

**LIMITATIONS TESTS FOR AN LLM
EXAMPLES**

Test Type	Sample Test
<p>Antonyms</p> <p>(Goal: Evaluate the ability of the model to identify and act on semantic relationships-crucial for interpreting information. With antonyms, we are evaluating the ability of the LLM to identify the opposite concepts in an antonymy semantic relationships and also understand how the model does that (note: some of the components in the test are meant to uncover that))</p>	<p>Add the correct missing antonyms, and illustrate each of them in a sentence.</p> <p>Pick up / / Go off / Turn off</p> <p>Bring in / Set up / Show up / Hold on / 2010 / / 3006</p> <p>Break down / Fill up / Look for / Look out / Look in / Break out / Take out</p> <p>0978 / / 0872x5 / Run into / Speed up</p> <p>Figure out / ??#%#% / PoetPoetPoet / Look up / Break up / Hang up / Cut off / Start up / / >>_99((&%\$#@@</p> <p>Manger/ /aimer / Sit down / Take away / Dress down / Come out</p>

<p>Inference and logical deduction</p> <p>(Goal: Evaluate the model's ability to infer implied information, make logical deductions, or draw conclusions based on the provided text).</p>	<p>Since 2010 the hospital at Minouche operates under constrained resources and has low capacity of beds for patients. Doctor Nora who joined the hospital in 2018, and who is paid 300k a year, works with their team to manage the free care for all their patients until 2027. This free care will be suspended this month and for two years, and then again will be resumed for free until January 2027. Their team includes nurses, lab staff, administrators and healthcare consultants. The doctor is the only one who makes the final decision, based on input from the team. This morning the team voiced their frustration when seeing a "little old lady or man" waiting in the emergency room for three or four days for a bed but not able to get one because someone who has "gone using drugs on the street for hours could just roll back in." All the team believes that this growing frustration will have ended in 2023 were more funds approved by Victoria.</p> <p>Answer these questions and provide a justification.</p> <ol style="list-style-type: none"> 1. For how many years has Nora been at Minouche? 2. Who is not part of the healthcare team at Minouche? 3. The patients care at Minouche will be kept free for how long? 4. Who is not included in the final decision making at Minouche? 5. When will the frustration end?
<p>Spatial reasoning</p> <p>(Goal: Evaluate the model's ability to reason about spatial relationships and objects in a mental space)</p>	<p>In a room with seven chairs placed around a large circular table, Sarah sits opposite to Alex. Alex is to the left of David, who is sitting to the right of Mary. To the left side of Sarah, Anna is sitting.</p> <ol style="list-style-type: none"> 1. Where is Sarah seated in relation to Mary? 2. Who is sitting to the left of Alex? 3. Who is sitting between Sarah and Alex? <p>On a shelf, there are five books stacked over each other. The red book is on top of the blue book, and the green book is between the red and blue books. The red book is below the yellow book. The yellow book is above the black book. The red book is above the black book.</p> <ol style="list-style-type: none"> 1. How many books are between the blue book and the black book? 2. How many books are between the yellow book and the blue book? 3. Which book is in the top most position?