

المخطوط العربي في زمن التسارع الرقمي
من أنطولوجيا النص التراثي إلى إبستمولوجيا التحرير الخوارزمي
دراسة تطبيقية في إشكاليات الرقمنة والتحقيق بالذكاء الاصطناعي

The Arabic Manuscript in the Age of Digital Acceleration: From the Ontology of the Heritage Text to the Epistemology of Algorithmic Editing A Applied Study on the Problematics of Digitization and AI-Assisted Critical Edition

د. صالح المهدي بن حمودة*

جامعة الزيتونة، تونس، البريد الإلكتروني: mahdihammouda@yahoo.fr

تاريخ النشر: 2026/10/15

تاريخ المراجعة: 2026/09/07

تاريخ الإيداع: 2026/08/09

ملخص:

تنهض هذه الدراسة على إشكالية تتقاطع فيها ثلاثة حقول ظلت متباعدة في الدرس العربي المعاصر: علم المخطوطات بمنهجه الكوديولوجي، والإنسانيات الرقمية بمشاريعها الأرشيفية، وأدوات الذكاء الاصطناعي ولا سيما نماذج التعرف الآلي على الخطوط (HTR). وتفترض أن التسارع الرقمي لا يغيّر وسائط حفظ المخطوط فحسب، بل يعيد تشكيل أنطولوجيا النص التراثي، ويزحزح مفاهيم «الأصل» و«النسخة الأم» و«السماع» و«الإجازة». وتبني الدراسة، بمقاربة نظرية تطبيقية، إطاراً إبستمولوجياً عربياً مؤصلاً، وتختتم بمقترح عملي: بروتوكول لبناء قاعدة بيانات عربية للمخطوطات تراعي الخصوصية الثقافية. وتخلص إلى أن الذكاء الاصطناعي لا يلغي التحقيق البشري بل يعيد توزيع مهامه، وأن الخصوصية العربية قابلة للنمذجة الرقمية إذا ما بُني نموذج عربي مؤصّل.

الكلمات المفتاحية: المخطوط العربي؛ الرقمنة؛ التحقيق النقدي؛ الذكاء الاصطناعي؛ قواعد البيانات التراثية؛ التسارع الرقمي.

* المؤلف المراسل: د. صالح المهدي بن حمودة

Abstract:

This study addresses a compound problematic at the intersection of three fields that have remained largely disconnected in contemporary Arab scholarship: classical codicology, digital humanities archival projects, and artificial intelligence tools, particularly Handwritten Text Recognition (HTR) engines. It advances the hypothesis that digital acceleration does not merely alter the media of manuscript preservation but reconfigures the very ontology of the heritage text, destabilizing foundational concepts of Arab editing: *aṣl* (original), *nuskha umm* (mother copy), *samāʿ* (audition), and *ijāza* (transmission license). Through a theoretical-applied approach, the study constructs a grounded Arab epistemological framework and concludes with a practical proposal: a protocol for building an Arabic manuscript database that respects cultural specificity (*isnād, samāʿ, ijāza, waqf*). It concludes that artificial intelligence does not abolish human editing but redistributes its tasks, and that Arab cultural specificity is digitally modelable provided that an authentically Arab data model is constructed.

Key words: Arabic Manuscript; Digitization; Critical Edition; Artificial Intelligence; Heritage Databases; Digital Acceleration.

المقدمة:

المخطوط العربي في تاريخنا المعرفي ليس حاملةً ماديةً للنص فحسب، بل مؤسسة معرفية متكاملة؛ يجتمع فيه صنعة الوراق، ودراية النسخ، وضبط المقابل، وتوثيق المحدث، وتوقيعات المجيزين، وشهادات السامعين، وصيغ التملك، وحجج الوقف. وقد لخص ابن خلدون هذه الحقيقة حين وصف الكتاب بأنه «وعاء العلم»¹؛ ومن ثم كان تحقيق النص التراثي في المنظومة العربية عملاً مركباً تتقاطع فيه الملكات اللغوية والفقهية والكوديكولوجية، وتضبطه منظومة نقدية صارمة بلورها جيل المحققين الكبار أحمد شاكر، وعبد السلام هارون، وعبد الرحمن المعلي اليماني، ومحمود محمد شاكر تقوم على تراتبية النسخ، ومركزية «النسخة الأم»، وتقديم السماع على الوجدادة².

غير أن هذه المنظومة وجدت نفسها في العقدين الأخيرين أمام تحول زمني جذري؛ إذ دخلت المخطوطات العربية قسراً أو طوعاً فضاء التسارع الرقمي الذي وصفه هارتموت روزا بأنه «إعادة هيكلة زمانية للحياة الاجتماعية»³، فلم يعد الزمن زمن المحققين القدماء، بل زمن الثانية والميلي ثانية، وزمن الخوارزمية التي تُنجز في ساعاتٍ ما كان يستغرق عقوداً. وفي هذا الفضاء برزت أدوات التعرف الآلي على الخطوط اليدوية (Handwritten HTR – t Recognition Tex) من قبيل Transkribus و Kraken و eScriptorium، ثم النماذج اللغوية الكبرى (LLMs) التي أضحت قادرة نظرياً على قراءة الخطوط العربية القديمة وتحويلها نصوصاً رقمية قابلة للبحث والتحرير⁴.

بيد أن هذا الاختراق التقني لم يقابله في المحيط الأكاديمي العربي تنظيمٌ إبستمولوجي يضبط حدوده، ولا تأصيلٌ منهجي يصله بتراث التحقيق العربي⁵؛ فغالِبُ الدراسات المتاحة وصفٌ تقنيٌ لمشاريع الرقمنة الكبرى، لا مساءلة فلسفية لمفهوم «الأصل» و«النسخة» في البيئة الرقمية، ولا نقدٌ لحجية النص المنتج آلياً. وهنا تتحدد إشكالية الدراسة وإضافتها العلمية؛ إذ تتجاوز الجرد الوصفي لمشاريع الرقمنة إلى أمرين: تأصيل إبستمولوجي عربي لمفهوم التحرير الخوارزمي، ونموذج تطبيقي هو مشروع «سند»، ويُحسَم موقعه هنا حسماً: فهو ليس نموذج بيانات منفرداً، ولا معيار ترميز مستقلاً، ولا بروتوكول تحقيق مجرداً، بل تصور متكامل يجمع المستويات

الثلاثة — نموذج البيانات (الكيانات والعلاقات)، ومعياري الترميز (Arabic-TEI)، وبروتوكول التحقيق الرقمي (دورة الجودة) — في بنية واحدة قابلة للاختبار على مادة مخطوطية حقيقية.

وتتعلق هذه الإشكالية بمصير المفاهيم التأسيسية للتحقيق العربي حين تنتقل النسخة من عالم المادة إلى عالم الرمز: هل يظل مفهوم «الأصل» قائماً في بيئة لا أصل فيها بالمعنى المادي؟ وكيف تنتظم العلاقة بين المحقق والآلة في زمن صار فيه التحقيق ممكناً بضغطة زر؟ وترتبط بهذه الإشكالية ثلاث فرضيات: أولاًهما أن التسارع الرقمي لا يغيّر وسائط الحفظ فحسب، بل يعيد تشكيل أنطولوجيا النص التراثي ويزحزح المفاهيم التأسيسية؛ وثانيتهما أن الخصوصية الثقافية العربية السند والسماح والإجازة والوقف قابلة للنمذجة الرقمية إذا بُني نموذج بيانات عربيّ مؤصّل لا يستنسخ النماذج الغربية؛ وثالثتهما أن بناء قاعدة بيانات عربية تراعي هذه الخصوصية ممكنٌ عملياً عبر بروتوكول محدّد المراحل يصون صرامة التحقيق ويستثمر التسارع. وقد لاحظ Cornelis van Lit أن الباحثين في التراث الإسلامي يعيشون «انفصاماً منهجياً» بين حقول المخطوطات والإنسانيات الرقمية والذكاء الاصطناعي⁷؛ وهذه الدراسة محاولة لردم هذا الانفصام.

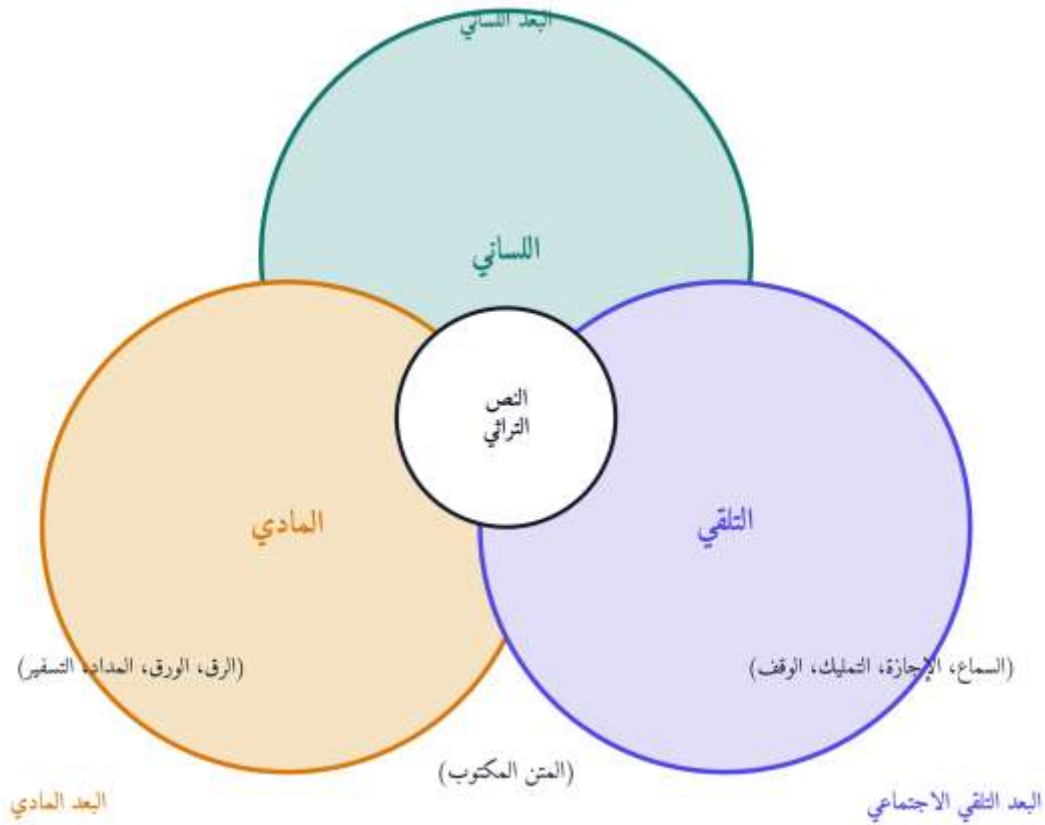
واعتمدت الدراسة المنهج الوصفي التحليلي في جرد المشاريع وتقويمها، والمنهج المقارن في دراسة معايير الوصف والترميز، والمنهج التطبيقي في بناء البروتوكول المقترح؛ وتتنظم في خمسة محاور يعالج أولها الإطار المفاهيمي، وثانيها التسارع الرقمي، وثالثها واقع الرقمنة العربية، ورابعها فرص الذكاء الاصطناعي ومزالقه، وخامسها الدراسة التطبيقية في مشروع «سند»؛ ويعقب ذلك خاتمة بالنتائج والتوصيات.

