

اختبار أكاكوس العربي

الإصدار 1.1، 15 أكتوبر 2026

الإصدار 1.1: أُضيفت 6 نماذج في 15 أكتوبر 2026

الملخص

اختبار صغير يثنتي عشرة مهمة بالعربية، أجريناه مرة واحدة على 21 نموذجًا بتاريخ 15 أكتوبر 2026. نعرض الدرجة مهمة مهمة، ولا نرتب النماذج في قائمة واحدة. حُكم على الإجابات آليًا، ولم يراجعها شخص بعد.

الإجابات كاملة، والطريقة، والبيانات المفتوحة على الصفحة:

<https://acacus.ly/research/arabic-test-2026-10>

كيف أجرينا الاختبار

- كتبنا بأنفسنا اثنتي عشرة مهمة: كتابة رسمية بالفصحى، وفهم سؤال بالدارجة الليبية، وترجمة في الاتجاهين، وتلخيص، وشرح درس، وسيرة ذاتية، ومسألة حسابية، وكود بتعليقات عربية، والتزام بشروط الشكل، ونص يختلط فيه العربي بالإنجليزي والأرقام.
- أرسلنا كل مهمة إلى كل نموذج مرة واحدة عبر واجهة أكاكوس البرمجية، كما يفعل أي عميل. بلا تعليمات خفية سوى جملة محايدة واحدة، وبإعدادات كل نموذج الافتراضية، من حرارة وتفكير. الحد الأقصى للجواب 8000 رمز لكل النموذج.
- قيّمنا كل إجابة بطريقتين. الأولى فحوص آلية موضوعية: عدّ النقاط والكلمات، ووجود نص معيّن حرفيًا، وتشغيل الكود فعليًا على حالات اختبار. والثانية محكّمان آليان، وهما نموذجان مختلفان، يقرأ كل منهما المهمة ومعايير التقييم وإجابة واحدة مجهولة المصدر وبترتيب عشوائي، ثم يضع درجة من 1 إلى 5.
- الدرجة في الجدول هي متوسط درجتَي المحكّمين، أو الدرجة التي صَحّحها القارئ البشري إن غيّرهما. إن اختلف المحكّمان بأكثر من درجة كتبنا ذلك بجانب الإجابة. وهذه الإجابات لم يراجعها شخص بعد. المحكّمان هما DeepSeek V4 Pro وXiaomi MiMo V2.6 Pro وDeepSeek V4 Pro من النماذج المختبرة أيضًا، فحكّم على إجاباته هو في الجولة الأولى، وهذا انحياز محتمل نذكره لك، والجولة الثانية بنموذج من شركة أخرى.

معنى الدرجات

- 5: ممتاز بلا ملاحظات تُذكر.
- 4: جيد مع ملاحظة صغيرة.
- 3: مقبول مع أخطاء ظاهرة.
- 2: ضعيف، يفي ببعض المطلوب.
- 1: لا يفي بالمطلوب، أو خاطئ، أو بلا جواب.

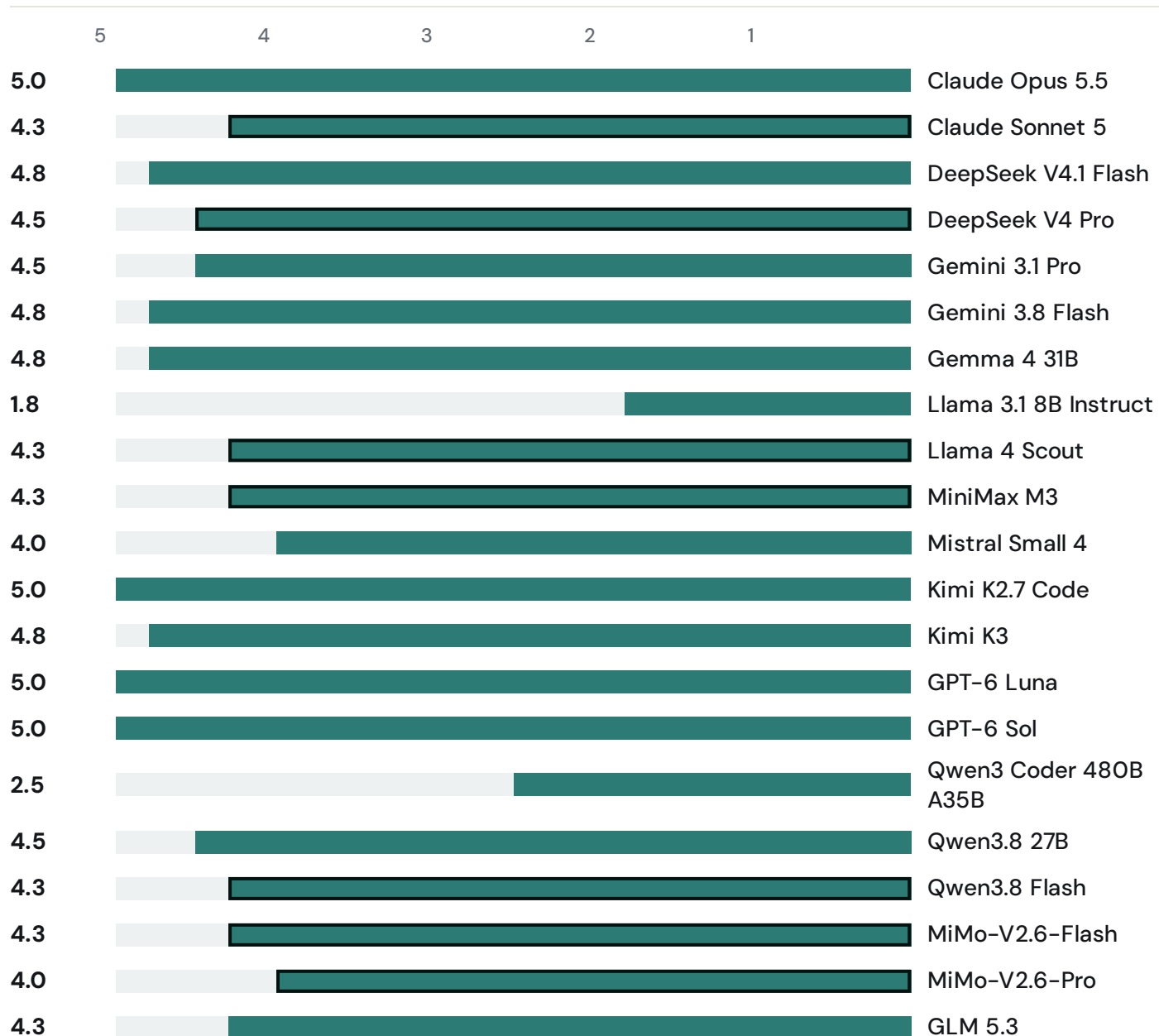
حدود هذا الاختبار: اختبار صغير، بمهام قليلة، ومرة واحدة لكل نموذج، في يوم واحد. النماذج وطرق تشغيلها تتغير، وقد يخطئ المحكم الآلي أو يميل إلى أسلوب دون آخر. لذلك لا نرتّب النماذج في قائمة واحدة ولا نُعلن نموذجًا متفوقًا على الجميع؛ انظر إلى المهمة التي تشبه مهمتك، وجرب النموذج على عملك الحقيقي.

