الذكاء الاصطناعي ومستقبل التحليل اللساني للغة العربية: الفرص والتحديات في دراسة اللغة الإسلامية

Ahmad Faiz Nurrosyid
Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan, Indonesia
E-mail: ahmadfaiznurrosyid@gmail.com

Muhammad Muslih
Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan, Indonesia
E-mail: muslihmuhammad2002@gmail.com

ملخص البحث

يسعى هذا المقال إلى تقديم قراءة نقدية لعلاقة الذكاء الاصطناعي، ولا سيما النماذج اللغوية الكبيرة ونظم معالجة اللغة الطبيعية، بحقل اللسانيات العربية وتحليل النصوص الإسلامية. وينطلق البحث من سؤال محوري مفاده: كيف يمكن للذكاء الاصطناعي أن يسهم في تحويل التحليل اللساني للغة العربية ضمن حقل الدراسات الإسلامية، وما الفرص والحدود والمخاطر المعرفية التي يفرزها هذا التوظيف؟ اعتمد البحث منهج الدراسة المكتبية بمقاربة كيفية وصفية تحليلية نقدية، تقوم على المراجعة الأدبية والتحليل المفاهيمي والمقارن لأدبيات نشرت في غالبيتها بين عامي 2020 و2026، مع الاستعانة بمراجع تأسيسية أقدم كلما دعت الحاجة إلى تأصيل نظري في النحو والصرف والدلالة. وتبيّن الدراسة أن اللغة العربية بوصفها لغة غنية الصرف، متعددة التراكيب، حافلة بالإعراب والاشتقاق، تفرض تحديات حاسوبية نوعية على أدوات معالجة اللغة الطبيعية، تزداد تعقيدًا حين يتعلق الأمر بالنصوص القرآنية والحديثية والتراثية. وتخلص الدراسة إلى أن الذكاء الاصطناعي يقدّم إمكانات واعدة في التحليل الصرفي والنحوي الآلي، وبناء المدونات اللغوية، والبحث الدلالي، واسترجاع المعلومات من النصوص الإسلامية، غير أن هذه الإمكانات تظل مقيدة بإشكاليات الهلوسة المعلوماتية، والتحيز في البيانات، ومحدودية تمثيل العربية الفصحى التراثية في مدونات التدريب، وقصور الفهم الحاسوبي عن بلوغ مرتبة الفهم اللغوي والتأويل الديني البشري. ويقترح البحث تصورًا لنموذج تكاملي بين الإنسان والآلة، توكل فيه للذكاء الاصطناعي مهام المعالجة الأولية والفرز والبحث، بينما يبقى الحكم اللغوي والتفسير الديني وقفًا على الخبير البشري، مؤكدًا أن الذكاء الاصطناعي أداة مساعدة للتحليل لا سلطة بديلة عن الاجتهاد اللغوي والفقهي.

الكلمات المفتاحية: الذكاء الاصطناعي؛ معالجة اللغة الطبيعية؛ اللسانيات العربية؛ التحليل الصرفي والنحوي؛ النصوص الإسلامية؛ النماذج اللغوية الكبيرة؛ التعاون بين الإنسان والذكاء الاصطناعي

المقدمة

شهد العقد الأخير تطورًا متسارعًا في تقنيات الذكاء الاصطناعي (الذكاء الاصطناعي)، وبخاصة في حقل معالجة اللغة الطبيعية (معالجة اللغة الطبيعية)، إذ انتقلت هذه التقنيات من النماذج الإحصائية التقليدية إلى نماذج التعلم العميق القائمة على بنية المحوّل (Transformer)، ثم إلى النماذج اللغوية الكبيرة (النماذج اللغوية الكبيرة) القادرة على توليد نصوص متماسكة والإجابة عن أسئلة معقدة والتفاعل بلغة طبيعية. ولم تعد هذه التقنيات حكرًا على اللغة الإنجليزية، بل امتد أثرها إلى لغات متعددة من بينها اللغة العربية، التي راكمت خلال العقدين الأخيرين مجموعة معتبرة من الأدوات والمدونات الحاسوبية والنماذج المخصصة لها.

