

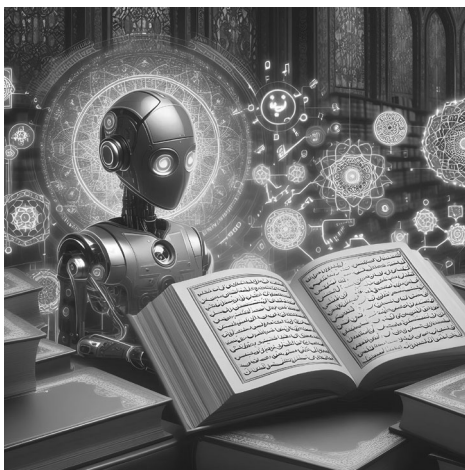
الذكاء الاصطناعي والتعامل الذكي

مع النصوص والمخطوطات العربية

Artificial Intelligence and Intelligent Dealing
with Arabic Texts and Manuscripts

أ.د/ أشرف عبد المحسن الشريف





الذكاء الاصطناعي والتعامل الذكي مع النصوص والمخطوطات العربية

جميع الحقوق محفوظة. لا يسمح بإعادة إصدار هذا الكتاب أو أي جزء منه، أو تخزينه في نطاق استعادة المعلومات، أو نقله بأي شكل من الأشكال، دون إذن مسبق من الناشر.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means without prior written permission of the publisher.

DAR AL FIKER

جميع حقوق الطبع محفوظة ©



www.daralfiker.com

كتابي الصغير
للنشر والتوزيع



www.ketabi.jo

المملكة الأردنية الهاشمية - عّان
شارع رشيد المدفعي - بجانب المسجد الحسيني
وسط البلد (د3) عمارة الحجيري
هاتف: +962 6 4621938
واتساب: +962 7 77695000
ص.ب: 183520 عّان 1118 الأردن
بريد إلكتروني: info@daralfiker.com
بريد المبيعات: sales@daralfiker.com

التصنيف: 411.40285

تأليف: أ. د أشرف عبد الحسن الشيريف

الذكاء الاصطناعي والتعامل الذكي مع النصوص
والخطوط العربية

رقم الإيداع: 2025/9/5610

الوصفات: الخطوط // الخط العربي // النسخ

المصورة // تطبيقات الحاسوب // الذكاء الاصطناعي

* أعدت دائرة المكتبة الوطنية بيانات الفهرسة والتصنيف الأولية
* يتحمل المؤلف كامل المسؤولية القانونية عن محتوى مصنفه ولا
يعبر هذا المصنف عن رأي دائرة المكتبة الوطنية أو أي جهة
حكومية أخرى.

الطبعة الأولى

1447- 2026

مواقع التواصل الاجتماعي



daralfiker1978@gmail.com



daralfiker/



dar_alfiker



dar al-fiker for publishing & distributing



pub/dar-alfiker

ISBN: 978-9957-92-336-5

الذكاء الاصطناعي والتعامل الذكي

مع النصوص والمخطوطات العربية

*Artificial Intelligence and Intelligent Dealing
with Arabic Texts and Manuscripts*

