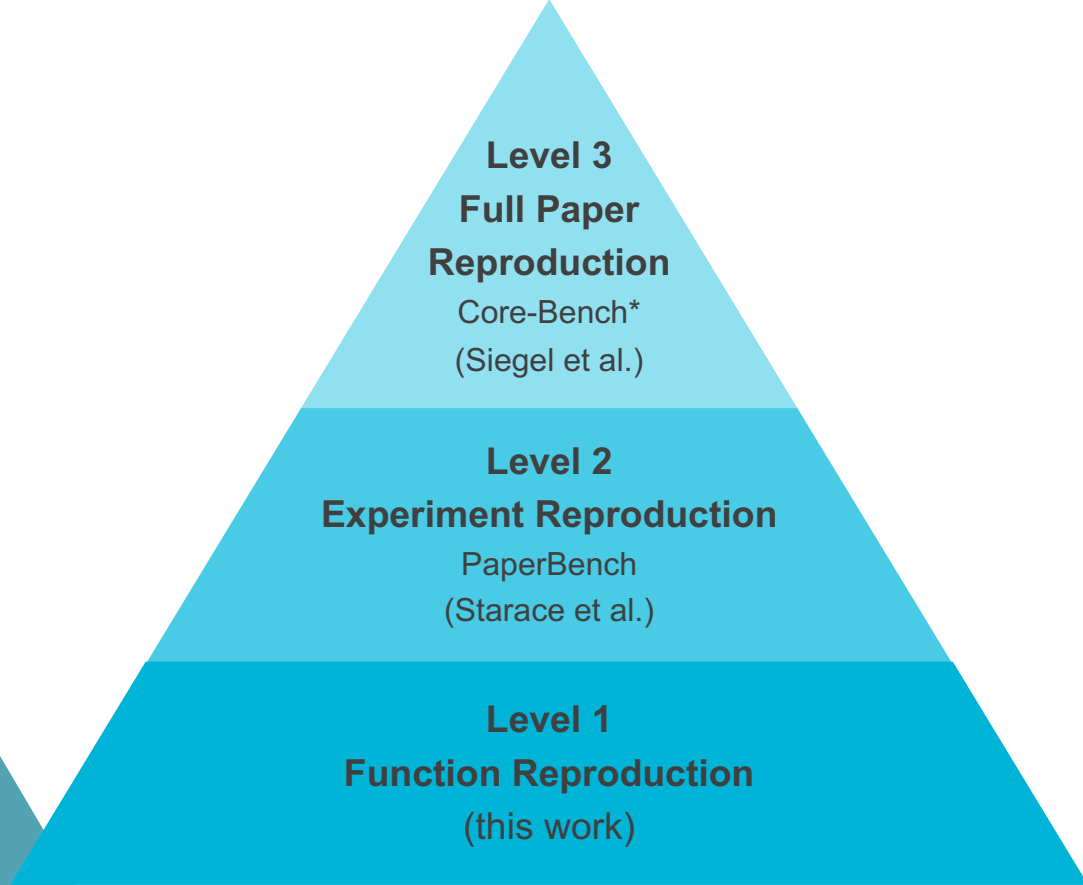**Capability Levels for AI Research Agents**
MLGym (Nathani et al.)

Benchmarks by Reproduction Level

* Core-Bench provides original code for reproduction.
All other benchmarks are based solely on papers.