

تقرير شامل: نماذج التعرف على الكلام العربي (Arabic ASR Models)

التاريخ: 18-02-2026 المشروع: إملاء صوتي عربي (arabic-voice-dictation) السياق: مشروعنا يستخدم whisper-rs مع Tauri 2

1. ترتيب النماذج العالمي (Open Universal Arabic ASR Leaderboard)

المصدر: [Arabic ASR Leaderboard](#) يُقيّم النماذج على 5 مجموعات بيانات متعددة اللهجات: SADA, Common Voice 18.0, MASC Clean, MASC Noisy, MGB-2

الترتيب	النموذج	متوسط WER	الحجم	ملاحظات
1	nvidia/conformer-ctc-large-ar (+LM)	25.71%	115M	الأفضل إطلاقاً
2	nvidia/conformer-ctc-large-ar (greedy)	27.46%	115M	language model بدون
3	openai/whisper-large-v3	29.87%	1,550M	Whisper أفضل نموذج
4	facebook/seamless-m4t-v2-large	32.55%	~2,300M	متعدد المهام
5	openai/whisper-large-v3-turbo	33.30%	809M	v3 من أسرع 6
6	openai/whisper-large-v2	34.04%	1,550M	-
7	openai/whisper-large	36.65%	1,550M	-
8	asafaya/hubert-large-arabic-ft	39.29%	~315M	Common على WER أقل Voice
9	openai/whisper-medium	39.60%	769M	-
10	facebook/mms-1b-all	47.86%	1,000M	لغة 1,107
11	whitefox123/w2v-bert-2.0-ft	48.62%	~600M	-
12	openai/whisper-small	52.18%	244M	-
13	jonatasgrosmann/wav2vec2-large-xlsr-ft	54.63%	315M	-
14	speechbrain/wav2vec2-large-ft	60.15%	315M	-

2. نماذج Whisper الأصلية - الأداء على العربية

أ. المقارنة حسب الحجم

النموذج	المعاملات	حجم الملف	RAM	متوسط (كل WER اللهجات)	FLEURS على WER (فصحى)
whisper-tiny	39M	75MB	~400MB	(تقدير) +65%	غير مختبر
whisper-base	74M	142MB	~500MB	(تقدير) +50%	غير مختبر

whisper-small	244M	466MB	~1GB	52.18%	30.11%
whisper-medium	769M	1.5GB	~2.5GB	39.60%	19.10%
whisper-large-v2	1,550M	3GB	~5GB	34.04%	17.14%
whisper-large-v3	1,550M	3GB	~5GB	29.87%	~11.5%
whisper-large-v3-turbo	809M	1.5GB	~2.6GB	33.30%	-

ب. ملاحظات مهمة

- القفزة الأكبر: من small إلى medium (~24-36% تحسن نسبي) - أعلى قيمة مقابل الحجم
- **large-v3-turbo**: قريب جداً من (33.30% vs 34.04%) large-v2 رغم أن لديه 4 طبقات decoder فقط بدلاً من 32
- مشكلة معروفة في **v3-turbo**: كلمة "نعم" تُكتب أحياناً كـ "Naah"/"Naahe"
- **Fine-tuning** أهم من الحجم: whisper-small المُدرَّب على MGB2 بـ WER 16.69% يتفوق على whisper-large-v2 بدون تدريب بـ WER 17.14%

3. نماذج Whisper المُدرَّبة على العربية (Fine-Tuned)

النموذج	الأساس	WER	بيانات التدريب	اللهجة
ARBML/whisper-largev2.1	large-v2	11.60% (FLEURS)	MGB2 (1200 ساعة)	فصحى
mohammed/whisper-large-arabic-cv-11	large	12.62%	Common Voice 11	فصحى
MohammedNasri/WHISPER-LARGE-ARABIC	large	15.93%	Common Voice 11	فصحى
speechbrain/asr-whisper-large-v2-commonvoice-ar	large-v2	16.96%	Common Voice 10	فصحى
anaszil/whisper-large-v3-turbo-darija	large-v3-turbo	24.88%	خاص	دارجة مغربية
HebArabNlpProject/WhisperLevantine	large-v3	33.00%	ساعة 1200	شامية
MohamedRashad/Arabic-Whisper-CodeSwitching-Edition	large-v2	غير منشور	عينة 12.5k	عربي-إنجليزي
MAdel121/whisper-small-egyptian-arabic	small	غير منشور	بيانات مصرية	مصرية
MAdel121/whisper-medium-egy	medium	غير منشور	ساعة 72	مصرية
Seyfelislem/whisper-medium-arabic	medium	18.29%	غير محدد	فصحى
ayoubkirouane/whisper-small-ar	small	غير منشور	Common Voice 11	فصحى

