

# إطار تقييم الذكاء الاصطناعي والنماذج اللغوية العربية: من اللغة إلى الثقافة

Assem Hijazi – Experience, Knowledge & Ideas

## Contents

|    |                                                                            |
|----|----------------------------------------------------------------------------|
| 1  | إطار تقييم الذكاء الاصطناعي والنماذج اللغوية العربية: من اللغة إلى الثقافة |
| 4  | المقدمة                                                                    |
| 5  | أعمدة تقييم نماذج اللغة العربية والذكاء الاصطناعي                          |
| 5  | الفئة الأولى: المهارات اللغوية الأساسية                                    |
| 6  | الفئة الثانية: الفهم السياقي والأسلوبي                                     |
| 6  | الفئة الثالثة: الكفاءة الثقافية والمنطقية                                  |
| 7  | (Grammar Accuracy) الركيزة الأولى: النحو                                   |
| 7  | :الركائز الفرعية                                                           |
| 7  | :معايير الاختبار                                                           |
| 7  | :ملاحظات التقييم                                                           |
| 8  | (Morphology Accuracy) الركيزة الثانية: الصرف                               |
| 8  | :الركائز الفرعية                                                           |
| 8  | :معايير الاختبار                                                           |
| 8  | :ملاحظات التقييم                                                           |
| 9  | (Spelling & Diacritics) الركيزة الثالثة: الإملاء والتشكيل                  |
| 10 | (Root & Derivation Competence) الركيزة الرابعة: الجذور والاشتقاق           |
| 10 | :الركائز الفرعية                                                           |
| 10 | :معايير الاختبار                                                           |
| 10 | :ملاحظات التقييم                                                           |
| 11 | (Lexical Richness) الركيزة الخامسة: الثراء المعجمي                         |
| 11 | :الركائز الفرعية                                                           |
| 11 | :معايير الاختبار                                                           |
| 11 | :ملاحظات التقييم                                                           |
| 12 | (Semantic Understanding) الركيزة السادسة: الفهم الدلالي                    |
| 12 | :الركائز الفرعية                                                           |
| 12 | :معايير الاختبار                                                           |
| 12 | :ملاحظات التقييم                                                           |

|                                                                                                  |    |
|--------------------------------------------------------------------------------------------------|----|
| الركيزة السابعة: البلاغة والأسلوب (Rhetoric & Style).....                                        | 13 |
| الركائز الفرعية :.....                                                                           | 13 |
| معايير الاختبار :.....                                                                           | 13 |
| ملاحظات التقييم :.....                                                                           | 13 |
| الركيزة الثامنة: فهم اللهجات (Dialectal Competence).....                                         | 14 |
| الركائز الفرعية :.....                                                                           | 14 |
| معايير الاختبار :.....                                                                           | 14 |
| الركيزة التاسعة: الحساسية الثقافية (Cultural Sensitivity) .....                                  | 15 |
| الركائز الفرعية :.....                                                                           | 15 |
| معايير الاختبار :.....                                                                           | 15 |
| ملاحظات التقييم :.....                                                                           | 15 |
| الركيزة العاشرة: الحساسية الدينية (Religious Sensitivity).....                                   | 16 |
| الركائز الفرعية :.....                                                                           | 16 |
| ملاحظات التقييم :.....                                                                           | 16 |
| الركيزة الحادية عشرة: التغطية الموضوعية (Domain Coverage) .....                                  | 17 |
| الركائز الفرعية :.....                                                                           | 17 |
| معايير الاختبار :.....                                                                           | 17 |
| ملاحظات التقييم :.....                                                                           | 17 |
| الركيزة الثانية عشرة: التفكير المنطقي (Logical Reasoning).....                                   | 18 |
| الركائز الفرعية :.....                                                                           | 18 |
| معايير الاختبار :.....                                                                           | 18 |
| ملاحظات التقييم :.....                                                                           | 18 |
| الركيزة الثالثة عشرة: شفافية مصادر المعرفة (Knowledge Source Transparency).....                  | 19 |
| الركائز الفرعية :.....                                                                           | 19 |
| معايير الاختبار :.....                                                                           | 19 |
| ملاحظات التقييم :.....                                                                           | 19 |
| الركيزة الرابعة عشرة: التعامل مع الوثائق المخصصة العربية (Custom Arabic Document Handling) ..... | 20 |
| الركائز الفرعية :.....                                                                           | 20 |
| معايير الاختبار :.....                                                                           | 20 |
| ملاحظات التقييم :.....                                                                           | 20 |
| الركيزة الخامسة عشرة: التكيف عبر الركائز (Adaptability Across Pillars) .....                     | 21 |
| الركائز الفرعية :.....                                                                           | 21 |

|                                  |    |
|----------------------------------|----|
| معايير الاختبار .....            | 21 |
| ملاحظات التقييم .....            | 21 |
| خطوات بناء المؤشر النهائي .....  | 22 |
| الخطوات .....                    | 22 |
| جدول الأوزان المقترح (مرن) ..... | 22 |

## المقدمة

تُعدّ اللغة العربية، التي يتحدث بها أكثر من 400 مليون إنسان، نظامًا لغويًا فريدًا يتميز بتعقيداته النحوية والصرفية، وبلاغته الثرية، وتنوع لهجاته، وحساسيته الثقافية والدينية. ورغم هذا الغنى، فإن نماذج الذكاء الاصطناعي اللغوي (LLMs) والتقنيات الذكية الأخرى المخصصة للغة العربية ما زالت تواجه تحديات كبيرة، ليس فقط في التطوير، بل أيضًا في تقييم أدائها. فالمعايير الغربية التقليدية لتقييم نماذج الذكاء الاصطناعي لا تعكس خصوصية اللغة العربية ولا تستجيب لاحتياجات المجتمعات الناطقة بها في العالم العربي.