الدرجات في رسوم

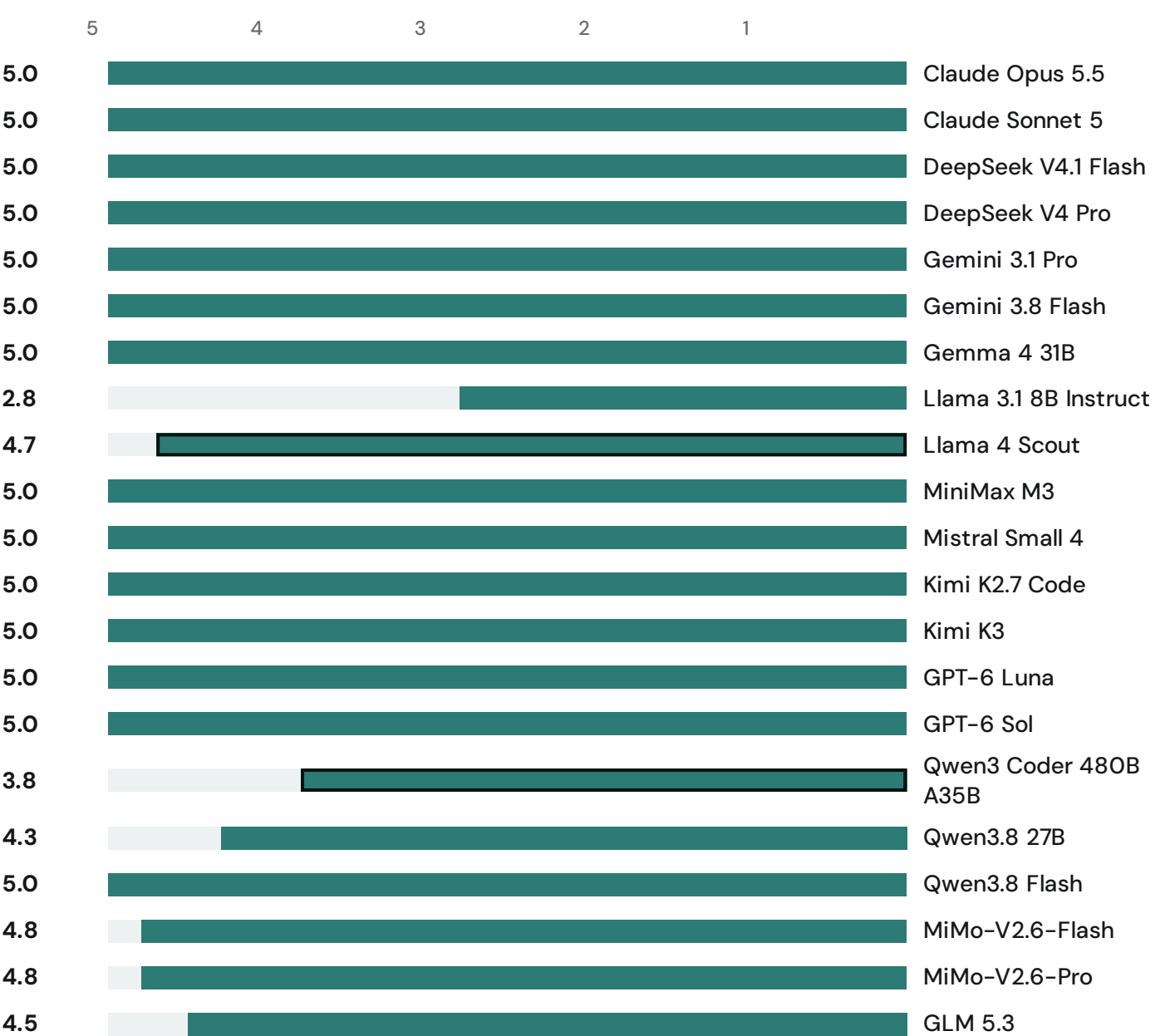
الرسوم نفسها أرقام الجداول أدناه. النماذج مرتبة أبعديًا لا بحسب الدرجة، ولا نجمع الدرجات في رقم واحد لكل نموذج. هي نتيجة اختبار صغير، مرة واحدة لكل نموذج.