غير أن موقع اللغة العربية في هذا المشهد التقني ما زال، بحسب أدبيات الحقل، موقعًا وسطًا بين اللغات ذات الموارد الوفيرة كالإنجليزية والصينية واللغات محدودة الموارد؛ إذ تحظى العربية باهتمام متزايد لكنها تظل تعاني من ندرة نسبية في المدونات المشروحة لسانيًا، ومن التعقيد الصرفي والنحوي الذي يميزها عن كثير من اللغات الأخرى. وتزداد هذه الإشكالية حدة حين يتعلق الأمر بالعربية الفصحى التراثية والنصوص القرآنية والحديثية، التي تمثل نسبة محدودة جدًا من مدونات التدريب المستخدمة في بناء النماذج اللغوية الكبيرة المعاصرة.

تكمن أهمية هذا البحث في أنه لا يتناول الذكاء الاصطناعي بوصفه أداة تعليمية عامة لتعلم اللغة العربية، وهو التوجه الذي استأثر بقسط وافر من الأدبيات السابقة، بل يتناوله من زاوية لسانية بحتة: كيف يتفاعل الذكاء الاصطناعي مع بنية اللغة العربية الصرفية والنحوية والدلالية؟ وكيف يوظَّف هذا التفاعل في تحليل النصوص الإسلامية على وجه الخصوص؟ وهذا التمييز جوهري لأن الخلط بين "تعليم اللغة" و"تحليل اللغة" يُفضي غالبًا إلى تناول سطحي لا يفي بعمق الإشكاليات اللسانية والمعرفية التي يثيرها توظيف الذكاء الاصطناعي في هذا الحقل.

وتنبع إشكالية البحث من ملاحظة أساسية: بينما تتقدم أبحاث معالجة اللغة العربية الطبيعية بوتيرة لافتة، من خلال نماذج مثل AraBERT (Antoun et al., 2020) وARBERT وMARBERT، و Jais (Sengupta dkk, 2023) وAceGPT (Huang dkk, 2024)، فإن هذا التقدم يتركز غالبًا في المهام التقنية (تصنيف النصوص، تحليل المشاعر، الترجمة الآلية) دون أن يواكبه بالقدر ذاته حوار نقدي بين حقلي اللسانيات العربية التقليدية ودراسات اللغة الإسلامية من جهة، وعلوم الحاسوب من جهة أخرى. ويمثل هذا التباعد المعرفي فجوة بحثية (research gap) تستدعي معالجة تكاملية.

تشير الأدبيات المتاحة إلى وجود فجوة بحثية محددة على النحو الآتي: فمعظم الدراسات حول الذكاء الاصطناعي والعربية تنطلق من زاوية تقنية حاسوبية بحتة، بينما تظل الدراسات التي تربط هذا التطور التقني بالتراث اللساني العربي (النحو والصرف

والدلالة) وبخصوصية النصوص الإسلامية أقل عددًا نسبيًا، وتحتاج إلى مقاربة بينية تجمع بين علم الحاسوب واللسانيات الحاسوبية والدراسات الإسلامية.

وعليه، يهدف هذا البحث إلى الإجابة عن السؤال الرئيس الآتي: كيف يمكن للذكاء الاصطناعي أن يسهم في تحويل التحليل اللساني للغة العربية في إطار دراسات اللغة الإسلامية، وما الفرص والحدود والمخاطر المعرفية التي يفرزها هذا التوظيف؟ ويتفرع عن هذا السؤال الرئيس اثنا عشر سؤالًا فرعيًا تتناول: أثر تطور الذكاء الاصطناعي ومعالجة اللغة الطبيعية على البحث اللساني العربي خصائص اللغة العربية التي تشكل تحديًا حاسوبيًا توظيف الذكاء الاصطناعي في التحليل الصرفي والنحوي معالجة الغموض المعجمي والتركيبي والدلالي؛ فهم السياق في العربية الكلاسيكية؛ فرص تحليل القرآن والحديث والتراث الأخطاء المحتملة في فهم النصوص الإسلامية أثر تحيز البيانات ومحدودية المدونة العربية والهلوسة المعلوماتية إمكانية إحلال الذكاء الاصطناعي محل اللغوي والعالم الديني؛ نموذج التعاون الأمثل بين الإنسان والآلة؛ واتجاهات البحث المستقبلية.

منهجية البحث.

