Voice Arena

The Voice Arena Entity Benchmark: Methodology and Results

November 2026 · Whitepaper v1.2 · Accented English · speech-to-text · voicearena.com

Summary

We measured how accurately **90 speech-to-text models** transcribe **entities**: names, street addresses, email addresses, codes and IDs, account numbers, amounts, dates and times. 57 are commercial APIs from 23 providers, called through their own endpoints or OpenRouter; 33 are open-weights models run with the Hugging Face Open ASR Leaderboard's own code and container images. The benchmark is **677 customer-service sentences** carrying 2,745 annotated entities, each read in **seven English accents** (Nigerian, Filipino, Indian, East African, Mexican, Portuguese and US): 4,739 clips, 30.0 hours of audio and 19,215 entity instances per model.

Zoom Scribe v2.1 Pro leads with a task success rate (TSR) of 31.9%, the share of clips with every entity right, ahead of ElevenLabs Scribe v2 Medical (28.3%) and ElevenLabs Scribe v3 (27.9%); the best open-weights model, Mistral Voxtral Small 24B, reaches 21.1%. Names and email addresses are the hardest entities (17.0% and 17.9% across all models), and accuracy falls from 66.5% for US speakers to 55.9% for Nigerian speakers. Every model heard the same 16 kHz audio (upsampled to 24 kHz for the two OpenAI Realtime models, which accept only that) at its provider's defaults; this paper documents every request, setting and check (sections 2–9), how a transcript is scored (section 10) and the results (section 11 and Appendix B).

90	4,739	19,215	30.0 h	31.9%
models: 57 API, 33 open weights	clips: 677 sentences × 7 accents	entity instances per model	of audio per model	best TSR: Zoom Scribe v2.1 Pro

Models	57 API models (29 batch, 28 streaming) from 23 providers · 33 open-weights models
Data	677 sentences in 12 customer-service domains with 2,745 entities, each read in 7 accents by 718 speakers; a public set of 133 sentences and a private set of 544 (section 2)
Settings	Model id, English and audio format only; everything else at the provider's default (section 6)
Runs	12–13 Nov 2026; AssemblyAI STT 3.7 Realtime Agent and the re-recorded clips: 17 Nov; FluidVoice Fluid Speech Pro and Salesforce SOV-Transcribe-1: 20 Nov; open source on HF Jobs H200 GPUs (sections 4, 7, 8)
Scoring	TSR (the ranking metric), entity accuracy, WER and CER; deterministic lookup in a reviewed map of accepted spellings; relaxed TSR and entity accuracy also accept sound-alike spellings (section 10)

Examples	11 worked examples with what the models wrote: an easy, a medium and a hard sentence, and 8 typical errors (section 1.2); one clip per entity type (section 2.1)
Results	Leaderboard, entity types, accents, tiers and domains (section 11); TSR by accent for every model in Appendix B
Open source	All harness code at github.com/VoiceArenaAI/Entity-eval: one standalone script per API model, plus the open-source job launcher (section 14)

Contents

1 Introduction 4
1.1 What we measure, and why 4
1.2 What the errors look like 4
1.3 The pipeline 11
2 Benchmark data 12
2.1 Sentences and entities 12
2.2 Accents, speakers and recordings 14
2.3 Audio preparation 16
2.4 Two dataset samples 16
3 Models evaluated 20
4 Software and hardware 20
5 The API harness 20
6 API models: settings 21
6.1 How English was set 21
6.2 Settings that differ from the defaults 22
6.3 How a streaming transcript is assembled 22
6.4 Output formats to expect 23
7 Open-source models 23
7.1 Approach 23
7.2 Data handling and jobs 23
7.3 Changes from the leaderboard defaults 24
8 Running the evaluation 24
8.1 What was captured 24
8.2 Coverage 24
9 Verification 25
10 Scoring and ranking 25
10.1 Metrics 25
10.2 When an entity counts as correct 27
10.3 Normalisation for WER and CER 28
10.4 Edge cases 28
10.5 Relaxed metric: sound-alike spellings 29
11 Results 30
11.1 Leaderboard 31
11.2 By entity type 34
11.3 By accent 36
11.4 By sentence difficulty, entity tier and domain 38
11.5 Entity accuracy and WER 39
11.6 Public and private sets 40
12 Vendor behaviour that shaped the settings 40
13 Limitations 41
14 Reproducing the work 42
A Appendix A: Models 44
B Appendix B: Results by accent and difficulty 45
C Appendix C: Public and private sets 51
D Appendix D: Public and private sets by accent 54

1 Introduction

1.1 What we measure, and why

General speech-to-text test sets are dominated by everyday words. The words that carry a voice transaction are different: a claim number, a street name, a licence plate, a dosage, an email address. They are rare or unseen in training data, many have no single spoken form, and a transcript that gets every other word right but one digit of an account number wrong has failed at its job. Accent compounds the problem, because a model has to map an unfamiliar pronunciation onto a string it has rarely or never seen.

This benchmark isolates that case. Every sentence is built around entities, and every sentence is read in seven accents, so the same entity can be compared across accents and across models. This round produces one transcript per model and clip under identical, documented conditions, then scores each transcript on the entities its sentence carries (section 10). Transcripts and scoring are kept apart, so the same transcripts can be re-scored as the metric evolves.

1.2 What the errors look like

The ten best models of section 11.1 get every entity right in some clips and the same entity wrong in others. Over all 19,215 entity instances, **all ten miss 3,070 (16.0%)**. For person names and email addresses it is about half: 49.3% and 50.2% of instances are wrong in all ten. Dates, digit sequences and amounts are close to solved.

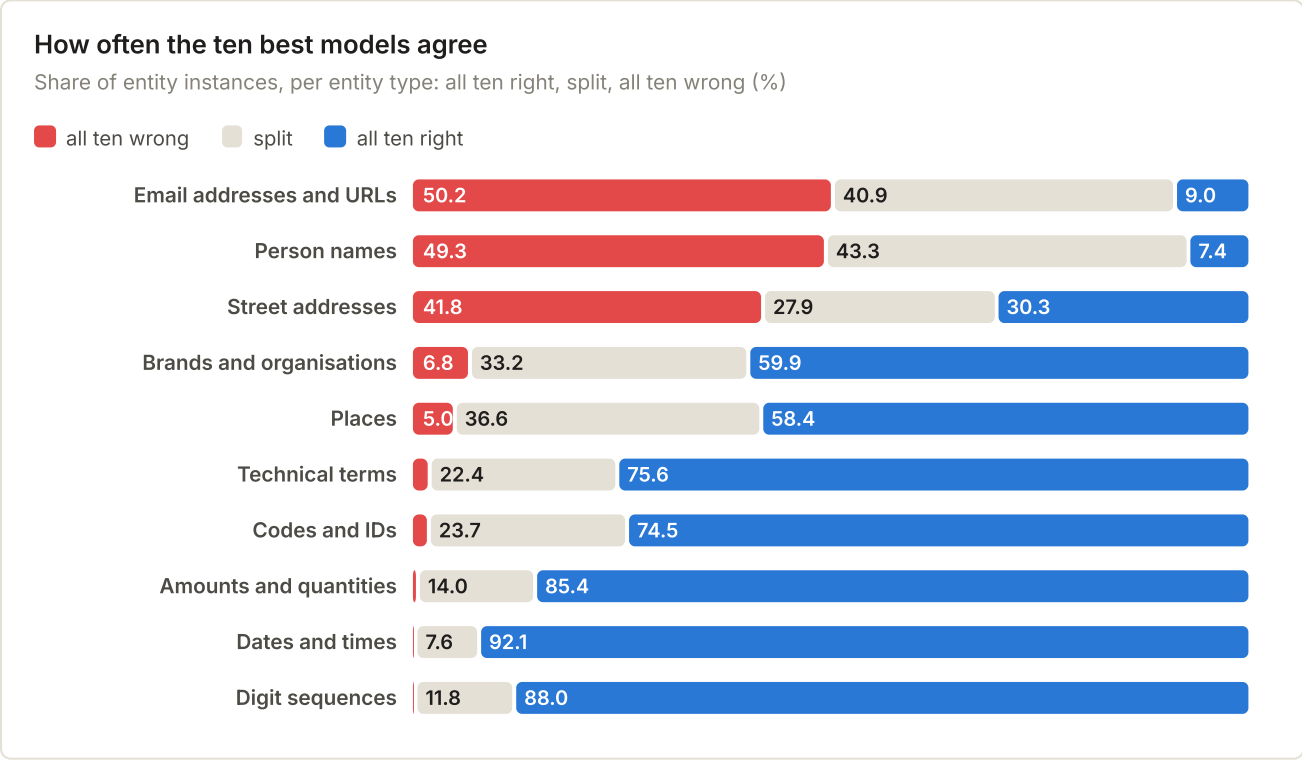

Figure 1. Each entity instance is one entity in one clip; ten models judge it with the rules of section 10.2.

Entity type	Entity accuracy, all 90 models	relaxed	Ten best	relaxed
Email addresses and URLs	17.9%	35.1%	28.0%	52.3%
Person names	17.0%	52.9%	26.0%	64.9%
Street addresses	33.6%	44.6%	42.3%	58.9%

Entity type	Entity accuracy, all 90 models	relaxed	Ten best	relaxed
Brands and organisations	62.9%	67.6%	80.3%	83.3%
Places	60.6%	63.5%	80.0%	82.2%
Technical terms	74.6%	75.8%	89.9%	90.5%
Codes and IDs	80.4%	80.4%	91.3%	91.3%
Amounts and quantities	84.7%	84.7%	95.4%	95.4%
Dates and times	89.4%	89.4%	97.9%	97.9%
Digit sequences	87.4%	87.4%	96.8%	96.8%

One easy, one medium and one hard clip

Easy, medium and hard. A sentence's difficulty is the share of its entity verdicts that are correct, over all 90 models and all seven accents. Sentences are split into thirds: **easy** 65.9% or more, **medium** 53.7–65.9%, **hard** below 53.7%. Examples below: S2088 is easy (77.7%), S0087 is medium (60.3%), S1610 is hard (46.5%).

One clip of each, with the raw output of the best batch API, streaming API and open-weights model, and how many of the clip's entities each wrote correctly. Every example in this paper gives both verdicts: the **strict** one that ranks the models, and the **relaxed** one (section 10.5), which also accepts a spelling that sounds the same. Each entity is underlined in green when strictly correct or red when not, with a second, yellow line beneath when the relaxed metric counts it.

EASY

An easy sentence

S2088 · East African English · Retail · 19.8 s · 3 entities

Price check request FKZB came from Merrick Familia, who noticed the shelf tag on the cereal did not match the register. Her rewards number is 27721, so please correct the tag on the shelf and give her a courtesy credit for the amount she was overcharged.

77.7% of entity verdicts correct on this sentence (all models, all accents); relaxed **85.3%**

76.3% in this clip; relaxed **80.0%**

37 of 90 models wrote every entity correctly; relaxed **45**

#1 Zoom Scribe v2.1 Pro best batch API	Price check request FKZB came from Merrick Familia, who noticed the shelf tag on the cereal did not match the register. Her rewards number is 27721, so please correct the tag on the shelf and give her a courtesy credit for the amount she was overcharged.	**3 of 3** entities, strict **3 of 3** relaxed
#6 Google Gemini 3.5 Transcribe Live best streaming API	Price check request FKZB came from Maryk Familia who noticed the shelf tag on the cereal did not match the register. Her rewards number is 27721 so please correct the tag on the shelf and give her a courtesy credit for the amount she was overcharged.	**2 of 3** entities, strict **3 of 3** relaxed
#14 Mistral Voxtral Small 24B best open weights	Price check request FKZB came from Merrick Familia who noticed the shelf tag on the cereal did not match the register. Her rewards number is 27721 so please correct the tag on the shelf and give her a Ketersee credit for the amount she was overcharged.	**3 of 3** entities, strict **3 of 3** relaxed

MEDIUM

A medium sentence

S0087 · Portuguese English · Healthcare · 21.8 s · 5 entities

Your appointment on November 3 at 3 PM is confirmed, a copy went to ashlyn@wahoo.us with booking code KAU, and the study team from Annovis Bio will meet you at 623 Elm Avenue.

60.3% of entity verdicts correct on this sentence (all models, all accents); relaxed **63.9%**

60.2% in this clip; relaxed **61.8%**

1 of 90 models wrote every entity correctly; relaxed **1**

#1 Zoom Scribe v2.1 Pro best batch API	Your appointment on November 3rd at 3 PM is confirmed. A copy went to ashlyn@wahoo.us with booking code KAU, and the study team from Anovis Bio will meet you at 623 Elm Avenue.	**4 of 5** entities, strict **4 of 5** relaxed
#6 Google Gemini 3.5 Transcribe Live best streaming API	Your appointment on November 3rd at 3:00 p.m. is confirmed. A copy went to ashley@wahoo.us with booking code KAU and the study team from Annovis Bio will meet you at 623 Elm Avenue.	**4 of 5** entities, strict **4 of 5** relaxed
#14 Mistral Voxtral Small 24B best open weights	Your appointment on November 3rd at 3 pm is confirmed. A copy went to ashlyn@wahoo.us with booking code KAU and the study team from Anovus Bio will meet you at 623 Elm Avenue.	**4 of 5** entities, strict **4 of 5** relaxed

HARD

A hard sentence

S1610 · Filipino English · Finance · 20.3 s · 4 entities

I am rebuilding my credit with a secured credit card I opened in Holladay, UT, and the letter about upgrading it lists offer code BLCK. Could Katia Baughan, who helped me last time, explain when my deposit would be returned?

46.5% of entity verdicts correct on this sentence (all models, all accents); relaxed **66.6%**

47.8% in this clip; relaxed **77.8%**

0 of 90 models wrote every entity correctly; relaxed **35**

#1 Zoom Scribe v2.1 Pro best batch API	I am rebuilding my credit with a secured credit card I opened in Holiday UT and a letter about upgrading it lists offer code BLCK. Could Katia Boughan, who helped me last time, explain when my deposit would be returned?	**2 of 4** entities, strict **4 of 4** relaxed
#6 Google Gemini 3.5 Transcribe Live best streaming API	I am rebuilding my credit with a secured credit card I opened in Holiday, UT and the letter about upgrading it lists offer code BLCK. Could Katie Boughan, who helped me last time, explain when my deposit would be returned?	**2 of 4** entities, strict **4 of 4** relaxed
#14 Mistral Voxtral Small 24B best open weights	I am rebuilding my credit with a secured credit card I opened in Holiday UT and a letter about upgrading it lists offer code BLCK. Could Katia Boghan who helped me last time explain when my deposit would be returned?	**2 of 4** entities, strict **4 of 4** relaxed

Eight typical errors

Each is a common error in its clip. Some are systematic, with the ten best models missing the entity in the other accents too; others are specific to one accent. The note under each gives how the same entity fared in the other six accents. Each grid shows the five best models, then the best-ranked models that wrote the entity correctly.

1 “7353 Lookabill Lane, Gifford, IA” became “7353 Lucasville Lane, Gifford, Iowa”

`S0960 · Portuguese English · Food · 19.4 s · Street addresses`

I'm traveling through Florence, KY this week, but my sister ordered a gift card from In-N-Out Burger for me. The confirmation code is 7C6B, and the card should be mailed to 7353 Lookabill Lane, Gifford, IA, so please double check that address.

#1 Zoom Scribe v2.1 Pro	7353 Lucabille Lane, Gifford, Iowa	≈
#2 ElevenLabs Scribe v2 Medical	7353 Lucasville Lane, Gifford Iowa	✗
#3 ElevenLabs Scribe v3	7353 Lucaville Lane, Gifford, Iowa.	✗
#4 Mistral Voxtral Small Transcribe 3	7-3-5-3 Luca Bill Lane, Gifford, Iowa	≈
#5 Microsoft MAI-Transcribe-2	7353 Lucabill Lane, Gifford, Iowa	≈
#7 AssemblyAI Universal-3.6 Pro	7353 Lookabill Lane, Gifford, Iowa	✓
#9 AssemblyAI STT 3.7 Realtime Agent	7353 Lookabill Lane, Gifford, Iowa	✓
#16 AssemblyAI Universal-3.6 Pro Realtime	7353 Lookabill Lane, Gifford, Iowa	✓
#24 OpenAI GPT Live Transcribe	seven three five three Lookabill Lane, Gifford, Iowa.	✓

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✗ wrong under both.

Strict: 4 of 90 models wrote it correctly (ten best: 2). **Relaxed** (section 10.5): 47 of 90 (ten best: 7); the most common error, “7353 Lucasville Lane, Gifford, Iowa” (5 models), is wrong under it too. 56 different forms overall. The same entity in the other six accents: 62 of 540 correct, relaxed 167 (0–36 per accent, relaxed 0–83).

2 “Jaxton Nakasone” became “Jackson Nakasone”

`S2594 · Mexican English · Legal / government · 19.1 s · Person names`

Jaxton Nakasone, a records custodian at Oak Ridge National Laboratory, has been asked to supply documents for a civil case, and he should send the certified copies to the clerk at tyanna.shatswell@courtrecords.us.

#1 Zoom Scribe v2.1 Pro	Jaxton Nakasone ✓	#2 ElevenLabs Scribe v2 Medical	Jackstone Nakasone ≈
#3 ElevenLabs Scribe v3	Jaxton Nakasone ✓	#4 Mistral Voxtral Small Transcribe 3	Jackston Nakasone ≈
#5 Microsoft MAI-Transcribe-2	Jaxton Nacason ≈	#13 OpenAI GPT Transcribe	Jaxton Nakasone ✓
#26 Cartesia Ink 2	Jaxton Nakasone ✓		

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✕ wrong under both.

Strict: 4 of 90 models wrote it correctly (ten best: 2). **Relaxed** (section 10.5): 28 of 90 (ten best: 5); the most common error, “Jackson Nakasone” (10 models), is wrong under it too. 58 different forms overall. The same entity in the other six accents: 18 of 540 correct, relaxed 113 (0–10 per accent, relaxed 0–31).

3 “gascon@larkmail.us” became “Gascon at blackmail.us”

`S2606 · East African English · Legal / government · 15.2 s · Email addresses and URLs`

Thank you for calling. Santiago Bushrod at the National Mediation Board is coordinating the airline labor session, and he asked both sides to submit position statements to gascon@larkmail.us so the mediator can read them in advance.

#1 Zoom Scribe v2.1 Pro	gascon at lacmail.us ✕	#2 ElevenLabs Scribe v2 Medical	gascon@lacmail.us ✕
#3 ElevenLabs Scribe v3	gascon@lackmail.us ✕	#4 Mistral Voxtral Small Transcribe 3	gascon @lockmail.us ✕
#5 Microsoft MAI-Transcribe-2	gascon@lacmail.us ✕	#16 AssemblyAI Universal-3.6 Pro Realtime	Gascon at larkmail.us ✓
#26 Cartesia Ink 2	Gascon at larkmail. us ✓	#28 Soniox STT Real-time v6	Gascon@larkmail.us ✓
#29 Soniox STT Async v6	gascon@larkmail.us ✓	#56 Reson8 Rumble-2 Flash	Gascon at larkmail.us ✓

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✕ wrong under both.

Strict: 5 of 90 models wrote it correctly (ten best: 0). **Relaxed** (section 10.5): 5 of 90 (ten best: 0); the most common error, “Gascon at blackmail.us” (8 models), is wrong under it too. 53 different forms overall. The same entity in the other six accents: 181 of 540 correct, relaxed 203 (4–57 per accent, relaxed 8–63).

4 “urijah.sowells@shieldmutual.com” became “araya.souls@shieldmutual.com”

`S1790 · Mexican English · Insurance · 17.5 s · Email addresses and URLs`

Following up on the email from urijah.sowells@shieldmutual.com, the $8.28 monthly premium quoted for Victor Ramnarine is for a basic Medicare Part D plan, and he can enroll online or by phone.

#1 Zoom Scribe v2.1 Pro	oriah.soules@shieldmutual.com	✗
#2 ElevenLabs Scribe v2 Medical	Araya.Souls@shieldmutual.com	✗
#3 ElevenLabs Scribe v3	aria.souls@shieldmutual.com	✗
#4 Mistral Voxtral Small Transcribe 3	araya.souls@shieldmutual.com	✗
#5 Microsoft MAI-Transcribe-2	oraya.souls@shieldmutual.com	✗

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✗ wrong under both.

Strict: 0 of 90 models wrote it correctly (ten best: 0). **Relaxed** (section 10.5): 0 of 90 (ten best: 0); the most common error, “araya.souls@shieldmutual.com” (10 models), is wrong under it too. 56 different forms overall. The same entity in the other six accents: 4 of 540 correct, relaxed 77 (0–2 per accent, relaxed 0–30).

5 “Pomonok” became “Pomona”

`S3101 · Nigerian English · Real estate · 16.8 s · Places`

Sign-ups are open for the March 26 at 1 PM tenant meeting at the Pomonok community center. Kian Hashagen from the housing advocacy group will explain the new repair request process and answer questions afterward.

#1 Zoom Scribe v2.1 Pro	Pomona	✗	#2 ElevenLabs Scribe v2 Medical	Pomona	✗
#3 ElevenLabs Scribe v3	Pomona	✗	#4 Mistral Voxtral Small Transcribe 3	Pomona	✗
#5 Microsoft MAI-Transcribe-2	Pomono	✗	#11 Google Cloud Chirp 3	Pomonok	✓
#13 OpenAI GPT Transcribe	Pomonok	✓	#22 OpenAI Whisper large-v3	Pomonok	✓
#23 Reson8 Resonant-1	Pomonok	✓	#26 Cartesia Ink 2	Pomonok	✓

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✗ wrong under both.

Strict: 25 of 90 models wrote it correctly (ten best: 0). **Relaxed** (section 10.5): 25 of 90 (ten best: 0); the most common error, “Pomona” (28 models), is wrong under it too. 18 different forms overall. The same entity in the other six accents: 134 of 540 correct, relaxed 135 (1–48 per accent, relaxed 1–49). Pomonok is a neighbourhood in Queens, New York; the models wrote the better-known city name.

6 “Pappadeaux” became “papa dough”

`S1023 · US English · Food · 25.9 s · Brands and organisations`

The pickup counter at 386 Pine Drive will have your Pappadeaux order ready on Saturday, June 8 at 2:50 PM. Show code GE-078X for the chicken-fried steak platter, and if anything is missing, call back with reference 9098 116 325.

#1 Zoom Scribe v2.1 Pro	Papa Doe	✗	#2 ElevenLabs Scribe v2 Medical	Papa Doe ✗
#3 ElevenLabs Scribe v3	Papa Dough	✗	#4 Mistral Voxtral Small Transcribe 3	Papa Doe ✗
#5 Microsoft MAI-Transcribe-2	Papa Dough	✗	#12 ElevenLabs Scribe v3 Live	Pappadeaux ✓

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✗ wrong under both.

Strict: 1 of 90 models wrote it correctly (ten best: 0). **Relaxed** (section 10.5): 1 of 90 (ten best: 0); the most common error, “papa dough” (38 models), is wrong under it too. 18 different forms overall. The same entity in the other six accents: 4 of 540 correct, relaxed 4 (0–2 per accent, relaxed 0–2). 2 of the ten best wrote “Papa Johns”, a different restaurant chain.

7 “425 Hill Way” became “425 Hillary”

`S1399 · East African English · Healthcare · 14.7 s · Street addresses`

When the home oxygen tech arrives, give him service ticket WHR, point him to the side door of 425 Hill Way, and use alarm code 6527 if the back gate is locked.

#1 Zoom Scribe v2.1 Pro	425 Hillway	✓	#2 ElevenLabs Scribe v2 Medical	425 Hillway ✓
#3 ElevenLabs Scribe v3	425 Hillway	✓	#4 Mistral Voxtral Small Transcribe 3	4-2-5 Hillary ✗
#5 Microsoft MAI-Transcribe-2	425 Hillway	✓	#6 Google Gemini 3.5 Transcribe Live	425 Hillway ✓
#10 Google Gemini 3.5 Transcribe	425 Hillway	✓	#12 ElevenLabs Scribe v3 Live	425 Hillway ✓
#13 OpenAI GPT Transcribe	425 Hillway	✓	#19 xAI Grok Voice Transcribe 2.0	four two five Hillway ✓

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✗ wrong under both.

Strict: 23 of 90 models wrote it correctly (ten best: 6). **Relaxed** (section 10.5): 23 of 90 (ten best: 6); the most common error, “425 Hillary” (22 models), is wrong under it too. 33 different forms overall. The same entity in the other six accents: 378 of 540 correct, relaxed 378 (52–84 per accent, relaxed 52–84).

8 “Triamcinolone” became “trimecanolin”

`S3505 · US English · Healthcare · 18.0 s · Technical terms`

Knee injection kit LJ54 includes Triamcinolone and a numbing agent, and Rosalva Appiah, the medical assistant at our Moline, IL office, will set it up on the procedure tray before the orthopedist arrives.

#1 Zoom Scribe v2.1 Pro	trimipramine ✗	#2 ElevenLabs Scribe v2 Medical	trimikanolin ✗
#3 ElevenLabs Scribe v3	triamcinolone ✓	#4 Mistral Voxtral Small Transcribe 3	trimacinolone ✗
#5 Microsoft MAI-Transcribe-2	trimiclanolin ✗	#6 Google Gemini 3.5 Transcribe Live	triamcinolone ✓
#10 Google Gemini 3.5 Transcribe	triamcinolone ✓	#11 Google Cloud Chirp 3	triamcinolone ✓
#13 OpenAI GPT Transcribe	triamcinolone ✓	#15 OpenAI GPT-4o Mini Transcribe	Triamcinolone ✓

The five best models (section 11.1), then the best-ranked models that wrote it correctly; raw output. ✓ correct · ≈ correct under the relaxed metric only (yellow line) · ✗ wrong under both.

Strict: 23 of 90 models wrote it correctly (ten best: 3). **Relaxed** (section 10.5): 23 of 90 (ten best: 3); a medical name, scored strictly under both. 49 different forms overall. The same entity in the other six accents: 298 of 540 correct, relaxed 298 (27–69 per accent, relaxed 27–69). The leader, Zoom Scribe v2.1 Pro, wrote “trimipramine”: the name of a different drug.

1.3 The pipeline

From prompt to leaderboard

Each stage is documented in the section shown in its box.

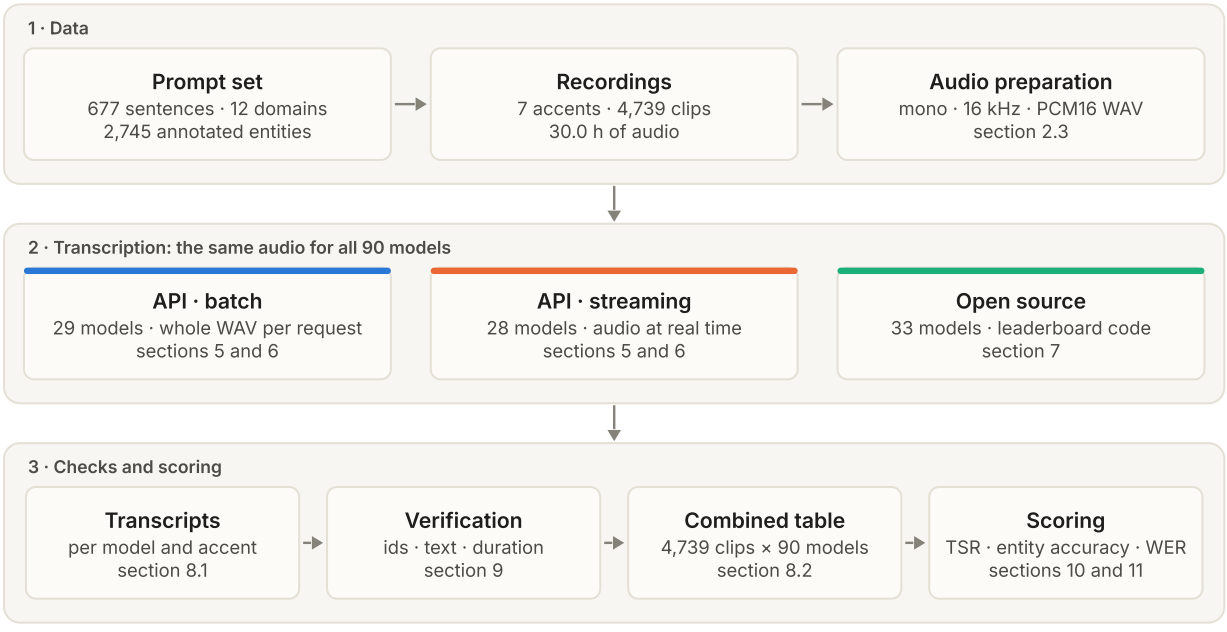

Figure 2. The evaluation pipeline. Batch, streaming and open-source models receive identical audio, and every transcript is checked against its clip before it enters the combined table.

2 Benchmark data

2.1 Sentences and entities

The evaluation set has **677 sentences**, each with a unique id and unique text, written as the things people say to, or hear from, customer-service lines: delivery notices, claim updates, appointment reminders, billing questions, support tickets. Each sentence was written around **3–6 entities** (mean 4.1), and each entity was annotated with a type and a difficulty tier at that point: **2,745 entities** in all, or 19,215 entity instances once every sentence is read in seven accents. The mean sentence has 37.3 words (range 22–54).

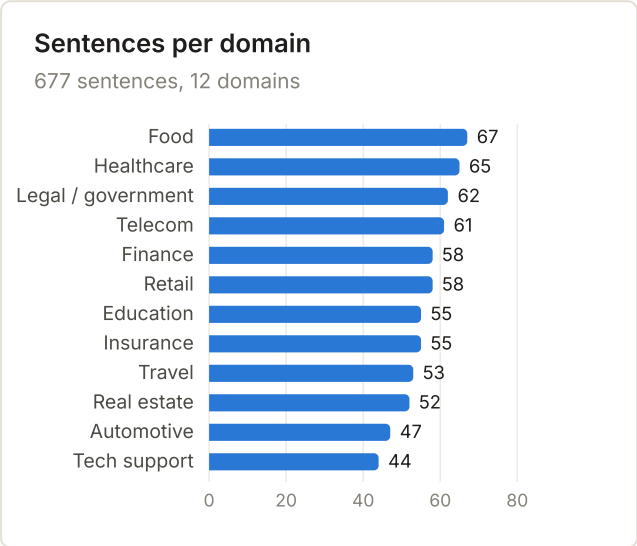

Figure 3. Domains are balanced within a factor of two (44–67 sentences).

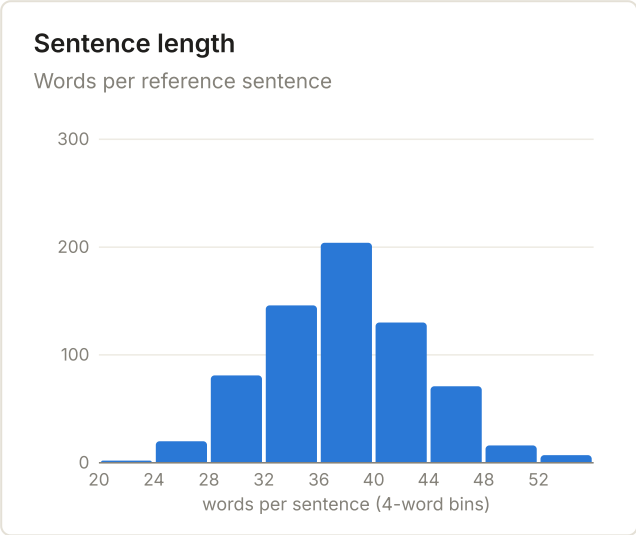

Figure 4. Median 37 words; every sentence is long enough to hold several entities.

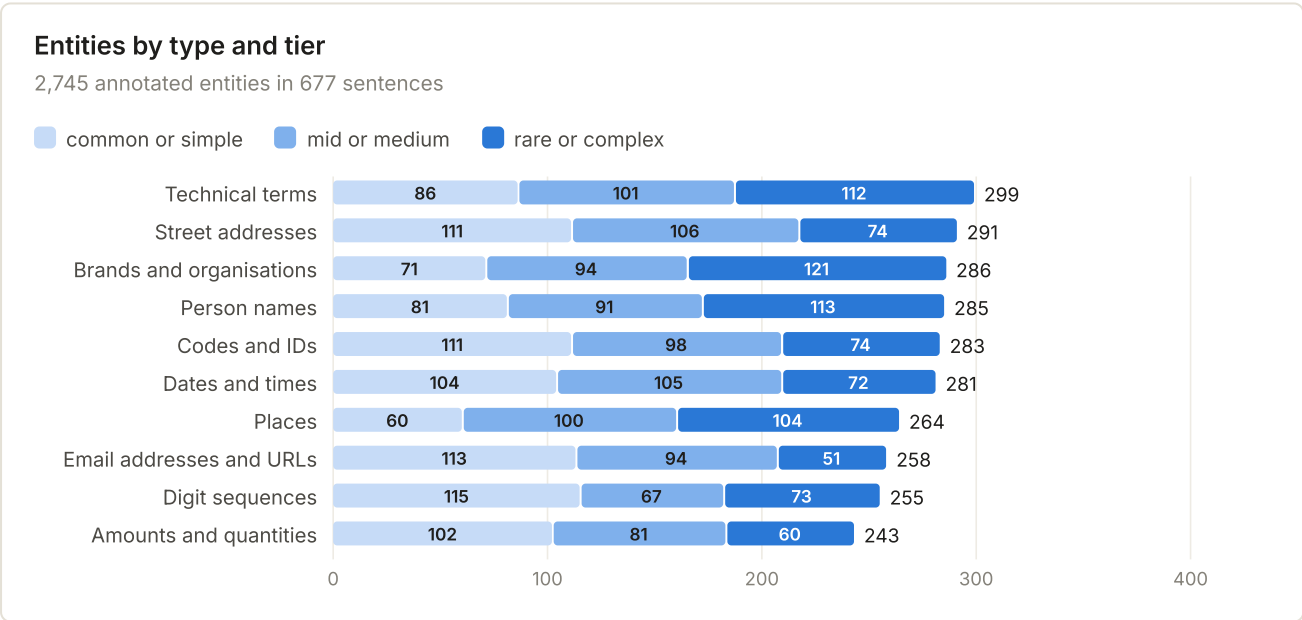

Figure 5. The ten entity types are close to evenly represented. Tiers grade difficulty within a type: common, mid and rare for names, brands, places and technical terms; simple, medium and complex for structured values.

Entity type	What it covers	Entities
Technical terms	Domain words and product names: medicines, devices, models, services	299
Street addresses	Full street addresses: number, direction, street, unit, city, state, ZIP	291

Entity type	What it covers	Entities
Brands and organisations	Businesses, organisations and brands	286
Person names	First and last names of customers, staff and contacts	285
Codes and IDs	Codes that mix letters and digits: bookings, references, plates, permits	283
Dates and times	Dates, weekdays and times	281
Places	Cities with their state, airports and other named places	264
Email addresses and URLs	Email addresses and web addresses	258
Digit sequences	Digit-only numbers: accounts, PINs, phone and claim numbers	255
Amounts and quantities	Money, and quantities with a unit	243

Hear each entity type

One clip per type, each a typical case: the entity whose share of correct models is closest to the average for its type.

Entity type	In its sentence	Models right (strict, relaxed)
Person names Rosalva Appiah	Knee injection kit LJ54 includes Triamcinolone and a numbing agent, and Rosalva Appiah, the medical assistant at our Moline, IL office, will set it up on the procedure tray before the orthopedist arrives. S3505 · Nigerian English · 15.6 s	16 of 90 relaxed 46 type average 15, relaxed 48
Brands and organisations Allianz Travel Insurance	Your appointment to review the claim is set for July 11 at 9 AM, and a representative from Allianz Travel Insurance named Elvira Proehl will call you on the number listed in your policy, so have your receipts ready. S1889 · Nigerian English · 15.0 s	57 of 90 relaxed 57 type average 57, relaxed 61
Places Leland, NC	Voicemail for the oyster roast organizers: the rental tables are scheduled for Sunday, July 13 at the park shelter in Leland, NC. Our booking code is JXHA, so mention it if you need to change the setup time. S4813 · US English · 19.3 s	56 of 90 relaxed 56 type average 55, relaxed 57
Technical terms Honda Ridgeline	Message for the front desk at 234 Park Street: key tag KJU belongs to Lindsay Aristy. She asked that the invoice go to fraleigh@kalm.us, and the Honda Ridgeline should be released only after payment for repair order 52868. S3058 · Indian English · 14.4 s	68 of 90 relaxed 85 type average 67, relaxed 68
Street addresses 527 Elm Boulevard	Room 6284 is where the evaluation committee will review applicants who have at least 65 credits. If you plan to attend, confirm by writing to sayavong@wahoo.us, and park in the visitor lot at 527 Elm Boulevard. S3337 · East African English · 19.1 s	30 of 90 relaxed 30 type average 30, relaxed 40
Email addresses and URLs emberway.com	Pharmacy pickup closes Sunday at 4 PM, so check emberway.com for weekend hours, or ask a neighbor at 534 Hill Boulevard to collect the Indaptus Therapeutics study pack labeled RAQ. S1476 · Mexican English · 15.6 s	16 of 90 relaxed 21 type average 16, relaxed 32
Codes and IDs HUY	Hey, it's me again about the prepaid card I bought for my grandson. The cashier said it would activate by December 5, but it still shows inactive, and the receipt lists a store code of HUY at the bottom. S2427 · Portuguese English · 16.2 s	72 of 90 relaxed 72 type average 72, relaxed 72
Digit sequences 82890	Hi, this is a voicemail for Shirley Rockmore. Your rental application number is 82890, and the landlord approved you for the cottage at 4686 Lepinski Lane, Salmon, ID. Please call us back to sign the lease. S1175 · Indian English · 18.1 s	79 of 90 relaxed 79 type average 79, relaxed 79

Entity type	In its sentence	Models right (strict, relaxed)
Amounts and quantities 0.62 mg	Melatonin at 0.62 mg was approved for the toddler, and Landry MacDougall from the sleep lab will phone the parents on Tuesday, June 25 at 2:05 PM to see how bedtime is going. S1422 · East African English · 13.2 s	76 of 90 relaxed 76 type average 76, relaxed 76
Dates and times 01/22/2028	Our coding summer camp awards up to 15 credits toward the district's digital skills badge. Early registration closes on 01/22/2028, so let us know soon if your child would like to save a spot. S3263 · Mexican English · 18.4 s	80 of 90 relaxed 80 type average 80, relaxed 80

How the set is built

- **Entity-first.** Every sentence carries 3–6 entities, each of a **different type**: no sentence repeats a type (284 sentences with 3, 165 sentences with 4, 135 sentences with 5, 93 sentences with 6).
- **Unique.** All 677 sentence ids and texts are distinct, all 2,745 entity strings are distinct, and every entity appears verbatim in its sentence, so each one can be located in the reference.
- **Graded.** Each entity carries a difficulty tier, set when the sentence was written. For structured values the tier follows the number of parts; for names, brands, places and terms it is a familiarity label, and entities in the three tiers are the same length (median 12–15 characters).

