Benchmarks

In order to measure and compare LLMs more holistically, you can make use of pre-existing datasets, and associated benchmarks that have been established by LLM researchers specifically for this purpose.

A screenshot of a computer

Description automatically generated

GLUE

A screenshot of a computer

Description automatically generated

SuperGLUE

A close up of a document

Description automatically generated

Benchmark for Massive Models

A diagram of different types of models

Description automatically generated

HELM ( Holistic Evaluation of Laguage Models)

A diagram of a model

Description automatically generated with medium confidence