MODUŁ 6 • ROZDZIAŁ 5

Błędy i porównanie

Karta pracy uczestnika

Czas pracy	Materiały	Rezultat
Około 75 min	Notebook R5 i karta	Pliki opisane w karcie

Poprawa systemu zaczyna się od ustalenia przyczyny problemu. Jedna kontrolowana zmiana pozwala sprawdzić hipotezę, zamiast zgadywać.

Cel i rezultat rozdziału

Zdiagnozujesz trzy rodzaje błędów i porównasz konfigurację bazową z jednym wariantem. Wybierzesz konfigurację końcową na podstawie dowodów oraz regresji.

Po zakończeniu masz

- Rejestr błędów retrieval, odpowiedzi lub źródeł oraz rozmowy lub narzędzi.
- Porównanie tych samych przypadków po zmianie jednego czynnika.
- Uzasadniony wybór final_config.json i zestaw regresji.

Zanim zaczniesz

Przygotuj wyniki R3 i R4. Offline możesz porównać top_k równe 3 i 5. Zmiana promptu wymaga nowych odpowiedzi w trybie Ollama.

Otwórz notebook

```
notebooki/M12_M6_R5_analiza_bledow_start.ipynb
```

Plik lub materiał	Do czego służy
evaluation/retrieval_runs.jsonl i assistant_results.jsonl	Surowe dowody z poprzednich ćwiczeń.
szablony/error_analysis.md i error_log_template.json	Formularze diagnozy i analizy.

Tryb: checkpoint lub Ollama. Slajdy M12-M6-v6: 284–358. Czas orientacyjny: 75 min.

Praca krok po kroku

1. Wybierz trzy problemy

- Znajdź po jednym problemie wyszukiwania, odpowiedzi lub źródeł oraz rozmowy lub narzędzi.
- Możesz wykorzystać T004 oraz historyczne tura_1 i tura_2; wskaż pochodzenie.

2. Oddziel objaw od przyczyny

- Zapisz etap, problem_code, dowód, hipotezę i skutek dla użytkownika.
- Nie nazywaj wszystkiego błędem modelu. Sprawdź ranking i kontekst przed odpowiedzią.

3. Wybierz jeden czynnik

- Skopiuj bazę do candidate_config.json i zmień tylko top_k z 3 na 5.
- Alternatywnie w Ollama zbadaj rag_conditions_v1; nie łącz obu zmian w jednym eksperymencie.

4. Powtórz te same testy

- Użyj identycznych kontraktów, źródeł i wybranego podzbioru.
- Nowe wyniki otrzymują nowe result_id. Porównuj przez test_id, turn_id i kontrakt.

5. Sprawdź regresję

- Zestaw korzyści, pogorszenia i problemy nierozwiązane. Oceń też kryteria ręczne.
- Lepsze pokrycie nie oznacza lepszej odpowiedzi; wzrost szumu może niwelować korzyść.

6. Podejmij decyzję

- Wpisz wynik eksperymentu i wybierz bazę albo wariant jako final_config.json.
- Przy braku dowodów zapisz decyzję nierozstrzygniętą. Nie ogłaszaj automatycznie poprawy.

Offline porównujesz wyszukiwanie, nie jakość nowych odpowiedzi ani wydajność modeli. Zapisz tę granicę we wnioskach.

Punkty kontrolne pracy

Kontrola	Pytanie	Dowód
Diagnoza	Czy jest dowód i etap błędu?	result_id lub jawny rekord historyczny.
Eksperyment	Czy zmienił się jeden czynnik?	compare_configs: dokładnie jedna różnica.
Decyzja	Czy uwzględniono regresję?	Identyczne testy i uzasadniony wybór.

Nazewnictwo i zapis wyników

evaluation/error_log.jsonl; experiment_log.jsonl; error_analysis.md; regression_suite.json; final_config.json

W R6 połączysz decyzję z wynikami, instrukcją uruchomienia i scenariuszem pokazu.

Lista kontroli przed zakończeniem

□ Opisałem trzy rodzaje problemów.
□ Rozdzieliłem objaw, hipotezę i dowód.
□ Zmieniłem tylko jeden czynnik.
□ Porównałem te same kontrakty.
□ Sprawdziłem problemy i regresję.
□ Uzasadniłem konfigurację końcową.

Gdy pojawi się problem

Problem	Co zrobić
Zmieniły się dwa parametry	Rozdziel zmiany na dwa eksperymenty.
Inne przypadki przed i po	Powtórz pomiar na wspólnym zestawie.
Nie ma poprawy	Zachowaj bazę i zapisz negatywny wynik.
Brakuje trzech własnych błędów	Użyj oznaczonych dowodów z M5 i R3.

Zachowaj wyniki i uzasadnienia. Oznacz pochodzenie danych oraz próby niewykonane.