مجلة التربية للعلوم الانسانية

مجلة علمية فصلية محكّمة، تصدر عن كلية التربية للعلوم الإنسانية / جامعة الموصل

الكشفُ الآليّ عن الصّورِ البلاغيّةِ في الشّعرِ العربيّ عبرَ تقاناتِ التُّعلّمِ العميقِ والرُّؤيةِ الحاسوبيّةِ

علاء عبد الخالق حسين [1] خالد عبود حمودي [2]

جامعة بغداد – كلّيّة العلوم الإسلاميّة [1، 2]

معلومات الارشفة

تاريخ القبول : 2025/11/27

تاريخ النشر : 2026/6/19

الكلمات المفتاحية :

الذكاء الاصطناعي، التّعلّم العميق، معالجة اللّغة الطبيعية، التّحليل البلاغي، الشّعر العربي، التّشبيه، الاستعارة، الكِناية، المنهجية الكمّية

معلومات الاتصال

علاء عبد الخالق

alaa.Abdulkhaleq@colaw.uobaghdad.edu.iq

الملخص

يسعى البحث إلى استكشاف الدور التحويلي لتقانات الذكاء الاصطناعي، ولا سيما التعلم العميق وتحليل الصور الرقمية، في الكشف التلقائي عن الصور البلاغية الأساسية (التشبيه، والاستعارة، والكناية) في الشّعر العربي القديم والحديث. تهدف الدراسة إلى التغلب على القيود التي تواجه التحليل البلاغي التقليدي، الذي يعاني من ذاتية التفسير وطول الوقت المطلوب، مما يعوق إجراء دراسات كمية شاملة لتطور الظواهر البلاغية عبر التاريخ. كما يسعى إلى ردم الفجوة بين التراث البلاغي العربي والأدوات الحاسوبية الحديثة، عبر تقديم منهجية تطبيقية تعزّز البحث الأدبي واللغوي بتحليلات قائمة على البيانات الرقمية.

اعتمدت الدراسة منهجًا كمّيًا تحليليًّا يعتمد على تقانات التعلم العميق، مع اتّباع سلسلة من الخطوات الدقيقة لضمان دقّة النتائج. بدأ العمل بجمع قاعدة بيانات واسعة من النصوص الشعرية، تغطي الشعر العربي قديمه وحديثه، مع الاعتماد على مصادر رقمية موثوقة ونصوص محقّقة. ثم أُخضعت هذه النصوص لعمليات معالجة لغوية دقيقة، شملت: تدقيق النصوص من الأغلاط، وتوحيد كتابة الهمزات والألفات، وإزالة التشكيل الزائد الذي لا يؤثر في الدلالة، وتقسيم النصوص على وحدات تحليلية (أبيات أو مقاطع دلالية)، وكانت المرحلة الحاسمة هي بناء قاعدة البيانات المُعلَّمة.

ولضمان دقة التصنيف، قام ثلاثة متخصصين في البلاغة العربية بتصنيف 20,000 بيت شعري يدويًا، مع وضع إرشادات تفصيلية لضمان الاتّساق في

التقويم. كما قيست درجة التوافق بين المصنفين عبر مقاييس إحصائية لضمان موثوقية البيانات.

وطُورت نماذج تعلّم عميق، مع التركيز على تقانات معالجة اللغة الطبيعية، ولا سيّما نموذج "AraBERT" المخصّص للغة العربية. واعتمدت النماذج على تمثيل النّصوص بمتّجهات سياقية غنية، مع دمج ميزات لغوية إضافية لتحسين الأداء. وتجدر الإشارة إلى أن تقانات تحليل الصور (رؤية الحاسوب) لم تُظهر فاعلية كبيرة في هذه المهمة، نظرًا لطبيعتها الدلالية المعقدة. كما شمل العمل توظيف تقانات متقدمة لمنع فرط التكيف، مثل التوقف المبكّر والتّسرب العشوائي.

وأظهرت النتائج أن النّموذج حقّق دقّة عالية في التّعرف على التشبيه، وأداءً جيدًا في اكتشاف الاستعارة، في حين واجه صعوبات كبرى في تحديد الكناية، نظرًا لارتباطها الوثيق بالسياق الثقافي والمعرفة الخارجية. كما كشفت الدراسة عن اختلافات واضحة في أنماط استخدام الصور البلاغية بين الشعر القديم والحديث. وأوصت الدراسة بمجموعة توصيات، منها: توسيع قاعدة البيانات البلاغية العربية لتشمل نصوصًا أكثر تنوعًا، وتحسين قدرة النماذج على فهم الكناية عبر دمج المعرفة السياقية، وتطوير نماذج أكثر شفافية لتفسير النتائج بلاغيًا، وإجراء دراسات مقارنة لتحليل تطور الأساليب البلاغية عبر العصور، وتوسيع نطاق البحث ليشمل صورًا بلاغية أخرى، مثل المجاز المرسل والطِباق

DOI: **********,, ©Authors, 2025, College of Education for Humanities University of Mosul.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

Journal of Education for Humanities

A peer-reviewed quarterly scientific journal issued by College of Education for Humanities / University of Mosul

Towards the Automatic Detection of Figurative Language in Arabic Poetry through Deep Learning and Computer Vision Techniques

Alaa Abdulkhaleq Hussein [1] **Khaled Abboud Hamoudi** [2]
University of Baghdad - College of Islamic Sciences [1,2]

Article information

Accepted : 27/11/2025
Published 19/6/2026

Keywords:
Artificial Intelligence, Deep Learning, Natural Language Processing, Rhetorical Analysis, Arabic Poetry, Simile, Metaphor, Metonymy, Quantitative Methodology

Correspondence:
Alaa Abdulkhaleq
alaa.Abdulkhaleq@colaw.uobaghdad.edu.iq

Abstract

This research seeks to explore the transformative role of artificial intelligence techniques, particularly deep learning and digital image analysis, in the automatic detection of basic rhetorical figures (simile, metaphor, metonymy) in Arabic poetry from the pre-Islamic era to the modern era. The study aims to overcome the limitations facing traditional rhetorical analysis, which suffers from subjective interpretation and the time required, hindering the conduct of comprehensive quantitative studies of the evolution of rhetorical phenomena throughout history. It also seeks to bridge the gap between Arabic rhetorical heritage and modern computational tools by presenting an applied methodology that enhances literary and linguistic research with data-driven analyses. The study adopted a quantitative analytical approach based on deep learning techniques, following a series of meticulous steps to ensure the accuracy of the results. The work began by collecting a large database of poetic texts, covering ancient poetry (from the pre-Islamic era to the Abbasid era) and modern poetry (from the Renaissance era to the present day), relying on reliable digital sources and verified texts. These texts then underwent a rigorous linguistic processing process, which included:
Removing errors and impurities from the texts.
Unifying the spelling of hamzas and alifs.
Removing redundant diacritics that do not affect the meaning.
Dividing the texts into analytical units (verses or semantic segments).

The critical stage: Building the labeled database

To ensure classification accuracy, three Arabic rhetoric experts manually classified 20,000 poetic verses, setting detailed guidelines to ensure consistency in evaluation. The degree of agreement between classifiers was also measured using statistical measures to ensure data reliability.

Advanced deep learning models were developed, focusing on natural language processing techniques, specifically the "AraBERT" model, which is tailored to the Arabic language. The models relied on representing texts with rich contextual vectors, incorporating additional linguistic features to improve performance. It is worth noting that image analysis techniques (computer vision) have not proven highly effective in this task, due to its complex semantic nature. Training also included the use of advanced techniques to prevent overfitting, such as early stopping and random dropout.

The results showed that the model achieved high accuracy in recognizing similes and performed well in detecting metaphors, while it encountered greater difficulties in identifying metonymy, given its close association with cultural context and external knowledge. The study also revealed clear differences in the patterns of rhetorical imagery usage between ancient and modern poetry. The study recommends the following:

Expanding the Arabic rhetorical database to include more diverse texts.

Improving the models' ability to understand metonymy by incorporating contextual knowledge.

Developing more transparent models for interpreting results rhetorically.

Conducting comparative studies to analyze the evolution of rhetorical techniques across the ages.

Expanding the scope of research to include other rhetorical images, such as metonymy and antithesis

DOI: **********,, ©Authors, 2025, College of Education for Humanities University of Mosul.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

الفصل الأول

المقدّمة والإطار النّظري

يُعدّ الشّعر العربي أحد أبرز مكونات التراث الأدبي واللغوي، إذ حافظ على مكانته عبر العصور، ليس بوصفه وسيلة تعبيرية فحسب، بل بوصفه مرآة تعكس ثقافة الأمّة وقيمها الجمالية والفكرية. ومن بين العناصر التي تُضفي على النّصّ الشّعري جماليته الخاصّة، تأتي الصور البلاغية لتحتل مكانة محورية، إذ تُمثّل نسيجًا دلاليًا معقّدًا يمنح النّصّ عمقًا فنيًا وقدرة على إثارة المشاعر والتّأمل. وتُعدّ دراسة هذه الصّور وتحليلها أمرًا مهمًا ليس لفهم الجماليات الأدبية فحسب، بل لاستكشاف الطبقات الدلالية والرّمزية التي تُخفيها النّصوص الشّعرية، والتي تعكس –بدورها– رؤى الشّعراء وتفاعلاتهم مع محيطهم الثّقافي والاجتماعي.

ولطالما اعتمدت الدّراسات البلاغية التّقليدية على المنهج التّحليلي الشخصي، الذي يتطلب خبرة لغوية وأدبية عميقة، فضلاً عن ذائقة فنية عالية تُمكّن الباحث من استنباط المعاني الضّمنية والعلاقات الدّلالية بين الكلمات. ومع ذلك، تواجه هذه الطريقة عدّة تحدّيات منهجية وعملية، أبرزها طابعها الذّاتي الذي قد يؤدّي إلى اختلافات في التّفسير بين الباحثين، وصعوبة تطبيقها على نطاق واسع نظرًا للوقت والجهد الكبيرين اللذين تتطلبهما عملية التحليل الشامل لمجموعات نصّيّة كبيرة. كما أنّ الاعتماد على الخبرة المتخصصة قد يُمثّل معوقًا أمام إجراء دراسات كمّية واسعة تهدف إلى تتبع تطوّر الصّور البلاغية عبر العصور أو مقارنة أنماط استخدامها بين المدارس الشعرية المختلفة. إذ أشار ابن الأثير (2018) إلى هذه الإشكالية حين ناقش محدودية الأدوات التقليدية في تحليل النصوص الأدبية، مؤكدًا الحاجة إلى منهجيات أكثر دقة وموضوعية.

