

Иван Дискин

**GEO: как попасть в ответы
искусственного интеллекта. Новая
эпоха поиска, где бренды
рекомендуют нейросети**

Иван ДИСКИН

**ГЕО: как попасть в ответы
искусственного интеллекта.
Новая эпоха поиска, где
бренды рекомендуют нейросети**

«Автор»

2026

Дискин И.

GEO: как попасть в ответы искусственного интеллекта. Новая эпоха поиска, где бренды рекомендуют нейросети / И. Дискин — «Автор», 2026

Поиск меняется: пользователь всё чаще получает не список ссылок, а готовый ответ от ChatGPT, Gemini, Perplexity и других ИИ-систем. В новой реальности брендам недостаточно занимать позиции в Google - нужно становиться источниками, которым доверяют нейросети. Эта книга объясняет, как работает Generative Engine Optimization (GEO), почему ИИ выбирает одни компании и игнорирует другие, как создавать контент, который попадает в ответы, выстраивать цифровую репутацию и измерять видимость бренда в эпоху AI Search. Внутри - практическая методология, чек-листы, метрики, примеры, дорожная карта на 90 дней и рабочие шаблоны для маркетологов, SEO-специалистов, руководителей и владельцев бизнеса.

© Дискин И., 2026

© Автор, 2026

Содержание

Выходные данные	6
Оглавление	7
Глава 1. Конец эпохи десяти синих ссылок	8
Глава 2. GEO vs SEO vs AEO – в чём разница на самом деле	11
Глава 4. Кому и зачем это нужно уже сейчас	18
Глава 5. RAG, эмбединги и ранжирование источников – простыми словами, но без лжи	22
Глава 6. Что такое «цитируемость» и как ИИ выбирает, кого упомянуть	25
Конец ознакомительного фрагмента.	26

Иван Дискин

GEO: как попасть в ответы искусственного интеллекта. Новая эпоха поиска, где бренды рекомендуют нейросети

GEO

Полное практическое руководство по оптимизации для ИИ-поиска

Как сделать так, чтобы ChatGPT, Gemini, Perplexity, YandexGPT и другие ИИ-системы рекомендовали именно ваш бренд. Теория с историей, практика с кейсами.

Иван Дискин

Эксперт в области SEO, GEO и интернет-маркетинга

В маркетинге с 2011 года

Москва

2026

Выходные данные

GEO. Полное практическое руководство по оптимизации для ИИ-поиска. Как сделать так, чтобы ChatGPT, Gemini, Perplexity, YandexGPT и другие ИИ-системы рекомендовали именно ваш бренд. Теория с историей, практика с кейсами.

© Иван Дискин, 2026

Все права защищены.

Никакая часть настоящего издания не может быть воспроизведена, распространена, переведена, размещена в информационных системах или передана в любой форме и любыми средствами – электронными, механическими, фотокопировальными, записывающими или иными – без предварительного письменного разрешения правообладателя, за исключением случаев, предусмотренных законодательством об авторском праве.

Настоящая книга носит исключительно информационный характер. Автор не гарантирует достижения конкретных коммерческих, маркетинговых или финансовых результатов при использовании изложенных подходов. Все примеры, оценки и рекомендации отражают состояние рынка на момент подготовки книги и могут изменяться по мере развития технологий и поисковых систем.

При подготовке книги использовались открытые исследования, официальная документация, материалы компаний и собственный практический опыт автора. Полный перечень источников приведён в разделе «Список источников по главам».