Entity type	Simple / common	Medium / mid	Complex / rare
Street addresses	House number and street only	Adds city and state	Adds direction, unit or suite and ZIP code (all of them)
Codes and IDs	3–4 letters, no digits	2–7 characters, mostly letters (22% with a digit)	2–13 characters, 78% with digits, 36% hyphenated
Digit sequences	4–5 digits	5–9 characters, 41% hyphenated	12–14 characters: long grouped numbers and phone numbers
Email addresses and URLs	name@domain, no dot or digit in the user name	first.last@domain (86%) or a bare domain	Digits in the address (85%), subdomains, URL paths (21%)
Amounts and quantities	Whole amounts and quantities, no cents	36% with cents	90% with cents, 27% in the thousands
Dates and times	A weekday, a date or an hour; no year	Adds a year (46%) or a time	Minute-level times (75%), numeric dates (25%)
Names, brands, places, terms	Familiar	Less familiar	Rare or unusual

Tier criteria as they appear in the 2,745 annotated entities (shares are of that tier). For names, brands, places and terms the tiers predict difficulty: across all models, rare entities score 45.0% against 64.0% for common ones. For structured values the simple tier is clearly easier (73.9%), while medium and complex score about the same (59.2% and 59.4%; section 11.4).

Writing and review. The 814 sentences were written by the Voice Arena team with the help of large language models, balancing domains, entity types and difficulty tiers across the set, and every sentence was reviewed twice before recording. **Entity sources.** First names come from the *babynames* dataset (github.com/hadley/babynames), each matched to a different last name, so every full name is invented. Email addresses and street names are made up. All other entities were collected from the web. All people, email addresses, street addresses, codes and scenarios are fictional; real organisations, products and places appear only as vocabulary. Every sentence printed in this paper is from the public set.

2.2 Accents, speakers and recordings

Every sentence was recorded in **seven English accents**: Nigerian, Filipino, Indian, East African, Mexican, Portuguese and US, 5,698 recordings of the 814 sentences originally written. Quality control compared each

recording with its sentence and removed **118** whose speech does not match it (missing, extra or changed words). So that every accent is scored on exactly the same sentences, a sentence that lost a recording in any accent was dropped from all of them.

Re-recording. After a first scoring pass, clips whose entities were missed by most models were reviewed by listening, and those flagged as possible speaker misreads were recorded again; a sentence whose entity was still missed by nearly every model after re-recording was removed from all seven accents. **297 clips** of the evaluation set use these re-recordings, chosen by a person listening to each flagged clip. In all, 137 of the 814 sentences were removed: 112 that lost a recording in quality control and 25 whose entity was still misread after re-recording. Leaving all 405 re-recorded clips (297 from this round and 108 from the public-set review below) out of the scoring hardly changes the ranking: Spearman ρ = 0.999 on TSR, the same 10 of the top 10, and no model moves more than 3 places.

Public set review. The sentences chosen for open release were matched to the full set on difficulty, domain, task type and entity types, and every one of their clips was checked again by listening. Clips that did not match their sentence were recorded again, and all 90 models transcribed the new recordings (17 November 2026; the 2 open-weights models added later, on 20 November 2026). A new recording is used only if it passes the public set's quality checks: no entity that almost no model gets in that clip while most get it in the other accents, no unusually high word error rate for the best models, and a speaking rate within 0.9–3.2 words per second.

Public and private sets. The evaluation set is therefore **677 sentences × 7 accents = 4,739 clips** and **29.97 hours** of audio: a **public** set of 133 sentences (931 clips), released openly, and a **private** set of 544 sentences (3,808 clips) that is not released. Results are reported on all 677 sentences, and on each set alone (section 11.6).

718 speakers recorded the set, 94–118 per accent. A speaker read a median of 6 clips and at most 45 (6.6% of an accent), so no single voice dominates an accent.

Accent	Speakers	Clips	Hours	Median clip	Most common native languages, self-reported (share of speakers)
Nigerian	98	677	4.52	22.9 s	English 39%, Yoruba 19%, Hausa 12%
Filipino	98	677	3.91	20.0 s	Tagalog (Filipino) 74%, not recorded 17%, English 5%
Indian	116	677	4.25	22.0 s	English 38%, Tamil 28%, Hindi 11%
East African	118	677	4.16	21.6 s	English 29%, Swahili (Kiswahili) 17%, not recorded 16%
Mexican	96	677	4.43	22.6 s	Spanish 86%, not recorded 10%
Portuguese	98	677	4.71	23.9 s	Portuguese 80%, not recorded 17%
US	94	677	3.99	20.3 s	English 93%, not recorded 5%

Clips run from 10.2 s to 54.5 s (median 21.8 s). 482 clips (10.2%) are longer than 30 s and 54 (1.1%) longer than 40 s, which matters for models with a fixed input window (section 7.3). The source files are 16-bit PCM WAV, mono, 44.1 kHz.

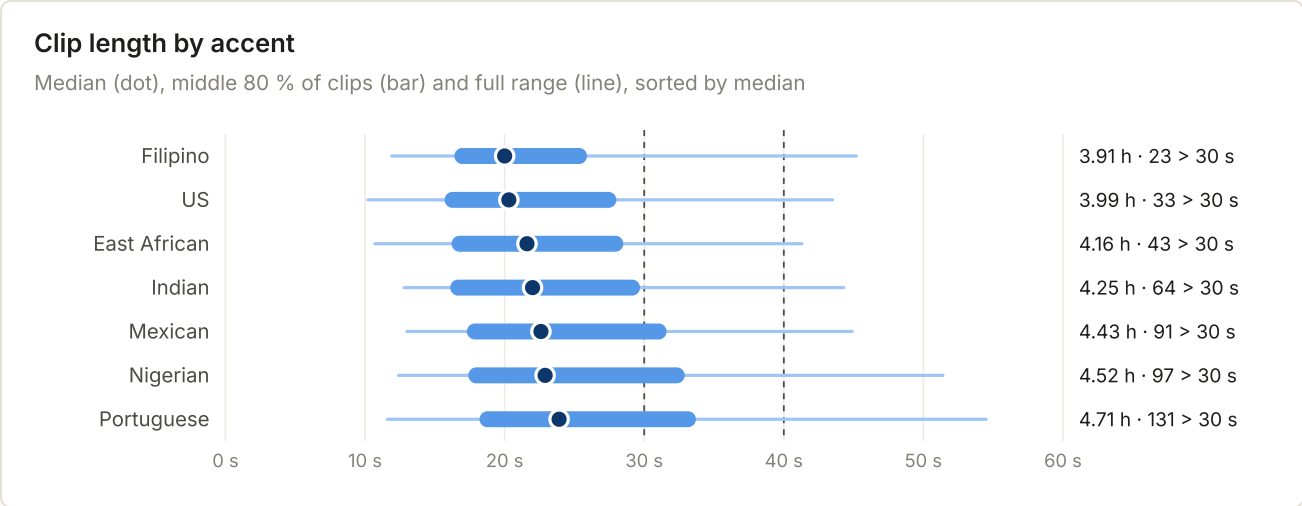

Figure 6. Clip durations per accent. Dashed lines mark 30 s and 40 s; the right column gives each accent's total audio and its clips over 30 s. Accents differ by up to 4 s in median length for the same sentences.

Speakers. Speakers were hired for the project and screened first; only those who qualified recorded the set.

2.3 Audio preparation

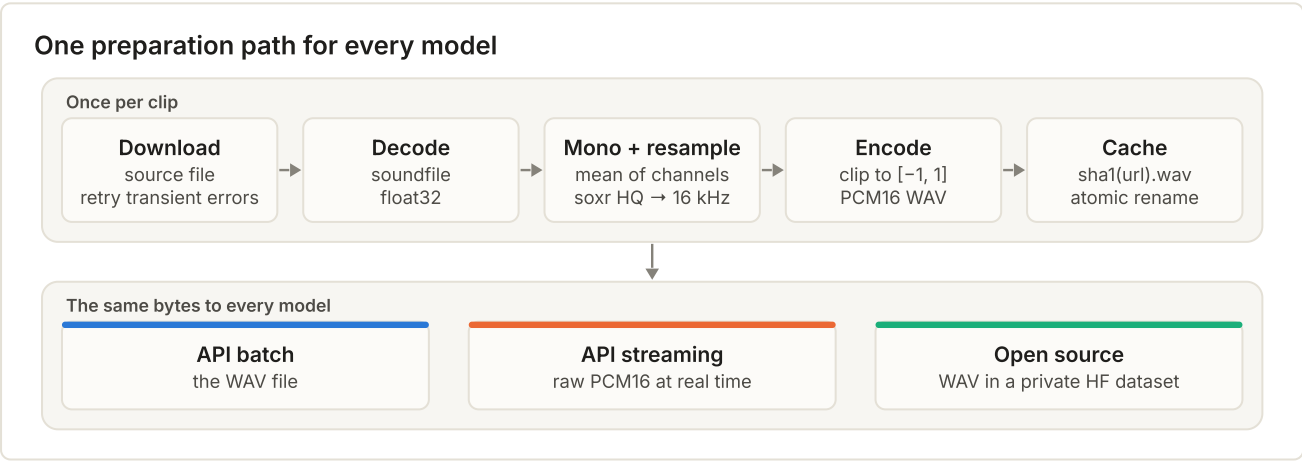

Figure 7. Each recording is fetched once, converted once and cached, so every model reads the same bytes.

- **Download:** transient errors are retried; a 403 or 404 is a permanent failure.
- **Convert:** decoded with `soundfile` to float32, channels averaged to mono, resampled to **16 kHz** with `soxr` (default “HQ” quality), clipped to [−1, 1] and written as 16-bit PCM WAV. This matches the Open ASR Leaderboard's `prepare_data`, which also feeds 16 kHz audio.
- **Cache:** stored as `audio_cache/<sha1(url)>.wav` through a temporary file and an atomic rename, so parallel processes never read a partial file.
- **Deliver:** batch APIs receive the WAV file, streaming APIs the same samples as raw PCM16, and open-source models the same WAV bytes through a private dataset (section 7.2). The one exception is the OpenAI Realtime API, which only accepts 24 kHz PCM: for its two models the 16 kHz signal is upsampled with `soxr`, which adds no information.

2.4 Two dataset samples

Two clips from the set, with every model's raw output, show what the benchmark is built to expose. The counts are the benchmark's own verdicts (section 10): an entity counts when its value or spelling is right in any accepted form, not only when it matches the prompt character for character.

Sample 1 — a healthcare sentence with a name

S3542 | US English | Healthcare | 23.6 s | 5 entities

REFERENCE (THE PROMPT AS WRITTEN)

Dietitian Irene Chhabra will hold a video visit with the client at 7362 Santell Lane, Centerville, PA from her Novato, CA office at 2:25 PM, and the fluid target under review is 660 mL per meal.

Entity	Type	Models correct, strict	Relaxed
Irene Chhabra	Person names	17 of 90	88 of 90
7362 Santell Lane, Centerville, PA	Street addresses	1 of 90	9 of 90
Novato, CA	Places	87 of 90	87 of 90
2:25 PM	Dates and times	84 of 90	84 of 90 (scored strictly)
660 mL	Amounts and quantities	79 of 90	79 of 90 (scored strictly)

“Novato, CA” and “2:25 PM” are easy (87 and 84 of 90 correct), but the name came back in 7 different forms. 17 of 90 were scored correct; the most common errors were “Irene Chabra” (69), “iren chabra” (1), “Irene Chabbra” (1). Under the relaxed metric (section 10.5), which accepts reviewed sound-alike spellings of made-up names, “Irene Chabra”, “iren chabra” and “Irene Chabbra” count: 88 of 90 models are correct. The hardest of the other entities was “7362 Santell Lane, Centerville, PA”: 1 of 90 correct.

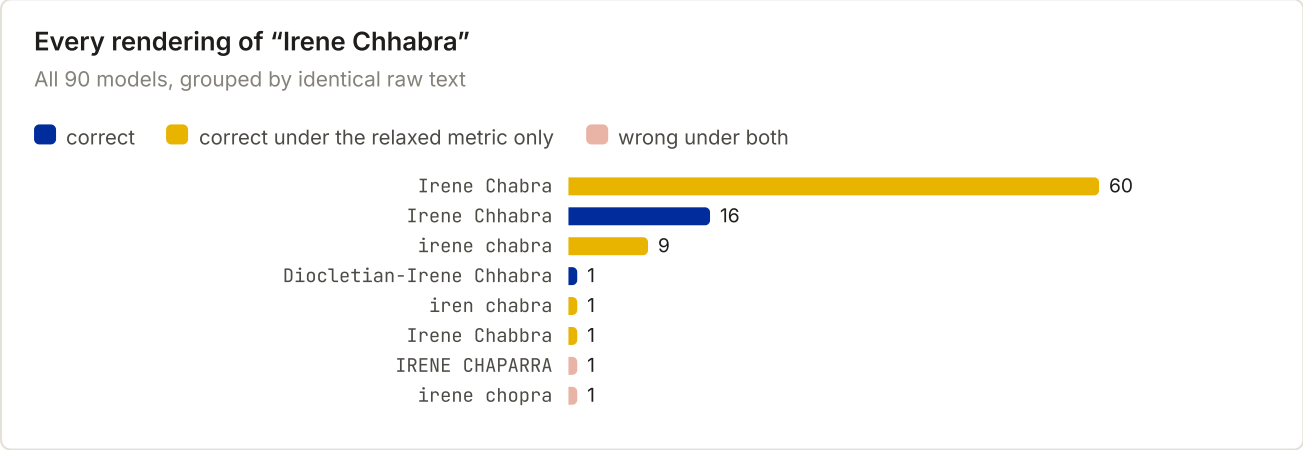

Figure 8. How the 90 models wrote the name in sample 1.

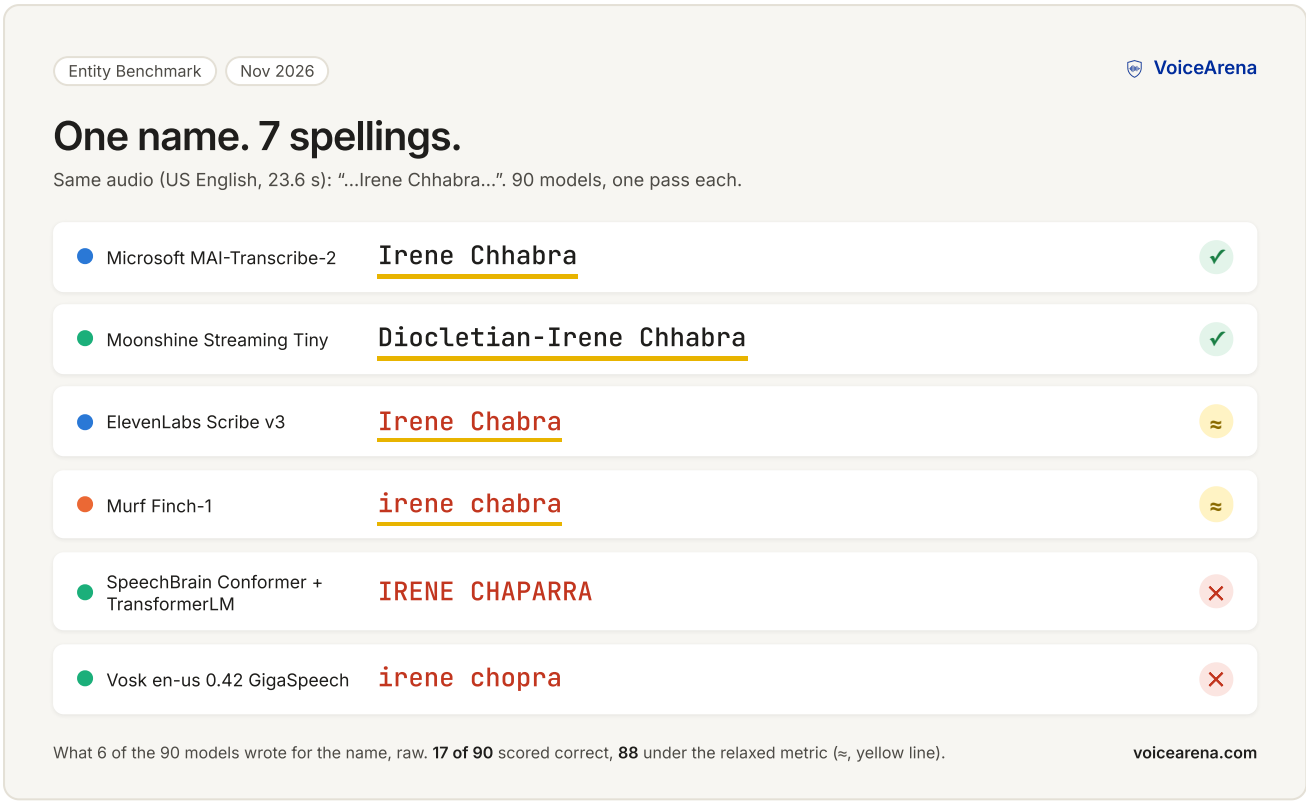

Figure 9. Sample 1 as it would be published: a selection of what the models wrote for the name.

Sample 2 — a finance sentence with an email address

S1656 · Filipino English · Finance · 22.1 s · 5 entities

REFERENCE (THE PROMPT AS WRITTEN)

For compensation report GSJY covering employee group 54734, the stock options were valued using the Black-Scholes model, and anyone with questions can write to faron@larkmail.us or speak directly with Delaney Ipina in human resources.

Entity	Type	Models correct, strict	Relaxed
GSJY	Codes and IDs	84 of 90	84 of 90 (scored strictly)
54734	Digit sequences	80 of 90	80 of 90 (scored strictly)
Black-Scholes model	Technical terms	57 of 90	57 of 90
faron@larkmail.us	Email addresses and URLs	46 of 90	72 of 90
Delaney Ipina	Person names	21 of 90	80 of 90

"GSJY" and "54734" are easy (84 and 80 of 90 correct), but the email address came back in 21 different forms (counting "at" and "dot" read out as @ and .). 46 of 90 were scored correct; the most common errors were "farron@larkmail.us" (21), "faron@larkmailus" (4), "faron@larcmail.us" (2). Under the relaxed metric (section 10.5), which accepts reviewed sound-alike spellings of made-up names, "farron@larkmail.us" and "faron@larcmail.us" count, but not "faron@larkmailus": 72 of 90 models are correct. The hardest of the other entities was "Delaney Ipina": 21 of 90 correct.

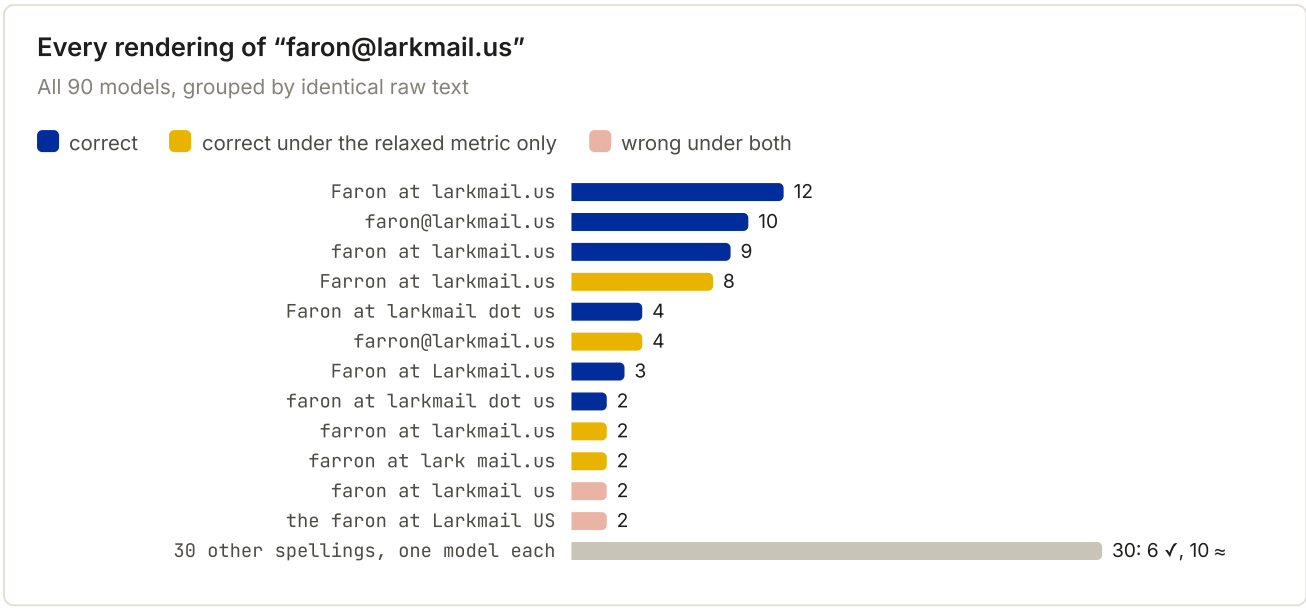

Figure 10. How the 90 models wrote the email address in sample 2.

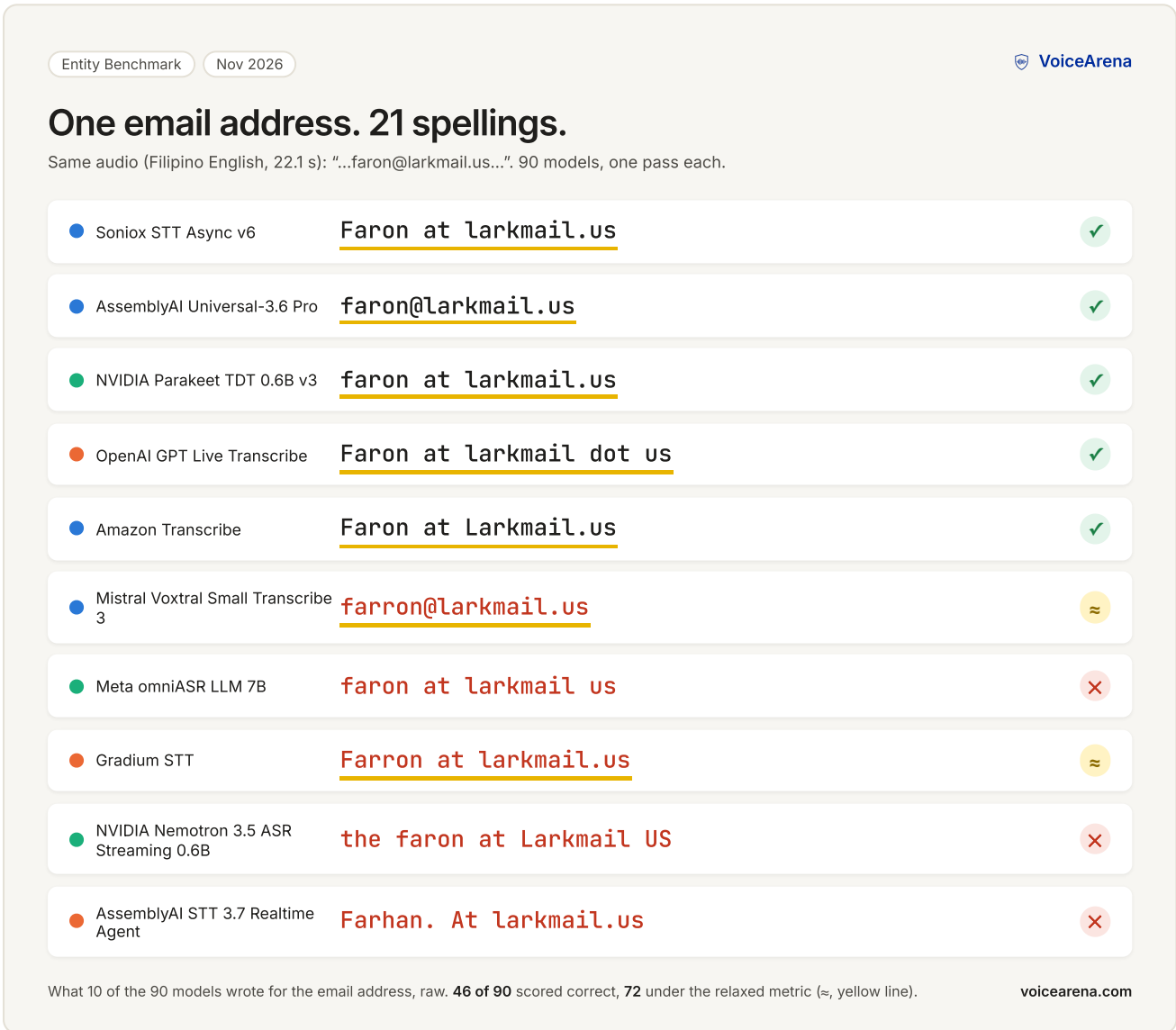

Figure 11. Sample 2 as it would be published: a selection of what the models wrote for the email address.

3 Models evaluated

The benchmark covers **90 current English speech-to-text models**: 57 API models from 23 providers and 33 open-weights models. A model whose API serves the same weights as an open-source release is evaluated once, as open source, under the leaderboard's settings.

29
API batch models; 12 of them through OpenRouter

28
API streaming models over WebSocket

33
open-weights models

23
API providers

Each API model is called the way a customer would call it: batch models receive the whole file in one request, streaming models receive audio over a WebSocket at real time. Where a model is listed on OpenRouter's speech-to-text catalogue, it was called there, with one request format for all of them; the others use the provider's own API. Open-source models run with the Open ASR Leaderboard's own evaluation code and container images (section 7); only the top Whisper model (large-v3) is kept, so the Whisper family does not crowd the table. Appendix A lists every model.

4 Software and hardware

API models run from a single machine and need only bandwidth: a streaming session sends 32 kB of audio per second for as long as the clip lasts. Open-source models run on Hugging Face Jobs: one NVIDIA H200 (141 GB) per job, and 32 vCPUs for Vosk, which runs on CPU.

Component	Version
API harness	Python 3.12, aiohttp 3.14.5, websockets 17.4, numpy 2.5.4, soundfile 0.14.0, soxr 1.1.0, msgpack 1.2.4, tqdm 4.70.1
Dataset and job tooling	huggingface_hub 2.4.0; datasets 3.6.0, the version the leaderboard images use, so the dataset loads in every image
Open ASR Leaderboard code	`huggingface/open_asr_leaderboard` at commit **859159c** (12 Nov 2026); **c469619** for the two models added on 20 Nov, whose launchers came later. The launcher refuses any other commit unless told otherwise
Open-source environments	The leaderboard's own Hugging Face Space images, one per model family

5 The API harness

Every API model is one self-contained script, `api/<model>.py`: the model-specific request at the top, and below it a runner that is **identical in every script** (checked by hash). One model's request can be audited without reading anything else, and a runner fix reaches every model at once.

batch **One request per clip**

The 16 kHz WAV in one HTTP request (multipart, raw body or base64, as the API requires). Asynchronous APIs take an upload and a job, are polled, and the upload is deleted afterwards.

streaming **One WebSocket session per clip**

Raw PCM16 on an absolute real-time schedule, in 100 ms chunks unless the provider documents another size; the session ends with the provider's end-of-audio message and counts only when its closing event arrives.

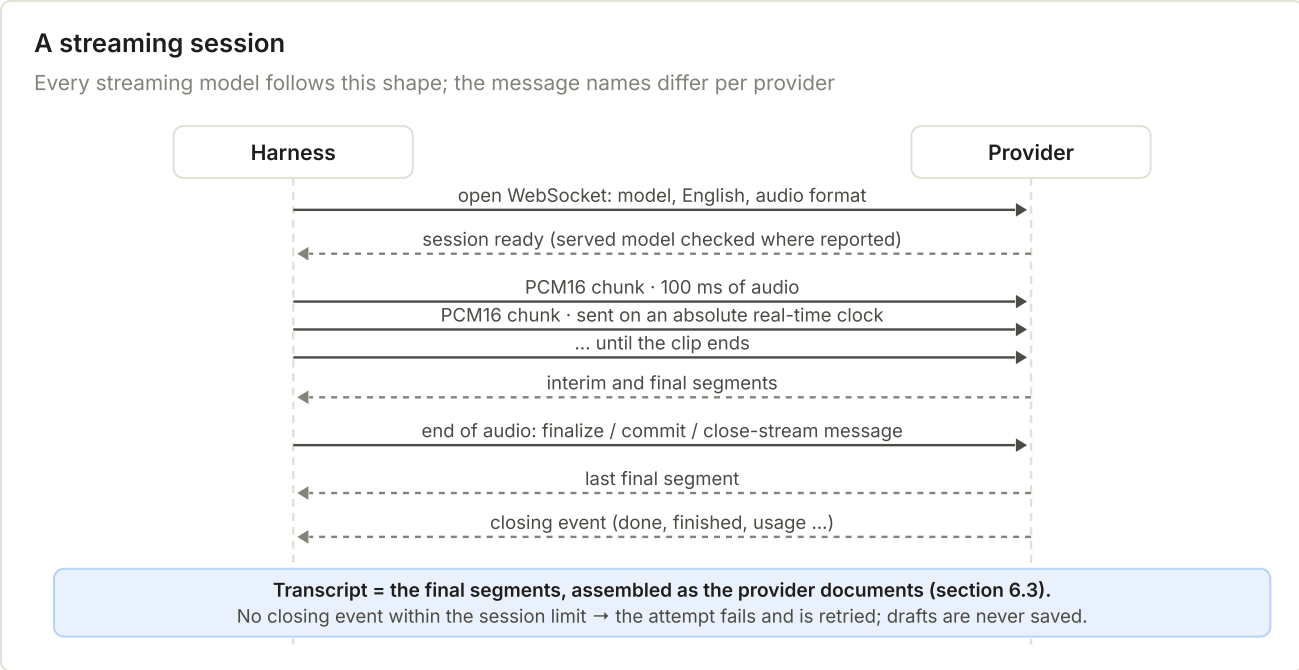

Figure 12. Streaming protocol, simplified.

- **Runner:** one process per model with an asyncio worker pool; the first clip runs alone, so a bad key or parameter fails once; session limit clip length + 60 s.
- **Retries:** 408, 409, 425, 429, 5xx and network errors are retried up to 7 times with exponential back-off (60 s cap, jitter), and a 429 pauses the whole pool. 401, 402 and 403, or 30 failures in a row, stop the run. Any other 4xx is recorded as an error for that clip.
- **Nothing finished is lost:** each clip is appended and `fsync`ed as it finishes; reruns skip successes; files are rewritten through an atomic rename; one writer per model (tested with `kill -9` and concurrent runs).
- **Nothing wrong is saved:** a failure is never stored as a transcript; a streaming session without the provider's final result is retried; where the API reports the serving model it must match the one requested; `check_results.py` flags empty, short, long and degraded rows, and only provider-degraded rows are resent.

6 API models: settings

The rule for every API model: **set the model, English and the audio format, and leave everything else at the provider's default**, as the Open ASR Leaderboard does. A setting differs from that rule only where the API requires it or the clip could not otherwise be transcribed whole; each case is listed below with its reason. Each script's docstring gives its endpoint and full request.

6.1 How English was set

Method	Models	Which, and how
language = en	42	Every other API model: `language`, `language_code`, `language_codes` or `languages` set to English
locale en-US	6	Amazon Transcribe (batch, streaming, Medical batch and streaming) and Zoom Scribe require a locale; Gemini Live takes BCP-47 codes (the Hugging Face Gemini provider maps `en` to `en-US` too)

Method	Models	Which, and how
English-only restriction	2	Soniox Async and Real-time: `language_hints=["en"]` with `language_hints_strict=true`, Soniox's documented way to restrict recognition to English
English-only model	2	Deepgram Flux English and Modulate Velma-2 English have no language parameter; the model or endpoint is English-only
hint or bias only	4	Speechmatics Melia 1 and Oak 1 accept only `language="multi"`, so English is a hint; Deepgram Flux Multilingual takes `language_hint=en`; Meta Muse Realtime takes `languageBias=["English"]`
auto-detect	1	Fish Audio Transcribe Stream 1 has no language parameter

6.2 Settings that differ from the defaults

Model	Setting	Reason
Amazon Transcribe Medical (batch, streaming)	`Specialty=PRIMARYCARE`, `Type=DICTATION`	Required fields. PRIMARYCARE is the only specialty valid for both modes; DICTATION means one speaker, as in this data. The batch Medical API returned the same text as standard Amazon Transcribe for 99.8% of clips
Amazon Transcribe and Medical, batch	Audio uploaded to a private S3 bucket and deleted after the job	The batch API only reads from S3; Medical also requires an output bucket, whose file is deleted too
Omi General-1, Medical-1, Medical Edge-1	`response_format=json`, `dictionary=false`	The default `diarized_json` only adds speaker structure; the default dictionary merges terms stored on the API key, and Omi recommends `false` for evaluation
OpenAI GPT Live Transcribe, GPT Realtime Whisper	`turn_detection: null`; audio upsampled to 24 kHz	Each clip is one turn, committed by the client at the end; the Realtime API accepts only 24 kHz PCM. The server's `session.updated` echo confirmed the language was applied
Deepgram Flux (both)	Silence sent after the clip until Flux has decoded all of it (≤ 1.5 s), then `ForceEndTurn`	Without it the last words were cut (section 12)
ElevenLabs Scribe v3 Live	Last chunk sent with `commit=true`; waits up to 15 s for the commit	Finalises the tail of the clip
Cartesia Ink 2	`/stt/websocket`, one manual `finalize` at the end	`/stt/turns` would add Cartesia's own turn detection; the batch endpoint serves Ink-Whisper, which is Whisper weights
Rime STT	`outputContract=REVISED_HYPOTHESES`	Required: the API has no default. Rime STT has no model id
Zoom Scribe v2.1 Pro	Fast mode (one synchronous request per file)	Batch mode only reads from and writes to customer S3 storage
Smallest.ai Pulse Pro	Requests started 2.4 s apart	The plan allows 25 requests per minute

6.3 How a streaming transcript is assembled

Streaming APIs return interim drafts and final segments. Only finals are kept, joined the way each provider documents, and the session counts only when its closing event arrives; the per-model details are in each script's docstring.

6.4 Output formats to expect

Transcripts are stored exactly as returned. These defaults shape the raw text and need normalisation before scoring:

Model(s)	Raw output
Deepgram Nova-3 streaming, Murf Finch-1	Lowercase, no punctuation; Murf also spells out letters and numbers (both API defaults)
xAI Grok Voice Transcribe 2.0	`format=false` (default): numbers come out as words
Fish Audio Transcribe 1 Pro, Stream 1	A `<|speaker:N|>` tag in the default output; strip it before scoring
Deepgram Nova-3 (OpenRouter)	Formatted by OpenRouter's integration, unlike the direct streaming call
Several open-source models	Normalised-style text: lowercase and unpunctuated (MMS, SpeechBrain in capitals, omniASR, Granite, Vosk)

7 Open-source models

7.1 Approach

Open-source models were run with the **Open ASR Leaderboard's own setup**, so their settings are the leaderboard's. Each model runs on Hugging Face Jobs in its family's leaderboard Space image. The leaderboard's eval script and `normalizer` package are injected at commit 859159c (c469619 for the two models added later), exactly as its `submit_jobs*.sh` scripts do by default, and each model takes the arguments of its own submit script (extracted into `hf_jobs/models.json` with the source file and line numbers). Where a submit script sets a value per dataset, the short-form default was used. **Only the dataset changes.**



Figure 13. From the private dataset to verified per-accent results. Each HF Job runs the leaderboard's own image and eval script with the arguments of the model's submit script.