إعداد

أ. د/ أشرف عبد المحسن الشريف

كلية الآداب - جامعة بني سويف

الطبعة الأولى

1447 - 2026

دار الفکر

ناشرون وموزعون

الفهرس

13 مقدمة الكتاب

الفصل الأول: التعلُّم العميق- المفهوم والأساسيات

15 مفهوم الذكاء الاصطناعي

16 نشأة التعلُّم العميق

17 أسباب انتشار التعلُّم العميق

19 تحديات التعلُّم العميق

22 أولاً: عدم توفر بيانات التدريب

22 ثانياً: إنشاء مجموعة بياناتٍ تدريبية يدوياً

25 مستقبل التعلُّم العميق

26 تطبيقات التعلُّم العميق

26 أولاً: المساعدون الافتراضيون

27 ثانياً: جمع الأخبار وكشف الاحتيال

27 ثالثاً: الذكاء العاطفي

27 رابعاً: الرعاية الصحية

28 طرق التعلُّم الآلي

36 إنشاء نموذج لاستخدام التعلُّم العميق

41 التصنيف الدلالي للنصوص باستخدام التعلُّم العميق

43 نماذج التعلُّم العميق

48 نماذج التعلُّم العميق الحديثة المتعلقة بمعالجة اللغة الطبيعية

الفصل الثاني: التعلُّم العميق والتعرف الذكي على نصوص المخطوطات الإسلامية

57	تصنيف المخطوطات المجهولة باستخدام التعلُّم العميق
63	تصميم تجربة لتصنيف الكلمات بالمخطوطات الإسلامية
65	1 - اقتراح إطار بديل للتجربة في المرحلة الأولى
70	2 - إجراء وتنفيذ التجربة
75	3 - نتائج التجربة
76	التعرف على المخطوطات العربية مجهولة المؤلف - تجربة عملية
78	أولاً: التنبؤ باستخدام الخوارزميات التقليدية
81	ثانياً: التنبؤ باسم مؤلف المخطوطة باستخدام خوارزميات التعلُّم العميق
83	الخطوات التنفيذية لإجراء التجربة
99	المقارنة بين نتائج نماذج التعلُّم العميق الأربعة المطورة
101	خاتمة التجربة
102	التعرف الذكي على المخطوطات المجهولة التاريخ
401	تاريخ أسلوب الكتابة العربية والمخطوطات
107	تجربة عملية للتعرف على تاريخ المخطوطة المفقود
109	مقارنة KERTAS مع مجموعات البيانات الموجودة

الفصل الثالث: التعلُّم العميق والتنبؤ بالحروف المكتوبة بخط اليد في المخطوطات الإسلامية

119	التنبؤ بأسلوب خط اليد في المخطوطات الإسلامية القديمة
123	تجربة عملية للتنبؤ بأشكال الكتابة اليدوية في المخطوطات العربية
124	1 - منهجية تطبيق التجربة
125	2 - إعداد وتطوير النموذج
128	3 - نتائج التجارب والاختبارات
139	تقنية ICR للتعرف على الحروف المكتوبة بخط اليد في الوثائق العربية

140	التعريف بتقنية ICR
140	الفرق بين تقنية (ICR) وتقنية (OCR)
141	الفرق بين تقنية ICR وتقنية HWR
141	فوائد استخدام تقنية ICR
142	المراحل العملية لتطبيق تقنية ICR على الوثائق العربية
142	1 - مرحلة ما قبل المعالجة
143	2 - مرحلة المعالجة للتعرف على الحروف
143	3 - مرحلة ما بعد المعالجة
144	تطبيق تقنية ICR على نماذج من الوثائق العربية في العصر العثماني
146	أولاً: المدخلات
147	ثانياً: المعالجة
149	ثالثاً: المخرجات

الفصل الرابع: أمثلة عملية على استخدام الذكاء الاصطناعي

155	الذكاء الاصطناعي وتحليل النصوص
159	المثال الأول: منصة Transkribus لقراءة وتحليل المخطوطات التاريخية
166	ثانياً: مشروع مصحف مسقط الإلكتروني
169	ثالثاً: مشروع التعرف الآلي على الحروف بالتعاون بين المكتبة البريطانية ومكتبة قطر
173	رابعاً: مبادرة النصوص الإسلامية المفتوحة (OpenITI)
179	الخاتمة
181	خاتمة الكتاب
185	قائمة المراجع
185	قائمة المصطلحات

فهرس الأشكال

- الشكل (1) يبين طريقة توزيع مجموعة البيانات في النصوص العربية 32
- الشكل (2) نموذج الشبكة العصبية التكرارية (ReNN) 44
- الشكل (3) يبين نموذجًا لشبكة عصبية عميقة 45
- الشكل (4) يبين نموذجًا لشبكة عصبية تلافيفية 46
- الشكل (5) نموذج اللغة العصبية المقترح بواسطة Bengio 50
- الشكل (6) يبين هندسة المحولات 52
- الشكل 7 التعلم الآلي التقليدي مقابل التعلم الانتقالي 54