أولاً: الإطار المفاهيمي: النص التراثي بين الأنطولوجيا والرقمنة:

لا سبيل إلى بناء تصور سليم لمسألة الرقمنة والتحرير الخوارزمي إلا بتحرير المفاهيم أولاً؛ فما النصُّ التراثيُّ؟ وما الأصل والنسخة؟ وما الكوديولوجيا في صورتها الكلاسيكية والرقمية؟ يضبط هذا المحور هذه المفاهيم ضبطاً إجرائياً في أربعة مطالب، تمهيداً للمعالجات اللاحقة.

1. أنطولوجيا المخطوط العربي: الأبعاد الثلاثة المتلازمة:

المخطوط قبل كل شيء ليس «حاوية» للنص يمكن استبدالها بأي حاوية أخرى، بل بنية أنطولوجية ثلاثية الأبعاد: بُعد لسانی (المتن المكتوب)، وبُعد مادي (الرق أو الورق، والمداد، والتفسير، والحبكة)، وبُعد تلقّي اجتماعي (السماعات والإجازات والتمليكات والوقوف)⁸. وقد نبّه Gacek إلى أن عناصر المخطوط من نوع المداد ومقاس الكراسة (bifolium) إلى العلامة المائية (watermark) وعلامات المقابلة تشكل «معجماً موازياً» لا يقل أهمية عن المتن نفسه. والمحدثون المسلمون، منذ القرن الثاني الهجري، وعوا هذه الأبعاد وعياً إبستمولوجياً مبكراً: بنوا نظرية الحديث على تلازم السند بالمتن، وقدموا السماع على الوجداء، وميّزوا الرواية باللفظ من الرواية بالمعنى⁹؛ ومن هنا كان إصرار المعلمي اليماني على تتبع مسار النسخة الأم من ناسخها إلى خزانها، وهو ما يسميه «Witkam أركيولوجيا المخطوط» (ygoloeahcra tpircsunam)¹⁰. ويجسّد شكل (1) هذه الأبعاد الثلاثة المتلازمة.



شكل (1): الأبعاد الثلاثة المتلازمة للنص التراثي المخطوط

المصدر: من إعداد الباحث، انطلاقاً من (Déroche (2005 و (Gacek (2009).

يكشف الرسم أن المتن اللساني وحده لا يقوم به النص التراثي؛ فإهمال البعد المادي أو التلقي الاجتماعي اختزالاً أنطولوجياً يفقد المخطوط دلالاته التاريخية، وهو معيارٌ يجب استحضاره عند تحويل المخطوط إلى صورة رقمية وعند تصميم قواعد البيانات الخاصة به. ويلزم التدقيق هنا في مفهوم «الأنطولوجيا» المعتمد في هذه الدراسة؛ فالرقمنة لا تغير طبيعة النص التراثي في ذاته — إذ يظل المخطوط المادي وثيقة تاريخية لا تُنسخ كينونتها — بل تغير أنماط تمثيله ووصفه وإتاحته وإعادة استخدامه داخل البيئة الرقمية؛ ومن ثمّ فالمقصود بـ«إعادة تشكيل الأنطولوجيا» تحوُّل بنية التمثيل المعرفي للنص وأدوات مقارنته، لا زوال كينونته المادية ولا تعويضها؛ وهو تمييزٌ يقي من الانزلاق من وصف تحوُّل تقنيٍّ إلى حكم أنطولوجيٍّ غير مؤسَّس.

2. الأصل والنسخة في المنظومة النقدية العربية:

لم يكن مفهوم «الأصل» في التراث النقدي العربي مفهوماً مجرداً، بل مفهوماً إجرائياً محكوماً بمعايير صارمة: الأصل عند ابن الصلاح «ما قرئ على الشيخ وسمع منه»، وهو مقدّم على الوجدادة والمناولة والإجازة¹¹. والنسخة الأم عند المحققين المعاصرين هي ما كتبه المؤلف بخطه، أو أملاه فأقرّ، أو قرئ عليه فسُمع¹². وقد اشترط عبد السلام هارون سبعة شروطٍ لاعتماد النسخة أصلاً: التقدم الزمني، وتوثيق الناسخ، والمقابلة، والسلامة من التصحيف، ووجود السماعات، وحسن الخط، وسلامة المادة¹³. وهذه الشروط كما سيظهر كلّها معطياتٌ قابلةٌ للترميز الرقمي، وهو ما يجعل المنظومة العربية قابلةً للنمذجة من حيث المبدأ.

3. من الكوديكولوجيا الكلاسيكية إلى الكوديكولوجيا الرقمية:

ميّز François Déroche في كتابه Islamic Codicology بين خمس طبقاتٍ من العناصر ينبغي للمحقق مقاربتها: المادة الحاملة، والخط، والتخطيط الصفحي (mise en page)، والزخرفة، والتسفير. غير أن هذا الإطار الكلاسيكي المصنّم لعصر الفهرسة اليدوية اصطدم بظهور «الكوديكولوجيا الرقمية» التي تضيف طبقاتٍ جديدة: البيانات الوصفية (metadata)، ومعايير الترميز (TEI و IIIIF)، وطبقات التحليل التخطيطي (Layout Analysis)¹⁴؛ فصار وصف المخطوط وصفاً آلياً قابلاً للاسترجاع والمقارنة، وانتقل علم المخطوطات من حرفة الفهارس إلى علم البيانات.

4. مراتب التمثيل الرقمي وحجية النسخة الرقمية:

هل النسخة الرقمية حجة في التحقيق؟ يميز الباحثون في التحرير الرقمي بين ثلاث مراتب للتمثيل: الفاكسيميلي التصويري (نقل صورة المخطوط كما هي) بحجية بصرية، والنقل الحرفي المطابق للنسخة بحجية وصفية، والتحرير النقدي الرقمي بحجية علمية كاملة، بشرط استيفاء معايير الترميز ومصادقة لجنة علمية¹⁵¹⁶؛ وهو اصطلاح متسق مع مصطلحات الدرس المخطوطي العربي من قبيل «النقل الحرفي» و«التحرير النقدي» كما جرى به العمل في فهارس المخطوطات العربية الحديثة. ويضيف John Bryant تصور «النص السائل» (diulf Text): النص تدفق من التنوعات (variants) تتزامن ولا تتعاقب¹⁷؛ وإسقاط هذا التصور على المخطوط العربي الذي عرف نسخاً متعددة متفاوتة الزمان والمكان والناسخ يفتح إمكان العرض التزامني (ic DisplaySynopt) لكل النسخ، بدل إخضاعها لنسخة أمّ مهيمنة¹⁸؛ غير أن العرض التزامني للنسخ لا يلغي التراتبية بين الشواهد؛ فإتاحة النسخ المتعددة متزامنة لا تعني مساواتها في القيمة التاريخية أو الحجية النقدية؛ ولذلك يُشترط في النموذج الرقمي الرصين ترميز مكانة كل نسخة — أمّا كانت أو مقرّبة من عصر المؤلف أو شاهداً متأخراً — ودرجة حجيتها ضمن بياناتها الوصفية، فيجمع بين إتاحة الشواهد جميعاً وضبط علاقاتها التاريخية والإسنادية، بدل استبدال تراتبية الشواهد بمنطق المساواة الرقمية بينها. وهذا التمييز لازمٌ للمحقق العربي قبل توظيف الصور الرقمية في تحقيقاته.