يهدف هذا الإطار، الذي يُقدّم لأول مرة في هذا المقال، إلى وضع **أعمدة تقييم شاملة** لتقييم نماذج الذكاء الاصطناعي اللغوي (LLMs) وجميع أنظمة الذكاء الاصطناعي التي تستهدف اللغة العربية أو الناطقين بها في العالم العربي. يمتد هذا الإطار من **اللغة إلى المنطق**، ومن **الأسلوب إلى الثقافة**، ومن **النصوص إلى السياق**، ليجيب عن سؤال أساسي: "هل يتحدث النموذج مثلنا؟ هل يفهمنا؟ هل يتفاعل بلغتنا، بمنطقنا، وثقافتنا؟"

**ملاحظة هامة:** يتطلب تطبيق هذا الإطار خبراء لغويين متخصصين في اللغة العربية، يمتلكون معرفة عميقة بالنحو، الصرف، البلاغة، اللهجات، والسياقات الثقافية والدينية العربية. لا يُفترض أن يكون المقيّمون خبراء تقنيين في الذكاء الاصطناعي، بل يجب أن يعتمدوا على خبرتهم اللغوية والثقافية لتقييم الأداء بدقة وموضوعية، دون الاعتماد على نماذج ذكاء اصطناعي أخرى أو النموذج المختبر نفسه. هذا الإطار موجه لتقييم جميع التقنيات الذكية، بما في ذلك النماذج اللغوية الكبيرة، التي تستهدف اللغة العربية أو المجتمعات الناطقة بها.

يُعدّ هذا المقال التمهيد الأول لأعمدة التقييم العربية، وسيتبعه مقال آخر يغطي أسباب اختيار هذه الأعمدة، والجهات المسؤولة عن إعدادها واختبارها، والتزامات مطوري الذكاء الاصطناعي الذين يدعمون اللغة العربية، إلى جانب دور المنطقة العربية في تعزيز هذه الجهود. من خلال هذا الإطار، نسعى إلى وضع معيار عالمي لتقييم الذكاء الاصطناعي العربي يحترم خصوصيات لغتنا وثقافتنا

# أعمدة تقييم نماذج اللغة العربية والذكاء الاصطناعي

جدول الأعمدة المحدث مع الركائز الفرعية والأمثلة

تستهدف هذه الأعمدة تقييم أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، التي تدعم اللغة العربية أو تستهدف الناطقين بها.

الفئة الأولى: المهارات اللغوية أساسية

| الرقم | الركيزة                                         | ما يقيسه                                        | الركائز الفرعية                                                                                                    |
|-------|-------------------------------------------------|-------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| 1     | النحو (Grammar Accuracy)                        | القواعد النحوية (إعراب، تركيب الجملة، المطابقة) | تحديد الوظائف النحوية، الإعراب، أنواع الجمل، المطابقة، تحويل الجمل، التراكيب المعقدة، الجمل الغامضة                |
| 2     | الصرف (Morphology Accuracy)                     | التصريف الصحيح للأفعال والأسماء                 | تصريف الأفعال حسب الزمن والضمير، تصريف الأسماء، صيغ المشتقات، التحويل بين الصيغ، كشف الأخطاء، الأشكال غير المنتظمة |
| 3     | الإملاء والتشكيل (Spelling & Diacritics)        | الكتابة الصحيحة وعلامات التشكيل                 | كتابة الهمزات، الألف اللينة، المدود، علامات الترقيم، التشكيل الكامل، الأخطاء الطباعية                              |
| 4     | الجزور والاشتقاق (Root & Derivation Competence) | استخراج الجذور وتوليد المشتقات                  | استخراج الجذر، تحديد الوزن، توليد المشتقات، التعامل مع الغموض الصرفي، الجذور غير الثلاثية                          |
| 5     | الثراء المعجمي (Lexical Richness)               | تغطية المفردات الشائعة والنادرة                 | تغطية التواتر، التغطية الموضوعية، اللهجات والدخيل، معدل خارج المعجم                                                |

## الفئة الثانية: الفهم السياقي والأسلوبي

| الرقم | الركيزة                                | ما يقيسه                           | الركائز الفرعية                                                                                          |
|-------|----------------------------------------|------------------------------------|----------------------------------------------------------------------------------------------------------|
| 6     | الفهم الدلالي (Semantic Understanding) | فهم المعنى العميق والتعدد الدلالي  | التعدد الدلالي، الفهم السياقي، الفهم الضمني، التراكيب المجازية، التفرقة بين المعاني، التعابير الاصطلاحية |
| 7     | البلاغة والأسلوب (Rhetoric & Style)    | الأساليب البلاغية والتعبير الملائم | التشبيه، الاستعارة، الكناية، تنوع الأسلوب، التعبير العاطفي، التصوير الفني                                |
| 8     | فهم اللهجات (Dialectal Competence)     | فهم وتفسير اللهجات العربية         | التعرف على اللهجة، تفسير المفردات، الاختلافات النحوية، التكيف الأسلوبي                                   |

