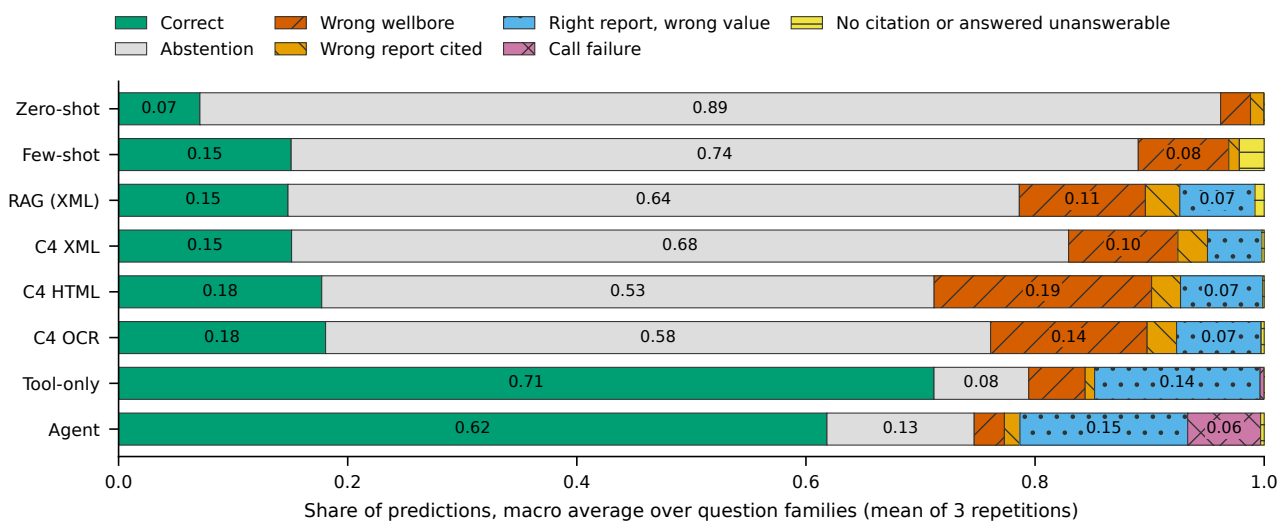

Correct
Abstention
Wrong wellbore
Wrong report cited
Right report, wrong value
Call failure
No citation or answered unanswerable
Zero-shot
Few-shot
RAG (XML)
C4 XML
C4 HTML
C4 OCR
Tool-only
Agent
0.07
0.89
0.15
0.74
0.08
0.15
0.64
0.11
0.07
0.15
0.68
0.10
0.18
0.53
0.19
0.07
0.18
0.58
0.14
0.07
0.71
0.08
0.14
0.62
0.13
0.15
0.06
0.0
0.2
0.4
0.6
0.8
1.0
Share of predictions, macro average over question families (mean of 3 repetitions)