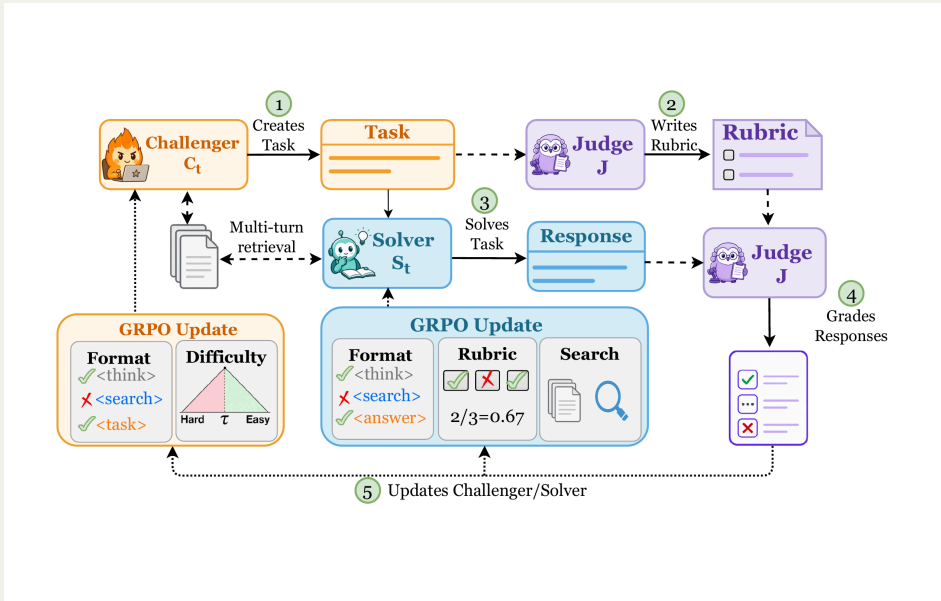


Learn / control



SCOPE

NeurIPS

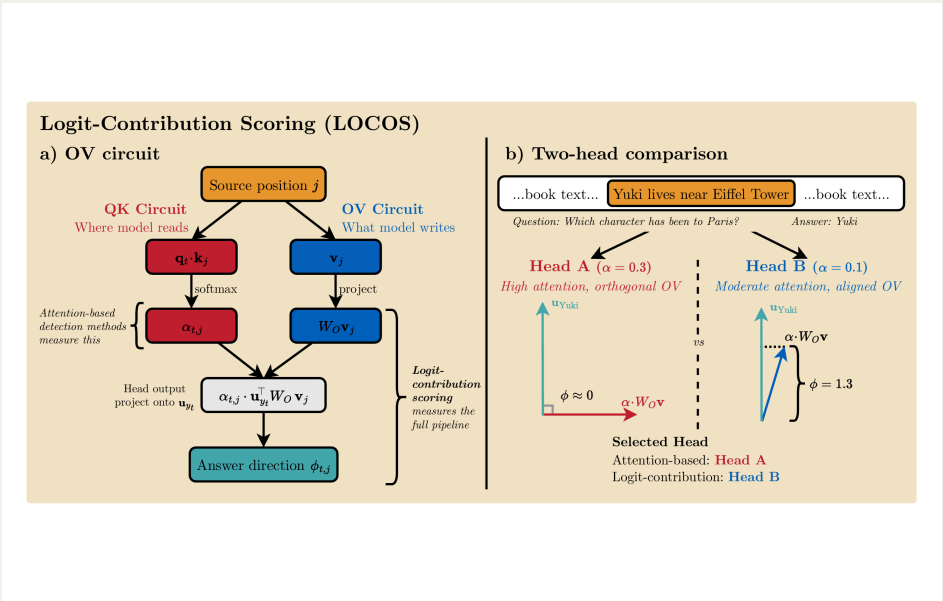
Composing steering vectors

EMNLP

Graph-to-table interfaces

LoG

Inspect



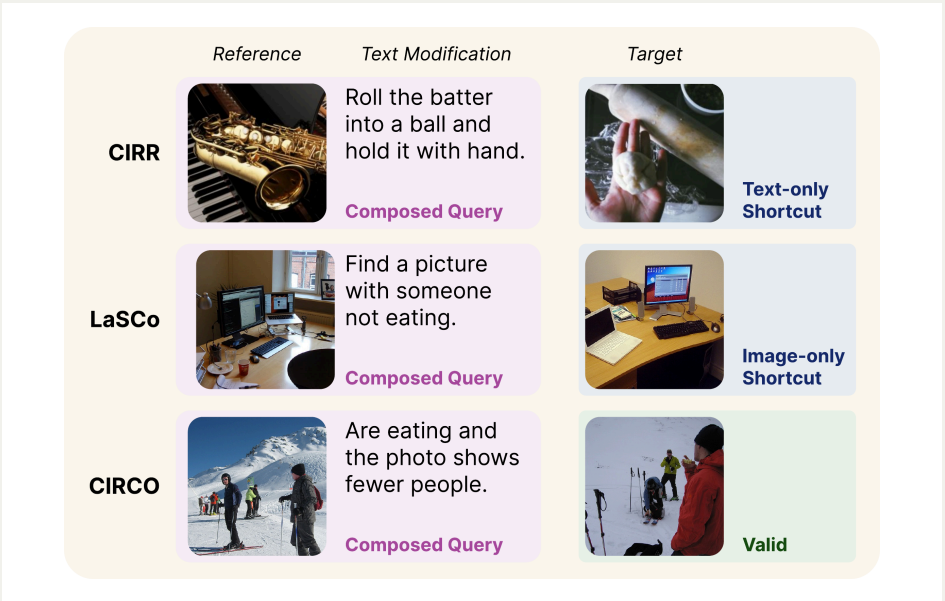
LOCOS

NeurIPS

Multimodal retrieval heads

EMNLP

Evaluate



Composed image retrieval

NeurIPS

AI psychophysics

NeurIPS

Hidden VLM instability

EMNLP

PREPRINTS

SPSD / FACE-Eval / VLM frequency filtering