Dot plot across models / quantisations, per Task

Accuracy

1.0
0.8
0.6
0.4
0.2
0.0

llama-2-chat
llama-3-instruct
openhermes-2.5
code-llama-instruct
mistral-instruct-v0.2
llama-3.1-instruct
gpt-4
claude-3-opus-20240229
claude-3-5-sonnet-20240620
gpt-3.5-turbo
gpt-4-turbo-2024-04-09
gpt-4o-mini-2024-07-18
chatglm3
mixtral-instruct-v0.1

multimodal_answer
query_generation
explicit_relevance_of_single_fragments
implicit_relevance_of_multiple_fragments
api_calling
property_exists
medical_exam
entity_selection
naive_query_generation_using_schema
sourcedata_info_extraction
end_to_end_query_generation
relationship_selection
property_selection

Task