى ذلك عدم توفر سجلات موثوقة للمؤلفات العربية وصعوبة تتبع مصادر المعلومات وتاريخها وتطوّرها، الأمر الذي تؤكده اليونسكو والإسكوا في توصياتهما الداعية إلى بناء منظومة عربية أكثر موثوقية لحفظ المحتوى وتوثيقه. ونتيجة لذلك، يبرز غياب نظام مركزي أو لامركزي موثوق لاعتماد جودة المحتوى العربي وضمان أصالته وموثوقيته في البيئة الرقمية.

في هذا السياق المعقد، يظهر دور تقنية البلوك–تشين (Blockchain) كأحد أهم التقنيات الثورية التي يمكن استثمارها لتطوير المنظومة اللغوية العربية. فالبلوك–تشين (Nakamoto, 2008) يقدم بديلاً نوعيًا لتوثيق البيانات والنصوص، ويمتاز بخصائص مناسبة جدًا للمحتوى اللغوي العربي، أبرزها: اللامركزية، الشفافية، التتبع الزمني، وعدم قابليّة التغيير أو عدم القدرة على التغيير (Immutability)، إلى جانب تنفيذ العقود الذكية(Zheng et al., 2020) .

إن هذه الخصائص تجعل البلوك–تشين مناسبًا لتوثيق النصوص وضمان أصالة البيانات اللغوية، وحماية الأعمال الفكرية والحقوق الفكرية، وتنظيم عمليات الاعتماد، وتمكين المتخصصين من التأكد من مصدر النصوص وتاريخها حيث تسعى هذه الدراسة إلى استكشاف كيف يمكن توظيف تقنية البلوك–تشين بشكل فعّال لتعزيز المحتوى العربي، وضمان جودة البيانات، وحماية الحقوق الفكرية، وتطوير منظومة لغوية رقمية مستدامة، بما يعزز الحوكمة الرقمية في المؤسسات اللغوية.

2. مشكلة الدراسة

تتمحور مشكلة الدراسة في سياق المحتوى العربي الرقمي حول غياب نظام رقمي موثوق وموحّد لتوثيق النصوص وإدارة المحتوى اللغوي. يهدف هذا النظام إلى حماية أعمال المؤلفين وضمان جودة البيانات المستخدمة في التطبيقات التقنية الحديثة. ينتج عن هذا الغياب مجموعة من الإشكالات الجوهرية التي تعيق التطور والموثوقية الأكاديمية والرقمية للمحتوى العربي:

1. ضعف آليات التوثيق وانتشار الانتحال: تكمن إحدى المشاكل الرئيسية في ضعف آليات التوثيق الرقمي، حيث لا تحمل العديد من النصوص تاريخًا واضحًا أو سجلًا لتعديلاتها، مما يسهل التلاعب بها. يؤدي غياب سجل موثوق إلى انتشار الانتحال وصعوبة إثبات أحقية المؤلف في أعماله.
2. التحديات المعجمية واللغوية: هناك مشكلة واضحة في عدم وجود معيار اعتماد موحّد للمعاجم، حيث تختلف المداخل اللغوية بين المعاجم، وتغيب آليات موحدة لاعتماد التعديلات.

3. ضعف الحماية القانونية والهوية الرقمية: يعاني المحتوى العربي من فقدان الهوية الرقمية، مما يؤدي إلى ضعف الحماية القانونية ويعقّد مسألة حماية الحقوق الخاصة بالمؤلفين والناشرين.
4. نقص البيانات الموثوقة للذكاء الاصطناعي: تبرز الحاجة المُلحة للتوثيق في سياق التكنولوجيا الحديثة، حيث يوجد غياب لبيانات موثوقة يمكن استخدامها لتدريب نماذج الذكاء الاصطناعي. وتشير الدراسات إلى أن جودة هذه النماذج تتأثر بشكل مباشر بجودة البيانات (Wang et al., 2022)، مما يجعل الحاجة للتوثيق أمرًا ضروريًا.

إن غياب هذا النظام الموثوق والموحد يؤثر سلبًا على عدة جوانب حيوية:

- دقة المحتوى العربي وموثوقية المخطوطات.
- جودة النماذج اللغوية المبنية على البيانات.
- حماية المؤلفين وضمان حقوقهم.
- اعتماد الشهادات والمعاجم العربية.

خلاصة القول، تتمثل مشكلة الدراسة في المعضلة الشاملة لافتقار البيئة الرقمية العربية لنظام متكامل يضمن الموثوقية اللغوية، ويدعم الأمن القانوني، ويكفل جودة البيانات اللازمة للتطبيقات التقنية المتقدمة.

الفجوة البحثية: تتمثل فجوة البحث في غياب منظومة رقمية عربية موثوقة توفّر توثيقاً دائماً ومحدّثاً للمحتوى العربي، وهو ما تحاول الدراسة معالجته عبر استكشاف إمكانات البلوك–تشين.

3. أهمية الدراسة

1.3 الأهمية النظرية

- تفتح الدراسة مجالًا جديدًا يجمع بين اللغويات الحاسوبية وتقنيات البلوك–تشين، وهو مجال حديث نسبيًا.
- دعم جهود بناء موارد عربية رقمية معتمدة.