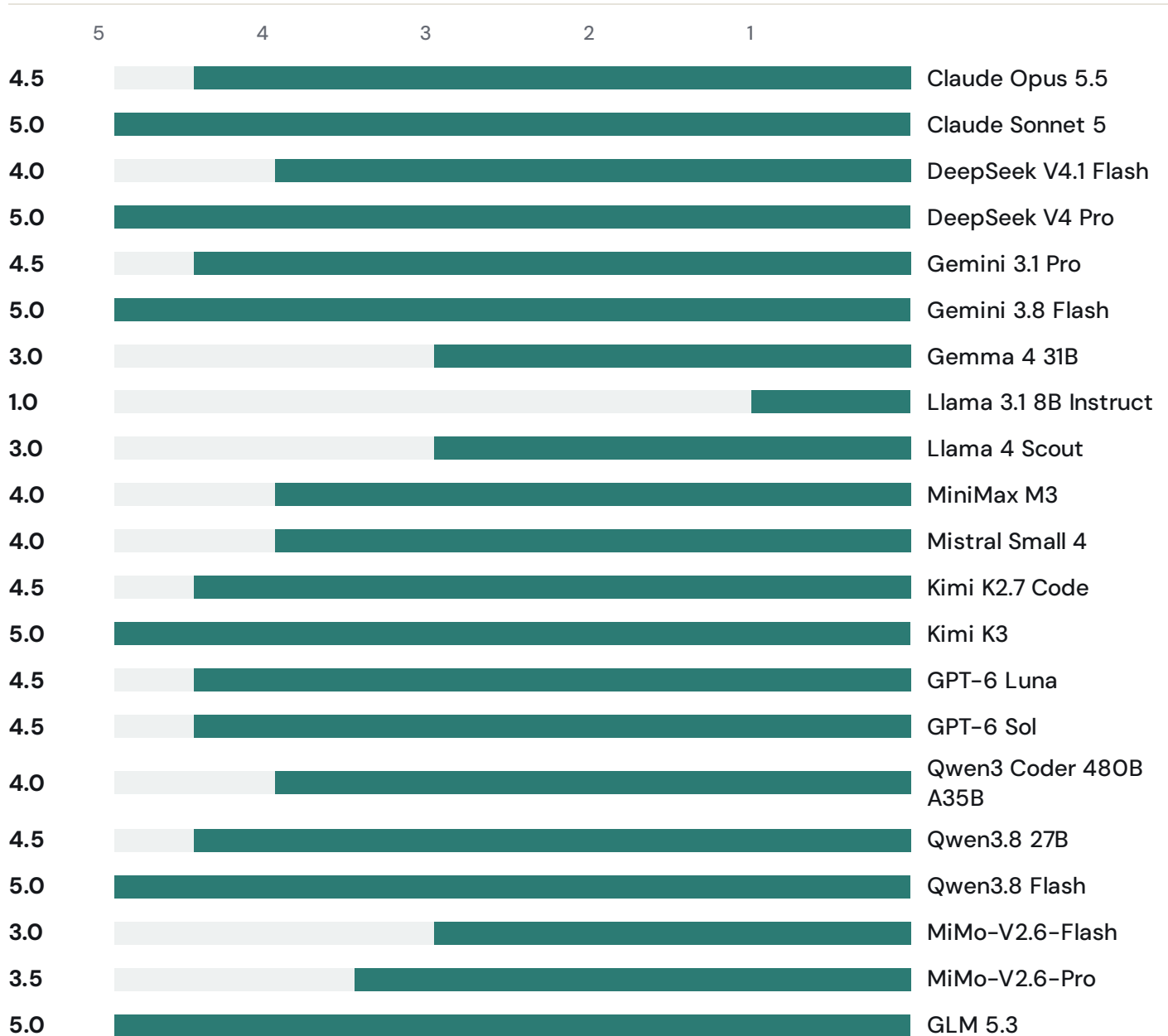
متوسط الدرجة في كل نوع من المهام

كل شريط متوسط درجات النموذج في مهام ذلك النوع، من 1 إلى 5. الشريط المخطط فيه إجابة اختلف عليها المحكمان بأكثر من درجة.

