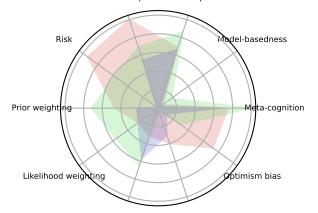
llama-2-70-chat human Random explo Directed exploration

random



Temporal discounting Learning rate