2.3 الأهمية التطبيقية

- يمكن تطبيق مخرجات البحث في الجامعات، المجامع اللغوية، دور النشر، المؤسسات الحكومية، وشركات التقنية.
- تطوير نظم معاجم موثقة ومعتمدة.
- حماية حقوق المؤلفين والباحثين.

3.3 الأهمية المستقبلية

- مع ظهور الجيل الثالث من الويب (Web3) والذكاء الاصطناعي اللغوي، تصبح الحاجة إلى بيانات موثوقة أمرًا أساسيًا.
- إنشاء قاعدة بيانات عربية موثوقة تخدم الذكاء الاصطناعي الناطق بالعربية.
- تعزيز الحوكمة الرقمية في المؤسسات اللغوية.

4. أهداف الدراسة وأسئلتها

1.4 أهداف الدراسة

1) تشخيص التحديات الرقمية التي تعترض مسيرة اللغة العربية في الفضاء الرقمي، وتحليل جذورها وآثارها.
2) تحديد وتحليل التحديات الرقمية المرتبطة بتوثيق المحتوى العربي.
3) تقييم إمكانات البلوك–تشين في معالجة فجوات التوثيق الرقمي.
4) بناء إطار نظري متكامل يربط بين المدخلات والعمليات والمخرجات.
5) تطوير نموذج تطبيقي يوضح آليات توظيف البلوك–تشين في مجالات لغوية مختارة.

2.4 أسئلة الدراسة

1. كيف يمكن للبلوك–تشين تحسين توثيق النصوص العربية؟
2. كيف يمكن للتقنية حماية حقوق المؤلفين؟
3. ما التطبيقات الممكنة في مجال المعاجم؟
4. كيف يمكن توثيق شهادات تعلم العربية؟
5. كيف يمكن ضمان جودة بيانات تدريب الذكاء الاصطناعي؟

5. الإطار النظري والتكنولوجي

1.5 مفهوم البلوك–تشين (Blockchain)

تُعرّف تقنية البلوك–تشين على أنها قاعدة بيانات موزعة (Distributed Ledger Technology – DLT)، تعمل بشكل لامركزي عبر شبكة من العُقد(Nodes). تتمثل وظيفتها الأساسية في تخزين المعاملات أو البيانات في "كُتل" مرتبطة ببعضها البعض باستخدام التشفير(Cryptographic Hashing)، مما يخلق سلسلة بيانات زمنية غير قابلة للتغيير [Tapscott & Tapscott, 2016]. إن أهم ما يميز هذه التقنية هو أنها لا تتطلب وسيطًا مركزيًا أو جهة موثوقة لإدارة وتأكيد البيانات. وبدلاً من ذلك، تعتمد على آليات التوافق (Consensus Mechanisms) بين المشاركين لضمان صحة البيانات المضافة إلى السلسلة، مما يضمن مستوى عالٍ من الثقة والموثوقية.

2.5 الخصائص الجوهرية لتقنية البلوك–تشين

توفر بنية البلوك–تشين عددًا من الخصائص التقنية التي تجعلها حلاً مثاليًا لتحديات الموثوقية والحفظ الرقمي:

I. اللامركزية (Decentralization): تُعد اللامركزية خاصية محورية، حيث يتم توزيع سجل البيانات عبر الشبكة، مما يعني إزالة الجهة المركزية التي قد تتلاعب بالمحتوى أو تكون نقطة فشل وحيدة. وهذا التوزيع يزيد من مقاومة النظام للرقابة والتلاعب.

II. الشفافية (Transparency): تضمن هذه الخاصية أن تكون جميع العمليات المُسجلة على السلسلة قابلة للعرض والتتبع من قبل جميع المشاركين في الشبكة. وعلى الرغم من أن هوية المشاركين قد تكون مُعمّاة، إلا أن التحقق من صحة المعاملات ومسارها الزمني متاح للجميع، مما يعزز الثقة في سلامة السجل.

III. عدم قابلية التعديل (Immutability): تُعد عدم قابلية التعديل أساس الموثوقية في البلوك–تشين. بمجرد تسجيل الكتلة والتحقق منها وإضافتها إلى السلسلة، يصبح تغييرها مستحيلاً من الناحية العملية بسبب الترابط التشفيري. يضمن هذا أن البيانات المُسجلة (مثل سجلات التوثيق) تبقى محفوظة بشكلها الأصلي إلى الأبد.

IV. العقود الذكية (Smart Contracts): العقود الذكية هي برامج حاسوبية ذاتية التنفيذ تُخزن وتُنفّذ على شبكة البلوك–تشين وتعمل وفق شروط محددة مسبقًا. تتيح هذه العقود أتمتة الإجراءات دون الحاجة إلى تدخل طرف ثالث، مثل تنفيذ شروط حقوق النشر أو إدارة سجلات التوثيق تلقائيًا.

