The GLUE Benchmarks

Lesson Downloads

Show TranscriptSummarize Video

The GLUE benchmarks serve as an essential tool to assess an AI's grasp of human language, covering diverse tasks, from grammar checking to complex sentence relationship analysis. By putting AI models through these varied linguistic challenges, we can gauge their readiness for real-world tasks and uncover any potential weaknesses.
Technical Terms Explained:

Semantic Equivalence: When different phrases or sentences convey the same meaning or idea.
Textual Entailment: The relationship between text fragments where one fragment follows logically from the other.

GLUE Tasks / Benchmarks

Short Name	**Full Name**	**Description**
CoLA	Corpus of Linguistic Acceptability	Measures the ability to determine if an English sentence is linguistically acceptable.
SST-2	Stanford Sentiment Treebank	Consists of sentences from movie reviews and human annotations about their sentiment.
MRPC	Microsoft Research Paraphrase Corpus	Focuses on identifying whether two sentences are paraphrases of each other.
STS-B	Semantic Textual Similarity Benchmark	Involves determining how similar two sentences are in terms of semantic content.
QQP	Quora Question Pairs	Aims to identify whether two questions asked on Quora are semantically equivalent.
MNLI	Multi-Genre Natural Language Inference	Consists of sentence pairs labeled for textual entailment across multiple genres of text.
QNLI	Question Natural Language Inference	Involves determining whether the content of a paragraph contains the answer to a question.
RTE	Recognizing Textual Entailment	Requires understanding whether one sentence entails another.
WNLI	Winograd Natural Language Inference	Tests a system's reading comprehension by having it determine the correct referent of a pronoun in a sentence, where understanding depends on contextual information provided by specific words or phrases.