Оглавление

О чем расскажет эта книга 7
Об авторе 8
Введение 9

Глава 0. От компонентов к системе: анатомия интеллектуального агента 19

0.1. Генеративные модели: рождение синтетического мышления 20
От дискриминативных к генеративным моделям 20
Трансформеры: архитектура, изменившая мир 20
БЯМ: от текста к универсальному интерфейсу 21
Мультимодальность: агенты воспринимают мир целостно 22
Ограничения генеративных моделей: где им нужна помощь 22
Мостик к агентам 23
0.2. Метакогнитивная промпт-инженерия: система управления разумом модели 23
От простого промптинга к метакогниции 23
Chain-of-Thought: визуализация рассуждений 24
ReAct: цикл «размышление → действие» 24
Метакогнитивные техники: инструментарий саморегуляции 25
Цикл метакогнитивного агента 26
Ограничения метакогнитивного промптинга 27
0.3. RAG-системы: органы восприятия агента 27
Проблема параметрических знаний 27
Архитектура RAG: retriever + generator 28
Векторные базы данных и семантический поиск 29
Гибридный поиск и ранжирование 29
От наивного RAG к агентному 30
Инструментализация: RAG как часть экосистемы 30
Качество данных: мусор на входе – мусор на выходе 31
Ограничения RAG: когда данных слишком много 31
Мостик к агентам 31

ЧАСТЬ I. ОСНОВЫ АГЕНТНОГО ПОДХОДА 35

Глава 1. Агенты в искусственном интеллекте 36

Глава 2. Архитектурные принципы построения агентов 45

Глава 3. Агенты и многоагентные системы 57

ЧАСТЬ II. ПРОЕКТИРОВАНИЕ АГЕНТНЫХ СИСТЕМ..... 71

Глава 4. Инструменты и их интеграция 72

4.1. Ландшафт инструментов для агентов на основе БЯМ..... 73
4.2. Архитектурные паттерны интеграции инструментов 77
4.3. Практическая реализация: пример агента с инструментальной интеграцией..... 80

Глава 5. Планирование действий и их выполнение..... 90

5.1. Цикл ReAct – от мысли к действию..... 90
5.2. Планирование, декомпозиция и управление зависимостями 94
5.3. Реализация ReAct-агента на языке Python..... 98

Глава 6. Управление состоянием и память 103

6.1. Краткосрочная и долгосрочная память – архитектурные принципы.. 104
6.2. Агентная память и динамическое управление знаниями 107
6.3. Метакогнитивная память в многоагентных системах – совместное использование знаний 110
6.4. Реализация агента с многоуровневой памятью 113

Глава 7. Руководство для разработчиков 118

7.1. Принципы разработки агентов 118
7.2. Паттерны проектирования агентных систем..... 122
Паттерны создания целей..... 123
Паттерны оптимизации промптов и ответов 124
Паттерны расширения контекста 125
Паттерны рефлексии и самоконтроля 125
Паттерны сотрудничества между агентами 126
Паттерны безопасности и соответствия 128
Паттерны управления инструментами и агентами 128
Паттерны планирования и выполнения..... 130
Паттерны взаимодействия с моделью 131
Применение паттернов: модель принятия решений 133

ЧАСТЬ III. КОНТРОЛЬ КАЧЕСТВА И БЕЗОПАСНОСТЬ..... 137

Глава 8. Метрики качества..... 138

8.1. Перечень метрик качества агентов 138
Метрики итоговых результатов..... 139
Метрики промежуточного поведения 140
Метрики производительности 140
Метрики надёжности и безопасности..... 141
Метрики многоагентных систем..... 142
Метрики пользовательского опыта..... 142
Специфические метрики для отдельных задач 142
8.2. Программная реализация контроля метрик качества 143

Глава 9. Принципы тестирования агентов 151

9.1. Функциональное тестирование – проверка корректности выполнения задач 152
Юнит-тестирование компонентов агента 152
Интеграционное тестирование взаимодействия с инструментами 153
Сквозное тестирование полного пользовательского пути 154
Регрессионное тестирование 156
Оценка через сравнение с эталонами и механизм «БЯМ как судья» 157
9.2. Тестирование надёжности – устойчивость к ошибкам и нештатным ситуациям 159
Симуляция отказов инструментов и внешних зависимостей 159
Тестирование восстановления после ошибок 160
Тестирование на противоречивых и неполных данных 162
Автоматизированное внедрение ошибок 162
Итеративное улучшение на основе результатов тестирования 163
9.3. Тестирование безопасности – защита от враждебных атак и утечек данных 164
Инъекции промптов и защита от них 164
Джейлбрейки и обход ограничений 165
Извлечение системных промптов и конфиденциальных данных 166
Тестирование на генерацию токсичного и вредоносного контента 167
Состязательное тестирование через диверсионный анализ 168
Защита данных и соблюдение политик доступа 168
9.4. Тестирование производительности и масштабируемости 169
Измерение задержки ответа и её компонентов 169
Нагрузочное тестирование параллельных сессий 171
Стресс-тестирование за пределами проектных ограничений 171
Тестирование выносливости и утечек ресурсов 172
Оптимизация на основе результатов тестирования 173
9.5. Прочие принципы тестирования агентов 174
Многоуровневая архитектура тестов 174
Репрезентативность тестовых наборов 174
Автоматизация оценки результатов 175
«Человек в цикле» для субъективной оценки 175
Непрерывность тестирования на всех этапах 175
Воспроизводимость результатов тестирования 176
Отказоустойчивость самих тестов 176
Документирование и трассировка результатов 176
Итеративное улучшение тестовых наборов 177