7.2 Data handling and jobs

- The prepared WAVs (section 2.3) are in a **private** Hugging Face dataset, one config per accent plus an `all` config; the launcher refuses to start unless the dataset and the results bucket are private. Every config was read back after upload (rows, unique ids, audio length).
- Fast models run one job on the `all` config (10 models); slow models, and scripts that write no clip id, run one job per accent (23 models). Each model first ran a 2-clip smoke test, and a relaunch skips any job already done or running.

7.3 Changes from the leaderboard defaults

Model(s)	Change	Reason
Whisper large-v3	Leaderboard **long-form** settings (`--longform` , batch size 32, from `submit_jobs_whisper_longform.sh`)	The short-form path pads or truncates to 30 s without warning, and 10.2% of clips are longer
MMS-1b-all	One-line patch to the injected `run_eval.py` : `gen_kwargs = {}`	At 859159c CTC models crash with `NameError: gen_kwargs`
Nemotron 3.5 Streaming 0.6B, Phonon-2	Batch size 64 → 16	At 64 the 10–60 s clips overflow PyTorch's 32-bit conv2d indexing. Batch size changes speed only
SpeechBrain conformer + TransformerLM	Batch size 8	Not in the leaderboard scripts; 32 (by analogy) ran out of memory on long clips
Granite 5.0 2B, 470M TurboCTC	Inject `run_eval_speculative_bpe.py` / `run_eval_ctc.py`	As the granite submit script does; the CTC script is not baked into the image
ASR4ALL L / M, Granite TurboCTC	One job per accent; rows matched by reference text	These scripts write `sample_N` instead of clip ids and sort by duration. Texts are unique within an accent, so each row matches exactly, and the duration check confirms it
ARK-ASR 3B, omniASR LLM 7B and CTC 7B	**Kept** the leaderboard's 40 s input limit	The leaderboard runs them this way; the 54 clips over 40 s lose their ending for these three models
Parakeet Ultra, Fun-ASR MLT Nano, SenseVoice Small, Vosk	Our runner, following the model card	Not supported by, or not in, the leaderboard repository; details in `hf_jobs/models.json`

8 Running the evaluation

8.1 What was captured

Each model writes `<model>/<accent>.jsonl` and `.csv` , one record per clip, with the Open ASR Leaderboard manifest fields (`audio_filepath` , reference `text` , raw `pred_text` , `duration`) plus `sentence_id` , `domain` , `accent` , `status` and `error` ; API rows also keep the retry history and provider metadata. `time` is wall time (streaming includes real-time playback; open-source rows carry a share of batch time), not latency.

8.2 Coverage

90
models complete (every model has at least 4,738 of 4,739 transcripts)

426,508
of 426,510 transcripts produced

2
clips failed on the provider side every time

34
empty transcripts, kept as returned

Every model in the combined table has a transcript for every clip except these 2, which failed on every attempt on the provider's side. They stay in the results as `status=error` with the reason, and are not filled with interim text.

Model	Clips	Missing clips	Reason
Deepgram Flux English	4,738	one East African clip	A turn never receives `EndOfTurn`, on every attempt
Rime STT	4,738	one Portuguese clip (a re-recording, 24.4 s, 45 words)	`resource_exhausted`: the transcript is cut by the model's generation token cap

9 Verification

A transcript enters the combined table only if it is provably the output for its clip.

API models

- Every clip is `ok` with a transcript, or `error` with its reason.
- Errors are retried until they succeed or are shown to fail on every attempt on the provider's side (section 8.2).
- `check_results.py` reports errors, pending, empty, short, long and provider-degraded rows per model.
- Every model's reference text matches the prompts CSV; 4,739 rows each, no duplicates.

Open-source models

- **Ids:** every clip id appears exactly once, and every clip in the prompts CSV is present.
- **Text:** the manifest's reference equals the CSV sentence.
- **Duration:** equals our 16 kHz file within 0.05 s; with the reference-text check, this shows the model received that clip.
- All 156,387 cells re-checked after download.

Combined table

- Built by `combine.py` from every results folder, one column per model.
- Refuses a model name that appears in two folders.
- Checked cell by cell against the separate API and open-source tables: identical.

10 Scoring and ranking

Every clip is scored on the entities its sentence was written around. A transcript gets credit for an entity when it contains the entity's exact value or spelling, in any written or spoken form a reader would recover it from. Scoring is **map-only**: an entity counts only when the transcript contains a rendering that was verified and entered in the entity map. The map was built and checked with rules in code; no language model is involved, and the same transcripts always give the same scores.

2,745
entities in 677 sentences, 3–6 per sentence

19,215
entity instances per model: each entity in its 7 clips

10
entity types, each with its own matching rules

225
test cases that fix what is accepted and rejected

10.1 Metrics

Metric	Answers	Computed as
Task success rate (TSR) ranking metric	How many clips need no correction	Clips with every entity correct ÷ 4,739
Entity accuracy	How many entities the model gets right	Correct entity instances ÷ all instances (19,215 per model), pooled over every clip and accent

Metric	Answers	Computed as
TSR and entity accuracy, relaxed	The same, when a sound-alike spelling of a made-up name, email or street name is not held against the model	As above, also accepting the reviewed sound-alike renderings of section 10.5; never used for ranking
WER	Word errors in the whole transcript	Substitutions + deletions + insertions ÷ reference words, summed over all clips, after the normalisation in 10.3
CER	Character errors in the whole transcript	As WER, on characters
WER (HF normaliser only)	WER as the Open ASR Leaderboard computes it	As WER, without the entity step of 10.3

How each number is computed

Entity accuracy	$\mathrm{EA} = 100 \times \dfrac{\sum_{c \in C} \sum_{e \in E(c)} [e \text{ correct in } c]}{\sum_{c \in C} \lvert E(c) \rvert}$	C: the clips in scope (all 4,739, or the 677 of one accent). E(c): the entities of the clip's sentence, or only those of one type or tier for a breakdown. The denominator is 19,215 over all accents and 2,745 per accent.
Task success rate	$\mathrm{TSR} = 100 \times \dfrac{1}{\lvert C \rvert} \sum_{c \in C} [\text{all of } E(c) \text{ correct}]$	The ranking metric. A clip with no transcript counts as a failure.
WER	$\mathrm{WER} = 100 \times \dfrac{\sum_c (S_c + D_c + I_c)}{\sum_c N_c}$	S, D, I: word substitutions, deletions and insertions between the normalised reference and transcript; N_c: words in the normalised reference. Summed over the corpus, so a long clip weighs more than a short one.
CER	$\mathrm{CER} = 100 \times \dfrac{\sum_c (S^{ch}_c + D^{ch}_c + I^{ch}_c)}{\sum_c L_c}$	The same on characters; L_c: characters of the normalised reference, spaces included.
Normalisation	$\mathrm{ref}' = H(r), \quad \mathrm{hyp}' = H(\Phi(\mathrm{clean}(h)))$	H: the Open ASR Leaderboard normaliser; Φ: the entity step (10.3); clean: model artifacts removed. WER (HF normaliser only) uses hyp′ = H(clean(*h*)).
Accent gap	$\mathrm{Gap} = \max_a \mathrm{EA}_a - \min_a \mathrm{EA}_a$	Over the seven accents *a*, for one model.
All models pooled	$\mathrm{EA}_{\mathrm{pooled}} = \dfrac{1}{90} \sum_m \mathrm{EA}_m$	Used in the bar charts. Every model is scored on the same instances, so pooling the instances equals the mean of the models.
95% interval	$\mathrm{TSR}^*_b = 100 \times \dfrac{\sum_s w_{s,b} k_s}{\sum_s w_{s,b} n_s}$ $\mathrm{CI} = [\, P_{2.5}, P_{97.5} \,] \text{ of the } \mathrm{TSR}^*_b$	For each of 2,000 resamples *b*, 677 sentences are drawn with replacement; $w_{s,b}$ is how often sentence *s* was drawn, k_s and n_s its fully correct and total clips over its seven accents. Entity accuracy uses the same formula on entity instances. The rank interval takes the same percentiles of the model's TSR rank in each resample.

Example. A clip whose sentence has 5 entities, of which a model gets 4 right, adds 4 to the numerator and 5 to the denominator of entity accuracy, and 0 to the task success rate: one wrong entity fails the clip. [*x*] is 1 when *x* is true and 0 otherwise.

- **Ranking.** Models are ranked by task success rate (TSR) over all accents, then by entity accuracy; identical scores share a rank. Every accent contributes the same 677 sentences, so pooling weights the accents equally.
- **Uncertainty.** 95% intervals come from 2,000 bootstrap resamples of the 677 sentences. The seven readings of a sentence are resampled together because they share their entities. The same resamples give each model a rank interval (in the results CSV).

10.2 When an entity counts as correct

The rule: formatting may differ; the value and the spelling may not. A rendering counts when a reader of the transcript would recover exactly the entity in the prompt: the same digits, the same letters, the same date, whether it is written as digits or words, spelled out, abbreviated or punctuated differently.

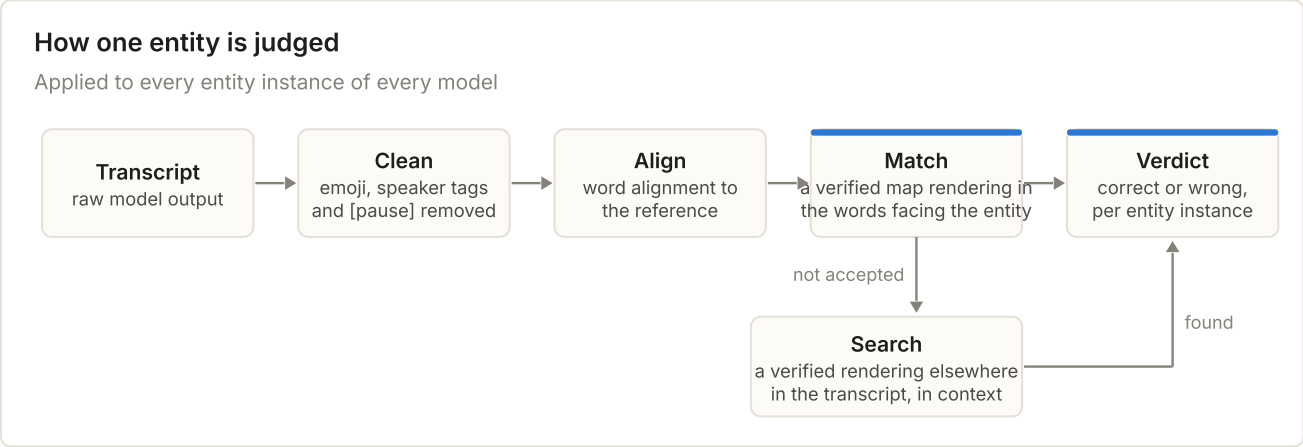

Figure 14. The transcript is aligned word by word to the reference; the words facing an entity form its span, which counts when it contains a rendering listed in the entity map for that entity. If it does not, a listed rendering found elsewhere in the transcript still counts when it sits in the right context.

The map is built with rules: each entity type has its own comparison key (digit strings for codes, letters for names, a value and unit for amounts, a normalised date and time, and so on), and a rendering is a candidate when its key matches the entity's. What the rules accept and reject:

Entity type	Accepted (same value or spelling)	Rejected (changed value or spelling)
Person names	Case, punctuation and accents; a split at a capital (“Mac Kenzie”); letter-by-letter spelling; a dropped possessive ’s	Any different letter, including a common spelling of an unusual name (“Augustine” for Augustin); an added apostrophe; one name split in two
Brands and organisations	“&” as “and”; standard abbreviations both ways (Inc, Co, St, Mt); singular or plural of the head noun	Any different letter; a British spelling of a proper noun
Places	Full state name, two-letter code, letter-spaced code or AP abbreviation (“Calif.”); Saint/St, Fort/Ft, Mount/Mt	A different state; a missing or extra part
Technical terms	British or American spelling; open or closed compounds of common words (“wait list”, “waitlist”); numbers as words; “+” as “plus”	A different word; a plural of an all-capitals acronym
Street addresses	USPS abbreviations both ways (Street/St, Apartment/Apt, North/N); any reading of the house number or ZIP whose digits are right (“forty-seven oh six” for 4706); “in” before the city; the street name written as one word or two (“Hillway” for Hill Way)	Any part missing or wrong; a house number written as a time, decimal or amount
Email addresses and URLs	Symbols spoken as words (“at”, “dot”, “underscore”, “dash”, “slash”); spaces; letter-by-letter spelling; numbers as words	A missing or extra dot, underscore or hyphen; any different letter

Entity type	Accepted (same value or spelling)	Rejected (changed value or spelling)
Codes and IDs	Separators ignored; digits as words; “double five”; the letter O and the digit 0 treated alike	A character missing, extra, changed or out of order; a currency sign in front
Digit sequences	Any grouping of the digits, as digits or words (“five four six nine”, “fifty-four sixty-nine”)	A missing or extra digit; a match that stops inside a longer number
Amounts and quantities	Any reading with the right value (“seven hundred forty-seven dollars” for $747); “point”; currency and units abbreviated or in full	A different value; a missing or different unit or currency; slang units (“gig”)
Dates and times	Cardinal or ordinal day; spoken years; “:00” optional; any am/pm spelling; day before month	A different day, month, year, weekday, hour or minute; a missing part

The map. Every rendering the rules accepted across the models of the first release was reviewed line by line, entity by entity, and entered in the map (3 were overruled and are listed as rejections). Renderings that appeared only later, from AssemblyAI STT 3.7 Realtime Agent, FluidVoice Fluid Speech Pro, Salesforce SOV-Transcribe-1 and the re-recorded clips, were checked against the same rules and added (1,666 renderings), and so was each entity's own spelling where no model had produced it (131). The map now holds 36,337 accepted renderings for the 2,745 entities. A rendering that is not in the map counts as wrong until it is verified and added, so a score can change only through a reviewed map entry; 225 test cases fix the rules used to build it.

10.3 Normalisation for WER and CER

1. **Model artifacts removed:** emoji, speaker tags such as `<|speaker:0|>` and bracketed events such as `[pause]`.
2. **Entity step:** every correctly rendered entity is replaced by its reference form, so a correct entity costs no word errors whatever form the model chose.
3. **Open ASR Leaderboard normaliser:** its `EnglishTextNormalizer` (commit `859159c`, unchanged) runs on reference and transcript: lowercasing, removal of bracketed text and fillers, British to American spelling, number words to digits, standard currency, punctuation and most symbols removed.
4. **Counting:** substitutions, deletions and insertions with `jiwer`, summed over the corpus.

Without the entity step

`gascon@larkmail.us` → `gascon larkmail us`
`gascon at larkmail dot us` → `gascon at larkmail dot us`
2 word errors for a correct email address.

With the entity step

`gascon at larkmail dot us` → `gascon@larkmail.us` → `gascon larkmail us`
identical to the reference after normalisation: **0 word errors**.

The leaderboard normaliser already maps spoken numbers, amounts and spelled codes to the written form, so the entity step matters most for emails, URLs, names with spacing variants and codes with separators. It lowers WER by 2.2 points on average (0.8–4.3 by model); both versions are reported.

10.4 Edge cases

- **No transcript, no credit.** Every entity in a clip with no transcript or an empty one counts as wrong. This covers the 2 clips that failed on the provider's side (Deepgram Flux English: 1; Rime STT: 1) and 34 empty transcripts.

10.5 Relaxed metric: sound-alike spellings

Many names, emails and street names in the set are made up, and real brands and places can share a sound with a more common spelling (Wisper ISP, Holladay); a speaker reading *Mathew*, *kalm.us* or *Wisper* gives the model no way to know the letters. Every results table therefore also reports a **relaxed** entity accuracy and TSR, which also accept a spelling that **sounds the same**. Rankings, WER and CER always use the strict rules.

Where it applies

Entity	May be spelled as it sounds	Must stay exact
Person names	every word of the name	n/a
Emails and URLs	the letter parts between @ . _ - / and digits (user name, domain words, path words)	symbols @ . _ - /, digits, the ending (.com, .us, .edu), `www`, and a label of up to four letters before .edu / .gov (fdu, wvu, va), which is spelled letter by letter
Street addresses	the made-up street name (one that is not all dictionary words) and, in this release, a city differing only in spelling	house number, direction, suffix, unit, state and ZIP; a common-word street (Maple, Cedar, Hill) stays strict
Brands, places and technical terms	the accepted spelling written with different spacing or punctuation (globelife, oakridge), and reviewed spellings that sound identical (whisper for Wisper ISP, rico for Ricoh, holiday for Holladay, UT, fair class for fare class)	**medical names**, always exact: drugs, conditions and procedures, and pharma, biotech and medical companies (61 entities); and every code, number, amount and date

How the scorer compares sounds

Relaxed scoring is map-only, like strict. A rendering counts when it is a strict map entry or an entry of the relaxed map (32,988 renderings), matched ignoring spacing and punctuation. The entries were proposed by the rule in `relaxed_keys.py` and reviewed before they were entered. The rule reduces both spellings to a consonant skeleton and compares the skeletons:

#	Step	Example
1	Lower case; accents and punctuation removed; a name is compared word by word and also joined, so spacing does not matter	julianbola = Julian Bhola
2	*th* kept as one sound, never t + a dropped h	Mathew: m-T
3	Same-sound spellings merged: ph → f, ck / q → k, x → ks, c → s before e/i/y and k otherwise, z → s, tch / cch → ch, silent kn- / wr-, silent h after b d g j k p, ee / ea / ie / ei / ey / y → i, oo → u, aa → a, a final h after a vowel and a final e after a consonant (words of 4+ letters) dropped, doubled letters single	pham = fam, kalm = calm, Bhola = bola, russel = russell
4	A w after a vowel is part of the vowel (aw, ow, ew)	yawlmail = yolmail
5	All vowels removed (a e i o u y): what is left is the consonant skeleton	Holly → h-l, holy → h-l
6	Two renderings match when their skeletons are identical; voiced/unvoiced pairs (t/d, b/p, f/v, g/k, j/ch) are **not** merged automatically, and each such pair was reviewed: accepted only where both can sound the same	Nakada = nakata (reviewed: a flapped t); Vasudevan ≠ vasudifan (v ≠ f)
7	A bare letter is not a rendering of a word	"sammy k" for Kha is wrong

Worked examples from the set, with skeletons computed by the scorer's own code (one group per word or email part; C = ch, T = th). *Rule* is the automatic proposal; *Relaxed* is what the scorer counts, from the map.

Entity	Skeleton	Written	Skeleton	Rule	Relaxed
Holly Muramoto	`hl-mrmt`	holy muramoto	`hl-mrmt`	✓ same	✓ counts
Elysia Hilson	`ls-hlsn`	alicia hilson	`ls-hlsn`	✓ same	✓ counts
stieg@quailpost.com	`stg-klpst-km`	stig@quailpost.com	`stg-klpst-km`	✓ same	✓ counts
posada@yawlmail.com	`psd-lml-km`	posada@yolmail.com	`psd-lml-km`	✓ same	✓ counts

Entity	Skeleton	Written	Skeleton	Rule	Relaxed
latonya@kalm.us	`ltn-klm-s`	latonya@calm.us	`ltn-klm-s`	✓ same	✓ counts
Zenaida Nakada	`snd-nkd`	zenaida nakata	`snd-nkt`	✕ differ	✓ counts (reviewed entry)
stieg@quailpost.com	`stg-klpst-km`	steve@quailpost.com	`stv-klpst-km`	✕ differ	✕ wrong
Owen Richarson	`n-rCrsn`	owen richardson	`n-rCrdsn`	✕ differ	✕ wrong
Karime Vasudevan	`krm-vsdvn`	karime vasudifan	`krm-vsdfn`	✕ differ	✕ wrong
posada@yawlmail.com	`psd-lml-km`	posada@yourmail.com	`psd-rml-km`	✕ differ	✕ wrong
fdu.edu	`fd-d`	fddu.edu	`fd-d`	✕ exact part	✓ counts (reviewed entry)

Still wrong under relaxed

Difference	Examples
Any other consonant change	Richarson ≠ richardson (an added d), yawlmail ≠ yourmail (r added)
A voiced/unvoiced swap that sounds different	Vasudevan ≠ vasudifan (v and f differ)
A bare letter as a word	“sammy k” for Kha
Exact parts	an email's symbols, digits, ending, www and acronym labels (fdu.edu ≠ fddu.edu, larkmail.us ≠ larkmail.com); an address's number, direction, suffix, unit, state and ZIP

- **Vowels are ignored on purpose.** The speaker's vowel cannot be checked without listening, so the relaxed metric does not judge vowels at all. That makes it consistent and auditable, but generous: some vowel mishearings (holy for Holly, stag for Stieg) count. It is also blind to accent: the same entry counts for every speaker. These are the reasons the ranking uses strict TSR.
- **What this release added to the relaxed map.** The strict-map additions of section 10.2 for names, emails and street addresses (1,025 entries), so a correct spelling that differs only in spacing or punctuation also counts; 24 reviewed sound-alikes for two brands no model wrote correctly (swift for Swyfft, easy fiber for Ezee Fiber); for the 788 brands, places and technical terms that are not medical, every accepted spelling compared without spacing and punctuation (5,035 entries) and 35 reviewed spellings that sound identical, each written by at least a tenth of the attempts (whisper isp, pinnacle assurance, windham rewards, holiday ut, greenville tn, fair class), while near misses with a different vowel or word stay wrong (shore for Shure, louis for Lowe's, holy grove for Hollygrove, photon for futon); 4,288 names and emails that differ from an accepted spelling only in spelling or in symbols read out as words (Chabbra for Chabra, “at” and “dot” for @ and .); and 2,983 renderings of 163 addresses whose made-up street name has the same skeleton (Aristy = aristi, Rockow = rockau) or whose city differs only in spelling (Fennville = fenville, Foresthill = forest hill, Centerville = centreville), with everything else exact. Over all models, 128,202 entity instances are correct only under the relaxed metric.

11 Results

All 90 models are scored on the same 4,739 clips and 19,215 entity instances. Section 11.1 gives the leaderboard over all accents; 11.2–11.4 break it down by entity type, accent, sentence difficulty, tier and domain; 11.5 compares entity accuracy with WER; and 11.6 compares the public and private sets; section 1.2 has worked examples. Appendix B gives every model's TSR in each accent, and the website table and CSVs carry every number.

31.9%
task success rate (TSR), Zoom Scribe v2.1 Pro: best of 90

21.1%
best open weights: Mistral Voxtral Small 24B (#14)

16.1%
TSR of the median model

76.2%
entity accuracy, Zoom Scribe v2.1 Pro

11.1 Leaderboard

- **Zoom Scribe v2.1 Pro leads** with a task success rate of 31.9% (95% CI 29.3–34.4), 3.6 points ahead of ElevenLabs Scribe v2 Medical and 4.0 ahead of ElevenLabs Scribe v3, and ranks first in 100% of bootstrap resamples. Places 4–10 lie within 1.4 points of each other and their intervals overlap: treat them as a group.
- **Streaming:** the best streaming model, Google Gemini 3.5 Transcribe Live (#6, TSR 25.0%), is within 0.8 points of its batch version (24.2%). Across 15 batch and streaming pairs from the same provider, streaming changes TSR by −2.5 points at the median (figure below).
- **Open weights:** Mistral Voxtral Small 24B is #14 with a TSR of 21.1%, ahead of 44 of the 57 API models. On average, batch APIs reach a TSR of 20.0%, streaming APIs 16.3% and open-weights models 9.9%.
- **Added on 20 November 2026:** Salesforce SOV-Transcribe-1 is #20 overall (#2 of 33 open-weights models), TSR 19.3%, entity accuracy 67.1%; FluidVoice Fluid Speech Pro is #59 overall (#9 of 33 open-weights models), TSR 13.8%, entity accuracy 61.6%.
- **Spread:** TSR runs from 31.9% down to 0.3% (Meta MMS 1B); the median model gets every entity right in only 16.1% of clips. Counted entity by entity, the leader gets 76.2% right.
- **Relaxed metric** (section 10.5): accepting sound-alike spellings of made-up names, emails and street names, and of real non-medical brands, places and terms, raises TSR by 10.9 points and entity accuracy by 7.4 points for the average model (0.6–21.0 points of TSR by model). Zoom Scribe v2.1 Pro also leads on relaxed TSR (52.5%); the strict and relaxed TSR rankings correlate at ρ = 0.982, share 10 of their top 10, and no model moves more than 20 places.

Leaderboard

All accents · 90 models · ranked by task success rate (TSR), then entity accuracy · relaxed: sound-alike spellings also count · all values in %

#	SYSTEM	TSR	95% CI	TSR RELAXED	ENTITY ACC.	EA RELAXED	WER	CER	WER HF NORM.
1	Zoom Scribe v2.1 Pro	31.9	29.3–34.4	52.5	76.2	84.8	5.3	2.4	7.2
2	ElevenLabs Scribe v2 Medical	28.3	25.7–31.0	47.7	74.2	82.9	5.7	2.5	7.6
3	ElevenLabs Scribe v3	27.9	25.3–30.5	46.9	73.8	82.3	5.6	2.5	7.4
4	Mistral Voxtral Small Transcribe 3	25.6	23.2–28.1	43.1	71.9	80.5	7.0	2.9	9.6
5	Microsoft MAI-Transcribe-2	25.2	22.7–27.7	46.2	72.0	82.0	6.0	2.6	7.4
6	Google Gemini 3.5 Transcribe Live	25.0	22.5–27.5	40.5	71.6	79.4	7.0	3.3	8.8
7	AssemblyAI Universal-3.6 Pro	24.9	22.5–27.5	43.2	71.6	80.8	6.3	2.8	8.5
8	Mistral Voxtral Mini Transcribe 3	24.3	21.9–26.8	42.4	71.3	80.2	6.6	2.8	8.7
9	AssemblyAI STT 3.7 Realtime Agent	24.3	21.8–26.9	41.6	71.2	79.8	6.8	3.1	8.9
10	Google Gemini 3.5 Transcribe	24.2	21.9–26.8	39.8	71.4	79.4	6.8	3.3	8.6
11	Google Cloud Chirp 3	22.5	20.2–25.0	36.8	70.4	77.9	7.8	3.7	9.6
12	ElevenLabs Scribe v3 Live	21.9	19.6–24.4	38.3	69.4	78.1	7.5	3.4	9.4
13	OpenAI GPT Transcribe	21.6	19.4–23.9	36.6	69.0	77.2	8.8	4.2	10.8
14	Mistral Voxtral Small 24B	21.1	19.0–23.4	37.4	68.4	77.4	8.3	3.6	10.3
15	OpenAI GPT-4o Mini Transcribe	21.0	18.7–23.3	35.3	68.7	76.3	7.7	3.5	9.0
16	AssemblyAI Universal-3.6 Pro Realtime	20.9	18.7–23.3	35.2	68.3	76.5	8.3	4.0	10.4
17	Reson8 Rumble-2	19.7	17.5–22.0	35.3	67.2	76.2	9.2	4.3	11.1
18	Reson8 Rumble-2 Realtime	19.7	17.5–21.9	35.1	67.2	76.2	9.3	4.4	11.2
19	xAI Grok Voice Transcribe 2.0	19.4	17.2–21.5	32.1	66.3	74.0	11.2	5.4	14.8

#	SYSTEM	TSR	95% CI	TSR RELAXED	ENTITY ACC.	EA RELAXED	WER	CER	WER HF NORM.
20	Salesforce SOV-Transcribe-1	19.3	17.2–21.6	32.3	67.1	75.0	8.8	4.3	11.4
21	Fish Audio Transcribe 1 Pro	19.2	17.1–21.4	33.1	66.9	74.6	9.7	4.5	11.7
22	OpenAI Whisper large-v3	19.0	16.9–21.2	34.2	66.5	75.5	9.2	4.1	10.9
23	Reson8 Resonant-1	18.9	16.8–21.1	34.5	65.9	75.7	9.1	4.0	10.8
24	OpenAI GPT Live Transcribe	18.7	16.5–20.9	30.7	66.1	73.6	11.2	5.5	14.2
25	Fish Audio Transcribe Stream 1	18.5	16.4–20.7	31.0	65.9	73.5	9.2	4.3	12.1
26	Cartesia Ink 2	18.4	16.4–20.4	31.7	65.7	74.3	10.3	4.5	12.8
27	xAI Grok Voice Transcribe 2.0 Streaming	18.3	16.2–20.4	30.0	65.3	72.8	12.2	5.9	15.8
28	Soniox STT Real-time v6	18.3	16.1–20.5	32.7	65.4	74.5	9.9	4.5	11.5
29	Soniox STT Async v6	18.2	16.1–20.4	32.6	66.0	74.9	8.5	3.9	9.9
30	Omi General-1	18.1	16.0–20.3	30.8	65.4	73.4	9.4	4.4	11.4
31	Qwen3-ASR 1.7B	17.9	15.9–20.1	31.2	65.4	73.7	9.4	4.4	11.5
32	Omi Medical-1	17.5	15.5–19.6	30.3	64.6	72.6	9.9	4.7	11.6
33	Omi Medical-1 Flash	17.5	15.5–19.6	30.2	64.5	72.5	9.9	4.7	11.7
34	Amazon Transcribe	17.2	15.1–19.3	30.4	65.3	73.7	8.9	3.9	10.9
35	Amazon Transcribe Medical	17.2	15.1–19.3	30.4	65.3	73.7	8.9	3.9	10.9
36	Meta Muse Voice Transcribe Realtime	17.1	15.2–19.1	29.2	64.3	72.0	11.2	5.4	13.7
37	Mistral Voxtral Mini Transcribe Realtime 3	17.1	15.0–19.0	30.6	64.6	73.0	11.4	5.1	13.5
38	Speechmatics Enhanced	16.7	14.6–18.9	24.2	64.9	69.8	11.1	5.4	13.7
39	OpenAI GPT Realtime Whisper	16.7	14.7–18.8	28.1	63.8	71.3	11.6	5.6	13.6
40	Omi Medical-1 Live	16.7	14.7–18.7	27.9	63.1	70.6	11.2	5.5	13.0
41	Mistral Voxtral Mini 4B Realtime	16.7	14.8–18.6	29.5	62.2	70.4	15.2	8.1	17.2
42	Speechmatics Enhanced Realtime	16.6	14.5–18.7	23.8	64.7	69.6	11.4	5.5	14.1
43	Deepgram Nova-3	16.4	14.4–18.4	27.3	63.8	71.5	12.3	5.7	15.9
44	Meta Muse Voice Transcribe 1.0	16.3	14.4–18.2	29.0	64.3	72.1	11.2	5.4	13.5
45	Reson8 Resonant-1 Realtime	16.1	14.3–18.0	29.7	63.2	72.4	10.4	4.6	12.1
46	Smallest.ai Pulse Pro	16.1	14.2–18.1	28.3	64.1	72.0	13.3	7.3	16.6
47	Inworld STT-1	15.8	13.9–17.7	26.7	62.7	70.1	13.0	6.6	15.2
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	15.7	13.8–17.6	26.2	63.4	70.8	10.2	5.1	12.8
49	Musci-ASR qwen-1019	15.7	13.8–17.7	27.0	63.2	70.8	11.2	5.4	14.5
50	ARK-ASR 3B	15.7	13.8–17.6	28.4	62.7	71.4	12.2	5.5	16.0
51	Omi General-1 Live	15.1	13.1–17.0	25.5	62.8	70.4	11.4	5.4	14.0
52	Speechmatics Melia 1	14.4	12.7–16.3	25.2	62.2	70.1	11.5	5.4	13.7
53	Speechmatics Linden 1	14.4	12.5–16.3	20.1	61.7	66.1	13.6	6.5	16.1
54	Speechmatics Oak 1	14.3	12.5–16.2	24.8	61.9	69.7	11.7	5.5	14.0
55	Amazon Transcribe Streaming	14.2	12.3–16.1	23.1	62.5	68.9	11.3	5.0	13.3
56	Reson8 Rumble-2 Flash	13.9	12.1–15.7	24.5	61.0	69.1	12.6	5.8	14.4
57	Deepgram Nova-3 Streaming	13.9	12.1–15.7	23.2	60.4	67.8	14.9	7.1	18.4
58	Amazon Transcribe Medical Streaming	13.8	12.0–15.7	22.2	61.4	67.9	11.8	5.1	13.8
59	FluidVoice Fluid Speech Pro	13.8	12.0–15.6	27.1	61.6	70.9	10.9	4.7	12.6
60	NVIDIA Canary 1B v2	13.7	11.9–15.5	25.7	61.1	69.8	11.3	5.1	12.9
61	Modulate Velma-2 STT English	13.5	11.8–15.4	24.9	59.9	68.7	12.3	5.8	14.5
62	Qwen3-ASR 0.6B	13.3	11.6–15.1	22.5	60.1	67.4	12.7	6.0	15.0
63	Moondream Parakeet Ultra	13.2	11.4–15.0	25.1	60.6	69.7	11.0	4.8	12.4
64	Deepgram Flux English	12.8	11.1–14.5	21.5	60.2	66.7	14.6	7.0	18.2
65	Omi Medical Edge-1	12.5	10.8–14.3	24.9	60.4	69.5	11.3	4.9	13.0
66	IBM Granite Speech 5.0 2B	12.4	10.7–14.2	22.0	58.7	66.7	11.3	4.9	12.9
67	Fun-ASR MLT Nano 2512	12.1	10.4–13.8	19.3	57.9	64.6	15.2	7.4	17.3
68	Gradium STT	12.0	10.4–13.7	22.1	57.8	66.0	14.1	6.7	16.1
69	Moonshine Streaming Medium	10.8	9.2–12.4	18.3	55.6	62.2	16.7	8.4	18.3
70	NVIDIA Parakeet TDT 0.6B v3	10.7	9.2–12.2	20.7	57.7	66.0	12.1	5.4	13.7
71	Deepgram Flux Multilingual	9.9	8.4–11.4	16.8	56.8	63.5	16.7	8.0	20.0

#	SYSTEM	TSR	95% CI	TSR RELAXED	ENTITY ACC.	EA RELAXED	WER	CER	WER HF NORM.
72	Cohere Transcribe 03-2026	9.6	8.3–11.0	17.3	51.5	58.6	15.2	8.4	17.5
73	MOSS Transcribe Preview 2B	9.5	8.1–11.1	15.0	55.3	60.5	16.1	8.3	19.1
74	Fun-ASR Nano 2512	9.5	8.0–11.1	15.8	54.6	61.0	17.2	8.3	19.6
75	Rime STT	9.2	7.7–10.8	14.7	54.2	60.1	16.1	7.5	18.8
76	Meta omniASR LLM 7B	8.7	7.3–10.1	14.6	53.5	60.7	14.5	6.0	16.8
77	Moonshine Base	7.6	6.3–8.9	13.4	51.2	57.3	19.4	9.4	20.9
78	Murf Finch-1	7.0	5.9–8.3	12.4	51.8	58.3	17.4	7.9	20.6
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	6.2	5.1–7.5	10.4	48.3	54.0	20.1	9.6	22.5
80	IBM Granite Speech 5.0 470M TurboCTC	6.1	4.8–7.3	9.7	51.0	55.5	13.1	5.6	14.6
81	NVIDIA FastConformer Hybrid Large	4.9	4.0–5.9	9.0	45.7	52.0	20.0	8.5	22.6
82	SenseVoice Small	3.9	3.0–5.0	6.1	43.6	47.2	23.3	10.1	27.5
83	Moonshine Streaming Tiny	3.4	2.7–4.2	5.3	36.0	39.3	37.7	21.0	39.0
84	FUTO ASR4All L	3.4	2.7–4.2	5.6	43.4	48.1	22.4	10.1	25.1
85	Fermion Phonon-2	3.3	2.6–4.1	6.8	40.2	47.6	15.9	6.9	17.6
86	Vosk en-us 0.42 GigaSpeech	3.1	2.4–3.8	4.5	37.5	40.0	36.6	18.4	38.7
87	Meta omniASR CTC 7B	3.1	2.4–3.9	5.9	40.0	46.5	22.7	8.8	24.4
88	FUTO ASR4All M	2.5	1.9–3.2	4.1	39.1	43.1	25.1	11.5	27.5
89	SpeechBrain Conformer + TransformerLM	1.5	1.1–2.0	2.4	28.0	30.5	42.0	20.4	43.9
90	Meta MMS 1B	0.3	0.1–0.5	0.9	17.2	22.8	35.6	15.2	36.4

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

Ranks use the unrounded TSR, then entity accuracy, so two models with the same TSR at one decimal can appear in either order. The same holds for every ranked table in this paper.