## الفئة الثالثة: الكفاءة الثقافية والمنطقية

| الرقم | الركيزة                                                              | ما يقيسه                                 | الركائز الفرعية                                                                 |
|-------|----------------------------------------------------------------------|------------------------------------------|---------------------------------------------------------------------------------|
| 9     | الحساسية الثقافية (Cultural Sensitivity)                             | الوعي بالعادات وتجنب الصور النمطية       | الوعي بالتنوع الثقافي، تجنب الصور النمطية، فهم العادات، اللغة المهذبة           |
| 10    | الحساسية الدينية (Religious Sensitivity)                             | الدقة والاحترام في المواضيع الدينية      | الاحترام في الخطاب، الدقة في المعلومات الشرعية، التعامل مع القضايا الفقهية      |
| 11    | التغطية الموضوعية (Domain Coverage)                                  | تغطية مجالات متعددة (ديني، قانوني، علمي) | التنوع المعرفي، الدقة، الملاءمة للسياق العربي، العمق مقابل التبسيط              |
| 12    | التفكير المنطقي (Logical Reasoning)                                  | التحليل والاستنتاج السليم                | الاستنتاج المنطقي، حل المشكلات، التحليل السببي، التمييز بين الحقائق والآراء     |
| 13    | شفافية مصادر المعرفة (Knowledge Source Transparency)                 | توضيح مصدر المعلومة                      | الإفصاح عن المصدر، موثوقية المصدر، نوع المصدر، التعامل مع الغموض                |
| 14    | التعامل مع الوثائق المخصصة العربية (Custom Arabic Document Handling) | قراءة وتحليل وثائق عربية                 | دعم صيغ الملفات، فهم المحتوى، التحليل وإعادة الصياغة، التعامل مع اللغة المتخصصة |
| 15    | التكيف عبر الركائز (Adaptability Across Pillars)                     | الدمج بين الركائز وتغيير الأسلوب         | الدمج بين الركائز، تغيير الأسلوب حسب المجال، استخدام المصادر المتعددة           |

ملاحظة إضافية: يجب أن يقوم التقييم خبراء لغويون متخصصون لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة الأولى: النحو (Grammar Accuracy)

- **الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على تطبيق قواعد النحو العربي بدقة في تحليل الجمل وتكوينها، مع مراعاة خصوصيات اللغة العربية.
- **ما يقيسه:** دقة النحو في الإعراب، تركيب الجمل، المطابقة بين العناصر، والتعامل مع التراكيب المعقدة والغامضة.

### الركائز الفرعية:

- **تحديد الوظائف النحوية:** القدرة على تحديد المبتدأ، الخبر، الفاعل، المفعول به، وغيرها.
- **الإعراب الصحيح:** تطبيق علامات الإعراب (الفتح، الضم، الكسر، السكون) بدقة.
- **أنواع الجمل:** فهم الفرق بين الجمل الاسمية والفعلية والشرطية.
- **المطابقة:** ضمان التوافق في الجنس، العدد، والإعراب بين المكونات.
- **تحويل الجمل:** تحويل الجمل بين الأنماط (مثل الجملة الخبرية إلى طلبية).
- **التعامل مع التراكيب المعقدة:** تحليل الجمل ذات التوابع أو الجمل المركبة.
- **التعامل مع الجمل الغامضة:** القدرة على تحليل الجمل ذات التفسيرات النحوية المتعددة (مثل: "رأى الرجل الفتى").

### معايير الاختبار:

- تحليل جمل متنوعة (بسيطة، مركبة، غامضة) وإعرابها.
- تصحيح الأخطاء النحوية في جمل معطاة.
- تكوين جمل جديدة بناءً على قواعد محددة (مثل تحويل جملة إلى صيغة الشرط).

### ملاحظات التقييم:

- **الدرجة 5:** إعراب دقيق، تحليل صحيح لجميع التراكيب، بما في ذلك الجمل الغامضة.
- **الدرجة 3:** أخطاء طفيفة في الإعراب أو تحليل الجمل المعقدة.
- **الدرجة 1:** أخطاء جوهريّة أو عدم القدرة على تحليل الجمل.
- **ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون متخصصون في النحو العربي لضمان الموضوعية، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة الثانية: الصرف (Morphology Accuracy)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على تصنيف الأفعال والأسماء وتوليد المشتقات بدقة، مع مراعاة التعقيدات الصرفية للغة العربية.