ثانياً: التسارع الرقمي: من زمن المحقق إلى زمن الخوارزم:

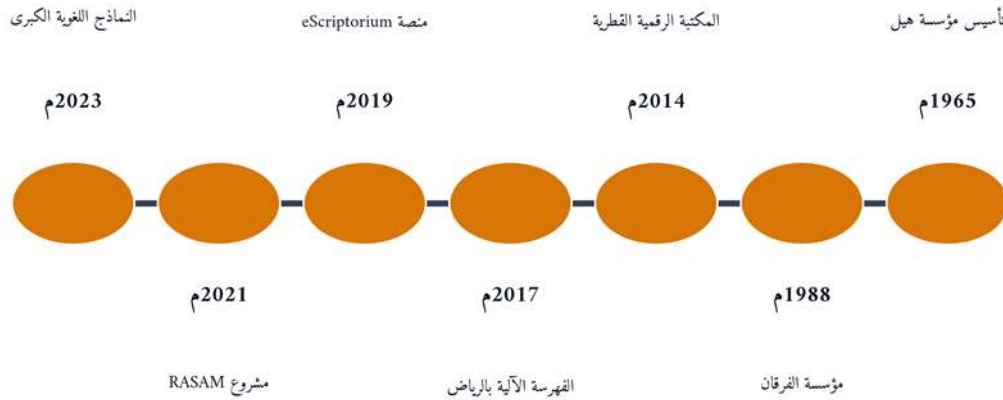
بعد ضبط المفاهيم، يفحص هذا المحور المناخ الزمني الجديد الذي صارت فيه الخوارزمية شريكاً في التحقيق؛ وذلك في أربعة مطالب.

1. نظرية التسارع عند هارتموت روزا:

يميز روزا بين ثلاثة أنحاءٍ للتسارع: تسارعٌ تقني، وتسارعٌ التغير الاجتماعي، وتسارعٌ إيقاع الحياة¹⁹. وإسقاطاً على حقل المخطوط الذي كان تاريخياً «فضاء البطء» بامتياز، إذ استغرق تحقيق «كتاب سيبويه» عند عبد السلام هارون عمراً مديداً نجد الحقل وقد انتقل إلى التسارع الأقصى: الخوارزمية تُنجز في ساعاتٍ ما كان يستغرق عقوداً، فيتحول الأمر من مسألة تقنية إلى إعادة تعريف معايير «الجودة العلمية» ذاتها²⁰.

2. محطات التسارع في خدمة المخطوط العربي:

يمكن تتبع هذا الانتقال في محطات موثقة: بدأت المؤسسات الغربية الحفظ الميكروفيلي للتراث الشرقي منذ ستينيات القرن الماضي (مؤسسة هيل للمخطوطات)²¹، ثم جاءت المؤسسات الإسلامية الكبرى كمؤسسة الفرقان للتراث الإسلامي²²، فالمكتبة الرقمية القطرية بملايين الصفحات المرقمنة²³، ثم مشاريع الفهرسة الآلية في المكتبات العربية، فأدوات التعرف الآلي المفتوحة، ثم مشاريع النماذج المخصصة (RASAM)، وأخيراً النماذج اللغوية الكبرى. ويرسم شكل (2) هذه المحطات على خط زمني يكشف عن التسارع المتصاعد.



شكل (2): محطات التسارع الرقمي في خدمة المخطوط العربي (1965-2023م)

يوضح الرسم انتقال الحقل من «زمن البطء المعرفي» إلى تسارع متصاعد، خاصة منذ 2014؛ فالخط الزمني ترجمة بصرية للفرضية الأولى: التسارع لا يغيّر الوسائط فحسب، بل يغيّر نسق العمل المعرفي نفسه.

3. مفارقة «الحقيقة المرجعية» في التعرف الآلي:

يقوم التعرف الآلي على الخطوط المخطوطة (HTR) على ثلاثة مكونات: تحليل التخطيط (Layout Analysis) الذي يميز الأسطر والأعمدة والحواشي، ونموذج التعرف (Recognition Model) المدرب على مجموعة بيانات مرجعية (Ground Truth)، والنموذج اللغوي (Language Model) الذي يصحح الإخراج بناءً على الاحتمالات السياقية²⁴²⁵. والمفهوم المفصلي هنا هو «الحقيقة المرجعية»: فهو ينطوي على مفارقة إستمولوجية حادة: كيف تكون نسخة يدوية من القرن العاشر «حقيقة مرجعية» للخوارزمية، في حين أنها عند المحققين شاهد من شهود متعددين لا الأصل المطلق؟ ولضبط هذا المفهوم إجرائياً في سياق المخطوط العربي، تميز هذه الدراسة بين ثلاثة مستويات للحقيقة المرجعية: (1) الحقيقة الصورية: الصفحة المصورة للشاهد بعينه؛ (2) الحقيقة النصية: القراءة النصية المستخرجة من ذلك الشاهد نفسه، لا النص المحقق المجرد عن شواهد؛ (3) الحقيقة النقدية: النص المحقق الذي تُسند فيه القراءات إلى شواهدا. ويبنى مشروع «سند» مجموعات التدريب على المستوى الثاني — قراءة كل شاهد على حدة مربوطة بصورته — حتى لا تُدرّب

الخوارزمية على نصّ محققٍ هو نفسه حصيلةُ حكمٍ بشريٍّ سابق، وحتى يظل قياسُ الدقة قياساً لقدرة الآلة على قراءة الشاهد، لا لقدرتها على إعادة إنتاج اختيارات المحقق.

4. مسار المعالجة ومكان الخطر:

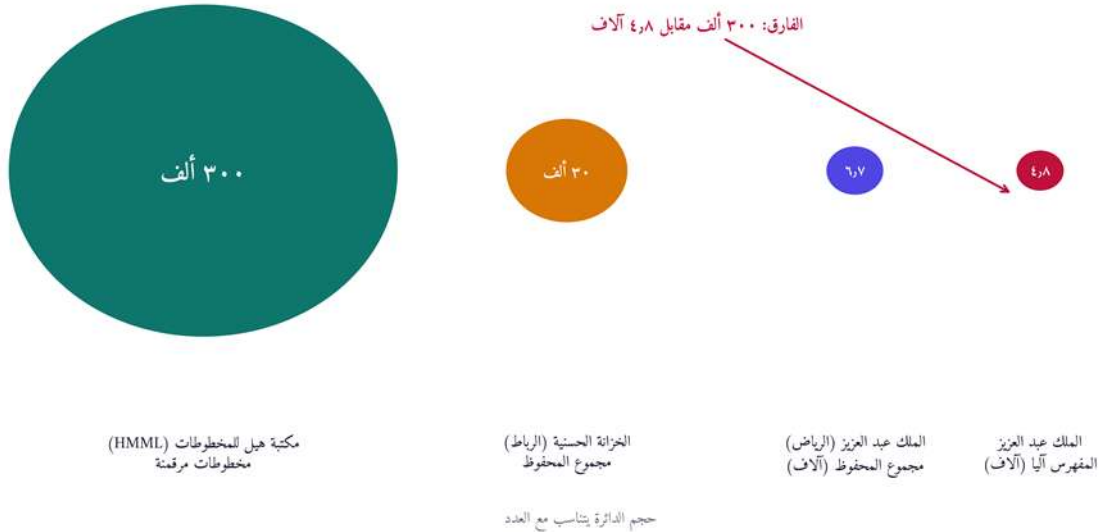
تمر المعالجة بست محطاتٍ متتابعة: التصوير الرقمي، ثم تحليل التخطيط، ثم التعرف الآلي، ثم التصحيح اللغوي الاحتمالي، ثم المراجعة البشرية المتدرجة، ثم النشر التفاعلي؛ ولكل محطة مخاطرها: التصوير يفقد بيانات، والتحليل التخطيطي قد يخطئ في تمييز الحواشي، ونموذج التعرف ينتج أخطاءً بصرية، والنموذج اللغوي قد «يهلس» في تصحيحاته الاحتمالية²⁶، ولا تلتقط هذه المخاطر إلا المراجعة البشرية المتدرجة؛ ومن ثمّ فهي المحطة التي لا يجوز تجاوزها في أي بروتوكول جاد. ويمكن تصنيف أخطاء المعالجة الآلية في أربعة أصنافٍ متميزة: خطأً بصرياً (كالخلط بين الحروف المتشابهة)، وخطأً نصيً (كإسقاط كلماتٍ أو زيادتها)، وخطأً لغويً احتماليً (كتصحيح سياقٍ يغيّر صيغةً أصليةً نادرةً إلى صيغةٍ متداولة)، وهلوسةً توليديةً (كاختلاق نصٍّ لا وجود له في الشاهد)؛ وهذا التمييز شرطٌ لتوجيه المراجعة البشرية توجيهاً علمياً، إذ لكل صنفٍ آلية كشفٍ ومعالجةٍ تخالف غيره.

ثالثاً: خرائط الرقمنة: الجرد النقدي:

لا يُبنى البروتوكول المقترح في فراغ؛ فلا بد من جرد ما أنجز في الميدان العربي والدولي، وتقويمه نقدياً، واستخلاص الدروس. يتناول هذا المحور ذلك في ثلاثة مطالب.

1. المشهد العربي: ثلاث فئات مؤسسية وأرقامها:

يمكن تصنيف مشاريع الرقمنة العربية في ثلاث فئات: مؤسساتٌ ملكيةٌ ووطنية (مكتبة الملك عبد العزيز العامة، والخزانة الحسنية، ومكتبة الأزهر، ودار الكتب المصرية)، ومؤسساتٌ متخصصة (مؤسسة الفرقان للتراث الإسلامي، ومركز جمعة الماجد، ومعهد المخطوطات العربية بالقاهرة)²⁷، ومشاريعٌ أكاديميةٌ جامعية (رقمنة الخزانة الزيتونية بتونس، ومخطوطات المكتبة الوطنية الجزائرية)²⁸. ويكشف شكل (3) التفاوت في نطاق هذه الرقمنة بمعطياتٍ موثقة²⁹: مكتبة هيل تحفظ نحو ٣٠٠ ألف مخطوط مرقم³⁰، بينما لا تتجاوز المخطوطات المفهرسةً ألياً في مكتبة الملك عبد العزيز ٤,٨ آلاف من أصل ٦,٦٦٣ محفوظاً³¹، والخزانة الحسنية على ما قرره مديرها أحمد شوقي بنين ما زالت في طور رقمنة نحو ٣٠ ألف مخطوط محفوظ³²، بينما تجاوزت الصفحات المرقمنة في مكتبة قطر الرقمية مليوني صفحة³³.



