

AI benchmarks: read real scores and compare models responsibly

LearnIA - teaching template - 2026-10-14

Use fictional data for the first trial. Adapt this template to your context and verify the result before using it.

Brief to complete

Objective and intended decision :

Audience and context :

Documents, versions and sources :

Constraints and boundaries :

Expected output format :

Information to confirm :

Reviewer :

Starter instruction

Prepare a worksheet comparing three models on a document task. Keep third-party published scores separate from our future trials. Include source, result date, exact version, settings, tools, test inputs, reference answers, ambiguity, critical errors, raw outputs, cost and total duration. Invent no measurements. Finish with conditions that would invalidate the conclusion.

Quality review

☐ Every score has a source and date.

☐ Different types of metrics stay separate.

☐ Failures and retries are retained.

☐ The conclusion states scope and limitations.

Verification log

Item	Source or calculation	Result	Correction

Approval and next step

Remaining errors :

Assumptions to resolve :

Retained version :

Next action and owner :

Full tutorial, example and sources : <https://learn-prompting.fr/blog/benchmarks-ia-comprendre-comparer>

This resource prepares a draft; it is not a certification, a grading decision, legal advice or a guarantee of results.