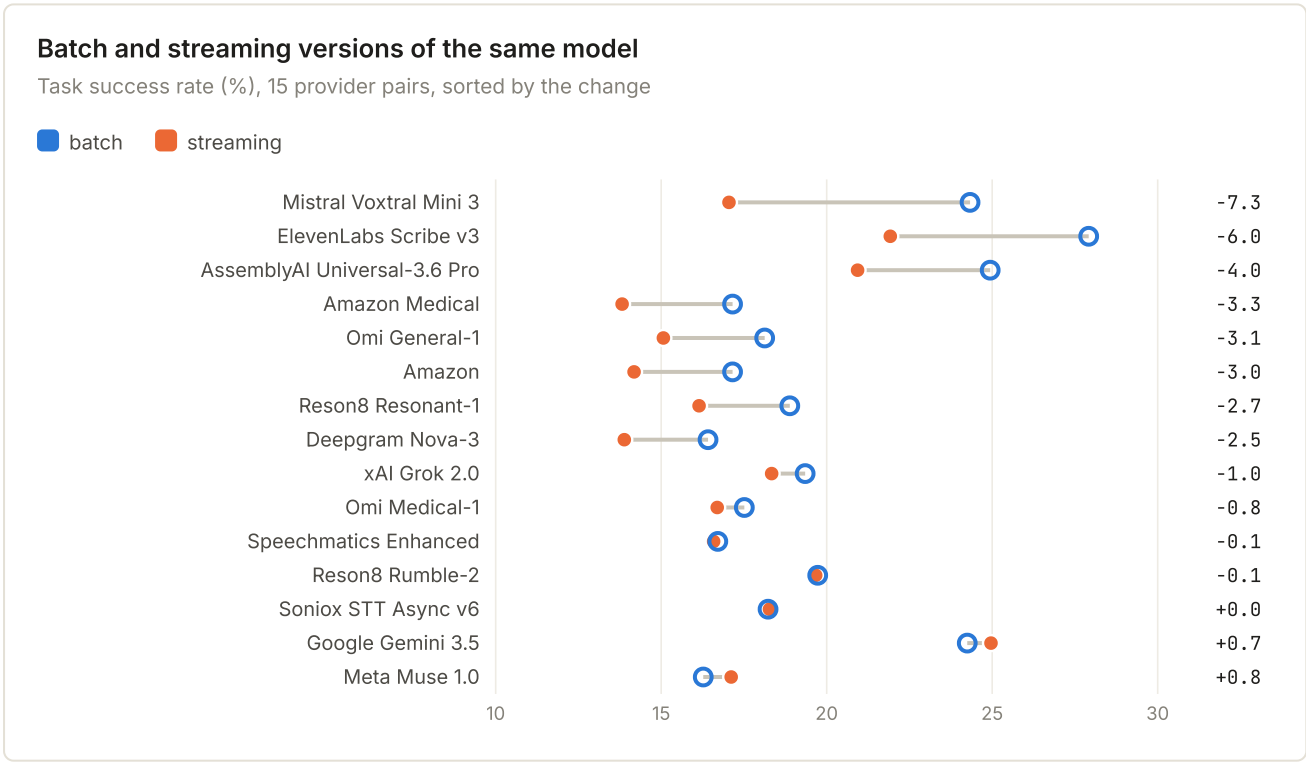

Figure 15. Pairs are a provider's batch and streaming versions of the same model; models sold as different products (e.g. OpenAI GPT Transcribe and GPT Live Transcribe) are not paired. Blue ring: batch; orange dot: streaming; right: the change from batch to streaming in points. Streaming lowers TSR for 12 of the 15 pairs.

11.2 By entity type

Names and email addresses are the hardest entities (17.0% and 17.9% across all models), followed by street addresses (33.6%), where one wrong part fails the whole address. Dates, digit sequences and amounts are mostly solved (89.4%, 87.4%, 84.7%). Email errors are almost all misspelled names in the user name or domain, not lost symbols: the speakers read the address as written, so a model has to spell an unfamiliar name it hears once. Zoom Scribe v2.1 Pro leads on 5 of the 10 types (person names, brands and organisations, technical terms, street addresses, email addresses and urls) and ties for first on codes and ids. Under the relaxed metric (section 10.5), which accepts sound-alike spellings with the same consonants, names rise from 17.0% to 52.9%, emails from 17.9% to 35.1% and addresses from 33.6% to 44.6%.

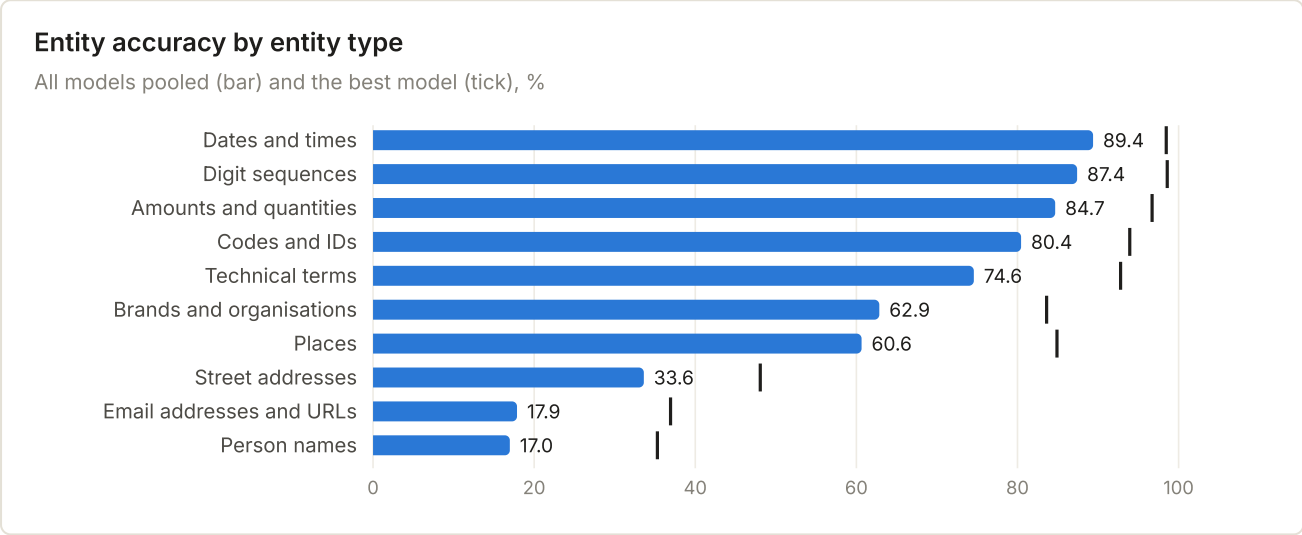

Figure 16. Pooled accuracy over all 90 models and all accents; the tick marks the best single model for that type.

Entity accuracy by entity type

All accents · %

#	SYSTEM	OVERALL	NAMES	BRANDS	PLACES	TERMS	ADDRESS	EMAIL	CODES	DIGITS	AMOUNTS	DATES
1	Zoom Scribe v2.1 Pro	76.2	35.3	83.6	82.7	92.8	48.1	36.9	93.9	97.6	95.9	97.3
2	ElevenLabs Scribe v2 Medical	74.2	28.7	82.2	83.8	91.6	43.6	31.8	90.8	97.4	96.4	98.2
3	ElevenLabs Scribe v3	73.8	26.4	82.8	84.9	92.0	43.2	29.1	90.7	98.6	94.9	98.2
4	Mistral Voxtral Small Transcribe 3	71.9	26.6	80.5	79.8	89.2	42.7	23.7	91.5	95.7	94.0	97.5
5	Microsoft MAI-Transcribe-2	72.0	25.2	78.4	77.6	87.4	41.4	26.9	93.9	98.3	95.5	98.5
6	Google Gemini 3.5 Transcribe Live	71.6	22.9	79.7	80.7	90.6	39.5	28.2	89.0	96.4	94.2	97.4
7	AssemblyAI Universal-3.6 Pro	71.6	23.5	79.2	77.6	88.2	41.3	26.6	91.4	96.3	96.7	98.2
8	Mistral Voxtral Mini Transcribe 3	71.3	26.5	77.5	76.9	87.2	43.2	22.8	92.7	95.9	95.2	97.7
9	AssemblyAI STT 3.7 Realtime Agent	71.2	23.4	78.9	76.4	88.8	40.7	26.5	90.2	95.6	96.3	98.0
10	Google Gemini 3.5 Transcribe	71.4	21.3	79.8	80.0	90.5	39.5	27.7	88.7	96.5	94.9	97.8
11	Google Cloud Chirp 3	70.4	17.4	80.6	82.3	88.0	37.8	25.6	87.8	95.9	94.2	97.6
12	ElevenLabs Scribe v3 Live	69.4	24.0	73.5	72.3	85.4	41.0	22.4	91.1	97.1	92.7	97.0
13	OpenAI GPT Transcribe	69.0	21.2	76.4	72.5	87.0	41.0	24.4	89.1	92.5	93.3	94.9
14	Mistral Voxtral Small 24B	68.4	22.8	70.6	69.2	82.7	39.9	25.8	89.9	96.1	94.8	95.7
15	OpenAI GPT-4o Mini Transcribe	68.7	19.2	77.7	78.4	88.3	37.4	24.1	84.7	89.6	95.9	94.9
16	AssemblyAI Universal-3.6 Pro Realtime	68.3	21.8	76.8	74.3	87.3	38.2	23.6	86.3	90.4	90.5	95.9
17	Reson8 Rumble-2	67.2	21.4	68.6	64.9	84.1	38.5	22.3	88.5	96.5	93.6	96.4
18	Reson8 Rumble-2 Realtime	67.2	21.3	69.2	65.4	83.7	38.4	22.5	88.8	96.5	93.3	96.1
19	xAI Grok Voice Transcribe 2.0	66.3	21.3	70.9	72.5	81.5	39.3	24.9	85.2	91.8	87.1	91.5
20	Salesforce SOV-Transcribe-1	67.1	18.6	74.1	70.1	84.5	40.0	23.9	84.9	93.8	89.9	93.8
21	Fish Audio Transcribe 1 Pro	66.9	19.9	71.4	69.3	84.6	38.6	20.3	86.8	94.8	90.1	96.0
22	OpenAI Whisper large-v3	66.5	22.2	67.6	65.6	78.9	36.6	22.5	90.0	95.4	93.5	95.8

#	SYSTEM	OVERALL	NAMES	BRANDS	PLACES	TERMS	ADDRESS	EMAIL	CODES	DIGITS	AMOUNTS	DATES
23	Reson8 Resonant-1	65.9	21.6	64.5	65.3	76.2	36.9	23.8	91.0	96.8	91.7	95.4
24	OpenAI GPT Live Transcribe	66.1	19.2	72.8	68.1	83.1	39.4	23.5	83.6	91.7	88.9	93.4
25	Fish Audio Transcribe Stream 1	65.9	18.7	71.1	70.4	86.1	37.5	20.1	86.0	92.5	86.7	92.0
26	Cartesia Ink 2	65.7	20.5	65.8	63.8	79.4	37.7	24.9	87.8	95.1	89.8	95.6
27	xAI Grok Voice Transcribe 2.0 Streaming	65.3	21.2	69.8	69.9	79.9	38.6	23.1	84.6	92.7	85.2	91.0
28	Soniox STT Real-time v6	65.4	18.5	64.6	59.1	77.7	37.8	23.3	90.6	97.1	91.2	97.7
29	Soniox STT Async v6	66.0	18.6	65.6	59.0	78.5	38.1	23.5	90.6	97.8	93.4	98.0
30	Omi General-1	65.4	17.5	68.4	64.7	82.3	38.6	19.6	86.7	93.3	92.0	94.3
31	Qwen3-ASR 1.7B	65.4	17.3	68.5	64.3	81.9	38.3	20.2	87.3	93.3	91.6	94.2
32	Omi Medical-1	64.6	16.9	67.4	64.3	81.5	37.4	19.5	86.6	93.6	87.5	93.8
33	Omi Medical-1 Flash	64.5	17.0	67.1	64.7	81.3	36.9	18.8	86.6	93.4	87.9	93.6
34	Amazon Transcribe	65.3	17.0	65.0	64.4	77.0	38.4	20.7	90.2	96.1	90.5	97.2
35	Amazon Transcribe Medical	65.3	17.0	65.0	64.4	77.0	38.4	20.7	90.2	96.1	90.4	97.2
36	Meta Muse Voice Transcribe Realtime	64.3	19.6	68.4	64.8	78.5	37.4	21.1	84.9	88.1	88.9	94.0
37	Mistral Voxtral Mini Transcribe Realtime 3	64.6	21.5	69.0	67.1	79.3	36.6	19.7	82.4	90.5	87.4	95.2
38	Speechmatics Enhanced	64.9	21.5	76.7	81.0	88.5	34.7	11.9	63.0	92.2	87.2	94.5
39	OpenAI GPT Realtime Whisper	63.8	18.7	70.2	63.9	81.2	35.7	18.7	82.3	88.4	87.5	93.8
40	Omi Medical-1 Live	63.1	16.3	65.6	61.9	80.3	36.8	17.8	86.2	91.8	83.7	92.1
41	Mistral Voxtral Mini 4B Realtime	62.2	19.6	63.4	60.7	73.9	34.7	20.3	81.6	91.9	87.5	91.5
42	Speechmatics Enhanced Realtime	64.7	21.1	76.1	79.9	88.4	34.3	11.8	63.9	92.3	87.4	94.4
43	Deepgram Nova-3	63.8	18.0	68.1	65.5	78.2	36.8	21.7	82.3	93.2	86.2	91.3
44	Meta Muse Voice Transcribe 1.0	64.3	19.5	68.6	63.9	78.4	37.6	20.3	85.5	86.0	90.2	95.1
45	Reson8 Resonant-1 Realtime	63.2	21.4	64.0	64.0	74.6	35.9	20.9	85.6	89.3	83.8	94.7
46	Smallest.ai Pulse Pro	64.1	17.5	68.1	65.1	81.7	37.7	21.8	81.3	92.5	85.7	91.7
47	Inworld STT-1	62.7	15.0	63.1	62.0	78.6	36.5	20.0	81.7	91.1	89.4	92.7
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	63.4	18.3	71.1	71.8	85.1	37.8	18.3	66.3	91.4	87.2	89.1
49	Musci-ASR qwen-1019	63.2	16.7	67.3	61.7	80.0	36.8	20.9	83.3	91.8	85.9	90.4
50	ARK-ASR 3B	62.7	16.3	63.3	57.8	78.5	37.5	24.5	84.0	91.1	86.2	90.2
51	Omi General-1 Live	62.8	16.7	65.2	61.2	80.2	37.3	16.8	84.5	91.4	85.7	91.2
52	Speechmatics Melia 1	62.2	18.3	62.6	62.0	74.4	34.1	20.8	86.3	88.7	85.0	92.8
53	Speechmatics Linden 1	61.7	20.7	74.0	75.7	85.8	33.6	8.3	62.0	88.0	78.2	91.3
54	Speechmatics Oak 1	61.9	17.6	61.9	61.2	75.6	33.9	20.1	85.3	88.5	85.0	92.8
55	Amazon Transcribe Streaming	62.5	15.8	62.0	63.0	75.3	35.2	14.8	86.0	94.0	86.6	95.1
56	Reson8 Rumble-2 Flash	61.0	16.3	60.4	53.8	72.9	31.4	18.4	84.1	93.1	89.3	93.5
57	Deepgram Nova-3 Streaming	60.4	15.4	63.8	58.5	74.0	35.1	19.5	79.1	89.3	83.5	88.8
58	Amazon Transcribe Medical Streaming	61.4	15.3	59.6	59.5	71.6	34.8	14.3	87.6	95.9	87.9	91.2
59	FluidVoice Fluid Speech Pro	61.6	18.0	59.3	56.1	67.7	34.2	20.8	89.0	93.9	90.3	90.4
60	NVIDIA Canary 1B v2	61.1	16.2	59.5	56.9	69.3	33.2	20.4	85.6	91.8	90.0	92.4
61	Modulate Velma-2 STT English	59.9	15.4	56.4	56.1	68.9	32.3	19.9	86.1	90.9	85.8	91.4
62	Qwen3-ASR 0.6B	60.1	14.7	59.6	55.2	73.2	33.3	18.1	82.4	90.5	85.8	91.7
63	Moondream Parakeet Ultra	60.6	16.3	57.3	54.2	67.1	32.6	18.2	88.0	93.8	90.5	92.6
64	Deepgram Flux English	60.2	11.7	60.1	60.4	73.1	35.1	19.7	77.7	91.6	85.2	90.8
65	Omi Medical Edge-1	60.4	14.8	54.9	53.7	67.8	33.0	18.6	86.9	95.7	89.2	93.7
66	IBM Granite Speech 5.0 2B	58.7	16.0	60.1	63.0	75.0	37.0	17.6	70.6	69.5	87.7	92.5
67	Fun-ASR MLT Nano 2512	57.9	13.7	57.7	54.9	68.4	28.0	10.7	79.5	91.8	85.8	92.3
68	Gradium STT	57.8	16.6	59.1	52.3	67.2	32.4	17.2	81.9	83.3	79.9	90.8
69	Moonshine Streaming Medium	55.6	12.6	54.2	51.5	64.0	25.2	15.2	75.5	87.3	84.7	90.6
70	NVIDIA Parakeet TDT 0.6B v3	57.7	14.0	54.9	50.9	66.0	31.2	15.7	82.7	90.0	87.0	88.3
71	Deepgram Flux Multilingual	56.8	8.6	51.4	50.3	67.0	34.1	16.7	78.3	91.7	84.5	89.3
72	Cohere Transcribe 03-2026	51.5	11.9	49.8	47.6	66.5	27.4	16.5	63.7	77.3	72.9	84.2
73	MOSS Transcribe Preview 2B	55.3	7.9	56.5	48.1	70.0	30.4	12.4	75.7	86.8	80.0	87.6
74	Fun-ASR Nano 2512	54.6	10.6	53.6	49.6	65.5	26.0	10.1	79.9	88.6	80.2	85.9

#	SYSTEM	OVERALL	NAMES	BRANDS	PLACES	TERMS	ADDRESS	EMAIL	CODES	DIGITS	AMOUNTS	DATES
75	Rime STT	54.2	11.2	55.7	50.3	72.0	33.7	7.8	77.9	64.2	81.5	88.7
76	Meta omniASR LLM 7B	53.5	15.2	49.3	50.4	62.3	30.8	1.6	86.9	85.6	73.1	81.2
77	Moonshine Base	51.2	8.9	47.9	40.7	58.0	22.4	13.2	74.5	84.1	79.7	87.0
78	Murf Finch-1	51.8	10.3	49.1	45.8	62.4	28.3	6.5	70.3	80.4	82.1	86.4
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	48.3	10.4	43.8	37.3	55.8	27.6	1.8	66.4	80.3	79.2	84.2
80	IBM Granite Speech 5.0 470M TurboCTC	51.0	12.5	51.5	54.3	62.5	31.4	0.2	66.1	84.2	81.1	69.8
81	NVIDIA FastConformer Hybrid Large	45.7	11.3	46.9	47.3	57.5	23.0	8.0	50.3	57.3	73.1	84.3
82	SenseVoice Small	43.6	3.8	36.2	32.4	49.6	20.1	4.9	74.4	85.9	63.6	68.9
83	Moonshine Streaming Tiny	36.0	5.2	30.6	26.3	42.0	11.8	7.4	53.4	59.3	60.2	66.9
84	FUTO ASR4All L	43.4	7.9	38.0	32.1	49.1	18.3	1.3	59.9	75.3	74.8	81.2
85	Fermion Phonon-2	40.2	13.5	50.7	47.4	60.6	6.8	15.0	79.9	32.1	57.8	37.6
86	Vosk en-us 0.42 GigaSpeech	37.5	5.6	44.6	39.6	56.9	12.5	4.0	31.8	54.3	60.1	67.3
87	Meta omniASR CTC 7B	40.0	11.5	38.3	36.5	49.9	16.3	0.0	80.7	66.8	44.7	55.2
88	FUTO ASR4All M	39.1	5.7	33.0	27.4	43.5	16.4	0.4	56.1	68.6	69.2	74.5
89	SpeechBrain Conformer + TransformerLM	28.0	2.4	22.0	21.3	27.6	12.2	0.7	26.7	46.6	54.7	70.1
90	Meta MMS 1B	17.2	5.9	26.3	22.7	34.4	2.8	0.8	44.0	7.6	11.7	12.6

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

11.3 By accent

Accuracy is highest for US speakers (66.5% across all models) and lowest for Nigerian speakers (55.9%), on identical sentences. The gap between a model's best and worst accent averages 10.8 points; Zoom Scribe v2.1 Pro has a gap of 7.5, and the most even of the top 20 is ElevenLabs Scribe v3 (5.7). Rankings are stable across accents: Zoom Scribe v2.1 Pro has the highest TSR on all seven.

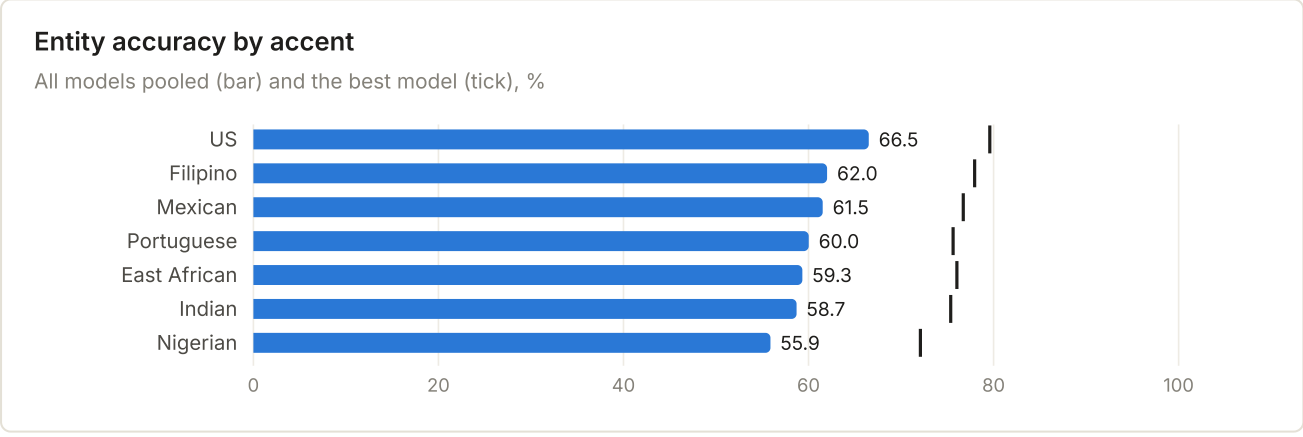

Figure 17. Every accent has the same 677 sentences, so differences reflect the speakers and recordings, not the content.

Entity accuracy by accent

% · gap = best accent minus worst accent (lower is more even)

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN	GAP
1	Zoom Scribe v2.1 Pro	76.2	79.6	78.0	76.7	75.6	76.0	75.4	72.1	7.5
2	ElevenLabs Scribe v2 Medical	74.2	77.6	75.9	74.4	73.5	74.2	72.5	71.3	6.3
3	ElevenLabs Scribe v3	73.8	77.2	75.2	74.0	73.1	73.6	72.5	71.4	5.7
4	Mistral Voxtral Small Transcribe 3	71.9	75.7	74.6	72.1	71.3	71.1	71.1	67.6	8.1
5	Microsoft MAI-Transcribe-2	72.0	76.4	73.0	73.2	71.7	70.9	71.2	67.7	8.7
6	Google Gemini 3.5 Transcribe Live	71.6	75.6	74.0	71.3	69.7	71.3	70.9	68.4	7.2
7	AssemblyAI Universal-3.6 Pro	71.6	75.7	72.4	72.0	70.4	72.3	70.3	68.1	7.6
8	Mistral Voxtral Mini Transcribe 3	71.3	76.0	73.4	71.9	71.5	70.1	70.0	66.6	9.5
9	AssemblyAI STT 3.7 Realtime Agent	71.2	75.8	72.1	71.8	70.3	71.1	70.1	67.2	8.5

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN	GAP
10	Google Gemini 3.5 Transcribe	71.4	75.1	72.8	70.7	70.1	71.5	71.0	68.5	6.6
11	Google Cloud Chirp 3	70.4	73.8	72.1	70.2	69.0	70.3	70.1	67.4	6.4
12	ElevenLabs Scribe v3 Live	69.4	73.7	71.4	69.5	68.4	69.0	68.2	65.6	8.0
13	OpenAI GPT Transcribe	69.0	73.6	69.7	69.7	68.6	67.5	68.0	65.9	7.7
14	Mistral Voxtral Small 24B	68.4	73.1	70.3	69.4	67.8	67.1	67.4	63.7	9.4
15	OpenAI GPT-4o Mini Transcribe	68.7	73.7	70.3	69.9	68.3	67.2	67.2	64.5	9.2
16	AssemblyAI Universal-3.6 Pro Realtime	68.3	72.9	69.7	69.3	66.9	68.2	66.6	64.6	8.3
17	Reson8 Rumble-2	67.2	72.0	69.4	67.2	67.2	66.2	65.6	62.8	9.3
18	Reson8 Rumble-2 Realtime	67.2	71.9	69.6	67.5	66.9	66.2	65.8	62.7	9.1
19	xAI Grok Voice Transcribe 2.0	66.3	71.0	68.1	66.2	66.0	65.8	65.5	61.8	9.2
20	Salesforce SOV-Transcribe-1	67.1	70.4	70.0	66.7	66.8	66.0	66.9	62.9	7.5
21	Fish Audio Transcribe 1 Pro	66.9	72.6	67.8	67.1	66.2	66.0	65.9	62.9	9.7
22	OpenAI Whisper large-v3	66.5	72.1	68.3	67.1	66.7	65.3	64.7	61.1	10.9
23	Reson8 Resonant-1	65.9	71.1	67.2	66.4	66.2	65.4	64.2	61.1	10.1
24	OpenAI GPT Live Transcribe	66.1	71.2	66.7	66.4	65.3	65.2	65.4	62.7	8.5
25	Fish Audio Transcribe Stream 1	65.9	70.3	67.8	65.8	65.5	64.9	65.2	61.9	8.5
26	Cartesia Ink 2	65.7	71.7	67.9	67.3	64.5	63.9	64.4	60.4	11.4
27	xAI Grok Voice Transcribe 2.0 Streaming	65.3	69.9	67.3	65.9	64.4	64.2	65.0	60.8	9.1
28	Soniox STT Real-time v6	65.4	69.9	67.1	67.3	64.7	64.3	64.1	60.7	9.3
29	Soniox STT Async v6	66.0	70.0	68.2	66.7	65.8	65.5	65.2	60.3	9.7
30	Omi General-1	65.4	71.3	66.7	65.6	64.8	64.8	62.8	62.0	9.3
31	Qwen3-ASR 1.7B	65.4	71.1	66.8	65.9	65.2	64.7	62.8	61.6	9.5
32	Omi Medical-1	64.6	70.7	66.0	65.5	63.9	63.7	61.4	61.0	9.7
33	Omi Medical-1 Flash	64.5	70.8	66.1	65.2	63.2	63.9	61.2	61.0	9.8
34	Amazon Transcribe	65.3	70.7	66.7	66.4	63.7	63.6	64.5	61.4	9.4
35	Amazon Transcribe Medical	65.3	70.7	66.7	66.4	63.7	63.6	64.5	61.4	9.4
36	Meta Muse Voice Transcribe Realtime	64.3	70.7	66.0	65.2	62.3	63.1	62.6	60.5	10.2
37	Mistral Voxtral Mini Transcribe Realtime 3	64.6	71.4	67.0	66.3	62.5	62.9	63.5	58.9	12.5
38	Speechmatics Enhanced	64.9	70.5	65.8	65.9	62.1	65.0	64.2	60.7	9.8
39	OpenAI GPT Realtime Whisper	63.8	70.4	64.3	63.9	62.4	63.4	62.4	60.1	10.3
40	Omi Medical-1 Live	63.1	69.6	65.4	63.5	61.0	62.9	60.2	58.9	10.7
41	Mistral Voxtral Mini 4B Realtime	62.2	70.2	63.6	66.1	62.6	59.9	55.4	57.3	14.8
42	Speechmatics Enhanced Realtime	64.7	70.7	65.2	66.2	62.1	64.7	64.6	59.7	11.1
43	Deepgram Nova-3	63.8	69.0	66.7	64.9	63.8	61.6	62.3	58.6	10.4
44	Meta Muse Voice Transcribe 1.0	64.3	71.3	65.6	65.3	61.8	63.0	62.6	60.3	11.0
45	Reson8 Resonant-1 Realtime	63.2	69.5	65.1	63.4	61.1	63.2	61.4	58.4	11.1
46	Smallest.ai Pulse Pro	64.1	68.9	65.5	63.8	63.5	63.9	62.7	60.3	8.6
47	Inworld STT-1	62.7	70.1	64.7	64.2	62.3	61.3	59.0	57.3	12.7
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	63.4	68.5	64.8	62.7	64.0	61.9	63.0	58.8	9.7
49	Musci-ASR qwen-1019	63.2	68.7	65.0	63.6	63.2	62.3	61.6	58.3	10.4
50	ARK-ASR 3B	62.7	66.9	64.8	63.3	62.2	62.0	61.9	57.6	9.3
51	Omi General-1 Live	62.8	68.1	65.2	62.6	61.2	62.2	61.6	58.8	9.3
52	Speechmatics Melia 1	62.2	68.5	64.0	63.9	61.7	61.4	59.6	56.4	12.1
53	Speechmatics Linden 1	61.7	67.4	63.9	62.7	58.2	61.4	61.9	56.1	11.2
54	Speechmatics Oak 1	61.9	68.2	63.8	63.5	61.5	61.1	59.3	56.2	12.1
55	Amazon Transcribe Streaming	62.5	69.3	65.1	63.0	60.2	61.1	61.7	57.0	12.2
56	Reson8 Rumble-2 Flash	61.0	66.8	63.4	61.6	60.5	58.7	59.1	56.5	10.3
57	Deepgram Nova-3 Streaming	60.4	66.8	62.7	61.7	59.7	58.2	57.9	56.0	10.8
58	Amazon Transcribe Medical Streaming	61.4	67.7	63.2	62.0	59.4	60.8	60.4	56.4	11.2
59	FluidVoice Fluid Speech Pro	61.6	67.7	62.5	63.5	60.3	60.8	59.3	56.9	10.7
60	NVIDIA Canary 1B v2	61.1	68.2	61.2	62.5	60.8	60.5	58.5	56.3	11.9
61	Modulate Velma-2 STT English	59.9	68.1	61.6	61.1	58.9	58.8	56.8	54.4	13.6

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN	GAP
62	Qwen3-ASR 0.6B	60.1	66.4	61.4	61.0	60.5	58.8	57.0	55.7	10.7
63	Moondream Parakeet Ultra	60.6	67.6	62.1	61.9	59.6	59.4	57.3	56.5	11.1
64	Deepgram Flux English	60.2	65.9	62.2	61.5	59.3	57.8	58.8	55.8	10.1
65	Omi Medical Edge-1	60.4	66.3	62.2	61.6	59.1	60.1	57.4	56.1	10.1
66	IBM Granite Speech 5.0 2B	58.7	63.5	59.7	59.9	58.0	58.7	56.7	54.2	9.3
67	Fun-ASR MLT Nano 2512	57.9	66.2	60.0	59.6	58.3	54.9	54.6	51.8	14.5
68	Gradium STT	57.8	65.8	59.3	60.2	57.4	55.6	54.9	51.8	14.0
69	Moonshine Streaming Medium	55.6	64.6	56.2	57.6	56.1	51.9	52.8	50.4	14.2
70	NVIDIA Parakeet TDT 0.6B v3	57.7	64.1	58.4	58.6	57.7	56.2	55.1	53.5	10.6
71	Deepgram Flux Multilingual	56.8	62.2	59.2	57.6	57.7	53.5	55.6	51.8	10.4
72	Cohere Transcribe 03-2026	51.5	51.9	55.3	53.8	54.0	42.6	55.1	47.9	12.8
73	MOSS Transcribe Preview 2B	55.3	63.3	53.8	56.5	55.7	53.5	54.1	50.1	13.2
74	Fun-ASR Nano 2512	54.6	63.2	56.0	57.2	54.1	51.5	51.5	48.8	14.5
75	Rime STT	54.2	59.2	55.0	55.6	54.5	53.8	51.0	50.6	8.6
76	Meta omniASR LLM 7B	53.5	59.1	55.2	55.4	53.1	53.4	49.7	48.2	10.9
77	Moonshine Base	51.2	60.3	52.3	52.7	51.4	47.4	47.2	47.1	13.3
78	Murf Finch-1	51.8	59.5	51.3	53.0	52.8	50.4	50.0	45.7	13.8
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	48.3	55.1	47.9	49.0	46.6	48.0	49.0	42.7	12.4
80	IBM Granite Speech 5.0 470M TurboCTC	51.0	55.3	52.8	51.8	49.7	50.8	49.7	46.7	8.5
81	NVIDIA FastConformer Hybrid Large	45.7	53.8	45.4	47.9	45.6	44.5	43.4	39.1	14.7
82	SenseVoice Small	43.6	52.1	46.0	44.0	43.1	41.9	40.7	37.6	14.5
83	Moonshine Streaming Tiny	36.0	49.0	35.7	38.8	38.4	32.0	30.5	27.4	21.6
84	FUTO ASR4All L	43.4	51.0	43.4	45.2	43.8	42.0	41.0	37.2	13.8
85	Fermion Phonon-2	40.2	43.4	38.9	38.4	38.6	43.2	39.2	39.3	5.0
86	Vosk en-us 0.42 GigaSpeech	37.5	49.9	38.9	43.9	40.3	33.2	31.0	25.1	24.8
87	Meta omniASR CTC 7B	40.0	45.2	41.7	37.9	38.4	40.9	36.8	38.9	8.3
88	FUTO ASR4All M	39.1	48.3	38.7	41.2	39.9	36.9	36.1	32.5	15.8
89	SpeechBrain Conformer + TransformerLM	28.0	41.6	29.5	31.7	32.2	22.1	22.0	17.1	24.5
90	Meta MMS 1B	17.2	23.4	17.7	19.7	19.6	14.2	13.6	12.2	11.2

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch API streaming open weights

11.4 By sentence difficulty, entity tier and domain

Sentences are split into thirds by how often all models get their entities right (section 1.2): 225 easy, 225 medium and 227 hard sentences, each read in all seven accents. **Task success falls steeply with difficulty:** across all models, TSR is 37.5% on easy sentences, 6.4% on medium and 1.8% on hard ones. Zoom Scribe v2.1 Pro gets every entity right in 65.3% of easy clips, 23.2% of medium and 7.5% of hard ones.

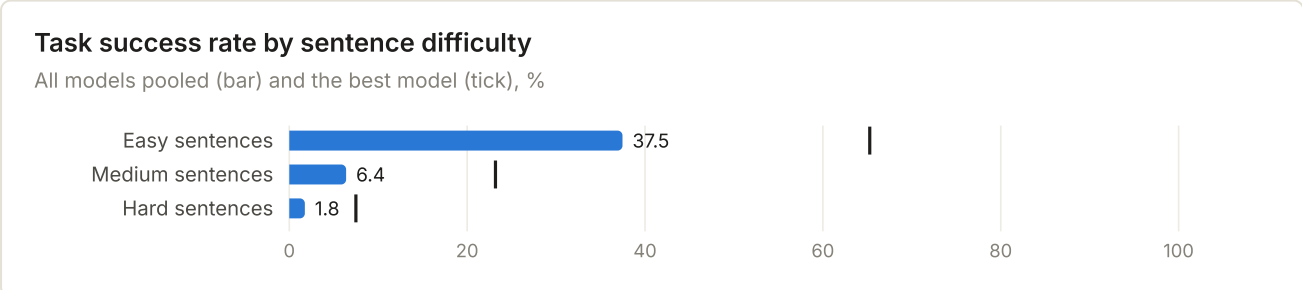

Figure 18. Difficulty is a property of the sentence, the same in every accent; the tick marks the best single model on that third.

Every model's TSR on easy, medium and hard sentences is in Appendix B.1.

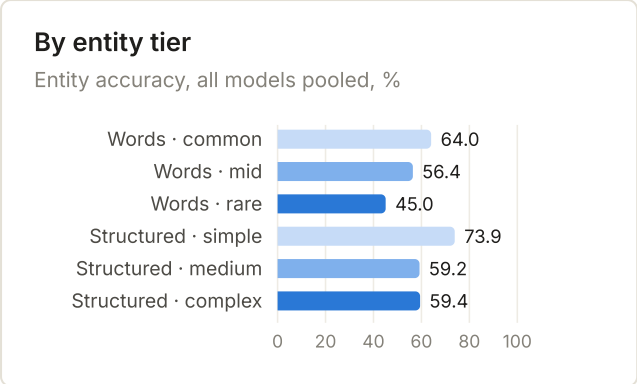

Figure 19. Rare names, brands, places and terms are the hardest tier; simple structured values the easiest.

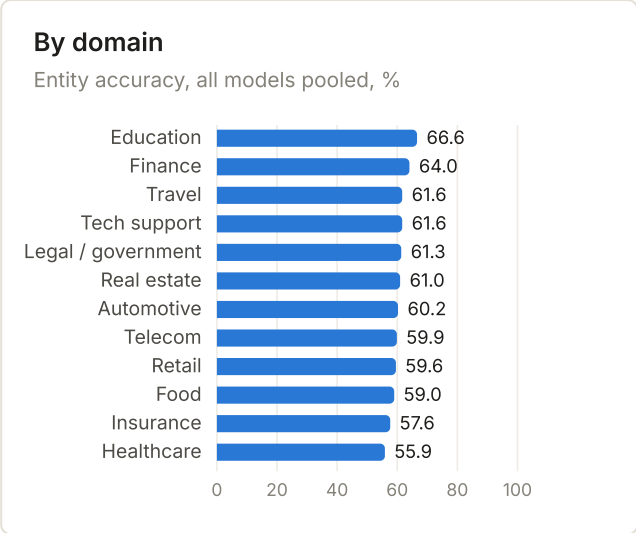

Figure 20. Domains differ by up to 10.7 points, following their entity mix.