شكل (3): نطاق الرقمنة في كبرى المؤسسات: معطيات موثقة

يرسم الشكل المقارنة بدوائر يتناسب حجمها مع العدد: الفارق بين نحو 300 ألف مخطوط مرقمة في مكتبة هيل، ونحو 30 ألف محفوظ في الخزانة الحسنية، وأقل من 5 آلاف مفهرس آيا في مكتبة الملك عبد العزيز؛ وهو ما يدعم الملاحظة النقدية بأن الرقمنة العربية تعمل في «جزر منعزلة» وبايقاع لا يواكب حجم التراث.

2. المشهد الدولي: خمسة مشاريع كبرى:

على المستوى الدولي تبرز خمسة مشاريع كبرى ذات صلة بالتراث العربي: المكتبة الرقمية القطرية (Qatar Digital Library) بأكثر من مليوني صفحة مرقمة من الأرشيف البريطاني؛ ومؤسسة هيل للمخطوطات (HMML) التي تحفظ نسخاً رقمية لنحو 300 ألف مخطوط شرقي؛ ومكتبة فرنسا الوطنية (Gallica) بمجموعتها الشرقية؛ ومكتبة ليدن؛ والمكتبة الفاتيكانيّة بمشروعها DigiVatLib³⁴. وهذه المشاريع تقدم دروساً في معايير التصوير والبيانات الوصفية، لكنها لا تعالج الخصوصية العربية علاجاً كافياً، كما سيبين المحور الخامس.

3. التقييم النقدي: ثلاث مشكلات بنيوية:

على الرغم من هذه الجهود، يكشف التحليل النقدي عن ثلاث مشكلات بنيوية: (أ) تركيز المشاريع على المسح الضوئي دون النشر المعرفي؛ فالمخطوط يُصوّر ويُرفع إلى المنصة دون بيانات وصفية متقدمة، ودون تحويل نصي يمكن من البحث داخل المتن³⁵؛ (ب) عمل كل مشروع في «جزيرة منعزلة» دون تبني معايير موحدة كـ IIF الذي يتيح التشارك العابر للمكتبات³⁶؛ (ج) ضعف مجموعات التدريب المرجعية لنماذج HTR العربية قياساً باللغات اللاتينية³⁷. وهذه المشكلات الثلاث هي التي يستهدف البروتوكول المقترح معالجتها. وإذا أُعيد تنظيم الجرد السابق وفق معايير موحدة — طبيعة المادة، وحجم المجموعة، ومستوى الفهرسة، ووجود النص القابل للبحث، واستخدام HTR، والربط بين الصورة والنص، وطبيعة البيانات الوصفية، ومستوى الإتاحة وإمكان إعادة الاستخدام — كشفت المقارنة عن تفاوت نوعي لا كمي فحسب: فالمكتبة الرقمية القطرية تقدم صوراً عالية الجودة وبيانات وصفية غنية وإتاحة حرة، لكنها دون نص مخطوطي قابل للبحث؛ ومؤسسة

هيل تحفظ أكبر مجموعة مع إتاحة إعادة الاستخدام عبر OPenn، لكن فهرستها متفاوتة العمق؛ فيما لا يتجاوز كثير من المشاريع العربية طبقة الصورة المسحوقة دون بيانات وصفية متقدمة. ويلزم التحفظ المنهجي هنا: فالأرقام الواردة لا تقيس وحدة واحدة؛ إذ يعد بعضها المخطوطات، وبعضها الصفحات، وبعضها المواد المفهرسة أو المرقمنة، فلا تصح المقارنة الكمية المباشرة بينها إلا بهذا الضبط.

رابعاً: الذكاء الاصطناعي: الإمكان والحدود:

بعد جرد الميدان، ينتقل هذا المحور إلى فحص الأدوات الخوارزمية نفسها: ما قدراتها الحقيقية؟ وما حدود مشروعيتها العلمية؟ وذلك في ثلاثة مطالب.

1. أدوات HTR للعربية: قراءة مقارنة:

تتصدر ثلاث أدوات مفتوحة المصدر مشهد التعرف الآلي على الخطوط العربية: Transkribus (طوره اتحاد READ الأوروبي، ويعتمد شبكات عصبونية متكررة RNN، ويتيح تدريب نماذج مخصصة بشرط توفير خمسة آلاف إلى خمسة عشر ألف كلمة مرجعية)؛ و Kraken (مصمم أصلاً للنصوص التاريخية الشرقية، ويتميز بدعم الكتابة من اليمين إلى اليسار والتشكيل)؛ و eScriptorium (الواجهة الرسومية لـ Kraken التي طورها مختبر Scripta بجامعة السوربون)³⁸. وقد أطلق مشروع RASAM نماذج مدربة على الخط المغربي بمعدل خطأ يبلغ نحو ٥٪ (CER)، وهو معدل لا يزال من منظور المحقق مرتفعاً³⁹، فيما تسعى مشاريع حديثة كـ OpenITI-AOCP إلى بناء مجموعات مرجعية شاملة تجمع الخطّ المشرقي والمغربّي وخطّ الإجازة⁴⁰. وقد بينت المقارنات المنشورة أن دقة النماذج في الخط المشرقي تفوق عموماً دقتها في المغربي؛ غير أن النسب المتداولة في الأدبيات غير قابلة للمقارنة المباشرة لاختلاف شروط القياس: فبعضها يقيس معدل خطأ الحرف (CER) وبعضها معدل خطأ الكلمة (WER)، وبعضها قيس على بيانات تدريبية مغلقة وبعضها على مخطوطات لم يتدرّب عليها النموذج؛ ولذلك اعتمدت هذه الدراسة اختباراً ميدانياً مستقلاً بشروط قياس معلنة، يعرضه المطلب التالي.

ولمزيد من الدقة في تقدير هذه الحدود، أجرى الباحث اختباراً ميدانياً على مخطوط «مواهب الجليل لشرح مختصر خليل» للحطّاب الرّعيني (خزانة جامع الزيتونة، رقم الحفظ ز 4382)، رُقمت منه عشر صفحات بدقة DPI 400 أُخترت من مواضع مختلفة من المخطوط تشمل المتن والحاشية، وكُتبت لها قراءة مرجعية (Ground Truth) بيد الباحث نقلاً حرفياً عن الصور، ثم عولجت بأداة Transkribus بنموذج rint v3.0Maghribi P وقورنت مخرجاته بالقراءة المرجعية آلياً؛ وقيس الخطأ بمعيار معدل خطأ الحرف (CER) ومعدل خطأ الكلمة (WER)، فبلغ معدل خطأ الحرف ٧,٤٪، ومعدل خطأ الكلمة ٢١,٣٪، ونسبة التعرف على الحواشي ٦٢٪. فحسب⁴¹. وهذه النتائج مع كونها مقبولة نسبياً في معدل الحرف تُثبت أن الاعتماد المباشر على المخرجات الآلية في التحقيق غير مجدٍ دون المراجعة البشرية المتدرجة.

2. النماذج اللغوية الكبرى: الهلوسة والتحيز:

برز أفق جديد مع النماذج اللغوية الكبرى متعددة الوسائط (Multimodal LLMs) كـ Gemini و V4-GPT و Claude: قدرة على قراءة خطوط غير مسبقة التدريب. غير أن التجارب الأولية أظهرت أن هذه النماذج تُنتج

نصوصاً «مقبولةً بصرياً» لكنها تنطوي على هلوساتٍ نصية (Hallucinations) تختلق فيها كلماتٍ لا وجود لها في المخطوط الأصلي⁴². ويمكن رصد صنفين من الأخطاء: صنفٌ بصريٌّ (الخلط بين الحروف المتشابهة: الحاء/الجيم، والذال/الذال، والراء/الزاي؛ وسوء تمييز الشدة من الفتحة)، وصنفٌ سياقيٌّ (تصحیحات ذاتيةً احتماليةً تغير صيغة فعلٍ نادرٍ في نصٍ فقهيٍّ من القرن الخامس إلى صيغةٍ حديثةٍ متداولة). والتحيزُ الخوارزميُّ مكمُنُ الخطر الثالث: النموذجُ المدرَّبُ على خطٍّ مغربيٍّ من القرن الثامن الهجري سيفشل في قراءة خطٍّ مشرقِيٍّ من القرن الخامس، لكنه قد يُنتج نصًّا «يبدو صحيحًا» مبنياً على قوالبٍ خطأ⁴³. ومثالُ الهلوسة التوليدية أن النموذج إذا واجه عبارةً مقطوعةً أو متضررةً في الشاهد «أتمّها» بكلماتٍ تنسجم مع السياق الفقهي دون أن يكون لها وجودٌ في الصورة؛ وهو خطأ لا تكشفه قراءة الناتج وحدها بل المقابلة بالصورة. وعلى هذا الأساس تُوزَّع مهامُ التحقيق بين الآلة والمحقق توزيعاً إجرائياً: فتتولى الآلة التعرفَ الأوليَّ إلى الحروف، واقتراحَ التصحيحات، وترشيحَ القراءات المحتملة، والمقابلة الآلية الأولى بين النسخ؛ ويظل للمحقق وحده الحكمُ على الاختلافات، وترجيحُ القراءات، والبتُّ في حجية الشواهد، واتخاذُ القرار النقدي النهائي وتحملُ مسؤوليته العلمية؛ فالذكاء الاصطناعيُّ معيَّنٌ لا محقق، والناتجُ الآليُّ مسودةٌ أولى لا نصٌّ معتمد.