اعتمد هذا البحث منهج الدراسة المكتبية (library research) القائم على مقاربة كيفية وصفية تحليلية-نقدية (descriptive-analytical-critical approach qualitative)، وذلك بالجمع بين إجراءات المراجعة الأدبية المنهجية (literature review)، والتحليل الموضوعاتي (thematic analysis) لتصنيف الأدبيات وفق محاور البحث، والتحليل المقارن (comparative analysis) بين مقاربات النماذج اللغوية المختلفة في التعامل مع العربية، والتحليل المفاهيمي (conceptual analysis) لتمييز مستويات الأداء الحاسوبي (المعالجة، التعرف على الأنماط، التحليل اللغوي) عن مستويات الفهم والتفسير الإنساني.

اعتمد البحث على مصادر بيانات ثانوية، تتمثل في مقالات علمية محكّمة، وأوراق مؤتمرات دولية متخصصة في معالجة اللغة الطبيعية واللسانيات الحاسوبية، ومجموعات بيانات (datasets) موثقة، وكتب مرجعية في اللسانيات العربية. وقد اعتُمدت معايير محددة لاختيار الأدبيات، أبرزها: الصلة المباشرة بالمحاور الاثني عشر للبحث؛ وثوقية المصدر (مجلات محكمة، مؤتمرات معترف بها مثل ACL Anthology، قواعد بيانات مثل Scopus وWeb of Science)؛ وإمكانية التحقق من بيانات النشر (المؤلفون، السنة، اسم الدورية أو المؤتمر، رقم المعرف الرقمي DOI متى توافر).

استفاد هذا البحث أيضًا من أدبيات منشورة باللغة الإندونيسية تتناول توظيف الذكاء الاصطناعي في تعليم اللغة العربية والدراسات القرآنية ضمن السياق الأكاديمي الإندونيسي، وهو سياق ذو دلالة خاصة بالنظر إلى كون إندونيسيا أكبر الدول من حيث عدد السكان المسلمين وموطنًا لتقليد راسخ في تدريس اللغة العربية وعلوم القرآن عبر المعاهد الدينية (Pesantren) والجامعات الإسلامية الحكومية. فقد تناولت دراسات إندونيسية حديثة تصورات طلبة الدراسات العليا واستخداماتهم للذكاء الاصطناعي في تعلم اللغة العربية (Rohmawaty dkk, 2024؛ Hasaniyah, & ,Oktari, Harahap

2025)، وأسهمت دراسات أخرى في مسح الفرص والتحديات المرتبطة بتوظيف الذكاء الاصطناعي في الدراسات القرآنية والحديثية من منظور إسلامي إندونيسي (Hidayat, Rahardyanto, & Hardjita, 2020؛ Mauluddin, 2024). وتفيد هذه الأدبيات، رغم تركيزها على السياق التعليمي أكثر من التحليل اللساني الصرف، في تقديم شواهد ميدانية إضافية على اتساع الاهتمام بتوظيف الذكاء الاصطناعي في حقل اللغة العربية والدراسات الإسلامية خارج الفضاءين الأنجلوسكسوني والعربي. ومن زاوية مقارنة، يوفر البحث التأسيسي في معالجة اللغة الإندونيسية الطبيعية، وبخاصة في مجال التحليل الصرفي الآلي القائم على استلاب الزوائد (stemming؛ Amien, 2023) (Adriani dkk, 2007 confix-stripping)، مثالًا لغويًا مقارنًا يبرز أن إشكاليات التعقيد الصرفي وندرة الموارد الحاسوبية ليست حكرًا على العربية، بل تتقاطع معها لغات أخرى غنية صرفيًا في المنطقة نفسها من العالم الإسلامي.

ومن أبرز محددات هذا البحث أنه دراسة مكتبية لا تتضمن تجربة حاسوبية ميدانية جديدة أو بناء نموذج حاسوبي مبتكر، بل تنصبّ على تركيب نقدي للأدبيات المتاحة؛ كما أن حداثة الحقل، وبخاصة في شقه المتعلق بالنماذج اللغوية الكبيرة، تجعل بعض الاستنتاجات عرضة للتجاوز السريع مع الإصدارات المتلاحقة لهذه النماذج؛ فضلًا عن أن تركيز البحث على العربية الفصحى والنصوص التراثية والإسلامية يحدّ من إمكانية تعميم النتائج على اللهجات العربية المحكية.