**ما يقيسه:** دقة تصنيف الأفعال والأسماء، التعامل مع الأوزان الصرفية، وتوليد المشتقات المناسبة سياقياً.

### الركائز الفرعية:

1. **تصنيف الأفعال حسب الزمن والضمير:** تصنيف الأفعال في الأزمنة المختلفة (ماضي، مضارع، أمر) مع الضمائر.
2. **تصنيف الأسماء:** تكوين المفرد، المثنى، الجمع (مذكر/مؤنث، سالم/مكسر).
3. **صيغ المشتقات:** توليد أسماء الفاعل، المفعول، المكان، الزمان، والآلة.
4. **التحويل بين الصيغ:** تحويل الأفعال بين المبني للمعلوم والمجهول أو بين الأوزان.
5. **كشف الأخطاء الصرفية:** تحديد وتصحيح الأخطاء في التصريف.
6. **التعامل مع الأشكال غير المنتظمة:** تصنيف الأفعال المهموزة، المضعفة، والناقصة، ومعالجة الأسماء غير القياسية.

### معايير الاختبار:

- تصنيف أفعال وأسماء في سياقات متنوعة (مثل: "كتب" إلى مضارع، جمع "كتاب").
- توليد مشتقات بناءً على جذر معين (مثل: "علم" إلى "معلم"، "عالم").
- تصحيح جمل تحتوي على أخطاء صرفية.
- معالجة أفعال غير منتظمة (مثل: "أكل" إلى مضارع ناقص).

### ملاحظات التقييم:

- **الدرجة 5:** تصنيف دقيق، مشتقات صحيحة، ومعالجة الأشكال غير المنتظمة بنجاح.
  - **الدرجة 3:** أخطاء طفيفة في التصريف أو المشتقات.
  - **الدرجة 1:** أخطاء جوهريّة أو عدم القدرة على التعامل مع الأوزان الصرفية.
- ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون متخصصون في الصرف العربي لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة الثالثة: الإملاء والتشكيل (Spelling & Diacritics)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على كتابة النصوص العربية بدقة إملائية واستخدام صحيح لعلامات التشكيل والترقيم.

**ما يقيسه:** دقة الكتابة الإملائية، تطبيق علامات التشكيل (الفتحة، الضمة، الكسرة، السكون، الشدة)، ومعالجة الأخطاء الشائعة.

**الركائز الفرعية:**

1. كتابة الهمزات: وضع الهمزة (ابتداء، وسط، نهاية) بشكل صحيح.
2. الألف اللينة والمدود: التمييز بين الألف اللينة (ي/ى) والمدود (أ، إ).
3. علامات الترقيم: استخدام الفواصل، النقاط، وعلامات الاستفهام والتعجب بشكل صحيح.
4. التشكيل الكامل: إضافة علامات التشكيل بدقة في النصوص غير المشكّلة.
5. تصحيح الأخطاء الإملائية: تحديد وتصحيح الأخطاء الشائعة (مثل: "كتبت" بدلاً من "كتب").
6. معالجة الأخطاء الطباعية: التعرف على الأخطاء الناتجة عن الإدخال الخاطئ (مثل: "كتاب" بدلاً من "كتلب").

**معايير الاختبار:**

- كتابة نصوص مع تشكيل كامل ودقيق.
- تصحيح نصوص تحتوي على أخطاء إملائية أو تشكيلية.
- تحليل نصوص بها أخطاء طباعية وإصلاحها.

**ملاحظات التقييم:**

- **الدرجة 5:** كتابة خالية من الأخطاء، تشكيل دقيق، ومعالجة الأخطاء الطباعية بنجاح.
- **الدرجة 3:** أخطاء طفيفة في التشكيل أو الإملاء.
- **الدرجة 1:** أخطاء جوهريّة أو عدم القدرة على تطبيق التشكيل.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون متخصصون في الإملاء والتشكيل العربي لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة الرابعة: الجذور والاشتقاق (Root & Derivation Competence)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على استخراج الجذور العربية وتوليد المشتقات بدقة وفق قواعد الاشتقاق.

**ما يقيسه:** دقة استخراج الجذور، تحديد الأوزان الصرفية، وتوليد المشتقات مع التعامل مع الجذور غير القياسية.

### الركائز الفرعية:

1. استخراج الجذر: تحديد الجذر من الكلمات (مثل: "معلم" إلى "علم").
2. تحديد الوزن: التعرف على الأوزان الصرفية (مثل: "فاعل"، "مفعول").
3. توليد المشتقات: إنتاج مشتقات صحيحة (مثل: "علم" إلى "عالم"، "معلم").
4. التعامل مع الغموض الصرفي: حل الغموض في الكلمات ذات الجذور المتشابهة.
5. الجذور غير القياسية: معالجة الجذور الرباعية أو غير الثلاثية (مثل: "ترجم" إلى "مترجم").

### معايير الاختبار:

- استخراج الجذر من كلمات معقدة أو مشتقة.
- توليد مشتقات بناءً على جذر أو وزن محدد.
- تحليل كلمات تحتوي على جذور غير قياسية أو غامضة.

### ملاحظات التقييم:

- **الدرجة 5:** استخراج وتوليد دقيق للجذور والمشتقات، بما في ذلك الجذور غير القياسية.
- **الدرجة 3:** أخطاء طفيفة في الاشتقاق أو معالجة الغموض.
- **الدرجة 1:** عدم القدرة على استخراج الجذور أو توليد المشتقات.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويين متخصصون في الصرف والاشتقاق العربي لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة الخامسة: الثراء المعجمي (Lexical Richness)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على تغطية مفردات اللغة العربية الشائعة والنادرة والمتخصصة، مع مراعاة التنوع اللغوي.