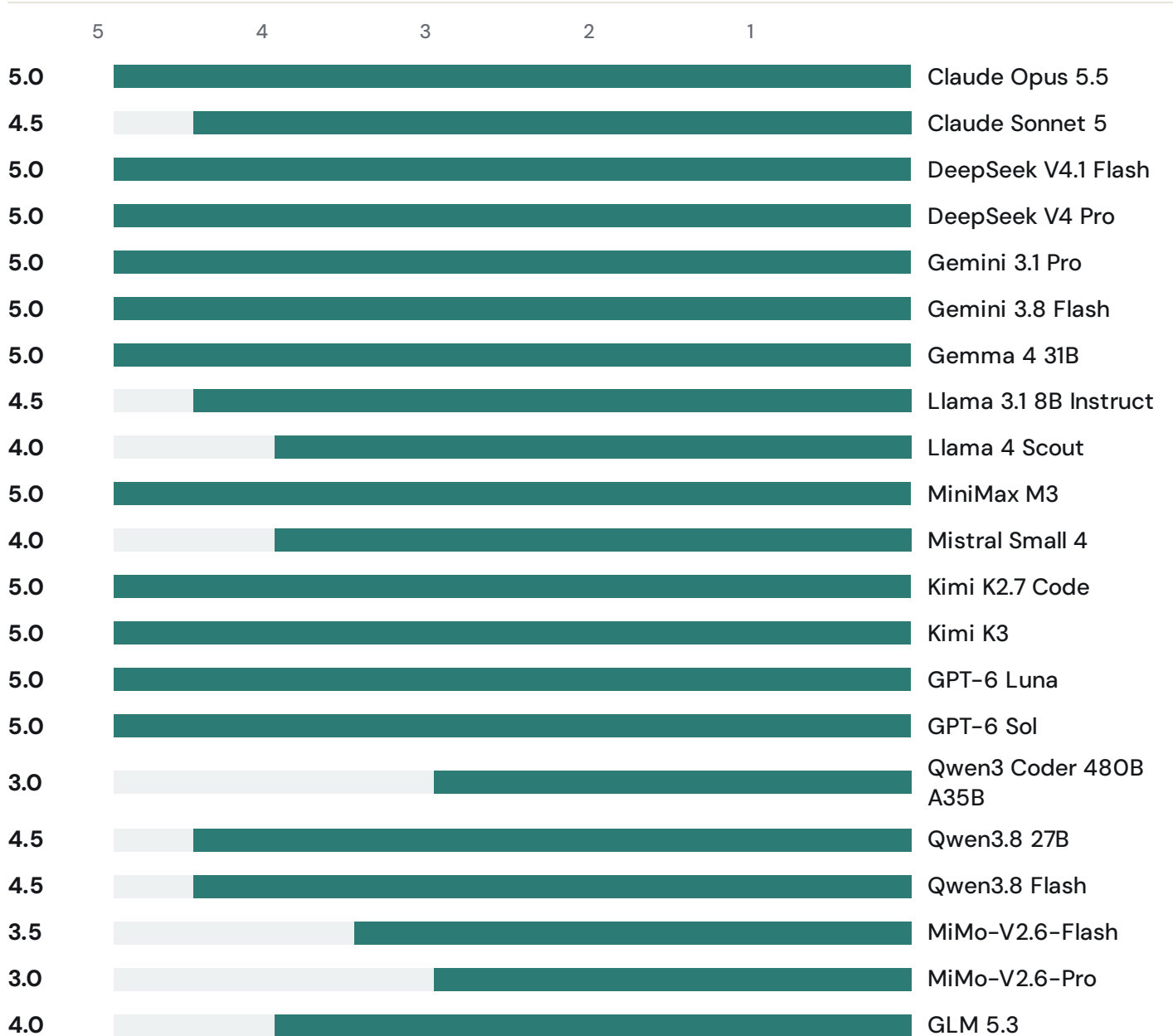


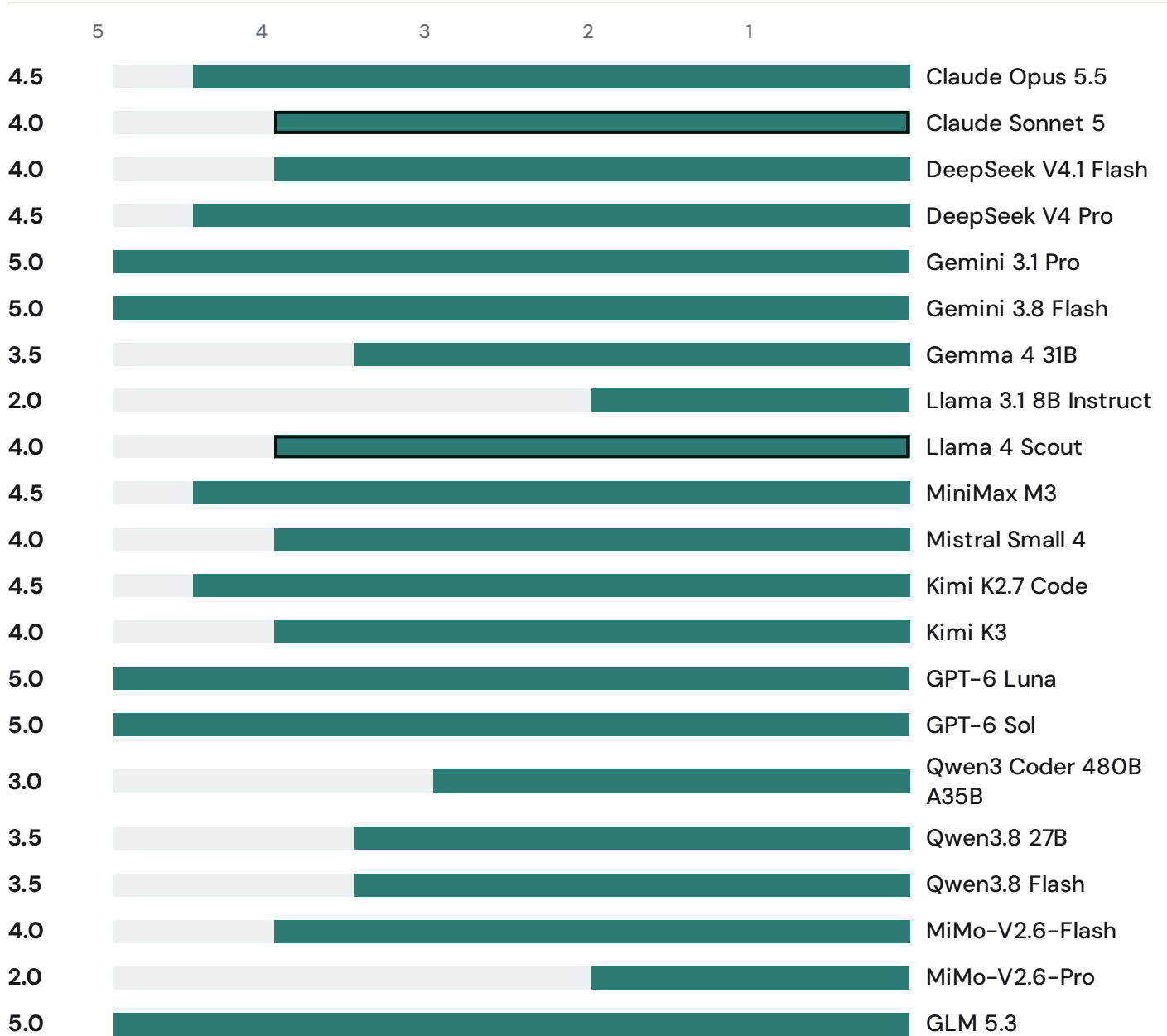


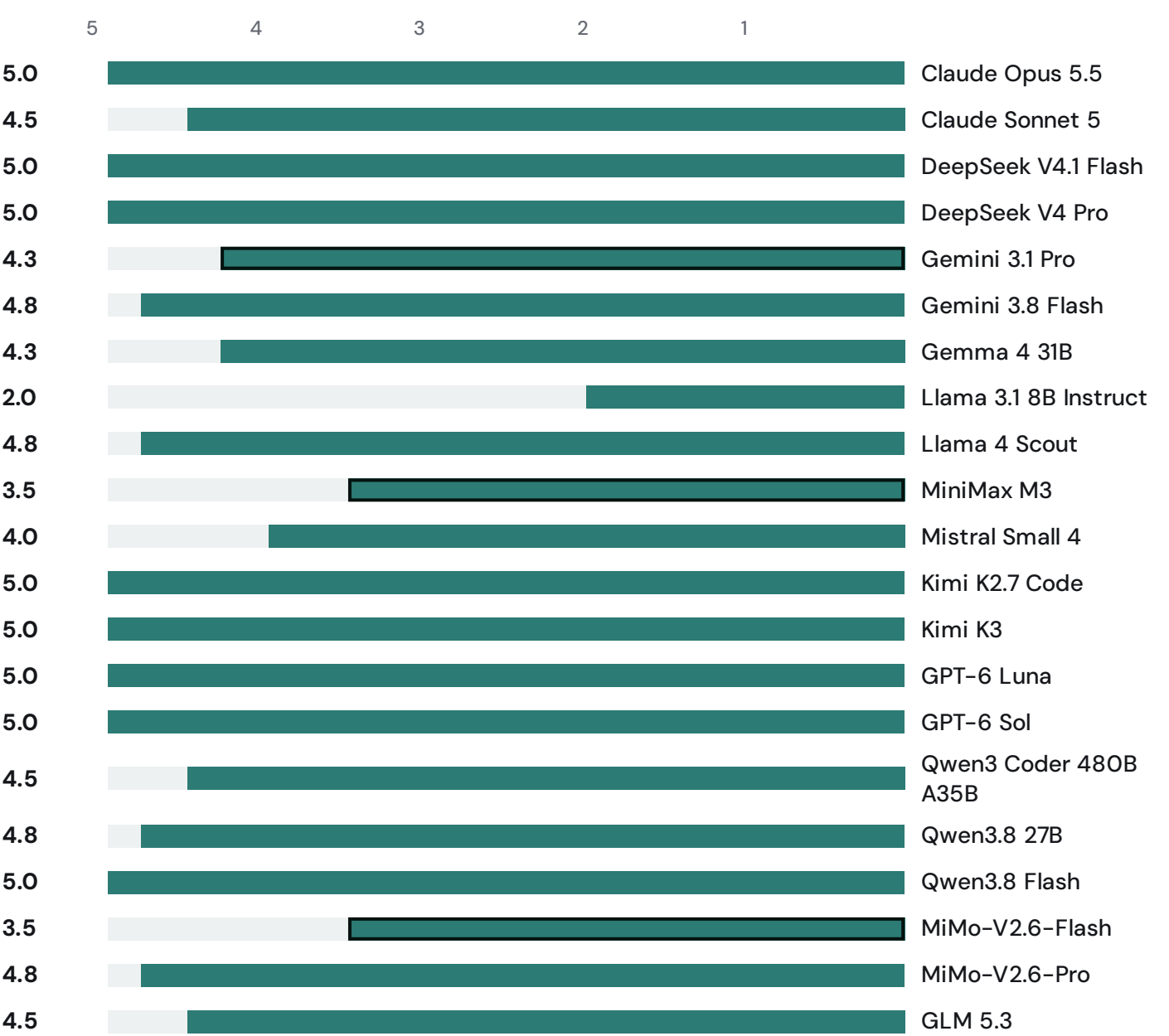












كل نموذج في كل مهمة

كل خلية درجة نموذج في مهمة واحدة. والأعمق أعلى. على الهاتف مرّر الجدول جانبياً.