❖ **السجلات الزمنية:**

تُعدّ السجلات الزمنية (Timestamping) إحدى الركائز الجوهرية في تقنية البلوك–تشين، حيث تمنح كل كتلة من البيانات «ختم وقت» دقيق وغير قابل للتعديل يحدد لحظة إنشاء المحتوى أو تعديله. وتُشكّل هذه الخاصية قيمة مضافة كبيرة في سياق اللغة العربية؛ إذ تمكّن الباحثين والمؤلفين والمؤسسات اللغوية من:

1. **إثبات أسبقية النشر: ضمان حقّ المؤلّف في نصّه الأصلي من خلال إثبات تاريخ قيامه بإنشاء المحتوى (جابر، 2020) .**
2. **توثيق مراحل التطوير: تسجيل جميع التعديلات والتحديثات على النصوص والمعاجم، مما يوفر "سجلاً تطورياً" يمكن الرجوع إليه (سيد، 2020) .**
3. **مقاومة الانتحال والتلاعب: لأن أي محاولة لاحقة لتغيير النص لا تتطابق مع الختم الزمني الأصلي، مما يكشف التلاعب فورًا.**
4. **تحسين موثوقية البيانات اللغوية:** يُسهّل تتبّع النسخة المعتمدة من النصّ، ويمكّن الباحثين من بناء نماذج لغوية تعتمد على بيانات نظيفة ومحدّدة التاريخ.
5. **توحيد النسخ في المؤسسات اللغوية:** السماح بإدارة نسخ متعددة من نفس المعجم أو النص، مع توثيق تاريخ كل تحديث واعتماد النسخة الرسمية.

وبالنظر إلى أن المحتوى العربي غالبًا ما يعاني من غياب ختم وقت واضح أو توثيق مؤسسي معيّن، فإن دمج خاصية الختم الزمني ضمن منظومة بلوك–تشين يمثل خطوة استراتيجية لتعزيز مصداقية المحتوى وحفظ الحقوق اللغوية والفكرية.

3.5 الإمكانات التقنية للبلوك–تشين في مواجهة تحديات المحتوى العربي الرقمي

بالنظر إلى خصائص تقنية البلوك–تشين، يمكن تبرير أهميتها كإطار نظري محتمل لحل الأزمة القائمة في توثيق المحتوى العربي. تتمثل مشكلة الدراسة الرئيسية في غياب نظام رقمي موثوق وموحّد لتوثيق النصوص العربية وإدارة المحتوى اللغوي، مما يؤثر سلبًا على موثوقية المخطوطات وحماية أعمال المؤلفين.

يمكن للبلوك–تشين أن تعالج التحديات الجوهرية المحددة في البيئة الرقمية العربية على النحو التالي:

1. معالجة ضعف التوثيق ومكافحة الانتحال: تُعد خاصية عدم قابلية التعديل الملازمة لتقنية السلاسل الكتلية حلاً فعالاً لمشكلة ضعف آليات التوثيق؛ حيث إن العديد من النصوص لا تحمل تاريخًا واضحًا أو سجلًا لتعديلاتها، مما يسهل التلاعب بها. إن تطبيق نظام توثيق مبني على البلوك–تشين يوفر سجلًا موثوقًا ودائمًا، وهذا يخدم بشكل مباشر الحاجة إلى إثبات أحقية المؤلف ويسهم في الحد من انتشار الانتحال الذي ينتج عن غياب سجل موثوق.

2. ضمان جودة البيانات لدعم الذكاء الاصطناعي: تبرز الحاجة الماسة للتوثيق الدقيق في سياق التطبيقات التقنية الحديثة، حيث تشير الدراسات إلى أن النماذج تتأثر بجودة البيانات، مما يجعل الحاجة للتوثيق أمرًا ضروريًا. إن غياب نظام التوثيق يؤدي إلى غياب بيانات موثوقة لتدريب نماذج الذكاء الاصطناعي. يوفر النظام المقترح البيانات الموثوقة اللازمة ويضمن جودة البيانات المستخدمة في التطبيقات التقنية الحديثة. وهذا يساهم في تحسين جودة النماذج اللغوية المبنية على البيانات.
3. دعم الحماية القانونية والمعايير اللغوية: تساهم البلوك–تشين في معالجة إشكالية فقدان الهوية الرقمية للمحتوى، وهو ما يعقّد حماية حقوقه ويؤدي إلى ضعف الحماية القانونية. يمكن أن يوفر الإطار المقترح آلية قوية لضمان حماية المؤلفين وحقوقهم. كما يمكن للامركزية والعقود الذكية أن تعالج التحدي المتمثل في غياب معيار اعتماد موحّد للمعاجم، حيث تختلف المداخل اللغوية وتغيب آليات موحدة لاعتماد التعديلات. وبالتالي، يدعم النظام الموثوق المقترح اعتماد الشهادات والمعاجم العربية.

6. الدراسات السابقة

تشير الدراسات الحديثة إلى أن البلوك–تشين يمكنه معالجة تحديات رئيسية في التوثيق والملكية الفكرية.

1.6 نماذج من الدراسات السابقة:

❖ **الدراسات التقنية**

أظهرت المراجعات المنهجية أن تقنية البلوك–تشين تمتلك إمكانات كبيرة في حفظ البيانات وضمان أصالتها (Casino et al., 2019).