4. نماذج غير Whisper

أ. NVIDIA FastConformer - الأفضل مطلقاً

الخاصية	القيمة
النموذج	nvidia/stt_ar_fastconformer_hybrid_large_pc_v1.0
الحجم	115M (Whisper Large! من x أصغر 13) فقط
WER	على المعايير القياسية، 25.71% متوسط على كل اللهجات 8-12%
الترخيص	CC-BY-4.0
التدريب	(MASC 690h + CV17 65h + FLEURS 5h) ساعة ~760
العيب	whisper-rs غير متوافق مع ،NeMo framework يحتاج

ب. Meta SeamlessM4T v2

الخاصية	القيمة
النموذج	facebook/seamless-m4t-v2-large
الحجم	~2,300M
WER	متوسط 32.55%
الميزة	يدعم الترجمة + التحويل + التوليد الصوتي، ~100 لغة
العيب	ضخم جداً

ج. MBZUAI ArTST - متخصص بالعربية

الخاصية	القيمة
النماذج	MBZUAI/artst_asr_v1 / v2 / v3
الحجم	فقط 200M
البنية	SpeechT5
الميزة	Whisper نموذج مخصص للهجات، نصف معدل خطأ 17
يدعم	مصري، خليجي، شامي، مغاربي + فصحي
MGB-2 WER	~12.8% (v1)

د. Meta MMS (Massively Multilingual Speech)

الخاصية	القيمة
النموذج	facebook/mms-1b-all
الحجم	1B
WER	متوسط 47.86%

الميزة	يدعم 1,107 لغة عبر محولات خاصة بكل لغة
العيب	جودة عربية متوسطة

هـ. Meta OmniASR (جديد 2025)

الخاصية	القيمة
النماذج	facebook/omniASR-CTC-7B / LLM-7B / CTC-1B / LLM-1B / CTC-300M / LLM-300M
اللغات	لغة +1,600
الميزة	تنوعاً لغوياً ASR أكثر نظام

و. Qwen3-ASR (جديد 2026)

الخاصية	القيمة
النموذج	Qwen/Qwen3-ASR-1.7B / 0.6B
اللغات	لغة 52
دقة تعريف اللغة	97.9%
الميزة	منافس للحلول التجارية

ز. Voxtral من Mistral (جديد 2025)

الخاصية	القيمة
النماذج	mistralai/Voxtral-Mini-4B-Realtime-2602 / Small-24B / Mini-3B
الترخيص	Apache 2.0
اللغات	لغة (العربية من ضمنها) 13
الميزة	خالص Rust يوجد تطبيق (Burn framework)
العيب	كبير نسبياً B جديد، 4

5. أداء اللهجات العربية بالتفصيل

أداء Whisper Large-v3 على اللهجات المختلفة

اللهجة	WER	مستوى الموارد	ملاحظات
الفصحى (MSA)	11.56%	عالي	أفضل أداء
الأردنية	39.68%	متوسط	الأقرب للفصحى صوتياً
المصرية	48.95%	متوسط	أكثر لهجة موارد
الفلسطينية	50.20%	منخفض	-
الإماراتية	52.88%	منخفض جداً	-

اليمنية	59.45%	منخفض جداً	-
المغربية	83.05%	منخفض	تأثير الفرنسية
الجزائرية	84.14%	منخفض	تأثير الفرنسية
الموريتانية	85.68%	منخفض جداً	الأسوأ أداءً

ملاحظة: Code-switching (خلط اللغات) يسبب فشل كامل - Whisper يسجل 90%+ WER على المقاطع المختلطة عربي-فرنسي

6. نماذج مخصصة لكل لهجة

المصرية (EGY) - الأفضل تغطية

- Whisper Small - MAde1121/whisper-small-egyptian-arabic مُدرَّب
- Whisper Medium - MAde1121/whisper-medium-egy ساعة بيانات
- ARBML/klaam - مكتبة Python تدعم المصرية ('lang='egy)
- FastConformer Egyptian - فاز بمسابقة MTC-AIC 2024

الشامية (LEV) - تغطية جيدة

- 1200 - HebArabNlpProject/WhisperLevantine ساعة، 33% WER
- ArtST v2 - يغطي سوريا، الأردن، لبنان، فلسطين
- IWSLT 2024 - مسار مخصص للشامية الشمالية