Глава 10. Безопасность и доверие 178

10.1. Архитектурные паттерны защитных гардов 179
Модель швейцарского сыра для многослойной защиты 179
Паттерн оптимистического выполнения 180
Паттерн «Политики прежде всего» 180
Каскадные гарды 181

Адаптивные гарды AGrail ... 182
Агенты-охранники GuardAgent ... 182
Контекстуализация правил безопасности ... 183
Алгоритм выбора паттерна для реализации гардов ... 184
10.2. Мультимодальная модерация и соответствие нормативным требованиям ... 186
10.3. Практическая реализация гардов ... 187

ЧАСТЬ IV. ПРАКТИЧЕСКИЕ ПРИМЕНЕНИЯ ... 199

Глава 11. Шаблоны для реальных задач ... 200
11.1. Агент поддержки клиентов – автоматизация первой линии ... 201
11.2. Медицинский агент – диагностика и поддержка пациентов ... 211
11.3. Финансовый аналитик – прогнозирование и управление рисками ... 227
11.4. Агент-программист – автоматизация разработки и сопровождения кода ... 242
11.5. Мультиязычный переводчик с культурной адаптацией ... 258

Глава 12. Примеры успешных реализаций ... 272

ЧАСТЬ V. БУДУЩЕЕ АГЕНТНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ... 281

Глава 13. Тренды и инновации ... 282
13.1. Нейросимвольные гибриды: соединение интуиции и логики ... 282
13.2. Автономность и самообучение: агенты, которые растут ... 286
13.3. Многоагентные экосистемы: от одиночек к цивилизациям ... 291
13.4. Агенты в метавселенных и виртуальных мирах ... 296
13.5. Регуляция и агенты-законодатели ... 298
13.6. Этика, безопасность и выравнивание ... 299

Глава 14. Видение будущего ... 303
14.1. Цифровое бессмертие через агентные системы ... 304
14.2. Симбиоз человека и агента: следующая ступень эволюции ... 309
14.3. Агенты как новая форма жизни: постбиологическая эволюция ... 315
14.4. Планетарные экосистемы агентов: цифровая ноосфера ... 322
14.5. Экзистенциальные риски и сценарии катастроф ... 326
Заключение ... 332
Глоссарий ... 336
Библиография ... 347

О чем расскажет эта книга

«Искусственный интеллект: современный агентный подход» – ваш путеводитель по новому тренду гипераутоматизации процессов, в рамках которого агенты становятся цифровыми помощниками, способными решать задачи без единой строчки кода. Эта книга раскрывает, как проектировать, внедрять и управлять интеллектуальными системами, которые автономно работают с данными, взаимодействуют с пользователями и оптимизируют бизнес-процессы. Вы узнаете, как создавать агентов на базе больших языковых моделей, используя визуальные конструкторы и бескодовые платформы, интегрировать их с внешними инструментами (от поиска в интернете, выполнения SQL-запросов до взаимодействия между агентами в рамках многоагентных систем) и настраивать под конкретные задачи – от поддержки клиентов и анализа финансовых отчётов до управления сложными информационными системами. Основное внимание уделяется практическим шаблонам: цепочкам действий, маршрутизации запросов, параллельному выполнению задач и автоматическому исправлению ошибок. Отдельные главы посвящены безопасности, тестированию и этическим аспектам, чтобы ваши агенты оставались надёжными и прозрачными. Книга написана для предпринимателей, руководителей и вообще любых энтузиастов и технооптимистов, которые хотят внедрить методы искусственного интеллекта в рабочие процессы без погружения в программирование. Она подойдёт тем, кто ищет готовые решения для автоматизации рутинных операций, и тем, кто мечтает создать умного помощника для бизнеса. С примерами из реальных кейсов, сравнением фреймворков и прогнозами на будущее это издание станет настольной книгой для всех, кто верит, что искусственный интеллект должен работать на человека – просто, эффективно и безопасно.

Об авторе

Роман Викторович Душкин – широко известный технологический евангелист и автор книги «Искусственный интеллект». Является генеральным директором аккредитованной ИТ-компании «А-Я эксперт», занимающейся разработкой систем искусственного интеллекта, старшим преподавателем кафедры 22 «Кибернетика» Национального исследовательского ядерного университета «МИФИ» и руководителем образовательной программы бакалавриата «Искусственный интеллект и моделирование когнитивных процессов в проектировании и разработке информационных систем». Роман Душкин создал первый на русском языке курс по генеративному искусственному интеллекту и промпт-инженерии, который стал основополагающим для многих специалистов в этой области. Его работы широко известны и востребованы среди тех, кто интересуется новейшими достижениями в сфере искусственного интеллекта и компьютерных технологий.

