

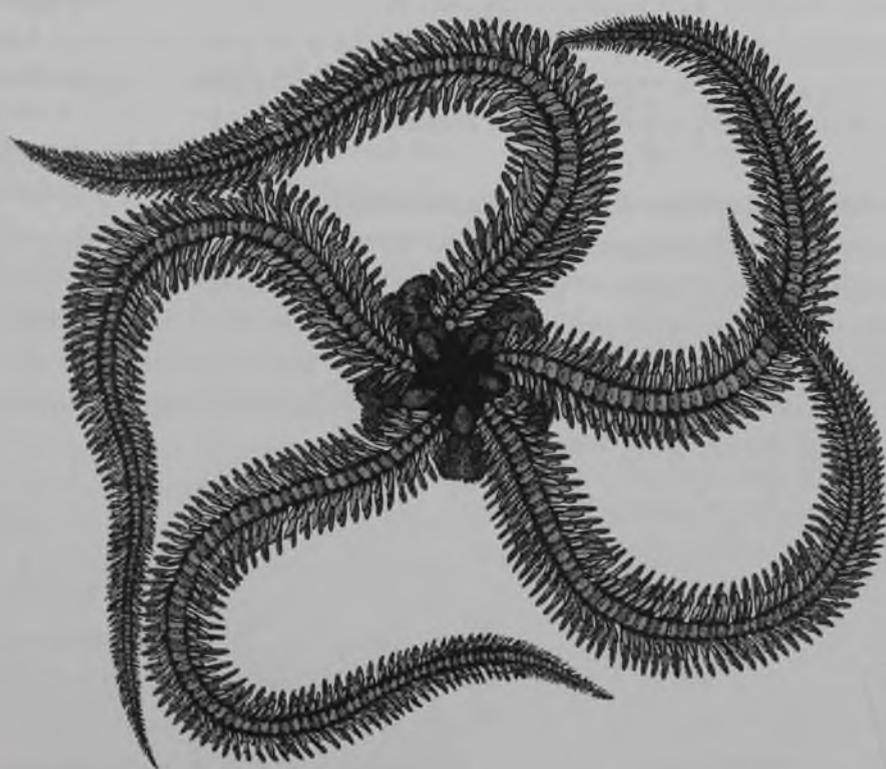
СН[№] 114230

O'REILLY®

НЕ СЕ ИЗНАСЯ

Разработка на приложения с GPT-4 и ChatGPT

Създаване на интелигентни чатботове, генератори на съдържание и много други



ИЗДАТЕЛСТВО
АСЕНЕВЦИ

Оливие Кейлън и
Мари-Алис Блете

к 93584 007.8

© издателство Асеновци, 2023

© Влади Владев, превод от английски, 2023

14 ноември 2023: актуализация

Посетете <http://oreilly.com/catalog/errata.csp?isbn=978109>

133474

БИБЛИОТЕКА - УНСС

Инв. №



120000133476

изации.

ISBN: 978-619-266-019-2

Първо издание. Лицензиран превод от английски език. Заглавие на оригинала:

Developing Apps with GPT-4 and ChatGPT ISBN 9781098152482

by Olivier Caelen and Marie-Alice Blete, published by O'Reilly Media, Inc.,

1005 Gravenstein Highway North, Sebastopol, CA 95472 in 2023

This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. Developing Apps with GPT-4 and ChatGPT, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

Този превод се публикува и продава с разрешение на O'Reilly Media, Inc., която притежава или контролира всички права за публикуване и продажба на същия.

Логото на O'Reilly е регистрирана търговска марка на O'Reilly Media, Inc. Изображението на корицата и свързания търговски вид са търговски марки на O'Reilly Media, Inc.

Изданието като цяло, както и нито една част от него, не може да се възпроизвежда, съхранява или разпространява под каквато и да е форма, и по какъвто и да е начин, без изричното писмено съгласие от издателя. Някои от наименованията, споменати в книгата са запазени марки и принадлежат на техните собственици.

Изключване на гаранциите и ограничаване на отговорността:

Информацията в тази книга се разпространява на основа "такава, каквато е" („as is“), без каквито и да било гаранции. Въпреки, че са взети всички необходими предпазни мерки при подготовката на тази книга, авторът и издателят не носят никаква отговорност спрямо което и да е физическо или юридическо лице за причинени или предполагаеми щети или вреди, възникнали пряко или косвено от инструкциите, съдържащи се в тази книга, или от компютърният софтуер или хардуер, описани в книгата.