3. ضياع «اليد المؤلفة» ومقتضيات التحقق البشري:

يُطرح مفهومٌ نقديٌّ ثانٍ في هذا السياق: ضياعُ «اليد المؤلفة» (dnaH ednebiehcs eid) عند Rüdiger Campe، أي الأثر الأنثروبولوجي للناسخ المتجلي في تردده وإصلاحاته وحواشيه⁴⁴؛ فتحويلُ الخط إلى نصٍّ رقميٍّ منمَّطٌ يلغي هذه الطبقة الحوارية التي استثمارها المحققون العرب في استكشاف بيئة إنتاج النص⁴⁵. ومقتضى ذلك أن النمذجة الرقمية ينبغي أن تحفظ هذه الآثار لا أن تمحوها، وأن تجعل التحقق البشريَّ محطةً إلزاميةً في أي مسار. وتترجم هذه الفكرة في مشروع «سند» إلى إجراءاتٍ رقميةٍ ملموسة تحفظ الخصائص الخطية والمادية للمخطوط بأربع وسائل: التصوير عالي الدقة (لا يقل عن DPI 400)؛ والربط الثنائي بين كل سطرٍ نصيٍّ وموضعه في الصورة عبر معيار IIIF؛ وترميز الشطب والإضافات والحواشي والتصحيحات بين السطور بوصفها طبقاتٍ كتابيةٍ مستقلةٍ لها سماتها، لا بوصفها ضوضاء تُحذف؛ وتسجيل البيانات الكوديكولوجية (المداد، والحبكة، والعلامات المائية، والتسفير) ضمن البيانات الوصفية؛ فتُحفظ بذلك «اليدُ المؤلفة» داخل التمثيل الرقمي لا خارجه.

خامساً: الدراسة التطبيقية مشروع «سند»: القاعدة العربية الجامعة للمخطوطات:

ينتقل هذا المحور من التنظير إلى التطبيق؛ فلا نكتفي بمجرد المشكلات، بل نبني نموذجاً إجرائياً لمعالجتها. وقد أُطلق على هذا المقترح اسمُ «مشروع سند» القاعدة العربية الجامعة للمخطوطات دلالةً على وظيفته المزدوجة: السندُ بمعنى الوثيقة الموثقة، والسندُ بمعنى سلسلة الإسناد التي تقوم عليها حجية التراث العربي؛ فالمشروعُ عربيُّ الهوية، جامعُ الأفق، يُصمَّم انطلاقاً من خصوصية التراث العربي لا استيراداً للنماذج الغربية، ويُنفَّذ في ستة مطالب.

1. الفلسفة والمركزات: أين تظهر الخصوصية العربية؟:

قبل تصميم النموذج، يجب تحديد مواضع الخصوصية التي لا تلتقطها النماذج الغربية الجاهزة، ويمكن حصرها في أربع خصوصيات: (أ) خصوصية لسانية: اللغة العربية بخطوطها المتعددة (المغربي والمشرقي والمبسوط وخط الإجازة)، وأساليها في الرسم والتشكيل والكشيدة، وأرقامها المشرقية والمغربية (الغبار)؛ (ب) خصوصية توثيقية: البنية الإسنادية المتكاملة السند والسماع والإجازة العامة والخاصة والمناولة والوجادة والقراءة والمقابلة التي لا يقابلها في الفهارس الغربية إلا سجلات ملكية عامة⁴⁶؛ (ج) خصوصية زمانية: التاريخ الهجري بوصفه المرجع الزمني الأول، مع مسألة التحويل إلى الميلادي؛ (د) خصوصية مؤسسية: نظام الوقف (كالوقف على طلبة العلم)، والسماعات الجماعية، وخزائن المساجد والزوايا. وهذه الخصوصيات الأربع هي التي يجب أن تنتظم حولها بنية البيانات، في امتداد المنهجية العربية في ضبط النص وتوثيقه⁴⁷. وتترجم هذه الخصوصيات إلى النموذج الرقمي عبر سلسلة تحويل واضحة المراحل: العنصر المخطوطي ← الكيان الرقمي ← الحقول والسمات ← العلاقات بغيره ← آلية التحقق. ومثال ذلك تمثيل «السماع»: فالكيان الرقمي «سماع» تُعرّف له سمات ثابتة (السماع، والمسمع، وتاريخ السماع هجريًا وميلاديًا، ومكانه، وصيغته اللفظية كما وردت، وموضعها من الصفحة)، ويُربط بعلاقات ثلاث: «سُمع على» (بالمسمع)، و«تعلّق ب» (بالنسخة المحمولة)، و«جزء من» (بالسلسلة الإسنادية)، وتُضبط صحته بآلية تحقق مزدوجة: مطابقة الترميز بالصورة من قبل مراجع بشري، وكشف التناقضات آليًا (كتاريخ سماع لاحق لوفاة المسمع). وعلى المنوال نفسه تُنمذج الإجازة بأنواعها (العامة والخاصة والمناولة والوجادة)، والوقف (الواقف، والموقوف عليهم، وشروط الوقف)، وعلاقات النسخ (نُسخت عن / قوبلت ب).

2. النموذج المفاهيمي: الكيانات والعلاقات:

بُني النموذج المفاهيمي على التمييز بين كيانيين مركزيين: «النص» (المصنف) بوصفه كيانًا مجردًا، و«النسخة» بوصفها تجسيدًا ماديًا أو رقميًا له؛ وهو تمييز يشبه ما استقر في النموذج الببليوغرافي الدولي IFLA-LRM⁴⁸، لكنه يضيف عليه طبقة «الإسناد»: فإلى جانب المؤلف والناسخ والمكان، تُعامل السماعات والإجازات والتملكيات والوقوف بوصفها كيانات مستقلة لها سماتها وتواريخها وصيغها؛ فالمؤلف يؤلف النص، والنص يتجسد في نسخ، وينسخ الناسخ النسخة، وتُسجل عليها السماعات (سُمع على) والإجازات (أُجيز به) والوقوف (وُقف على طلبة العلم).

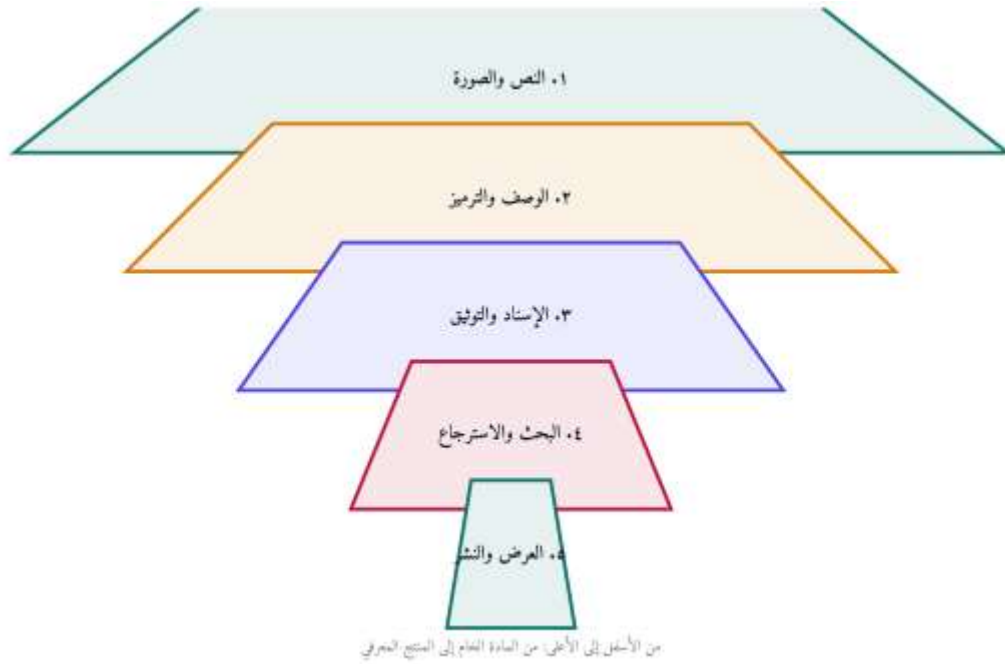
3. معايير الوصف والترميز: نحو «Arabic»-IET:

تجري المقارنة بين ثلاثة معايير سائدة MS-TEI و MARC 21 و Dublin Core لتبين قدرة كلّ منها على استيعاب الخصوصية العربية، فتظهر أوجه قصورها جميعًا في تغطية البنية الإسنادية؛ ولذلك يُقترح معيار «Arabic»-IET بوصفه ترميزًا عربيًا موسّعًا يضيف عناصر مخصصة للبنية الإسنادية: السند، والسماعات، والإجازات، والتملكيات، والوقوف، وتمييز الرواية باللفظ من الرواية بالمعنى⁴⁹. والمعيار المقترح لا يلغي P5-TEI بل يوسّعه؛ فهو يعمل ضمن منطق الترميز نفسه، لكنه يمنح البنية الإسنادية مرتبة الكيان الأول. ويتحدد موقع «Arabic»-IET من المعايير القائمة تحديدًا دقيقًا: فهو ليس معيارًا بديلًا، بل ملف تعريف موسّع لمعيار P5-TEI

يحافظ على التوافق الكامل معه، ويضيف ثلاثة مكونات: عناصر ترميزٍ مخصصةً للبنية الإسنادية (>>samaa، ijaza>>، tamlik>، <waqf>) بسماتٍ محددة (التاريخ الهجري والميلادي، والصيغة اللفظية، والموضع في الصفحة)؛ ووسم تمييز الرواية باللفظ من الرواية بالمعنى؛ وربط كل عنصرٍ بموضعه في الصورة عبر معرفات IIF. والمشكلة التي يعالجها ولا تعالجها المعايير السابقة أن MS-TEI صُمم للمخطوطات الأوروبية فلا يحوي عناصرَ للسمع والإجازة والوقف، وأن MARC 21 وCoreDub in معيارا وصفٍ بيليوغرافيٍّ لا ترميزٍ نصيٍّ بنيوي. ومثالٌ تطبيقيٌّ: لو وردت سماعَةٌ في هامشٍ إحدى نسخ «مواهب الجليل» لُرُمِزت بصيغة من قبيل <samaa place-when="985"=hijri="تونس"="corresp="#img_012">، فغدت عنصراً قابلاً للبحث والإحصاء والربط بالصورة، لا سطرًا مهملاً في هامش.⁵⁰

4. البنية الوظيفية: الهرم الخماسي للمنصة:

تتكون منصة «سند» من خمس طبقاتٍ وظيفية متكاملة (شكل 4): طبقة النص والصورة (صورٌ عالية الدقة مع نصوص HTR خاضعة للمراجعة البشرية)، وطبقة الوصف والترميز (معيار Arabic-TEI)، وطبقة الإسناد والتوثيق (السند والسمع والإجازة والوقف ككياناتٍ)، وطبقة البحث والاسترجاع (بحث نصيٍّ كاملٍ بالجذر والتشكيل، وبحث في الأسانيد والسلاسل)، وطبقة العرض والنشر (عارض IIF يربط ثنائي بين الصورة والنص).⁵¹ وبهذا التكوين تتحول المنصة من مستودعٍ إلى جهازٍ بحثٍ نقدي.