أشكال الفصل الثاني

- الشكل (1) أمثلة لبطاقات الفهرس ذات المعلومات والتخطيط المختلفين 64
- الشكل (2) الإطار القائم على الصور 65
- الشكل (3) شبكة R-CNN أسرع لمعالجة بطاقات المخطوطة 66
- الشكل (4) مثال لمنطقة الاهتمام بصورة واحدة 68
- الشكل (5) شريحة واحدة من مخطوطة غير معروفة والبطاقة المرتبطة بها 70
- الشكل (6) مثال على بطاقات الكتالوج: (أ) مخطوطة غير معروفة (ب) مخطوطة معروفة 71
- الشكل (7) المناطق الناتجة باستخدام R-CNN الأسرع 74
- الشكل (8) بنية النماذج المطورة 86
- الشكل (9) الصور الأصلية. الشكل 8. ب الصور المعززة 87
- الشكل (10) بنية طبقات الشبكات العصبية التلافيفية (CNN) 91
- الشكل (10) النقش الكوفي العربي ما قبل الإسلام 104
- الشكل (12) مخطوطة قرآنية مبكرة محفوظة في مكتبة جامعة برمنجهام 105
- الشكل (13) أمثلة على المسائل الرياضية من مكتبة قطر الوطنية من القرن الخامس 106

- الشكل (14) يبين هيكل الدليل لقاعدة بيانات كيرناس، ب- محتوى المخطوطة رقم 0101 في القرن الأول، ج- ملف XML مع بياناتٍ وصفية للمخطوطة.....110
- الشكل (15) صفحة مثال تحتوي على الأشكال وب صفحة مثال تحتوي على الجداول من مكتبة ويلكوم.....112

أشكال الفصل الثالث

- الشكل (1) يبين بنية النموذج المطوّر MobileNet126
- الشكل (2) صور المخطوطات المكتوبة باستخدام أنماط مختلفة من الخط العربي.....127
- الشكل (3). مصفوفة الارتباك الناتجة عن نسب مجموعات البيانات للتدريب بنسبة 60 % والاختبار بنسبة 40 % والتحقق من الصحة ..133
- الشكل (4). مصفوفة الارتباك الناتجة عن نسب مجموعات البيانات التدريبية بنسبة 70 % والتحقق من الصحة. 134
- الشكل (5). مصفوفة الارتباك الناتجة عن نسب مجموعات البيانات للتدريب بنسبة 80 % والاختبار بنسبة 20 % والتحقق من الصحة ..135
- الشكل (6). واجهة المستخدم الرسومية للنموذج المقترح.136
- الشكل (7) يوضح الواجهة الرئيسية للتطبيق النموذج المطوّر للتنبؤ137
- الشكل (8) وثيقة مفردة من محكمة الباب العالي في العصر العثماني 1174هـ.....145
- الشكل (9) يبين نماذج من الوثائق في العصر المملوكي والعصر العثماني.....146
- الشكل (10) نماذج من صور الوثائق عينة التطبيق.147
- الشكل (11) نماذج تعرض نتائج تقطيع سطور النص وكلماته في الوثائق عينة التطبيق.148

أشكال الفصل الرابع

- الشكل (1) يوضح الصفحة الرئيسية لمنصة Transkribus161
- الشكل (2) يبين جهاز ScanTent الذي يقوم بإجراء المسح الضوئي بمنصة Transkribus162
- الشكل (3) يبين استخدام جهاز المسح الضوئي من قبل الباحثين164
- الشكل (4) يبين طريقة تحميل صور المخطوط على موقع Transkribus165
- الشكل (5) يبين قراءة الوثائق التاريخية المكتوبة بخط اليد إلى نصّ آليّ166
- الشكل (6) يبين موقع مصحف مسقط على شبكة الإنترنت167