**ما يقيسه:** مدى اتساع المفردات، التعامل مع المصطلحات الموضوعية، والمفردات الدخيلة، واللهجات.

### الركائز الفرعية:

1. **تغطية التواتر:** استخدام المفردات الشائعة في السياقات اليومية.
2. **التغطية الموضوعية:** فهم واستخدام مصطلحات متخصصة (مثل: طبية، قانونية).
3. **اللهجات والدخيل:** التعامل مع المفردات اللهجية والكلمات المستعارة.
4. **معدل خارج المعجم:** القدرة على التعامل مع كلمات نادرة أو غير مألوفة.
5. **الكلمات المستعارة:** فهم واستخدام الكلمات الدخيلة بشكل صحيح (مثل: "تلفزيون").

### معايير الاختبار:

- إنتاج نصوص تحتوي على مفردات متنوعة (شائعة، نادرة، متخصصة).
- تفسير مصطلحات لهجية أو دخيلة في سياقات محددة.
- إعادة صياغة نصوص باستخدام مفردات بديلة.

### ملاحظات التقييم:

- **الدرجة 5:** تغطية شاملة للمفردات، استخدام دقيق للمصطلحات والكلمات الدخيلة.
- **الدرجة 3:** أخطاء طفيفة في اختيار المفردات أو تفسيرها.
- **الدرجة 1:** تغطية محدودة أو أخطاء جوهريّة في المفردات.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون متخصصون في المعجم العربي لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة السادسة: الفهم الدلالي (Semantic Understanding)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على استيعاب المعاني العميقة للنصوص العربية، مع مراعاة التعدد الدلالي والسياق.

**ما يقيسه:** فهم المعاني الحرفية والمجازية، التفرقة بين المعاني المختلفة، والتعامل مع التعبيرات الاصطلاحية والغموض السياقي.

### الركائز الفرعية:

1. **التعدد الدلالي:** تحديد المعاني المتعددة للكلمات بناءً على السياق (مثل: "عين" بمعنى العين أو الماء).
2. **الفهم السياقي:** استيعاب المعنى بناءً على سياق النص أو الجملة.
3. **الفهم الضمني:** استنباط المعاني غير المباشرة (مثل: السخرية، التلميح).
4. **التراكيب المجازية:** تفسير الاستعارات والكنائيات (مثل: "قلبه من حجر").
5. **التفرقة بين المعاني:** التمييز بين المعاني المتقاربة (مثل: "نظر" و"رأى").
6. **التعابير الاصطلاحية:** فهم واستخدام التعبيرات الشائعة (مثل: "ضرب عصفورين بحجر").
7. **الغموض السياقي:** معالجة النصوص التي تحتل تفسيرات متعددة بناءً على السياق.

### معايير الاختبار:

- تحديد المعنى الصحيح لكلمات متعددة المعاني في سياقات مختلفة.
- تفسير نصوص مجازية أو تحتوي على تعابير اصطلاحية.
- إعادة صياغة نصوص غامضة مع توضيح المعنى المقصود.

### ملاحظات التقييم:

- **الدرجة 5:** فهم دقيق للمعاني الحرفية، المجازية، والاصطلاحية مع معالجة الغموض.
- **الدرجة 3:** أخطاء طفيفة في تفسير السياق أو التعبيرات.
- **الدرجة 1:** عدم القدرة على فهم المعاني أو التمييز بينها.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون متخصصون في الدلالات العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة السابعة: البلاغة والأسلوب (Rhetoric & Style)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على استخدام الأساليب البلاغية والتعبير بأسلوب ملائم للسياق في النصوص العربية.

**ما يقيسه:** إنتاج وتحليل النصوص البلاغية باستخدام التشبيه، الاستعارة، الكناية، والأساليب المناسبة للجمهور.

### الركائز الفرعية:

1. **التشبيه:** استخدام وتحليل التشبيهات (مثل: "وجهه كالقمر").
2. **الاستعارة:** إنتاج وتفسير الاستعارات (مثل: "البحر من الدموع").
3. **الكناية:** فهم واستخدام الكنايات (مثل: "طال لسانه" بمعنى الكلام الكثير).
4. **تنويع الأسلوب:** تغيير الأسلوب حسب السياق (رسمي، غير رسمي، أدبي).
5. **التعبير العاطفي:** إنتاج نصوص تعبر عن العواطف بدقة.
6. **التصوير الفني:** خلق صور بصرية أو حسية في النصوص.
7. **التكيف الأسلوبي مع الجمهور:** تعديل الأسلوب حسب الجمهور (مثل: أكاديمي، عام، أطفال).

### معايير الاختبار:

- إنتاج نصوص تحتوي على تشبيهات، استعارات، أو كنايات.
- تحليل نصوص بلاغية وتحديد الأساليب المستخدمة.
- إعادة صياغة نصوص بأسلوب يناسب جمهورًا محددًا.