نتائج البحث

بعد استعراض نقدي متوازن لأدبيات الحقل، إلى أن الذكاء الاصطناعي يمثل أداة واعدة لكنها محدودة في آن معًا ضمن حقل التحليل اللساني للغة العربية ودراسات اللغة الإسلامية. فمن جهة، أسهمت تقنيات معالجة اللغة الطبيعية والنماذج اللغوية الكبيرة في تطوير أدوات فعّالة للتحليل الصرفي والنحوي الآلي، وبناء مدونات لغوية موثقة كمدونة القرآن الحاسوبية ومدونة MASAQ، وتجيب هذه الخلاصة عن السؤال الرئيس للبحث بتأكيد أن الذكاء الاصطناعي يسهم في تحويل التحليل اللساني للعربية عبر تسريع المعالجة الأولية وتوسيع نطاق البحث الكمي، دون أن يرقى –في حدود ما تسمح به التقنية الحالية– إلى مرتبة الفهم اللغوي أو الديني الكامل الذي يقوم به الخبير البشري؛ وهو ما يستدعي اعتماد نموذج تكاملي واضح المعالم بين الإنسان والآلة، يوكل فيه للذكاء الاصطناعي دور المعالجة والفرز والترشيح، بينما يبقى الحكم اللغوي النهائي والتفسير الديني وقفًا على الخبير البشري المؤهل.

ومن الناحية النظرية، يسهم هذا البحث في تعميق الحوار البيني بين اللسانيات العربية التقليدية وعلوم الحاسوب، عبر تقديم إطار تحليلي يميّز بوضوح بين مستويات "القدرة الحاسوبية" و"التعرف على الأنماط" و"التحليل اللغوي" و"فهم السياق" و"تأويل المعنى"، وهو تمييز يمكن أن يفيد أبحاثًا مقبلة في ضبط مستوى الثقة الممنوح لمخرجات الذكاء الاصطناعي بحسب طبيعة المهمة المطلوبة.

أما من الناحية التطبيقية، فيوصي البحث الباحثين والمؤسسات الأكاديمية والدينية بعدم التعامل مع مخرجات النماذج اللغوية الكبيرة بوصفها أحكامًا لغوية أو فقهية جاهزة، بل بوصفها مقترحات أولية تستوجب التحقق والمراجعة من خبراء مؤهلين؛ كما يوصي بالاستثمار في بناء مدونات عربية تراثية وقرآنية وحديثية موسعة وموثقة تحت إشراف علمي مزدوج (لغوي وشرعي)، وبتطوير أدوات تحقق آلي متخصصة تحد من مخاطر الهلوسة المعلوماتية في السياقات الدينية.

ومن أبرز محددات هذا البحث طبيعته المكتبية التي لم تتضمن تجربة حاسوبية تطبيقية مباشرة، إضافة إلى التطور المتسارع لحقل النماذج اللغوية الكبيرة الذي قد يتجاوز بعض المعطيات التقنية الواردة هنا خلال فترة وجيزة؛ ولذا يوصي البحث بدراسات مستقبلية تجريبية تقيّم أداء نماذج عربية محددة على مهام تحليل صرفي ونحوي ودلالي دقيقة للنصوص القرآنية والحديثية، وبدراسات نوعية تستطلع تصورات اللغويين والعلماء الشرعيين أنفسهم إزاء توظيف الذكاء الاصطناعي في حقل اختصاصهم، بما يرسخ مقاربة تكاملية متوازنة بين عمق التراث اللغوي العربي الإسلامي وإمكانات التقنية الحديثة.

الذكاء الاصطناعي ومعالجة اللغة العربية.

مرت معالجة اللغة الطبيعية بمراحل متعاقبة من الأنظمة القائمة على القواعد (rule-based systems)، إلى النماذج الإحصائية (statistical models)، ثم التعلم الآلي (Machine Learning)، فالتعلم العميق (Deep Learning) القائم على الشبكات العصبية المتكررة والتلافيفية، وصولًا إلى بنية المحوّل (Transformer) التي أحدثت نقلة نوعية سنة 2017، ومهّدت لظهور النماذج اللغوية الكبيرة (النماذج اللغوية الكبيرة) والذكاء الاصطناعي التوليدي (الذكاء الاصطناعي التوليدي) في السنوات الأخيرة.