❖ **الدراسات الثقافية والتراث، وتشمل:**

- مشاريع توثيق المخطوطات
- توثيق المصحف
- حماية الذاكرة الرقمية.

استخدمت تقنية البلوك–تشين لتوثيق التراث الثقافي وحفظ المخطوطات (Hassani et al., 2018).

❖ **دراسات حماية الملكية الفكرية، وتشمل:**

- توثيق الأعمال الأدبية والشعرية.
- مكافحة الانتحال الرقمي.

تناولت دراسات عديدة دور التقنية في منع السرقة الأدبية (Gürkaynak et al., 2018).

❖ **دراسات التعليم، وتشمل:**

- تخزين الشهادات عبر بلوك–تشين.
- اعتماد الساعات التعليمية.

أُقترحت أنظمة توثيق للشهادات عبر بلوك–تشين لتعزيز مصداقيتها(Sharples & Domingue, 2016) .

❖ **دراسات الذكاء الاصطناعي اللغوي، وتشمل:**

- توثيق بيانات تدريب النماذج.
- ضمان عدم التلاعب بالبيانات.

أكدت الدراسات أهمية تتبع مصدر البيانات المستخدمة في تدريب النماذج(Wang et al., 2022).

2.6 أوجه القصور في الدراسات السابقة

تشترك الدراسات السابقة في عدة قصور رغم تنوع مجالات تطبيق البلوك–تشين، أبرزها:

1. إهمال المحتوى العربي: غياب التركيز على توظيف التقنية لخدمة اللغة العربية وتحدياتها الخاصة.
2. نقص النماذج التطبيقية: افتقاد نماذج عملية تبين كيفية تنفيذ نظام توثيق عربي باستخدام البلوك–تشين.
3. ضعف الربط بالذكاء الاصطناعي: عدم معالجة العلاقة بين جودة البيانات العربية وتطوير النماذج الذكية.
4. غياب الرؤية المؤسسية: عدم تناول المتطلبات المؤسسية والتشريعية لتطبيق نظام عربي موحد.
5. إهمال البعد الزمني: عدم المعالجة المعمقة لدور التوثيق الزمني في حماية النصوص العربية من العبث.
6. ندرة الدراسات العربية: قلة الدراسات المتخصصة التي تُعنى بخصوصية اللغة العربية في هذا المجال.

3.6 تميّز الدراسة الحالية

تميزت هذه الدراسة بمعالجة الفجوات السابقة عبر:

1. ربط فريد: الجمع بين اللسانيات الحاسوبية وتقنية البلوك–تشين بشكل غير مسبوق.
2. نموذج متكامل: تقديم إطار شامل (المدخلات، العمليات، المخرجات) يقدم رؤية بنيوية متكاملة.
3. معالجة التحديات العربية: التركيز الصريح على تحديات اللغة العربية مثل تضارب النسخ وضعف المعاجم.
4. تعزيز الذكاء الاصطناعي: إبراز دور البيانات الموثقة في تطوير النماذج اللغوية العربية الذكية.
5. تطبيقات عملية: طرح استخدامات قابلة للتنفيذ مثل توثيق المخطوطات وحماية الحقوق الفكرية.
6. رؤية مستقبلية: تقديم منظور استراتيجي طويل المدى وتوصيات مؤسسية لتطبيق واسع النطاق.

7. المنهجية

تعتمد الدراسة على منهج تصميم الأنظمة (DSR) ويتضمن:

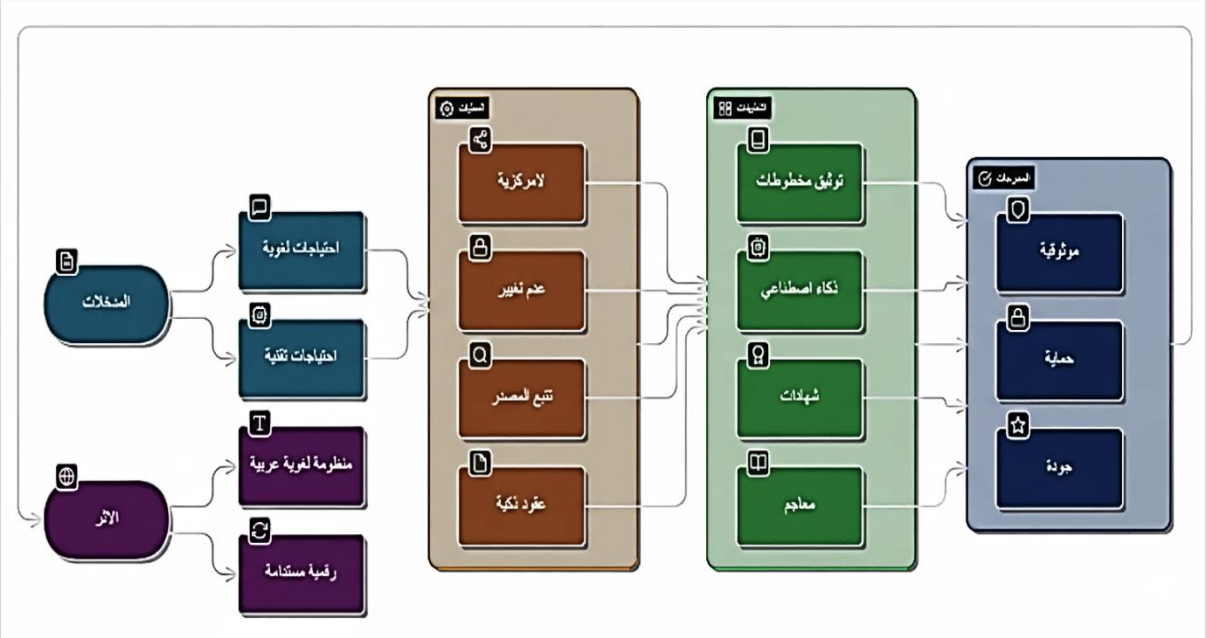
1. تحديد المشكلة	2. تحليل المتطلبات
3. تصميم النموذج	4. اختبار النموذج
5. تقييم الإطار	6. تطوير النموذج النهائي