وفي ظل التّطورات المتسارعة التي يشهدها مجال الذكاء الاصطناعي، ولا سيّما في فرعي معالجة اللغات الطبيعية ورؤية الحاسوب، تبرز فرصة حقيقية لتجاوز القيود التي تفرضها المنهجيات التقليدية. إذ تُتيح التّقانات الحديثة إمكان معالجة كميات هائلة من النّصوص بدقة وكفاءة غير مسبوقتين، ممّا يفتح آفاقًا جديدة لتحليل الصور البلاغية بنحو موضوعي وكمي. ولم يعد الذكاء الاصطناعي مجرّد أداة نظرية، بل تحوّل إلى واقع عملي يُحدث ثورة في شتى التخصصات، بما في ذلك العلوم الإنسانية والدراسات الأدبية. فعن طريق نماذج التعلم الآلي والتعلم العميق، صار بالإمكان تحليل الأنماط اللغوية المعقّدة، وفهم العلاقات الدلالية بين الكلمات، وتحديد السياقات التي تُستخدم فيها الصور البلاغية بكفاءة عالية.

وتُعزّز تقانات رؤية الحاسوب هذه الإمكانات عبر تحويل النصوص إلى تمثيلات بصرية، مما يُتيح تحليل الجوانب المرئية للغة الشعرية، مثل التشكيل الطباعي وتوزيع الكلمات، التي قد تُسهم في الكشف عن أنماط بلاغية خفية. إذ أشار سزيلسكي (2010) إلى أن معالجة النصوص كصور يمكن أن يُضيف بُعدًا

جديدًا لفهم البنية اللغوية، ولا سيما في النصوص التي تعتمد على الترتيب البصري للكلمات، كما هي الحال في بعض القصائد الحديثة.

وانطلاقًا من هذه الرؤية، يهدف هذا البحث إلى تطوير منهجية جديدة تعتمد على تقانات الذكاء الاصطناعي والرؤية الحاسوبية للكشف عن الصور البلاغية الرئيسة في الشعر العربي، وهي: التشبيه، والاستعارة، والكناية. كما يسعى إلى إجراء مقارنة تحليلية بين استخدام هذه الصور في الشعر القديم والحديث، سعيًا لفهم التحولات الأسلوبية والدلالية التي شهدتها البلاغة عبر الزمن. ويُتوقع أن يُسهم هذا المنهج في تقديم رؤى جديدة بشأن تطور الشعر العربي، كما أنه يُمهّد الطريق لبناء أدوات حاسوبية متقدمة تُساعد الباحثين على تحليل النصوص الأدبية بفعالية كبرى.

وتتمثل الأهداف الرئيسة لهذا البحث في:

1. تطوير نماذج ذكاء اصطناعي قادرة على التعرف التلقائي على الصور البلاغية في النصوص الشعرية العربية.

2. إنشاء قاعدة بيانات شاملة للصور البلاغية في عيّنة ممثلة من الشعر العربي القديم والحديث.

3. تحليل الفروقات والتشابهات في استخدام الصور البلاغية بناءً على النتائج المستخلصة من النماذج الحاسوبية.

4. الإسهام في تطوير منهجية علمية جديدة لدراسة البلاغة العربية باستخدام التقانات الحديثة.

ويُمثّل تحقيق هذه الأهداف معًا إطارًا بحثيًا متكاملًا لا يقتصر على توظيف التقانات الحاسوبية فحسب، بل يسعى أيضًا إلى إثراء الفهم الأكاديمي للجماليات الشعرية وطرائق تحليلها.

الصور البلاغية

تُعدّ الصور البلاغية من أهم الأدوات التي يستخدمها الشاعر لتحقيق الإبداع الفني في نصّه، فهي تمنح اللغة قوة تعبيرية تتجاوز المعنى المباشر، وتُسهم في إثراء النّصّ بمستويات متعدّدة من الدلالة. إذ تعتمد هذه الصور على العلاقات المركبة بين الكلمات التي تُنتج معاني مجازية أو رمزية تُضفي على النّصّ جماليته الخاصة. ولكي يكون التحليل الحاسوبي لهذه الصور دقيقًا، لا بد من فهم طبيعتها النظرية وتصنيفاتها المختلفة، إذ إن كل نوع منها يخضع لقواعد وأسس تُميزه من غيره.

التشبيه

يُعدّ التشبيه من أكثر الصور البلاغية وضوحًا وانتشارًا في الشّعر العربي، ويعرّفه الزّمخشري (1998، ص.345) بأنّه "عقد مشاركة بين شيئين في صفة أو أكثر، بأداة تشبيه ملفوظة أو مقدّرة". ويهدف التشبيه إلى توضيح المعاني المجردة أو غير المألوفة عبر ربطها بأشياء معروفة أو محسوسة، مما يُسهّل على المتلقي فهمها واستيعابها. ويتكوّن التشبيه من أربعة أركان أساسية، هي: المشبّه، والمشبّه به، وأداة التشبيه، ووجه الشبه.

وتتنوّع أنواع التشبيه وفقًا للأركان الموجودة أو المحذوفة، مما يُنتج صورًا بلاغية متعدّدة المستويات. فالتشبيه التام، الذي يُذكر فيه جميع الأركان، إذ يُقدّم صورة واضحة ومباشرة، مثل قول المتنبي (1983، ص.120): "**كأنّما الشمس دينارٌ يُضيء في سماءٍ زرقاء**". في حين يُقدّم التشبيه المؤكّد، الذي تُحذف فيه أداة التشبيه، صورة أكثر قوّة، كما في قول أبي تمام (1997، ص.50): "**أنتَ بحرٌ في الكرم**". أما التشبيه البليغ، فهو الذي يُحذف فيه كل من الأداة ووجه الشبه، فيُنتج دلالة قوية على اتّحاد المشبّه والمشبّه به، مثل قولنا: **العلم نور**.

وتُظهر الدراسات البلاغية التقليدية، كما أشار ضيف (1968)، أن التشبيه كان يُستخدم بكثافة في الشعر القديم لتوضيح المعاني الحسية، بينما يميل الشعر الحديث إلى استخدام تشبيهات أكثر تعقيدًا أو غير مباشرة، مما يعكس تحولًا في الأساليب الفنية.

الاستعارة

تُعدّ الاستعارة صورة بلاغية أكثر تركيزًا من التشبيه، إذ يُحذف أحد طرفيه (المشبّه أو المشبّه به)، ممّا يُضفي قوّة تعبيرية كبرى. وتُقسّم الاستعارة على نوعين رئيسين، هما: التصريحية، التي يُحذف فيها المشبّه ويُصرّح بالمشبّه به، مثل: **رأيت أسدًا يُحاضر**، والمكنية، التي يُحذف فيها المشبّه به ويُرمز إليه بصفة من صفاته، مثل: **الدهرُ غادِرٌ**.

وتُسهم الاستعارة في تكثيف المعنى وتقديمه بطريقة موجزة ومؤثرة، وغالبًا ما تُستخدم في الشعر القديم لتضخيم الصّفات، في حين يُوظفها الشعر الحديث لخلق صور أكثر تجريدًا ورمزية.

الكِناية

تعتمد الكِناية على التعبير غير المباشر، إذ يُذكر لازم المعنى ويُقصد به المعنى الأصلي. وتُقسّم الكناية على ثلاثة أنواع: الكناية عن صفة، مثل: **كثير الرماد**، للدلالة على الكرم (البغدادي، 2000، ص.115)،

والكناية عن موصوف، مثل: **حامل لواء الشعر**، للإشارة إلى المتنبي (ابن منظور، 1994، ص.100)، والكناية عن نسبة، مثل: "**الجود بين برديه**" (المرادي، 1999، ص.80).

وتُضفي الكناية على النّصّ طابعًا من الغموض الجميل، وتُثير فضول المتلقي لتأويل المعنى. وفي حين كانت الكنايات في الشعر القديم تعتمد على رموز متّفق عليها، فإن الشّعر الحديث يوظفها بطرائق أكثر إبداعًا وتجريدًا.

الدراسات السابقة

على الرغم من التقدم الكبير في مجال معالجة اللغات الطبيعية، فإن الدراسات التي تركّز على تحليل الصور البلاغية في الشّعر العربي ما تزال محدودة. إذ درست بعض الأبحاث تحليل الأوزان والقوافي (Al-Tamimi & Al-Khalifa, 2018)، أو مشاعر القصائد (El-Haj & Al-Khalifa, 2015)، لكنّها لم تُعالج تعقيدات الصور البلاغية بنحوٍ وافٍ.

أما في مجال الذكاء الاصطناعي، فقد حقّقت النماذج مثل BERT (Devlin et al., 2019) تقدمًا كبيرًا في فهم النّصوص، لكنّ تطبيقها على الشّعر العربي يحتاج إلى تكييف خاص. كما أنّ الدراسات التي حاولت تحليل الاستعارة أو التشبيه في اللّغات الأخرى (Shutova, 2010) لا تُغطي خصوصيات اللّغة العربيّة.

من هنا، يسعى هذا البحث إلى سدّ هذه الفجوة عبر تطوير منهجية شاملة تعتمد على أحدث تقانات الذكاء الاصطناعي لتحليل الصور البلاغية في الشعر العربي، وهو ما يُعدّ إضافة نوعية في مجال الأدب والدراسات الحاسوبية معًا.

الفصل الثاني

الذكاء الاصطناعي ورؤية الحاسوب وتطبيقاتها اللغوية

يشهد هذا العصر تطورًا متسارعًا في مجال الذكاء الاصطناعي (AI)، الذي لم يعد يقتصر على أنّه فرع من فروع علم الحاسوب فحسب، بل صار يمثل إطارًا شاملًا يضم مجموعة متكاملة من التقانات والمنهجيات التي تهدف إلى محاكاة العمليات المعرفية البشرية، بما في ذلك التعلم والاستدلال وحل المشكلات وفهم اللغة الطبيعية (Russell & Norvig, 2020). وأدّى هذا التطور إلى ظهور تطبيقات متقدّمة في مجال تحليل النّصوص الأدبية، ولا سيّما الشعر، إذ يوفر الذكاء الاصطناعي أدوات تحليلية متطورة تمكّن الباحثين من تجاوز القيود التقليدية للتحليل اليدوي، ممّا يفتح آفاقًا جديدة لاكتشاف الأنماط اللغوية والبلاغية بدقّة غير مسبوقة.