### ملاحظات التقييم:

- **الدرجة 5:** استخدام بلاغي دقيق وأسلوب ملائم للسياق والجمهور.
- **الدرجة 3:** أخطاء طفيفة في الأساليب البلاغية أو التكيف مع الجمهور.
- **الدرجة 1:** عدم القدرة على إنتاج أو تحليل نصوص بلاغية.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون متخصصون في البلاغة العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

# الركيزة الثامنة: فهم اللهجات (Dialectal Competence)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على فهم وتفسير اللهجات العربية المختلفة والتفاعل بها بشكل طبيعي.

**ما يقيسه:** التعرف على اللهجات، تفسير المفردات والتراكيب اللهجية، والتكيف مع الاختلافات الأسلوبية والنحوية.

## الركائز الفرعية:

1. التعرف على اللهجة: تحديد اللهجة المستخدمة (مثل: مصرية، خليجية، مغربية).
2. تفسير المفردات اللهجية: فهم معاني الكلمات الخاصة باللهجات (مثل: "زين" بمعنى جيد في الخليجية).
3. الاختلافات النحوية: التعامل مع التراكيب النحوية الخاصة باللهجات (مثل: "برجع" بدلاً من "سأعود").
4. التكيف الأسلوبي: إنتاج نصوص أو ردود تتماشى مع أسلوب اللهجة المحددة.
5. التعامل مع التبدل اللغوي: فهم وإنتاج نصوص تخلط بين الفصحى واللهجة أو لهجات متعددة.

## معايير الاختبار:

- تحديد اللهجة في نص أو محادثة معطاة.
- تفسير جمل لهجية وإعادة صياغتها بالفصحى أو لهجة أخرى.
- إنتاج ردود بلهجة محددة بناءً على السياق.

## ملاحظات التقييم:

- الدرجة 5: تعرف دقيق على اللهجات، تفسير صحيح، وتكيف أسلوبي مناسب.
- الدرجة 3: أخطاء طفيفة في التعرف أو التفسير.
- الدرجة 1: عدم القدرة على فهم أو إنتاج نصوص لهجية.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون متخصصون في اللهجات العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة التاسعة: الحساسية الثقافية (Cultural Sensitivity)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على التفاعل مع النصوص والسياقات العربية باحترام للتنوع الثقافي وتجنب الصور النمطية.

**ما يقيسه:** فهم العادات والتقاليد العربية، استخدام لغة محترمة، والتعامل مع الاختلافات الثقافية الإقليمية.

### الركائز الفرعية:

1. الوعي بالتنوع الثقافي: فهم الاختلافات بين الثقافات العربية (مثل: الخليجية، المغاربية).
2. تجنب الصور النمطية: عدم إنتاج محتوى يعزز التحيزات أو الصور النمطية.
3. فهم العادات: التعرف على العادات والتقاليد المحلية (مثل: آداب الضيافة).
4. اللغة المهذبة: استخدام تعابير محترمة ومناسبة ثقافيًا.
5. الاختلافات الثقافية الإقليمية: التكيف مع الفروق الثقافية بين المناطق العربية (مثل: اللباس التقليدي).

### معايير الاختبار:

- إنتاج ردود تحترم السياقات الثقافية المتنوعة.
- تحليل نصوص لتحديد الصور النمطية أو التعابير غير المناسبة.
- الرد على أسئلة تتطلب معرفة بالعادات الإقليمية.

### ملاحظات التقييم:

- **الدرجة 5:** تفاعل محترم مع التنوع الثقافي وتجنب كامل للصور النمطية.
- **الدرجة 3:** أخطاء طفيفة في التكيف الثقافي أو اللغة.
- **الدرجة 1:** عدم الوعي الثقافي أو إنتاج محتوى غير لائق.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون وثقافيون متخصصون في الثقافة العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة العاشرة: الحساسية الدينية (Religious Sensitivity)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على التعامل مع المواضيع الدينية في اللغة العربية باحترام ودقة.

**ما يقيسه:** الدقة في المعلومات الدينية، الاحترام في الخطاب، والتكيف مع السياقات الدينية المتنوعة.

### الركائز الفرعية:

1. الاحترام في الخطاب: استخدام لغة محترمة عند مناقشة المواضيع الدينية.
2. الدقة في المعلومات الشرعية: تقديم معلومات دينية صحيحة (مثل: أحكام فقهية).
3. التعامل مع القضايا الفقهية: الإجابة عن أسئلة دينية مع مراعاة الفروق المذهبية.
4. التكيف مع السياقات الدينية المتعددة: التعامل مع النصوص التي تشمل سياقات إسلامية وغير إسلامية.

### معايير الاختبار:

- الإجابة عن أسئلة دينية بمعلومات دقيقة ومحترمة.
- تحليل نصوص دينية لتحديد مدى الاحترام والدقة.
- الرد على استفسارات تتعلق بالفروق المذهبية أو السياقات متعددة الأديان.

### ملاحظات التقييم:

- **الدرجة 5:** خطاب محترم، معلومات دقيقة، وتكيف مناسب مع السياقات الدينية.
- **الدرجة 3:** أخطاء طفيفة في الدقة أو التكيف الديني.
- **الدرجة 1:** معلومات غير دقيقة أو خطاب غير لائق.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون ودينيون متخصصون في الثقافة والشريعة العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

# الركيزة الحادية عشرة: التغطية الموضوعية (Domain Coverage)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على تغطية مجالات معرفية متنوعة باللغة العربية بدقة وملاءمة.