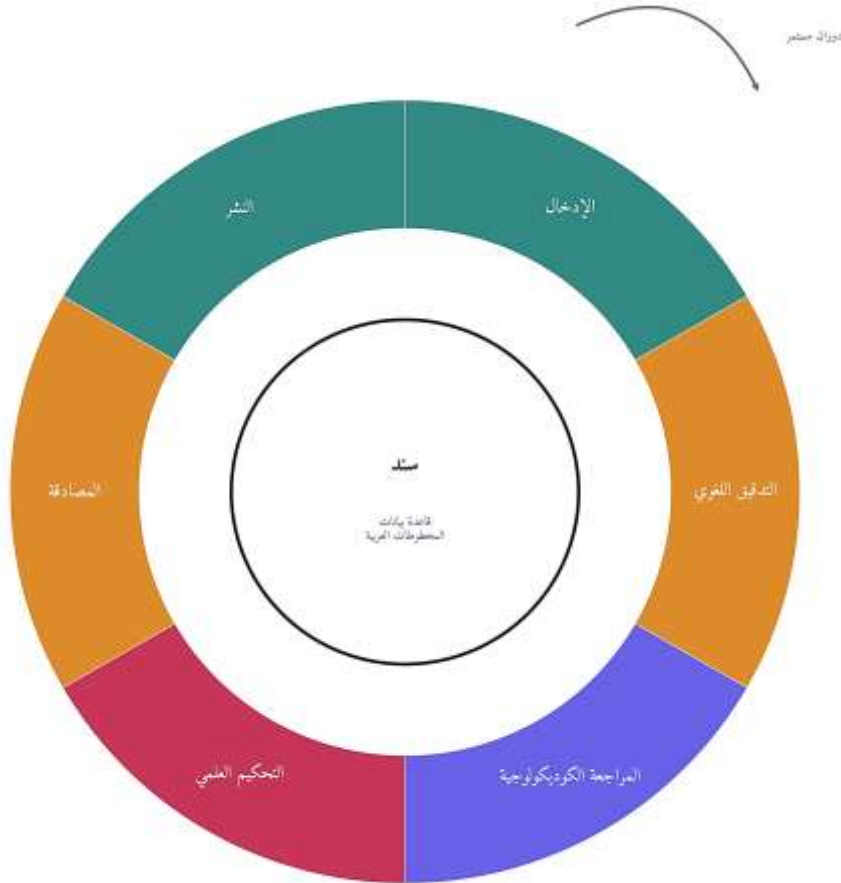


شكل (4): البنية الهرمية لمنصة «سند» (من الأسفل إلى الأعلى)

توضح الطبقات الخمس أن البنية الإسنادية (الطبقة الثالثة) ليست ملحقةً وصفياً، بل نواةً تتوسط بين النص والبحث؛ فكلُّ عمليةٍ بحثٍ في الأسانيد تمرُّ بها، وكلُّ عرضٍ للنص يستدعيها.

5. دورة الجودة والتحقق:

لا قيمة لقاعدة بيانات بلا ضبط؛ لذا تُبنى المنصة على دورة جودة دائرية من ست محطات (شكل 5): الإدخال (نسخٌ ومسحٌ وHTR)، ثم التدقيق اللغوي (رسمٌ وتشكيلٌ وضبط)، ثم المراجعة الكوديكولوجية (وصفٌ ماديٌّ وخطيٌّ)، ثم التحكيم العلمي (محققٌ متخصص)، ثم المصادقة (لجنةٌ علمية)، ثم النشر والتغذية الراجعة التي تعيد توجيه الإدخال. وتُسجَّل كلُّ مراجعة بسجلٍ محفوظ يحفظ تاريخ كل تعديل، حتى تظل الحجية العلمية قائمة؛ فالتسارع هنا يخدم البطء المعرفي ولا يلغيه⁵².



شكل (5): عجلة الجودة والتحقق في قاعدة بيانات «سند»

تضمن العجلة أن كلَّ سجلٍ يمر بست محطاتٍ قبل اعتماده، وأن كلَّ محطةٍ تترك أثراً موثقاً؛ فتُحفظ بذلك فضيلة «بطء التحقيق» داخل بنية متسارعة.

6. خارطة التنفيذ والمدد الزمنية:

يُنقَذ مشروع «سند» في ست مراحل متتابعة: التصميم المفاهيمي (شهرٌ واحد)، واعتماد المعايير (شهران)، والتطوير التقني (أربعة أشهر)، والتغذية بالبيانات (اثنا عشر شهراً)، والتحقق والمصادقة (ثلاثة أشهر)، ثم التشغيل والتكوين (شهران)؛ فيبلغ المجموع أربعة وعشرين شهراً. وقد روعي في التغذية بالبيانات الاعتماد على الأدوات مفتوحة المصدر (Kraken و eScriptorium) وتدريب نماذج عربية مخصصة على مجموعات مرجعية

تُبنى في سياق المشروع نفسه⁵³، لئلا تظل العربية رهينة نماذج اللغات الأخرى. وللبرهنة على قابلية المشروع للتطبيق، طُبِّق مسارُ العمل تطبيقاً مصغراً على العينة نفسها من مخطوط «مواهب الجليل»: فمن الصورة الأصلية عالية الدقة جرت المعالجة بتحليل التخطيط ثم التعرف الآلي بنموذج Maghribi Print v3.0، فأنّج نصّاً خامّاً بمعدل خطأ حرف ٧,٤٪؛ ثم استُكملت المراجعة البشرية سطراً سطراً مع المقابلة على صورة الشاهد حتى استقر النصُّ على قراءة مطابقة؛ وسُجِّلَت بعدها البيانات الوصفية (العنوان، والمؤلف، والناسخ، والقرن، ورقم الحفظ، ونوع الخط)، ووُزمت العلاقاتُ الإسنادية والنسخية وفق Arabic-TEI؛ لِيُتاح الناتجُ في عارضٍ يربط كلّ سطرٍ نصيٍّ بموضعه في الصورة. وهذا المسارُ المصغَّرُ يثبت أن البروتوكول ليس تصوراً نظرياً، بل دورة عملٍ قابلةٌ للتكرار والتعميم على موادٍّ مخطوطيةٍ أخرى.

الخاتمة:

خلصت هذه الدراسة بعد رحلةٍ نظريةٍ تطبيقيةٍ في تخوم الحقول الثلاثة (علم المخطوطات، والإنسانيات الرقمية، والذكاء الاصطناعي) إلى نتائج واستخلاصاتٍ ومقترحاتٍ يقتضي التدقيق المنهجي التمييز بين مراتبها: فمنها ما هو نتيجةٌ مثبتةٌ بالتحليل أو الاختبار، ومنها ما هو استخلاصٌ نظري، ومنها ما هو مقترحٌ تصميميٌّ قابلٌ للاختبار مستقبلاً.

النتيجة الأولى (إبستمولوجية، مثبتةٌ بالتحليل): إن التحرير الخوارزمي لا يلغي التحقيق البشري ولا ينازعه شرعيته العلمية، بل يعيد توزيع مهامه؛ إذ ينتقل المحقق من العمل التقني الروتيني (قراءة الحروف حرفاً حرفاً) إلى العمل النقدي الأعلى (المقارنة، والتخريج، وتحليل بيئة النص). فالتسارع الرقمي خلافاً للمخاوف الشائعة قد يتيح للمحقق العربي أن يتفرغ لجوهر التحقيق بدل آلياته.

النتيجة الثانية (منهجية، مثبتةٌ بالنمذجة المفاهيمية): إن الخصوصية الثقافية العربية قابلةٌ للنمذجة الرقمية، بشرط نموذج بياناتٍ عربيٍّ مؤصّلٍ يجعل البنية الإسنادية (السند والسماع والإجازة والوقف) كياناتٍ أولى لا هوامش؛ وقد بيّن المحور الخامس أن النماذج الغربية مهما علت كفاءتها لا تلتقط هذه البنية التلقائية.