وفي هذا السياق، يبرز دور معالجة اللغات الطبيعية (NLP) كأحد الفروع الأساسية للذكاء الاصطناعي، إذ يركّز على تمكين الحواسيب من فهم اللغة البشرية وتفسيرها وتوليدها بطريقة تحاكي الفهم الإنساني (Jurafsky & Martin, 2009). وتكمن أهمية هذا المجال في تحليل الشعر العربي، إذ لا يقتصر الأمر على التّعرف على الكلمات أو القواعد النّحوية فحسب، بل يتعدّاه إلى فهم السياقات الدلالية العميقة التي تتجاوز المعنى الحرفي للنّصّ. إذ تعتمد البلاغة الشعرية بنحوٍ كبير على الانزياحات اللغوية والمجازات والإيحاءات التي تمثل تحديًا كبيرًا للتحليل الآلي. ومع ذلك، فإن التطورات الحديثة في نماذج التعلم العميق، ولا سيّما تلك المستوحاة من بنية الشّبكات العصبية البشريّة؛ قد أظهرت قدرة ملحوظة على معالجة هذه التّعقيدات.

كما تؤدّي رؤية الحاسوب (Computer Vision) دورًا مكمّلاً في هذا المجال، إذ توفر إمكانية معالجة النّصوص كبيانات بصرية، ممّا يسهم في تحليل الخصائص المورفولوجية والتّركيبية للنّصوص الشّعرية. ويهدف هذا الفصل إلى استعراض المفاهيم الأساسية للذكاء الاصطناعي والرؤية الحاسوبية، مع التركيز على التّقانات ذات الصّلة بتحليل الشّعر العربي، وكيفية توظيفها للكشف عن الصور البلاغية المعقّدة، مثل التّشبيه، والاستعارة، والكِناية.

الذكاء الاصطناعي ومعالجة اللّغات الطبيعية

تُعدّ معالجة اللغات الطبيعية حجر الزاوية في أي محاولة لتحليل النّصوص البشرية آليًّا، إذ تهدف إلى تمكين الحواسيب من فهم اللغة الطبيعية ومعالجتها بطريقة ذكية. وتنطوي هذه العملية على سلسلة من المراحل المتتابعة التي تمثل ما يعرف بـ"خطوات معالجة اللغات الطبيعية"، التي تتراوح بين التحليل الصرفي الأساسي للكلمات وصولاً إلى استخلاص المعاني الدلالية والسياقية الأكثر تعقيدًا (Manning & Schütze, 1999).

تبدأ عملية التحليل بالمرحلة الصّرفية (Morphological Analysis)، التي تركّز على تفكيك الكلمات إلى مكوناتها الأساسية، مثل الجذور واللواحق والسّوابق. وتكتسب هذه المرحلة أهميّة خاصّة في اللغة العربية نظرًا لطبيعتها الاشتقاقية والتّصريفية الغنية. فعلى سبيل المثال، يمكن تحليل كلمة "ويكتبون" إلى الجذر "كتب" مع إضافة السوابق واللواحق الدالة على الزمن والفاعل والعدد. ويسهم هذا التحليل في تحديد الصيغ الصرفية التي قد تكون ذات دلالة بلاغية، مثل صيغ المبالغة أو أسماء الفاعلين والمفعولين (Al-Khalifa & Al-Qurashi, 2010).

تلي ذلك مرحلة التحليل النحوي (Syntactic Analysis)، التي تُعنى بدراسة بنية الجملة والعلاقات بين مكوناتها. وتُستخدم في هذه المرحلة تقانات متقدّمة مثل التحليل الإعرابي (Parsing) لتحديد الأدوار

النحوية للكلمات، مثل الفاعل والمفعول به، وفهم التراكيب النّحوية المعقّدة. وفي سياق الشعر العربي، يمكن أن يساعد هذا التحليل في الكشف عن الانزياحات النّحوية، مثل التقديم والتأخير أو الحذف، التي غالباً ما تحمل دلالات بلاغية عميقة (Khoja & Garside, 1999).

أمّا المرحلة الدّلالية (Semantic Analysis) فتركّز على فهم المعنى الحقيقي للكلمات والجمل، متجاوزةً المعنى الحرفي إلى استيعاب العلاقات الدّلالية بين المفردات، مثل التّرادف، والتّضاد، والاشتراك اللّفظي. وتُعدّ هذه المرحلة حاسمة في تحليل الصّور البلاغية، إذ تعتمد على التّمييز بين الاستخدام الحرفي والمجازي للكلمات، وهو ما يمثّل أساسًا لفهم التّشبيهات والاستعارات والكنايات (Jurafsky & Martin, 2009).

وتأتي المرحلة البراجماتية (Pragmatic Analysis) كخاتمة لهذه العملية، إذ تركّز على فهم المعنى في سياقه الواسع، بما في ذلك نيّة المتكلم والظروف المحيطة بالنّصّ. وفي الشّعر، تساعد هذه المرحلة في تفسير الأغراض البلاغية الكامنة وراء الصور الشعرية، مثل المدح أو الذّم أو التّهكم، ممّا يضيف بعدًا آخر لفهم النّصّ.

التعلم الآلي: الأسس والتطبيقات

يمكن تصنيف تقانات التعلم الآلي إلى عدة أنواع رئيسة، لكل منها تطبيقاته في مجال معالجة اللغة:

التّعلم المراقب (Supervised Learning): هو النوع الأكثر شيوعًا في معالجة النصوص، إذ يعتمد على وجود بيانات مدروسة ومصنفة مسبقًا. ويمكن استخدامه في تصنيف الصور البلاغية، إذ يتم تدريب النموذج على مجموعة من الأبيات الشعرية مع تحديد نوع الصورة البلاغية الموجودة فيها (Bishop, 2006).

التّعلم غير المراقب (Unsupervised Learning): يستخدم عندما لا تتوفر بيانات مصنّفة، إذ يعتمد على اكتشاف الأنماط الخفية في البيانات، مثل تجميع النّصوص المتشابهة في الأسلوب أو الموضوع (Hastie et al., 2009).

التّعلم شبه المراقب (Semi-supervised Learning): يجمع بين كمية صغيرة من البيانات المصنّفة وكمية كبيرة من البيانات غير المصنّفة، ممّا يجعله مفيدًا في حالات ندرة البيانات المدروسة.

التّعلم المعزّز (Reinforcement Learning): يركّز على تحسين الأداء عبر التّغذية الرّاجعة، ويمكن أن تكون له تطبيقات في توليد النّصوص الشّعرية (Sutton & Barto, 2018).

التّعلم العميق: الشبكات العصبية والتطبيقات اللّغوية

يُعدّ التّعلم العميق امتدادًا للتعلم الآلي، إذ يعتمد على الشّبكات العصبية العميقة (Deep Neural Networks) لتحليل البيانات المعقّدة مثل النّصوص والصّور (Lecun et al., 2015). ومن أهم النماذج المستخدمة في هذا المجال:

الشبكات العصبية المتكرّرة (RNNs): صمّمت لمعالجة البيانات المتسلسلة مثل النصوص، لكنّها تعاني من مشكلة "تلاشي التدرج" (Bengio et al., 1994).

شبكات الذّاكرة طويلة المدى (LSTMs): وهي تطوير للشّبكات المتكرّرة، إذ تتغلّب على مشكلة التّبعيات طويلة المدى (Hochreiter & Schmidhuber, 1997).

نماذج المحوّلات (Transformers): مثل BERT وGPT، التي تعتمد على آلية الانتباه (Attention) لفهم السياق الدّلالي للنّصوص (Vaswani et al., 2017). وقد أظهرت نماذج مثل AraBERT كفاءة عالية في معالجة اللّغة العربيّة (Attia et al., 2019).

الرّؤية الحاسوبيّة وتطبيقاتها في تحليل النّصوص

على الرّغم من أنّ رؤية الحاسوب ترتبط تقليديًّا بتحليل الصّور، إلا أن تطبيقاتها في معالجة النّصوص المكتوبة آخذة في التّوسع، ولا سيّما في مجال تحليل المخطوطات القديمة (Szeliski, 2010). ومن أبرز هذه التّطبيقات:

التّعرف البصري على الحروف (OCR): الذي يمكن من تحويل النّصوص المطبوعة أو المخطوطة إلى نصوص رقمية قابلة للمعالجة (Smith, 2007).

تحليل الخطوط: لفهم الخصائص البصرية للنصوص القديمة.

دمج الرؤية الحاسوبية مع معالجة اللغة: لتحسين فهم النّصوص، ولا سيّما في حالة المخطوطات ذات الجودة المنخفضة.

تطبيقات الذكاء الاصطناعي في تحليل الشعر

شهدت السّنوات الأخيرة تطورًا ملحوظًا في تطبيقات الذّكاء الاصطناعي في تحليل الشّعر، منها:

تحليل المشاعر: لتصنيف القصائد بحسب العواطف (El-Haj & Al-Khalifa, 2015).

توليد الشِّعر: باستخدام نماذج مثل GPT-3 (Brown et al., 2020).

تحليل الوزن والقافية: باستخدام تقانات معالجة اللغة (Al-Tamimi & Al-Khalifa, 2018).

الكشف عن الصّور البلاغية: مثل التّشبيه، والاستعارة، إذ تظهر الدراسات إمكانية تطبيق هذه التقانات على اللغة العربية (Shutova, 2010).

وخلاصة القول: إنّ التطورات الحديثة في الذكاء الاصطناعي ومعالجة اللغات الطبيعية تقدم أدوات واعدة لتحليل الشعر العربي، ولا سيما في الكشف عن الصّور البلاغية المعقدة، ممّا يفتح آفاقًا جديدة للبحث في هذا المجال.

الفصل الثالث

جمع البيانات وبناء قاعدة المعرفة

يُعدّ جمع البيانات عالية الجودة وبناء قاعدة معرفية مُعلّمة بدقة من المراحل الأساسية التي تُحدّد نجاح أي مشروع بحثي يعتمد على تقانات الذكاء الاصطناعي والتعلم العميق، ولا سيّما في مجال معالجة اللغات الطبيعية. فجودة النّموذج النهائي ومدى كفاءته في أداء المهام المطلوبة ترتبط ارتباطًا وثيقًا بجودة البيانات المستخدمة في تدريبه وكميتها (Goodfellow et al., 2016). وفي سياق هذا البحث، الذي يهدف إلى تطوير نموذج قادر على الكشف عن الصور البلاغية في الشعر العربي القديم والحديث باستخدام تقانات الذّكاء الاصطناعي ورؤية الحاسوب؛ تكتسب عملية جمع النصوص الشعرية وتدقيقها وتصنيفها أهمية استثنائية؛ إذ إنها تُمثل الأساس الذي تُبنى عليه جميع المراحل اللاحقة.

