

<b>Wstęp .....</b>	<b>9</b>
<b>1. Podstawy modeli językowych z wykorzystaniem LangChaina .....</b>	<b>25</b>
Konfiguracja projektu LangChain	27
Stosowanie modeli LLM w bibliotece LangChain	29
Tworzenie promptów nadających się do wielokrotnego użycia	32
Uzyskiwanie od modeli LLM wyników w określonych formatach	37
Dane wyjściowe w formacie JSON	37
Inne formaty czytelne dla oprogramowania z wykorzystaniem parserów wyjściowych	38
Łączenie wielu elementów aplikacji korzystającej z modelu LLM	39
Wykorzystanie interfejsu Runnable	39
Łączenie imperatywne	41
Łączenie deklaratywne	44
Podsumowanie	45
<b>2. RAG, część I: Indeksowanie danych .....</b>	<b>47</b>
Cel: wybór odpowiedniego kontekstu dla modeli językowych	48
Osadzenia: zamiana tekstu na liczby	49
Osadzenia przed erą modeli LLM	49
Osadzenia oparte na modelach językowych	51
Wyjaśnienie osadzeń semantycznych	51
Konwersja dokumentów na tekst	54
Dzielenie tekstu na mniejsze fragmenty	56
Generowanie osadzeń tekstów	60
Przechowywanie osadzeń w wektorowej bazie danych	63
Konfiguracja PGVector	64
Praca z magazynami wektorów	65
Śledzenie zmian w dokumentach	68

Optymalizacja indeksowania	72
MultiVectorRetriever	72
RAPTOR — rekurencyjne przetwarzanie abstrakcyjne dla wyszukiwania w strukturach drzewiastych	76
ColBERT — optymalizacja osadzeń	77
Podsumowanie	79
<b>3. RAG, część II: Rozmawianie z własnymi danymi .....</b>	<b>81</b>
Wprowadzenie do generowania wspomaganego wyszukiwaniem	81
Wyszukiwanie odpowiednich dokumentów	83
Generowanie predykcji modeli LLM na podstawie istotnych dokumentów	86
Przekształcanie zapytań	91
Przepisz – wyszukaj – przeczytaj	92
Wyszukiwanie danych z użyciem wielu zapytań	94
RAG-Fusion	97
Hipotetyczne osadzanie dokumentów	101
Trasowanie zapytań	104
Trasowanie logiczne	104
Trasowanie semantyczne	108
Tworzenie zapytań	110
Filtr tekst na metadane	110
Przetwarzanie języka naturalnego na zapytania SQL	113
Podsumowanie	115
<b>4. Dodanie pamięci do chatbota z wykorzystaniem LangGraphu .....</b>	<b>117</b>
Tworzenie systemu pamięci dla chatbotów	118
Wprowadzenie do LangGraphu	120
Tworzenie obiektu grafu stanowego	123
Dodawanie pamięci do obiektu grafu stanowego	126
Modyfikowanie historii czatu	129
Przycinanie wiadomości	129
Filtrowanie wiadomości	131
Łączenie następujących po sobie komunikatów	133
Podsumowanie	135
<b>5. Architektury poznawcze z wykorzystaniem LangGraphu .....</b>	<b>137</b>
Architektura nr 1. Wywołanie LLM	140
Architektura nr 2. Sekwencja	142
Architektura nr 3. Router	147
Podsumowanie	154

<b>6. Architektura agentowa .....</b>	<b>155</b>
Pętla planowania i działania	156
Tworzenie agenta opartego na grafach językowych	158
Zawsze w pierwszej kolejności używaj narzędzi	162
Praca z wieloma narzędziami	167
Podsumowanie	171
<b>7. Agenty II .....</b>	<b>173</b>
Refleksja	173
Podgrafy w LangGraphie	179
Bezpośrednie wywołanie podgrafu	180
Wywoływanie podgrafu przy użyciu funkcji	181
Architektury wieloagentowe	183
Architektura z nadzorcą	184
Podsumowanie	187
<b>8. Wzorce efektywnego wykorzystania wielkich modeli językowych .....</b>	<b>189</b>
Ustrukturyzowane dane wyjściowe	191
Wynik pośredni	193
Strumieniowanie wyników modelu LLM token po tokenie	196
Tryby rozwiązań typu human in the loop	197
Wielozadaniowość modeli językowych	203
Podsumowanie	206
<b>9. Wdrażanie: uruchamianie aplikacji AI w środowisku produkcyjnym .....</b>	<b>207</b>
Wymagania wstępne	207
Instalacja zależności	208
Duży model językowy	208
Magazyn wektorowy	208
API backendu	213
Tworzenie konta LangSmith	215
Prezentacja API LangGraph Platform	216
Modele danych	216
Możliwości	217
Wdrażanie aplikacji AI w LangGraph Platform	219
Utwórz konfigurację API LangGraph	219
Przetestuj swoją aplikację LangGraph lokalnie	220
Wdrażanie z wykorzystaniem interfejsu użytkownika LangSmith	222
Uruchomienie LangSmith Studio	224
Bezpieczeństwo	227
Podsumowanie	229

<b>10. Testowanie: ocena, monitorowanie i ciągłe doskonalenie .....</b>	<b>231</b>
Techniki testowania w cyklu rozwoju aplikacji	
opartych na modelach językowych	231
Etap projektowania: RAG z autokorektą	232
Etap przed wdrożeniem	239
Tworzenie zestawów danych	239
Określanie kryteriów oceny	241
Testy regresyjne	246
Ocena całościowej wydajności agenta	247
Produkcja	259
Śledzenie	259
Zbieranie informacji zwrotnych w środowisku produkcyjnym	260
Klasyfikacja i znakowanie	261
Monitorowanie i usuwanie błędów	261
Podsumowanie	261
<b>11. Budowanie z wykorzystaniem modeli językowych .....</b>	<b>263</b>
Interaktywne chatboty	264
Wspólna praca nad tekstem z wykorzystaniem modeli LLM	265
Przetwarzanie w otoczeniu	267
Podsumowanie	268