

قضايا بحثية 3 - التكاليف 10

Prepared by: **Abobakr Mohammed Baras**
ID: **253002656**

Aug 2026



10- اعداد ملخص تنفيذي لأحد المشاريع البحثية الحديثة في مجال التخصص.

يطلب من الباحث إعداد ملخص تنفيذي لأحد المشاريع البحثية الحديثة ذات الصلة بمجال تخصصك ذه المهمة تتطلب البحث والاطلاع على أحد المشاريع البحثية الحديثة في مجال التخصص، ثم إعداد ملخص تنفيذي له.

الملخص التنفيذي يجب أن يكون موجزا ودقيقا ويسلط الضوء على الجوانب الرئيسية للمشروع البحثي كالأهداف والمنهجية والنتائج والتوصيات

ملخص تنفيذي

تقييم النماذج اللغوية الكبيرة العربية مسح لمعايير القياس والأساليب والفجوات

مشروع بحثي حديث في مجال معالجة اللغات الطبيعية (NLP)

معهد الابتكار التكنولوجي (TII) — أبوظبي، الإمارات | أكتوبر 2025

بطاقة تعريف المشروع البحثي

الورقة العلمية المرفقة مع العرض

Evaluating Arabic Large Language Models: A Survey of Benchmarks, Methods, and Gaps

عنوان البحث

تقييم النماذج اللغوية الكبيرة العربية: مسح لمعايير القياس والأساليب والفجوات

الترجمة العربية

أحمد الزبيدي وشيخة السويدي وبسمة الأمل و 5 باحثين آخرين (8 مؤلفين)

المؤلفون

معهد الابتكار التكنولوجي (TII) — أبوظبي، الإمارات

جهة البحث

15 أكتوبر 2025

تاريخ النشر

علوم الحاسب — اللسانيات الحاسوبية ومعالجة اللغات الطبيعية

مجال التصنيف

مراجعة منهجية شاملة (Survey) لمعايير تقييم النماذج اللغوية العربية

نوع الدراسة

المشكلة البحثية: لماذا هذا المشروع؟

1

فجوة في اللغة والموارد

العربية لغة نحو 500 مليون متحدث، لكن محتواها الرقمي شحيح ومحدود التنوع، ما يرفع كلفة بناء معايير التقييم.

2

تعقيد لغوي و لهجي

تعقيد صرفي عالٍ وأكثر من 20 لهجة تكاد تعمل كلغات مستقلة، ما يجعل التقييم الشامل صعباً.

3

ضعف البدائل السريعة

ترجمة المعايير الإنجليزية تُنتج اختبارات غير متوائمة ثقافياً، والتوليد الاصطناعي يورث تحيز النموذج المولّد.

4

غياب مرجع جامع

لم يوجد قبل هذا العمل مسح شامل لمعايير تقييم النماذج اللغوية العربية رغم تسارع إصدارها.

أهداف المشروع البحثي

ثلاث مساهمات رئيسية أعلنها الباحثون

بناء تصنيف مرجعي

1

اقتراح تصنيف (Taxonomy) ينظم معايير التقييم العربية ضمن أربع فئات كبرى.

تحليل أكثر من 40 معياراً

2

تحليل تفصيلي لأكثر من 40 معيار تقييم عربياً ليكون مرجعاً موحداً لمجتمع معالجة اللغات الطبيعية.

رصد الأساليب والفجوات

3

مناقشة أساليب البناء وأدوات التقييم والاتجاهات الحديثة، وتحديد الفجوات الحرجة مع مستودع مفتوح يجمع الروابط.

منهجية المشروع

مراجعة منهجية للأدبيات وليست دراسة تجريبية

1

الحصر

جمع ومراجعة أكثر من 40 معياراً عربياً
منشوراً حتى مطلع عام 2025.

2

التصنيف

اشتقاق تصنيف رباعي من الأنماط المشتركة
بين المعايير.

3

التحليل المقارن

مقارنة المعايير من حيث مصدر البيانات
والحجم ومقاييس التقييم ومدى إتاحتها.