إن التّحدي الرّئيس في هذه العملية يكمن في الطبيعة الغنية والمعقّدة للشّعر العربي، الذي يتميز بتنوّعه التاريخي والجغرافي، وتعقيد بنيته اللّغوية والبلاغيّة، ممّا يتطلّب عناية فائقة في إعداد البيانات ومعالجتها. كما أنّ هناك عقباتٍ أخرى تتعلّق بندرة النّصوص الشّعرية الرّقمية المُشكّلة والمُعلّمة، ولا سيّما تلك التي تعود إلى الحِقب القديمة، ممّا يُضيف صعوبة أخرى إلى عملية جمع البيانات.

يركّز هذا الفصل على تفصيل المنهجية المتّبعة في جمع النّصوص الشّعرية من مصادرها الموثوقة، بدءًا من تحديد هذه المصادر وصولًا إلى عمليات التّدقيق والمعالجة الأولية التي تُهيّئ البيانات للتّحليل الآلي. كما سيقوم الفصل بشرح عملية بناء قاعدة المعرفة للصّور البلاغية، التي تشمل التصنيف اليدوي الدقيق للنصوص بوساطة خبراء في اللغة العربية والبلاغة، وهي خطوة حاسمة لإعداد البيانات المطلوبة لتدريب نماذج التعلم الآلي. أخيرًا، سيناقش الفصل آليات تمثيل هذه البيانات اللغوية في صيغ عددية تُمكّن النماذج الحاسوبية من معالجتها بكفاءة، فضلاً عن استراتيجيات تقسيم البيانات لضمان تقويم موضوعي لأداء النماذج.

جمع البيانات الشعرية

تتطلّب مهمّة الكشف عن الصور البلاغيّة في الشّعر العربي القديم والحديث؛ جمع عينة تمثيلية وشاملة من النّصوص الشّعرية العربية. إذ يجب أن تكون هذه العيّنة متنوّعة بما يكفي لتعكس الأساليب الشّعرية المختلفة، والمدارس الأدبيّة المتنوّعة، والتّعبيرات البلاغية التي سادت في كلّ عصر. كما يجب أن تشمل النّصوص التي تمثّل مراحل تطوّر الشّعر العربي، بدءًا من العصر الجاهلي مرورًا بالعصور الإسلامية والأُموية والعباسيّة، وصولًا إلى العصر الحديث بتياراته المختلفة مثل مدرسة الإحياء والديوان وأبولو والشعر الحر وقصيدة النثر.

تحديد مصادر الشعر العربي القديم

يُقصد بالشعر العربي القديم ذلك الشعر الذي نُظم في تلك الحقبة الممتدّة من العصر الجاهلي وحتى نهاية العصر العباسي، وهي حقبة زمنية شهدت تطورًا كبيرًا في الأساليب الشّعرية والبلاغيّة. تشمل تلك الحقبة عصورًا متعدّدة مثل العصر الجاهلي، والعصر الإسلامي المبكّر (عصر صدر الإسلام)، والعصر الأموي، والعصر العباسي الذي يُعدّ ذروة الازدهار الأدبي والبلاغي في التاريخ العربي.

تُمثّل الشعر في تلك الحقبة مخزونًا هائلًا من النّصوص الشعرية التي تُمثل أساس البلاغة العربية، ممّا يجعل جمعها وتحليلها أمرًا بالغ الأهمية لأي دراسة تهدف إلى فهم الصور البلاغية وتطورها. إلا أن عملية جمع النصوص من هذه الحقبة التاريخية تواجه عدّة تحدّيات، أبرزها صعوبة توثيق النّصوص القديمة وندرة المصادر الرّقمية المُشكّلة، فضلًا عن الاختلافات في الرّوايات والنّسخ المخطوطة التي قد تؤثر في دقّة النّصوص.

أبرز المصادر المعتمدة للشعر العربي القديم

المعلّقات الشعرية: تُعدّ المعلّقات من أقدم النماذج الشعرية في العصر الجاهلي وأهمها، إذ تمثل قصائد مُختارة من أفضل ما قيل في ذلك العصر. وتتميز المعلقات ببلاغتها العالية وغناها بالصور الفنية، مما يجعلها نموذجًا مثاليًا لدراسة الصور البلاغية في الشعر الجاهلي (الزوزني، 1999).

دواوين شعراء العصر الجاهلي وصدر الإسلام: تشمل هذه الدواوين أعمال شعراء بارزين مثل عنترة بن شداد، وطرفة بن العبد، وزهير بن أبي سلمى في العصر الجاهلي، وحسّان بن ثابت، وكعب بن زهير في عصر صدر الإسلام (ديوان حسان بن ثابت، 2004). وتُقدّم هذه الدواوين مادة غنية لدراسة التطور البلاغي في تلك الحقب، إذ تظهر فيها بوضوح المؤثرات الدّينية والاجتماعية في الصّور الشّعرية.

دواوين شعراء العصر الأُموي: مثل ديوان جرير، والفرزدق، والأخطل، الذين اشتُهروا بالشّعر السّياسي والهجاء، ممّا أضاف أبعادًا جديدة للبلاغة العربية (ديوان جرير، 1999).

دواوين شعراء العصر العباسي: يُعدّ العصر العباسي العصر الذهبي للشعر العربي، إذ برز شعراء كبار مثل أبي نواس، وأبي تمام، والمتنبي، والبحتري، وأبي العلاء المعري (ديوان المتنبي، 1983؛ ديوان أبي العلاء المعري، 2002). وتمتاز قصائد هؤلاء الشّعراء بتعقيدها البلاغي وغناها بالصور الفنية، ممّا يجعلها ضرورية لأيّ دراسة في البلاغة الشعرية.

كتب الأدب والبلاغة القديمة: مثل "البيان والتبيين" للجاحظ (2002)، و"دلائل الإعجاز" لعبد القاهر الجرجاني (1987)، و"العمدة في محاسن الشعر وآدابه" لابن رشيق القيرواني (1981). إذ تحتوي هذه الكتب على شروح بلاغية مستفيضة للشعر العربي، فضلاً عن شواهد شعرية مُصنّفة، ممّا يجعلها مصدرًا قيّمًا للباحثين.

تحديد مصادر الشّعر العربيّ الحديث

يُقصد بالشعر العربي الحديث ذلك الشعر الذي نُظم في تلك الحقبة التي تبدأ بعصر النّهضة في القرن التاسع عشر، وتستمر حتى هذا العصر. إذ شهدت هذه الحقبة تحوّلات كبيرة في الشّكل والمضمون الشّعري، إذ ظهرت مدارس جديدة مثل مدرسة الإحياء والبعث، ومدرسة الدّيوان، ومدرسة أبولو، ثم تطوّرت إلى الشّعر الحر وقصيدة النّثر. ويتميز الشعر الحديث بتنوّعه الأسلوبي واختلافه عن الشّعر القديم في استخدام الصّور البلاغية، ممّا يستدعي جمع عينات تمثيلية من شتّى التّيارات (غنيمي، 2005).

أبرز المصادر المعتمدة للشّعر العربي الحديث

دواوين شعراء مدرسة الإحياء: مثل محمود سامي البارودي، وأحمد شوقي، وحافظ إبراهيم، الذين حاولوا إحياء الأسلوب الكلاسيكي مع إضافة لمسات عصرية (ديوان أحمد شوقي، 1998).

دواوين شعراء المدارس التجديدية: مثل عبّاس محمود العقّاد، وإبراهيم ناجي، وعلي محمود طه، الذين قدّموا رؤى جديدة في الشّعر (ديوان العقاد، 2007).

دواوين شعراء الشّعر الحرّ وقصيدة النّثر: مثل بدر شاكر السياب، ونازك الملائكة، ومحمود درويش، وأدونيس، الذين كسروا الأوزان التقليدية وابتكروا أشكالًا جديدة (ديوان السياب، 1971؛ ديوان محمود درويش، 2009).

المجلّات الأدبيّة والمواقع الإلكترونية الموثوقة: تُعدّ هذه المصادر ضرورية للحصول على نماذج من الشّعر المعاصر، إذ تُنشر قصائد جديدة تعكس الاتجاهات الحديثة في الصّور البلاغية.

معايير اختيار النّصوص وتحدّيات جمع البيانات

يجب أن تخضع عملية اختيار النّصوص الشعرية لمعايير دقيقة لضمان جودة البيانات وتمثيلها، ومن أهم هذه المعايير:

التمثيل التاريخ والأدبي: يجب أن تغطي النصوص المختارة حقبًا زمنيّة متنوّعة ومدارس شعرية مختلفة لضمان شمولية التّحليل.

الدّقة والأصالة: يجب الاعتماد على نصوص محقّقة ومراجعة بعناية لتجنب الأخطاء النّصّية التي قد تؤثر في التّحليل البلاغي.

التّنوع الجغرافي والثّقافي: يُفضّل أن تشمل العيّنة شعراء من شتّى المناطق العربية لضمان تمثيل أوسع للغة الشّعرية.

التّشكيل (التنقيط): يُعد التّشكيل تحدّيًا كبيرًا في النّصوص القديمة، إذ إن كثيرًا منها غير مشكّل بالكامل، ممّا قد يؤدي إلى التباس في المعنى. يُفضّل استخدام نصوص مشكّلة أو الاعتماد على أدوات التشكيل الآلي مع المراجعة اليدوية (Habash et al., 2009).

تنوع الخطوط والتنسيقات: في حال جمع النصوص من مصادر مصورة، يجب استخدام تقانات التّعرف الضّوئي على الحروف (OCR) بصورة دقيقة لتحويلها إلى نصوص رقمية (Smith, 2007).

تدقيق البيانات الأولية ومعالجتها

بعد جمع النصوص الشعرية، تأتي مرحلة تدقيق البيانات ومعالجتها، وهي خطوة حاسمة لإعدادها للتحليل الآلي. إذ تهدف هذه المرحلة إلى إزالة الأغلاط وتوحيد تنسيق النصوص لضمان دقة المعالجة اللاحقة.

