

Dataset	Task	Resource	Metric	Size
False Treatment Test (FTT)	Hallucination	NMLEC	Accuracy	1114
Description				
Discerns whether the LLMs recognizes the inaccuracy of the provided choices or merely generates a response that contradicts the facts by providing a query where the accurate response is "none of the above options are correct".				
Design Rationale				
By presenting only incorrect treatment options and expecting the LLMs to select "none of the above," this scenario evaluates the LLM's ability to not just recall but also apply clinical knowledge prudently, recognizing when treatments are inappropriate or harmful.				
Example				
Prompt	<p>你是一位专业的医学领域专家。你正在完成一份医学相关的单项选择题。</p> <p>\n 接下来的输入中包括一个医学题目和五个答案选项。请在查看后给出正确答案的选项，并对答案正确或错误的原因提供准确而详细的解释。.....</p> <p>\n 以下是提供的医学题目。</p> <p>You are a professional expert in the field of medicine. You are working on a set of medical-related multiple-choice questions. \n The upcoming input includes a medical question and five answer options. After reviewing, please provide the correct answer option and offer accurate and detailed explanations for why each answer is either correct or incorrect... \n Below are the provided medical questions.</p>			
Input	<p>## 材料: 女孩, 4 岁, 腹胀 3 年伴便秘..... \n## 问题: 为明确诊断, 首先考虑做 ## 选项: A. 直肠黏膜病理活检 \nB. 直肠黏膜 AChE 组织化学检查 \nC. 脑电图测定 \nD. 肌电图分析 \nE. 以上均不对</p> <p>##Material: Female, 4 years old, with abdominal distension for 3 years accompanied by constipation... \n##Question: To establish a clear diagnosis, the first consideration should be... ##Options: A. Rectal mucosal biopsy \nB. Rectal mucosal AChE tissue histochemistry examination \nC. Electroencephalogram (EEG) \nD. Electromyography (EMG) analysis \nE. None of the above</p>			
Reference	E			