وقد استفادت اللغة العربية من هذا المسار التقني، وإن بوتيرة متفاوتة قياسًا بالإنجليزية. ففي مرحلة النماذج المدرَّبة مسبقًا القائمة على بنية BERT، ظهرت نماذج مثل AraBERT (Antoun dkk, 2020)، ثم ARBERT وMARBERT (Abdul-Mageed dkk, 2021)، وهي نماذج أسهمت في تحسين أداء مهام الفهم اللغوي كالتصنيف والتعرف على الكيانات المسماة والإجابة عن الأسئلة. وقد رافق ذلك تطوير أدوات معالجة أساسية متخصصة بالعربية مثل CAMeL Tools (Obeid et al., 2020)، الذي يوفر وحدات للتحليل الصرفي، والتعرف على اللهجات، والتعرف على الكيانات المسماة، وتحليل المشاعر؛ وأداة Farasa للتجزئة الصرفية السريعة والدقيقة (Mubarak, & Darwish 2016).

أما في مرحلة النماذج اللغوية الكبيرة، فقد شهدت السنوات 2023–2026 إطلاق نماذج عربية مخصصة من أبرزها Jais وJais-chat، وهو نموذج بحجم 13 مليار معامل قائم على بنية GPT-3، دُرِّب على مزيج من النصوص العربية والإنجليزية، وأظهر أداءً متفوقًا على النماذج العربية المفتوحة المصدر المتاحة آنذاك (Sengupta et al., 2023)؛ ثم توالت إصدارات

أحدث من عائلة Jais بأحجام أكبر. كما ظهر نموذج AceGPT الذي اعتمد استراتيجية "توطين" (localization) خاصة، تجمع بين التدريب الإضافي على نصوص عربية، والضبط الدقيق بتعليمات عربية أصيلة، والتعلم المعزز بالتغذية الراجعة من الذكاء الاصطناعي (RLAIF) لمواءمة مخرجات النموذج مع القيم والسياق الثقافي العربي (Huang et al., 2024).

ولا يقتصر التقدم على النماذج التأسيسية العامة، بل يمتد إلى معايير تقييم مخصصة لقياس الكفاءة الثقافية والدينية للنماذج اللغوية، من أبرزها مهمة PalmX 2025 المشتركة، التي صُممت لتقييم مدى فهم النماذج اللغوية الكبيرة للثقافة العربية والإسلامية عبر أسئلة اختيار من متعدد تغطي 22 دولة عربية، وقد بيّنت نتائجها أن أفضل النماذج المشاركة لم تتجاوز دقة 72.15% في فرع الثقافة العامة، و84.22% في فرع المعرفة الإسلامية، وهو ما يكشف عن فجوة معرفية ثقافية–دينية لا تزال قائمة رغم التقدم التقني (Alwajih dkk, 2025).

ولفهم هذا التطور فهمًا دقيقًا، لا بد من التمييز بين مستويات متعددة من "الأداء" الحاسوبي: فثمة القدرة الحاسوبية المحضة (computational capability) المتمثلة في معالجة النصوص وتخزينها واسترجاعها؛ وثمة التعرف على الأنماط اللغوية (pattern recognition) الذي تحققه النماذج الإحصائية والعصبية عبر رصد الانتظامات في البيانات؛ وثمة التحليل اللغوي (linguistic analysis) بمعناه الفني الدقيق، أي تفكيك البنية الصرفية والنحوية وفق قواعد صريحة أو مكتسبة؛ وثمة فهم السياق (contextual understanding)، وهو أكثر تعقيدًا لأنه يستلزم ربط الدلالة بالمقام والملابسات؛ وأخيرًا ثمة تأويل المعنى (interpretation of meaning) والتفسير الديني (religious exegesis)، وهما مستويان يستلزمان حكمًا معرفيًا وقيميًا يتجاوز ما تقدمه النماذج الحاسوبية الحالية. وسيتكرر هذا التمييز عبر فصول هذا البحث بوصفه إطارًا تحليليًا ناظمًا.

الخصائص اللسانية للغة العربية وتحديات الذكاء الاصطناعي.

