

Reproduction of SYSTRAN/faster-whisper: large-v3 WER, speed and memory benchmarks

K-Veritas Team

Submitted by K-Veritas Team

kv:2610.00042v1 · Published 2026-10-18

Independent reproduction · Benchmark or leaderboard entry · Natural language processing

Abstract

Ran the three scripts in the repository's benchmark/ folder at the pinned commit, unchanged, with the Whisper large-v3 model on CUDA: wer_benchmark.py over all 2703 utterances of LibriSpeech clean validation (streamed from the Hugging Face hub, normalized with the bundled English normalizer), speed_benchmark.py on the bundled benchmark.m4a (3 repeats of 10 transcriptions, minimum time reported) and memory_benchmark.py with --gpu_memory (peak GPU memory and power while transcribing benchmark.m4a). The tests/ and docker/ folders are withheld as hash-only entries. CTranslate2 4.8 with cuBLAS and cuDNN 9 from PyPI, datasets 3.6. One NVIDIA RTX 5060 Ti.

About this record

Title, authors and abstract were provided by the authors and reviewed for completeness, not for scientific soundness. Every number in this record is copied from the signed data of the sealed reports [2], which remain authoritative.

This cover page is not signed. Check it against the SHA-256 listed at <https://kveritas.org/records/2610.00042v1>, and verify each report with its code bundle: kveritas verify <report.pdf> --bundle <bundle.zip>.

1 Sealed reports

The experiments were sealed with K-Veritas [1], which binds each result to the code, hardware and time that produced it.

Report	Label	Sealed	Runs	Invocations	Signer
r1 [2]	faster-whisper large-v3: WER,...	2026-10-11	1	1	Signed by the K-Veritas server

2 Results

r1: faster-whisper large-v3: WER, speed, memory

Run	Metric	Final value	Step
1	librispeech_clean_val_wer	2.92	

Run	Inline claim	Value	Phase
1	librispeech_clean_val_wer	2.92	
1	benchmark_m4a_min_transcribe_s	46.408	
1	benchmark_m4a_max_gpu_mem_mib	5348	

3 Setup

Report	Item	Value
r1	Operating system	linux / amd64
r1	CPU	Intel(R) Xeon(R) W-2145 CPU @ 3.70GHz, 16 cores
r1	GPU	NVIDIA GeForce RTX 5060 Ti, 16311 MiB
r1	Memory	31.0 GB
r1	Total run time	2721.3 s

4 Files

Each report and code bundle must hash to these values. A bundle belongs to the report whose signed data names its hash.

File	SHA-256	Size
r1.pdf	98fe97cab8d9840046127c9096e99023c4dac5950a223a30e41e0e4cc5491476	6.6 MB
r1.kvbundle.zip	b14744aef44990309f121f667f1fd986232b576b91ea12a01a0b561dc2a7e5a6	54 KB

5 Links and licence

Reproduction of: <https://github.com/SYSTRAN/faster-whisper/tree/a52b026ac844c02117865e61968547639b60a6f8>

Record: CC-BY 4.0. Code: MIT, as declared by the authors. Tags: speech-recognition, whisper, ctranslate2, benchmark.

Listed, not endorsed. Review checks the artifact, never the science. Reports are never edited; changing them makes a new version. Other fields are corrected in place, and every correction is logged in metadata.yaml.

References

[1] M. K. Keita and C. Homan. Computer Science Conferences Should Require Nonrepudiable Experimental Results. NeurIPS 2026 Position Paper Track. arXiv:2605.08586.

[2] Sealed experiment report r1 (faster-whisper large-v3: WER, speed, memory), kv:2610.00042v1/r1. Data hash f2b4af0bfbf1b20524d96f22a17b1ef7f63b8d1026cf8dbdf80420c0dbd52f21.

<https://kveritas.org/records/2610.00042v1/r1.pdf>