
Оглавление

От издательства	13
О научном редакторе русского издания	13
Предисловие научного редактора русского издания	13
Отзывы о книге	15
Предисловие	16
О чем эта книга	17
Кому эта книга не подойдет	18
Для кого эта книга	18
Структура книги	19
Условные обозначения	20
Использование исходного кода примеров	21
Благодарности	22
Глава 1. Знакомство с агентами	24
Что такое ИИ-агенты?	24
Революция в предварительном обучении	25
Типы агентов	27
Выбор модели	29
От синхронных операций к асинхронным	30
Практические применения и сценарии использования	31
Рабочие потоки и агенты	33
Принципы построения эффективных агентных систем	36
Организационные основы успешного создания агентных систем	37
Агентные фреймворки	38
LangGraph	38
AutoGen	39
CrewAI	39
OpenAI Agents Software Development Kit (SDK)	39
Заключение	40

Глава 2. Проектирование агентных систем	41
Наша первая агентная система	41
Основные компоненты агентных систем.....	44
Выбор модели	45
Инструменты.....	49
Проектирование функциональности для конкретных задач.....	49
Интеграция инструментов и модульность	50
Память.....	50
Краткосрочная память.....	50
Долгосрочная память.....	51
Управление памятью и извлечение данных.....	51
Оркестрация	52
Компромиссы проектирования	52
Производительность: скорость / доля верных результатов.....	52
Масштабируемость: проектирование масштабируемых агентных систем.....	53
Надежность: обеспечение надежного и последовательного поведения агента	55
Затраты: баланс между производительностью и расходами	56
Архитектурные паттерны проектирования	57
Одноагентные архитектуры	57
Мультиагентные архитектуры: сотрудничество, параллелизм и координация	58
Лучшие практики	60
Итеративное проектирование.....	60
Стратегия оценки	61
Тестирование в реальных условиях.....	63
Заключение	65
Глава 3. UX-дизайн для агентных систем	66
Модальности взаимодействий.....	67
Текст	68
Графические интерфейсы	71
Речевые и голосовые интерфейсы.....	75
Интерфейсы на основе видео	79
Объединение модальностей для бесшовного взаимодействия.....	80
Регулировка автономности.....	81
Синхронное и асинхронное взаимодействие с агентами	83
Принципы проектирования синхронных взаимодействий	84
Принципы проектирования асинхронных взаимодействий.....	85

Баланс между проактивностью и назойливостью.....	85
Удержание контекста и непрерывность	86
Поддержание состояния между взаимодействиями	87
Персонализация и адаптируемость	88
Информирование о возможностях агента	89
Передача уверенности и неуверенности.....	91
Запрос указаний и пользовательского ввода.....	92
Корректное преодоление сбоев.....	92
Доверие при проектировании взаимодействий	93
Заключение	95
Глава 4. Инструменты.....	97
Основы LangChain.....	98
Локальные инструменты	99
Инструменты на базе API.....	101
Плагины	105
MCP	108
Инструменты с состоянием	111
Автоматизированная разработка инструментов	112
Фундаментальные модели для создания инструментов	113
Генерация кода в реальном времени	113
Конфигурация инструментов.....	115
Заключение	116
Глава 5. Оркестрация	117
Типы агентов	117
Рефлекторный агент	118
Агент ReAct.....	118
Агент «планировщик-исполнитель».....	119
Агент декомпозиции запросов.....	119
Агент с рефлексией и метарассуждениями	120
Агент глубокого исследования.....	120
Выбор инструмента	122
Стандартный выбор инструментов.....	122
Семантический выбор инструментов.....	125
Иерархический выбор инструментов	129
Выполнение инструментов.....	134
Топологии инструментов	134

Выполнение одного инструмента	135
Параллельное выполнение инструментов	135
Цепочки	136
Графы	138
Контекст-инжиниринг	141
Заключение	143
Глава 6. Знания и память	144
Основные подходы к работе с памятью	145
Управление окнами контекста	145
Традиционный полнотекстовый поиск	147
Семантическая память и векторные хранилища	148
Семантический поиск	148
Реализация семантической памяти на базе векторных хранилищ	149
RAG	150
Семантическая память с учетом накопленного опыта	152
GraphRAG	152
Использование графов знаний	153
Построение графов знаний	154
Перспективы и риски динамических графов знаний	160
Ведение заметок (Note-Taking)	162
Заключение	163
Глава 7. Обучение в агентных системах	164
Непараметрическое обучение	164
Непараметрическое обучение по образцу	164
Рефлексия	166
Обучение через опыт	170
Параметрическое обучение: тонкая настройка	175
Тонкая настройка больших фундаментальных моделей	176
Перспективы малых моделей	180
Тонкая настройка с учителем	183
Прямая оптимизация предпочтений	187
Обучение с подкреплением с проверяемым вознаграждением	191
Заключение	192
Глава 8. От одного агента ко многим	193
Сколько нужно агентов?	193