تُعَدّ اللغة العربية من أكثر اللغات الطبيعية تعقيدًا من الناحية الحاسوبية، وذلك لعدد من الخصائص اللسانية المتشابكة. فعلى المستوى الصرفي، تقوم العربية على نظام اشتقاقي جذري قائم على الجذر (الجذر) والوزن (pattern/wazan)، بحيث تتولد من جذر ثلاثي أو رباعي واحد عشرات الصيغ الصرفية المختلفة عبر الاشتقاق (derivation) والتصريف (inflection)، ولحوق الزوائد الصرفية القبلية والبعدية والداخلية (and affixes clitics)، وهو ما يجعل الكلمة العربية الواحدة قد تعادل جملة كاملة من حيث المعلومات النحوية والدلالية التي تحملها. وقد أظهرت دراسة CAMeL Tools أن صيغ الفعل في العربية الفصحى الحديثة وحدها قد تتجاوز 5400 صيغة صرفية (Obeid dkk, 2020)، وهو رقم يعكس حجم التحدي الحاسوبي في هذا المستوى.

ويضاف إلى ذلك الغموض الرسمي (orthographic ambiguity) الناتج عن اختيارية علامات التشكيل في الكتابة العربية المعاصرة؛ إذ تُكتب النصوص عمومًا دون تشكيل، وهو ما يجعل الكلمة الواحدة عرضة لتعدد القراءات الصرفية والدلالية

المحتملة، ولا يُحسم الاختيار بينها إلا بالسياق أو بالتشكيل الصريح. ويشتد هذا الإشكال في النصوص التراثية والقرآنية، حيث يكون التشكيل الدقيق حاسمًا في تحديد المعنى والإعراب.

وعلى المستوى الدلالي، تتسم العربية بظاهرة تعدد المعنى (polysemy) الواسعة، إذ يحمل كثير من الجذور والألفاظ معاني متعددة يحددها السياق، فضلًا عن الترادف (synonymy) والاشتراك اللفظي (homonymy)، وهي ظواهر تتطلب من نظم معالجة اللغة الطبيعية نمذجة السياق الدلالي بدقة تتجاوز مجرد التعرف على شكل الكلمة.

الجدول: مقارنة بين أبرز خصائص اللغة العربية وتحدياتها الحاسوبية أمام الذكاء الاصطناعي

التحدي الحاسوبي المقابل	الخصيصة الأساسية	المستوى اللساني
صعوبة التحليل الصرفي الآلي الدقيق وتفكيك الزوائد والسوابق واللواحق	الاشتقاق من الجذر والوزن، وتعدد الصيغ (أكثر من 5400 صيغة فعلية في الفصحى الحديثة)	الصرف
غموض في تحديد الحركات الإعرابية والقراءة الصحيحة للكلمة	اختيارية التشكيل في الكتابة المعاصرة	الرسم الإملائي
صعوبة التحليل التركيبي الآلي (parsing) وتحديد الوظائف النحوية دون تشكيل	حرية نسبية في ترتيب الجملة، ونظام إعرابي دقيق (iʿrāb)	النحو
صعوبة تفكيك الغموض الدلالي دون سياق غني	اتساع تعدد المعنى (polysemy) والاشتراك اللفظي	الدلالة
ضعف تمثيل النصوص الكلاسيكية والقرآنية في مدونات تدريب النماذج المعاصرة	الفارق بين الفصحى الحديثة والفصحى التراثية/الكلاسيكية	التنوع الزمني

المصدر: تركيب الباحث استنادًا إلى (Obeid dkk, 2020؛ Bashir dkk., 2023؛ Mashaabi dkk., 2024).

ويضاف إلى ما سبق التفاوت الكبير بين العربية الفصحى الحديثة (Standard Arabic Modern) والعربية الفصحى التراثية أو الكلاسيكية (Classical Arabic)، من حيث المفردات والتراكيب والأساليب البلاغية؛ فالنصوص القرآنية والحديثية والتراثية تحمل خصائص أسلوبية وصرفية ونحوية تختلف عن نظيرتها في العربية المعاصرة، بينما تتركز غالبية مدونات التدريب المستخدمة في بناء النماذج اللغوية الكبيرة على العربية الفصحى الحديثة والمحتوى الرقمي المعاصر، مما يخلق فجوة تمثيلية واضحة تنعكس سلبًا على أداء هذه النماذج حين تُطبَّق على النصوص الكلاسيكية (Bashir dkk., 2023؛ Mashaabi dkk., 2024).