إزالة الأغلاط الإملائية

تُعدّ هذه الخطوة أساسية لضمان نقاء البيانات، إذ تتم مراجعة النّصوص لإزالة الأغلاط الإملائية والطباعية، وكذلك علامات الترقيم الزائدة أو غير الموحّدة، والرّموز والعلامات غير النّصّيّة التي قد تظهر في أثناء عملية التّحويل الرّقمي.

التّطبيع اللّغوي (Text Normalization)

تهدف هذه العملية إلى توحيد أشكال الكلمات لتكون قابلة للمعالجة الآلية، وتشمل:

1. **توحيد الهمزات والألفات:** مثل تحويل "أ"، "إ"، "آ" إلى "ا" في الحالات التي لا تؤثر في المعنى.

2.توحيد التاء المربوطة والمفتوحة: في السياقات التي لا تؤثر في الدلالة النحوية.

3.إزالة التشكيل غير الضروري: إلا إذا كان التشكيل ضروريًا لفهم المعنى البلاغي.

4.توحيد الفواصل والعلامات النّصّيّة: لضمان تناسق التّنسيق.

5.تقسيم النّصوص على وحدات تحليلية: في الشعر، تُعدّ القصيدة أو البيت الشعري وحدة التحليل الأساسية. لذلك، يتم تقسيم النصوص على أبيات، مع مراعاة الحفاظ على السياق عند الحاجة.

معالجة النّصوص المشكّلة: إذا كانت النّصوص غير مشكّلة بالكامل، يمكن استخدام أدوات التّشكيل الآلي مثل "Mishkal" (Habash et al., 2009)، مع ضرورة المراجعة اليدوية لضمان الدقة، ولا سيّما في المواضع التي يؤثر فيها التشكيل في المعنى البلاغي.

بناء قاعدة المعرفة للصّور البلاغية

تُعدّ هذه المرحلة الأكثر أهمية، إذ يتم فيها تصنيف النّصوص يدويًّا وفقًا لأنواع الصور البلاغية من (تشبيه، واستعارة، وكِناية).

التّصنيف اليدوي (Annotation)

ويتضمّن:

اختيار العيّنة التعليمية: يتم اختيار عينة تمثيلية من النصوص المُجمعة لتصنيفها.

تطوير إرشادات التصنيف: يتم وضع دليل واضح يُحدد أنواع الصور البلاغية وأمثلتها لضمان اتّساق التّصنيف بين المُعلّمين.

قياس الاتّفاق بين المصنّفين: باستخدام مقاييس مثل كابا كوهين (Cohen, 1960) لضمان موثوقية البيانات.

تمثيل البيانات وتقسيمها

ويتضمّن:

تمثيل الكلمات بمتجهات (Word Embeddings): مثل تقانات Word2Vec و AraBERT (Mikolov et al., 2013؛ Attia et al., 2019).

تقسيم البيانات: على مجموعات تدريب (70–80%)، تحقق (10–15%)، واختبار (10–15%) لضمان تقويم موضوعي (Hastie et al., 2009).

يمثّل هذا الجهد أساسًا متينًا لبناء نماذج ذكاء اصطناعي قادرة على تحليل الصّور البلاغية في الشّعر العربي بدقّة وموثوقيّة.

الفصل الرابع

تطوير النّماذج والمنهجيات

يمثّل هذا الفصل محورًا رئيسًا في الدراسة، إذ ينتقل البحث من مرحلة جمع البيانات وإعدادها إلى مرحلة تصميم النماذج الحاسوبية القادرة على اكتشاف الصّور البلاغية في الشّعر العربي. وتعتمد هذه المرحلة على الأسس النظرية والعملية التي تم وضعها في الفصول السابقة، إذ يتم تحويل البيانات النصية المُصنفة إلى نماذج ذكاء اصطناعي قادرة على التعرف على التشبيه، والاستعارة، والكِناية في النّصوص الشّعرية القديمة والحديثة (Goodfellow et al., 2016).

وتتميّز هذه المرحلة بتحدّيات فريدة، أبرزُها قدرة النّماذج على فهم الفروق الدّقيقة بين الاستخدام الحرفي والمجازي للغة، واستيعاب السّياقات الثّقافية واللّغوية المعقّدة التي تمثّل جوهر البلاغة العربية. كما يتطلّب الأمر تكييف خوارزميات التّعلّم العميق مع الخصائص المميزة للشّعر العربي، من حيث الوزن الشعري، والقافية، والبنية الدلالية، ممّا يجعل عملية تصميم النماذج أكثر تعقيدًا مقارنة بتحليل النثر الاعتيادي.

اختيار معماريات النماذج وتصميمها

نماذج التعلم المراقب (Supervised Learning Models)

نظرًا لأنّ المهمة الرئيسة تتمثّل في تصنيف الأبيات الشعرية إلى فئات بلاغية محددة (تشبيه، واستعارة، وكناية، أو لا شيء)، فإن نماذج التعلم المراقب تُعد الخيار الأمثل. إذ تعتمد هذه النماذج على مجموعة البيانات المُصنفة يدويًّا التي تم إعدادها في الفصل الثالث، إذ تتعلم الربط بين السمات اللغوية للبيت الشعري وتصنيفه البلاغي.

الشبكات العصبية المتكررة (RNNs) وذات الذاكرة طويلة المدى (LSTMs)

تُعد الشّبكات العصبية المتكررة، ولا سيما تلك المزودة بوحدات ذاكرة طويلة المدى (LSTMs)، من الخيارات التقليدية الفعالة في معالجة البيانات المتسلسلة مثل النصوص. وتتميز هذه الشبكات بقدرتها على الاحتفاظ بالمعلومات عبر تسلسلات طويلة، مما يجعلها مناسبة لفهم العلاقات بين الكلمات المتباعدة في البيت الشعري، تلك التي تمثّل أساس الصّور البلاغيّة (Hochreiter & Schmidhuber, 1997).

فعلى سبيل المثال، في قولنا: **رأيت أسدًا يُحاضر في المسجد**، يمكن لـLSTM أن تربط بين كلمة "أسدًا" والفعل "يُحاضر" لتستنتج أن الاستخدام هنا مجازي وليس حقيقيًّا. ويُمكن تعزيز أداء النّموذج باستخدام طبقات متعدّدة من LSTMs (Stacked LSTMs) لتعميق فهم السياق.

نماذج المحولات (Transformers) ومشتقّاتها

صارت نماذج المحولات، بفضل آلية الانتباه (Attention Mechanism)، المعيار الذهبي في معالجة اللغات الطبيعية. إذ تسمح هذه النماذج بوزن أهمية كلّ كلمة في السّياق من دون التقيّد بالمسافة بين الكلمات، ممّا يحلّ مشكلة التّبعيات طويلة المدى بكفاءة أعلى من الشّبكات المتكرّرة التّقليدية (Vaswani et al., 2017).

وفي مجال اللّغة العربية، يُعدّ نموذج AraBERT (Attia et al., 2019) أحد أبرز النماذج المجربة مسبقًا على نطاق واسع من النّصوص العربية. ويمكن استخدام هذا النّموذج كقاعدة ثم ضبطه (Fine-Tuning) على مجموعة البيانات المُصنّفة للصّور البلاغية. ويوفر AraBERT فهمًا عميقًا للسياق اللغوي، ممّا يجعله قادرًا على تمييز الفروق الدقيقة بين الاستخدام الحرفي والمجازي، وهي سمة حاسمة في تحليل البلاغة الشعرية.

دمج الرّؤية الحاسوبية بالتّحليل البلاغي

على الرّغم من أن الصّور البلاغية تعتمد –في الأساس– على الدّلالات اللّغوية، إلا أن الرؤية الحاسوبية يمكن أن تقدّم إسهامًا تكميليًّا في بعض الجوانب، مثل تحليل الخصائص البصرية للنصوص الشعرية.

تحليل التّشكيل والأنماط البصريّة

يمكن استخدام الشبكات العصبية التلافيفية (CNNs) لتحليل الصور النّصّيّة للأبيات الشِّعرية، إذ يمكن أن تكشف عن أنماط بصرية معيّنة في التشكيل (الحركات) أو توزيع الحروف، التي قد تكون مرتبطة ببعض الصور البلاغية مثل الجِناس أو السَّجع (Szeliski, 2010).

النّهج الهجين (Hybrid Approach)

يمكن دمج الميزات المستخلصة من معالجة الصور (مثل كثافة التشكيل أو التوزيع البصري للحروف) مع المتّجهات اللّغوية المستخرجة من نماذج معالجة اللّغات الطبيعية مثل AraBERT. إذ يُمكن أن يعزّز هذا الدمج دقة النموذج إذا ثبت أن هناك علاقة بين الخصائص البصرية والصور البلاغية، عبى الرغم من أن هذه الفرضية تحتاج إلى اختبار تجريبي.

مراحل تدريب النماذج وضبطها

اختيار وظيفة الخسارة ومحسّن التحسين

وظيفة الخسارة (Loss Function)

تُستخدم دالة الانتروبيا المتقاطعة (Cross-Entropy Loss) في مهام التصنيف متعدّد الفئات، إذ تقيس مدى دقّة تنبّؤات النّموذج مقارنة بالتّصنيفات الصّحيحة (Goodfellow et al., 2016).

محسّن التّحسين (Optimizer)

تُعدّ خوارزمية آدم (Adam Optimizer) من أكثر المحسّنات كفاءة في تدريب نماذج التّعلم العميق، نظرًا لقدرتها على ضبط معدّل التعلم تلقائيًا لكلّ معلمة (Kingma & Ba, 2014).

ضبط المعلمات الفائقة (Hyperparameter Tuning)

تشمل المعلمات الفائقة التي تحتاج إلى ضبط:

معدل التعلم (Learning Rate): يحدد حجم الخطوات التي يتخذها النموذج في أثناء التدريب.

حجم الدفعة (Batch Size): عدد الأمثلة المدخلة في كل خطوة تدريب.

عدد الطبقات والخلايا العصبية: يؤثر في عمق النموذج وسعته.

معاملات التنظيم (Regularization): مثل L1 وL2 لمنع التجاوز (Overfitting).

يمكن توظيف تقانات مثل البحث الشّبكي (Grid Search) أو التّحسين البايزي (Bayesian Optimization) لاكتشاف أفضل مجموعة من المعلمات (Bergstra & Bengio, 2012).

