

acceptance Rate

1.0  
0.8  
0.6  
0.4  
0.2  
0.0

Invalid

Valid

OLMo-7B-Inst

- Unbelievable
- Believable
- Control

