

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

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

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

المصدر: SADA, Common Voice 18.0, MASC Clean, Arabic ASR Leaderboard يقيّم النماذج على 5مجموعات بيانات متعددة للهجات: 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	بدون language model
3	<b>openai/whisper-large-v3</b>	<b>29.87%</b>	1,550M	أفضل نموذج Whisper
4	facebook/seamless-m4t-v2-large	32.55%	~2,300M	متعدد المهام
5	<b>openai/whisper-large-v3-turbo</b>	<b>33.30%</b>	809M	6 أسرع من v3
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	أقل WER على Common 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	jonatasgrosman/wav2vec2-large-xlsr-ft	54.63%	315M	-
14	speechbrain/wav2vec2-large-ft	60.15%	315M	-

## 2. نماذج الأصلية - الأداء على العربية

### أ. المقارنة حسب الحجم

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

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

#### بـ. ملاحظات مهمة

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

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

النموذج	الأساس	WER	بيانات التدريب	اللهجة
ARBML/whisper-largev2.1	large-v2	<b>11.60%</b> (FLEURS)	MGB2 (1200 ساعة)	فصحي
mohammed/whisper-large-arabic-cv-11	large	<b>12.62%</b>	Common Voice 11	فصحي
MohammedNasri/WHISPER-LARGE-ARABIC	large	15.93%	Common Voice 11	فصحي
speechbrain/asr-whisper-large-v2-commonvoice-ar	large-v2	16.96%	Common Voice 10	فصحي
anaszil/whisper-large-v3-turbo-darija	large-v3-turbo	24.88%	خاص	دارجة مغربية
HebArabNlpProject/WhisperLevantine	large-v3	33.00%	1200 ساعة	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	72 ساعة	مصرية
SeyfeliSem/whisper-medium-arabic	medium	18.29%	غير محدد	فصحي
ayoubkirouane/whisper-small-ar	small	غير منشور	Common Voice 11	فصحي

### 4. نماذج غير Whisper

### أ. NVIDIA FastConformer .

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

### ب. Meta SeamlessM4T v2 .

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

### ج. MBZUAI ArTST .

الخاصية	القيمة
النموذج	MBZUAI/artst_asr_v1 / v2 / v3
الحجم	فقط 200M
البنية	SpeechT5
الميزة	نموذج مخصص للهجات، نصف معدل خطأ 17 Whisper
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	47.86٪ متوسط

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

#### ٤. (2025 جديد) Meta OmniASR

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

#### ٥. (2026 جديد) Qwen3-ASR

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

#### ٦. (2025 جديد) Mistral من Voxtral

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

### ٥. أداء اللهجات العربية بالتفصيل

#### أداء Whisper Large-v3 على اللهجات المختلفة

اللهمة	WER	مستوى الموارد	ملاحظات
الفصحي (MSA)	<b>11.56%</b>	عالي	أفضل أداء
الأردنية	<b>39.68%</b>	متوسط	الأقرب للفصحي صوتيًا
المصرية	48.95%	متوسط	أكثر لهجة مواردًا
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	<b>85.68%</b>	منخفض جداً	الأسوأ أداء

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

## 6. نماذج مخصصة لكل لهجة

### المصرية (EGY) - الأفضل تغطية

- MAdel121/whisper-small-egyptian-arabic - Whisper Small
- MAdel121/whisper-medium-egy - Whisper Medium
- Python - مكتبة ARBML/klaam
- MTC-AIC 2024 - فاز بمسابقة FastConformer Egyptian

### الشامية (LEV) - تغطية جيدة

- HebArabNLpProject/WhisperLevantine - 1200 ساعة، WER 33%
- 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

### المغاربية (NOR) - الأصعب

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

## 7. مقارنة شاملة: المميزات والعيوب

### نماذج Whisper (متوفقة مع مشروعنا عبر whisper-rs)

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

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

النموذج	المميزات	العيوب
NVIDIA FastConformer	(115M) الأفضل مطلقاً (25.71%)، صغير جداً	غير متوافق مع NeMo، يحتاج whisper-rs
ArTST	(200M) متخصص عربي، 17 لهجة، صغير	غير متوافق مع SpeechT5، يحتاج whisper-rs
Voxtral	موجود Rust تطبيق 2.0، 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) ضخم
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	وبيطىء ولكن 3GB أقل WER (29.87%)
أفضل توازن ★	openai/whisper-large-v3-turbo	1.5GB، أسرع 6x، WER 33.30%
للفصحي فقط	mohammed/whisper-large-arabic-cv-11	على الفصحي 12.62%
خفيف	openai/whisper-medium	1.5GB، WER 39.60%

### للمستقبل (يستحق المتابعة)

1. Rust (Burn framework)، Apache 2.0 - يوجد تطبيق Voxtral.

2. NVIDIA FastConformer - إذا أمكن تحويله لـ ONNX واستخدامه من Rust.

3. ArTST - لهجة عربية متخصصة.

### 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 اللهجات.

### 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](#)