**ما يقيسه:** التنوع في المجالات (ديني، قانوني، علمي، أدبي)، الدقة في المصطلحات، والتكيف مع السياقات العربية.

## الركائز الفرعية:

1. **التنوع المعرفي:** تغطية مجالات متعددة (مثل: الفقه، القانون، العلوم).
2. **الدقة في المصطلحات:** استخدام المصطلحات المتخصصة بشكل صحيح.
3. **الملاءمة للسياق العربي:** تقديم محتوى يتماشى مع الثقافة والاحتياجات العربية.
4. **العمق مقابل التبسيط:** التوازن بين المعلومات التفصيلية والمبسطة حسب السياق.
5. **التعامل مع المواضيع متعددة التخصصات:** الجمع بين مجالات مختلفة (مثل: أخلاقيات التكنولوجيا).

## معايير الاختبار:

- الإجابة عن أسئلة في مجالات متنوعة (مثل: شرح مصطلح قانوني، مفهوم علمي).
- إنتاج نصوص متخصصة بمصطلحات دقيقة.
- معالجة أسئلة تجمع بين مجالات متعددة.

## ملاحظات التقييم:

- **الدرجة 5:** تغطية شاملة، مصطلحات دقيقة، وتكيف مناسب للسياقات.
- **الدرجة 3:** أخطاء طفيفة في الدقة أو التنوع.
- **الدرجة 1:** تغطية محدودة أو مصطلحات غير صحيحة.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون ومتخصصون في المجالات المعرفية العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## الركيزة الثانية عشرة: التفكير المنطقي (Logical Reasoning)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على التحليل والاستنتاج المنطقي في النصوص والأسئلة باللغة العربية.

**ما يقيسه:** القدرة على الاستنتاج المنطقي، حل المشكلات، والتمييز بين الحقائق والآراء، مع معالجة المغالطات المنطقية.

### الركائز الفرعية:

1. **الاستنتاج المنطقي:** اشتقاق استنتاجات صحيحة من مقدمات معطاة.
2. **حل المشكلات:** معالجة الألغاز أو المسائل المنطقية (مثل: مسائل الترتيب).
3. **التحليل السببي:** تحديد العلاقات بين الأسباب والنتائج.
4. **التمييز بين الحقائق والآراء:** التفرقة بين المعلومات الموضوعية والذاتية.
5. **معالجة المغالطات المنطقية:** التعرف على المغالطات (مثل: المغالطة السببية) وتصحيحها.

### معايير الاختبار:

- حل ألغاز منطقية أو مسائل رياضية بسيطة باللغة العربية.
- تحليل نصوص تحتوي على مغالطات منطقية وتصحيحها.
- الإجابة عن أسئلة تتطلب التمييز بين الحقائق والآراء.

### ملاحظات التقييم:

- **الدرجة 5:** استنتاجات دقيقة، حلول صحيحة، ومعالجة المغالطات بنجاح.
- **الدرجة 3:** أخطاء طفيفة في الاستنتاج أو تحليل المغالطات.
- **الدرجة 1:** عدم القدرة على الاستنتاج أو التعرف على المغالطات.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون ومتخصصون في التفكير المنطقي باللغة العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

# الركيزة الثالثة عشرة: شفافية مصادر المعرفة (Knowledge Source Transparency)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على توضيح مصادر المعلومات المقدمة باللغة العربية ومدى موثوقيتها.

**ما يقيسه:** الإفصاح عن مصادر المعلومات، تقييم موثوقيتها، والتعامل مع المعلومات الغامضة أو غير القابلة للتحقق.

## الركائز الفرعية:

1. الإفصاح عن المصدر: توضيح مصدر المعلومة (مثل: كتاب، مقال، قاعدة بيانات).
2. موثوقية المصدر: تقييم مدى مصداقية المصادر المستخدمة.
3. نوع المصدر: تحديد نوع المصدر (علمي، شعبي، ديني).
4. التعامل مع الغموض: معالجة المعلومات غير الواضحة المصدر.
5. التعامل مع المصادر غير القابلة للتحقق: التعامل مع المعلومات التي لا يمكن التحقق من مصدرها.

## معايير الاختبار:

- تقديم معلومات مع ذكر المصدر (مثل: "بحسب الموسوعة الفلانية").
- تقييم موثوقية مصدر معطى في سياق عربي.
- الرد على أسئلة تحتوي على معلومات غامضة أو غير موثقة.

## ملاحظات التقييم:

- **الدرجة 5:** إفصاح كامل عن المصادر، موثوقية عالية، ومعالجة الغموض بنجاح.
  - **الدرجة 3:** أخطاء طفيفة في الإفصاح أو تقييم الموثوقية.
  - **الدرجة 1:** عدم الإفصاح عن المصادر أو الاعتماد على مصادر غير موثوقة.
- ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون ومتخصصون في تقييم المصادر باللغة العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

# الركيزة الرابعة عشرة: التعامل مع الوثائق المخصصة العربية (Custom Arabic Document Handling)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على قراءة، تحليل، ومعالجة الوثائق العربية بمختلف الصيغ والسياقات.