Введение

Технологии искусственного интеллекта развиваются с такой скоростью, что даже автору этой книги пришлось взяться за новый труд, едва завершив предыдущий (оригинал-макет книги «Генеративный искусственный интеллект» [Душкин, 2025] ушёл в типографию ровно тогда, когда я сел за эти строки). Впрочем, потом я его отложил на неопределённое время, так как реально скорость изменений была такова, что написанное становилось неактуальным уже через неделю. Поэтому внимание было переключено на более устоявшуюся тему – RAG-системы [Душкин, 2025b]. Впрочем, до этого я успел написать книгу, посвящённую метакогниции и её применению в промпт-инженерии [Душкин, 2025a]. Всё это стало более крепкой основой для того, чтобы перейти к работе над книгой об агентном подходе в искусственном интеллекте.

Мир стоит на пороге эры гиперавтоматизации, когда ключевую роль играют не алгоритмы-одиночки, а многоагентные системы, вдохновлённые самой природой [Darwish, 2024]. Представьте муравейник: тысячи крошечных существ без централизованного управления создают сложные сети туннелей, распределяют ресурсы и защищают свою колонию. Или пчелиный рой, в котором каждая особь, следуя простым правилам, участвует в построении идеальных шестиугольных ячеек. Эти биологические системы – не просто метафоры, а прототипы будущего глобального искусственного интеллекта. Их эффективность основана на коллективном, или *роевом*, *интеллекте*, в рамках которого отдельные агенты взаимодействуют, адаптируются и самоорганизуются без жёстких управляющих сигналов из центра [Shahzad, 2022].

Самоорганизация – ключевой принцип, который позволяет многоагентным системам достигать сложных целей без явного управления [Ye et al., 2017]. В природе она проявляется в самых разных формах: от стай птиц, синхронно меняющих направление полёта, до нейронных сетей мозга, в котором миллиарды клеток взаимодействуют, чтобы порождать мысли и эмоции. Эти процессы кажутся хаотичными, однако на самом деле они подчиняются определённым правилам, которые можно перенести в мир технологий. Но природа – не единственный учитель. Человечество давно создаёт аналогичные структуры: от корпораций с их отделами и рабочими процессами до онлайн-сообществ, в которых пользователи формируют контент через спонтанное сотрудничество. Именно эти принципы – *децентрализация, специализация и синергия* – легли в основу современных искусственных когнитивных агентов [Tallam, 2025]. Как муравьи координируются через феромоны, так цифровые агенты обмениваются данными через шлюзы и интерфейсы, а их «роли» определяются не кодом, а настройками в визуальном конструкторе. Сегодня, благодаря no-code платформам, создать такую систему может даже тот, кто никогда не написал и строчки программы, – достаточно понять, как превратить хаос взаимодействий в упорядоченный процесс [Sido & Emon, 2024]. И это лишь начало пути, на котором биология, технологии и бизнес сливаются в единый поток прогресса…

Но если природа и человеческие организации подсказывают нам, *как* строить многоагентные системы, то ответ на вопрос, *почему* они работают, кроется в устройстве самого разума. Ещё в 1986 году Марвин Мински в своей книге «Сообщество разума» выдвинул революционную идею: *мозг – это не единый, монолитный «компьютер», а сообщество тысяч простых агентов*, каждый из которых решает свою микроскопическую задачу [Минский, 2018]. Нейроны, синапсы, отделы мозга – всё это «агенты», взаимодействующие через электрические импульсы и химические сигналы. Но даже на более высоком уровне супервентности – наши мысли, казалось бы цельные, на поверку оказываются результатом конкуренции и сотрудничества этих крошечных «агентов мыслительной деятельности». Современные исследования подтверждают: память, креативность, принятие решений – всё это возникает из динамики взаимодействия агентов, будь то клетки, нейронные ансамбли или ментальные состояния [Levin, 2019].

Эта концепция стала фундаментом для одного из подходов к искусственному интеллекту. Можно предположить, что ранние попытки создать искусственный разум проваливались именно потому, что учёные и инженеры пытались втиснуть интеллект в единый алгоритм, пусть даже и основанный на машинном обучении. Но сегодня, с развитием больших языковых моделей, мы наконец можем имитировать «сообщество разума». Представьте: один агент анализирует данные, другой генерирует гипотезы, третий проверяет их на ошибки – и всё это без жёсткого кода, через диалог на естественном языке. Большие языковые модели стали тем универсальным «клеем», который позволяет агентам – как цифровым, так и биологическим – находить общий язык [Gao et al., 2024]. И если раньше формализация таких взаимодействий упёрлась в тупик (как попытки описать человеческую речь грамматиками), то теперь большие языковые модели и агенты на их основе сами учатся правилам коммуникации, впитывая их из миллионов диалогов [Szabo & Teo, 2015]. Мы стоим на пороге эры, когда многоагентные системы – уже не абстракция, а рабочий инструмент, доступный каждому через текстовые и графические интерфейсы. И это не просто технологический скачок – это возвращение к истокам, к тому, как устроена сама жизнь.