8. النموذج المفاهيمي

يقترح البحث نموذجًا يتضمن خمس طبقات مترابطة:

1) المدخلات	• احتياجات لغوية • احتياجات تقنية	• بيئة رقمية
2) العمليات	• التشفير • اللامركزية	• العقود الذكية • تتبع المصدر
3) التطبيقات	• توثيق المحتوى • حماية الحقوق الفكرية • اعتماد المداخل اللغوية	• تعليم العربية • دعم الذكاء الاصطناعي
4) المخرجات	• بيانات دقيقة • محتوى موثوق	• شهادات معتمدة
5) الأثر	منظومة لغوية عربية مستدامة	

والشكل رقم 1 يمثل هذا النموذج:



الشكل 1: رسم للنموذج المقترح

9. التطبيقات العملية

1) توثيق المخطوطات العربية

يسمح البلوك–تشين بحفظ نسخ المخطوطات وتوليد بصمة رقمية (Hash) لكل نسخة، مما يمنع التلاعب.

2) حماية الملكية الفكرية

يُسجّل النص في سلسلة الكتل ويصبح ملكًا موثقًا بزمن الإنشاء.

3) اعتماد المعاجم

كل مدخل لغوي يمكن تتبّع تاريخه وتعديله ومراجعته.

4) تعليم العربية للناطقين بغيرها

توثيق الشهادات يحسّن القبول والمصداقية عالميًا.

5) دعم الذكاء الاصطناعي

تتبع مصدر البيانات يعالج مشكلة "البيانات الملوثة" ويزيد من جودة النماذج.

10. التحديات

*** تحديات تقنية**

- استهلاك الطاقة.
- كلفة التشغيل.
- قابلية التوسع.

*** تحديات لغوية**

- اختلاف المدارس.
- نقص الموارد.
- غياب المعايير.

*** تحديات مؤسسية**

- ضعف التنسيق.
- عدم جاهزية البنية التحتية.
- نقص التشريعات.

11. النتائج المتوقعة والتوصيات

1.11 النتائج المتوقعة

1. رفع جودة المحتوى العربي بشكل عام.
2. حماية النصوص
3. تعزيز جودة البيانات
4. تقليل الانتحال
5. دعم الذكاء الاصطناعي بتحسين أداء النماذج اللغوية الذكية
6. توحيد مرجعيات لغوية
7. زيادة الاعتماد على مصادر موثوقة.
8. حماية حقوق المؤلفين.
9. .تطوير منظومة عربية رقمية متكاملة.

2.11 التوصيات

- ✓ إنشاء هيئة عربية لتوثيق المحتوى عبر البلوك–تشين.
- ✓ تطوير مشروع عربي موحد لحفظ المعاجم والمخطوطات.
- ✓ تشجيع الجامعات على تبني التقنية.
- ✓ تعزيز التعاون بين المؤسسات.

12. حدود الدراسة وتوجهات البحث المستقبلية

1.12 حدود الدراسة

تجدر الإشارة إلى أنه على الرغم من اعتماد هذه الدراسة على منهجية البحث في علوم التصميم (DSR)، وهي منهجية فعّالة لبناء نماذج نظرية وتطبيقية، إلا أنها تخضع لبعض القيود البحثية:

أ: يُركز النموذج المقترح على التوثيق اللغوي والمعجمي دون إجراء اختبارات تجريبية واسعة النطاق داخل المؤسسات اللغوية الفعلية أو المنصات الرقمية القائمة. وهذا يُقيد القدرة على قياس الأداء الفعلي للنظام المقترح.

ب: لم تتناول الدراسة بشكل مُوسّع التباينات التشريعية والقانونية بين الدول العربية فيما يتعلق باستخدام تقنية البلوك تشين لحماية الملكية الفكرية. وقد تؤثر هذه الاختلافات بشكل مباشر على التطبيق العملي للنموذج.

