















relevant for applications and
limited measuring specificity


- ChemBench  
- LabSafety Bench   
- MaScQA
- MolPuzzle 

relevant for applications and
specific measurement

- SciKnowEval 
- LAB-Bench   
- SciAssess 
- MaCBench   

- CAMEL – Chemistry

limited applicability and
measuring specificity

- LlaSMol 
- ChemLLM
- ChemLLMBench

limited applicability and
specific measurement



end-to-end runnable



maintanance



multi modality