HAUSABENCH · JULY 2026

The State of Hausa AI Quality

Four leading AI models. Ten practical Hausa tasks. Every tested model produced at least one error serious enough to block publication.

4	10	100%
MODELS	PUBLIC PROMPTS	HAD A BLOCKER

> Fluent Hausa is not the same as safe, accurate, or publication-ready Hausa.

Abba Bello Kanwa · Native Hausa quality specialist

WHAT WE TESTED

A practical human-QA benchmark

On 17 July 2026, Hausa AI Studio tested GPT-5.2, Gemini 2.5 Flash, Claude Sonnet 5, and DeepSeek V3.2 through their APIs on default settings. Each model received the same ten public prompts.

Tasks

Translation in both directions, Hausa instructions, Nigerian localization judgment, code-mixed Hausa-English, a medical safety scenario, and long-form Hausa writing.

Scoring

Meaning accuracy, fluency, tone, cultural fit, hallucination risk, and safety. Defects were labeled and severity-rated.

Model	Overall	Verdict
GPT-5.2	4.83	**Good**
Claude Sonnet 5	4.57	**Good**
DeepSeek V3.2	4.48	**Mixed**
Gemini 2.5 Flash	4.37	**Mixed**

Important: averages hide release-blocking defects. This is practical expert guidance for real-world Hausa use, not an academic leaderboard.

FIVE FINDINGS THAT MATTER

The errors looked fluent

1. Everyday Hausa was misunderstood.

Two models failed the common phrase “Dan Allah”, translating it as an invocation or inventing a person named Dan.

2. Fluent outputs contained nonexistent or wrong words.

Examples included “akulla” for updated and “fatalwa” (ghost) where “fartanya” (hoe) was intended.

3. A model invented a date.

A community announcement gained an unrequested and stale date that users could act on.

4. Quality weakened as output became longer.

Hooked letters dropped, grammar loosened, and explicit length or language constraints were ignored.

5. Safety behavior improved, but terminology still matters.

All models directed a fever case toward professional care, while one rendered antibiotics in a way that can read as vaccines.

WHAT BUYERS SHOULD DO

Use models to draft. Use expert Hausa QA to release.

AI products

Review every customer-facing Hausa flow for meaning, safety, terminology, register, and constraint compliance.

Localization

Do not use fluency as the acceptance signal. Verify idiom, lexical reality, and audience fit.

Data and evaluation

Catch format drift, invented details, and English leakage before they enter training or evaluation sets.

Speech and transcripts

Validate terminology, named entities, segmentation, and meaning before scaling.

Start with a focused Hausa QA pilot.

Severity-rated findings, recommended corrections, and a clear readiness decision.

hausaai.studio/intake

Scope: 4 models, 10 public prompts, one API run on 17 July 2026. No client content was used. HausaBench is small by design and should be read as directional operational evidence.