ج: تقتصر الدراسة على اللغة العربية الفصحى الحديثة (MSA)، ولا تشمل اللهجات العربية أو المحتوى متعدد اللهجات. وهذا أمر بالغ الأهمية، نظراً لأن هذا المحتوى يُشكل نسبة كبيرة من الموارد العربية الرقمية، ويُمثل تحديات إضافية للتوثيق المُوحد والتحكم في الإصدارات.

د: لم تتضمن الدراسة منصات اختبار أو تجارب حاسوبية فعلية لنموذج سلسلة الكتل المقترح. وبالتالي، لا تزال جوانب مثل كفاءة الأداء، وتكاليف التشغيل، وقابلية التوسع، وتوافق التكامل مع تطبيقات الذكاء الاصطناعي اللغوي غير مُختبرة.

2.12 اتجاهات البحث المستقبلية المقترحة

تفتح نتائج هذه الدراسة آفاقًا لمسارات بحثية مستقبلية متعددة، أبرزها:

- ✓ التطبيق التجريبي: تطوير منصة سلسلة كتل تجريبية لتوثيق النصوص والمعاجم العربية، واختبارها في بيئة واقعية داخل أكاديميات اللغات أو الجامعات.
- ✓ الأطر القانونية: تحليل الجوانب التشريعية لحقوق الطبع والنشر في المنطقة العربية واقتراح إطار قانوني لدعم اعتماد تقنيات التوثيق اللامركزية.
- ✓ توسيع نطاق اللهجات: توسيع نطاق النموذج ليشمل اللهجات العربية، مع تطوير آليات موحدة لتوثيق الاختلافات اللهجة عبر سلسلة الكتل.
- ✓ تكامل الذكاء الاصطناعي: تطوير نماذج ذكاء اصطناعي عربية تعتمد على بيانات موثقة بتقنية سلسلة الكتل، ودراسة تأثير الختم الزمني ومصدر البيانات على دقة هذه النماذج. الويب 3 والحوكمة: دراسة تكامل الويب 3 والحوكمة اللغوية من خلال استكشاف آليات تطوير هوية لغوية عربية رقمية قائمة على اللامركزية والأصالة الرقمية.
- ✓ تحليل قابلية التوسع: تقييم تكاليف وقابلية تطبيق الأنظمة المقترحة على مستوى الوطن العربي، لا سيما فيما يتعلق بإدارة ملايين النصوص أو مدخلات المعاجم.

13. الختام

تقدم تقنية البلوك–تشين فرصة حقيقية لنهضة اللغة العربية رقميًا. فهي لا توثق المحتوى فحسب، بل تحميه، وتعطيه قيمة زمنية وسجلًا موثوقًا يمكن الرجوع إليه. ويُتوقع أن تتوسع استخدامات البلوك–تشين في المجال اللغوي مع تطور تطبيقات الجيل الثالث من الويب (Web3) والذكاء الاصطناعي. تشكّل هذه الدراسة نقطة انطلاق نحو بناء منظومة لغوية عربية تعتمد على الثقة والشفافية والاستدامة.

14. المراجع

المراجع العربية:

[1]

جابر، أشرف (2020). «البلوك تشين والإثبات الرقمي في مجال حق المؤلف». مجلة الدولية للفقه والقضاء والتشريع، العدد 1، الصفحات 32–58.

[2]

سيد، فايز رحاب (2020). «تقنية البلوك تشين وتوثيق الإنتاج الفكري العربي: دراسة تحليلية تقييمية لمحرك «إيداع»». مجلة المكتبات والمعلومات العربية، المجلد 40، العدد 2.

المراجع الأجنبية

[3]

Alzubaidi, M., & AlKhateeb, A. (2022). Blockchain applications in digital content protection: A review. *Journal of Information Security*, 11(3), 145–160.

[4]

Grech, A., & Camilleri, A. (2017). Blockchain in education. *European Commission Report*.

[5]

Gürkaynak, G., Yılmaz, İ., & Yeşilaltay, B. (2018). Intellectual property law and blockchain. *Computer Law & Security Review*, 34(4), 850–861.

[6]

Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in the heritage sector. *Technological Forecasting and Social Change*, 129, 77–87.

[7]

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

[8]

Tapscott, D., & Tapscott, A. (2016). *Blockchain revolution*. Penguin.

[9]

Wang, Y., Zhang, H., & Chen, X. (2022). Blockchain-based data provenance for machine learning. *IEEE Transactions on Computers*, 71(8), 1802–1815.

[10]

Xie, J., Tang, H., & Huang, T. (2019). A survey of blockchain technology applied to data management. *ACM Computing Surveys*, 51(6), 1–34.

[11]

Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2020). Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, 16(4), 352–375.

[12]

Casino, F., Dasaklis, T., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications. *Sensors*, 19(22), 4870.

[13]

Xu, X., Weber, I., & Staples, M. (2019). *Architecture for blockchain applications*. Springer.

[14]

Monrat, A., Schelén, O., & Andersson, K. (2019). A survey of blockchain applications in the energy sector. *IEEE Access*, 7, 21842–21865.

[15]

Savelyev, A. (2018). Copyright in the blockchain era. *Computer Law & Security Review*, 34(3), 550–561.