Одноагентные сценарии.....	193
Мультиагентные сценарии.....	200
Рой.....	208
Принципы добавления агентов.....	209
Мультиагентная координация.....	211
Демократическая координация.....	212
Координация менеджерами.....	212
Иерархическая координация.....	213
Методы «актер — критик».....	214
Автоматизированное проектирование агентных систем.....	215
Методы коммуникаций.....	220
Локальные и распределенные коммуникации.....	221
Протокол «агент — агент».....	221
Брокеры сообщений и шины событий.....	225
Фреймворки акторов: Ray, Orleans и Akka.....	228
Оркестрация и механизмы рабочих потоков.....	232
Управление состоянием и долгосрочное хранение.....	234
Заключение.....	236

Глава 9. Проверки и измерения 238

Измерение характеристик агентных систем.....	238
Измерение как краеугольный камень.....	239
Интеграция оценки в жизненный цикл разработки.....	240
Создание и масштабирование оценочных наборов.....	240
Оценка компонентов.....	242
Оценка инструментов.....	242
Оценка планирования.....	243
Оценка памяти.....	245
Оценка обучения.....	246
Целостная оценка.....	247
Производительность в сквозных сценариях.....	248
Согласованность.....	250
Связность.....	252
Галлюцинации.....	252
Обработка непредвиденного ввода.....	254
Подготовка к развертыванию.....	255
Заключение.....	256

Глава 10. Мониторинг на стадии эксплуатации	257
Мониторинг как механизм обучения.....	258
Стеки мониторинга	261
Grafana с OpenTelemetry, Loki и Tempo.....	262
Стек ELK (Elasticsearch, Logstash/Fluentd, Kibana)	262
Arize Phoenix.....	263
SigNoz	264
Langfuse.....	264
Выбор оптимального стека.....	265
Инструментация OTel	266
Визуализация и оповещения.....	268
Паттерны мониторинга.....	271
Теневого режим	271
Канареечные развертывания.....	271
Сбор трассировки регрессий.....	272
Самовосстанавливающиеся агенты	272
Обратная связь от пользователей как сигнал наблюдаемости.....	273
Смещения распределения	273
Принадлежность метрик и кросс-функциональное управление.....	276
Заключение	279
Глава 11. Циклы улучшения.....	280
Пайплайны обратной связи	282
Автоматизированное обнаружение проблем и анализ корневых причин	287
Ревью с участием человека	289
Уточнение промптов и инструментов.....	292
Агрегирование и приоритеты улучшений	297
Эксперименты	299
Теневое развертывание	300
A/B-тестирование.....	301
Байесовские бандиты.....	303
Непрерывное обучение	305
Контекстное обучение	305
Автономное повторное обучение.....	307
Заключение	308

Глава 12. Защита агентных систем	310
Уникальные риски агентных систем	311
Новые векторы угроз	313
Защита фундаментальных моделей	315
Методы защиты	316
Красная команда	319
Моделирование угроз с MAESTRO	321
Защита данных в агентных системах	324
Конфиденциальность данных и шифрование	324
Происхождение и целостность данных	326
Обработка чувствительных данных	328
Защита агентов	330
Средства защиты	330
Защита от внешних угроз	332
Защита от внутренних сбоев	335
Заключение	339
Глава 13. Взаимодействия «человек — агент»	340
Роли и автономность	340
Меняющаяся роль человека в агентных системах	340
Согласование целей со стейкхолдерами и содействие принятию	342
Масштабирование сотрудничества	344
Область действия агентов и организационные роли	345
Общая память и границы контекстов	347
Доверие, управление и комплаенс	349
Жизненный цикл доверия	349
Схемы подотчетности	351
Механизм эскалации и контроль	354
Приватность и комплаенс	355
Заключение: будущее взаимодействий «человек — агент»	357
Глоссарий	360
Об авторе	365
Обложка	366