Plan ewaluacji

Karta pracy uczestnika

Czas pracy	Materiały	Rezultat
Około 45 min	Notebook R1 i karta	Pliki opisane w karcie

Działający prototyp nie jest jeszcze sprawdzonym systemem. Ustal, co i na jakiej podstawie oceniasz, zanim obejrzysz odpowiedzi asystenta.

Cel i rezultat rozdziału

Przygotujesz plan oceny DocAssist oraz zamrozisz konfigurację bazową. Oddzielisz wymagania dotyczące wyszukiwania, odpowiedzi, źródeł i zachowania asystenta.

Po zakończeniu masz

- Plan z zakresem, kryteriami i warunkami krytycznymi.
- Konfiguracja bazowa z index_id, model_digest i wersją promptu.
- Jawny tryb pracy oraz ograniczenia wybranej ścieżki.

Zanim zaczniesz

Rozpakuj pakiet Student. Uruchom notebook 00 i sprawdź dostarczony indeks M3 oraz materiały M5. Nie zmieniaj jeszcze parametrów systemu.

Otwórz notebook

```
notebooki/M12_M6_R1_plan_ewaluacji_start.ipynb
```

Plik lub materiał	Do czego służy
Student/START.md i notebook 00	Środowisko, tryb i sprawdzenie pakietu.
szablony/evaluation_plan.md	Formularz planu i zasad oceny.

Tryb: checkpoint lub Ollama. Slajdy M12-M6-v6: 8–77. Czas orientacyjny: 45 min.

Praca krok po kroku

1. Określ cel

- Opisz użytkownika, zakres dokumentów i dopuszczalne działania DocAssist.
- Wymień również pytania i działania pozostające poza zakresem.

2. Wybierz ścieżkę

- Checkpoint odtwarza zapisane wywołania. Ollama wykonuje nowe wywołania lokalnych modeli.
- Zapisz tryb w planie. Nie łącz obu trybów w jeden pomiar czasu lub jakości.

3. Zamroź bazę

- Zapisz baseline_config.json z faktycznie zastosowanej konfiguracji.
- Digest embeddingu pochodzi z manifestu indeksu. Tagi modeli nie zastępują digestów.

4. Ustal kryteria

- Przeczytaj dane/rubryki.json i rozdziel kryteria kontekstu od kryteriów odpowiedzi.
- Ustal znaczenie ocen 0–2 i N/D. Błąd krytyczny oznacza niezaliczenie przypadku.

5. Zapisz plan

- Uzupełnij evaluation_plan.md: zakres, dane, tryb, kryteria, minimum dowodów i ograniczenia.
- Określ, które próby będą nowe, odtworzone lub importowane z M5.

6. Sprawdź gotowość

- Potwierdź, że inna osoba potrafi odtworzyć konfigurację na podstawie zapisanych plików.
- Zapisz notebook. Plan zatwierdź przed ocenianiem wyników w R3 i R4.

Konfiguracja bazowa jest punktem odniesienia. Kolejne warianty zapisuj osobno, bez nadpisywania bazy.

Punkty kontrolne pracy

Kontrola	Pytanie	Dowód
Plan	Czy wiadomo, co sprawdzamy?	Zakres, kryteria, warunki krytyczne.
Konfiguracja	Czy parametry są rzeczywiste?	index_id, digests, tryb i prompt_version.
Pochodzenie	Czy źródła wyników są jawne?	Nowy pomiar / checkpoint / import M5.

Nazewnictwo i zapis wyników

evaluation/evaluation_plan.md; evaluation/baseline_config.json

W R2 zbudujesz kontrakty przypadków zgodne z tym planem.

Lista kontroli przed zakończeniem

□ Mam jawny zakres i granice działania.
□ Znam znaczenie 0, 1, 2 i N/D.
□ Ustaliłem warunki krytyczne.
□ Zapisałem baseline_config.json.
□ Nie zmieniłem bazy po obejrzeniu wyników.
□ Wskazałem ograniczenia trybu pracy.

Gdy pojawi się problem

Problem	Co zrobić
Brak Ollama	Wybierz checkpoint i zadeklaruj zakres offline.
Niezgodny digest	Użyj całego dostarczonego indeksu i checkpointów.
Niejasne kryteria	Dopisz przykład oceny i warunek krytyczny.
Nie wiem, co oddać	Sprawdź minimum dowodów w instrukcji modułu.

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