11.5 Entity accuracy and WER

The two rankings agree broadly (Spearman ρ = 0.94 over 90 models) but not model by model. xAI Grok Voice Transcribe 2.0 Streaming, Smallest.ai Pulse Pro, OpenAI GPT Live Transcribe rank 25, 23 and 18 places higher on entity accuracy than on WER; Moondream Parakeet Ultra, FluidVoice Fluid Speech Pro, IBM Granite Speech 5.0 2B rank 24, 21 and 20 places lower. A model can transcribe the sentence around an entity well and still get the entity wrong, which is why WER alone does not rank models for this task.

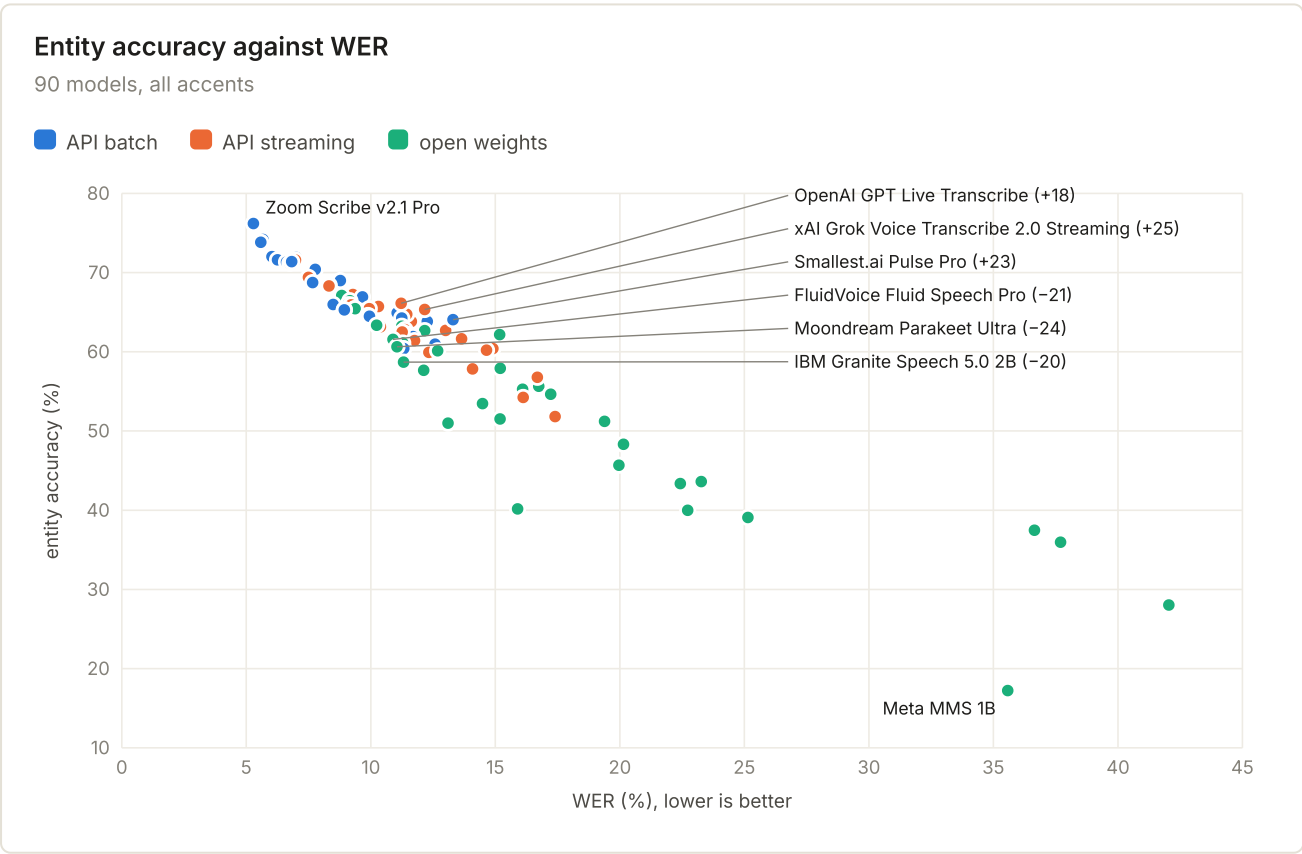

Figure 21. Each dot is a model. Labels mark the leader, the last model and the six largest disagreements between the two rankings, with the number of places a model gains (+) or loses (−) on entity accuracy compared with its WER rank.

11.6 Public and private sets

Every model is also scored on the public set alone (931 clips, 3,619 entity instances per model) and on the private set alone (3,808 clips, 15,596 instances). The two rankings agree closely: Spearman ρ = 0.98 on task success rate, and 10 of the ten best models are the same on both sets. The public set scores slightly higher, by 1.52 points of TSR and 0.58 of entity accuracy on average, so public and private numbers should be compared within a set, not across sets. The relaxed metric (section 10.5) behaves the same way: Spearman ρ = 0.99 on relaxed TSR and 0.99 on relaxed entity accuracy, with the public set higher by 3.07 and 1.46 points. Every model's strict and relaxed scores on both sets are in Appendix C.

Metric	Spearman ρ	Mean gap (public − private, points)	Largest gap	Same top 10
Task success rate (TSR)	0.98	+1.52	3.50	10 of 10
TSR, relaxed	0.99	+3.07	5.98	9 of 10
Entity accuracy	0.99	+0.58	2.06	10 of 10
Entity accuracy, relaxed	0.99	+1.46	3.36	10 of 10
WER	0.99	−0.94	4.51	10 of 10
CER	0.99	−0.45	3.00	10 of 10

#	Model (top 10 overall)	Pub. #	Pub. TSR	TSR rel.	Pub. EA	EA rel.	Priv. #	Priv. TSR	TSR rel.	Priv. EA	EA rel.
1	Zoom Scribe v2.1 Pro	1	34.4	55.0	77.1	85.9	1	31.3	51.9	76.0	84.5
2	ElevenLabs Scribe v2 Medical	2	29.3	48.9	74.5	83.4	2	28.1	47.5	74.1	82.8
3	ElevenLabs Scribe v3	3	28.9	48.4	74.1	83.1	3	27.7	46.6	73.8	82.1
4	Mistral Voxtral Small Transcribe 3	7	26.4	45.3	72.5	81.4	4	25.4	42.5	71.8	80.3
5	Microsoft MAI-Transcribe-2	4	26.9	49.8	72.8	83.4	5	24.8	45.3	71.8	81.7
6	Google Gemini 3.5 Transcribe Live	8	26.4	43.6	72.0	80.9	6	24.6	39.7	71.5	79.1
7	AssemblyAI Universal-3.6 Pro	6	26.5	45.2	71.8	81.6	7	24.6	42.7	71.6	80.6
8	Mistral Voxtral Mini Transcribe 3	9	26.0	44.4	71.8	81.2	9	23.9	41.9	71.2	80.0
9	AssemblyAI STT 3.7 Realtime Agent	5	26.7	45.5	71.6	81.4	10	23.7	40.6	71.1	79.5
10	Google Gemini 3.5 Transcribe	10	25.3	40.9	71.6	80.7	8	24.0	39.5	71.3	79.1

TSR: task success rate, EA: entity accuracy, both in %; rel.: relaxed (section 10.5). Ranks within each set by strict TSR, as in section 11.1.

12 Vendor behaviour that shaped the settings

Speech-to-text APIs differ in more than their models. The behaviours below decided how a session is closed, when a transcript counts as final, or whether a model could be run at all. Each is handled the same way for every clip of that model.

Vendor or model	Behaviour	How it is handled	Effect on data
AssemblyAI STT 3.7 Realtime Agent	Most clips end within a tenth of a second of the last word, so the last turn is often still open at `Terminate`, which finalises it. In replays, the final text of a long turn sometimes left out words its interim drafts had shown	A turn with text that never receives `end_of_turn=true` fails the attempt and is retried; finals are kept as returned, as for every model	None: no clip ended with an open turn
Deepgram Flux	`ForceEndTurn` finalises only the audio decoded so far, and an empty trailing turn never receives `EndOfTurn`	Silence is sent until Flux has decoded the whole clip (at most 1.5 s), then `ForceEndTurn`; only turns with text must finalise	1 clip never finalises (section 8.2)
Google Gemini 3.5 Transcribe Live	A session can end with a turn still open, leaving only an interim draft	Wait until no turn is open; otherwise retry. Drafts are never saved	None in the evaluation set: the 2 clips that never finalised were removed in QC
Rime STT	One 24.4 s Portuguese clip (a re-recording) is cut by the model's generation token cap on every attempt (`resource_exhausted`)	Left as an error; the audio is not split	1 clip missing

Handled the same way, with no effect on the data: Amazon Transcribe Streaming (leaves WebSocket keepalive pings unanswered, so a client with default pings drops sessions about 40 s in; clips longer than about 35 s failed); ElevenLabs Scribe v3 Live (the final commit can fail to arrive after the last chunk); Omi Live models (the closing `final_transcript` differs from the confirmed turns); Inworld STT-1 (after `closeStream` the server sends the last final and a `usage` message but leaves the socket open); xAI Grok streaming (`transcript.done` carries only the text still pending at `audio.done`).

13 Limitations

- **One pass per model.** Streaming and LLM-based models are not fully deterministic; a rerun can differ slightly.
- **Language control varies.** Five API models cannot be forced to English: four take a hint or bias and one auto-detects (section 6.1).
- **Input windows.** ARK-ASR 3B and both omniASR 7B models cut audio at 40 s, the leaderboard setting; 54 clips (1.1%) lose their ending for these three (section 7.3).
- **Raw formats differ.** Some models return lowercase, unpunctuated or spelled-out text (section 6.4); scoring must normalise both sides.
- **2 clips are missing and 34 transcripts are empty.** Both are kept as returned and score as wrong (sections 8.2 and 10.4).
- **Entity scoring is map-only.** Every accepted rendering is a reviewed map entry. A new model can write a correct form nobody has produced before; it counts as wrong until it is checked and added. For this release 1,666 such renderings, from the three models added later and the re-recorded clips, were checked against the rules and added (section 10.2).
- **The relaxed metric ignores vowels.** It cannot hear the speaker, so it accepts any vowel difference in a made-up name, email or street name; some accepted spellings are vowel mishearings (section 10.5). Real brands, places and non-medical terms get only reviewed identical-sounding spellings and spacing variants; medical names are never relaxed.

- **Accent and speaker go together.** Each accent is read by 94–118 speakers, so an accent result also reflects those voices and recording conditions.
- **Cloud services change.** API models were measured on 12–13 November 2026 (AssemblyAI STT 3.7 Realtime Agent: 17 November 2026; re-recorded clips: 17 November 2026) and depend on the provider's network and model at that time.
- **Time is not latency.** Streaming time includes real-time playback and open-source time is a share of batch time (section 8.1).

14 Reproducing the work

The transcription harness is open source under the MIT licence. Each API script runs on its own with uv, which installs the pinned dependencies listed at the top of the script, and the key named in `.env.example`; the open-source pipeline needs a Hugging Face account with Jobs. The private set's audio and references are not released; the public set is released separately.

```
Entity-eval/
├── api/                    # one standalone script per API model (57)
├── hf_jobs/
│   ├── make_dataset.py     # private HF dataset from the prepared audio
│   ├── models.json         # per-model image, script, arguments and their source lines
│   ├── run_jobs.py         # launch and track HF Jobs (refuses public data)
│   ├── collect.py          # download and verify manifests
│   └── runners/            # runners for models outside the leaderboard repo
├── check_results.py        # QA report for API results
├── combine.py              # wide table: one column per model
├── run_all.sh              # every API model, one process each
├── prompts.example.csv     # the format of the prompts CSV
├── .env.example            # names of the API keys (values never committed)
└── LICENSE                 # MIT
```

```
git clone https://github.com/VoiceArenaAI/Entity-eval && cd Entity-eval

# API models: one standalone script per model, keys in .env (uv installs each script's dependencies)
uv run api/<model>.py --limit 2                    # smoke test: 2 sentences x 7 accents
./run_all.sh                                       # every API model
uv run check_results.py                            # QA report

# Open-source models: HF Jobs; the dataset and bucket must be private
git clone https://github.com/huggingface/open_asr_leaderboard && git -C open_asr_leaderboard checkout
859159c
uv run hf_jobs/make_dataset.py --repo <ns>/<private-dataset>
uv run hf_jobs/run_jobs.py launch --dataset <ns>/<private-dataset> --bucket <ns>/<private-bucket> \
    --leaderboard open_asr_leaderboard --limit 2 --configs us     # smoke test
uv run hf_jobs/run_jobs.py launch --dataset <ns>/<private-dataset> --bucket <ns>/<private-bucket> \
    --leaderboard open_asr_leaderboard                             # full run
uv run hf_jobs/collect.py --bucket <ns>/<private-bucket>

# One table of every model
uv run combine.py --results results results_oss --out all_models.csv
```

Repository: github.com/VoiceArenaAI/Entity-eval (MIT licence).

References

1. Hugging Face. *Open ASR Leaderboard*, evaluation code. github.com/huggingface/open_asr_leaderboard, commit 859159c.

2. H. Wickham. *babynames: US baby names 1880–2017*. github.com/hadley/babynames (source of the first names, section 2.1).
3. OpenRouter. Speech-to-text API. openrouter.ai/docs/guides/overview/multimodal/stt.
4. Each provider's API documentation, as cited in the docstring of every model script.

A Appendix A: Models

Every API model's endpoint and full request is in the docstring of its script, and every open-source model's image, script and arguments in `hf_jobs/models.json` (section 14).

A.1 API models (57, 23 providers)

- **Amazon:** Transcribe (batch); Transcribe Medical (batch); Transcribe Medical Streaming (streaming); Transcribe Streaming (streaming)
- **AssemblyAI:** Universal-3.6 Pro (batch, OpenRouter); STT 3.7 Realtime Agent (streaming); Universal-3.6 Pro Realtime (streaming)
- **Cartesia:** Ink 2 (streaming)
- **Deepgram:** Nova-3 (batch, OpenRouter); Flux English (streaming); Flux Multilingual (streaming); Nova-3 Streaming (streaming)
- **ElevenLabs:** Scribe v2 Medical (batch, OpenRouter); Scribe v3 (batch); Scribe v3 Live (streaming)
- **Fish Audio:** Transcribe 1 Pro (batch, OpenRouter); Transcribe Stream 1 (streaming)
- **Google:** Cloud Chirp 3 (batch, OpenRouter); Gemini 3.5 Transcribe (batch, OpenRouter); Gemini 3.5 Transcribe Live (streaming)
- **Gradium:** STT (streaming)
- **Inworld:** STT-1 (streaming)
- **Meta:** Muse Voice Transcribe 1.0 (batch, OpenRouter); Muse Voice Transcribe Realtime (streaming)
- **Microsoft:** MAI-Transcribe-2 (batch, OpenRouter)
- **Mistral:** Voxtral Mini Transcribe 3 (batch, OpenRouter); Voxtral Small Transcribe 3 (batch, OpenRouter); Voxtral Mini Transcribe Realtime 3 (streaming)
- **Modulate:** Velma-2 STT English (streaming)
- **Murf:** Finch-1 (streaming)
- **Omi Health:** Omi General-1 (batch); Omi Medical Edge-1 (batch); Omi Medical-1 (batch); Omi Medical-1 Flash (batch); Omi General-1 Live (streaming); Omi Medical-1 Live (streaming)
- **OpenAI:** GPT Transcribe (batch, OpenRouter); GPT-4o Mini Transcribe (batch, OpenRouter); GPT Live Transcribe (streaming); GPT Realtime Whisper (streaming)
- **Reson8:** Resonant-1 (batch); Rumble-2 (batch); Rumble-2 Flash (batch); Resonant-1 Realtime (streaming); Rumble-2 Realtime (streaming)
- **Rime:** STT (streaming)
- **Smallest.ai:** Pulse Pro (batch)
- **Soniox:** STT Async v6 (batch); STT Real-time v6 (streaming)
- **Speechmatics:** Enhanced (batch); Melia 1 (batch); Oak 1 (batch); Enhanced Realtime (streaming); Linden 1 (streaming)
- **xAI:** Grok Voice Transcribe 2.0 (batch); Grok Voice Transcribe 2.0 Streaming (streaming)
- **Zoom:** Scribe v2.1 Pro (batch)

A.2 Open-source models (33)

Leaderboard script on an H200 GPU unless marked.

- Qwen/Qwen3-ASR-1.7B-hf
- Qwen/Qwen3-ASR-0.6B-hf
- Musci-research/Musci-ASR-qwen-1019-hf
- CohereLabs/cohere-transcribe-03-2026
- mistralai/Voxtral-Small-24B-2507
- mistralai/Voxtral-Mini-4B-Realtime-2602
- moonshine-ai/moonshine-streaming-medium
- moonshine-ai/moonshine-streaming-tiny
- moonshine-ai/moonshine-base
- openai/whisper-large-v3
- nvidia/nemotron-3.5-asr-streaming-0.6b
- FunAudioLLM/Fun-ASR-Nano-2512-hf
- facebook/mms-1b-all (patched)
- FermionResearch/Phonon-2
- Edge0/ARK-ASR-3B
- nvidia/parakeet-tdt-0.6b-v3
- nvidia/canary-1b-v2
- nvidia/stt_en_fastconformer_hybrid_large_pc
- ibm-granite/granite-speech-5.0-2b
- ibm-granite/granite-speech-5.0-470m-turboctc
- OpenMOSS-Team/MOSS-Transcribe-preview-2B
- futo-org/asr4all-l
- futo-org/asr4all-m
- nvidia/NVIDIA-Nemotron-3.5-Transcribe-30B-A4B
- speechbrain/asr-conformer-transformerlm-librispeech
- facebook/omniASR-LLM-7B
- facebook/omniASR-CTC-7B
- moondream/parakeet-ultra (our runner)
- FunAudioLLM/Fun-ASR-MLT-Nano-2512 (our runner)
- FunAudioLLM/SenseVoiceSmall (our runner)
- Vosk/vosk-model-en-us-0.42-gigaspeech (our runner) (CPU)
- FluidVoice-Labs/fluid-speech-pro
- salesforce-sovereign-intelligence/SOV-Transcribe-1

B Appendix B: Results by accent and difficulty

Every model's TSR on each accent alone (677 sentences per accent), strict and relaxed (section 10.5), on the easy, medium and hard thirds of the sentences (section 11.4), and its relaxed entity accuracy by accent. Rows are in overall rank order; colours rank each column. Entity accuracy by accent is in section 11.3, and the website CSVs carry every accent's full breakdown (entity types, WER, CER). Best per accent: US: Zoom Scribe v2.1 Pro, Filipino: Zoom Scribe v2.1 Pro, Mexican: Zoom Scribe v2.1 Pro, Portuguese: Zoom Scribe v2.1 Pro, E. African: Zoom Scribe v2.1 Pro, Indian: Zoom Scribe v2.1 Pro, Nigerian: Zoom Scribe v2.1 Pro.

B.1 Task success rate by accent and by sentence difficulty

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN	EASY	MEDIUM	HARD
1	Zoom Scribe v2.1 Pro	31.9	38.6	34.6	33.2	31.8	30.6	28.5	26.1	65.3	23.2	7.5
2	ElevenLabs Scribe v2 Medical	28.3	34.0	30.4	28.7	28.8	27.9	23.9	24.5	63.1	16.6	5.4
3	ElevenLabs Scribe v3	27.9	34.1	29.7	28.8	27.3	27.9	24.1	23.5	61.7	17.5	4.8
4	Mistral Voxtral Small Transcribe 3	25.6	31.8	28.5	25.6	24.2	25.4	23.6	20.2	57.8	14.5	4.7
5	Microsoft MAI-Transcribe-2	25.2	32.2	27.2	26.1	25.1	23.5	23.0	19.1	58.6	13.1	4.0
6	Google Gemini 3.5 Transcribe Live	25.0	29.7	28.7	25.1	23.2	24.7	22.9	20.5	57.5	13.0	4.6
7	AssemblyAI Universal-3.6 Pro	24.9	31.0	25.4	27.5	22.9	24.8	23.2	19.8	58.6	12.7	3.7
8	Mistral Voxtral Mini Transcribe 3	24.3	31.2	27.3	24.8	25.3	23.2	20.1	18.5	56.6	13.0	3.5
9	AssemblyAI STT 3.7 Realtime Agent	24.3	30.6	25.6	27.6	22.9	22.6	22.0	18.8	57.7	12.2	3.2
10	Google Gemini 3.5 Transcribe	24.2	28.8	25.8	23.3	23.6	25.8	23.2	19.1	55.9	12.8	4.3
11	Google Cloud Chirp 3	22.5	27.2	23.6	23.2	22.2	22.3	20.2	18.9	53.8	10.3	3.7
12	ElevenLabs Scribe v3 Live	21.9	27.9	23.9	22.7	22.0	21.0	19.5	16.4	52.1	10.9	3.0
13	OpenAI GPT Transcribe	21.6	27.9	21.9	22.7	21.7	20.1	19.5	17.4	51.0	11.1	2.8
14	Mistral Voxtral Small 24B	21.1	27.8	23.9	23.0	19.6	19.9	17.4	16.2	50.1	10.9	2.6
15	OpenAI GPT-4o Mini Transcribe	21.0	28.5	22.7	22.7	21.1	18.3	17.6	16.0	49.0	11.8	2.4
16	AssemblyAI Universal-3.6 Pro Realtime	20.9	27.2	22.6	24.1	19.6	21.0	17.1	14.9	50.2	10.2	2.5
17	Reson8 Rumble-2	19.7	26.0	20.8	20.7	20.5	19.5	15.5	15.1	48.8	8.5	2.1
18	Reson8 Rumble-2 Realtime	19.7	26.1	21.3	21.0	20.2	18.8	15.4	14.9	48.5	8.6	2.0
19	xAI Grok Voice Transcribe 2.0	19.4	25.8	20.2	20.1	19.5	17.7	17.3	14.8	47.0	9.1	2.1
20	Salesforce SOV-Transcribe-1	19.3	23.5	23.2	19.8	19.6	17.6	17.1	14.2	46.9	8.3	2.8
21	Fish Audio Transcribe 1 Pro	19.2	26.7	20.2	21.0	18.2	17.4	16.4	14.5	47.6	8.5	1.7
22	OpenAI Whisper large-v3	19.0	25.4	21.4	21.6	19.5	16.5	15.4	13.1	46.7	8.4	2.0
23	Reson8 Resonant-1	18.9	25.4	20.7	19.9	18.0	18.0	16.1	14.0	45.8	8.5	2.5
24	OpenAI GPT Live Transcribe	18.7	25.3	17.9	19.2	19.2	17.7	16.2	15.2	45.7	8.4	2.0
25	Fish Audio Transcribe Stream 1	18.5	23.3	20.5	19.9	18.8	16.8	15.2	14.8	45.8	7.8	2.0
26	Cartesia Ink 2	18.4	25.6	20.8	20.5	17.9	16.4	15.2	12.6	45.5	7.6	2.4
27	xAI Grok Voice Transcribe 2.0 Streaming	18.3	24.1	19.8	19.4	18.5	16.7	15.7	14.3	45.0	8.1	2.1
28	Soniox STT Real-time v6	18.3	22.3	19.1	20.7	19.2	17.6	16.1	13.0	46.0	6.7	2.2
29	Soniox STT Async v6	18.2	22.3	18.6	21.1	17.9	18.2	17.4	12.1	46.7	6.1	2.0
30	Omi General-1	18.1	25.7	18.6	19.8	18.6	16.8	13.0	14.3	46.0	6.4	2.1
31	Qwen3-ASR 1.7B	17.9	24.8	18.2	19.6	18.8	17.0	13.1	13.7	45.3	6.7	1.9
32	Omi Medical-1	17.5	23.8	18.9	19.5	17.9	16.1	12.1	14.3	44.6	6.5	1.5
33	Omi Medical-1 Flash	17.5	23.8	19.1	19.1	17.6	16.1	12.6	14.0	44.7	6.3	1.5
34	Amazon Transcribe	17.2	23.8	19.2	18.2	16.2	16.1	13.9	12.7	44.0	6.2	1.4
35	Amazon Transcribe Medical	17.2	23.8	19.2	18.2	16.2	16.1	13.9	12.7	44.0	6.2	1.4
36	Meta Muse Voice Transcribe Realtime	17.1	24.8	19.1	19.1	15.8	13.4	15.7	12.0	41.1	8.1	2.2
37	Mistral Voxtral Mini Transcribe Realtime 3	17.1	24.4	20.1	18.5	15.2	15.8	14.8	10.6	43.0	6.3	2.0
38	Speechmatics Enhanced	16.7	21.6	17.6	17.3	15.4	17.3	14.8	13.1	39.4	8.2	2.6
39	OpenAI GPT Realtime Whisper	16.7	24.2	16.1	17.0	15.8	14.9	15.1	13.9	41.1	7.1	2.0
40	Omi Medical-1 Live	16.7	23.2	18.6	19.1	16.1	16.2	11.7	12.0	42.7	6.1	1.4

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN	EASY	MEDIUM	HARD
41	Mistral Voxtral Mini 4B Realtime	16.7	26.3	17.0	20.5	17.3	13.1	10.8	11.8	41.0	7.2	2.0
42	Speechmatics Enhanced Realtime	16.6	21.9	17.4	18.5	15.1	17.1	14.8	11.4	38.7	8.4	2.8
43	Deepgram Nova-3	16.4	22.3	19.1	18.2	16.2	14.6	13.0	11.5	41.7	6.0	1.6
44	Meta Muse Voice Transcribe 1.0	16.3	24.4	17.3	19.2	14.2	13.7	13.1	12.0	40.3	6.6	2.0
45	Reson8 Resonant-1 Realtime	16.1	22.9	17.9	16.2	14.2	16.5	12.9	12.4	39.4	7.0	2.1
46	Smallest.ai Pulse Pro	16.1	22.0	16.4	16.2	16.4	16.0	13.3	12.6	40.4	6.0	2.0
47	Inworld STT-1	15.8	23.5	15.4	18.2	15.2	14.0	12.0	12.1	40.1	6.0	1.3
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	15.7	21.9	15.8	15.5	17.0	14.5	13.6	11.8	38.9	6.6	1.8
49	Musci-ASR qwen-1019	15.7	21.7	16.0	17.1	16.5	14.5	12.7	11.5	40.0	5.7	1.6
50	ARK-ASR 3B	15.7	21.1	17.4	16.2	15.5	14.9	13.1	11.2	40.1	5.2	1.8
51	Omi General-1 Live	15.1	21.4	16.1	16.0	13.7	13.6	12.1	12.6	39.6	4.4	1.3
52	Speechmatics Melia 1	14.4	20.7	16.2	17.1	13.6	12.7	11.1	9.6	36.1	5.7	1.6
53	Speechmatics Linden 1	14.4	19.9	15.8	15.7	12.7	14.5	12.7	9.3	33.9	7.0	2.3
54	Speechmatics Oak 1	14.3	20.5	16.0	16.7	13.6	13.0	11.1	9.5	35.7	5.8	1.6
55	Amazon Transcribe Streaming	14.2	20.5	15.7	15.4	12.6	12.7	12.0	10.5	36.9	5.1	0.7
56	Reson8 Rumble-2 Flash	13.9	19.6	14.9	16.8	12.4	10.9	12.4	10.0	36.5	4.1	1.2
57	Deepgram Nova-3 Streaming	13.9	19.9	16.0	14.8	13.3	12.3	11.4	9.6	35.4	4.9	1.4
58	Amazon Transcribe Medical Streaming	13.8	18.6	15.1	14.8	11.8	14.9	11.7	9.9	36.6	4.3	0.7
59	FluidVoice Fluid Speech Pro	13.8	20.1	13.6	15.8	12.9	12.7	11.8	9.6	35.3	5.0	1.2
60	NVIDIA Canary 1B v2	13.7	21.1	13.4	16.0	14.0	13.0	10.0	8.1	34.9	5.1	1.1
61	Modulate Velma-2 STT English	13.5	21.7	14.9	14.2	12.6	12.4	10.2	8.7	34.1	5.6	1.0
62	Qwen3-ASR 0.6B	13.3	18.9	13.6	14.5	13.7	12.3	10.2	9.7	36.1	3.0	0.8
63	Moondream Parakeet Ultra	13.2	19.6	13.1	14.9	12.0	11.8	10.6	10.0	33.9	4.8	0.9
64	Deepgram Flux English	12.8	19.6	13.7	13.9	11.8	12.1	10.0	8.4	32.4	4.7	1.4
65	Omi Medical Edge-1	12.5	18.3	13.9	13.4	9.9	13.1	9.3	9.6	32.3	4.5	0.9
66	IBM Granite Speech 5.0 2B	12.4	16.1	12.0	15.7	12.6	11.7	10.5	8.7	31.9	4.5	1.0
67	Fun-ASR MLT Nano 2512	12.1	19.6	12.3	14.3	11.7	9.5	9.0	8.1	32.1	3.4	0.8
68	Gradium STT	12.0	18.5	12.1	13.3	12.0	10.0	9.3	9.0	30.9	4.4	0.9
69	Moonshine Streaming Medium	10.8	19.2	10.6	11.4	11.5	7.8	7.1	7.7	27.7	3.7	0.9
70	NVIDIA Parakeet TDT 0.6B v3	10.7	16.4	10.2	11.5	10.2	9.2	8.4	9.0	28.3	3.1	0.8
71	Deepgram Flux Multilingual	9.9	15.4	10.0	11.5	10.8	7.7	7.4	6.4	26.7	2.7	0.3
72	Cohere Transcribe 03-2026	9.6	10.6	12.4	10.2	12.1	4.1	9.7	7.8	24.3	3.5	1.0
73	MOSS Transcribe Preview 2B	9.5	15.2	8.1	10.6	9.5	8.1	7.7	7.5	25.9	2.3	0.5
74	Fun-ASR Nano 2512	9.5	16.1	8.7	12.1	9.5	7.1	7.7	5.5	26.3	2.2	0.2
75	Rime STT	9.2	11.7	8.1	10.3	9.6	9.7	7.4	7.5	24.9	2.2	0.6
76	Meta omniASR LLM 7B	8.7	12.3	9.6	9.3	8.3	8.0	6.8	6.4	23.4	2.1	0.5
77	Moonshine Base	7.6	14.6	8.7	7.5	7.5	4.3	5.6	4.7	20.6	1.8	0.3
78	Murf Finch-1	7.0	12.0	6.5	8.0	7.4	6.2	5.0	4.1	19.4	1.3	0.4
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	6.2	10.0	5.6	6.4	5.0	5.6	5.6	5.5	17.9	0.8	0.1
80	IBM Granite Speech 5.0 470M TurboCTC	6.1	7.7	6.4	6.4	6.2	5.9	5.5	4.4	16.5	1.5	0.2
81	NVIDIA FastConformer Hybrid Large	4.9	8.7	4.7	5.8	4.6	3.4	3.7	3.4	13.5	1.0	0.3
82	SenseVoice Small	3.9	7.1	4.0	3.8	4.1	3.0	3.4	2.2	11.4	0.4	0.1
83	Moonshine Streaming Tiny	3.4	7.7	3.2	4.7	2.8	1.6	2.2	1.8	9.8	0.3	0.3
84	FUTO ASR4All L	3.4	6.6	3.2	3.8	2.5	3.1	2.4	1.9	9.7	0.5	0.0
85	Fermion Phonon-2	3.3	4.6	3.2	3.1	2.8	4.1	2.8	2.4	8.8	1.0	0.1
86	Vosk en-us 0.42 GigaSpeech	3.1	8.1	2.2	3.8	3.4	2.1	1.0	1.2	8.8	0.6	0.0
87	Meta omniASR CTC 7B	3.1	3.7	3.8	3.5	2.8	2.1	3.0	2.7	8.6	0.4	0.2
88	FUTO ASR4All M	2.5	5.8	2.1	2.7	2.2	1.2	1.8	1.8	7.3	0.1	0.1
89	SpeechBrain Conformer + TransformerLM	1.5	3.8	1.9	1.8	1.6	0.1	0.7	0.4	4.4	0.1	0.0
90	Meta MMS 1B	0.3	0.6	0.1	0.4	0.6	0.0	0.1	0.0	0.8	0.1	0.0

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

B.2 Relaxed task success rate by accent

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN
1	Zoom Scribe v2.1 Pro	52.5	63.1	60.6	53.3	51.4	50.5	49.5	39.0
2	ElevenLabs Scribe v2 Medical	47.7	58.8	53.9	48.2	46.8	45.5	42.8	38.1
3	ElevenLabs Scribe v3	46.9	55.7	52.6	47.6	45.8	45.3	44.0	37.5
4	Mistral Voxtral Small Transcribe 3	43.1	54.2	50.2	44.8	41.8	40.2	38.4	32.1
5	Microsoft MAI-Transcribe-2	46.2	57.9	53.0	46.4	45.3	42.7	44.5	33.7
6	Google Gemini 3.5 Transcribe Live	40.5	47.6	48.4	40.3	37.4	40.9	37.1	31.6
7	AssemblyAI Universal-3.6 Pro	43.2	54.8	49.3	46.5	39.4	41.8	39.0	31.5
8	Mistral Voxtral Mini Transcribe 3	42.4	53.9	49.9	43.7	41.1	39.0	37.5	31.3
9	AssemblyAI STT 3.7 Realtime Agent	41.6	54.5	47.1	44.6	39.6	38.1	38.4	28.7
10	Google Gemini 3.5 Transcribe	39.8	46.2	46.8	38.8	37.4	41.7	36.8	30.7
11	Google Cloud Chirp 3	36.8	47.0	42.1	35.7	34.9	34.7	34.9	28.4
12	ElevenLabs Scribe v3 Live	38.3	48.4	45.2	38.7	35.5	36.3	35.9	28.1
13	OpenAI GPT Transcribe	36.6	48.0	40.0	37.1	36.9	33.5	33.7	27.2
14	Mistral Voxtral Small 24B	37.4	50.4	44.3	40.5	35.9	34.0	32.3	24.5
15	OpenAI GPT-4o Mini Transcribe	35.3	47.4	38.8	37.5	36.0	32.5	30.6	24.2
16	AssemblyAI Universal-3.6 Pro Realtime	35.2	47.6	40.0	38.8	30.3	34.3	31.2	24.2
17	Reson8 Rumble-2	35.3	48.0	40.0	37.1	35.6	32.2	30.0	24.5
18	Reson8 Rumble-2 Realtime	35.1	47.7	40.3	37.7	34.1	32.3	30.1	23.5
19	xAI Grok Voice Transcribe 2.0	32.1	43.1	37.8	32.6	31.5	30.4	26.9	22.5
20	Salesforce SOV-Transcribe-1	32.3	40.0	39.6	32.6	32.5	29.0	30.1	22.6
21	Fish Audio Transcribe 1 Pro	33.1	49.3	36.0	35.2	32.2	29.5	27.2	22.2
22	OpenAI Whisper large-v3	34.2	49.6	40.8	36.9	32.3	29.5	28.8	21.7
23	Reson8 Resonant-1	34.5	47.3	38.4	37.1	33.4	32.1	30.1	23.5
24	OpenAI GPT Live Transcribe	30.7	43.7	30.3	31.8	31.0	29.1	25.7	23.0
25	Fish Audio Transcribe Stream 1	31.0	42.1	35.5	32.9	32.9	27.2	25.0	21.6
26	Cartesia Ink 2	31.7	46.7	35.5	35.6	30.6	27.2	26.6	19.6
27	xAI Grok Voice Transcribe 2.0 Streaming	30.0	40.6	36.2	30.4	29.7	27.5	25.3	20.2
28	Soniox STT Real-time v6	32.7	43.0	37.1	35.6	32.2	30.0	29.2	21.9
29	Soniox STT Async v6	32.6	40.9	36.6	36.3	31.8	31.3	31.0	20.2
30	Omi General-1	30.8	44.3	33.7	31.8	31.3	27.3	25.7	21.3
31	Qwen3-ASR 1.7B	31.2	43.6	35.2	31.8	31.9	29.0	25.7	21.4
32	Omi Medical-1	30.3	42.8	34.7	31.8	29.7	27.5	23.3	22.3
33	Omi Medical-1 Flash	30.2	43.9	34.0	31.6	28.2	28.1	23.8	21.7
34	Amazon Transcribe	30.4	42.5	34.0	32.9	27.9	27.8	26.9	21.0
35	Amazon Transcribe Medical	30.4	42.5	34.0	32.9	27.9	27.8	26.9	21.0
36	Meta Muse Voice Transcribe Realtime	29.2	41.2	31.8	31.9	27.8	26.3	26.1	19.1
37	Mistral Voxtral Mini Transcribe Realtime 3	30.6	43.7	37.5	33.8	28.1	26.7	27.0	17.6
38	Speechmatics Enhanced	24.2	32.8	24.8	25.6	20.5	24.8	22.7	18.0
39	OpenAI GPT Realtime Whisper	28.1	43.1	29.4	28.4	26.1	25.8	23.5	20.5
40	Omi Medical-1 Live	27.9	40.2	32.9	30.9	25.1	25.8	21.4	19.2
41	Mistral Voxtral Mini 4B Realtime	29.5	45.9	32.5	35.5	29.4	23.2	21.7	18.0
42	Speechmatics Enhanced Realtime	23.8	34.0	24.4	26.7	20.5	23.2	22.0	15.8
43	Deepgram Nova-3	27.3	37.4	35.3	29.5	26.3	22.7	22.3	17.7
44	Meta Muse Voice Transcribe 1.0	29.0	41.8	32.8	31.5	26.9	26.3	24.2	19.2
45	Reson8 Resonant-1 Realtime	29.7	43.4	34.6	30.7	26.3	28.7	24.2	19.9
46	Smallest.ai Pulse Pro	28.3	39.7	30.7	28.1	27.0	26.7	25.7	19.8
47	Inworld STT-1	26.7	38.8	29.2	29.8	26.4	22.7	21.9	17.7
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	26.2	37.1	28.8	27.6	27.0	23.0	22.3	17.4
49	Musci-ASR qwen-1019	27.0	38.4	30.6	27.5	27.5	24.5	23.0	17.6
50	ARK-ASR 3B	28.4	40.0	31.5	30.4	28.2	25.6	25.3	18.2