المراجع

Abdul-Mageed, M., Elmadany, A., & Nagoudi, E. M. B. (2021). ARBERT & MARBERT: Deep bidirectional transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics* (pp. 7085–7105). Association for Computational Linguistics.

Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S. M. M., & Widyantoro, D. H. (2007). Stemming Indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing (TALIP), 6*(4), 1–33. https://doi.org/10.1145/1316450.1316452

Alwajih, F., El Mekki, A., Mubarak, H., Hawasly, M., Mohamed, A., & Abdul-Mageed, M. (2025). PalmX 2025: The first shared task on benchmarking LLMs on Arabic and Islamic culture. In *Proceedings of the Arabic NLP 2025 Shared Tasks.* Association for Computational Linguistics. arXiv:2509.02550.

Amien, M. (2023). Sejarah dan perkembangan teknik Natural Language Processing (NLP) bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia. *ELANG: Journal of Interdisciplinary Research*. arXiv:2304.02746.

Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based model for Arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT4)* (pp. 9–15). European Language Resources Association.

Bashir, M. H., Azmi, A. M., Nawaz, H., Zaghouani, W., Diab, M., Al-Fuqaha, A., & Qadir, J. (2023). Arabic natural language processing for Qur'anic research: A systematic review. Artificial *Intelligence Review, 56*(7), 6801–6854. https://doi.org/10.1007/s10462-022-10313-2

Darwish, K., & Mubarak, H. (2016). Farasa: A new fast and accurate Arabic word segmenter. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 1070–1074). European Language Resources Association.

Hidayat, R., Rahardyanto, S., & Hardjita, P. W. (2020). Survey paper: Tantangan dan peluang kecerdasan buatan dalam bidang Islam, Qur'an dan Hadits. *Prosiding Konferensi Integrasi Interkoneksi Islam dan Sains, 2*, 343–346.

Huang, H., Yu, F., Zhu, J., Sun, X., Cheng, H., Song, D., Chen, Z., Alharthi, A., An, B., He, J., Liu, Z., Zhang, Z., Chen, J., Li, J., Wang, B., Zhang, L., Sun, R., Wan, X., Li, H., & Xu, J. (2024). AceGPT, localizing large language models in Arabic. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 8139–8163). Association for Computational Linguistics.

Mauluddin, M. (2024). Kontribusi Artificial Intelligence (AI) pada studi Al-Qur'an di era digital; peluang dan tantangan. Madinah: *Jurnal Studi Islam, 11*(1), 99–113. https://doi.org/10.58518/madinah.v11i1.2518

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020). CAMeL Tools: An open source Python toolkit for Arabic natural language processing. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)* (pp. 7022–7032). European Language Resources Association.

Oktari, S. J., Harahap, D. H., & Hasaniyah, N. (2025). Artificial intelligence and the transformation of Arabic meaning: Challenges and opportunities. *Abjadia: International Journal of Education, 10*(1), 58–71.

Rohmawaty, E. N., Hilmi, D., Salimul Uqba, M. S., & Saleh, U. S. (2024). Peran Artificial Intelligence (AI) dalam pembelajaran Bahasa Arab mahasiswa pascasarjana UIN Maulana Malik Ibrahim Malang. *Khatulistiwa: Jurnal Pendidikan dan Sosial Humaniora, 4*(3), 316–328. https://doi.org/10.55606/khatulistiwa.v4i3.4023

Sengupta, N., Sahu, S. K., Jia, B., Katipomu, S., Li, H., Koto, F., Marshall, W., Gosal, G., Liu, C., Chen, Z., Afzal, O. M., Kamboj, S., Pandit, O., Pal, R., Pradhan, L., Mujahid, Z. M., Baali, M., Han, X., Bsharat, S. M., Aji, A. F., Shen, Z., Liu, Z., Vassilieva, N., Hestness, J., Hock, A., Feldman, A., Lee, J., Jackson, A., Ren, H. X., Nakov, P., Baldwin, T., & Xing, E. (2023). *Jais and Jais-chat: Arabic-centric foundation and instruction-tuned open generative large language models.* arXiv:2308.16149.