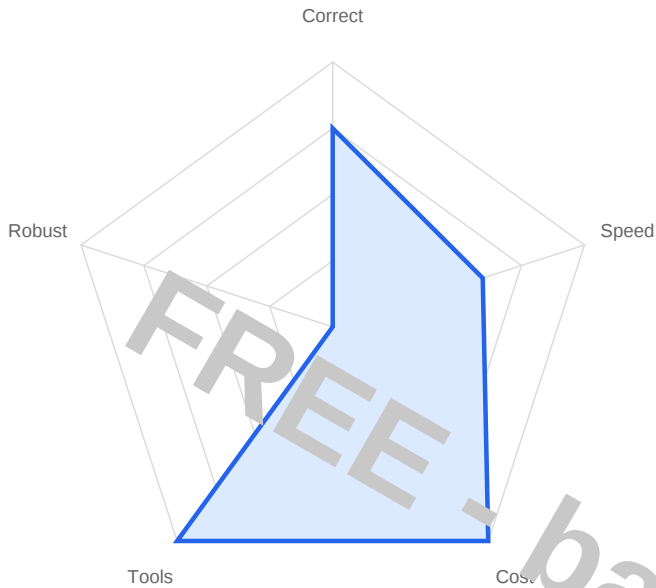




Strengths & weak spots

BEST TASK

Calculator Tool Use

100% success

WEAKEST TASK

Unstructured Text Extraction

0% success