Идея о том, что разум – это не монолит, а сложная система взаимодействующих агентов, нашла отражение в трудах многих современных исследователей. Например, Джефф Хокинс в своей книге «Об интеллекте» [Хокинс, 2007] предложил модель, в которой мозг работает как иерархия предсказательных механизмов. Каждый уровень этой иерархии – это своего рода агент, который обрабатывает информацию, делает прогнозы и передаёт результаты на следующий уровень. Хокинс подчёркивает, что именно взаимодействие между этими уровнями позволяет нам воспринимать мир, учиться и принимать решения. Его теория предсказательного кодирования стала важным шагом в понимании того, как множество простых процессов могут складываться в сложное поведение.

Рэй Курцвейл в книге «Эволюция разума» [Курцвейл, 2015] пошёл ещё дальше, утверждая, что человеческий разум – это результат эволюции множества специализированных модулей, каждый из которых отвечает за определённую функцию. Эти модули, по сути, являются агентами, которые обмениваются информацией и конкурируют за ресурсы. Курцвейл также предсказал, что будущее искусственного интеллекта будет связано с созданием аналогичных систем, в которых множество узкоспециализированных агентов будут работать

вместе, чтобы решать сложные задачи. Его идеи о «сингулярности» и слиянии человеческого и машинного интеллектов во многом основаны на этой многоагентной парадигме.

Косвенно эту тему затрагивали и другие исследователи. Например, в работах Дэвида Иглмана [Иглмен, 2019] подчёркивается, что наше сознание – это результат конкуренции и сотрудничества множества «нейронных ансамблей». А в трудах Дэниела Деннета [Деннет, 2021] разум рассматривается как продукт эволюции, когда каждый новый уровень сложности возникает благодаря взаимодействию более простых элементов.

Эти идеи не только вдохновляют, но и дают практическую основу для разработки современных систем искусственного интеллекта. Если мозг – это сообщество агентов, то почему бы не построить искусственную интеллектуальную систему по тому же принципу? Сегодня, благодаря большим языковым моделям, мы можем создавать агентов, которые не просто выполняют задачи, но и учатся взаимодействовать друг с другом, адаптироваться к новым условиям и даже «договариваться» о распределении ресурсов [Hua et al., 2024]. Это открывает новые горизонты для гипераватоматизации, где каждый агент становится частью глобальной сети, способной решать задачи, которые раньше казались недостижимыми.

История искусственного интеллекта – это в том числе и история многоагентных систем, которые всегда стояли на фронтире научного поиска [Тарасов, 2002]. С первых дней развития искусственного интеллекта исследователи как будто бы интуитивно чувствовали, что сложные задачи требуют распределённого подхода. В 1980-х годах, ещё до бума искусственных нейронных сетей, учёные экспериментировали с распределённым искусственным интеллектом, в рамках которого агенты-«эксперты» решали узкие задачи, а затем координировали результаты [Bond & Gasser, 2014]. Например, проект «Blackboard System» имитировал команду специалистов, работающих над общей проблемой, как учёные у доски [Nii, 1986]. В 1990-х, с расцветом интернета, появились первые многоагентные платформы для электронной коммерции, на которых автономные боты торговались за ресурсы, словно участники биржи [Mazumdar & Roy, 2021]. Даже робототехника шла по этому пути: рои дронов, способных строить карты местности или тушить пожары, стали наглядным воплощением принципов роевого интеллекта [Alsammak et al., 2023].

Не остался в стороне и советский научный ландшафт. Ещё в 1960-х годах, когда кибернетика в СССР переживала период реабилитации, учёные задумывались о системах, напоминающих современные многоагентные подходы. Проект ОГАС – общегосударственная автоматизированная система управления экономикой – под руководством Виктора Михайловича Глушкова предполагал создание распределённой сети вычислительных центров, которые координировали бы данные от предприятий, министерств и регионов [Глушков & Валах, 1981]. По сути, это была попытка построить «агентов» для оптимизации плановой экономики. В Институте кибернетики АН УССР разрабатывались модели коллективного поведения роботов, способных совместно решать задачи в изменяющейся среде. А работы Николая Михайловича Амосова по моделированию мозга и социальных систем предвосхитили идеи о том, что интеллект рождается из взаимодействия простых элементов [Амосов, 1965]. Однако

многие из этих идей, опередивших своё время, так и остались в стадии экспериментов – отчасти из-за технологических ограничений эпохи, отчасти из-за идеологического недоверия к «децентрализованным» системам. Тем не менее они стали важным этапом в эволюции многоагентного подхода, доказав, что даже в условиях жёсткого централизованного управления наука стремилась к моделям, имитирующим природную самоорганизацию.