الخليجية (GLF) - تغطية ضعيفة

- لا يوجد نموذج مخصص مفتوح المصدر على HuggingFace
- ArtST v2 يغطي السعودية، الكويت، الإمارات، عُمان، قطر، العراق، اليمن
- Munsit (تجاري) - أفضل خيار للخليجية
- Lisan + DeepVA - يغطي الكويتية، البحرينية، القطرية، الإماراتية، السعودية

المغربية (NOR) - الأصعب

- 24.88% WER - anaszil/whisper-large-v3-turbo-darija (دارجة مغربية)
- ychafiqui/whisper-medium-darija - دارجة مغربية
- MGB-5 Challenge - أفضل نظام حقق 59.4% WER على المغربية
- مشكلة code-switching مع الفرنسية تسبب فشل كبير

7. مقارنة شاملة: المميزات والعيوب

نماذج Whisper (متوافقة مع مشروعتنا عبر whisper-rs)

العيوب	المميزات	النموذج
بطيء، هلوسات في الصمت، (3GB) ضخمة	للعربية (29.87%)، يدعم كل اللهجات WER أفضل	whisper-large-v3
"مشكلة" نعم، v3 أضعف قليلاً من v2	v2 قريب من WER، (1.5GB) حجم معقول، سريع 6	whisper-large-v3-turbo
ضعيف على اللهجات (39.60%)	حجم معقول، أداء مقبول للفصحى	whisper-medium
عالي (52%)، غير عملي للهجات WER	وسريع (466MB) خفيف	whisper-small
غير صالح للعربية	صغير جداً	whisper-tiny/base

نماذج غير Whisper (غير متوافقة مباشرة مع whisper-rs)

النموذج	المميزات	العيوب
NVIDIA FastConformer	(115M) الأفضل مطلقاً (25.71%)، صغير جداً	whisper-rs غير متوافق مع NeMo، يحتاج
ArTST	(200M) متخصص عربي، 17 لهجة، صغير	whisper-rs غير متوافق مع SpeechT5، يحتاج
Voxtral	موجود تطبيق Rust، Apache 2.0 ترخيص	كبير B جديد، 4
Qwen3-ASR	لغة، دقة عالية 52	جديد 1.7B،
SeamlessM4T	متعدد المهام، 100 لغة	ليس الأفضل للعربية، (2.3B) ضخمة
MMS	لغة 1,107	جودة عربية متوسطة (47.86%)

8. التوصية النهائية لمشروعنا

للاستخدام الحالي (whisper-rs)

الأولوية	النموذج	السبب
الأفضل جودة	openai/whisper-large-v3	وبطيء GB لكن 3، WER (29.87%) أقل
أفضل توازن ⭐	openai/whisper-large-v3-turbo	WER 33.30%، فقط، أسرع 6 GB 1.5
للفصحي فقط	mohammed/whisper-large-arabic-cv-11	WER 12.62% على الفصحى
خفيف	openai/whisper-medium	WER 39.60%، GB 1.5

للمستقبل (يستحق المتابعة)

1. Voxtral - يوجد تطبيق Apache 2.0 (Burn framework) Rust
2. NVIDIA FastConformer - إذا أمكن تحويله لـ ONNX واستخدامه من Rust
3. 17 - ArTST لهجة عربية مخصصة

النتيجة الأهم

whisper-large-v3-turbo هو الخيار الأفضل لتطبيقنا حالياً - يوفر توازن ممتاز بين الجودة (WER 33.30%) والسرعة (6x أسرع) والحجم (1.5GB). وهو بالفعل من ضمن النماذج المدعومة في مشروعنا.

لتحسين جودة اللهجات: الحل الأفضل هو استخدام AI Refiner (Claude/OpenAI/Gemini) لتصحيح أخطاء Whisper، خاصة مع سياق اللهجة.

المصادر

- [Open Universal Arabic ASR Leaderboard](#)
- [ArXiv Paper: Arabic ASR Leaderboard](#)
- [N-Shot Benchmarking of Whisper on Arabic \(Interspeech 2023\)](#)
- [Casablanca: Multidialectal Arabic ASR \(EMNLP 2024\)](#)
- [ArTST: Arabic Text-Speech Transformer](#)
- [ARBML/whisperar Benchmarks](#)
- [Munsit Arabic ASR](#)
- [Overcoming Data Scarcity in Multi-Dialectal Arabic ASR](#)

- [Munsit at NADI 2025](#)
- [NVIDIA FastConformer Arabic](#)
- [Qwen3-ASR](#)
- [Voxtral Rust Implementation](#)
- [ARBML/klaam](#)