

- [APPS benchmark](#) - coding
-
-

Alternative phrasings

-

Related

- [What is intelligence?](#)
 - [How does the intelligence of current AI compare to that of animals?](#)
 - [What benchmarks exist for evaluating the safety of AI systems?](#)
-

Scratchpad

Deleted benchmark pages (these are both safety and capabilities, I haven't classified them: See

[What benchmarks exist for evaluating the safety of AI systems?](#)

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