[16]

De Filippi, P., & Wright, A. (2018). *Blockchain and the Law: The Rule of Code*. Harvard University Press.

[17]

Chen, Y. (2018). Blockchain tokens and the potential democratization of creativity. *Management Science*, 64(7), 3096–3111.

[18]

Sharples, M., & Domingue, J. (2016). The blockchain for open science and education. *Journal of Learning Analytics*, 3(1), 47–62.

[19]

Turkanović, M., et al. (2018). EduCTX: A blockchain-based platform for higher education credits. *IEEE Access*, 6, 5112–5127.

[20]

Lionel, M. (2020). Blockchain-driven educational credentialing. *Computers & Education*, 153, 103891.

[21]

Kumar, D., & Simwanza, S. (2021). Blockchain for AI data governance. *AI Magazine*, 42(2), 27–40.

[22]

Rieffel, J., et al. (2019). Secure provenance for machine learning datasets. *arXiv preprint arXiv:1903.04919*.

[23]

Rahman, A., et al. (2022). Blockchain-empowered NLP frameworks. *Journal of Big Data*, 9(1), 1–22.

[24]

Ide, N., & Pustejovsky, J. (2017). *Handbook of Linguistic Annotation*. Springer.

[25]

Kilgarriff, A. (2016). Corpus linguistics and big data. *Language Resources and Evaluation*, 50(3), 1–20.

[26]

Maamouri, M., et al. (2014). Developing Arabic linguistic resources. *LREC Proceedings*.

[27]

Habash, N. (2010). *Introduction to Arabic Natural Language Processing*. Morgan & Claypool.

[28]

Zaghouani, W. (2017). Critical review of Arabic NLP resources and tools. *Journal of Artificial Intelligence Research*, 60, 1–38.

[29]

Alqahtani, M. (2020). Digital Arabic content creation challenges. *Arab Journal of Information Technology*, 17(4), 577–590.

[30]

Hassani, H., et al. (2020). Blockchain for heritage documentation. *Journal of Cultural Heritage*, 42, 180–189.

[31]

Tsiaras, S., et al. (2021). Preserving cultural heritage in the digital age using blockchain. *Digital Applications in Archaeology and Cultural Heritage*, 21, e00182.

[32]

Crosby, M. et al. (2016). Blockchain technology: Beyond bitcoin. *Applied Innovation Review*, 2, 6–19.

[33]

Mougayar, W. (2016). *The Business Blockchain*. Wiley.

[34]

Lansiti, M., & Lakhani, K. (2017). The truth about blockchain. *Harvard Business Review*, 95(1), 118–128.

[35]

ESCWA. (2013). Boosting digital Arabic content: New report examines challenges and opportunities. United Nations Economic and Social Commission for Western Asia. www.unescwa.org/news/boosting-digital-arabic-content-%E2%80%93-new-report-examines-challenges-and-opportunities

[36]

ESCWA. (2023). Digital Arabic content framework. United Nations Economic and Social Commission for Western Asia. https://www.unescwa.org/digital-arabic-content

[37]

UNESCO. (2023). UNESCO celebrates the heritage of the Arabic language and highlights its role in shaping new knowledge. UNESCO. www.unesco.org/en/articles/unesco-celebrates-heritage-arabic-language-and-highlights-its-role-shaping-new-knowledge

[38]

Zawacki-Richter, O., Kondakci, Y., Bedenlier, S., & Alturki, U. (2020). Open educational resources in the Arab region: A study of 22 countries. Smart Learning Environments, 7(24). https://slejournal.springeropen.com/articles/10.1186/s40561-020-00120-z

[39]

Abubaker, H. B., Salah, K., Al-Muhairi, H., & Bentiba, A. (2015, July). Digital Arabic content: Challenges and opportunities. *International Conference on IT Research and Innovation (ICTRC 2015).* https://doi.org/10.1109/ICTRC.2015.7156489.

[40]

Economic and Social Commission for Western Asia (ESCWA). (2012). Status of the digital Arabic content industry in the region (E/ESCWA/ICTD/2012/Technical Paper.4). United Nations ESCWA. https://www.unescwa.org/publications/status-digital-arabic-content-industry-region.

[41]

Economic and Social Commission for Western Asia (ESCWA). (2013). Business models for digital Arabic content (E/ESCWA/ICTD/2013/Technical Paper.1). United Nations ESCWA. https://www.unescwa.org/publications/business-models-digital-arabic-content.

[42]

Tlili, A., Jemni, M., Khribi, M. K., Huang, R., Chang, T.-W., & Liu, D. (2020). Current state of open educational resources in the Arab region: An investigation in 22 countries. *Smart Learning Environments, 7*(11). https://doi.org/10.1186/s40561-020-00120-z.