Парадоксально, но попытки создать «идеальный язык» для общения агентов повторяли ошибки лингвистов, столкнувшихся с естественной речью. В 1970-х, когда многоагентные системы только зарождались, инженеры пытались формализовать коммуникацию агентов через строгие протоколы вроде KQML (Knowledge Query and Manipulation Language) – аналога формальных грамматик для обработки естественного языка [Labrou & Finin, 1994]. Каждое сообщение между агентами должно было соответствовать шаблону: тип действия (запрос, утверждение, команда), отправитель, получатель, контент в виде логических формул. Это напоминало попытки описать человеческий диалог правилами типа «подлежащее + сказуемое + дополнение» – система работала в лаборатории, но разваливалась в реальном мире, когда контекст, ирония и недоговорённости делали между людьми общение «живым».

Яркий пример – проект FIPA (Foundation for Intelligent Physical Agents) 1990-х. Разработанные им стандарты коммуникации напоминали юридические документы: агенты обменивались сообщениями в формате «перформативов» («запрос», «согласие», «отказ»), в которых каждое слово имело чёткий семантический вес [Pitt & Mamdani, 1999]. Но как только система сталкивалась с нестандартной ситуацией (например, агент-логист получал противоречивые запросы от двух других агентов), она заходила в тупик, требуя вмешательства человека. Это напоминало ранние чат-боты вроде ELIZA, которые имитировали психотерапевта, но пасовали перед простым вопросом «Что ты думаешь о закате?» [King et al., 2025].

Прорыв произошёл, когда учёные и инженеры осознали: *естественный язык – не баг, а фича*. Вместо того чтобы навязывать агентам искусственные протоколы, стоит позволить им общаться так, как это делают люди, – свободно, контекстно, с долей неопределённости. Большие языковые модели стали «переводчиками», которые превращают неструктурированные запросы в действия. Сегодня агент может получить инструкцию «Найди самые рискованные акции за последнюю неделю и предложи стратегию хеджирования» – и, не требуя формальных спецификаций, разбить задачу на этапы, делегируя части запроса другим агентам через обычный текст.

Тупик формализации обернулся новыми возможностями. То, что когда-то считалось слабостью естественного языка, – гибкость, многозначность, контекстная зависимость – стало основой для синергии между людьми и системами искусственного интеллекта. Агенты теперь не просто исполняют код, а ведут переговоры, учатся на ошибках и даже проявляют подобие креативности – всё потому, что их «общение» больше не ограничено жёсткими рамками, а напоминает живой диалог, в котором каждая фраза открывает новые горизонты для сотрудничества.

Появление больших языковых моделей стало поворотным моментом, который перевернул представление о возможностях многоагентных систем. Эти

модели превратились в универсальных «коммуникаторов», способных не только понимать контекст, но и генерировать осмысленные действия [Becker, 2024]. Сегодня большие языковые модели позволяют агентам говорить на языке людей, а не машин, что стирает границы между техническими специалистами и конечными пользователями. Представьте врача, который формулирует запрос на естественном языке: «Проанализируй последние исследования по лечению диабета 2-го типа и подготовь персонализированные рекомендации для пациента 55 лет». Система, состоящая из агентов-аналитиков, агентов-синтезаторов и агентов-валидаторов, распределяет задачу, находит релевантные данные, сравнивает методики и формирует вывод – всё это без единой строчки кода, через цепочку диалогов.

Сфера применения таких систем безгранична. В образовании агенты становятся персональными тьюторами: один адаптирует материал под уровень ученика, другой проверяет задания, третий осуществляет мониторинг прогресса и корректирует программу. В финансах многоагентные системы прогнозируют риски, оптимизируют инвестиционные портфели и автоматизируют переговоры с контрагентами – представьте агентов-трейдеров, которые на естественном языке обсуждают стратегии, как это делают трейдеры-люди. Даже в креативных индустриях большие языковые модели открывают новые горизонты: рекламные агентства используют «команды» агентов для генерации идей, где один отвечает за слоганы, другой за визуальные метафоры, а третий анализирует целевую аудиторию.

Ключевая сила больших языковых моделей заключается в их способности стать мозгом для каждого агента, наделив его не только навыками решения задач, но и «социальным интеллектом». Агенты теперь умеют договариваться, идти на компромиссы и даже проявлять эмпатию. Например, в сервисах поддержки клиентов искусственный когнитивный агент может распознать раздражение в тексте пользователя, делегировать запрос к более опытному «коллеге» и предложить компенсацию, сохраняя лояльность. Это стало возможным потому, что большие языковые модели впитывают не только факты, но и культурные паттерны, этические нормы и тонкости человеческих отношений [Pawar et al., 2024].

Но главное – большие языковые модели делают многоагентные системы доступными. No-code платформы вроде LangFlow или Voiceflow позволяют любому пользователю собрать команду агентов, как из конструктора LEGO [Qdrant, 2024]. Бизнес-аналитик без навыков программирования может создать систему для автоматизации отчётности, соединив агента-сборщика данных из CRM, агента-визуализатора и агента-редактора, который готовит презентации. Это максимально расширяет сферу применения технологии, превращая гиперавтоматизацию из удела корпораций в инструмент малого бизнеса, образования и даже бытового использования.