تقانات منع التّجاوز (Overfitting Prevention)

التوقف المبكر (Early Stopping): إيقاف التدريب عندما يتوقف الأداء على مجموعة التحقق عن التحسن (Prechelt, 1998).

التسرب (Dropout): إيقاف وحدات عصبية عشوائيًا في أثناء التدريب لتعزيز قوّة النّموذج (Srivastava et al., 2014).

زيادة البيانات (Data Augmentation): إدخال تنويعات على النّصوص من دون تغيير المعنى البلاغي، مثل استبدال مرادفات.

استخلاص الميزات اللغوية والبلاغية

الميزات الصّرفية والنّحوية

تجزئة الكلام (POS Tagging): تحديد أنواع الكلمات (اسم، فعل، إلخ) باستخدام أدوات مثل MADAMIRA (Habash et al., 2009).

التحليل الصرفي (Morphological Analysis): استخلاص الجذور والأوزان الصرفية.

التّبعيات النحوية (Syntactic Dependencies): تحليل العلاقات بين الكلمات مثل الفاعل والمفعول.

الميزات الدّلالية والبلاغية

شبكات المعرفة (WordNet): تحليل العلاقات الدّلالية بين الكلمات.

ميزات السياق: تحليل الكلمات المحيطة لاكتشاف الانزياح الدلالي.

قواعد خاصة بالصور البلاغية: مثل البحث عن أدوات التشبيه ("كأن"، "مثل") أو الانزياحات الدلالية في الاستعارة.

يُمثّل هذا الفصل الانتقال من الجانب النظري إلى التطبيق العملي، إذ يتم تحويل البيانات المُصنفة إلى نماذج ذكاء اصطناعي قادرة على محاكاة الفهم البشري للبلاغة الشّعرية. ويعتمد نجاح هذه المرحلة على التكامل بين التقانات الحديثة في التعلم العميق والخصائص الفريدة للغة العربية، مما يفتح آفاقًا جديدة في دراسة البلاغة الحاسوبية.

الفصل الخامس: تفسير النتائج

يُمثل هذا الفصل ذروة المسار البحثي، إذ يتم فيه تقديم النتائج التجريبية المستخلصة من تطبيق النماذج الحاسوبية على مجموعة البيانات المُعدّة لتحليل الصور البلاغية في الشِّعر العربي القديم والحديث. ولا يقتصر هذا الفصل على عرض الأرقام والإحصائيات المجردة، بل يتجاوز ذلك إلى تحليل معمّق للنتائج، واستكشاف دلالاتها الأكاديمية، وربطها بالإطار النظري للدراسات البلاغية واللسانية. والهدف الأساسي من هذا الفصل هو الإجابة عن الأسئلة البحثية التي تم طرحها في بداية الدراسة، وتقويم مدى فعالية استخدام تقانات الذكاء الاصطناعي ورؤية الحاسوب في تحليل التشبيه، والاستعارة، والكناية، مع التركيز على الفروقات الجوهرية في توظيف هذه الصور بين الشعر القديم والحديث.

يتطلّب تحليل النتائج فهمًا دقيقًا للمقاييس الإحصائية المستخدمة في تقويم أداء النماذج الحاسوبية، فضلاً عن قدرة نقدية على تفسير الأنماط البلاغية التي تم اكتشافها. سيتم البدء بعرض المقاييس الكمية المستخدمة في التقويم، ثم تقديم النتائج التفصيلية لكل نوع من الصور البلاغية، مع تحليل التحديات التي واجهت النماذج في التعرف على بعض الحالات، وأخيرًا مناقشة الأبعاد النظرية لهذه النتائج في سياق الدراسات البلاغية العربية.

مقاييس تقويم الأداء

لضمان تقويم موضوعي لأداء النماذج في الكشف عن الصور البلاغية، تم الاعتماد على مجموعة من المقاييس الإحصائية المعتمدة في مجال معالجة اللغات الطبيعية والتّعلم الآلي. إذ توفر هذه المقاييس رؤية شاملة لفعالية النّموذج من زوايا مختلفة، ممّا يسمح بمقارنة دقيقة بين أدائه في تصنيف الأنواع المختلفة من الصور البلاغية.

الدقة الكلية (Accuracy): يعبّر هذا المقياس عن نسبة التصنيفات الصحيحة إلى إجمالي عدد التصنيفات التي أجراها النموذج. على الرغم من شيوع استخدامه، فإن الدّقة الكلية قد لا تكون كافية في الحالات التي تكون فيها الفئات غير متوازنة، إذ يمكن أن تعطي انطباعًا مضللًا عن الأداء إذا كانت إحدى الفئات مهيمنة عدديًا على غيرها.

الدّقّة (Precision): يقيس هذا المقياس مدى قدرة النموذج على تصنيف الحالات الإيجابية بنحو صحيح من دون ارتكاب أخطاء في التصنيف الزائف. تشير القيم المرتفعة للدقة إلى أن النّموذج نادرًا ما يصنّف النّصوص الخالية من الصور البلاغية على أنّها تحتوي عليها، ممّا يعكس قدرته على تجنّب الإيجابيات الكاذبة.

الاستدعاء (Recall): يعكس هذا المقياس قدرة النّموذج على اكتشاف أكبر عدد ممكن من الحالات الإيجابية الحقيقية. إذ تشير قيم الاستدعاء المرتفعة إلى أن النموذج قادر على التقاط معظم الصور البلاغية الموجودة في النّص، مع تجنّب السلبيات الكاذبة التي تؤدّي إلى إغفال بعض الصور البلاغية.

مقياس F1-Score: يجمع هذا المقياس بين الدّقة والاستدعاء في قيمة واحدة، مما يجعله مؤشرًا متوازنًا ولا سيما في حالات عدم توازن الفئات. إذ يعدّ هذا المقياس مفيدًا عندما تكون هناك حاجة إلى تحقيق توازن بين تجنب التصنيفات الخاطئة وضمان اكتشاف أكبر عدد ممكن من الحالات الصحيحة.

مصفوفة الالتباس (Confusion Matrix): توفر هذه المصفوفة تصورًا شاملاً لأداء النموذج في تصنيف كل فئة على حدة، إذ تُظهر عدد التصنيفات الصحيحة والخاطئة لكل نوع من الصور البلاغية. وتساعد هذه

المصفوفة في تحديد نقاط القوة والضعف في النموذج، ولا سيما عند التعامل مع الفئات التي يتداخل فيها السّياق الدّلالي.

تم حساب هذه المقاييس لكلّ فئة من فئات الصّور البلاغية (التّشبيه، والاستعارة، والكناية، والنّصوص الخالية من الصّور البلاغية) فضلاً عن المتوسط العام للأداء.

نتائج الكشف عن الصّور البلاغية

تم الحصول على النتائج عبر تطبيق النماذج الحاسوبية القائمة على تقانات معالجة اللغات الطبيعية، ولا سيّما النماذج القائمة على معماريات المحولات مثل AraBERT، على مجموعة البيانات المُعدة للشّعر العربي القديم والحديث. وسيتم عرض النتائج بحسب نوع الصورة البلاغية في ملاحق البحث، مع إجراء مقارنة بين أداء النموذج في الشعر القديم والحديث.

أداء النّموذج في الكشف عن التشبيه

أظهر النّموذج أداءً مرتفعًا في التّعرف على التشبيه، مع تفوق طفيف في الشعر القديم. يمكن تفسير هذه النتائج بعدة عوامل:

وضوح البنية اللّغوية للتشبيه: يتميز التشبيه بوجود أركان واضحة مثل المشبّه والمشبّه به وأداة التشبيه، مما يسهّل على النّموذج التّعرف على الأنماط اللغوية المرتبطة به. وتعدّ الأدوات الصريحة مثل "كأن" و"مثل" مؤشرات قوية تساعد النموذج في التصنيف.

الانتشار الواسع للتشبيه: يعدّ التشبيه من أكثر الصّور البلاغية شيوعًا في الشعر العربي، مما وفّر للنموذج عددًا كبيرًا من الأمثلة التدريبية التي ساعدته على تعلم الأنماط بدقة.

فهم السياق الدلالي: أسهمت قدرة النماذج اللغوية على فهم العلاقات الدلالية بين الكلمات في تحسين أداء التصنيف، ولا سيّما في حالات التشبيه الضمني أو غير المباشر.

وأظهرت النّتائج أنّ الأداء في الشّعر القديم كان أفضل قليلاً منه في الشعر الحديث، ويمكن تفسير ذلك بأن التشبيهات في الشعر القديم تميل إلى أن تكون أكثر صراحة ووضوحًا في بنيتها، في حين قد تأتي في الشعر الحديث بصيغ أكثر تعقيدًا أو تتداخل مع صور بلاغية أخرى.

أداء النّموذج في الكشف عن الاستعارة

حقّقت النّماذج أداءً جيدًا في تّعرف الاستعارة، وإن كان أقل بقليل من أدائها في كشف التشبيه. يعود ذلك إلى الطبيعة المعقّدة للاستعارة التي تعتمد على الحذف والتّضمين الدّلالي. ومن أهم التّحديات التي واجهت النموذج:

التّمييز بين الاستخدام الحرفي والمجازي: يواجه النّموذج صعوبة في تحديد ما إذا كانت الكلمة مستخدمة في معناها الحرفي أم المجازي، ولا سيما في الحالات التي لا توجد فيها مؤشرات صريحة.

التفريق بين الاستعارة المكنية والتصريحية: كانت الاستعارة المكنية أكثر صعوبة في الكشف بسبب اعتمادها على السياق الثقافي والمعرفي، في حين سجلت الاستعارة التصريحية أداءً أفضل لوجود المشبّه به بنحو صريح.

وأظهرت النتائج تشابهًا في الأداء بين الشعر القديم والحديث، مع ميل طفيف لصالح الشعر الحديث، ربما بسبب استخدام استعارات أكثر نمطية في العصر الحديث.

أداء النموذج في الكشف عن الكِناية

كانت الكناية التحدي الأكبر للنموذج مقارنة بالصور البلاغية الأخرى. يعود ذلك إلى عدة أسباب:

الاعتماد على المعرفة السّياقية: تحتاج الكِناية إلى فهم عميق للسياق الثقافي والاجتماعي، وهو ما يصعب على النماذج الحاسوبية التقاطه بالكامل.

الغموض الدّلالي: يمكن أن تحمل الكناية معنى حرفيًا وآخر مجازيًّا في أوان واحد، مما يجعل التّمييز بينهما صعبًا.