[43]

Hosman, L., Gómez-Zermeño, M. G., & Alemán de la Garza, L. (2024). An overview of Arabic language open educational resources (OER) for primary and secondary education and their use in offline environments. *Electronic Journal of e-Learning, 22*(9). https://files.eric.ed.gov/fulltext/EJ1439900.pdf

[44]

Farahat, F., & Hamouda, A. (2015). Detecting plagiarism in Arabic e-learning system using text mining techniques. *Journal of Advances in Mathematics and Computer Science.*

[45]

Al-Thwaib, N., Alnajjar, K., Al-Qatamin, A., & Kanaan, G. (2020). An academic Arabic corpus for plagiarism detection in compilations of electronically available Bachelor's and Master's theses and dissertations. *International Journal of Electrical and Computer Engineering (IJECE).*

[46]

Global Investigative Journalism Network. (2020). Plagiarism is widespread on the Arabic net, and it's hurting investigative journalism. Global Investigative Journalism Network.

[47]

Baccar Ben Amar, M., Hadrich Belguith, L., & Ben Hamadou, A. (2008). LMF standardized model for the editorial electronic dictionaries of Arabic. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008).*

[48]

Habash, N., Bouamor, H., & Rambow, O. (2018). Unified guidelines and resources for Arabic dialect orthography. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

15. جدول الالفاظ

English	عربي
Applied Framework	إطار مفاهيمي وتطبيقي
Blockchain technology	تقنية البلوك–تشين
Arabic digital content	المحتوى العربي الرقمي
Digital documentation	التوثيق الرقمي
Linguistic data integrity	سلامة البيانات اللغوية
Intellectual property	الملكية الفكرية

16. السيرة الذاتية للباحثين

17.

	جميل اطميزي يحمل رتبة أستاذ مشارك، حصل على درجة الدكتوراه في علوم الحاسوب (تطوير برامج التعليم الالكتروني) من جامعة غرناطة بإسبانيا سنة 2005، ويعمل محاضرا في كلية الهندسة وتكنولوجيا المعلومات، وقد عمل عميدا للبحث العلمي ومديرا لمركز التعليم الإلكتروني. د. اطميزي باحث نشيط، قام بتأليف أكثر من 17 كتاب، ونشر أكثر من 65 بحثا حول مجالات مثل التعليم الإلكتروني، والذكاء الاصطناعي، والموارد التعليمية المفتوحة، وهو مُقيّم برامج أكاديمية بوزارة التعليم العالي الفلسطينية، ويشارك في عضوية العديد من المؤتمرات والدوريات وقد قيم عشرات الأبحاث في عديد المؤتمرات الدولية، وبتحكيم العشرات من رسائل الماجستير والدكتوراه. وشارك د. اطميزي في تدريس عديد الدورات التدريبية، وفي وضع خطط للبرامج الأكاديمية، وكتب عشرات المقالات التكنولوجية، وقد فاز بعدة منح وجوائز، وحضر عشرات الورش والمؤتمرات.
	محمد طميزه باحث وأكاديمي فلسطيني متخصص في تقنيات المعلومات والاتصالات، يعمل محاضراً في كلية الهندسة وتكنولوجيا المعلومات في جامعة فلسطين الأهلية، ويشغل منصب عميد القبول والتسجيل. يحمل درجة الدكتوراه من جامعة غرناطة في مجال تقنيات المعلومات والاتصالات، ودرجة الماجستير في علوم الحاسوب، ويركز في أبحاثه على تطبيقات البلوك–تشين وأمن المعلومات وتخزين البيانات اللامركزي. نشر عدة أبحاث في مجلات ومؤتمرات دولية، وله خبرة واسعة في تطوير البرمجيات، العقود الذكية، تكامل نظم التعليم الإلكتروني، والإشراف على مشاريع تقنية. يدرس منذ 2011 في مجالات البرمجة وقواعد البيانات وأمن المعلومات، وشارك بلجان للمؤتمرات ولخدمة المجتمع.

الخلاصة باللغة الانجليزية

Employing Blockchain Technology in Supporting the Arabic Language: A Conceptual and Applied Framework for Enhancing Digital Documentation

Jamil Itmazi, and Mohammed Tmaizeh,
Faculty of Engineering and IT, Palestine Ahliya University, Bethlehem, Palestine
j.itmazi@gmail.com and mohammedt@paluniv.edu.ps

Abstract:
Over the past two decades, the Arabic language has witnessed significant digital expansion reflected in the growth of educational, literary, research, and media content; however, this rapid development has also generated major challenges related to content reliability, intellectual property protection, standardized lexicons, text documentation, and the quality of linguistic data used in artificial intelligence applications. In this context, blockchain technology emerges as a transformative model based on decentralization and immutability, offering enhanced transparency and trust within digital ecosystems. This study aims to explore the potential of employing blockchain to support the Arabic language by developing a comprehensive theoretical and applied framework that presents practical implementation models and examines the institutional and technical factors influencing adoption. The findings indicate that integrating blockchain into the management of Arabic linguistic resources can lead to a qualitative shift by improving digital content quality, safeguarding authors' rights, and supporting the development of more accurate and reliable AI solutions for Arabic. The study adopts the Design Science Research (DSR) methodology and proposes a conceptual framework that links inputs, processes, applications, and outputs, supported by an extensive literature review and an in-depth analysis of existing challenges, concluding with actionable recommendations to promote the digital sustainability of the Arabic language.

Keywords: Blockchain technology, Arabic digital content, digital documentation, linguistic data integrity, intellectual property protection.