Большие языковые модели сегодня становятся не просто инструментом – они являются *метафорой нового этапа эволюции искусственного интеллекта*, когда интеллект возникает не из одиночных алгоритмов, а из синергии множества агентов, общающихся на естественном языке, понятном человеку. И это только начало: с каждым улучшением больших языковых моделей границы возможного раздвигаются, открывая эру, в которой технологии служат

не заменой людям, их естественному языку и коллективному разуму, а их логическим продолжением.

Однако стремительный прогресс многоагентных систем на базе больших языковых моделей имеет и обратную сторону – языковое неравенство. По данным исследования OpenAI, более 92 % данных для обучения таких моделей, включая GPT-4, составляют англоязычные тексты [OpenAI et al., 2024]. Это создаёт петлю положительной обратной связи: чем лучше модели «понимают» английский, тем активнее учёные используют их для генерации статей, патентов и документации на том же языке, что ещё больше увеличивает долю англоязычного контента в обучающих выборках. Например, система Galactica, заявленная как «модель для науки», при запуске в 2022 году демонстрировала на английском языке способность писать академические статьи, но на русском или китайском выдавала крайне поверхностные или ошибочные выводы [Muennighoff et al., 2022].

Этот дисбаланс превращает английский язык в лингва-франка искусственного интеллекта, что имеет далеко идущие последствия. Учёные, пишущие на других языках, сталкиваются с двойной нагрузкой: им приходится либо тратить ресурсы на адаптацию англоязычных моделей (что требует глубоких технических навыков), либо использовать зарубежные сервисы, которые игнорируют локальный контекст. Так, например, в 2023 году исследователи из Сингапура провели эксперимент: они попросили модель GPT-4 составить план борьбы с диабетом для азиатских стран. Модель предложила диету с овсянкой и авокадо – продукты, малодоступные в регионе, – проигнорировав местные аналоги вроде риса или кокосового молока [Menezes et al., 2025]. Это показывает, как культурно-языковая предвзятость больших языковых моделей может искажать решения даже в критических сферах.

Ситуация усугубляется тем, что научная коммуникация всё чаще перекладывается на системы искусственного интеллекта. Агенты-ассистенты вроде Elicit или Scite анализируют англоязычные статьи, генерируют обзоры и даже предлагают гипотезы – но для русскоязычных исследований эти инструменты остаются слепыми [Preisler, 2023]. В итоге учёные из неанглоязычных стран вынуждены тратить до 40 % времени на перевод и адаптацию чужих результатов вместо инноваций, что замедляет локальные технологические экосистемы. Как отмечается многими современными лингвистами, «язык – это не просто оболочка для идей, а матрица, формирующая само мышление» [Бородай, 2019]. Если системы искусственного интеллекта не научатся работать с разноязычными текстами наравне с английскими, мы рискуем получить мир, в котором «второстепенные» языки станут цифровыми резервациями, а их носители – пассивными потребителями чужих инноваций. Это то, что можно назвать языковым *ИИ-неоколониализмом*.

Но выход есть. Проекты вроде российских моделей YandexGPT и GigaChat или китайских моделей Qwen, DeepSeek, Ernie показывают, что развитие больших языковых моделей на национальных языках возможно – хотя и требует колоссальных ресурсов (а также здесь можно отметить и тот нюанс, что названия всех перечисленных моделей пишутся на латинице). Ключом становится создание открытых многоязычных наборов данных и законодательная поддержка. Например, в Европейском союзе программа Digital Europe

выделяет 800 млн евро на развитие искусственного интеллекта, включая модели для малых языков [Ruohonen & Timmers, 2025]. В России инициативы вроде Национального корпуса русского языка могли бы стать основой для обучения агентов, понимающих не только русский, но и региональные языки [Минеева, 2015]. Это вопрос не только технологий, но и цифрового суверенитета: многоагентные системы будущего должны отражать плюрализм культур, а не монополию одного языка.

Языковое неравенство в искусственном интеллекте – это не техническая проблема, а вызов глобального масштаба. Преодолеть его можно только через кооперацию: сообщества открытого исходного кода, государственные программы и транснациональные альянсы, которые превратят многоязычие из барьера в преимущество. Как когда-то интернет объединил мир, так и многоагентные системы на основе больших языковых моделей могут стать мостом между языками – но для этого им нужно научиться говорить на всех наречиях человечества, а не только на одном.

Моя мотивация как автора этой книги коренится в глубоком убеждении: *технологический суверенитет начинается с языка*. Когда знания и инструменты существуют только на английском, целые поколения специалистов оказываются в зависимом положении – они вынуждены прикладывать дополнительные усилия на изучение и использование чужого языка с чуждыми культурными конструктами вместо освоения сути технологий. А ещё можно вспомнить, как в 2022 году компания OpenAI заблокировала доступ к GPT для российских пользователей, а GitHub ввёл санкционные ограничения. Это сделали и тысячи других технологических компаний Запада. Эти шаги показали: надеяться на «нейтральность» зарубежных платформ наивно. Как Китай создаёт «великий файрвол» для данных, а США ограничивают экспорт чипов для искусственного интеллекта, так и России нужны собственные образовательные экосистемы – от школьных учебников до продвинутых курсов по большим языковым моделям и промпт-инженерии.