4

الاستخلاص

رصد الفجوات والمشكلات المنهجية ثم
صياغة توصيات للمجتمع البحثي.

ثلاثة أساليب لبناء المعايير جرت مقارنتها: جمع محتوى عربي أصلي، وترجمة معايير إنجليزية، وتوليد اصطناعي بنماذج لغوية — وقد نوقشت مفاضلاتها من حيث الأصالة والحجم والكلفة.

التصنيف الرباعي لمعايير التقييم

المخرج المنهجي الأول للمشروع

مهام معالجة اللغة

من المهام المفردة كالتصنيف وفهم المقروء إلى المعايير الشاملة متعددة المهام.

المعرفة

معايير المعرفة العامة والعلوم والرياضيات، إضافة إلى مجالات متخصصة كالقانون والطب والشعر.

معايير مخصصة

قياس خصائص بعينها: السلامة، والهلوسة، واتباع التعليمات، والقدرات متعددة الوسائط.

الثقافة واللهجات

قياس الوعي الثقافي وفهم اللهجات العربية وتوليدها والتمييز بينها.

أبرز النتائج

21

معيّاراً فقط تتيح
البيانات والكود معاً

82%

منها صدرت خلال
عامي 2024 و2025

41

معيّار تقييم عربي
جرى تحليلها

تحوّل واضح في المعايير الحديثة من الترجمة إلى المحتوى العربي الأصلي، استجابة لمشكلة عدم المواءمة الثقافية.

التحكيم بنموذج لغوي يرتبط بالحكم البشري ارتباطاً قوياً في أحد المعايير، لكنه نادراً ما يُتحقق منه في السياق العربي.

الفجوات الحرجة التي رصدتها المشروع

التقييم الزمني

لا تقيس المعايير الحالية تعامل النماذج مع تغيّر المعلومات عبر الزمن.

الحوار متعدد الأدوار

قياس ضعيف لقدرة النموذج على متابعة محادثة ممتدة.

التبديل اللغوي

غياب معايير للخلط بين العربية والإنجليزية أو الفرنسية داخل النص.

اللهجات قليلة الموارد

تغطية غير متوازنة تُهمل لهجات مثل السودانية والموريتانية.

الفهم التداولي

السخرية والتلميح والمعنى غير المباشر خارج نطاق القياس الحالي.

تلوث البيانات

اتساع بيانات التدريب يهدد صدق المعايير الثابتة وقابلية الوثوق بنتائجها.

توصيات المشروع للباحثين

ثمانى توصيات لتطوير معايير تقييم عربية موثوقة

1 إعطاء الأولوية للمحتوى العربي الأصلي مع مراجعة ثقافية إلزامية

2 إتاحة البيانات وكود التقييم كاملاً لضمان قابلية التكرار

3 مواجهة تلوث البيانات باختبارات مغلقة وترشيح تنافسي

4 توحيد صياغة التلقين وتنسيق أسئلة الاختيار من متعدد

5 التحقق من أساليب التقييم مقابل أحكام بشرية متنوعة

6 توسيع التغطية اللهجية بشكل منهجي ومتوازن

7 تطبيق إجراءات صارمة لضبط جودة البيانات قبل النشر

8 تطوير آليات تقييم زمنية وديناميكية تتجدد باستمرار

الخلاصة ومصدر البحث

يقدم المشروع أول مرجع منهجي شامل لمعايير تقييم النماذج اللغوية العربية: يصنفها في أربع فئات، ويحلل أكثر من أربعين معياراً، ويكشف أن التسارع الكمي في إصدارها لم يقابله انضباط في الشفافية وقابلية التكرار والمواثيق الثقافية.

المصدر

Alzubaidi, A., Alsuwaidi, S., Boussaha, B. E. A., AlQadi, L., Alkaabi, O., Alyafeai, M., Alobeidli, H., & Hacid, H. (2025). Evaluating Arabic Large Language Models: A Survey of Benchmarks, Methods, and Gaps. arXiv:2510.13430

<https://arxiv.org/abs/2510.13430>

<https://arxiv.org/pdf/2510.13430>



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You

