

1 **Supplementary Data 1:** Responses for all multimodal models

2 The file "model_responses.tsv" is a tab-delimited file containing the following columns:

3 **id:** Unique identifier for each entry representing the NEJM image challenge, that can be
4 accessed through "<https://www.nejm.org/image-challenge?ci=id>", **model_name:** name of the
5 multimodal model used, **model_response:** the response generated by the model,
6 **correct_answer:** The correct answer for the given query (A – E), **response_letter:** The letter
7 associated with the model's response (A – E), the complete response was provided wherever
8 the model was not responding with an answer, **answer_type:** The type or category of the
9 answer (either ANSWER or NOANSWER), **correct:** indicates whether the model's response
10 is correct or not (boolean: true or false). The models involved are 'internvl-1.5-8bit', 'gemini-
11 pro-vision', 'cogvlm-chat-hf-chat-template', 'claude-3-sonnet-20240229', 'llava-1.5-13b-hf',
12 'claude-3-haiku-20240307', 'llava-34b-v1.6-fp16', 'claude-3-opus-20240229', and 'gpt-4-1106-
13 vision-preview'.