Съдържание

Предисловие	5	<i>„Морган Стенли“</i>	26
Конвенции, използвани в книгата.....	6	<i>Khan Academy</i>	26
Използване на примерните кодове.....	6	<i>Duolingo</i>	27
<i>Онлайн обучение с O'Reilly</i>	7	<i>Yabble</i>	28
<i>Как да се свържете с нас?</i>	7	<i>Waymark</i>	28
<i>Благодарности</i>	8	<i>Inworld AI</i>	29
Основи на GPT-4 и ChatGPT	9	Пазете се от халюцинациите на ИИ: ограничения и съображения	29
Въведение в големите езикови модели	9	Оптимизиране на GPT моделите с плъгини и фина настройка	33
<i>Преглед на основите на езиковите модели и NLP</i>	10	Обобщение	34
<i>Разбиране на архитектурата на трансформатора и ролята ѝ в LLM</i>	12	Подробен обзор на интерфейсите на GPT-4 и ChatGPT	36
<i>Разкриване тайната на токенизацията и предвиждането в GPT моделите</i>	16	Основни понятия	37
Кратка история: от GPT-1 до GPT-4	17	Достъпни модели в OpenAI API.....	37
<i>GPT-1</i>	17	Тестване на GPT моделите с изралната площадка на OpenAI	39
<i>GPT-2</i>	19	Начало на работата: OpenAI Python Library	45
<i>GPT-3</i>	19	<i>Достъп до OpenAI и API ключ</i>	45
<i>От GPT-3 до инструктиран GPT</i>	21	<i>Пример Hello World</i>	47
<i>GPT-3.5, GPT-4 и ChatGPT</i>	23	Използване на ChatGPT и GPT-4	49
<i>GPT-4</i>	24	<i>Входни параметри за Chat Completion</i>	50
Приложение на LLM и примерни продукти	25	<i>Форматиране на резултата за Chat Completion</i>	53
<i>Be My Eyes</i>	25		

<i>От завършване на текст до функции</i>	54
Използване на други модели за завършване на текст	57
<i>Входни параметри за крайната точка за завършване на текст</i> ..	58
<i>Формат на извода на крайната точка за завършване на текст</i> ..	59
Съобщения	59
<i>Ценообразуване и токен ограничения</i>	59
<i>Сигурност и конфиденциалност: Внимание!</i>	60
Други API интерфейси на OpenAI и функции	61
<i>Вграждания</i>	61
<i>Модел модератор</i>	64
<i>Whisper и DALL-E</i>	67
Обобщение (и пищов)	67

Създаване на приложения с GPT-4 и ChatGPT 69

Основни моменти при разработката на приложения	69
<i>Управление на API ключа</i>	69
<i>Безопасност и конфиденциалност на данните</i>	72
Принципи за проектиране на софтуерна архитектура	72
Уязвимости на приложенията с LLM	73
<i>Анализ на входящите и изходящите данни</i>	74
<i>Неизбежността на промпт инжекциите</i>	75
Примерни проекти	75
<i>Проект 1: Създаване на решение за генериране на новини</i>	76
<i>Проект 2: Резюмиране на YouTube videa</i>	79
<i>Проект 3: Създаване на експерт за Zelda BOTW</i>	82
<i>Събираме всичко заедно</i>	87
<i>Проект 4: Гласов контрол</i>	89
Обобщение	96

Интелигентни техники за работа с GPT-4 и ChatGPT 97

Промпт инженеринг	97
<i>Проектиране на ефективни промптове</i>	98
<i>Ролята</i>	104
<i>Мислене стъпка по стъпка</i>	105
<i>Имплементиране на обучение с няколко изстрела</i>	108
<i>Повишаване на ефективността на промптовете</i>	110
Фина настройка	112
<i>Подготовка за работа</i>	112
<i>Фина настройка с OpenAI API</i> ..	115
<i>Приложение на фина настроените модели</i>	118
<i>Генератор на новинарски статии по нишови теми</i>	121
<i>Генериране на фина настроени синтетични данни за маркетинг кампания с имейли</i>	121
<i>Фина настройка на модел със синтетични данни</i>	126
<i>Колко струва фината настройка?</i>	130
Обобщение	130

Усъвършенстване свойствата на LLM с фреймуърка LangChain и плъгини 133

Фреймуъркът LangChain	133
<i>Динамични промптове</i>	135
<i>Агенти и инструменти</i>	136
<i>Памет</i>	140
<i>Вграждания</i>	142
GPT-4 плъгини	146
<i>Преглед</i>	148
<i>API 148</i>	
<i>Манифестът на плъгина</i>	150
<i>OpenAPI спецификация</i>	151
<i>Описания</i>	152
Обобщение	153
Заклучение	154

Речник на термините 155