Пример Китая здесь показателен. В 2025 году правительство Поднебесной запретило ведущим исследователям в области искусственного интеллекта из компании DeepSeek покидать страну без разрешения, а компании Tencent и Alibaba обязали делиться алгоритмами с государством. Это не просто протекционизм – это осознанная стратегия построения цифровой автаркии. Но Китай может себе это позволить: объём публикаций на китайском языке по искусственному интеллекту вырос на 300 % за последние 5 лет, а модели вроде уже упомянутых Ernie или Qwen конкурируют с GPT-4. В России же, по данным РАН, лишь 17 % статей по искусственному интеллекту публикуются на русском языке – остальные уходят в англоязычные журналы, усиливая технологическую неоколониальную зависимость. К тому же это вообще является последствиями требований к университетам – публикация в западных журналах и наращивание наукометрических показателей по данным западных баз данных.

Эту пропасть можно преодолеть только через массовую просвещённость. Когда я запускал первый (первый в мире) русскоязычный курс по генеративному искусственному интеллекту в начале 2023 года, 80 % моих слушателей признавались: даже базовые концепции вроде файн-тюнинга или промпт-инженерии они разбирали через переводчики и кустарные справочники. Сегодня, благода-

ря проектам вроде AI 360 или Школы анализа данных Яндекса, ситуация меняется – но этого недостаточно. Каждая книга, статья или видеоурок на русском – это кирпичик в фундаменте технологической независимости. Как в 1930-х СССР создавал индустриализацию через ликбезы и рабфаки, так сегодня нам нужен новый ликбез эпохи искусственного интеллекта – доступный каждому, кто хочет строить будущее, а не бездумно потреблять чужие технологии.

Эта книга – часть такой миссии. В ней вы не найдёте слепого копирования западных подходов. Вместо этого – анализ, адаптированный к реалиям русского мира: как работать с данными на кириллице и как обучать агентов для локальных рынков. Моя цель – чтобы читатель, будь то студент из Тамбова или предприниматель из Твери, мог создать конкурентоспособную систему, не становясь «цифровым эмигрантом» в англоязычной экосистеме.

Технологии – новая территория геополитики. И если мы хотим, чтобы на этой территории звучала русская речь, нужны не только ответы на налагаемые на нас санкции, но и созидание. Каждый специалист, обученный на родном языке, каждый проект, реализованный на отечественных платформах, – это шаг к миру, в котором искусственный интеллект служит не интересам отдельных корпораций и государств, возомнивших себя гегемонами, а многообразию глобальных культур. И я верю, что эта книга станет одним из таких шагов.

Вместе с тем стоит отметить, что обывательское понимание такой новой дисциплины, как промпт-инженерия, является крайне поверхностным. Промпт-инженерия – это гораздо больше, чем искусство «правильно спрашивать у ChatGPT». Это *новая парадигма проектирования информационных систем*, при применении которой естественный язык становится полноценным инструментом программного инженера. Представьте, что вы не пишете код на языке программирования, а описываете архитектуру многоагентной системы в диалоге с большой языковой моделью: «Создай трёх агентов – аналитика данных, проверяющего на аномалии, генератора отчётов и оптимизатора рабочих процессов. Пусть они обмениваются результатами каждые 10 минут, а при критических ошибках запрашивают человеческое вмешательство». Большая языковая модель превращает это описание в работающую систему, словно переводя мысли в исполняемую логику.

Такая методология меняет сам подход к разработке. Традиционное программирование требует точности синтаксиса, а промпт-инженерия – *точности семантики*. Например, промпт для агента-модератора контента может выглядеть как многоуровневая инструкция:

1. Анализируй комментарии на токсичность по шкале от 1 до 5.
2. Если оценка ⩾ 4 – делегируй решение агенту-этику.
3. При спорных случаях (3 балла) спрашивай второе мнение у агента-психолога.

Здесь каждый пункт – не просто запрос, а *архитектурное решение*, определяющее роли, потоки данных и эскалацию проблем. Как инженер-строитель чертит схемы коммуникаций между этажами здания, так и промпт-инженер проектирует диалоговые сценарии между агентами.

Это особенно критично для многоагентных систем, в которых сложность растёт экспоненциально. В 2023 году исследователи из Stanford CRFM показали, что добавление всего одного агента-арбитра, управляемого промптами,

снижает количество конфликтов в системе примерно на 40 % [Wornow et al., 2023]. Например, в логистической системе промпт вида: «Разрешай споры между агентами доставки и склада по принципу “клиент всегда прав”, но учитывай задержки из-за форс-мажоров» – становится не абстракцией, а рабочим механизмом балансировки интересов.

Но сила промпт-инженерии – не только в гибкости. Это мост между предметными экспертами и системами искусственного интеллекта. Врач, педагог или юрист, не обладая навыками программирования, могут через текстовые инструкции «настроить» команду агентов под свои нужды. Представьте себе учителя биологии, который поручает своему искусственному когнитивному агенту-ассистенту: «Создай 4 агента: один генерирует тесты по генетике, второй проверяет ответы, третий адаптирует сложность под успеваемость класса, а четвёртый готовит визуализации ДНК» – и получает готовый учебный модуль.