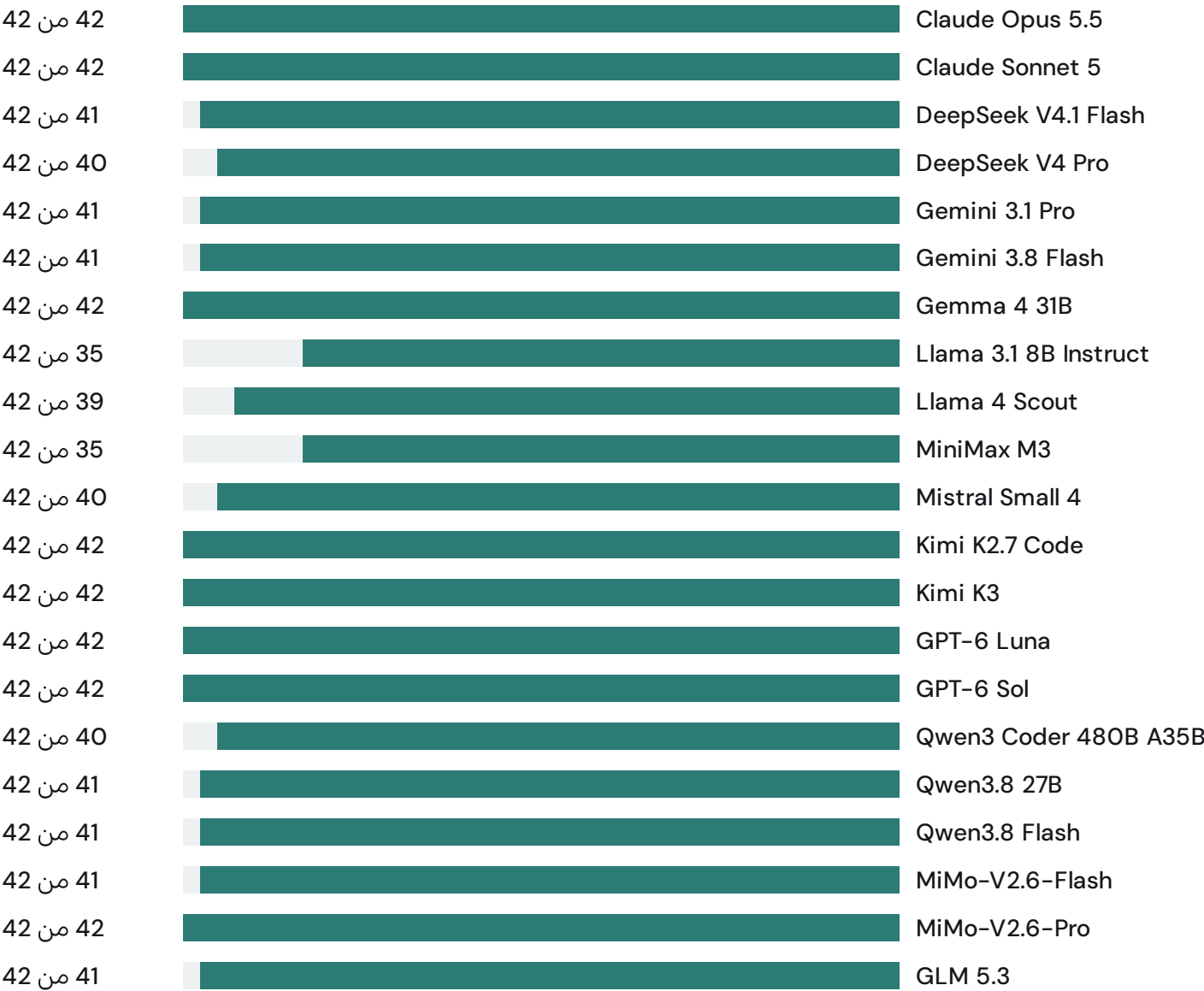
ال	النموذج					
	الكتابة بالفصحى	رسالة رسمية بالفصحى	تقرير قصير بالنسب الصحيحة	اللغة الليبية	الترجمة العربية إلى الإنجليزية	القراءة والتلخيص
شرح خط بخد	5.0	5.0	5.0	5.0	5.0	4.5
	4.5	4.0	4.5	4.5	5.0	5.0