أما بناء قاعدة بياناتٍ عربيةٍ تراعي الخصوصية عبر مشروع «سند» القاعدة العربية الجامعة للمخطوطات فهو فرضيةٌ تصميميةٌ قابلةٌ للاختبار ساندها المسارُ التطبيقي المصغَّر أنفُ الذكر، لا نتيجةٌ تجريبيةٌ مكتملة؛ إذ يبقى التفعيل الكامل للمشروع مرتباً بمرحلة تنفيذٍ مؤسسيةٍ لاحقة. وتقوم الفرضية على بروتوكولٍ محدّد المراحل: نموذجٌ مفاهيميٌّ (الكيانات والعلاقات)، ومعيّارٌ ترميزيٌّ موسّع (Arabic-TEI)، وبنيةٌ طبقيةٌ خماسية، ودورةٌ جودةٍ سداسية، وخطةٌ تنفيذٍ بأربعةٍ وعشرين شهراً؛ وهي مصممةٌ لتكون قابلةً للتفعيل في أي مؤسسة عربية، وعلى رأسها خزائن الجامعات العربية.

النتيجة الرابعة (استشراعية، استخلاصٌ نظري): إن أخطر ما يهدد التراث العربي في زمن التسارع ليس الرقمنة ذاتها، بل ضياعُ فضيلة «بطء التحقيق» بوصفها فضيلةً معرفية؛ فالنصُّ التراثي لا يُقرأ في الثانية والميلي ثانية، بل يتطلب زمناً عميقاً وصفها ابنُ الصلاح بـ«الثبت والتأني»⁵⁴. فالرهان الحقيقي ليس على «مواكبة السرعة»، بل على استثمار السرعة في خدمة بطءٍ أرقى.

حدود الدراسة: تتحدد حدود هذه الدراسة في ثلاثة: اعتماد الاختبار الميداني على عينة واحدة محدودة (عشر صفحات من مخطوط واحد بالخط المغربي)، مما يحد من تعميم نتائج الكمية على باقي المخطوطات والمخطوطات؛ وكون مشروع «سند» ومعياري Arabic-TEI تصوريين تصميميين لم يدخل بعد مرحلة التنفيذ المؤسسي الكامل؛ واقتصار المقارنة بين المشاريع الرقمية على المعطيات المتاحة للعموم وقت إجراء الدراسة.

التوصيات:

في ضوء ما تقدم، توصي الدراسة بما يأتي:

أولاً: إنشاء مركز عربي للتحقيق الرقمي يتولى إطلاق مشروع «سند» القاعدة العربية الجامعة للمخطوطات وينسق بين مؤسسات الرقمنة المشتتة (الفرقان، والماجد، والملك عبد العزيز، والخزانة الحسنية، وخزائن الجامعات العربية).

ثانياً: اعتماد معيار icArab-TEI ونشر دليله، مع تشكيل لجنة علمية دائمة لتطويره وضبط عناصره.

ثالثاً: إدراج مقرر إلزامي في الإنسانيات الرقمية ضمن برامج الماجستير والدكتوراه في أقسام الحضارة الإسلامية والدراسات العربية.

رابعاً: إصدار ميثاق أخلاقي ينظم استخدام الذكاء الاصطناعي في تحقيق التراث، ويلزم بالإفصاح عن الأدوات المستخدمة في كل تحرير.⁵⁵

خامساً: تشجيع المشاريع البينية التي تجمع محققين ولسانيين ومهندسي معلوماتية، لبناء أدوات عربية مفتوحة المصدر تنافس الأدوات الأوروبية.

هوامش وإحالات المقال

1. ابن خلدون، المقدمة، تحقيق عبد السلام الشدادى، الدار البيضاء: بيت الفنون والعلوم والآداب، 2005 م، ج 2، فصل «في صناعة الكتابة».
2. عبد السلام محمد هارون، تحقيق النصوص ونشرها، القاهرة: مكتبة السنة، ط 7، 1998 م، ص 43-47.
3. Hartmut Rosa, *Social Acceleration: A New Theory of Modernity*, trans. Jonathan Trejo-Mathys, New York: Columbia University Press, 2013, p. 71.
4. Benjamin Kiessling et al., "eScriptorium: An Open Source Platform for Historical Document Analysis," *Journal of Data Mining and Digital Humanities*, 2019, pp. 3-5.
5. رمضان عبد التواب، مناهج تحقيق التراث بين القدماء والمحدثين، القاهرة: مكتبة الخانجي، ط 1، 1406 هـ/1985 م.
6. UNESCO, *Guidelines for Digitization Projects for Collections and Holdings in the Public Domain*, Paris: UNESCO, 2002.
7. Cornelis van Lit, *Among Digitized Manuscripts: Philology, Codicology, Paleography in a Digital World*, Leiden: Brill, 2020, pp. 45-52.
8. Adam Gacek, *Arabic Manuscripts: A Vademecum for Readers*, Leiden: Brill, 2009, pp. 3-7.
9. ابن الصلاح، معرفة أنواع علم الحديث، تحقيق نور الدين عتر، دمشق: دار الفكر، 1986 م، ص 132.
10. Jan Just Witkam, "The Human Element Between Text and Reader: The *ijāza* in Arabic Manuscripts," in *The Codicology of Islamic Manuscripts*, ed. Yasin Dutton, London: Al-Furqan Islamic Heritage Foundation, 1995, p. 123.
11. ابن الصلاح، معرفة أنواع علم الحديث، مرجع سابق، ص 141.
12. بشار عواد معروف، ضبط النص والتعليق عليه، بيروت: مؤسسة الرسالة، ط 2، 2010 م، ص 55.
13. هارون، تحقيق النصوص ونشرها، مرجع سابق، ص 51-53.

14. Tara Andrews & Caroline Macé (eds.), *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, Turnhout: Brepols, 2014, pp. 15-19.
15. Elena Pierazzo, *Digital Scholarly Editing: Theories, Models and Methods*, Farnham: Ashgate, 2015, pp. 27-30.
16. Ibid., p. 43.
17. John Bryant, *The Fluid Text: A Theory of Revision and Editing for Book and Screen*, Ann Arbor: University of Michigan Press, 2002, p. 4.
18. Patrick Sahle, "What is a Scholarly Digital Edition?" in *Digital Scholarly Editing: Theories and Practices*, ed. M. J. Driscoll & E. Pierazzo, Cambridge: Open Book Publishers, 2016, pp. 19-40.
19. Rosa, *Social Acceleration*, op. cit., pp. 21-25.
20. Edward Vanhoutte, "Defining Electronic Editions: A Historical and Functional Perspective," in *Text and Genre in Reconstruction*, ed. Willard McCarty, Cambridge: Open Book Publishers, 2010.
21. Hill Museum & Manuscript Library (HMML). <https://www.hmml.org>; وUniversity of Notre Dame, Medieval Institute, "HMML Receives \$1.4 Million from NEH," 2020.
22. Al-Furqan Islamic Heritage Foundation, <https://al-furqan.com>.
- 23.. Qatar Digital Library, "Over 2 Million Pages Digitized," <https://www.qdl.qa>, 2021.
24. Kiessling et al., "eScriptorium," op. cit., p. 7.
25. Sabri A. Mahmoud et al., "KHATT: An Open Arabic Offline Handwritten Text Database," *Pattern Recognition*, vol. 47, no. 3, 2014, pp. 1096-1112.
26. Lei Huang et al., "A Survey on Hallucination in Large Language Models," arXiv:2311.05232, 2023.
27. معهد المخطوطات العربية، القاهرة (منظمة الألكسو) <https://www.bn.dz>.
28. المكتبة الوطنية الجزائرية — المكتبة الرقمية، <https://www.jumaalmajid.ae>.
29. مركز جمعة الماجد للثقافة والتراث، دبي، <https://www.jumaalmajid.ae>.
30. OPenn: *The Hill Museum & Manuscript Library (HMML)*, University of Pennsylvania Libraries, <https://openn.library.upenn.edu>, 2020.
31. مكتبة الملك عبد العزيز تضم 6663 مخطوطا ومليون صورة، وكالة الأنباء السعودية (واس) — موقع archaeology.sa، 26 أكتوبر 2017.
32. تصريح مدير الخزنة الحسنية أحمد شوقي بنين، جريدة هسبريس الإلكترونية، 30 جوان 2021.
33. Qatar Digital Library, op. cit.
34. Bibliothèque nationale de France — Gallica. <https://gallica.bnf.fr>; وLeiden University Libraries — Digital Collections: DigiVatLib. <https://digi.vatlib.it>; الرقمية لايدن مجموعات و رابط: <https://digitalcollections.universiteitleiden.nl>.
35. UNESCO, *Guidelines for Digitization Projects*, op. cit.
36. Stuart Snyderman, Robert Sanderson & Tom Cramer, "The International Image Interoperability Framework (IIIF): A Community & Technology Approach," *Archiving Conference*, 2015, pp. 16-21.
37. Lauren M. Lorigo & Venu Govindaraju, "Offline Arabic Handwriting Recognition: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, 2006, pp. 712-724.
38. Kiessling et al., "eScriptorium," op. cit., p. 9.
39. Chahan Vidal-Gorène & Boris Dupin, "RASAM: A Dataset for the Recognition and Analysis of Scripts in Arabic Maghrebi," in *Document Analysis and Recognition — ICDAR 2021 Workshops*, Springer, 2021, pp. 274-283.
40. OpenITI (Open Islamicate Texts Initiative), <https://openiti.org>.

