Knowing Is Not Doing

A Benchmark for Therapeutically Relevant Errors of Large Language Models

Coming soon.