النموذج	رسالة رسمية بالفصحى	تقرير قصير بالنسب الصحيحة	سؤال بالدارجة الليبية	ترجمة من العربية إلى الإنجليزية	ترجمة من الإنجليزية إلى العربية	القراءة والتلخيص	الكتابة بالفصحى	اللهجة الليبية	الترجمة	القراءة والتلخيص	النموذج
DeepSeek V4.1 Flash	5.0	4.5	4.5	5.0	5.0	4.0	5.5	5.0	5.0	5.0	DeepSeek V4.1 Flash
DeepSeek V4 Pro	5.0	4.0	3.3	5.0	5.0	5.0	5.0	5.0	5.0	5.0	DeepSeek V4 Pro
Gemini 3.1 Pro	4.5	4.5	5.0	5.0	5.0	4.5	5.0	5.0	5.0	4.5	Gemini 3.1 Pro
Gemini 3.8 Flash	4.5	5.0	4.0	5.0	5.0	5.0	5.0	5.0	5.0	4.5	Gemini 3.8 Flash
Gemma 4 31B	4.5	5.0	4.5	5.0	5.0	3.0	5.5	5.0	5.0	3.0	Gemma 4 31B
Llama 3.1 8B Instruct	2.0	1.5	1.0	3.5	2.0	1.0	5.0	5.0	5.0	1.0	Llama 3.1 8B Instruct
Llama 4 Scout	4.5	4.0	3.0	5.0	4.3	3.0	5.0	5.0	5.0	3.0	Llama 4 Scout
MiniMax M3	4.5	4.0	4.0	5.0	5.0	4.0	5.0	5.0	5.0	4.0	MiniMax M3
Mistral Small 4	3.5	4.5	2.5	5.0	5.0	4.0	5.0	5.0	5.0	4.0	Mistral Small 4
Kimi K2.7 Code	5.0	5.0	3.0	5.0	5.0	4.5	5.0	5.0	5.0	4.5	Kimi K2.7 Code
Kimi K3	5.0	4.5	4.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	Kimi K3
GPT-6 Luna	5.0	5.0	5.0	5.0	5.0	4.5	5.0	5.0	5.0	4.5	GPT-6 Luna
GPT-6 Sol	5.0	5.0	5.0	5.0	5.0	4.5	5.0	5.0	5.0	4.5	GPT-6 Sol
Qwen3 Coder 480B A35B	3.0	2.0	1.5	3.0	4.5	4.0	5.0	5.0	4.5	4.0	Qwen3 Coder 480B A35B
Qwen3.8 27B	4.5	4.5	2.5	5.0	3.5	4.5	5.0	5.0	4.5	4.5	Qwen3.8 27B
Qwen3.8 Flash	4.5	4.0	3.5	5.0	5.0	5.0	5.0	5.0	5.0	5.0	Qwen3.8 Flash
MiMo-V2.6-Flash	4.0	4.5	3.0	5.0	4.5	3.0	5.0	5.0	4.5	3.0	MiMo-V2.6-Flash
MiMo-V2.6-Pro	4.0	4.0	3.0	5.0	4.5	3.5	5.0	5.0	4.5	3.0	MiMo-V2.6-Pro
GLM 5.3	4.0	4.5	4.5	5.0	5.0	4.0	5.0	5.0	5.0	5.0	GLM 5.3

الدرجة: 1 2 3 4 5 خط متقطع: اختلف الحكمان بأكثر من درجة

الفحوص الآلية التي نجح فيها كل نموذج

الفحوص الآلية عدّ وتحقق حرفي وتشغيل كود. ومجموعها 42 فحوصًا في المهام الاثنتي عشرة لكل نموذج. دون حكم محكم.



اتفاق المحكمين

كل خلية عدد الإجابات التي أعطاهها المحكم الأول الدرجة في الصف والثاني الدرجة في العمود. القطر هو الاتفاق التام.

درجة المحكم الثاني					درجة المحكم الأول
5	4	3	2	1	
0	0	0	1	3	1
0	2	1	8	1	2

درجة المحكّم الثاني					درجة المحكّم الأول
5	4	3	2	1	
7	8	11	1	0	3
31	20	7	2	0	4
117	25	7	0	0	5

اتفق المحكّمان في 159 إجابة من 252 على الدرجة نفسها، واختلفا بأكثر من درجة في 20 إجابة (خط متقطع).

الدرجات بحسب نوع المهمة

في كل خلية الدرجة من 5 فوق، وتحتها عدد الفحوص الآلية التي نجحت من مجموعها. النماذج مرتبة أبجديًا لا بحسب الدرجة.

