Porównanie czatu LLM i systemu z własną wiedzą

Od trzech niezależnych prób do uzasadnionego porównania

Czas	Pracujesz z	Rezultat
około 30 min	notebookiem i lokalnym modelem lub przykładami demo	porownanie_AB.json i opis_projektu.md

Porównasz odpowiedzi na to samo pytanie: bez dokumentu, z pełną procedurą i z nietrafnym fragmentem. Ocenisz zachowanie modelu oraz wskazanie źródła, zachowując niezmienione wyniki prób.

Cel i rezultat rozdziału

Porównasz warianty A, B i C przy tym samym modelu i pytaniu oraz uzasadnisz ocenę.

Po zakończeniu masz

- trzy surowe odpowiedzi oraz oceny w dwóch skalach 0–2;
- wniosek wskazujący wynik i ograniczenia eksperymentu;
- porownanie_AB.json, opis_projektu.md i uzupełniony notebook R6.

Zanim zaczniesz

Rozpakuj cały pakiet i pracuj na własnej kopii katalogu Student. Zachowaj razem foldery notebooki, dane, kod i wyniki. Ścieżki w tej karcie są liczone od Student; pliki danych pozostają bez zmian.

Otwórz notebook

```
notebooki/M12_M1_R6_eksperyment_ab_start.ipynb
```

Materiał wejściowy	Do czego służy
dane/konteksty.json	Pełna sekcja 5.2 oraz nietrafna sekcja 6.1.
dane/instrukcja.txt	Źródło do weryfikacji procedury.
konfiguracja.json	Wspólny model i parametry całego zestawu.

Wybierz demo albo ollama. Demo nie wywołuje modelu; ollama wymaga wcześniej przygotowanego stanowiska. Sposób przygotowania opisuje M12_M1_instrukcja_pracy.docx. Prezentacja M12-M1-v2: slajdy 85–100 przed podziałem na rozdziały.

Praca krok po kroku

Krok 1 Ustal warunki porównania — Uruchom Przygotowanie w wybranym trybie.

- Dla własnego eksperymentu ustaw MODE na ollama. W demo analizujesz gotowe przykłady.
- Zachowaj model, parametry i pytanie w całym zestawie. Zanotuj katalog OUT.

Krok 2 Wykonaj A B i C — Uruchom komórkę Wykonaj porównanie A B C.

- A nie ma dokumentu. B otrzymuje pełną sekcję 5.2. C otrzymuje sekcję 6.1 o kopii przed aktualizacją.
- Każda próba zaczyna nową rozmowę. Zachowaj surowa_A.json, surowa_B.json i surowa_C.json.

Krok 3 Porównaj treść ze źródłem — Sprawdź B względem sekcji 5.2, a C względem sekcji 6.1.

- W B szukaj restartu, polecenia is-active, wyniku active i 50 wpisów dziennika przy wyniku innym niż active.
- W C pożądane jest wskazanie braku danych do pytania. Nie nagradzaj dopisania procedury spoza przekazanego fragmentu.

Krok 4 Nadaj oceny z uzasadnieniem — Uzupełnij pola zgodnosc_0_2 oraz zrodla_0_2 zgodnie z rubryką notebooka.

- Zgodność: 0 oznacza sprzeczność lub dopowiedzenia; 1 częściowe wykonanie; 2 pełne B albo uczciwe zgłoszenie braków w C.
- Źródła: 0 brak lub pozorny przypis; 1 niepełne wskazanie; 2 właściwy plik, wersja, sekcja i wsparcie treści. Opisz dodatki oraz uzasadnienie.

Krok 5 Sformułuj wniosek i zapisz wyniki — Uzupełnij wniosek i uruchom zapis obu plików wynikowych.

- Dla A oceniasz odpowiedź względem instrukcji dostępnej Tobie. Nie twierdź, że instrukcję otrzymał model.
- A/B zmienia kontekst i instrukcję; B/C zachowuje instrukcję, zmieniając trafność kontekstu. Zaznacz to ograniczenie.

Krok 6 Zachowaj cały zestaw — Przeczytaj porownanie_AB.json i opis_projektu.md, następnie zapisz notebook.

- Błędny wynik modelu nie oznacza źle wykonanej pracy, jeśli wykrywasz go i uzasadniasz ocenę.
- Nie wybieraj tylko najlepszej odpowiedzi. Kolejne próby zachowuj osobno; jednego zestawu nie uogólniaj na wszystkie modele.

Ważne: ponowne uruchomienie komórki Przygotowanie tworzy nowy katalog. Jeśli ponowisz generowanie, otrzymane odpowiedzi oceń ponownie. Po samej korekcie komentarzy uruchom komórkę oceny i zapisu, bez nowego wywołania modelu.

Punkty kontrolne pracy

Po etapie	Masz sprawdzić	Gdy wynik nie pasuje
Przed próbami	To samo pytanie, model i ustawienia.	Sprawdź konfigurację i pełne wiadomości.
Po ocenie	Są dwie oceny 0–2 i uzasadnienia dla A–C.	Zastąp None oraz UZUPEŁNIJ własną oceną.
Po zapisie	Wyniki zawierają trzy próby i wniosek.	Otwórz oba pliki z tego samego katalogu OUT.

Nazewnictwo i zapis wyników

Notebook zapisuje pliki w wyniki/R6/<data_id>/. Zapis <data_id> oznacza katalog nadany automatycznie, a nie nazwę do ręcznego utworzenia. Odczytaj rzeczywistą ścieżkę z komunikatu Wyniki. Zachowaj nazwy: porownanie_AB.json i opis_projektu.md.

Zapisz notebook przez Ctrl+S. Nie zastępuj nazw słowami final lub ostateczny. Do przekazania wybierz uzupełniony notebook i właściwy katalog wyników według ustaleń prowadzącego; karta nie dodaje nowych zasad zaliczenia.

Lista kontroli przed zakończeniem

- □ Otworzyłem właściwy notebook i zachowałem dane źródłowe.
- □ Wszystkie pola UZUPEŁNIJ są zastąpione własną analizą.
- □ Rozróżniam niepełny kontekst R3 i nietrafny kontekst R6.
- □ Nie utożsamiam punktów za wynik modelu z oceną mojej pracy.
- □ Otworzyłem plik wynikowy i sprawdziłem zgodność z analizowaną próbą.
- □ Zapisałem notebook oraz oznaczyłem rzeczywisty sposób pracy.

Gdy pojawi się problem

Objaw	Najpierw sprawdź
Nie można znaleźć pakietu lub pliku	Rozpakuj cały katalog Student. Sprawdź, czy obok notebooki znajdują się dane oraz kod.
W raporcie pozostało UZUPEŁNIJ	Zmień pola w komórce oceny, uruchom ją ponownie i przeczytaj zapisany raport.
Brak połączenia z modelem	Sprawdź uruchomienie Ollama i konfiguracja.json. Zmianę na demo oznacz jawnie; program nie robi tego automatycznie.
C poprawnie zgłasza brak danych	Oceń uczciwe rozpoznanie braku według rubryki; brak pełnej procedury nie oznacza automatycznie 0 punktów.

Zasada końcowa: zachowaj oryginalny wynik i własne uzasadnienie. Jeśli model się pomylił, opisz błąd zamiast poprawiać surową odpowiedź.

Raport porównawczy ze slajdu 100 ma charakter opcjonalny. Warunki przekazania pracy ustala prowadzący.