

Exercise 1-2: Limitation Hunt

Objective

Discover the real-world limitations of large language models by finding questions they get wrong.

Prerequisites

- **Exercise 1-1** completed (you should have a working Chat Trigger + AI Agent + OpenAI Chat Model workflow)

Setup

Use the same workflow from Exercise 1-1 (no changes needed).

Demo Prompts

Try each of these prompts and observe where the LLM struggles. Run them **multiple times** to see different results.

Knowledge Cutoff

- “How many cities was Changi connected to as at 31 March 2025?”

Math Errors

- “What is $7,847 \times 6,329$?”
- “Check and Pay saves employers 20,000 hours annually. If efficiency gains grow at 12% per year compounded, how many hours will it save in year 5?”
- “Calculate: 17% of \$7,400, then multiply by 12 months, and add 4% interest compounded annually for 3 years”

Character/Token Processing

- “Remove all the ‘t’s from ‘refrigerator’”

Your Task

Find your **own** Changi-related question that the LLM gets wrong.

Valid Limitation Criteria

Your question must meet ALL of these:

1. The question has a **factually correct answer** (verifiable from Changi documentation)
2. The LLM gives a **different or wrong answer**
3. The error is due to an **LLM constraint**: knowledge cutoff, math errors, or reasoning failures

Examples of Valid Limitations

- “What is Changi’s total Scope 3 emissions?”
- “Who is the current Chairman of CAG?”
- “What does CAG do in one sentence?”

What to Submit

1. The question you asked
2. The LLM’s response
3. The correct answer (with source)
4. Which limitation category it falls under (knowledge cutoff, math, reasoning, etc.)