الكتابة بالفصحى

النموذج	رسالة رسمية بالفصحى	تقرير قصير بالنسب الصحيحة
Claude Opus 5.5	5.0 4 من 4	5.0 4 من 4
Claude Sonnet 5	4.5 4 من 4	4.0 4 من 4
DeepSeek V4.1 Flash	5.0 4 من 4	4.5 4 من 4
DeepSeek V4 Pro	5.0 4 من 4	4.0 3 من 4
Gemini 3.1 Pro	4.5 4 من 4	4.5 4 من 4
Gemini 3.8 Flash	4.5 4 من 4	5.0 4 من 4
Gemma 4 31B	4.5 4 من 4	5.0 4 من 4
Llama 3.1 8B Instruct	2.0 4 من 4	1.5 3 من 4
Llama 4 Scout	4.5 4 من 4	4.0 4 من 4
MiniMax M3	4.5 4 من 4	4.0 4 من 4

رسالة رسمية بالفصحى		النموذج
تقرير قصير بالنسب الصحيحة		
4.5 4 من 4	3.5 4 من 4	<u>Mistral Small 4</u>
5.0 4 من 4	5.0 4 من 4	<u>Kimi K2.7 Code</u>
4.5 4 من 4	5.0 4 من 4	<u>Kimi K3</u>
5.0 4 من 4	5.0 4 من 4	GPT-6 Luna
5.0 4 من 4	5.0 4 من 4	GPT-6 Sol
2.0 4 من 3	3.0 4 من 4	<u>Qwen3 Coder 480B A35B</u>
4.5 4 من 4	4.5 4 من 4	<u>Qwen3.8 27B</u>
4.0 4 من 4	4.5 4 من 4	<u>Qwen3.8 Flash</u>
4.5 4 من 4	4.0 4 من 4	<u>MiMo-V2.6-Flash</u>
4.0 4 من 4	4.0 4 من 4	<u>MiMo-V2.6-Pro</u>
4.5 4 من 4	4.0 4 من 4	<u>GLM 5.3</u>

اللهجة الليبية

سؤال بالدارجة الليبية		النموذج
5.0 2 من 2		Claude Opus 5.5
4.5 2 من 2		Claude Sonnet 5
4.5 2 من 2		<u>DeepSeek V4.1 Flash</u>
3.3 2 من 2		<u>DeepSeek V4 Pro</u>
5.0 2 من 2		Gemini 3.1 Pro
4.0 2 من 2		Gemini 3.8 Flash
4.5 2 من 2		<u>Gemma 4 31B</u>

1.0	Llama 3.1 8B Instruct
2 من 1	
3.0	Llama 4 Scout
2 من 2	
4.0	MiniMax M3
2 من 2	
2.5	Mistral Small 4
2 من 2	
3.0	Kimi K2.7 Code
2 من 2	
4.0	Kimi K3
2 من 2	
5.0	GPT-6 Luna
2 من 2	
5.0	GPT-6 Sol
2 من 2	
1.5	Qwen3 Coder 480B A35B
2 من 2	
2.5	Qwen3.8 27B
2 من 2	
3.5	Qwen3.8 Flash
2 من 2	
3.0	MiMo-V2.6-Flash
2 من 2	
3.0	MiMo-V2.6-Pro
2 من 2	
4.5	GLM 5.3
2 من 2	

الترجمة

5.0	5.0	Claude Opus 5.5
3 من 3	4 من 4	
5.0	5.0	Claude Sonnet 5
3 من 3	4 من 4	
5.0	5.0	DeepSeek V4.1 Flash
3 من 3	4 من 4	
5.0	5.0	DeepSeek V4 Pro
3 من 3	4 من 4	

ترجمة من العربية إلى الإنجليزية		النموذج
5.0 3 من 3	5.0 4 من 4	Gemini 3.1 Pro
5.0 3 من 3	5.0 4 من 4	Gemini 3.8 Flash
5.0 3 من 3	5.0 4 من 4	<u>Gemma 4 31B</u>
2.0 3 من 3	3.5 4 من 4	<u>Llama 3.1 8B Instruct</u>
4.3 3 من 3	5.0 4 من 4	<u>Llama 4 Scout</u>
5.0 3 من 1	5.0 4 من 4	<u>MiniMax M3</u>
5.0 3 من 3	5.0 4 من 4	<u>Mistral Small 4</u>
5.0 3 من 3	5.0 4 من 4	<u>Kimi K2.7 Code</u>
5.0 3 من 3	5.0 4 من 4	<u>Kimi K3</u>
5.0 3 من 3	5.0 4 من 4	GPT-6 Luna
5.0 3 من 3	5.0 4 من 4	GPT-6 Sol
4.5 3 من 3	3.0 4 من 4	<u>Qwen3 Coder 480B A35B</u>
3.5 3 من 3	5.0 4 من 4	<u>Qwen3.8 27B</u>
5.0 3 من 3	5.0 4 من 4	<u>Qwen3.8 Flash</u>
4.5 3 من 3	5.0 4 من 4	<u>MiMo-V2.6-Flash</u>
4.5 3 من 3	5.0 4 من 4	<u>MiMo-V2.6-Pro</u>
4.0 3 من 3	5.0 4 من 4	<u>GLM 5.3</u>

القراءة والتلخيص

تلخيص نص عربي		النموذج
4.5 3 من 3		Claude Opus 5.5

5.0 3 من 3	Claude Sonnet 5
4.0 3 من 3	DeepSeek V4.1 Flash
5.0 3 من 3	DeepSeek V4 Pro
4.5 3 من 3	Gemini 3.1 Pro
5.0 3 من 3	Gemini 3.8 Flash
3.0 3 من 3	Gemma 4 31B
1.0 3 من 2	Llama 3.1 8B Instruct
3.0 3 من 2	Llama 4 Scout
4.0 3 من 3	MiniMax M3
4.0 3 من 3	Mistral Small 4
4.5 3 من 3	Kimi K2.7 Code
5.0 3 من 3	Kimi K3
4.5 3 من 3	GPT-6 Luna
4.5 3 من 3	GPT-6 Sol
4.0 3 من 3	Qwen3 Coder 480B A35B
4.5 3 من 3	Qwen3.8 27B
5.0 3 من 3	Qwen3.8 Flash
3.0 3 من 3	MiMo-V2.6-Flash
3.5 3 من 3	MiMo-V2.6-Pro
5.0 3 من 3	GLM 5.3