Однако такая мощь требует тщательности. Промпт-инженерия – это не магия, а *дисциплина с чёткими принципами*:

1. *Контекстная вложенность.* Промпты должны учитывать предыдущие взаимодействия агентов.
2. *Безопасность.* Ограничения вроде «никогда не предлагай медицинские рекомендации без проверки лицензированным врачом».
3. *Адаптивность.* Правила вроде «если агент трижды ошибся – передай управление человеку».

Эти принципы уже воплощаются в инструментах вроде Microsoft Guidance или LangChain, в которых промпты становятся узлами в графах автоматизируемых рабочих процессов. Но главное – они стирают границу между «технарями» и «гуманитариями», позволяя каждому стать архитектором систем искусственного интеллекта. В этом и есть суть технологического суверенитета: когда создавать агентов могут не только избранные, а все, кто мыслит ясно и системно – хоть на русском, хоть на суахили.

Технологии искусственного интеллекта развиваются с такой скоростью, что даже в процессе написания этой книги я сталкиваюсь с парадоксом: текст, который я пишу сегодня, завтра может оказаться устаревшим. Каждую неделю появляются новые модели, фреймворки и подходы, которые переворачивают представления о возможностях искусственного интеллекта. Например, пока я планировал эту книгу в начале марта 2025 года, компания OpenAI выпустила модель GPT-4.5, а компания Google представила новую версию своей модели Gemini, способной одновременно обрабатывать текст, аудио и видео. Пока я писал введение, компания Baidu выпустила следующую версию своей модели Ernie. Эти изменения настолько стремительны, что даже самые актуальные примеры и варианты использования рискуют стать историей ещё до выхода книги.

Но в этом и заключается ценность фиксации знаний. Даже если какие-то детали устареют, фундаментальные принципы – будь то архитектура многоагентных систем, основы промпт-инженерии или подходы к языковому суверенитету – останутся актуальными. Эта книга не претендует на то, чтобы быть энциклопедией всех технологий, но она станет точкой отсчёта, которая поможет читателям понять, как сегодня устроены технологии искусственного интеллекта, а потому позволит подготовиться к его завтрашним изменениям.

Это как будто бы карта, которая не успевает за ландшафтом, но всё равно указывает верное направление.

И если вы, читая эти строки, заметите, что какие-то технологии уже ушли вперёд, – не расстраивайтесь. Это не недостаток книги, а её особенность. Ведь главное – не запомнить каждый факт, а научиться мыслить в терминах искусственного интеллекта, промпт-инженерии и многоагентных систем, чтобы самостоятельно адаптироваться к новым вызовам. И я надеюсь, что эта книга станет для вас не просто источником знаний, а компасом в мире, в котором ускоряющиеся изменения – единственная константа.

Сегодня, когда большие языковые модели позволяют агентам общаться на естественном языке, многоагентные системы переживают ренессанс. Они уже становятся базой для автономных научных открытий, когда агенты-исследователи генерируют гипотезы, планируют эксперименты и анализируют данные [Wang et al., 2023]. Они управляют умными городами, балансируя энергопотребление и транспортные потоки. Они даже проникают в креативные индустрии – представьте сценарий, написанный «командой» агентов: один отвечает за диалоги, другой – за сюжетные повороты, третий – за эмоциональную глубину. Кажется, будто вся эволюция искусственного интеллекта вела нас к этому моменту, когда вместо гонки за «универсальным алгоритмом» мы собираем интеллект из кирпичиков-агентов, как когда-то природа собирала человеческий разум из нейронов. И если раньше это было уделом лабораторий, то сегодня no-code инструменты превращают создание таких систем в искусство, доступное каждому, кто готов мыслить как архитектор коллективного разума.

Эта книга – не просто сборник знаний, а стартовая площадка для вашего путешествия в широчайшую область современного искусственного интеллекта. В эпоху, когда технологии меняются быстрее, чем мы успеваем их осмыслить, важно не просто идти в ногу со временем, а опережать его. Здесь вы найдёте не только ответы на сегодняшние вопросы, но и инструменты, которые помогут вам задавать правильные вопросы завтрашнего дня. Пусть она станет вашим проводником при разработке многоагентных систем, использовании промпт-инженерии и гиперавтоматизации в новом дивном мире, в котором каждый, кто готов учиться и мыслить системно, может стать творцом будущего. Ведь искусственный интеллект – это не только технологии, это возможность для каждого из нас переосмыслить, как мы работаем, учимся и создаём. И первый шаг к этому будущему – у вас в руках.

Выражаю благодарность Еремихиной Марии Глебовне за неоценимую помощь с иллюстрациями.

Р. В. Душкин,
Серпухов – Москва,
2026 год

Конец ознакомительного фрагмента.
Приобрести книгу можно
в интернет-магазине
«Электронный универс»
e-Univers.ru