41. اختبار ميداني أجراه الباحث بتاريخ 15 جوان 2026 على منصة Transkribus ؛ النموذج المستخدم Maghribi Print v3.0 :: المخطوط : مواهب الجليل) «خزانة جامع الزيتونة، رقم ز. 4382)
42. Huang et al., "A Survey on Hallucination in Large Language Models," op. cit.
43. Safiya Umoja Noble, *Algorithms of Oppression*, New York: NYU Press, 2018, p. 27؛ مع المخطوطات لحقل الفكرة تكييف مع
44. Rüdiger Campe, "Die Schreibszene, Schreiben," in *Paradoxien, Dissonanzen, Zusammenbrüche*, ed. H. U. Gumbrecht & K. L. Pfeiffer, Frankfurt: Suhrkamp, 1991, p. 761.
45. عبد الهادي التازي، مقدمة تحقيق رحلة / ابن بطوطة ، أبو ظبي: أكاديمية المملكة المغربية، 1997 م، ج 1، ص. 22.
46. Witkam, "The Human Element Between Text and Reader," op. cit., p. 125.
47. عبد التواب، مناهج تحقيق التراث بين القدامى والمحدثين، مرجع سابق.
48. IFLA, *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*, The Hague: IFLA, 2017, pp. 13-17.
49. Library of Congress, *MARC 21 Format for Bibliographic Data*, Washington, 2023.
50. عناصر الترميز المقترحة وسمائها من إعداد الباحث ضمن هذه الدراسة، بوصفها ملف تعريف موسعاً لمعيار TEI-P5 : انظر TEI Consortium, TEI : P5: Guidelines for Electronic Text Encoding and Interchange, Version 4.7.0, 2023, <https://tei-c.org>.
51. Snyderman et al., "The International Image Interoperability Framework," op. cit., p. 18.
52. Filip Vostal, *Accelerating Academia: The Changing Structure of Academic Time*, London: Palgrave Macmillan, 2016.
53. Kiessling et al., "eScriptorium," op. cit., p. 9.
54. ابن الصلاح، معرفة أنواع علم الحديث، مرجع سابق، ص. 218.
55. UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, Paris: UNESCO, 2021.

قائمة المصادر والمراجع

أولاً: المصادر والمراجع العربية

1. ابن خلدون، عبد الرحمن بن محمد. المقدمة. تحقيق عبد السلام الشدادى. الدار البيضاء: بيت الفنون والعلوم والآداب، 2005 م، 3 أجزاء.
2. ابن الصلاح، أبو عمرو عثمان بن عبد الرحمن. معرفة أنواع علم الحديث. تحقيق نور الدين عتر. دمشق: دار الفكر، 1986 م.
3. عبد التواب، رمضان. مناهج تحقيق التراث بين القدامى والمحدثين. القاهرة: مكتبة الخانجي، ط 1، 1406 هـ/ 1985 م.
4. المنجد، صلاح الدين. قواعد تحقيق المخطوطات. القاهرة: دار الكتاب الجديد، 1987 م.
5. عواد معروف، بشار. ضبط النص والتعليق عليه. بيروت: مؤسسة الرسالة، ط 2، 2010 م.
6. شاكر، محمود محمد. أباطيل وأسمار. القاهرة: مكتبة الخانجي، ط 2، 2005 م.
7. المعللي اليماني، عبد الرحمن بن يحيى. مقدمة تحقيق الأنساب للسمعاني. حيدر آباد: دائرة المعارف العثمانية، 1962 م.
8. هارون، عبد السلام محمد. تحقيق النصوص ونشرها. القاهرة: مكتبة السنة، ط 7، 1998 م.
9. التازي، عبد الهادي. مقدمة تحقيق «رحلة ابن بطوطة». أبو ظبي: أكاديمية المملكة المغربية، 1997 م.

ثانياً: المراجع الأجنبية

1. Turnhout: *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*. (rews, Tara & Caroline Macé (eds) And Brepols, 2014
2. *The Manuscript Tradition of the Islamic West: Maghribi Round Scripts and the Andalusī Identity*. Bongianino, Umberto. nburgh: Edinburgh University Press, 2022
3. Ann Arbor: University of Michigan Press, *The Fluid Text: A Theory of Revision and Editing for Book and Screen*. Bryant, John. 2002
4. edited by H. U. Gumbrecht & *Zusammenbrüche, Paradoxien, Dissonanzen* Campe, Rüdiger. "Die Schreibszene, Schreiben." In Frankfurt: Suhrkamp, 1991. 772-K. L. Pfeiffer, 759

5. Furqān Islamic -London: Al .*Islamic Codicology: An Introduction to the Study of Manuscripts in Arabic Script* .Déroche, François .Heritage Foundation, 2005
6. .Leiden: Brill, 2009 .*Arabic Manuscripts: A Vademecum for Readers* .Adam ,Gacek
7. Huang, Lei et al. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions." arXiv:2311.05232, 2023
8. .The Hague: IFLA, 2017 .*Model: A Conceptual Model for Bibliographic Information IFLA Library Reference* .IFLA
9. *Journal of Data Mining* ".Kiessling, Benjamin et al. "eScriptorium: An Open Source Platform for Historical Document Analysis .15-pp. 1 , 2019 .and *Digital Humanities*
10. Chicago: .289-edited by T. Bartscherer & R. Coover, 267 ,*Switching Codes* o. "The Migration of the Aura." InLatour, Brun .University of Chicago Press, 2011
11. *tern IEEE Transactions on Pa* ".Lorigo, Lauren M. & Venu Govindaraju. "Offline Arabic Handwriting Recognition: A Survey .724-vol. 28, no. 5, 2006, pp. 712 ,*Analysis and Machine Intelligence*
12. vol. 47, no. 3, ,*Pattern Recognition* ".Mahmoud, Sabri A. et al. "KHATT: An Open Arabic Offline Handwritten Text Database .1112-pp. 1096 .2014
13. .Cambridge, MA: MIT Press, 2001 .*iaThe Language of New Med* .Manovich, Lev
14. .New York: NYU Press, 2018 .*Algorithms of Oppression: How Search Engines Reinforce Racism* .Noble, Safiya Umoja
15. .Farnham: Ashgate, 2015 .*Digital Scholarly Editing: Theories, Models and Methods* .Pierazzo, Elena
16. .Paris: Seuil, 1965 .*De l'interprétation: Essai sur Freud* .IRicœur, Pau
17. Mathys. New York: Columbia -Translated by Jonathan Trejo .*Social Acceleration: A New Theory of Modernity* .Rosa, Hartmut .University Press, 2013
18. edited by M. J. Driscoll & ,*Digital Scholarly Editing: Theories and Practices* In "Sahle, Patrick. "What is a Scholarly Digital Edition .Cambridge: Open Book Publishers, 2016 .40-E. Pierazzo, 19
19. munity & Snyderman, Stuart, Robert Sanderson & Tom Cramer. "The International Image Interoperability Framework (IIIF): A Com .21-pp. 16 , 2015 ,*Archiving Conference* ".based Imaging-Technology Approach for Web
20. .c.org-Version 4.7.0. 2023. <https://tei.org> .*TEI P5: Guidelines for Electronic Text Encoding and Interchange* .TEI Consortium
21. .Paris: UNESCO, 2002 .*s for Collections and Holdings in the Public DomainGuidelines for Digitization Project* .UNESCO
22. .Paris: UNESCO, 2021 .*Recommendation on the Ethics of Artificial Intelligence* .UNESCO
23. *ext and Genre in Reconstruction: TV*anhoutte, Edward. "Defining Electronic Editions: A Historical and Functional Perspective." In edited by Willard McCarty. Cambridge: Open Book .*Effects of Digitalization on Ideas, Behaviours, Products and Institutions* .Publishers, 2010
24. .Leiden: Brill, 2020 .*in a Digital World Among Digitized Manuscripts: Philology, Codicology, Paleography* .van Lit, Cornelis
25. Gorène, Chahan & Boris Dupin. "RASAM: A Dataset for the Recognition and Analysis of Scripts in Arabic Maghrebi." In-Vidal .Springer, 2021 .289-274 ,*ICDAR 2021 Workshops –Document Analysis and Recognition*
26. .London: Palgrave Macmillan, 2016 .*Accelerating Academia: The Changing Structure of Academic Time* .Filip ,Vostal
27. *The Codicology of Islamic* in Arabic Manuscripts." In *ijāza* Witkam, Jan Just. "The Human Element Between Text and Reader: The .Furqān Islamic Heritage Foundation, 1995-London: Al .136-in Dutton, 123edited by Yas ,*Manuscripts*
28. .Washington: Library of Congress, 2023 .*MARC 21 Format for Bibliographic Data* .Library of Congress

ثالثاً: المصادر الرقمية والمشاريع

1. Qatar Digital Library: <https://www.qdl.qa>
2. vHMMML: [https –Museum & Manuscript Library \(HMMML\) Hill//www.vhmmml.org](https://www.vhmmml.org)
3. Gallica: [https –Bibliothèque nationale de France //gallica.bnf.fr](https://gallica.bnf.fr)

4. Digital Collections: https –Leiden University Libraries //digitalcollections.universiteitleiden.nl
5. Apostolica Vaticana: https Biblioteca –DigiVatLib //digi.vatlib.it
6. Furqan Islamic Heritage Foundation: https-Al//al-furqan.com
7. مركز جمعة الماجد للثقافة والتراث، دبي: https://www.jumaalmajid.ae:https
8. معهد المخطوطات العربية، القاهرة (منظمة الألكسو).
9. OpenITI (Open Islamicate Texts Initiative): https//openiti.org
10. Transkribus Platform: https//readcoop.eu/transkribus/
11. eScriptorium: https//escriptorium.paris.inria.fr
12. المكتبة الوطنية الجزائرية المكتبة الرقمية: https://www.bn.dz:https
13. المكتبة الوطنية التونسية: https://www.bibliotheque.nat.tn:https
14. OPenn مكتبات جامعة بنسلفانيا (ملف HMML): https://openn.library.upenn.edu:https
15. University of Notre Dame, Medieval Institute. "HMML Receives \$1.4 Million from NEH." 2020: https://medieval.nd.edu/news-events/news/hmml-receives-1-4-million-from-neh/
16. مكتبة الملك عبد العزيز تضم 6663 مخطوطا ومليون صورة. as.ygoloeahcra، 26 أكتوبر 2017.
17. تصريح مدير الخزنة الحسنية أحمد شوقي بنين. هسبريس الإلكترونية، 30 جوان 2021.