Legend:
- Agent Benchmarks
- In-Context Learning & Prompting
- Agents & Tools

2024

2023

2022

ToolFormer · API-Bank · ToolLLM · DSPy

ART · ToT · ToolBench · WebArena

Reflexion · Gorilla · Lang Graph · FireAct

TaskMatrix · API Bench · Agent Bench · Agent Tuning

HuggingGPT · Mind2Web · MINT · GAIA

Self-Refine · ToolQA · Meta Tool · LATM

Web Shop · SC

CoT · Constitutional AI

Talm

WebGPT · ReAct

ToolkenGPT · Work Arena

CodeAct · BFCL

ToolACE · xLAM

Agent Board · Hammer

SWE Bench · Tape Agents

τ-bench