

O autorze

O korektorach merytorycznych

Wprowadzenie

Rozdział 1. Wprowadzenie do dużych modeli językowych

- Czym jest duży model podstawowy i duży model językowy?
 - Przesunięcie paradygmatu sztucznej inteligencji - wprowadzenie do modeli podstawowych
 - Budowa dużego modelu językowego
- Najpopularniejsze architektury oparte na transformerach dużego modelu językowego
 - Wczesne eksperymenty
 - Wprowadzenie do architektury transformera
- Trenowanie dużego modelu językowego i jego ocena
 - Trenowanie dużego modelu językowego
 - Ocena modelu
- Modele podstawowe kontra modele dostosowane do własnych potrzeb
 - Jak można dostosować model do własnych potrzeb?
- Podsumowanie
- Odwołania

Rozdział 2. Duże modele językowe w aplikacjach wspomaganych przez sztuczną inteligencję

- Sposoby, w jakie duże modele językowe zmieniły branżę tworzenia oprogramowania
- System copilot
- Wprowadzenie do frameworków koordynowania sztucznej inteligencji umożliwiających osadzanie dużych modeli językowych w aplikacjach
 - Najważniejsze komponenty frameworków koordynowania sztucznej inteligencji
 - LangChain
 - Haystack
 - Semantic Kernel
 - Jak wybrać framework?
- Podsumowanie
- Odwołania

Rozdział 3. Wybór dużego modelu językowego dla aplikacji

- Ogólne omówienie najbardziej obiecujących dużych modeli językowych dostępnych na rynku
 - Modele własnościowe
 - Modele otwartoźródłowe
- Nie tylko modele językowe

- Kryteria stosowane podczas wyboru właściwego dużego modelu językowego

- Rozważania
- Studium przypadku

- Podsumowanie
- Odwołania

Rozdział 4. Prompt engineering

- Wymagania techniczne
- Czym jest prompt engineering?
- Reguły prompt engineering
 - Jasne instrukcje
 - Podziel skomplikowane zadania na mniejsze
 - Zapytaj o uzasadnienie
 - Wygenerowanie wielu danych wyjściowych, a następnie użycie modelu do wybrania tych najlepszych
 - Powtórzenie instrukcji na końcu
 - Używanie ograniczników
- Techniki zaawansowane
 - Technika few-shots
 - Technika chain of thought
 - ReAct
- Podsumowanie
- Odwołania

Rozdział 5. Osadzanie dużych modeli językowych w aplikacjach

- Wymagania techniczne
- Wprowadzenie do frameworka LangChain
- Rozpoczęcie pracy z frameworkiem LangChain
 - Modele i prompty
 - Połączenia danych
 - Pamięć
 - Łańcuch
 - Agenty
- Praca z dużymi modelami językowymi poprzez platformę Hugging Face Hub
 - Tworzenie tokena dostępu użytkownika Hugging Face Hub
 - Przechowywanie kluczy tajnych użytkownika w pliku .env
 - Rozpoczęcie pracy z otwartoźródłowymi dużymi modelami językowymi
- Podsumowanie

- Odwołania

Rozdział 6. Tworzenie aplikacji konwersacyjnych

- Wymagania techniczne
- Rozpoczęcie pracy z aplikacjami konwersacyjnymi
 - Tworzenie zwykłego bota
 - Dodawanie pamięci
 - Dodawanie wiedzy nieparametrycznej
 - Dodawanie narzędzi zewnętrznych
- Opracowanie frontendu za pomocą Streamlit
- Podsumowanie
- Odwołania

Rozdział 7. Używanie dużych modeli językowych do tworzenia silników wyszukiwania i rekomendacji

- Wymagania techniczne
- Wprowadzenie do systemu rekomendacji
- Istniejące systemy rekomendacji
 - Algorytm K najbliższych sąsiadów
 - Rozkład macierzy
 - Sieci neuronowe
- Jak duże modele językowe zmieniają systemy rekomendacji?
- Implementowanie systemu rekomendacji opartego na dużym modelu językowym
 - Wstępne przygotowanie danych
 - Tworzenie chatbota rekomendacji w scenariuszu typu zimny start
 - Tworzenie systemu opartego na treści
- Opracowanie frontendu za pomocą biblioteki Streamlit
- Podsumowanie
- Odwołania

Rozdział 8. Używanie dużych modeli językowych w połączeniu z danymi strukturyzowanymi

- Wymagania techniczne
- Czym są dane strukturyzowane?
- Rozpoczęcie pracy z relacyjnymi bazami danych
 - Wprowadzenie do relacyjnych baz danych
 - Ogólne omówienie bazy danych Chinook
 - Jak za pomocą Pythona pracować z relacyjną bazą danych?
- Implementacja aplikacji DBCopilot z użyciem frameworka LangChain
 - Agenty frameworka LangChain i agent SQL
 - Prompt engineering
 - Dodawanie kolejnych narzędzi

- Opracowanie frontendu za pomocą biblioteki Streamlit
- Podsumowanie
- Odwołania

Rozdział 9. Praca z kodem źródłowym

- Wymagania techniczne
- Wybór odpowiedniego dużego modelu językowego przeznaczonego do pracy z kodem źródłowym
- Analizowanie i generowanie kodu źródłowego
 - Falcon LLM
 - CodeLlama
 - StarCoder
- Działanie w charakterze algorytmu
- Wykorzystanie API Code Interpreter
- Podsumowanie
- Odwołania

Rozdział 10. Tworzenie wielomodalnych aplikacji wykorzystujących duże modele językowe

- Wymagania techniczne
- Dlaczego wielomodalność?
- Tworzenie agenta wielomodalnego za pomocą frameworka LangChain
- Opcja 1. Użycie standardowego zbioru narzędzi dla Azure AI Services
 - Rozpoczęcie pracy z AzureCognitiveServicesToolkit
- Opcja 2. Połączenie narzędzi w postaci pojedynczego agenta
 - Narzędzia YouTube i Whisper
 - DALL-E i generowanie tekstu
 - Połączenie wszystkiego w całość
- Opcja 3. Podejście na stałe zdefiniowane w kodzie z użyciem łańcucha sekwencyjnego
- Porównanie wszystkich trzech opcji
- Opracowanie frontendu za pomocą biblioteki Streamlit
- Podsumowanie
- Odwołania

Rozdział 11. Dostrajanie dużych modeli językowych

- Wymagania techniczne
- Czym jest dostrajanie?
- Kiedy dostrajanie modelu jest konieczne?
- Rozpoczęcie dostrajania
 - Pobranie zbioru danych
 - Tokenizacja danych

- Dostrajanie modelu
- Używanie wskaźników oceny modelu
- Trenowanie i zapisywanie

- Podsumowanie
- Odwołania

Rozdział 12. Odpowiedzialna sztuczna inteligencja

- Czym jest odpowiedzialna sztuczna inteligencja i dlaczego jej potrzebujemy?
- Architektura odpowiedzialnej sztucznej inteligencji
 - Warstwa modelu
 - Warstwa metapromptu
 - Warstwa interfejsu użytkownika
- Regulacje prawne dotyczące odpowiedzialnej sztucznej inteligencji
- Podsumowanie
- Odwołania

Rozdział 13. Najnowsze trendy i innowacje

- Najnowsze trendy w modelach językowych i generatywnej sztucznej inteligencji
 - GPT-4V(ision)
 - DALL-E 3
 - AutoGen
 - Małe modele językowe
- Firmy wykorzystujące generatywną sztuczną inteligencję
 - Coca-Cola
 - Notion
 - Malbek
 - Microsoft
- Podsumowanie
- Odwołania