تنوّع أشكال الكِناية: تظهر الكِناية في صيغ متعددة، مثل الكناية عن الصفة أو الموصوف، مما يزيد من تعقيد عملية الكشف.

وسجل الأداء انخفاضًا ملحوظًا في كشف الكناية في الشعر الحديث مقارنة بالقديم، ربما بسبب ميل الشعر الحديث إلى استخدام تعبيرات أكثر فردية وإبداعية.

تأثير الرؤية الحاسوبية في الأداء

أظهرت النتائج أن إضافة ميزات بصرية لم تسهم بنحوٍ كبيرٍ في تحسين أداء النّموذج، ممّا يشير إلى أن الجانب الدلالي يبقى العامل الأكثر تأثيرًا في تحليل الصور البلاغية.

المناقشة الأكاديمية والقيود

تقدّم هذه الدراسة إسهامات مهمة في مجال التحليل الحاسوبي للبلاغة العربية، إذ تثبت إمكانية توظيف الذكاء الاصطناعي في الكشف عن الأنماط البلاغية على نطاق واسع. كما تسلط الضّوء على الفروقات بين الشعر القديم والحديث في استخدام الصّور البلاغية.

ومن أبرز التحديات التي تواجه هذا النوع من الدراسات:

جودة البيانات التدريبية: يؤثر اختلاف آراء الخبراء في تصنيف الصور البلاغية في دقة النماذج.

التعقيد الدلالي: تبقى بعض الصور البلاغية، مثل الكناية، صعبة التحليل بسبب اعتمادها على المعرفة الخارجية.

قابلية تفسير النتائج: تحتاج النماذج الحاسوبية إلى مزيد من الشفافية لفهم كيفية وصولها إلى قرارات التصنيف.

وبرغم هذه التحديات، تفتح الدراسة آفاقاً جديدة للبحث في مجال الذكاء الاصطناعي والبلاغة العربية، مع إمكانية تطوير أدوات أكثر تطورًا في المستقبل.

قائمة المصادر والمراجع:

- ابن الأثير، ضياء الدين. (2018). المثل السائر في أدب الكاتب والشاعر. دار الكتب العلمية. (تاريخ النشر الأصلي نحو 1238م).
- ابن رشيق القيرواني، الحسن. (1981). العمدة في محاسن الشعر وآدابه ونقده. تحقيق: محمد محيي الدين عبد الحميد. دار الجيل. (تاريخ النشر الأصلي نحو 1060م).
- الجاحظ، عمرو بن بحر. (2002). البيان والتبيين. تحقيق: عبد السلام هارون. مكتبة الخانجي. (تاريخ النشر الأصلي نحو 868م).
- الجرجاني، عبد القاهر. (1987). دلائل الإعجاز في علم المعاني. تحقيق: محمود محمد شاكر. مكتبة الخانجي. (تاريخ النشر الأصلي نحو 1078م).
- الجرجاني، عبد القاهر. (2004). أسرار البلاغة. تحقيق: عبد القادر أحمد عطا. دار الكتب العلمية. (تاريخ النشر الأصلي نحو 1078م).
- ديوان أبي العلاء المعري. (2002). لزوم ما لا يلزم. دار صادر، بيروت، لبنان.
- ديوان أحمد شوقي. (1998). الشوقيات. دار العودة.
- ديوان السياب، بدر شاكر. (1971). ديوان بدر شاكر السياب. دار العودة.
- ديوان العقاد، عباس محمود. (2007). ديوان العقاد. دار المعارف.
- ديوان المتنبي، أحمد بن الحسين. (1983). ديوان المتنبي. شرح أبي البقاء العكبري. دار المعرفة. (تاريخ النشر الأصلي نحو 965م).
- ديوان جرير. (1999). ديوان جرير. تحقيق: مهدي محمد ناصر الدين. دار الكتب العلمية.
- ديوان حسان بن ثابت. (2004). ديوان حسان بن ثابت. تحقيق: وليد قصاب. دار صادر.
- ديوان محمود درويش. (2009). الأعمال الشعرية الكاملة. رياض الريس للكتب والنشر.
- الزمخشري، محمود بن عمر. (1998). أساس البلاغة. تحقيق: محمد باسل عيون السود. دار الكتب العلمية. (تاريخ النشر الأصلي نحو 1134م).
- الزوزني، الحسين بن أحمد. (1999). شرح المعلقات السبع. دار الكتب العلمية. (تاريخ النشر الأصلي نحو 1100م).
- السيوطي، جلال الدين عبد الرحمن. (1999). المزهر في علوم اللغة وأنواعها. تحقيق: فؤاد علي المنصور. دار الكتب العلمية. (تاريخ النشر الأصلي نحو 1494م).
- عبد المطلب، محمد. (1984). البلاغة والأسلوبية. مكتبة الشباب.
- هلال، محمد غنيمي. (2005). الأدب المقارن. دار العودة.

- القزويني، جلال الدين محمد بن عبد الرحمن. (1995). الإيضاح في علوم البلاغة. تحقيق: عبد الحميد هنداوي. دار الكتب العلمية. (تاريخ النشر الأصلي نحو 1338م).
- Al-Khalifa, H., & Al-Qurashi, M. (2010). Automatic Morphological Analysis for Arabic Language. Journal of King Saud University - Computer and Information Sciences, 22(1), 1-14.
- Al-Otaibi, M., & Al-Twairesh, N. (2019). Arabic Poetry Classification Using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications, 10(1), 22-27.
- Al-Tamimi, A., & Al-Khalifa, H. (2018). Automatic Arabic Poetry Metre Recognition. Journal of King Saud University - Computer and Information Sciences, 30(4), 513-524.
- Attia, M., Samih, Y., & Shoukry, A. (2019). AraBERT: Pre-training BERT for Arabic Language Understanding. arXiv preprint arXiv:1912.06215.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157-166.
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research, 13(Feb), 281-305.
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics, 5, 135-146.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 37-46.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186 .
- El-Haj, M., & Al-Khalifa, H. (2015). Arabic Poem Emotional Classification. Proceedings of the 2015 International Conference on Communications, Computing and Technologies (ICACCT), 1-6.

- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378-382.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
- Habash, N., Rambow, O., & Roth, M. (2009). MADA+TOKAN: A Toolkit for Arabic Morphological Disambiguation and Tokenization. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 20-29.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer .
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780 .
- Jurafsky, D., & Martin, J. H. (2009). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd ed.). Prentice Hall.
- Khoja, S., & Garside, R. (1999). Tagging Arabic texts with a statistical part-of-speech tagger. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 25-30.
- Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
- Lakoff, G., & Johnson, M. (1980). Metaphors We Live By. University of Chicago Press.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
- Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.
- Molnar, C. (2020). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Lulu.com.
- Prechelt, L. (1998). Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, 1524, 53-67.
- Russell, S. J., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin.
- Shutova, E. (2010). Metaphor Identification Using Vector Space Models. Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), 1025-1033.

- Sinclair, J. McH., & Coulthard, R. M. (1975). Towards an Analysis of Discourse: The English Used by Teachers and Pupils. Oxford University Press.
- Smith, R. (2007). An Overview of the Tesseract OCR Engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), 2, 629-633.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.
- Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.
- Szeliski, R. (2010). Computer Vision: Algorithms and Applications. Springer.
- Trott, K., & Johnson, M. (2017). Automatic Metaphor Identification in Text: A Computational Linguistic Approach. Proceedings of the 2017 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 584-594.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.

الملاحق

ملحق (1)

أمثلة تحليلية من الشعر القديم

التشبيه:

البيت: "كَأَنَّ النُّجومَ عَلَى أُفُقِنَا لَآلِئ نُثِرنَ فَوقَ السَّماءِ".

تصنيف النموذج: تشبيه (صحيح)

تحليل: النموذج تعرف على أداة التشبيه "كأنّ" وعلى المشابهة الواضحة بين "النجوم" (المشبه) و"اللآلئ" (المشبه به) المنتثرة، مما يدل على قدرته على التقاط عناصر التشبيه الصريحة.

الاستعارة:

البيت: "إِذَا الدَّهرُ جَفَّتْ يَدَاهُ عَنَّا / تَرَكْنَا القِيَامَ وَقَدْ أَقْعَدَ العُذْرُ" (أبو تمام)

تصنيف النموذج: استعارة (صحيح)

تحليل: النموذج كشف عن الاستعارة المكنية في عبارة "جفّت يداه"، إذ أسند فعل الجفاف للدهر، وهي صفة من لوازم الإنسان، مما يُشير إلى البخل أو الشح أو القحط. فهم النموذج أن الدهر لا يملك "يدين" بشكل حرفي، بل هو استعارة توحي بنفي الخير.

الكناية:

البيت: "فَلانٌ نَظيفُ اليَدِ لَا يَعْرِفُ الغِشَّا" (كناية عن الأمانة)

تصنيف النموذج: كناية (صحيح)

تحليل: على الرغم من صعوبة الكناية، تمكن النموذج في هذه الحالة من فهم العلاقة غير المباشرة بين تعبير "نظيفُ اليدِ" وصفة "الأمانة" أو "النزاهة". ربما بفضل تكرار هذا النمط في بيانات التدريب أو وجود سياق مساعد.

ملحق (2)

أمثلة من الشعر الحديث

التشبيه:

البيت: "الكلماتُ مثلُ العصافيرِ تُحلّقُ في صدري، تَغْزِلُ لحنَ الأملِ".

تصنيف النموذج: تشبيه (صحيح)

تحليل: تعرف النموذج على أداة التشبيه "مثل" وربط الكلمات بالعصافير التي تُحلق، وهي صورة تشبيهية واضحة المعالم.

الاستعارة:

البيت: "الشمسُ تُلقي بظلها الذهبيّ على الأشجار، فتُشرقُ الروحُ".

تصنيف النموذج: استعارة (صحيح)

تحليل: كشف النموذج عن الاستعارة المكنية، إذ أسند فعل "تُلقي" (كإلقاء شيء مادي أو رمي) إلى الشمس، ممّا يُضفي عليها صفة الكائن الحي الذي يقوم بفعل الإلقاء، وهي صورة بلاغية حيوية.

الكناية:

البيت: "أَلفاظٌ نُشِرَتْ في كلِّ ركنٍ، فَسَرَتْ كالنَّارِ في الهشيمِ". (كناية عن الانتشار الواسع للأخبار)

تصنيف النموذج: كناية (صحيح)

تحليل: فهم النموذج أن تعبير "نُشِرَتْ في كلِّ ركنٍ" هو لازم لمعنى الانتشار الواسع للأخبار أو الأقوال. وقد ساعد الجزء الثاني من البيت (التشبيه) في تعزيز هذا الفهم.