المذاكرة والحساب

النموذج	شرح درس خطوة بخطوة	مسألة حسابية بالعربية
Claude Opus 5.5	4.5 3 من 3	5.0 2 من 2
Claude Sonnet 5	4.5 3 من 3	5.0 2 من 2
DeepSeek V4.1 Flash	3.5 3 من 2	5.0 2 من 2
DeepSeek V4 Pro	5.0 3 من 2	5.0 2 من 2
Gemini 3.1 Pro	5.0 3 من 2	5.0 2 من 2
Gemini 3.8 Flash	4.5 3 من 3	5.0 2 من 2
Gemma 4 31B	3.5 3 من 3	5.0 2 من 2
Llama 3.1 8B Instruct	1.0 3 من 2	4.0 2 من 2
Llama 4 Scout	4.0 3 من 2	5.0 2 من 2
MiniMax M3	4.0 3 من 1	4.5 2 من 2
Mistral Small 4	4.5 3 من 3	5.0 2 من 2
Kimi K2.7 Code	4.5 3 من 3	5.0 2 من 2
Kimi K3	4.5 3 من 3	5.0 2 من 2
GPT-6 Luna	4.0 3 من 3	5.0 2 من 2
GPT-6 Sol	5.0 3 من 3	5.0 2 من 2
Qwen3 Coder 480B A35B	3.5 3 من 2	5.0 2 من 2
Qwen3.8 27B	4.5 3 من 2	5.0 2 من 2
Qwen3.8 Flash	4.5 3 من 2	5.0 2 من 2
MiMo-V2.6-Flash	2.0 3 من 3	5.0 2 من 2
MiMo-V2.6-Pro	3.0 3 من 3	5.0 2 من 2

النموذج	شرح درس خطوة بخطوة	مسألة حسابية بالعربية
GLM 5.3	5.0 2 من 3	5.0 2 من 2

العمل والسيرة الذاتية

النموذج	قسم من سيرة ذاتية
Claude Opus 5.5	5.0 4 من 4
Claude Sonnet 5	4.5 4 من 4
DeepSeek V4.1 Flash	5.0 4 من 4
DeepSeek V4 Pro	5.0 4 من 4
Gemini 3.1 Pro	5.0 4 من 4
Gemini 3.8 Flash	5.0 4 من 3
Gemma 4 31B	5.0 4 من 4
Llama 3.1 8B Instruct	4.5 4 من 4
Llama 4 Scout	4.0 4 من 3
MiniMax M3	5.0 4 من 4
Mistral Small 4	4.0 4 من 3
Kimi K2.7 Code	5.0 4 من 4
Kimi K3	5.0 4 من 4
GPT-6 Luna	5.0 4 من 4
GPT-6 Sol	5.0 4 من 4
Qwen3 Coder 480B A35B	3.0 4 من 4
Qwen3.8 27B	4.5 4 من 4

4.5 4 من 4	<u>Qwen3.8 Flash</u>
3.5 4 من 3	<u>MiMo-V2.6-Flash</u>
3.0 4 من 4	<u>MiMo-V2.6-Pro</u>
4.0 4 من 4	<u>GLM 5.3</u>

البرمجة

4.5 3 من 3	Claude Opus 5.5
4.0 3 من 3	Claude Sonnet 5
4.0 3 من 3	<u>DeepSeek V4.1 Flash</u>
4.5 3 من 3	<u>DeepSeek V4 Pro</u>
5.0 3 من 3	Gemini 3.1 Pro
5.0 3 من 3	Gemini 3.8 Flash
3.5 3 من 3	<u>Gemma 4 31B</u>
2.0 3 من 2	<u>Llama 3.1 8B Instruct</u>
4.0 3 من 3	<u>Llama 4 Scout</u>
4.5 3 من 3	<u>MiniMax M3</u>
4.0 3 من 3	<u>Mistral Small 4</u>
4.5 3 من 3	<u>Kimi K2.7 Code</u>
4.0 3 من 3	<u>Kimi K3</u>
5.0 3 من 3	GPT-6 Luna

النموذج		دالة برمجية بتعليقات عربية
النموذج	GPT-6 Sol	5.0 3 من 3
	Qwen3 Coder 480B A35B	3.0 3 من 3
	Qwen3.8 27B	3.5 3 من 3
	Qwen3.8 Flash	3.5 3 من 3
	MiMo-V2.6-Flash	4.0 3 من 3
	MiMo-V2.6-Pro	2.0 3 من 3
	GLM 5.3	5.0 3 من 3

الشكل والاتجاه

النموذج		الالتزام بشروط الشكل	نص مختلط عربي وإنجليزي وأرقام
النموذج	Claude Opus 5.5	5.0 5 من 5	5.0 5 من 5
	Claude Sonnet 5	5.0 5 من 5	4.0 5 من 5
	DeepSeek V4.1 Flash	5.0 5 من 5	5.0 5 من 5
	DeepSeek V4 Pro	5.0 5 من 5	5.0 5 من 5
	Gemini 3.1 Pro	4.0 5 من 5	4.5 5 من 5
	Gemini 3.8 Flash	4.5 5 من 5	5.0 5 من 5
	Gemma 4 31B	5.0 5 من 5	3.5 5 من 5
النموذج	Llama 3.1 8B Instruct	2.0 5 من 4	2.0 5 من 4
	Llama 4 Scout	5.0 5 من 5	4.5 5 من 5
	MiniMax M3	4.0 5 من 2	3.0 5 من 5
	Mistral Small 4	3.5 5 من 4	4.5 5 من 5

النموذج	الالتزام بشروط الشكل	نص مختلط عربي وإنجليزي وأرقام
<u>Kimi K2.7 Code</u>	5.0 5 من 5	5.0 5 من 5
<u>Kimi K3</u>	5.0 5 من 5	5.0 5 من 5
GPT-6 Luna	5.0 5 من 5	5.0 5 من 5
GPT-6 Sol	5.0 5 من 5	5.0 5 من 5
<u>Qwen3 Coder 480B A35B</u>	4.5 5 من 5	4.5 5 من 5
<u>Qwen3.8 27B</u>	5.0 5 من 5	4.5 5 من 5
<u>Qwen3.8 Flash</u>	5.0 5 من 5	5.0 5 من 5
<u>MiMo-V2.6-Flash</u>	3.0 5 من 5	4.0 5 من 5
<u>MiMo-V2.6-Pro</u>	5.0 5 من 5	4.5 5 من 5
<u>GLM 5.3</u>	4.5 5 من 5	4.5 5 من 5

كيف تستشهد بهذا الاختبار

مختبر أكاكوس (2026). اختبار أكاكوس العربي، الإصدار 1.1. شفرة التكوين. <https://acacus.ly/research/arabic-test-2026-10>

Acacus Lab (2026). Acacus Arabic Test, v1.1. Shafra AI-Takween.
<https://acacus.ly/research/arabic-test-2026-10>