**ما يقيسه:** دعم صيغ الوثائق، فهم المحتوى، إعادة الصياغة، والتعامل مع النصوص متعددة اللغات.

## الركائز الفرعية:

1. **دعم صيغ الملفات:** التعامل مع صيغ مثل PDF ، Word ، وصور النصوص العربية.
2. **فهم المحتوى:** استخراج المعلومات الرئيسية من الوثائق (مثل: عقود، تقارير).
3. **التحليل وإعادة الصياغة:** تلخيص أو إعادة صياغة الوثائق بدقة.
4. **التعامل مع اللغة المتخصصة:** فهم المصطلحات التقنية أو القانونية.
5. **الوثائق متعددة اللغات:** معالجة الوثائق التي تخط العربية بلغات أخرى (مثل: العربية والإنجليزية).

## معايير الاختبار:

- استخراج معلومات من وثيقة عربية بصيغة معقدة.
- تلخيص أو إعادة صياغة وثيقة قانونية أو علمية.
- تحليل وثيقة تحتوي على نصوص عربية وإنجليزية.

## ملاحظات التقييم:

- **الدرجة 5:** معالجة دقيقة للوثائق بجميع الصيغ واللغات مع تحليل صحيح.
- **الدرجة 3:** أخطاء طفيفة في الفهم أو الصياغة.
- **الدرجة 1:** عدم القدرة على معالجة الوثائق أو فهم المحتوى.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويين ومتخصصون في تحليل الوثائق باللغة العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

# الركيزة الخامسة عشرة: التكيف عبر الركائز (Adaptability Across Pillars)

**الهدف:** تقييم قدرة أنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، على دمج وتكييف الركائز المختلفة (لغوية، ثقافية، منطقية) في النصوص العربية حسب السياق.

**ما يقيسه:** القدرة على الجمع بين الركائز، تغيير الأسلوب حسب المجال، ومعالجة التعارضات بين الركائز.

## الركائز الفرعية:

1. **الدمج بين الركائز:** الجمع بين النحو، البلاغة، والثقافة في نص واحد.
2. **تغيير الأسلوب حسب المجال:** التكيف بين السياقات (مثل: علمي، ديني، أدبي).
3. **استخدام المصادر المتعددة:** الاعتماد على مصادر متنوعة لإنتاج محتوى متكامل.
4. **معالجة التعارضات بين الركائز:** حل التناقضات (مثل: أسلوب لهجي في سياق رسمي).

## معايير الاختبار:

- إنتاج نص يجمع بين النحو، البلاغة، والحساسية الثقافية.
- تحليل نصوص تتطلب التكيف بين مجالات مختلفة.
- معالجة سيناريوهات تحتوي على تعارض بين الركائز.

## ملاحظات التقييم:

- **الدرجة 5:** دمج سلس للركائز، تكيف دقيق، ومعالجة التعارضات بنجاح.
- **الدرجة 3:** أخطاء طفيفة في الدمج أو التكيف.
- **الدرجة 1:** عدم القدرة على دمج الركائز أو معالجة التعارضات.

**ملاحظة إضافية:** يجب أن يقوم التقييم خبراء لغويون ومتخصصون في التكامل اللغوي والثقافي باللغة العربية لضمان الدقة، دون الاعتماد على أنظمة ذكاء اصطناعي أخرى.

## خطوات بناء المؤشر النهائي

**الهدف:** حساب مؤشر تقييم شامل لأنظمة الذكاء الاصطناعي، بما في ذلك النماذج اللغوية الكبيرة، بناءً على الأداء في الـ15 ركيزة، مع مراعاة الأوزان والدرجات.

### الخطوات:

1. تحديد أوزان الركائز: تخصيص أوزان لكل ركيزة بناءً على الأولوية (المجموع 100%).
2. نظام الدرجات: درجة من 0 إلى 5 لكل ركيزة بناءً على الأداء (5: ممتاز، 3: جيد، 1: ضعيف).
3. حساب الدرجة: (متوسط درجات الركيزة) × (وزنها / 100).
4. المؤشر النهائي: مجموع درجات الركائز الموزونة.
5. تصنيف النظام: 90-100: ممتاز، 80-89: جيد جدًا، 70-79: جيد، 60-69: متوسط، <60: ضعيف.

### جدول الأوزان المقترح (مرن):

| الوزن % | الركيزة                    |
|---------|----------------------------|
| 10      | النحو                      |
| 8       | الصرف                      |
| 7       | الإملاء والتشكيل           |
| 7       | الجنور والاشتقاق           |
| 7       | الثراء المعجمي             |
| 8       | الفهم الدلالي              |
| 6       | البلاغة والأسلوب           |
| 5       | فهم اللهجات                |
| 5       | الحساسية الثقافية          |
| 7       | الحساسية الدينية           |
| 8       | التغطية الموضوعية          |
| 7       | التفكير المنطقي            |
| 6       | شفافية مصادر المعرفة       |
| 8       | التعامل مع الوثائق المخصصة |
| 5       | التكيف عبر الركائز         |

ملاحظة إضافية: يجب أن يقوم التقييم خبراء لغويون لضمان الدقة، مع إمكانية تعديل الأوزان حسب السياق.

