**1. Funkčné testy**

Cieľ: overiť, či model dokáže robiť to, čo deklaruje výrobca.

**Možné scenáre:**

* **Porozumenie kontextu**
  + Zadáme dlhý rozhovor a budeme sledovať, či drží kontext aj po desiatkach výmen.
* **Presnosť faktov**
  + Dá sa mu úloha s historickými aj aktuálnymi údajmi, a overíme vo viacerých dôveryhodných zdrojoch.
* **Viacjazyčnosť**
  + Mix jazykov v jednej otázke (slovenčina + nemčina + angličtina), či správne prepína a chápe.
* **Kritické myslenie**
  + Zadáme úlohu s chytákmi, polopravdami a dezinformáciami, a sledujeme, či ich odhalí.
* **Formátovanie výstupu**
  + Požiadame presne definovaný formát (napr. .md s tabuľkou) a skontrolujeme, či ho dodrží.

**2. Nefunkčné testy**

Cieľ: zmerať, ako dobre sa správa mimo bežných prípadov použitia.

**Možné scenáre:**

* **Rýchlosť odpovedí**
  + Stopnúť čas pri jednoduchých vs. komplexných otázkach.
* **Stabilita**
  + Spustiť viacero ťažkých úloh naraz a sledovať, či nedôjde k “halucináciám” alebo strate formátu.
* **Konzistentnosť**
  + Položiť rovnakú otázku viackrát s časovým odstupom a porovnať odpovede.
* **Práca s veľkým vstupom**
  + Zadať dlhý text (napr. 20 normostrán) a sledovať, či ho vie analyzovať alebo zhrnúť.

**3. Stresové testy**

Cieľ: odhaliť hranice a slabiny.

**Možné scenáre:**

* **Neúplné inštrukcie** – ako si poradí, keď niečo zámerne nedopovieme.
* **Zámerne konfliktné požiadavky** – napr. “napíš text bez písmena ‘a’, ale v ňom musí byť názov Bratislava”.
* **Extrémne špecifické znalosti** – úzka odborná téma s veľkou šancou, že bude musieť priznať nevedomosť.

**4. Hodnotenie**

Navrhol by som si pripraviť **testovaciu tabuľku** so stĺpcami:

* Názov testu
* Typ testu (funkčný/nefunkčný/stres)
* Vstup
* Očakávaný výstup
* Reálny výstup
* Hodnotenie (Pass/Fail + poznámky)

Navrhol by som tieto prvé testy, aby sme pokryli rôzne typy – funkčné, nefunkčné aj stresové – a rýchlo zistili, kde má GPT 5.0 svoje silné a slabé stránky:

| **Číslo testu** | **Názov testu** | **Typ testu** | **Vstup** | **Očakávaný výstup** | **Poznámky** |
| --- | --- | --- | --- | --- | --- |
| 1 | Udržanie dlhého kontextu | Funkčný | Poslať sériu 15 navzájom nadväzujúcich otázok v rámci jednej témy, posledná odkazuje na informáciu z prvej. | Správna odpoveď na poslednú otázku obsahuje detail z prvej správy. | Overenie pamäte v rámci jednej relácie. |
| 2 | Overenie faktov z viacerých zdrojov | Funkčný | Požiadať o aktuálny údaj (napr. počet obyvateľov konkrétneho mesta k roku 2025). | Hodnota v súlade aspoň s 2 nezávislými dôveryhodnými zdrojmi. | Test presnosti a schopnosti uviesť zdroje. |
| 3 | Viacjazyčný vstup | Funkčný | Zadanie obsahujúce slovenčinu, nemčinu a angličtinu v jednej otázke. | Správne pochopenie všetkých častí a odpoveď v zodpovedajúcom jazyku podľa požiadavky. | Test prepínania jazykov a porozumenia. |
| 4 | Konzistentnosť odpovedí | Nefunkčný | Položiť rovnakú otázku 3× v časovom odstupe. | Rovnaký alebo minimálne veľmi podobný význam odpovedí. | Sleduje stabilitu generovania. |
| 5 | Stresový test s konfliktnými požiadavkami | Stresový | „Napíš vetu bez písmena ‘a’, ale musí obsahovať názov Bratislava.“ | Logická odpoveď, ktorá splní obidve podmienky, alebo vysvetlenie nemožnosti. | Test logiky pri protichodných podmienkach. |