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN
51	Omi General-1 Live	25.5	36.6	30.6	25.6	23.8	22.3	22.6	17.0
52	Speechmatics Melia 1	25.2	36.6	30.3	28.5	23.3	22.2	19.9	15.2
53	Speechmatics Linden 1	20.1	28.5	22.0	22.0	17.1	19.2	19.2	12.6
54	Speechmatics Oak 1	24.8	36.6	29.2	27.2	23.5	22.3	19.4	15.4
55	Amazon Transcribe Streaming	23.1	36.8	26.6	23.5	20.5	19.8	20.5	14.3
56	Reson8 Rumble-2 Flash	24.5	37.4	28.2	28.7	22.0	19.8	20.7	14.9
57	Deepgram Nova-3 Streaming	23.2	34.4	27.9	25.7	20.8	19.1	18.9	15.4
58	Amazon Transcribe Medical Streaming	22.2	33.5	24.2	23.2	18.0	22.2	20.1	13.9
59	FluidVoice Fluid Speech Pro	27.1	41.1	29.4	29.7	25.4	24.8	22.9	16.7
60	NVIDIA Canary 1B v2	25.7	39.9	26.7	29.7	25.7	23.0	20.4	14.5
61	Modulate Velma-2 STT English	24.9	39.9	28.5	27.0	22.0	22.2	19.9	15.1
62	Qwen3-ASR 0.6B	22.5	35.3	22.9	24.5	23.2	19.4	17.3	14.6
63	Moondream Parakeet Ultra	25.1	38.1	27.0	26.6	24.7	22.5	21.3	15.4
64	Deepgram Flux English	21.5	31.8	25.1	23.8	19.5	18.9	17.3	13.9
65	Omi Medical Edge-1	24.9	38.8	29.2	26.4	21.9	22.6	19.4	16.2
66	IBM Granite Speech 5.0 2B	22.0	30.9	24.5	26.1	21.9	19.4	17.6	13.4
67	Fun-ASR MLT Nano 2512	19.3	32.3	20.7	23.9	17.6	15.5	13.9	10.9
68	Gradium STT	22.1	36.0	23.3	26.0	21.0	17.7	17.3	13.6
69	Moonshine Streaming Medium	18.3	32.3	17.3	21.0	19.9	13.7	12.9	10.8
70	NVIDIA Parakeet TDT 0.6B v3	20.7	32.3	21.6	21.6	20.4	18.0	16.1	14.9
71	Deepgram Flux Multilingual	16.8	26.6	20.1	17.7	17.4	11.8	13.7	10.0
72	Cohere Transcribe 03-2026	17.3	18.3	23.0	20.7	21.9	7.2	18.5	11.7
73	MOSS Transcribe Preview 2B	15.0	28.4	13.0	17.1	12.9	13.1	10.8	9.9
74	Fun-ASR Nano 2512	15.8	28.4	16.5	19.2	14.8	10.9	11.8	9.0
75	Rime STT	14.7	21.1	13.6	16.0	15.2	14.8	11.2	11.1
76	Meta omniASR LLM 7B	14.6	21.9	15.4	15.5	13.7	14.0	12.4	9.5
77	Moonshine Base	13.4	26.0	16.1	13.4	11.7	8.4	10.8	7.2
78	Murf Finch-1	12.4	23.0	12.6	13.7	12.4	9.6	9.6	6.2
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	10.4	18.3	9.2	10.9	7.5	9.2	10.6	7.2
80	IBM Granite Speech 5.0 470M TurboCTC	9.7	12.6	12.3	9.9	9.5	8.7	8.7	6.5
81	NVIDIA FastConformer Hybrid Large	9.0	18.5	8.9	10.3	7.7	5.8	6.9	4.7
82	SenseVoice Small	6.1	12.6	6.4	5.6	5.5	3.8	4.9	3.7
83	Moonshine Streaming Tiny	5.3	13.6	5.2	6.1	4.6	2.4	3.4	2.1
84	FUTO ASR4All L	5.6	11.5	4.7	6.4	4.9	5.2	4.1	2.4
85	Fermion Phonon-2	6.8	9.0	7.8	5.6	6.4	8.0	6.8	3.8
86	Vosk en-us 0.42 GigaSpeech	4.5	12.6	3.4	5.5	4.6	2.5	1.5	1.5
87	Meta omniASR CTC 7B	5.9	9.3	7.5	5.6	5.6	3.5	5.3	4.4
88	FUTO ASR4All M	4.1	9.3	3.7	4.9	3.4	2.1	3.0	2.5
89	SpeechBrain Conformer + TransformerLM	2.4	6.1	3.4	2.5	2.7	0.3	1.2	0.6
90	Meta MMS 1B	0.9	2.2	0.7	1.3	1.0	0.0	0.6	0.1

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

B.3 Relaxed entity accuracy by accent

Strict entity accuracy by accent is in section 11.3.

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN
1	Zoom Scribe v2.1 Pro	84.8	88.7	88.1	85.4	83.9	84.3	83.9	79.3
2	ElevenLabs Scribe v2 Medical	82.9	87.2	85.8	83.5	82.0	82.4	81.2	78.4
3	ElevenLabs Scribe v3	82.3	85.9	84.8	82.6	81.7	81.6	81.1	78.5
4	Mistral Voxtral Small Transcribe 3	80.5	85.4	84.1	81.2	80.0	79.2	79.6	74.3
5	Microsoft MAI-Transcribe-2	82.0	86.9	84.8	82.9	81.5	80.1	81.5	76.3

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN
6	Google Gemini 3.5 Transcribe Live	79.4	83.8	82.9	78.8	77.8	78.8	79.0	75.1
7	AssemblyAI Universal-3.6 Pro	80.8	86.0	83.3	81.9	78.7	80.7	79.0	75.7
8	Mistral Voxtral Mini Transcribe 3	80.2	85.4	83.4	81.3	79.8	78.7	78.7	74.2
9	AssemblyAI STT 3.7 Realtime Agent	79.8	85.8	82.6	80.6	78.7	79.2	78.1	74.0
10	Google Gemini 3.5 Transcribe	79.4	83.1	82.4	78.6	78.1	79.0	79.0	75.3
11	Google Cloud Chirp 3	77.9	82.3	81.2	77.5	76.3	77.0	77.8	73.2
12	ElevenLabs Scribe v3 Live	78.1	83.0	81.7	78.1	76.8	77.1	77.5	72.6
13	OpenAI GPT Transcribe	77.2	82.7	79.4	77.4	76.8	75.2	76.1	72.7
14	Mistral Voxtral Small 24B	77.4	83.4	80.8	78.4	77.0	75.4	76.1	70.7
15	OpenAI GPT-4o Mini Transcribe	76.3	81.9	79.2	77.4	76.2	75.0	74.3	70.2
16	AssemblyAI Universal-3.6 Pro Realtime	76.5	82.4	79.4	78.2	74.1	76.1	74.2	71.0
17	Reson8 Rumble-2	76.2	82.4	79.3	76.5	75.9	74.3	74.7	70.0
18	Reson8 Rumble-2 Realtime	76.2	82.0	79.6	77.0	75.2	74.5	74.9	69.8
19	xAI Grok Voice Transcribe 2.0	74.0	79.4	77.2	73.9	73.8	73.5	72.7	67.6
20	Salesforce SOV-Transcribe-1	75.0	79.0	79.3	74.6	74.6	73.0	74.9	69.4
21	Fish Audio Transcribe 1 Pro	74.6	82.4	75.9	75.1	74.2	73.2	72.6	68.9
22	OpenAI Whisper large-v3	75.5	82.6	79.2	76.7	75.5	73.6	73.3	67.8
23	Reson8 Resonant-1	75.7	82.2	78.6	76.8	75.8	74.4	73.7	68.6
24	OpenAI GPT Live Transcribe	73.6	80.0	74.4	73.7	72.7	72.5	72.6	69.4
25	Fish Audio Transcribe Stream 1	73.5	79.9	76.1	74.2	73.3	71.7	71.0	68.0
26	Cartesia Ink 2	74.3	81.4	77.3	76.3	72.9	72.0	72.9	67.5
27	xAI Grok Voice Transcribe 2.0 Streaming	72.8	78.4	76.3	73.3	71.9	71.2	72.3	65.9
28	Soniox STT Real-time v6	74.5	80.4	77.7	76.0	73.7	72.7	73.1	68.1
29	Soniox STT Async v6	74.9	79.8	78.3	76.0	74.5	73.6	74.4	67.8
30	Omi General-1	73.4	80.6	75.9	73.6	72.6	71.7	71.2	68.3
31	Qwen3-ASR 1.7B	73.7	80.3	76.2	74.0	73.0	72.2	71.5	68.3
32	Omi Medical-1	72.6	80.1	75.2	73.6	71.6	71.2	69.2	67.0
33	Omi Medical-1 Flash	72.5	80.5	75.2	73.4	70.9	71.6	69.1	67.1
34	Amazon Transcribe	73.7	80.1	76.2	75.3	71.7	71.0	73.0	68.4
35	Amazon Transcribe Medical	73.7	80.1	76.2	75.3	71.7	71.0	73.0	68.4
36	Meta Muse Voice Transcribe Realtime	72.0	78.7	74.6	73.2	70.2	70.8	69.8	66.5
37	Mistral Voxtral Mini Transcribe Realtime 3	73.0	80.6	77.0	75.0	70.7	70.9	71.6	65.4
38	Speechmatics Enhanced	69.8	76.3	71.2	70.7	67.0	69.6	69.3	64.5
39	OpenAI GPT Realtime Whisper	71.3	79.5	72.3	71.4	70.2	70.3	69.3	66.2
40	Omi Medical-1 Live	70.6	78.3	74.5	71.2	67.8	70.0	67.2	65.0
41	Mistral Voxtral Mini 4B Realtime	70.4	80.0	73.0	74.6	70.5	67.1	63.2	63.9
42	Speechmatics Enhanced Realtime	69.6	76.8	70.7	70.6	67.2	68.9	69.3	63.5
43	Deepgram Nova-3	71.5	77.7	76.4	72.9	71.1	69.0	69.3	64.3
44	Meta Muse Voice Transcribe 1.0	72.1	79.9	74.9	73.0	69.9	70.6	69.8	66.6
45	Reson8 Resonant-1 Realtime	72.4	80.1	76.4	73.1	69.7	71.7	70.5	65.2
46	Smallest.ai Pulse Pro	72.0	78.2	74.2	72.1	70.9	71.0	70.8	66.9
47	Inworld STT-1	70.1	78.8	73.2	71.8	69.7	67.9	66.2	63.5
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	70.8	77.4	73.5	70.7	71.2	68.3	69.5	64.7
49	Musci-ASR qwen-1019	70.8	77.6	74.0	70.9	70.5	69.1	69.3	64.2
50	ARK-ASR 3B	71.4	77.4	74.1	72.3	71.0	70.4	70.1	64.4
51	Omi General-1 Live	70.4	76.8	74.1	70.4	69.0	68.5	69.5	64.6
52	Speechmatics Melia 1	70.1	77.8	73.4	72.1	69.3	68.7	67.0	62.4
53	Speechmatics Linden 1	66.1	72.7	68.8	67.1	63.0	65.5	66.4	59.5
54	Speechmatics Oak 1	69.7	77.4	72.9	71.6	68.8	68.3	66.4	62.3
55	Amazon Transcribe Streaming	68.9	77.4	72.5	69.6	66.4	66.8	67.8	61.7
56	Reson8 Rumble-2 Flash	69.1	76.9	72.9	70.0	67.9	66.8	66.3	63.0
57	Deepgram Nova-3 Streaming	67.8	75.9	71.4	69.4	66.5	64.7	65.0	61.8

#	SYSTEM	ALL	US	FILIPINO	MEXICAN	PORTUGUESE	E. AFRICAN	INDIAN	NIGERIAN
58	Amazon Transcribe Medical Streaming	67.9	76.0	70.2	69.5	64.9	66.7	66.9	61.5
59	FluidVoice Fluid Speech Pro	70.9	79.2	73.2	72.3	69.4	69.8	67.9	64.3
60	NVIDIA Canary 1B v2	69.8	78.3	70.8	71.8	69.6	69.1	66.5	62.6
61	Modulate Velma-2 STT English	68.7	78.7	71.7	70.8	66.8	67.1	64.9	61.4
62	Qwen3-ASR 0.6B	67.4	75.6	68.8	68.8	68.1	65.5	63.8	61.1
63	Moondream Parakeet Ultra	69.7	78.5	72.7	70.6	68.6	68.3	66.0	63.4
64	Deepgram Flux English	66.7	73.6	70.1	68.4	65.6	63.8	65.0	60.9
65	Omi Medical Edge-1	69.5	77.8	72.5	70.7	67.7	68.2	65.9	63.6
66	IBM Granite Speech 5.0 2B	66.7	73.8	69.0	68.8	66.0	65.3	64.1	59.9
67	Fun-ASR MLT Nano 2512	64.6	74.4	67.8	66.4	64.7	61.3	60.7	56.6
68	Gradium STT	66.0	75.6	68.4	68.4	65.4	63.5	62.2	58.3
69	Moonshine Streaming Medium	62.2	73.2	62.5	64.0	63.0	58.2	59.3	55.4
70	NVIDIA Parakeet TDT 0.6B v3	66.0	74.0	68.1	66.6	66.1	64.2	63.0	60.1
71	Deepgram Flux Multilingual	63.5	70.4	66.8	64.7	64.3	59.1	62.1	56.8
72	Cohere Transcribe 03-2026	58.6	58.9	63.9	61.6	62.3	47.8	62.8	53.0
73	MOSS Transcribe Preview 2B	60.5	71.4	59.2	61.4	60.8	58.2	58.4	53.9
74	Fun-ASR Nano 2512	61.0	71.5	63.2	63.6	60.2	57.6	56.9	54.1
75	Rime STT	60.1	66.7	61.7	61.4	60.2	59.3	56.2	55.6
76	Meta omniASR LLM 7B	60.7	67.9	62.6	62.6	60.5	60.2	56.8	54.4
77	Moonshine Base	57.3	68.9	59.6	58.9	56.7	52.9	52.7	51.5
78	Murf Finch-1	58.3	68.1	58.0	59.6	59.9	56.2	56.2	50.3
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	54.0	62.7	53.8	54.3	51.9	53.4	55.3	46.7
80	IBM Granite Speech 5.0 470M TurboCTC	55.5	60.8	58.4	56.1	53.4	54.8	54.5	50.5
81	NVIDIA FastConformer Hybrid Large	52.0	62.0	52.3	54.1	51.9	49.9	49.7	43.8
82	SenseVoice Small	47.2	57.7	50.2	47.4	47.1	44.6	43.7	40.0
83	Moonshine Streaming Tiny	39.3	54.8	39.0	42.2	41.9	34.7	33.2	29.5
84	FUTO ASR4All L	48.1	57.2	47.9	49.9	47.9	46.9	46.3	40.7
85	Fermion Phonon-2	47.6	51.9	47.6	45.5	46.0	50.3	46.7	45.1
86	Vosk en-us 0.42 GigaSpeech	40.0	54.9	41.6	46.4	43.0	35.0	32.5	26.6
87	Meta omniASR CTC 7B	46.5	53.0	48.7	44.1	45.1	47.0	43.4	44.0
88	FUTO ASR4All M	43.1	53.6	42.7	45.6	43.6	40.7	40.4	35.4
89	SpeechBrain Conformer + TransformerLM	30.5	46.1	32.5	33.8	34.9	23.9	23.5	18.5
90	Meta MMS 1B	22.8	30.8	24.1	25.4	25.1	19.0	19.1	15.9

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch · API streaming · open weights

C Appendix C: Public and private sets

Every model on the public set alone (133 sentences, 931 clips) and on the private set alone (544 sentences, 3,808 clips): task success rate (TSR) and entity accuracy (EA), strict and relaxed (rel., section 10.5), in %. "#" is the model's rank within that set (strict TSR, then entity accuracy). Rows are in overall rank order; colours rank each column.

C.1 Every model

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	34.4	55.0	77.1	85.9	1	31.3	51.9	76.0	84.5
2	ElevenLabs Scribe v2 Medical	2	29.3	48.9	74.5	83.4	2	28.1	47.5	74.1	82.8
3	ElevenLabs Scribe v3	3	28.9	48.4	74.1	83.1	3	27.7	46.6	73.8	82.1
4	Mistral Voxtral Small Transcribe 3	7	26.4	45.3	72.5	81.4	4	25.4	42.5	71.8	80.3
5	Microsoft MAI-Transcribe-2	4	26.9	49.8	72.8	83.4	5	24.8	45.3	71.8	81.7
6	Google Gemini 3.5 Transcribe Live	8	26.4	43.6	72.0	80.9	6	24.6	39.7	71.5	79.1
7	AssemblyAI Universal-3.6 Pro	6	26.5	45.2	71.8	81.6	7	24.6	42.7	71.6	80.6
8	Mistral Voxtral Mini Transcribe 3	9	26.0	44.4	71.8	81.2	9	23.9	41.9	71.2	80.0
9	AssemblyAI STT 3.7 Realtime Agent	5	26.7	45.5	71.6	81.4	10	23.7	40.6	71.1	79.5
10	Google Gemini 3.5 Transcribe	10	25.3	40.9	71.6	80.7	8	24.0	39.5	71.3	79.1
11	Google Cloud Chirp 3	11	25.0	40.7	71.3	79.8	11	21.9	35.8	70.2	77.5
12	ElevenLabs Scribe v3 Live	12	23.8	41.7	70.2	79.7	13	21.5	37.5	69.2	77.8
13	OpenAI GPT Transcribe	16	21.8	39.3	69.1	78.5	12	21.6	36.0	69.0	76.9
14	Mistral Voxtral Small 24B	15	22.7	39.6	68.8	78.4	14	20.8	36.9	68.3	77.1
15	OpenAI GPT-4o Mini Transcribe	14	22.8	38.9	68.9	77.9	15	20.6	34.4	68.7	76.0
16	AssemblyAI Universal-3.6 Pro Realtime	13	22.8	39.1	69.0	78.2	16	20.5	34.2	68.2	76.1
17	Reson8 Rumble-2	17	21.6	38.8	67.4	77.3	18	19.3	34.5	67.2	75.9
18	Reson8 Rumble-2 Realtime	18	21.3	38.2	67.5	77.5	17	19.3	34.3	67.2	75.8
19	xAI Grok Voice Transcribe 2.0	25	19.7	33.4	66.1	74.4	19	19.3	31.8	66.4	73.9
20	Salesforce SOV-Transcribe-1	21	20.7	34.6	66.9	76.2	20	18.9	31.8	67.2	74.7
21	Fish Audio Transcribe 1 Pro	23	20.6	34.6	67.8	75.9	21	18.9	32.7	66.7	74.3
22	OpenAI Whisper large-v3	20	20.7	37.8	67.8	77.4	24	18.6	33.4	66.2	75.1
23	Reson8 Resonant-1	22	20.7	39.0	65.8	77.4	25	18.4	33.5	66.0	75.4
24	OpenAI GPT Live Transcribe	32	18.9	31.3	66.0	74.2	22	18.6	30.5	66.2	73.5
25	Fish Audio Transcribe Stream 1	38	18.0	31.1	65.8	74.2	23	18.6	31.0	65.9	73.3
26	Cartesia Ink 2	19	21.1	34.2	66.3	75.7	31	17.8	31.1	65.6	74.0
27	xAI Grok Voice Transcribe 2.0 Streaming	31	19.0	32.8	65.5	73.8	26	18.2	29.3	65.3	72.5
28	Soniox STT Real-time v6	26	19.7	37.1	65.1	75.9	28	17.9	31.6	65.5	74.2
29	Soniox STT Async v6	24	19.7	36.0	66.5	76.6	29	17.9	31.8	65.9	74.5
30	Omi General-1	34	18.8	33.2	65.8	74.4	27	18.0	30.2	65.4	73.2
31	Qwen3-ASR 1.7B	39	18.0	33.9	65.7	74.8	30	17.9	30.5	65.4	73.4
32	Omi Medical-1	36	18.3	33.1	65.3	73.9	32	17.3	29.6	64.4	72.2
33	Omi Medical-1 Flash	35	18.4	33.3	65.3	73.9	33	17.2	29.4	64.3	72.2
34	Amazon Transcribe	29	19.1	35.2	66.7	76.4	37	16.7	29.3	65.0	73.0
35	Amazon Transcribe Medical	30	19.1	35.2	66.7	76.4	37	16.7	29.3	65.0	73.0
36	Meta Muse Voice Transcribe Realtime	40	17.8	31.0	64.4	72.8	34	16.9	28.7	64.3	71.8
37	Mistral Voxtral Mini Transcribe Realtime 3	37	18.3	31.6	64.8	73.7	35	16.8	30.4	64.6	72.9
38	Speechmatics Enhanced	33	18.8	26.7	66.1	70.9	42	16.2	23.6	64.6	69.5
39	OpenAI GPT Realtime Whisper	47	17.0	30.5	64.6	73.0	39	16.6	27.5	63.7	70.9
40	Omi Medical-1 Live	48	16.6	30.2	63.6	71.6	36	16.7	27.4	62.9	70.3
41	Mistral Voxtral Mini 4B Realtime	42	17.6	32.0	62.2	71.0	40	16.5	28.8	62.2	70.2
42	Speechmatics Enhanced Realtime	27	19.3	26.5	66.0	70.5	43	15.9	23.1	64.5	69.3

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
43	Deepgram Nova-3	28	19.2	31.1	64.7	73.3	46	15.7	26.4	63.7	71.1
44	Meta Muse Voice Transcribe 1.0	53	15.6	29.8	64.5	73.1	41	16.4	28.8	64.2	71.9
45	Reson8 Resonant-1 Realtime	43	17.5	33.0	63.2	73.6	45	15.8	28.9	63.1	72.1
46	Smallest.ai Pulse Pro	44	17.4	30.4	64.2	72.4	44	15.8	27.7	64.0	71.9
47	Inworld STT-1	41	17.6	30.6	63.1	71.7	49	15.3	25.7	62.6	69.8
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	49	16.3	27.2	63.7	71.6	48	15.6	25.9	63.3	70.6
49	Musci-ASR qwen-1019	51	16.0	29.8	63.1	71.8	47	15.7	26.3	63.3	70.6
50	ARK-ASR 3B	45	17.2	31.9	63.1	72.7	50	15.3	27.6	62.6	71.1
51	Omi General-1 Live	54	15.5	27.6	62.3	70.7	51	15.0	25.0	62.9	70.4
52	Speechmatics Melia 1	57	15.1	27.6	62.8	71.4	52	14.3	24.6	62.1	69.8
53	Speechmatics Linden 1	46	17.1	23.0	63.0	67.2	55	13.7	19.4	61.3	65.9
54	Speechmatics Oak 1	60	14.9	27.1	62.5	71.0	53	14.2	24.2	61.8	69.4
55	Amazon Transcribe Streaming	50	16.0	27.4	64.2	71.3	54	13.7	22.1	62.1	68.4
56	Reson8 Rumble-2 Flash	56	15.3	27.5	61.7	70.8	57	13.6	23.8	60.8	68.7
57	Deepgram Nova-3 Streaming	62	14.7	24.9	60.4	68.3	56	13.7	22.7	60.4	67.7
58	Amazon Transcribe Medical Streaming	52	15.8	26.5	63.0	70.6	59	13.3	21.1	61.0	67.3
59	FluidVoice Fluid Speech Pro	58	15.1	30.4	61.0	71.5	58	13.4	26.3	61.7	70.7
60	NVIDIA Canary 1B v2	55	15.5	29.9	61.4	71.6	60	13.2	24.7	61.1	69.4
61	Modulate Velma-2 STT English	59	15.0	28.4	60.5	70.4	61	13.2	24.1	59.8	68.4
62	Qwen3-ASR 0.6B	61	14.9	25.9	60.3	68.6	63	12.9	21.6	60.1	67.1
63	Moondream Parakeet Ultra	65	14.1	28.9	60.6	70.7	62	12.9	24.1	60.6	69.5
64	Deepgram Flux English	66	13.6	23.0	60.2	67.5	64	12.6	21.1	60.2	66.6
65	Omi Medical Edge-1	68	13.1	28.1	60.7	71.3	65	12.4	24.2	60.3	69.1
66	IBM Granite Speech 5.0 2B	63	14.3	24.3	58.8	67.8	66	12.0	21.4	58.7	66.5
67	Fun-ASR MLT Nano 2512	64	14.2	22.1	58.7	65.9	68	11.6	18.6	57.7	64.2
68	Gradium STT	67	13.5	24.5	58.6	67.1	67	11.7	21.6	57.7	65.7
69	Moonshine Streaming Medium	72	11.3	20.0	57.0	64.1	69	10.6	17.9	55.3	61.8
70	NVIDIA Parakeet TDT 0.6B v3	69	12.8	24.4	57.9	67.3	70	10.2	19.8	57.6	65.7
71	Deepgram Flux Multilingual	71	11.3	20.4	57.7	65.1	71	9.5	15.9	56.6	63.1
72	Cohere Transcribe 03-2026	70	11.5	20.2	52.0	60.0	74	9.1	16.6	51.4	58.3
73	MOSS Transcribe Preview 2B	75	10.3	16.0	56.1	61.2	72	9.3	14.8	55.1	60.3
74	Fun-ASR Nano 2512	74	10.4	17.3	55.2	61.8	73	9.3	15.4	54.5	60.8
75	Rime STT	73	10.8	17.6	55.7	61.9	75	8.8	14.0	53.9	59.7
76	Meta omniASR LLM 7B	76	10.1	16.4	54.0	61.9	76	8.3	14.2	53.3	60.5
77	Moonshine Base	79	7.8	14.8	51.4	57.7	77	7.5	13.0	51.2	57.2
78	Murf Finch-1	77	8.6	15.1	52.7	59.7	78	6.6	11.8	51.6	58.0
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	78	7.9	11.8	48.5	54.2	80	5.8	10.1	48.3	54.0
80	IBM Granite Speech 5.0 470M TurboCTC	80	6.3	10.4	50.4	55.4	79	6.0	9.6	51.1	55.5
81	NVIDIA FastConformer Hybrid Large	82	5.6	10.6	46.6	53.4	81	4.7	8.6	45.5	51.6
82	SenseVoice Small	81	5.9	8.3	44.3	48.1	82	3.5	5.5	43.5	47.0
83	Moonshine Streaming Tiny	85	4.2	6.1	36.8	40.0	84	3.3	5.1	35.8	39.2
84	FUTO ASR4All L	84	4.2	6.6	43.3	48.6	85	3.2	5.4	43.4	48.0
85	Fermion Phonon-2	87	3.4	7.7	39.0	47.4	83	3.3	6.5	40.4	47.7
86	Vosk en-us 0.42 GigaSpeech	83	4.3	5.6	38.9	41.2	87	2.8	4.2	37.1	39.7
87	Meta omniASR CTC 7B	86	3.5	7.6	40.9	47.9	86	3.0	5.5	39.8	46.2
88	FUTO ASR4All M	88	3.0	4.5	39.7	44.0	88	2.4	4.0	38.9	43.0
89	SpeechBrain Conformer + TransformerLM	89	1.5	2.5	28.9	31.6	89	1.5	2.4	27.8	30.2
90	Meta MMS 1B	90	0.2	1.1	17.8	23.9	90	0.3	0.8	17.1	22.5

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

C.2 By accent

Each accent within each set: all 90 models pooled, and the best model on that accent and set (strict TSR, then entity accuracy). Every model's scores on each accent in each set, strict and relaxed, are in Appendix D.

Accent	Pub. TSR	TSR rel.	Pub. EA	EA rel.	Priv. TSR	TSR rel.	Priv. EA	EA rel.
US	20.4	37.2	65.6	75.0	20.9	36.7	66.7	75.4
Filipino	16.9	33.7	61.8	71.7	15.9	28.2	62.1	70.2
Mexican	17.6	30.5	61.1	69.0	16.1	27.0	61.6	69.1
Portuguese	18.4	30.0	63.8	71.9	13.9	23.6	59.1	66.2
East African	14.6	22.8	58.9	65.7	13.9	23.6	59.4	66.3
Indian	12.9	24.8	58.6	67.0	12.5	21.8	58.7	65.6
Nigerian	13.8	20.3	57.5	63.9	10.7	16.9	55.5	61.2

All 90 models pooled, in %. rel.: relaxed (section 10.5).

Accent	Set	Best model	TSR	TSR rel.	EA	EA rel.
US	Public	Zoom Scribe v2.1 Pro	36.1	56.4	77.2	86.1
	Private	Zoom Scribe v2.1 Pro	39.2	64.7	80.2	89.3
Filipino	Public	Zoom Scribe v2.1 Pro	39.8	70.7	80.3	91.3
	Private	Zoom Scribe v2.1 Pro	33.3	58.1	77.4	87.3
Mexican	Public	Zoom Scribe v2.1 Pro	33.1	54.9	76.0	85.3
	Private	Zoom Scribe v2.1 Pro	33.3	52.9	76.9	85.5
Portuguese	Public	Zoom Scribe v2.1 Pro	39.1	57.1	80.1	88.0
	Private	Zoom Scribe v2.1 Pro	30.0	50.0	74.6	83.0
East African	Public	Zoom Scribe v2.1 Pro	30.1	45.1	74.7	81.4
	Private	Zoom Scribe v2.1 Pro	30.7	51.8	76.3	84.9
Indian	Public	Zoom Scribe v2.1 Pro	31.6	57.1	77.4	87.4
	Private	Zoom Scribe v2.1 Pro	27.8	47.6	74.9	83.0
Nigerian	Public	Zoom Scribe v2.1 Pro	30.8	43.6	74.1	81.6
	Private	Zoom Scribe v2.1 Pro	25.0	37.9	71.6	78.7

D Appendix D: Public and private sets by accent

Every model on each accent alone, in the public set (133 sentences, one clip each per accent) and in the private set (544 sentences): task success rate (TSR) and entity accuracy (EA), strict and relaxed (rel., section 10.5), in %. "#" is the model's rank on that accent within that set (strict TSR, then entity accuracy). Rows are in overall rank order; colours rank each column. The same numbers are in the website CSV (`leaderboard_overall_public_private.csv`).

D.1 US English

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	36.1	56.4	77.2	86.1	1	39.2	64.7	80.2	89.3
2	ElevenLabs Scribe v2 Medical	2	33.1	57.9	77.0	86.8	3	34.2	59.0	77.8	87.3
3	ElevenLabs Scribe v3	3	33.1	52.6	75.0	84.1	2	34.4	56.4	77.6	86.3
4	Mistral Voxtral Small Transcribe 3	5	31.6	54.1	74.9	85.1	5	31.8	54.2	75.9	85.5
5	Microsoft MAI-Transcribe-2	9	29.3	58.6	75.2	86.7	4	32.9	57.7	76.7	86.9
6	Google Gemini 3.5 Transcribe Live	12	28.6	48.1	74.7	84.1	9	30.0	47.4	75.9	83.7
7	AssemblyAI Universal-3.6 Pro	8	30.1	51.1	73.7	84.5	7	31.2	55.7	76.2	86.3
8	Mistral Voxtral Mini Transcribe 3	7	30.1	51.9	73.9	84.1	6	31.4	54.4	76.5	85.6
9	AssemblyAI STT 3.7 Realtime Agent	6	30.1	54.1	74.7	85.3	8	30.7	54.6	76.0	85.9
10	Google Gemini 3.5 Transcribe	14	27.8	45.9	74.5	83.6	10	29.0	46.3	75.2	83.0
11	Google Cloud Chirp 3	10	29.3	48.9	73.3	82.6	17	26.7	46.5	73.9	82.2
12	ElevenLabs Scribe v3 Live	4	32.3	54.9	74.7	84.5	15	26.8	46.9	73.4	82.7
13	OpenAI GPT Transcribe	18	25.6	47.4	71.2	81.4	11	28.5	48.2	74.1	82.9
14	Mistral Voxtral Small 24B	16	27.1	47.4	71.2	81.6	13	27.9	51.1	73.6	83.8
15	OpenAI GPT-4o Mini Transcribe	13	28.6	48.9	73.5	83.6	12	28.5	47.1	73.7	81.5
16	AssemblyAI Universal-3.6 Pro Realtime	15	27.8	47.4	72.1	81.6	14	27.0	47.6	73.1	82.6
17	Reson8 Rumble-2	21	24.1	46.6	70.0	81.6	19	26.5	48.3	72.5	82.6
18	Reson8 Rumble-2 Realtime	22	24.1	47.4	69.8	81.6	18	26.7	47.8	72.4	82.1
19	xAI Grok Voice Transcribe 2.0	30	23.3	42.9	69.6	78.9	20	26.5	43.2	71.3	79.5
20	Salesforce SOV-Transcribe-1	38	22.6	38.3	67.7	77.2	38	23.7	40.4	71.1	79.4
21	Fish Audio Transcribe 1 Pro	11	29.3	51.9	73.1	84.3	21	26.1	48.7	72.5	82.0
22	OpenAI Whisper large-v3	19	24.8	47.4	70.6	81.4	26	25.6	50.2	72.4	82.9
23	Reson8 Resonant-1	24	24.1	51.9	69.1	83.2	25	25.7	46.1	71.6	82.0
24	OpenAI GPT Live Transcribe	41	21.8	40.6	69.1	78.7	23	26.1	44.5	71.7	80.3
25	Fish Audio Transcribe Stream 1	45	21.1	41.4	69.8	81.0	37	23.9	42.3	70.4	79.7
26	Cartesia Ink 2	28	23.3	44.4	70.4	79.9	22	26.1	47.2	72.0	81.8
27	xAI Grok Voice Transcribe 2.0 Streaming	31	23.3	40.6	69.1	78.3	34	24.3	40.6	70.2	78.4
28	Soniox STT Real-time v6	54	20.3	45.1	67.7	80.3	42	22.8	42.5	70.4	80.5
29	Soniox STT Async v6	32	23.3	44.4	68.7	80.3	45	22.1	40.1	70.3	79.7
30	Omi General-1	20	24.8	45.1	69.8	79.9	24	25.9	44.1	71.7	80.7
31	Qwen3-ASR 1.7B	42	21.8	42.9	68.9	79.3	27	25.6	43.8	71.6	80.5
32	Omi Medical-1	40	21.8	41.4	70.6	80.1	33	24.3	43.2	70.8	80.1
33	Omi Medical-1 Flash	44	21.1	42.1	70.6	80.3	31	24.4	44.3	70.9	80.6
34	Amazon Transcribe	34	22.6	41.4	69.4	79.5	35	24.1	42.8	71.1	80.3
35	Amazon Transcribe Medical	34	22.6	41.4	69.4	79.5	35	24.1	42.8	71.1	80.3
36	Meta Muse Voice Transcribe Realtime	17	26.3	43.6	71.4	79.3	32	24.4	40.6	70.5	78.6
37	Mistral Voxtral Mini Transcribe Realtime 3	53	20.3	39.8	68.1	77.8	28	25.4	44.7	72.1	81.2
38	Speechmatics Enhanced	26	23.3	32.3	71.6	76.8	54	21.1	32.9	70.3	76.2
39	OpenAI GPT Realtime Whisper	46	21.1	39.8	69.2	78.3	29	25.0	43.9	70.7	79.8
40	Omi Medical-1 Live	47	21.1	37.6	68.3	76.8	39	23.7	40.8	69.9	78.6
41	Mistral Voxtral Mini 4B Realtime	25	24.1	44.4	67.5	78.1	16	26.8	46.3	70.8	80.5