- الشكل (7) يبين موقع مشروع رقمنة المخطوطات العربيّة بالمكتبة البريطانية 169
- الشكل (8) يبين المخطوطات العربيّة التي تمت رقمنتها ورفعها على موقع المشروع 172
- الشكل (9) يبين موقع مبادرة Openiti على شبكة الإنترنت 175
- الشكل (10) يبين نسخة رقمية نموذجية لنص عمر بن سعيد على الويب وصفحة المخطوطة
المقابلة لها 177
- الشكل (11) نسخة رقمية نموذجية من صفحة من رسالة فارسية مع صفحة المخطوطة المقابلة لها 177
- الشكل (12) يبين فيديو على قناة المبادرة على موقع اليوتيوب طريقة ترجمة وتحليل المخطوطات 178

فهرس الجدول

الجدول 1 قائمة المكتبات المستخدمة.....	72
الجدول 2 قائمة المؤلفين وتفاصيل المخطوطات المرتبطة بهم.....	84
الجدول 3 نتائج التنفيذ الأولية للنماذج الأربعة.....	93
الجدول 4 MOBILENETV1 ضمن معدلات التعلم المختلفة.....	94
الجدول 5 RESNET50 ضمن معدلات تعلم مختلفة.....	94
الجدول 6 شبكة DENSENET201 ضمن معدلات تعلم مختلفة.....	94
الجدول 7 شبكة VGG19 ضمن معدلات تعلم مختلفة.....	95
الجدول 8 MOBILENETV1 من خلال عدد مختلف من الخلايا العصبية.....	98
الجدول 9 RESNET50 من خلال عدد مختلف من الخلايا العصبية.....	98
الجدول 10 DENSENET201 من خلال عدد مختلف من الخلايا العصبية.....	98
الجدول 11 VGG19 من خلال عدد مختلف من الخلايا العصبية.....	98
الجدول 12 يبين نماذج التعلم العميق ومدى تحقيق نتائج متقدمة.....	100
الجدول 13 مصادر مجموعة بيانات KERTAS.....	108
الجدول 14 مجموعات البيانات المماثلة لـ KERTAS.....	111
الجدول 15 نتائج التقييم لخوارزمية الكشف عن عمر المخطوطة المستندة إلى التمثيل المتفرق.....	116
الجدول 16 نتائج السمات المستندة إلى أسلوب الكتابة على مجموعة بيانات KERTAS.....	116

جداول الفصل الثالث

الجدول 1 معلمات التقييم لكل نمط من أنماط الكتابة اليدوية.....	130
الجدول 2 المقاييس المحسوبة لنسب مجموعات البيانات المختلفة.....	132
الجدول 3 المقارنة مع الأساليب الحديثة.....	138

مقدمة الكتاب

يعدُّ اكتشاف التشابه الدلالي في اللغة العربيّة تحديًا كبيرًا، ويصبح أكثر تعقيدًا في حالة اللغة العربيّة الفصحى. ويقدم القرآن مثالاً فريدًا للغة العربيّة الفصحى في أقصى أشكالها، وقد ركزت معظم الأبحاث في اكتشاف التشابه الدلالي في اللغة الإنجليزيّة، على الرغم من تناول الدراسات الحديثة في اللغة العربيّة موضوع التشابه الدلالي، والقليل من الأعمال بحثت في التشابه الدلالي في النص القرآني، واستخدم بعضها الترجمة الإنجليزيّة للقرآن، بينما نظر البعض الآخر في فصولٍ محددة من النص، من أجل الكشف عن التشابه الدلالي في نصوص القرآن.

وقد عمل علماء الحاسب الآليّ بالتعاون مع المتخصصين في اللغة العربيّة إلى تطوير أساليب حاسوبية جديدة للكشف عن التشابه الدلالي في اللغة العربيّة، والمساهمة في الجهود المستدامة للتعامل مع هذه النصوص. إنّ تقنيّة الذكاء الاصطناعي ممثلة في التعلّم العميق لديها القدرة على الكشف عن البنية الجوهرية وراء الكتاب القديم، والتي لم يكن من الممكن أن يكشفها التحقيق اليدوي، كما يمكنها أيضًا الكشف عن بعض النتائج الحاسوبية الجديدة التي لا يغطيها أي تفسير مما يضيف إلى مجموعة المعرفة الإسلاميّة.