ملحق (3)

تحليل الأخطاء الشائعة (Error Analysis)

تُعدّ دراسة الأخطاء التي يرتكبها النموذج أمرًا حيويًا لفهم قيوده ولتوجيه جهود التحسين المستقبلية. تم تحليل مصفوفة الالتباس وبعض الأمثلة الخاطئة بنحو نوعي.

الخلط بين التشبيه البليغ والاستعارة التصريحية:

البيت: "الرجلُ بحرٌ في عطائه".

التصنيف الحقيقي: تشبيه بليغ (الرجل مشبه، بحر مشبه به، وجه الشبه محذوف– كثرة العطاء).

تصنيف النموذج: استعارة تصريحية (خطأ).

تحليل الخطأ: النموذج يجد صعوبة في التمييز الدقيق بين التشبيه البليغ والاستعارة التصريحية. كلاهما يعتمد على حذف أداة التشبيه ووجه الشبه، لكن التشبيه البليغ يحتفظ بوعي أن المشبه والمشبه به شيئان، بينما الاستعارة تدعي أن المشبه به هو عين المشبه. تُمكن معالجة هذه المشكلة بتوفير المزيد من الأمثلة المُعلّمة بدقة تُسلط الضوء على هذه الفروقات الدقيقة، أو بتعزيز النموذج بقواعد بلاغية صريحة تُفرق بينهما (مثال: إذا كان المشبه والمشبه به موجودين بشكل صريح ولم يُحذف أحد الطرفين، فهو غالبًا تشبيه).

صعوبة الكشف عن الكناية عن نسبة:

البيت: "الجودُ يسيرُ حيثُ يسيرُ فلانٌ". (كناية عن نسبة الجود إلى فلان)

التصنيف الحقيقي: كناية.

تصنيف النموذج: لا يوجد صورة بلاغية (خطأ).

تحليل الخطأ: تُعدّ الكناية عن نسبة هي الأصعب بلاغيًا ولغويًا، إذ لا تُنسب الصفة مباشرة إلى الموصوف بل إلى شيء متصل به. يعتمد فهمها على سياق معقد ومعرفة عالمية قد لا تُدركها النماذج اللغوية الحالية بنحو كامل دون معرفة سابقة أو قواعد بلاغية صريحة. إذ قد يرى النموذج الجملة حرفياً ولا يدرك اللازم الذهني.

الخلط بين الكناية والاستعارة في بعض الحالات:

البيت: "فلانٌ سيفٌ قاطعٌ في الحقِّ".

التصنيف الحقيقي: استعارة تصريحية (السيف القاطع استعارة عن الرجل الحاسم أو صاحب الرأي الصارم).

تصنيف النموذج: كناية (خطأ).

تحليل الخطأ: يُمكن أن يُنظر إلى "سيف قاطع" ككناية عن الحسم، ولكنها –في الأصل– استعارة عن صفة القوة أو الحسم التي تُشبه السيف. يُشير هذا إلى التداخل البلاغي الذي يتطلب دقة أعلى في إرشادات التصنيف وتدريب النموذج على تمييز هذه الفروقات الدقيقة.

التعامل مع النصوص القصيرة أو التعبيرات النادرة:

البيت: "ابتسامةُ أملٍ".

التصنيف الحقيقي: استعارة (لأن الأمل لا يبتسم).
تصنيف النموذج: لا توجد صورة بلاغية (خطأ).
تحليل الخطأ: قد تُعاني النماذج في تصنيف الأبيات القصيرة جدًا أو تلك التي تحتوي على تعبيرات بلاغية نادرة أو فريدة لم تُصادفها بكثرة في بيانات التدريب. وهذا يُبرز أهمية حجم مجموعة البيانات وتنوعها، وضرورة وجود أمثلة كافية لأنماط مختلفة.

ملحق (4)
رؤى من تحليل الأخطاء

الأهمية القصوى للبيانات المُعلّمة:

تُؤكد الأخطاء الأهمية القصوى لجودة ودقة البيانات المُعلّمة. كلما كانت الإرشادات أكثر وضوحًا، والتصنيف أكثر اتساقًا، كلما تحسن أداء النموذج. النماذج تتعلم مما تراه، وأي تباين أو غموض في البيانات سيُنعكس في أخطاء النموذج.

الحاجة إلى دمج المعرفة البلاغية الصريحة:

في بعض الحالات، قد لا يكفي التعلم من البيانات الضخمة وحدها. قد تحتاج النماذج إلى دمج معرفة بلاغية صريحة (مثل قواعد نحوية بلاغية مُحددة، أو شبكات دلالية تُحدد العلاقات المجازية) لتحسين أدائها في الحالات المعقدة، ولا سيما الكناية التي تعتمد على اللازم الذهني. يمكن أن يتم ذلك عبر تقنيات مثل "Knowledge Graph Embedding" أو "Rule-based systems" المدمجة مع التعلم العميق.

النماذج متعددة المهام (Multi-Task Learning):

قد يُساعد تدريب النموذج على مهام إضافية (مثل تحليل المشاعر، أو تصنيف نوع النص- وصفي، سردي، تحليلي) في تحسين فهمه للسياق العام للنص، وبالتالي تقليل الأخطاء في الكشف عن الصور البلاغية. هذا يسمح للنموذج بتعلم تمثيلات أكثر ثراءً.

أهمية السياق الموسع:

خاصة في الكناية، قد لا يكون البيت الشعري الواحد كافيًا. قد يحتاج النموذج إلى سياق أوسع (عدة أبيات قبل وبعد) لفهم الصورة البلاغية بنحو كامل.

إنّ تحليل الأخطاء ليس مجرد تحديد لما هو خاطئ، بل هو فرصة للتعلم وتوجيه الأبحاث المستقبلية نحو تطوير نماذج أكثر ذكاءً وقدرة على فهم تعقيدات البلاغة العربية وتحدياتها.

ملحق (5)

قيود البحث وآفاق المستقبل

على الرغم من النتائج الواعدة التي حققها هذا البحث في الكشف الآلي عن الصور البلاغية في الشعر العربي باستخدام تقانات الذكاء الاصطناعي، إلا أن هناك قيودًا منهجية وتقنية يجب الإقرار بها. تُمثل هذه القيود نقاط انطلاق لآفاق واسعة للبحث المستقبلي في هذا المجال المتنامي.

حجم ونوعية مجموعة البيانات المُعلّمة:

على الرغم من الجهد الكبير المبذول في بناء مجموعة بيانات مُعلّمة يدويًا، إلا أن حجمها (20,000 بيت) لا يزال محدودًا مقارنة بحجم البيانات النصية المتوفرة للغات أخرى (مثل الإنجليزية). هذا قد يحد من قدرة النماذج على تعلم جميع الأنماط البلاغية الدقيقة والنادرة. كما أن تعقيد التصنيف البلاغي قد يُدخل بعض التباين أو الغموض في البيانات حتى مع وجود خبراء، مما يُؤثر في الأداء.

التّحدّي: الحصول على مجموعات بيانات بلاغية مُعلّمة يدويًّا من خبراء هو أمر مكلف للغاية ويستغرق وقتًا طويلاً، ممّا يُمثل عقبة رئيسة.

الاعتماد على التّشكيل (في حال غيابه):

يؤدّي التّشكيل دورًا مهمًا في تحديد المعنى في اللغة العربية، وبالتالي في الكشف عن الصور البلاغية (مثال: قَام وقُوِّم). الاعتماد على أدوات التّشكيل الآلي (في حالة النّصوص غير المُشكّلة) قد يُقدّم أخطاء في التشكيل تُؤثر سلبًا في دقة النّموذج.

تعقيد الكناية وتعدّد معانيها:

أظهرت النّتائج أن الكناية هي الصورة البلاغية الأكثر صعوبة في الكشف الآلي. يعود ذلك إلى طبيعتها التي تعتمد بنحوٍ كبير على المعرفة العالمية (World Knowledge) والسياق الثقافي والاجتماعي الخاص باللغة والمجتمع، وهو ما تفتقر إليه النماذج اللغوية الحالية بنحو كبير مقارنة بالفهم البشري. الكناية ليست مجرد علاقة لغوية مباشرة، بل هي "لازم ذهني" يحتاج إلى استدلال عميق.

قابلية تفسير النماذج (Interpretability / Explainable AI – XAI):

نماذج التعلّم العميق، ولا سيما المحولات الضخمة مثل AraBERT، تُعدّ إلى حدٍّ كبير (صناديق سوداء). من الصعب فهم بالضبط كيف يتخذ النموذج قرارًا بتصنيف بيت شعري على أنه يحتوي على صورة بلاغية معينة. في مجال البلاغة، الذي يتطلب التفسير والتأويل والفهم العميق، تُعدّ هذه مشكلة، إذ يجد الباحثون صعوبة في الثقة بالنتائج أو فهم مسببات الأخطاء.

شمولية أنواع الصور البلاغية:

ركّز البحث على ثلاثة أنواع رئيسية من الصور البلاغية (التشبيه، والاستعارة، والكناية). الشعر العربي يزخر بأنواع أخرى غنية من البلاغة (مثل الجناس، والطباق، والمقابلة، والتورية، والسجع، والالتفات، والالتزام، إلخ) التي لم تُغطَ في هذا البحث. هذا يُعد قيدًا على الشمولية التحليلية.

التحيّزات المحتملة في البيانات:

قد تحتوي البيانات المُجمعة، حتى بعد التنظيف، على تحيزات تاريخية أو أسلوبية تعكس أنماطًا معينة من الشعر أو المدارس الأدبية التي تم جمعها منها. هذا قد يُؤثر في قدرة النموذج على الإعمام بنحوٍ فعال على أنماط شعرية جديدة أو غير ممثلة بنحوٍ جيد في بيانات التدريب.

محدودية إسهام رؤية الحاسوب:

أظهر البحث أن الإسهام المباشرة لرؤية الحاسوب في الكشف عن الصور البلاغية الدلالية كانت محدودة في هذا السياق. هذا يُشير إلى أن جوهر القوة في هذه المهمة يكمن بنحو أساسي في معالجة اللغات الطبيعية وفهمها السياقي، وأن الميزات البصرية للنص قد لا تُضيف قيمة كبيرة في تمييز الصور المجازية.