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
42	Speechmatics Enhanced Realtime	27	23.3	33.8	71.4	76.6	50	21.5	34.0	70.6	76.9
43	Deepgram Nova-3	37	22.6	39.1	68.5	78.3	43	22.2	36.9	69.1	77.6
44	Meta Muse Voice Transcribe 1.0	28	23.3	37.6	70.4	78.7	30	24.6	42.8	71.5	80.1
45	Reson8 Resonant-1 Realtime	55	20.3	46.6	67.3	80.5	40	23.5	42.6	70.1	80.0
46	Smallest.ai Pulse Pro	48	21.1	39.8	68.1	77.2	44	22.2	39.7	69.0	78.5
47	Inworld STT-1	23	24.1	41.4	69.2	78.7	41	23.3	38.2	70.2	78.8
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	36	22.6	36.1	68.9	77.9	49	21.7	37.3	68.4	77.2
49	Musci-ASR qwen-1019	49	21.1	39.1	66.7	76.2	46	21.9	38.2	69.2	77.9
50	ARK-ASR 3B	61	19.5	40.6	64.4	76.6	51	21.5	39.9	67.5	77.6
51	Omi General-1 Live	58	19.5	36.1	66.5	75.4	47	21.9	36.8	68.5	77.1
52	Speechmatics Melia 1	70	16.5	36.1	66.3	77.4	48	21.7	36.8	68.9	77.9
53	Speechmatics Linden 1	43	21.8	28.6	68.1	72.7	62	19.5	28.5	67.2	72.7
54	Speechmatics Oak 1	67	17.3	37.6	66.3	77.8	52	21.3	36.4	68.7	77.4
55	Amazon Transcribe Streaming	52	20.3	37.6	68.3	76.8	56	20.6	36.6	69.5	77.6
56	Reson8 Rumble-2 Flash	57	20.3	36.8	66.2	76.6	63	19.5	37.5	67.0	76.9
57	Deepgram Nova-3 Streaming	60	19.5	33.8	65.4	75.2	59	20.0	34.6	67.1	76.0
58	Amazon Transcribe Medical Streaming	66	17.3	31.6	67.1	75.6	64	18.9	34.0	67.8	76.0
59	FluidVoice Fluid Speech Pro	59	19.5	45.1	65.6	78.7	57	20.2	40.1	68.1	79.3
60	NVIDIA Canary 1B v2	49	21.1	45.1	66.7	78.7	55	21.1	38.6	68.5	78.1
61	Modulate Velma-2 STT English	33	23.3	46.6	67.3	80.3	53	21.3	38.2	68.2	78.3
62	Qwen3-ASR 0.6B	39	22.6	39.8	66.7	77.2	68	18.0	34.2	66.3	75.2
63	Moondream Parakeet Ultra	56	20.3	38.3	66.5	76.6	61	19.5	38.1	67.8	78.9
64	Deepgram Flux English	68	17.3	34.6	65.0	74.3	58	20.2	31.1	66.1	73.4
65	Omi Medical Edge-1	64	18.8	42.1	64.8	79.1	66	18.2	38.1	66.6	77.5
66	IBM Granite Speech 5.0 2B	72	15.0	30.1	62.1	72.9	69	16.4	31.1	63.9	74.1
67	Fun-ASR MLT Nano 2512	63	18.8	33.8	65.2	74.5	60	19.9	32.0	66.5	74.4
68	Gradium STT	62	19.5	36.1	64.2	74.1	67	18.2	36.0	66.1	75.9
69	Moonshine Streaming Medium	51	21.1	36.1	66.2	75.0	65	18.8	31.4	64.2	72.8
70	NVIDIA Parakeet TDT 0.6B v3	65	18.0	35.3	62.7	73.5	70	16.0	31.6	64.4	74.1
71	Deepgram Flux Multilingual	69	17.3	31.6	62.1	71.4	73	14.9	25.4	62.2	70.2
72	Cohere Transcribe 03-2026	79	11.3	21.1	51.6	60.0	78	10.5	17.6	52.0	58.6
73	MOSS Transcribe Preview 2B	73	14.3	27.8	63.1	71.4	72	15.4	28.5	63.4	71.4
74	Fun-ASR Nano 2512	71	16.5	28.6	61.9	70.4	71	16.0	28.3	63.6	71.8
75	Rime STT	77	12.0	21.8	59.8	66.5	77	11.6	21.0	59.1	66.7
76	Meta omniASR LLM 7B	78	12.0	21.8	58.2	66.2	75	12.3	21.9	59.3	68.4
77	Moonshine Base	74	14.3	27.1	60.2	68.5	74	14.7	25.7	60.4	69.0
78	Murf Finch-1	75	13.5	28.6	59.6	68.7	76	11.6	21.7	59.5	67.9
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	80	9.8	18.8	54.4	61.1	79	10.1	18.2	55.3	63.1
80	IBM Granite Speech 5.0 470M TurboCTC	82	9.0	13.5	52.8	58.4	81	7.4	12.3	55.8	61.4
81	NVIDIA FastConformer Hybrid Large	83	9.0	19.5	52.2	60.7	80	8.6	18.2	54.2	62.3
82	SenseVoice Small	85	6.8	13.5	52.0	58.0	82	7.2	12.3	52.1	57.6
83	Moonshine Streaming Tiny	81	9.8	16.5	50.3	56.1	83	7.2	12.9	48.7	54.5
84	FUTO ASR4All L	84	9.0	12.0	50.9	57.3	85	6.1	11.4	51.0	57.1
85	Fermion Phonon-2	87	5.3	10.5	43.3	53.0	87	4.4	8.6	43.4	51.7
86	Vosk en-us 0.42 GigaSpeech	76	12.8	17.3	52.2	56.7	84	7.0	11.4	49.4	54.5
87	Meta omniASR CTC 7B	89	3.8	9.0	42.6	50.3	88	3.7	9.4	45.8	53.7
88	FUTO ASR4All M	86	5.3	7.5	47.8	53.2	86	5.9	9.7	48.5	53.7
89	SpeechBrain Conformer + TransformerLM	88	4.5	8.3	42.6	47.2	89	3.7	5.5	41.4	45.8
90	Meta MMS 1B	90	0.0	1.5	24.4	31.7	90	0.7	2.4	23.2	30.6

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

D.2 Filipino English

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	39.8	70.7	80.3	91.3	1	33.3	58.1	77.4	87.3
2	ElevenLabs Scribe v2 Medical	2	30.8	57.9	75.6	86.5	3	30.3	52.9	75.9	85.7
3	ElevenLabs Scribe v3	8	26.3	53.4	74.3	85.5	2	30.5	52.4	75.4	84.6
4	Mistral Voxtral Small Transcribe 3	9	25.6	56.4	75.4	87.4	4	29.2	48.7	74.4	83.3
5	Microsoft MAI-Transcribe-2	4	27.8	60.2	73.5	86.8	7	27.0	51.3	72.9	84.3
6	Google Gemini 3.5 Transcribe Live	3	30.1	54.1	73.7	85.3	5	28.3	47.1	74.1	82.4
7	AssemblyAI Universal-3.6 Pro	7	27.1	54.9	72.1	84.7	10	25.0	48.0	72.4	83.0
8	Mistral Voxtral Mini Transcribe 3	5	27.1	54.1	73.7	85.5	6	27.4	48.9	73.3	82.9
9	AssemblyAI STT 3.7 Realtime Agent	6	27.1	54.9	72.7	85.5	9	25.2	45.2	71.9	81.9
10	Google Gemini 3.5 Transcribe	10	25.6	49.6	71.8	83.4	8	25.9	46.1	73.0	82.1
11	Google Cloud Chirp 3	17	22.6	42.9	72.0	82.0	12	23.9	41.9	72.1	81.1
12	ElevenLabs Scribe v3 Live	18	22.6	47.4	71.8	83.2	11	24.3	44.7	71.3	81.4
13	OpenAI GPT Transcribe	23	21.8	47.4	69.8	82.2	17	21.9	38.2	69.7	78.8
14	Mistral Voxtral Small 24B	11	25.6	49.6	69.8	81.4	13	23.5	43.0	70.5	80.6
15	OpenAI GPT-4o Mini Transcribe	27	20.3	42.1	68.7	79.9	14	23.3	38.1	70.6	79.1
16	AssemblyAI Universal-3.6 Pro Realtime	20	22.6	46.6	69.1	81.0	16	22.6	38.4	69.8	79.0
17	Reson8 Rumble-2	24	21.8	42.9	67.9	79.3	21	20.6	39.3	69.7	79.3
18	Reson8 Rumble-2 Realtime	19	22.6	44.4	69.4	81.0	19	21.0	39.3	69.6	79.3
19	xAI Grok Voice Transcribe 2.0	32	19.5	39.1	67.5	77.6	23	20.4	37.5	68.2	77.1
20	Salesforce SOV-Transcribe-1	12	24.8	46.6	69.6	81.6	15	22.8	37.9	70.1	78.7
21	Fish Audio Transcribe 1 Pro	44	17.3	36.8	67.1	77.6	20	21.0	35.8	67.9	75.5
22	OpenAI Whisper large-v3	13	24.8	51.1	68.9	81.4	22	20.6	38.2	68.2	78.7
23	Reson8 Resonant-1	16	24.1	45.1	66.7	79.9	25	19.9	36.8	67.3	78.3
24	OpenAI GPT Live Transcribe	63	14.3	30.8	64.2	73.9	31	18.8	30.1	67.3	74.6
25	Fish Audio Transcribe Stream 1	48	16.5	35.3	66.5	76.8	18	21.5	35.5	68.1	75.9
26	Cartesia Ink 2	15	24.1	41.4	68.1	80.1	24	20.0	34.0	67.9	76.7
27	xAI Grok Voice Transcribe 2.0 Streaming	26	21.1	45.1	67.5	79.5	27	19.5	34.0	67.2	75.5
28	Soniox STT Real-time v6	28	20.3	45.9	67.5	80.9	32	18.8	34.9	67.0	76.9
29	Soniox STT Async v6	31	19.5	45.1	69.4	81.2	36	18.4	34.6	67.9	77.6
30	Omi General-1	49	16.5	35.3	65.4	76.8	29	19.1	33.3	67.0	75.7
31	Qwen3-ASR 1.7B	56	15.8	38.3	66.2	77.6	33	18.8	34.4	66.9	75.9
32	Omi Medical-1	41	18.0	39.8	65.4	77.2	30	19.1	33.5	66.1	74.7
33	Omi Medical-1 Flash	45	17.3	39.8	65.0	77.0	28	19.5	32.5	66.4	74.8
34	Amazon Transcribe	21	22.6	42.9	68.1	79.7	37	18.4	31.8	66.4	75.4
35	Amazon Transcribe Medical	21	22.6	42.9	68.1	79.7	37	18.4	31.8	66.4	75.4
36	Meta Muse Voice Transcribe Realtime	29	20.3	36.1	66.3	77.0	34	18.8	30.7	65.9	74.1
37	Mistral Voxtral Mini Transcribe Realtime 3	25	21.8	42.1	67.3	79.5	26	19.7	36.4	66.9	76.4
38	Speechmatics Enhanced	50	16.5	24.8	64.8	70.4	39	17.8	24.8	66.1	71.4
39	OpenAI GPT Realtime Whisper	65	14.3	32.3	64.0	74.9	47	16.5	28.7	64.3	71.7
40	Omi Medical-1 Live	40	18.0	38.3	65.8	77.4	35	18.8	31.6	65.3	73.8
41	Mistral Voxtral Mini 4B Realtime	34	19.5	38.3	62.5	72.1	48	16.4	31.1	63.9	73.2
42	Speechmatics Enhanced Realtime	51	16.5	24.1	63.8	69.6	42	17.6	24.4	65.5	70.9
43	Deepgram Nova-3	14	24.8	49.6	68.3	80.3	40	17.6	31.8	66.3	75.5
44	Meta Muse Voice Transcribe 1.0	57	15.8	34.6	65.2	75.6	41	17.6	32.4	65.7	74.7
45	Reson8 Resonant-1 Realtime	30	20.3	39.8	65.2	78.1	43	17.3	33.3	65.0	75.9
46	Smallest.ai Pulse Pro	58	15.8	32.3	64.4	74.5	46	16.5	30.3	65.7	74.2
47	Inworld STT-1	59	15.8	32.3	63.8	74.7	55	15.3	28.5	64.9	72.8
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	63	14.3	27.1	64.2	72.1	49	16.2	29.2	64.9	73.8
49	Musci-ASR qwen-1019	68	12.8	30.8	62.7	73.9	45	16.7	30.5	65.5	74.1
50	ARK-ASR 3B	36	18.8	38.3	64.2	74.9	44	17.1	29.8	65.0	74.0

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
51	Omi General-1 Live	42	18.0	35.3	64.0	75.0	51	15.6	29.4	65.5	73.8
52	Speechmatics Melia 1	33	19.5	40.6	64.8	76.6	53	15.4	27.8	63.8	72.7
53	Speechmatics Linden 1	61	15.0	22.6	61.7	67.3	50	16.0	21.9	64.4	69.2
54	Speechmatics Oak 1	42	18.0	36.8	64.0	75.2	54	15.4	27.4	63.7	72.4
55	Amazon Transcribe Streaming	39	18.0	33.1	66.7	76.6	56	15.1	25.0	64.7	71.6
56	Reson8 Rumble-2 Flash	53	16.5	33.1	63.1	72.7	57	14.5	27.0	63.5	73.0
57	Deepgram Nova-3 Streaming	46	17.3	33.1	63.1	73.1	52	15.6	26.7	62.6	71.0
58	Amazon Transcribe Medical Streaming	35	18.8	32.3	64.6	74.3	58	14.2	22.2	62.9	69.3
59	FluidVoice Fluid Speech Pro	60	15.8	35.3	62.3	74.1	62	13.1	27.9	62.6	72.9
60	NVIDIA Canary 1B v2	54	16.5	36.1	62.9	74.5	64	12.7	24.4	60.8	69.9
61	Modulate Velma-2 STT English	37	18.8	35.3	62.7	74.5	59	14.0	26.8	61.4	71.1
62	Qwen3-ASR 0.6B	69	12.8	22.6	58.8	66.3	60	13.8	23.0	62.0	69.4
63	Moondream Parakeet Ultra	61	15.0	34.6	61.7	73.7	63	12.7	25.2	62.3	72.5
64	Deepgram Flux English	38	18.8	29.3	61.9	70.6	65	12.5	24.1	62.3	70.0
65	Omi Medical Edge-1	52	16.5	38.3	63.6	75.6	61	13.2	27.0	61.8	71.8
66	IBM Granite Speech 5.0 2B	74	9.8	24.1	57.8	69.1	66	12.5	24.6	60.2	69.0
67	Fun-ASR MLT Nano 2512	55	16.5	27.1	60.2	68.9	68	11.2	19.1	59.9	67.5
68	Gradium STT	66	13.5	30.8	59.6	70.4	67	11.8	21.5	59.3	67.9
69	Moonshine Streaming Medium	71	10.5	20.3	57.6	64.6	70	10.7	16.5	55.9	62.0
70	NVIDIA Parakeet TDT 0.6B v3	67	13.5	27.8	58.6	70.0	72	9.4	20.0	58.4	67.6
71	Deepgram Flux Multilingual	70	12.0	24.8	59.0	66.7	71	9.6	18.9	59.2	66.9
72	Cohere Transcribe 03-2026	47	17.3	27.1	53.2	62.7	69	11.2	22.1	55.8	64.1
73	MOSS Transcribe Preview 2B	72	10.5	16.5	55.9	61.1	77	7.5	12.1	53.3	58.7
74	Fun-ASR Nano 2512	75	9.0	19.5	55.3	62.9	75	8.6	15.8	56.1	63.3
75	Rime STT	75	9.0	19.5	55.3	63.6	76	7.9	12.1	54.9	61.3
76	Meta omniASR LLM 7B	73	10.5	18.8	54.5	63.8	73	9.4	14.5	55.3	62.3
77	Moonshine Base	78	8.3	18.0	50.5	58.4	74	8.8	15.6	52.7	59.9
78	Murf Finch-1	81	5.3	13.5	49.3	57.6	78	6.8	12.3	51.8	58.1
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	79	8.3	12.8	47.4	54.4	80	5.0	8.3	48.0	53.6
80	IBM Granite Speech 5.0 470M TurboCTC	80	7.5	16.5	53.8	60.2	79	6.1	11.2	52.6	57.9
81	NVIDIA FastConformer Hybrid Large	84	3.8	9.0	44.5	52.4	81	5.0	8.8	45.6	52.3
82	SenseVoice Small	77	9.0	9.8	45.6	50.9	86	2.8	5.5	46.1	50.0
83	Moonshine Streaming Tiny	85	3.8	4.5	34.6	38.1	85	3.1	5.3	36.0	39.2
84	FUTO ASR4All L	89	1.5	3.0	41.6	46.4	82	3.7	5.1	43.8	48.2
85	Fermion Phonon-2	86	3.0	10.5	37.9	47.8	84	3.3	7.2	39.2	47.5
86	Vosk en-us 0.42 GigaSpeech	83	4.5	6.0	38.7	41.8	88	1.7	2.8	38.9	41.6
87	Meta omniASR CTC 7B	82	5.3	13.5	43.5	51.5	83	3.5	6.1	41.3	48.1
88	FUTO ASR4All M	88	2.3	3.8	38.7	42.0	87	2.0	3.7	38.6	42.9
89	SpeechBrain Conformer + TransformerLM	87	3.0	4.5	28.0	31.3	89	1.7	3.1	29.8	32.7
90	Meta MMS 1B	90	0.0	0.8	15.5	22.1	90	0.2	0.7	18.3	24.6

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch API streaming open weights

D.3 Mexican English

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	33.1	54.9	76.0	85.3	1	33.3	52.9	76.9	85.5
2	ElevenLabs Scribe v2 Medical	2	30.1	50.4	74.5	83.2	3	28.3	47.6	74.4	83.5
3	ElevenLabs Scribe v3	5	27.8	48.1	73.3	82.6	2	29.0	47.4	74.1	82.6
4	Mistral Voxtral Small Transcribe 3	6	27.8	48.1	72.0	80.9	7	25.0	43.9	72.1	81.2
5	Microsoft MAI-Transcribe-2	4	29.3	51.9	74.9	84.7	6	25.4	45.0	72.8	82.5

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
6	Google Gemini 3.5 Transcribe Live	9	25.6	41.4	70.6	78.7	8	25.0	40.1	71.5	78.8
7	AssemblyAI Universal-3.6 Pro	8	27.1	46.6	70.8	80.9	4	27.6	46.5	72.3	82.1
8	Mistral Voxtral Mini Transcribe 3	7	27.1	45.9	71.8	80.9	9	24.3	43.2	71.9	81.4
9	AssemblyAI STT 3.7 Realtime Agent	3	30.1	46.6	72.0	81.2	5	27.0	44.1	71.8	80.4
10	Google Gemini 3.5 Transcribe	15	23.3	40.6	69.8	79.1	12	23.3	38.4	70.9	78.5
11	Google Cloud Chirp 3	11	24.8	39.8	70.0	77.9	14	22.8	34.7	70.3	77.4
12	ElevenLabs Scribe v3 Live	10	25.6	42.1	69.2	77.9	16	22.1	37.9	69.6	78.2
13	OpenAI GPT Transcribe	24	21.8	38.3	69.6	77.9	13	23.0	36.8	69.7	77.2
14	Mistral Voxtral Small 24B	33	20.3	37.6	67.9	77.6	11	23.7	41.2	69.8	78.6
15	OpenAI GPT-4o Mini Transcribe	16	23.3	39.1	68.9	78.1	15	22.6	37.1	70.1	77.3
16	AssemblyAI Universal-3.6 Pro Realtime	17	23.3	42.1	68.5	78.9	10	24.3	38.1	69.5	78.0
17	Reson8 Rumble-2	12	24.8	42.1	67.3	77.6	26	19.7	35.8	67.1	76.2
18	Reson8 Rumble-2 Realtime	13	24.1	41.4	67.9	77.4	24	20.2	36.8	67.4	76.9
19	xAI Grok Voice Transcribe 2.0	36	20.3	30.1	63.6	71.6	25	20.0	33.3	66.7	74.5
20	Salesforce SOV-Transcribe-1	23	22.6	36.1	65.6	74.1	30	19.1	31.8	67.0	74.7
21	Fish Audio Transcribe 1 Pro	25	21.8	36.1	67.3	74.7	20	20.8	34.9	67.1	75.2
22	OpenAI Whisper large-v3	27	21.1	36.8	67.7	77.0	17	21.7	36.9	66.9	76.7
23	Reson8 Resonant-1	51	18.0	39.8	64.6	77.4	23	20.4	36.4	66.8	76.7
24	OpenAI GPT Live Transcribe	30	21.1	32.3	65.4	72.9	35	18.8	31.6	66.7	73.8
25	Fish Audio Transcribe Stream 1	31	21.1	35.3	65.2	74.7	27	19.7	32.4	65.9	74.1
26	Cartesia Ink 2	28	21.1	34.6	66.2	75.0	22	20.4	35.8	67.6	76.6
27	xAI Grok Voice Transcribe 2.0 Streaming	42	19.5	29.3	62.7	69.8	28	19.3	30.7	66.6	74.1
28	Soniox STT Real-time v6	38	19.5	37.6	65.2	75.2	19	21.0	35.1	67.8	76.2
29	Soniox STT Async v6	39	19.5	38.3	64.8	76.4	18	21.5	35.8	67.2	75.9
30	Omi General-1	20	22.6	36.1	66.7	74.9	31	19.1	30.7	65.3	73.2
31	Qwen3-ASR 1.7B	20	22.6	36.8	66.7	75.0	33	18.9	30.5	65.8	73.8
32	Omi Medical-1	22	22.6	36.1	66.3	74.5	36	18.8	30.7	65.3	73.4
33	Omi Medical-1 Flash	19	22.6	36.8	66.9	75.0	40	18.2	30.3	64.8	73.1
34	Amazon Transcribe	49	18.0	35.3	66.2	75.8	38	18.2	32.4	66.5	75.2
35	Amazon Transcribe Medical	49	18.0	35.3	66.2	75.8	38	18.2	32.4	66.5	75.2
36	Meta Muse Voice Transcribe Realtime	40	19.5	36.8	64.4	73.3	34	18.9	30.7	65.4	73.2
37	Mistral Voxtral Mini Transcribe Realtime 3	58	16.5	32.3	64.0	73.9	32	18.9	34.2	66.8	75.2
38	Speechmatics Enhanced	37	19.5	27.1	66.2	69.6	44	16.7	25.2	65.8	71.0
39	OpenAI GPT Realtime Whisper	52	18.0	31.6	64.4	72.5	45	16.7	27.6	63.8	71.1
40	Omi Medical-1 Live	32	21.1	35.3	62.9	70.6	37	18.6	29.8	63.6	71.4
41	Mistral Voxtral Mini 4B Realtime	35	20.3	37.6	64.4	73.7	21	20.6	34.9	66.5	74.9
42	Speechmatics Enhanced Realtime	14	24.1	30.8	67.1	70.0	42	17.1	25.7	65.9	70.7
43	Deepgram Nova-3	18	23.3	34.6	64.4	73.3	43	16.9	28.3	65.0	72.8
44	Meta Muse Voice Transcribe 1.0	43	18.8	35.3	65.8	73.9	29	19.3	30.5	65.2	72.8
45	Reson8 Resonant-1 Realtime	61	15.0	30.8	61.3	72.1	47	16.5	30.7	63.9	73.3
46	Smallest.ai Pulse Pro	34	20.3	36.1	64.8	73.1	56	15.3	26.1	63.5	71.9
47	Inworld STT-1	26	21.8	33.8	62.3	69.8	41	17.3	28.9	64.7	72.3
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	62	15.0	26.3	60.9	69.1	53	15.6	27.9	63.2	71.1
49	Musci-ASR qwen-1019	29	21.1	34.6	65.8	72.9	51	16.2	25.7	63.1	70.5
50	ARK-ASR 3B	44	18.8	36.8	63.6	73.5	52	15.6	28.9	63.2	72.1
51	Omi General-1 Live	46	18.8	30.8	62.7	71.0	57	15.3	24.3	62.6	70.3
52	Speechmatics Melia 1	41	19.5	33.1	63.2	72.5	46	16.5	27.4	64.0	71.9
53	Speechmatics Linden 1	53	18.0	24.8	62.9	66.3	59	15.1	21.3	62.7	67.2
54	Speechmatics Oak 1	45	18.8	32.3	63.2	72.0	50	16.2	25.9	63.6	71.5
55	Amazon Transcribe Streaming	60	15.0	27.8	63.4	70.6	55	15.4	22.4	62.9	69.4
56	Reson8 Rumble-2 Flash	47	18.8	31.6	61.3	69.2	48	16.4	27.9	61.7	70.2
57	Deepgram Nova-3 Streaming	54	18.0	30.8	60.9	69.8	63	14.0	24.4	61.8	69.3

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
58	Amazon Transcribe Medical Streaming	65	14.3	27.8	62.5	70.0	60	14.9	22.1	61.9	69.3
59	FluidVoice Fluid Speech Pro	66	14.3	30.8	61.1	71.8	49	16.2	29.4	64.0	72.5
60	NVIDIA Canary 1B v2	55	17.3	34.6	61.7	72.0	54	15.6	28.5	62.7	71.7
61	Modulate Velma-2 STT English	69	13.5	30.8	58.8	70.4	62	14.3	26.1	61.6	70.9
62	Qwen3-ASR 0.6B	56	17.3	30.8	61.3	69.4	64	13.8	23.0	61.0	68.7
63	Moondream Parakeet Ultra	67	13.5	31.6	61.3	71.2	58	15.3	25.4	62.0	70.5
64	Deepgram Flux English	63	15.0	27.1	59.8	67.3	65	13.6	23.0	61.9	68.6
65	Omi Medical Edge-1	68	13.5	31.6	60.2	71.0	67	13.4	25.2	61.9	70.7
66	IBM Granite Speech 5.0 2B	48	18.8	31.6	60.7	70.6	61	14.9	24.8	59.7	68.4
67	Fun-ASR MLT Nano 2512	57	17.3	27.1	59.2	66.3	66	13.6	23.2	59.7	66.5
68	Gradium STT	59	15.8	28.6	59.8	67.3	68	12.7	25.4	60.3	68.7
69	Moonshine Streaming Medium	72	12.8	23.3	56.7	63.2	72	11.0	20.4	57.9	64.1
70	NVIDIA Parakeet TDT 0.6B v3	74	11.3	25.6	56.7	65.8	70	11.6	20.6	59.0	66.8
71	Deepgram Flux Multilingual	71	12.8	22.6	58.2	66.0	71	11.2	16.5	57.5	64.4
72	Cohere Transcribe 03-2026	76	10.5	20.3	52.0	59.8	73	10.1	20.8	54.2	62.0
73	MOSS Transcribe Preview 2B	70	13.5	17.3	55.7	59.2	74	9.9	17.1	56.7	61.9
74	Fun-ASR Nano 2512	73	12.0	21.8	56.3	63.2	69	12.1	18.6	57.5	63.7
75	Rime STT	64	15.0	23.3	58.2	63.8	75	9.2	14.2	55.0	60.8
76	Meta omniASR LLM 7B	75	11.3	19.5	56.1	63.8	76	8.8	14.5	55.3	62.3
77	Moonshine Base	78	8.3	15.8	52.0	57.1	78	7.4	12.9	52.9	59.3
78	Murf Finch-1	77	9.8	15.8	51.5	57.4	77	7.5	13.2	53.4	60.1
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	80	6.8	13.5	47.0	52.4	80	6.2	10.3	49.4	54.8
80	IBM Granite Speech 5.0 470M TurboCTC	82	5.3	9.0	49.7	54.2	79	6.6	10.1	52.3	56.6
81	NVIDIA FastConformer Hybrid Large	79	8.3	15.8	49.9	56.1	81	5.1	9.0	47.4	53.7
82	SenseVoice Small	84	5.3	7.5	44.5	47.0	86	3.5	5.1	43.9	47.5
83	Moonshine Streaming Tiny	81	6.0	8.3	41.0	43.5	82	4.4	5.5	38.3	41.9
84	FUTO ASR4All L	83	5.3	9.8	45.1	49.9	85	3.5	5.5	45.2	49.9
85	Fermion Phonon-2	88	2.3	6.8	34.2	42.7	87	3.3	5.3	39.4	46.2
86	Vosk en-us 0.42 GigaSpeech	85	4.5	5.3	43.5	44.7	83	3.7	5.5	44.0	46.8
87	Meta omniASR CTC 7B	87	3.0	6.0	40.0	45.6	84	3.7	5.5	37.4	43.8
88	FUTO ASR4All M	86	3.8	6.8	41.4	45.5	88	2.4	4.4	41.2	45.7
89	SpeechBrain Conformer + TransformerLM	89	0.8	1.5	34.0	36.9	89	2.0	2.8	31.1	33.1
90	Meta MMS 1B	90	0.8	3.0	20.1	25.5	90	0.4	0.9	19.7	25.4

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

D.4 Portuguese English

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	39.1	57.1	80.1	88.0	1	30.0	50.0	74.6	83.0
2	ElevenLabs Scribe v2 Medical	5	30.1	48.1	74.5	83.8	2	28.5	46.5	73.2	81.6
3	ElevenLabs Scribe v3	3	31.6	48.1	75.2	83.9	3	26.3	45.2	72.6	81.2
4	Mistral Voxtral Small Transcribe 3	11	27.8	45.9	73.7	81.8	6	23.3	40.8	70.7	79.5
5	Microsoft MAI-Transcribe-2	8	28.6	51.9	74.3	84.9	4	24.3	43.8	71.1	80.7
6	Google Gemini 3.5 Transcribe Live	9	28.6	44.4	73.5	82.6	8	21.9	35.7	68.8	76.7
7	AssemblyAI Universal-3.6 Pro	4	30.8	48.9	74.9	83.8	11	21.0	37.1	69.4	77.5
8	Mistral Voxtral Mini Transcribe 3	6	29.3	43.6	73.7	81.4	5	24.3	40.4	71.0	79.4
9	AssemblyAI STT 3.7 Realtime Agent	2	32.3	51.9	74.3	84.5	12	20.6	36.6	69.4	77.3
10	Google Gemini 3.5 Transcribe	10	28.6	42.1	73.1	82.6	7	22.4	36.2	69.4	77.1
11	Google Cloud Chirp 3	7	29.3	45.1	73.1	81.6	13	20.4	32.4	68.0	75.1
12	ElevenLabs Scribe v3 Live	17	24.8	39.8	71.6	80.7	9	21.3	34.4	67.7	75.9

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
13	OpenAI GPT Transcribe	20	24.1	39.8	70.8	79.9	10	21.1	36.2	68.0	76.1
14	Mistral Voxtral Small 24B	13	26.3	45.1	71.2	81.8	21	18.0	33.6	67.0	75.9
15	OpenAI GPT-4o Mini Transcribe	18	24.8	42.9	70.6	80.1	14	20.2	34.4	67.8	75.4
16	AssemblyAI Universal-3.6 Pro Realtime	14	26.3	38.3	71.0	79.3	22	18.0	28.3	66.0	72.8
17	Reson8 Rumble-2	16	25.6	42.9	69.8	80.5	15	19.3	33.8	66.6	74.8
18	Reson8 Rumble-2 Realtime	23	24.1	39.1	69.4	79.5	16	19.3	32.9	66.3	74.2
19	xAI Grok Voice Transcribe 2.0	29	22.6	35.3	68.7	77.8	17	18.8	30.5	65.4	72.9
20	Salesforce SOV-Transcribe-1	19	24.8	39.8	70.2	79.3	18	18.4	30.7	66.0	73.6
21	Fish Audio Transcribe 1 Pro	35	20.3	35.3	69.1	77.6	26	17.6	31.4	65.6	73.4
22	OpenAI Whisper large-v3	12	26.3	43.6	72.1	82.2	24	17.8	29.6	65.4	73.9
23	Reson8 Resonant-1	15	26.3	44.4	70.0	81.4	36	16.0	30.7	65.3	74.5
24	OpenAI GPT Live Transcribe	22	24.1	36.8	69.8	77.9	23	18.0	29.6	64.2	71.5
25	Fish Audio Transcribe Stream 1	32	21.1	36.1	68.9	76.8	19	18.2	32.2	64.8	72.5
26	Cartesia Ink 2	24	24.1	35.3	69.2	78.1	32	16.4	29.4	63.4	71.6
27	xAI Grok Voice Transcribe 2.0 Streaming	33	21.1	31.6	67.5	75.8	25	17.8	29.2	63.6	71.0
28	Soniox STT Real-time v6	27	23.3	39.1	68.1	78.3	20	18.2	30.5	64.0	72.6
29	Soniox STT Async v6	37	20.3	33.8	68.9	77.9	29	17.3	31.2	65.1	73.7
30	Omi General-1	25	23.3	39.8	69.4	77.0	28	17.5	29.2	63.8	71.5
31	Qwen3-ASR 1.7B	21	24.1	40.6	70.0	78.1	27	17.5	29.8	64.0	71.8
32	Omi Medical-1	30	22.6	36.1	67.9	76.0	30	16.7	28.1	63.0	70.6
33	Omi Medical-1 Flash	28	23.3	35.3	67.7	75.2	35	16.2	26.5	62.2	69.8
34	Amazon Transcribe	35	20.3	36.8	69.1	77.9	38	15.3	25.7	62.5	70.2
35	Amazon Transcribe Medical	37	20.3	36.8	68.9	77.8	38	15.3	25.7	62.5	70.2
36	Meta Muse Voice Transcribe Realtime	54	16.5	29.3	65.8	75.4	37	15.6	27.4	61.4	69.0
37	Mistral Voxtral Mini Transcribe Realtime 3	44	19.5	33.1	66.7	75.2	48	14.2	26.8	61.5	69.6
38	Speechmatics Enhanced	43	19.5	25.6	67.3	72.3	47	14.3	19.3	60.9	65.8
39	OpenAI GPT Realtime Whisper	47	18.8	29.3	67.3	75.8	42	15.1	25.4	61.3	68.9
40	Omi Medical-1 Live	45	19.5	28.6	64.6	71.4	40	15.3	24.3	60.2	67.0
41	Mistral Voxtral Mini 4B Realtime	34	21.1	39.1	66.9	76.2	33	16.4	27.0	61.5	69.2
42	Speechmatics Enhanced Realtime	49	18.0	24.8	66.5	71.8	46	14.3	19.5	61.1	66.2
43	Deepgram Nova-3	26	23.3	32.3	68.9	76.2	44	14.5	24.8	62.6	69.9
44	Meta Muse Voice Transcribe 1.0	65	14.3	32.3	66.0	75.4	49	14.2	25.6	60.8	68.7
45	Reson8 Resonant-1 Realtime	42	20.3	33.8	64.6	74.5	54	12.7	24.4	60.2	68.6
46	Smallest.ai Pulse Pro	31	21.8	31.6	66.9	74.1	41	15.1	25.9	62.7	70.1
47	Inworld STT-1	51	17.3	32.3	66.2	74.3	43	14.7	25.0	61.4	68.6
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	40	20.3	30.8	66.3	74.7	34	16.2	26.1	63.4	70.4
49	Musci-ASR qwen-1019	57	16.5	30.1	65.0	73.7	31	16.5	26.8	62.8	69.7
50	ARK-ASR 3B	39	20.3	31.6	66.5	76.0	45	14.3	27.4	61.2	69.8
51	Omi General-1 Live	69	13.5	24.1	61.9	70.2	50	13.8	23.7	61.0	68.7
52	Speechmatics Melia 1	66	14.3	25.6	65.8	73.3	51	13.4	22.8	60.8	68.4
53	Speechmatics Linden 1	48	18.8	23.3	63.8	68.5	64	11.2	15.6	56.9	61.7
54	Speechmatics Oak 1	63	15.0	24.1	66.3	72.7	52	13.2	23.3	60.3	67.9
55	Amazon Transcribe Streaming	55	16.5	27.1	65.6	72.3	58	11.6	18.9	58.9	65.0
56	Reson8 Rumble-2 Flash	61	15.8	30.1	65.0	74.1	57	11.6	20.0	59.5	66.5
57	Deepgram Nova-3 Streaming	62	15.8	21.8	62.5	68.5	55	12.7	20.6	59.1	66.1
58	Amazon Transcribe Medical Streaming	55	16.5	26.3	65.6	72.1	68	10.7	16.0	57.9	63.2
59	FluidVoice Fluid Speech Pro	46	19.5	36.1	64.4	75.4	62	11.2	22.8	59.3	68.0
60	NVIDIA Canary 1B v2	41	20.3	34.6	65.4	76.2	56	12.5	23.5	59.7	68.0
61	Modulate Velma-2 STT English	58	16.5	31.6	64.0	72.9	59	11.6	19.7	57.7	65.4
62	Qwen3-ASR 0.6B	53	17.3	30.8	62.9	72.0	53	12.9	21.3	60.0	67.2
63	Moondream Parakeet Ultra	59	16.5	31.6	63.2	73.9	66	10.8	23.0	58.8	67.4
64	Deepgram Flux English	67	14.3	23.3	64.2	71.6	63	11.2	18.6	58.2	64.2

