Alternative phrasings

-

Related

-

Scratchpad

Deleted benchmark pages (these are both safety and capabilities, I haven't classified them: See What benchmarks exist for measuring the intelligence of AI systems?

- MACHIAVELLI
- MINST
- abstraction reasoning challenge (ARC)
- Switchboard
- SuperGLUE
- GLUE
- SQuAD
- ImageNet
- ConceptARC