ويقوم الباحثون المتخصصون باستخدام الذكاء الاصطناعي بكل فروعهِ من التعلّم العميق والشبكات العصبيّة العميقة ومعالجة اللغات الطبيعيّة، وذلك للحكم على صحة الحديث متّناً وسندًا بمنتهى الدقة والحيادية، وهنا يكون الاجتهادُ في تطوير «علم الآلة» ليتمكننا من معرفة درجة صحة الحديث النبوي بنسبة مئوية، وليس من خلال أسلوبٍ لفظي أو رمزي.

ولم يقف تأثيرُ الذكاء الاصطناعي ممثلًا في التعلّم العميق على الكتب الدينية المقدسة، سواء القرآن الكريم أو الحديث النبوي الشريف، بل تعدى الأمرُ إلى المخطوطات التاريخية التي أنتجتها الحضارة العربيّة الإسلاميّة طوال قرون ازدهارها. وتشير بعض التقديرات إلى وجود ما لا يقل عن 3 ملايين مخطوطٍ عربي في المكتبات المتخصصة المنتشرة في مختلف مناطق العالم. وقد ظلّ الوصول إليها مقصورًا على الباحثين والمتخصصين في مجال المخطوطات، لذلك لم تشمل جهودُ تحقيقها وطباعتها سوى بضعة آلاف منها.

وقد مكن التقدم العلمي والتقني من تطوير تقنيات «التعرف البصري على الأحرف» (OCR) التي تتيح التحويل الإلكتروني لصور النص المكتوب، أو المكتوب بخط اليد إلى نص قابلٍ للتعديل. لم يقف الأمر في التعرف البصري على الحروف فقط، بل ظهرت تقنية التعرف الذكي على الحروف المكتوبة بخط اليد وهي تقنية (ICR)، والتي تستخدم الشبكات العصبية العميقة ليس فقط في التعرف على الحروف بل التنبؤ بها.

وقد تناول هذا الكتاب تلك الموضوعات في فصوله الخمس التي جاءت كما يأتي:

- **الفصل الأول:** (التعلم العميق، نشأته وتطوره)، وتناول هذا الفصل مفهوم التعلم العميق ونشأته وأسباب انتشاره والتحديات التي يواجهها، وتطبيقات التعلم العميق، وطرق التعلم الآلي، مثل التعلم الخاضع للإشراف وغير الخاضع للإشراف، وأخيراً نماذج التعلم العميق.
- **الفصل الثاني:** موضوع (التعلم العميق والتعرف الذكي على النصوص العربية)، وتناول تصنيف المخطوطات باستخدام التعلم العميق، تصميم وتطبيق تجربة لتصنيف الكلمات بالمخطوطات الإسلامية، وتصميم وتطبيق تجربة للتعرف الذكي على نصوص المخطوطات الإسلامية، التعرف على المخطوطات مجهولة المؤلف باستخدام التعلم العميق، التنبؤ باستخدام خوارزميات التعلم العميق- تجربة عملية، التعرف على المخطوطات مجهولة المؤلف، وأخيراً تطبيق تجربة للتعرف على تاريخ المخطوطة.
- **الفصل الثالث:** (التعلم العميق والتعرف الذكي على الحروف المكتوبة بخط اليد في المخطوطات الإسلامية)، تضمن التنبؤ بأسلوب خط اليد في المخطوطات الإسلامية، واستكشاف أسلوب خط اليد باستخدام التعلم العميق، وتقنية ICR، والتعرف الذكي على الحروف المكتوبة بخط اليد في قراءة الوثائق العربية والإسلامية، كما تناول التطبيق العملي لتقنية ICR على نماذج من الوثائق العربية في العصر العثماني، وأخيراً تناول نمذجة التعرف على الحروف في المخطوطات المكتوبة بخط اليد Handwritten.
- **الفصل الرابع:** (استخدامات عملية الذكاء الاصطناعي في التعامل مع النصوص العربية والمخطوطات الإسلامية)، وتناول هذا الفصل أمثلة تطبيقية على منصات إلكترونية، ومن أشهرها منصة Transkribus، التي تستخدم الذكاء الاصطناعي عند التعامل مع النصوص العربية، سواء كانت مكتوبة بخط اليد أو مطبوعة.