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
65	Omi Medical Edge-1	73	12.0	28.6	63.1	73.7	72	9.4	20.2	58.2	66.3
66	IBM Granite Speech 5.0 2B	60	16.5	28.6	61.9	71.8	60	11.6	20.2	57.1	64.7
67	Fun-ASR MLT Nano 2512	64	15.0	21.8	62.1	70.0	67	10.8	16.5	57.4	63.4
68	Gradium STT	52	17.3	28.6	63.1	71.6	69	10.7	19.1	56.1	63.9
69	Moonshine Streaming Medium	71	12.8	24.8	61.9	70.8	65	11.2	18.8	54.7	61.2
70	NVIDIA Parakeet TDT 0.6B v3	50	18.0	30.1	63.2	72.5	75	8.3	18.0	56.4	64.6
71	Deepgram Flux Multilingual	72	12.8	20.3	60.0	67.7	70	10.3	16.7	57.1	63.6
72	Cohere Transcribe 03-2026	68	14.3	30.1	58.2	70.0	61	11.6	19.9	53.0	60.5
73	MOSS Transcribe Preview 2B	74	10.5	14.3	58.4	64.0	74	9.2	12.5	55.0	60.1
74	Fun-ASR Nano 2512	77	9.0	14.3	58.8	66.3	71	9.6	14.9	53.1	58.8
75	Rime STT	74	10.5	18.8	58.4	64.8	73	9.4	14.3	53.5	59.2
76	Meta omniASR LLM 7B	70	13.5	18.8	56.3	65.6	77	7.0	12.5	52.4	59.4
77	Moonshine Base	78	9.0	15.0	56.9	63.1	76	7.2	10.8	50.1	55.2
78	Murf Finch-1	76	10.5	16.5	56.9	65.4	78	6.6	11.4	51.9	58.6
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	81	7.5	9.8	48.4	53.2	80	4.4	7.0	46.2	51.6
80	IBM Granite Speech 5.0 470M TurboCTC	79	7.5	11.3	51.8	56.1	79	5.9	9.0	49.2	52.8
81	NVIDIA FastConformer Hybrid Large	80	7.5	12.0	50.5	58.0	81	3.9	6.6	44.4	50.5
82	SenseVoice Small	82	7.5	9.0	46.8	51.3	82	3.3	4.6	42.3	46.1
83	Moonshine Streaming Tiny	86	3.8	5.3	43.9	47.6	85	2.6	4.4	37.1	40.5
84	FUTO ASR4All L	87	2.3	5.3	47.6	51.8	84	2.6	4.8	43.0	47.0
85	Fermion Phonon-2	84	5.3	10.5	40.6	49.7	87	2.2	5.3	38.1	45.2
86	Vosk en-us 0.42 GigaSpeech	85	4.5	6.0	45.1	47.2	83	3.1	4.2	39.2	42.1
87	Meta omniASR CTC 7B	83	5.3	9.8	41.4	49.3	88	2.2	4.6	37.7	44.2
88	FUTO ASR4All M	88	1.5	2.3	42.4	45.8	86	2.4	3.7	39.4	43.1
89	SpeechBrain Conformer + TransformerLM	89	1.5	2.3	36.6	39.7	89	1.7	2.8	31.2	33.8
90	Meta MMS 1B	90	0.8	2.3	22.2	30.0	90	0.6	0.7	19.0	23.9

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

D.5 East African English

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	30.1	45.1	74.7	81.4	1	30.7	51.8	76.3	84.9
2	ElevenLabs Scribe v2 Medical	4	25.6	41.4	72.0	80.1	2	28.5	46.5	74.7	82.9
3	ElevenLabs Scribe v3	2	27.1	43.6	73.1	80.5	3	28.1	45.8	73.7	81.9
4	Mistral Voxtral Small Transcribe 3	5	24.8	36.1	69.8	76.6	5	25.6	41.2	71.4	79.8
5	Microsoft MAI-Transcribe-2	11	21.8	39.8	69.8	78.3	8	23.9	43.4	71.2	80.6
6	Google Gemini 3.5 Transcribe Live	9	23.3	38.3	70.0	77.8	7	25.0	41.5	71.5	79.0
7	AssemblyAI Universal-3.6 Pro	6	24.1	38.3	71.0	78.9	6	25.0	42.6	72.7	81.1
8	Mistral Voxtral Mini Transcribe 3	12	21.8	35.3	69.1	77.4	9	23.5	39.9	70.3	78.9
9	AssemblyAI STT 3.7 Realtime Agent	16	21.1	36.8	69.8	78.5	10	23.0	38.4	71.4	79.4
10	Google Gemini 3.5 Transcribe	3	26.3	37.6	71.2	78.5	4	25.7	42.6	71.6	79.1
11	Google Cloud Chirp 3	8	23.3	33.1	70.4	77.8	11	22.1	35.1	70.3	76.8
12	ElevenLabs Scribe v3 Live	7	24.1	37.6	69.8	77.2	13	20.2	36.0	68.9	77.1
13	OpenAI GPT Transcribe	18	21.1	33.8	67.5	75.6	14	19.9	33.5	67.5	75.1
14	Mistral Voxtral Small 24B	13	21.8	33.8	66.3	74.9	16	19.5	34.0	67.3	75.5
15	OpenAI GPT-4o Mini Transcribe	10	22.6	36.1	68.1	76.4	23	17.3	31.6	67.1	74.7
16	AssemblyAI Universal-3.6 Pro Realtime	17	21.1	35.3	68.7	76.8	12	21.0	34.0	68.0	75.9
17	Reson8 Rumble-2	24	18.8	28.6	65.8	73.3	15	19.7	33.1	66.3	74.5
18	Reson8 Rumble-2 Realtime	24	18.8	30.1	65.8	73.9	17	18.8	32.9	66.3	74.7
19	xAI Grok Voice Transcribe 2.0	35	17.3	30.8	63.6	70.6	18	17.8	30.3	66.2	74.1

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
20	Salesforce SOV-Transcribe-1	33	17.3	27.8	63.8	72.3	20	17.6	29.2	66.6	73.1
21	Fish Audio Transcribe 1 Pro	18	21.1	28.6	67.5	73.7	29	16.5	29.8	65.7	73.0
22	OpenAI Whisper large-v3	30	18.0	30.1	64.6	73.3	36	16.2	29.4	65.4	73.6
23	Reson8 Resonant-1	29	18.8	32.3	64.2	73.5	19	17.8	32.0	65.7	74.6
24	OpenAI GPT Live Transcribe	26	18.8	29.3	65.4	72.7	22	17.5	29.0	65.2	72.5
25	Fish Audio Transcribe Stream 1	31	17.3	24.8	64.2	70.6	28	16.7	27.8	65.0	71.9
26	Cartesia Ink 2	14	21.8	30.8	65.0	72.7	40	15.1	26.3	63.6	71.9
27	xAI Grok Voice Transcribe 2.0 Streaming	32	17.3	30.8	64.0	71.4	31	16.5	26.7	64.2	71.2
28	Soniox STT Real-time v6	23	20.3	34.6	63.8	73.5	27	16.9	28.9	64.4	72.5
29	Soniox STT Async v6	22	20.3	33.1	64.8	73.9	21	17.6	30.9	65.7	73.5
30	Omi General-1	41	15.8	24.1	64.4	70.4	25	17.1	28.1	64.9	72.0
31	Qwen3-ASR 1.7B	43	15.8	25.6	63.6	70.8	24	17.3	29.8	64.9	72.5
32	Omi Medical-1	47	14.3	25.6	63.8	72.0	32	16.5	27.9	63.6	71.1
33	Omi Medical-1 Flash	45	15.0	26.3	64.2	71.8	34	16.4	28.5	63.8	71.5
34	Amazon Transcribe	27	18.8	28.6	64.8	72.9	38	15.4	27.6	63.4	70.5
35	Amazon Transcribe Medical	27	18.8	28.6	64.8	72.9	38	15.4	27.6	63.4	70.5
36	Meta Muse Voice Transcribe Realtime	49	13.5	24.8	61.5	69.4	51	13.4	26.7	63.5	71.1
37	Mistral Voxtral Mini Transcribe Realtime 3	15	21.8	28.6	63.8	71.4	45	14.3	26.3	62.7	70.7
38	Speechmatics Enhanced	21	20.3	28.6	65.8	70.0	30	16.5	23.9	64.9	69.5
39	OpenAI GPT Realtime Whisper	39	16.5	29.3	63.1	70.2	43	14.5	25.0	63.4	70.3
40	Omi Medical-1 Live	55	12.8	23.3	62.5	70.0	26	17.1	26.5	63.0	70.0
41	Mistral Voxtral Mini 4B Realtime	48	14.3	21.8	59.0	66.7	57	12.9	23.5	60.1	67.2
42	Speechmatics Enhanced Realtime	20	20.3	24.8	66.7	70.2	33	16.4	22.8	64.2	68.5
43	Deepgram Nova-3	52	13.5	18.0	59.2	66.7	42	14.9	23.9	62.2	69.5
44	Meta Muse Voice Transcribe 1.0	59	12.0	25.6	60.9	69.4	48	14.2	26.5	63.5	70.9
45	Reson8 Resonant-1 Realtime	37	17.3	29.3	62.7	71.0	35	16.4	28.5	63.3	71.9
46	Smallest.ai Pulse Pro	42	15.8	22.6	63.8	70.4	37	16.0	27.8	63.9	71.1
47	Inworld STT-1	33	17.3	24.1	63.8	70.8	54	13.2	22.4	60.7	67.2
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	46	15.0	24.8	62.1	69.1	46	14.3	22.6	61.9	68.2
49	Musci-ASR qwen-1019	60	12.0	21.1	60.3	67.5	41	15.1	25.4	62.7	69.5
50	ARK-ASR 3B	40	16.5	25.6	61.5	69.2	44	14.5	25.6	62.1	70.6
51	Omi General-1 Live	51	13.5	19.5	60.7	66.5	50	13.6	23.0	62.5	69.0
52	Speechmatics Melia 1	58	12.0	18.0	61.3	67.1	56	12.9	23.2	61.4	69.1
53	Speechmatics Linden 1	36	17.3	20.3	62.9	66.0	49	13.8	18.9	61.0	65.4
54	Speechmatics Oak 1	64	11.3	18.0	60.7	67.1	52	13.4	23.3	61.2	68.6
55	Amazon Transcribe Streaming	50	13.5	21.1	61.1	67.5	59	12.5	19.5	61.1	66.6
56	Reson8 Rumble-2 Flash	62	12.0	21.1	58.0	66.2	66	10.7	19.5	58.9	67.0
57	Deepgram Nova-3 Streaming	57	12.8	20.3	57.1	64.2	62	12.1	18.8	58.4	64.8
58	Amazon Transcribe Medical Streaming	38	17.3	24.1	61.9	68.5	47	14.3	21.7	60.5	66.3
59	FluidVoice Fluid Speech Pro	54	13.5	22.6	58.2	66.9	58	12.5	25.4	61.4	70.4
60	NVIDIA Canary 1B v2	65	11.3	18.0	58.4	67.5	53	13.4	24.3	61.0	69.4
61	Modulate Velma-2 STT English	61	12.0	18.0	58.4	66.0	60	12.5	23.2	58.8	67.3
62	Qwen3-ASR 0.6B	53	13.5	19.5	58.4	66.0	64	11.9	19.3	58.9	65.4
63	Moondream Parakeet Ultra	66	11.3	21.8	57.8	66.5	63	11.9	22.6	59.8	68.7
64	Deepgram Flux English	67	10.5	14.3	55.1	60.2	61	12.5	20.0	58.4	64.6
65	Omi Medical Edge-1	56	12.8	21.8	59.2	67.5	55	13.2	22.8	60.4	68.4
66	IBM Granite Speech 5.0 2B	44	15.8	20.3	57.3	63.2	65	10.7	19.1	59.0	65.8
67	Fun-ASR MLT Nano 2512	63	12.0	17.3	56.7	62.5	70	8.8	15.1	54.5	61.1
68	Gradium STT	72	9.8	16.5	54.5	63.1	67	10.1	18.0	55.8	63.6
69	Moonshine Streaming Medium	74	8.3	12.8	50.9	57.3	72	7.7	14.0	52.2	58.4
70	NVIDIA Parakeet TDT 0.6B v3	67	10.5	16.5	55.1	62.7	69	8.8	18.4	56.5	64.6
71	Deepgram Flux Multilingual	73	8.3	12.8	54.0	59.8	73	7.5	11.6	53.4	58.9

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
72	Cohere Transcribe 03-2026	81	5.3	8.3	43.1	47.6	81	3.9	7.0	42.4	47.8
73	MOSS Transcribe Preview 2B	77	6.8	13.5	52.6	57.4	71	8.5	13.1	53.7	58.4
74	Fun-ASR Nano 2512	76	7.5	12.0	50.5	56.9	75	7.0	10.7	51.8	57.8
75	Rime STT	69	10.5	15.8	54.5	60.0	68	9.6	14.5	53.6	59.1
76	Meta omniASR LLM 7B	71	9.8	13.5	54.9	60.7	74	7.5	14.2	53.0	60.1
77	Moonshine Base	79	5.3	7.5	45.6	51.6	79	4.0	8.6	47.8	53.2
78	Murf Finch-1	74	8.3	11.3	50.9	56.1	77	5.7	9.2	50.3	56.2
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	70	10.5	11.3	48.0	53.4	78	4.4	8.6	48.0	53.5
80	IBM Granite Speech 5.0 470M TurboCTC	83	3.8	6.8	48.9	52.8	76	6.4	9.2	51.3	55.2
81	NVIDIA FastConformer Hybrid Large	80	5.3	6.8	45.1	50.5	82	2.9	5.5	44.3	49.8
82	SenseVoice Small	84	3.0	5.3	41.0	44.1	83	2.9	3.5	42.1	44.7
83	Moonshine Streaming Tiny	88	0.8	0.8	30.2	32.9	87	1.8	2.8	32.4	35.1
84	FUTO ASR4All L	78	6.0	7.5	41.6	47.2	84	2.4	4.6	42.1	46.9
85	Fermion Phonon-2	82	5.3	7.5	40.4	46.8	80	3.9	8.1	43.9	51.1
86	Vosk en-us 0.42 GigaSpeech	87	0.8	1.5	33.8	36.2	85	2.4	2.8	33.0	34.7
87	Meta omniASR CTC 7B	85	2.3	3.8	41.4	47.2	86	2.0	3.5	40.8	46.9
88	FUTO ASR4All M	86	2.3	3.0	37.9	42.6	88	0.9	1.8	36.6	40.2
89	SpeechBrain Conformer + TransformerLM	89	0.0	0.0	21.1	22.4	89	0.2	0.4	22.4	24.2
90	Meta MMS 1B	90	0.0	0.0	14.9	20.1	90	0.0	0.0	14.0	18.7

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

D.6 Indian English

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	31.6	57.1	77.4	87.4	1	27.8	47.6	74.9	83.0
2	ElevenLabs Scribe v2 Medical	3	27.1	45.1	74.1	82.4	3	23.2	42.3	72.1	81.0
3	ElevenLabs Scribe v3	2	27.8	51.9	74.3	84.3	4	23.2	42.1	72.1	80.3
4	Mistral Voxtral Small Transcribe 3	10	22.6	39.1	72.0	81.0	2	23.9	38.2	70.9	79.2
5	Microsoft MAI-Transcribe-2	5	26.3	46.6	72.5	83.4	7	22.2	43.9	70.9	81.1
6	Google Gemini 3.5 Transcribe Live	4	27.1	43.6	72.5	81.0	8	21.9	35.5	70.5	78.5
7	AssemblyAI Universal-3.6 Pro	7	24.8	42.9	71.2	81.0	6	22.8	38.1	70.2	78.5
8	Mistral Voxtral Mini Transcribe 3	11	21.1	41.4	70.8	80.9	10	19.9	36.6	69.8	78.2
9	AssemblyAI STT 3.7 Realtime Agent	9	23.3	41.4	69.4	78.5	9	21.7	37.7	70.2	78.1
10	Google Gemini 3.5 Transcribe	6	24.8	39.1	71.8	80.9	5	22.8	36.2	70.9	78.5
11	Google Cloud Chirp 3	8	24.1	42.9	72.0	81.2	13	19.3	32.9	69.6	77.0
12	ElevenLabs Scribe v3 Live	16	18.0	36.8	67.1	78.1	11	19.9	35.7	68.4	77.4
13	OpenAI GPT Transcribe	12	19.5	40.6	68.3	77.9	12	19.5	32.0	68.0	75.7
14	Mistral Voxtral Small 24B	22	15.8	33.1	69.1	78.1	14	17.8	32.2	67.0	75.6
15	OpenAI GPT-4o Mini Transcribe	13	19.5	35.3	67.3	76.0	17	17.1	29.4	67.2	73.9
16	AssemblyAI Universal-3.6 Pro Realtime	14	19.5	37.6	66.9	77.0	19	16.5	29.6	66.5	73.6
17	Reson8 Rumble-2	24	15.8	34.6	65.4	75.4	27	15.4	28.9	65.7	74.6
18	Reson8 Rumble-2 Realtime	25	15.8	34.6	65.2	76.0	29	15.3	29.0	65.9	74.6
19	xAI Grok Voice Transcribe 2.0	20	16.5	29.3	65.8	74.5	16	17.5	26.3	65.5	72.3
20	Salesforce SOV-Transcribe-1	23	15.8	29.3	66.7	76.8	15	17.5	30.3	66.9	74.4
21	Fish Audio Transcribe 1 Pro	17	18.0	30.1	65.4	72.9	24	16.0	26.5	66.0	72.5
22	OpenAI Whisper large-v3	50	11.3	28.6	65.4	74.7	21	16.4	28.9	64.5	72.9
23	Reson8 Resonant-1	33	15.0	30.8	62.9	74.7	21	16.4	30.0	64.5	73.5
24	OpenAI GPT Live Transcribe	26	15.8	27.8	65.0	73.5	20	16.4	25.2	65.5	72.4
25	Fish Audio Transcribe Stream 1	32	15.0	24.8	65.0	71.4	30	15.3	25.0	65.3	71.0
26	Cartesia Ink 2	21	16.5	30.1	63.2	72.9	31	14.9	25.7	64.6	72.8

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
27	xAI Grok Voice Transcribe 2.0 Streaming	38	13.5	27.8	65.6	74.5	23	16.2	24.6	64.9	71.9
28	Soniox STT Real-time v6	18	18.0	32.3	63.4	73.9	25	15.6	28.5	64.2	72.9
29	Soniox STT Async v6	15	19.5	34.6	65.8	76.6	18	16.9	30.1	65.1	73.9
30	Omi General-1	44	12.8	30.8	62.1	71.8	44	13.1	24.4	63.0	71.1
31	Qwen3-ASR 1.7B	52	11.3	29.3	61.3	71.4	38	13.6	24.8	63.2	71.5
32	Omi Medical-1	36	14.3	28.6	61.5	69.6	54	11.6	22.1	61.4	69.1
33	Omi Medical-1 Flash	36	14.3	28.6	61.5	69.8	49	12.1	22.6	61.2	68.9
34	Amazon Transcribe	30	15.0	35.3	65.8	77.2	36	13.6	24.8	64.2	72.1
35	Amazon Transcribe Medical	30	15.0	35.3	65.8	77.2	36	13.6	24.8	64.2	72.1
36	Meta Muse Voice Transcribe Realtime	29	15.8	28.6	61.3	69.1	26	15.6	25.6	62.9	70.0
37	Mistral Voxtral Mini Transcribe Realtime 3	35	14.3	26.3	62.7	71.6	32	14.9	27.2	63.7	71.6
38	Speechmatics Enhanced	27	15.8	24.8	63.8	69.6	33	14.5	22.2	64.3	69.2
39	OpenAI GPT Realtime Whisper	39	13.5	29.3	62.9	71.6	28	15.4	22.1	62.3	68.8
40	Omi Medical-1 Live	57	10.5	24.1	60.9	69.1	51	11.9	20.8	60.0	66.8
41	Mistral Voxtral Mini 4B Realtime	69	8.3	20.3	54.5	62.7	56	11.4	22.1	55.7	63.3
42	Speechmatics Enhanced Realtime	19	17.3	26.3	64.2	69.6	34	14.2	21.0	64.7	69.3
43	Deepgram Nova-3	49	12.0	20.3	62.9	70.6	41	13.2	22.8	62.1	69.0
44	Meta Muse Voice Transcribe 1.0	51	11.3	22.6	62.5	70.0	39	13.6	24.6	62.6	69.7
45	Reson8 Resonant-1 Realtime	53	11.3	23.3	59.6	69.8	42	13.2	24.4	61.8	70.6
46	Smallest.ai Pulse Pro	40	13.5	28.6	60.9	70.0	40	13.2	25.0	63.2	71.0
47	Inworld STT-1	60	10.5	24.1	57.8	67.3	48	12.3	21.3	59.3	66.0
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	48	12.0	24.8	63.6	71.8	35	14.0	21.7	62.8	69.0
49	Musci-ASR qwen-1019	42	13.5	30.1	60.2	70.2	47	12.5	21.3	61.9	69.1
50	ARK-ASR 3B	44	12.8	30.1	62.1	72.7	42	13.2	24.1	61.8	69.4
51	Omi General-1 Live	58	10.5	26.3	59.6	69.2	46	12.5	21.7	62.0	69.6
52	Speechmatics Melia 1	46	12.8	21.8	59.4	67.3	59	10.7	19.5	59.6	66.9
53	Speechmatics Linden 1	28	15.8	24.8	61.7	66.7	50	11.9	17.8	61.9	66.4
54	Speechmatics Oak 1	46	12.8	22.6	59.4	67.5	60	10.7	18.6	59.2	66.2
55	Amazon Transcribe Streaming	34	14.3	25.6	63.6	70.2	55	11.4	19.3	61.3	67.2
56	Reson8 Rumble-2 Flash	54	11.3	22.6	59.4	69.8	45	12.7	20.2	59.1	65.5
57	Deepgram Nova-3 Streaming	62	9.0	19.5	56.7	65.2	53	11.9	18.8	58.1	64.9
58	Amazon Transcribe Medical Streaming	41	13.5	25.6	60.5	69.8	57	11.2	18.8	60.4	66.2
59	FluidVoice Fluid Speech Pro	56	11.3	23.3	57.6	67.3	52	11.9	22.8	59.7	68.0
60	NVIDIA Canary 1B v2	59	10.5	21.8	58.0	67.7	64	9.9	20.0	58.6	66.2
61	Modulate Velma-2 STT English	68	8.3	18.0	54.7	63.8	61	10.7	20.4	57.3	65.1
62	Qwen3-ASR 0.6B	66	8.3	20.3	56.7	65.8	62	10.7	16.5	57.0	63.4
63	Moondream Parakeet Ultra	63	9.0	22.6	56.3	66.3	58	11.0	21.0	57.6	65.9
64	Deepgram Flux English	61	9.8	17.3	56.9	64.0	63	10.1	17.3	59.3	65.2
65	Omi Medical Edge-1	74	6.8	20.3	56.9	66.7	65	9.9	19.1	57.5	65.7
66	IBM Granite Speech 5.0 2B	43	13.5	20.3	56.9	65.4	66	9.7	16.9	56.7	63.8
67	Fun-ASR MLT Nano 2512	64	9.0	14.3	54.2	61.7	69	9.0	13.8	54.8	60.5
68	Gradium STT	71	7.5	16.5	54.2	61.9	67	9.7	17.5	55.0	62.3
69	Moonshine Streaming Medium	79	5.3	12.0	52.2	60.0	72	7.5	13.1	52.9	59.2
70	NVIDIA Parakeet TDT 0.6B v3	70	8.3	18.0	54.2	63.4	70	8.5	15.6	55.3	62.9
71	Deepgram Flux Multilingual	67	8.3	18.8	56.3	64.2	75	7.2	12.5	55.4	61.6
72	Cohere Transcribe 03-2026	55	11.3	21.1	57.8	66.0	68	9.4	17.8	54.5	62.0
73	MOSS Transcribe Preview 2B	75	6.8	9.0	56.1	60.3	71	7.9	11.2	53.6	58.0
74	Fun-ASR Nano 2512	65	9.0	12.0	51.1	56.3	74	7.4	11.8	51.6	57.1
75	Rime STT	76	6.8	9.8	52.6	57.6	73	7.5	11.6	50.6	55.8
76	Meta omniASR LLM 7B	73	7.5	15.0	49.9	58.6	76	6.6	11.8	49.7	56.4
77	Moonshine Base	80	5.3	13.5	47.0	53.6	77	5.7	10.1	47.3	52.5
78	Murf Finch-1	72	7.5	12.8	52.0	59.4	80	4.4	8.8	49.5	55.5

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	77	6.0	9.0	49.9	56.7	78	5.5	11.0	48.8	55.0
80	IBM Granite Speech 5.0 470M TurboCTC	78	6.0	10.5	48.5	54.4	79	5.3	8.3	50.0	54.6
81	NVIDIA FastConformer Hybrid Large	82	3.0	7.5	44.3	51.5	81	3.9	6.8	43.2	49.3
82	SenseVoice Small	81	4.5	6.0	40.0	42.7	83	3.1	4.6	40.8	43.9
83	Moonshine Streaming Tiny	85	3.0	5.3	28.6	31.3	86	2.0	2.9	30.9	33.6
84	FUTO ASR4All L	86	2.3	5.3	39.5	47.0	85	2.4	3.9	41.3	46.1
85	Fermion Phonon-2	83	3.0	6.8	37.3	45.1	84	2.8	6.8	39.7	47.1
86	Vosk en-us 0.42 GigaSpeech	88	1.5	1.5	30.4	31.9	88	0.9	1.5	31.1	32.6
87	Meta omniASR CTC 7B	87	1.5	6.0	36.4	44.3	82	3.3	5.1	36.9	43.2
88	FUTO ASR4All M	84	3.0	5.3	36.2	42.0	87	1.5	2.4	36.1	40.0
89	SpeechBrain Conformer + TransformerLM	89	0.8	0.8	21.1	22.8	89	0.7	1.3	22.3	23.7
90	Meta MMS 1B	90	0.0	0.0	14.1	20.7	90	0.2	0.7	13.5	18.8

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights

D.7 Nigerian English

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
1	Zoom Scribe v2.1 Pro	1	30.8	43.6	74.1	81.6	1	25.0	37.9	71.6	78.7
2	ElevenLabs Scribe v2 Medical	2	28.6	41.4	73.9	81.0	2	23.5	37.3	70.7	77.8
3	ElevenLabs Scribe v3	3	28.6	41.4	73.7	81.0	3	22.2	36.6	70.9	78.0
4	Mistral Voxtral Small Transcribe 3	5	24.8	37.6	70.0	76.8	6	19.1	30.7	67.1	73.7
5	Microsoft MAI-Transcribe-2	6	24.8	39.8	69.6	78.9	9	17.6	32.2	67.2	75.7
6	Google Gemini 3.5 Transcribe Live	8	21.8	35.3	69.2	76.4	4	20.2	30.7	68.2	74.8
7	AssemblyAI Universal-3.6 Pro	9	21.8	33.8	68.7	77.2	5	19.3	30.9	68.0	75.4
8	Mistral Voxtral Mini Transcribe 3	4	25.6	38.3	69.6	78.1	12	16.7	29.6	65.8	73.3
9	AssemblyAI STT 3.7 Realtime Agent	7	23.3	33.1	68.5	76.2	10	17.6	27.6	67.0	73.4
10	Google Gemini 3.5 Transcribe	12	21.1	31.6	69.4	76.6	7	18.6	30.5	68.3	75.0
11	Google Cloud Chirp 3	9	21.8	32.3	68.7	75.6	8	18.2	27.4	67.1	72.6
12	ElevenLabs Scribe v3 Live	15	19.5	33.1	67.3	76.6	13	15.6	26.8	65.3	71.6
13	OpenAI GPT Transcribe	17	18.8	27.8	66.7	74.7	11	17.1	27.0	65.7	72.2
14	Mistral Voxtral Small 24B	11	21.8	30.8	66.2	73.5	15	14.9	23.0	63.2	70.1
15	OpenAI GPT-4o Mini Transcribe	14	20.3	27.8	65.0	71.0	14	14.9	23.3	64.4	70.1
16	AssemblyAI Universal-3.6 Pro Realtime	17	18.8	26.3	66.7	72.5	19	14.0	23.7	64.1	70.7
17	Reson8 Rumble-2	13	20.3	33.8	65.4	73.5	24	13.8	22.2	62.2	69.2
18	Reson8 Rumble-2 Realtime	16	19.5	30.8	65.0	73.3	23	13.8	21.7	62.2	69.0
19	xAI Grok Voice Transcribe 2.0	21	18.0	26.3	63.6	70.0	22	14.0	21.5	61.4	67.0
20	Salesforce SOV-Transcribe-1	23	17.3	24.1	64.4	71.8	27	13.4	22.2	62.6	68.9
21	Fish Audio Transcribe 1 Pro	25	16.5	23.3	64.8	70.8	20	14.0	21.9	62.4	68.5
22	OpenAI Whisper large-v3	19	18.8	27.1	65.2	71.6	38	11.8	20.4	60.2	67.0
23	Reson8 Resonant-1	20	18.8	28.6	63.2	71.6	30	12.9	22.2	60.6	67.9
24	OpenAI GPT Live Transcribe	29	16.5	21.1	62.9	69.8	16	14.9	23.5	62.7	69.3
25	Fish Audio Transcribe Stream 1	44	14.3	20.3	61.1	68.3	17	14.9	21.9	62.0	68.0
26	Cartesia Ink 2	30	16.5	22.6	62.3	71.0	40	11.6	18.9	59.9	66.7
27	xAI Grok Voice Transcribe 2.0 Streaming	24	17.3	24.1	62.3	67.5	26	13.6	19.3	60.5	65.5
28	Soniox STT Real-time v6	36	15.8	24.8	59.8	69.1	31	12.3	21.1	60.9	67.8
29	Soniox STT Async v6	38	15.0	22.6	62.9	69.8	42	11.4	19.7	59.7	67.3
30	Omi General-1	33	15.8	21.1	63.1	70.0	21	14.0	21.3	61.8	67.9
31	Qwen3-ASR 1.7B	37	15.0	24.1	63.4	71.2	28	13.4	20.8	61.1	67.6
32	Omi Medical-1	43	14.3	24.1	61.3	67.9	18	14.3	21.9	61.0	66.8
33	Omi Medical-1 Flash	39	15.0	24.1	61.1	67.9	25	13.8	21.1	61.0	66.9

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
34	Amazon Transcribe	26	16.5	26.3	63.8	71.8	35	11.8	19.7	60.8	67.6
35	Amazon Transcribe Medical	26	16.5	26.3	63.8	71.8	35	11.8	19.7	60.8	67.6
36	Meta Muse Voice Transcribe Realtime	52	12.8	18.0	60.2	65.8	37	11.8	19.3	60.5	66.7
37	Mistral Voxtral Mini Transcribe Realtime 3	48	13.5	18.8	60.7	66.7	51	9.9	17.3	58.5	65.1
38	Speechmatics Enhanced	28	16.5	24.1	63.1	67.3	33	12.3	16.5	60.2	63.9
39	OpenAI GPT Realtime Whisper	31	16.5	21.8	61.1	67.9	29	13.2	20.2	59.8	65.8
40	Omi Medical-1 Live	50	13.5	24.1	60.0	66.3	41	11.6	18.0	58.6	64.6
41	Mistral Voxtral Mini 4B Realtime	35	15.8	22.6	60.5	67.3	46	10.8	16.9	56.6	63.2
42	Speechmatics Enhanced Realtime	34	15.8	21.1	62.1	66.0	50	10.3	14.5	59.1	62.9
43	Deepgram Nova-3	41	15.0	24.1	60.5	67.9	47	10.7	16.2	58.1	63.4
44	Meta Muse Voice Transcribe 1.0	47	13.5	20.3	61.1	68.5	39	11.6	18.9	60.1	66.2
45	Reson8 Resonant-1 Realtime	22	18.0	27.1	61.9	69.4	44	11.0	18.2	57.6	64.3
46	Smallest.ai Pulse Pro	49	13.5	21.8	60.3	67.5	32	12.3	19.3	60.2	66.7
47	Inworld STT-1	32	16.5	26.3	58.2	66.2	45	11.0	15.6	57.1	62.9
48	NVIDIA Nemotron 3.5 Transcribe 30B-A4B	42	15.0	20.3	60.2	66.7	43	11.0	16.7	58.4	64.2
49	Musci-ASR qwen-1019	39	15.0	22.6	61.1	68.5	48	10.7	16.4	57.7	63.2
50	ARK-ASR 3B	51	13.5	20.3	59.0	66.0	49	10.7	17.6	57.3	64.0
51	Omi General-1 Live	45	14.3	21.1	60.5	67.3	34	12.1	16.0	58.4	64.0
52	Speechmatics Melia 1	61	11.3	18.0	58.4	65.4	58	9.2	14.5	55.9	61.7
53	Speechmatics Linden 1	53	12.8	16.5	59.8	62.9	63	8.5	11.6	55.3	58.8
54	Speechmatics Oak 1	62	11.3	18.0	57.4	64.6	60	9.0	14.7	55.9	61.7
55	Amazon Transcribe Streaming	46	14.3	19.5	60.3	64.8	52	9.6	13.1	56.3	61.0
56	Reson8 Rumble-2 Flash	58	12.0	17.3	59.2	66.7	53	9.6	14.3	55.9	62.1
57	Deepgram Nova-3 Streaming	66	10.5	15.0	57.1	62.3	55	9.4	15.4	55.8	61.7
58	Amazon Transcribe Medical Streaming	54	12.8	18.0	58.6	63.6	57	9.2	12.9	55.9	61.0
59	FluidVoice Fluid Speech Pro	59	12.0	19.5	58.0	66.5	59	9.0	16.0	56.7	63.7
60	NVIDIA Canary 1B v2	63	11.3	18.8	56.9	65.0	70	7.4	13.4	56.1	62.0
61	Modulate Velma-2 STT English	56	12.8	18.0	57.4	65.0	67	7.7	14.3	53.7	60.5
62	Qwen3-ASR 0.6B	55	12.8	17.3	57.6	63.4	61	9.0	14.0	55.3	60.6
63	Moondream Parakeet Ultra	56	12.8	21.8	57.4	66.7	54	9.4	13.8	56.2	62.6
64	Deepgram Flux English	70	9.8	15.0	58.2	64.4	66	8.1	13.6	55.3	60.1
65	Omi Medical Edge-1	63	11.3	14.3	56.9	65.6	56	9.2	16.7	56.0	63.2
66	IBM Granite Speech 5.0 2B	67	10.5	15.0	54.7	61.5	65	8.3	13.1	54.1	59.6
67	Fun-ASR MLT Nano 2512	68	10.5	13.5	53.8	57.6	68	7.5	10.3	51.3	56.3
68	Gradium STT	65	11.3	14.3	54.5	61.5	64	8.5	13.4	51.1	57.5
69	Moonshine Streaming Medium	74	8.3	10.5	53.4	57.4	69	7.5	10.8	49.7	54.9
70	NVIDIA Parakeet TDT 0.6B v3	71	9.8	17.3	54.7	63.4	62	8.8	14.3	53.2	59.4
71	Deepgram Flux Multilingual	75	7.5	12.0	54.2	60.0	75	6.1	9.6	51.3	56.0
72	Cohere Transcribe 03-2026	69	10.5	13.5	47.8	54.2	71	7.2	11.2	47.9	52.7
73	MOSS Transcribe Preview 2B	73	9.8	13.5	50.9	54.9	72	7.0	9.0	49.9	53.6
74	Fun-ASR Nano 2512	72	9.8	12.8	52.2	56.9	78	4.4	8.1	48.0	53.5
75	Rime STT	60	12.0	14.3	51.3	57.1	73	6.4	10.3	50.4	55.2
76	Meta omniASR LLM 7B	77	6.0	7.5	47.8	54.7	74	6.4	9.9	48.3	54.3
77	Moonshine Base	81	4.5	6.8	47.8	51.8	77	4.8	7.4	46.9	51.5
78	Murf Finch-1	78	5.3	7.5	48.9	53.6	80	3.9	5.9	44.9	49.6
79	NVIDIA Nemotron 3.5 ASR Streaming 0.6B	76	6.8	7.5	44.5	48.4	76	5.1	7.2	42.3	46.3
80	IBM Granite Speech 5.0 470M TurboCTC	79	5.3	5.3	47.2	51.8	79	4.2	6.8	46.6	50.2
81	NVIDIA FastConformer Hybrid Large	85	2.3	3.8	39.5	44.7	81	3.7	5.0	39.0	43.5
82	SenseVoice Small	80	5.3	6.8	40.0	42.6	86	1.5	2.9	37.0	39.4
83	Moonshine Streaming Tiny	86	2.3	2.3	28.6	30.4	85	1.7	2.0	27.1	29.3
84	FUTO ASR4All L	83	3.0	3.0	36.8	40.4	84	1.7	2.2	37.3	40.8
85	Fermion Phonon-2	88	0.0	1.5	39.5	46.4	82	2.9	4.4	39.2	44.8

#	SYSTEM	PUB. #	PUB. TSR	PUB. TSR REL.	PUB. EA	PUB. EA REL.	PRIV. #	PRIV. TSR	PRIV. TSR REL.	PRIV. EA	PRIV. EA REL.
86	Vosk en-us 0.42 GigaSpeech	87	1.5	1.5	28.4	30.2	88	1.1	1.5	24.3	25.7
87	Meta omniASR CTC 7B	82	3.8	5.3	40.8	46.8	83	2.4	4.2	38.5	43.4
88	FUTO ASR4All M	84	3.0	3.0	33.8	36.8	87	1.5	2.4	32.2	35.1
89	SpeechBrain Conformer + TransformerLM	89	0.0	0.0	19.1	21.1	89	0.6	0.7	16.7	18.0
90	Meta MMS 1B	90	0.0	0.0	13.2	17.0	90	0.0	0.2	12.0	15.6

BEST WORST Cells are coloured by rank within each column (solid green is best; below the top 15, white). API batch
API streaming open weights