Автор: Дискин Иван Сергеевич

Сайт: ivandiskin.ru

Проект AI Presence: ai-presence.ru

Энциклопедия MarketWiki: marketwiki.ru

Москва, 2026 год

Оглавление

Выходные данные	
Оглавление	
Глава 1. Конец эпохи десяти синих ссылок	
Глава 2. GEO vs SEO vs AEO – в чём разница на самом деле	
Глава 3. Как устроены ChatGPT, Perplexity, Gemini, Copilot, YandexGPT: откуда берутся	
ответы и цитаты	
Глава 4. Кому и зачем это нужно уже сейчас	
Глава 5. RAG, эмбединги и ранжирование источников – простыми словами, но без лжи	
Глава 6. Что такое «цитируемость» и как ИИ выбирает, кого упомянуть	
Глава 7. Роль E-E-A-T, авторства и цифрового следа в эпоху ИИ	
Глава 8. Кейс: как я выстраивал видимость личного бренда практически с нуля	
Глава 9. Структура контента, «цитируемая» по умолчанию	
Глава 10. Техническое GEO: schema.org, llms.txt, доступность для краулеров ИИ	
Глава 11. Работа с внешним контуром: PR, Habr, Cossa, отраслевые площадки, форумы, Reddit/Пикабу и аналоги	
Глава 12. E-commerce и высококонкурентные ниши: кейс GEO-пилота в жёстко регулируемой категории	
Глава 13. AEO в высококонкурентных нишах: кейс захвата нулевой выдачи в высококонкурентной нише	
Глава 14. Как мониторить упоминания бренда в ИИ-поиске	
Глава 15. Метрики GEO: что считать вместо кликов и позиций	
Глава 16. Обзор инструментов рынка и их ограничения	
Глава 17. Как объяснить SEO, зачем деньги в GEO, а не в контекстную рекламу	
Глава 18. GEO-стратегия под ключ: как выглядит дорожная карта на 3–4 недели	
Глава 19. Бренд или перформанс: стратегия маркетинга для роста	
Глава 20. Место GEO в оргструктуре: кто отвечает за это в компании	
Глава 21. Куда идёт рынок: агентный поиск, ИИ-браузеры, покупки через ассистентов	
Глава 22. Что будет дальше с классическим SEO	
Глава 23. Практический чек-лист и дорожная карта для читателя на 90 дней	
Заключительное слово	
Глоссарий терминов	
Приложение. Шаблоны и рабочие документы	
Шаблон 1. Чек-лист первичного GEO-аудита (первые 30 дней дорожной карты, главы 18 и 23)	7
Шаблон 2. Бриф на создание контента под answer-first структуру (глава 9)	
Шаблон 3. Библиотека промптов для мониторинга (глава 14)	
Шаблон 4. GEO-стратегия для бренда с ограниченным доступом к сайту	
Шаблон 5. Питч: зачем GEO-направление нужно параллельно существующему SEO/ORM-процессу	
Список источников по главам	

Глава 1. Конец эпохи десяти синих ссылок

Как ИИ-системы меняют потребление информации

Ещё десять лет назад «поиск» и «список из десяти синих ссылок» были синонимами. Вы вводили запрос, получали страницу результатов выдачи, кликали, читали, возвращались за следующим кликом. Вся экономика интернета – реклама, SEO, контент-маркетинг – в значительной степени строилась вокруг клика как основной единицы привлечения пользователя. Сегодня значение этой метрики стремительно снижается.

По оценкам Similarweb и других аналитических компаний, к началу 2026 года ChatGPT обрабатывал сотни миллионов поисковых запросов в неделю и входил в первую десятку самых посещаемых сайтов планеты. При этом темпы распространения ChatGPT были беспрецедентно быстрыми. По крайней мере, если сравнивать с классическими поисковыми продуктами вроде Bing. По публичным заявлениям OpenAI, к началу 2026 года ChatGPT использовали сотни миллионов пользователей еженедельно, а количество обращений к сервису измерялось миллиардами в месяц.

Но дело не в размере аудитории ИИ-сервисов самой по себе. По-настоящему важно то, что происходит с запросом дальше – и здесь имеет место настоящий разрыв с привычной моделью поиска.

Zero-click: поиск без клика как новая норма

Термин «zero-click search» (поиск без перехода на сайт) появился задолго до бума генеративного ИИ. В 2019 году Рэнд Фишкин (SparkToro) опубликовал исследование, показавшее, что чуть больше половины поисковых сессий Google заканчиваются без перехода на внешний сайт. Тогда это объясняли блоками с прямыми ответами, картами и вставками из Википедии.

К 2026 году ситуация качественно другая. По совокупности данных нескольких исследовательских компаний:

Поисковые сессии в ИИ-продуктах – ChatGPT Search, Perplexity, Google AI Mode и других – в зависимости от платформы и типа запроса завершаются без перехода на сайт примерно в 60–90% случаев.

В самом Google по состоянию на 2026 год без клика заканчивается порядка 60% всех поисковых сессий, а появление блоков AI Overview в выдаче может снижать CTR первой органической позиции на десятки процентов, а в отдельных исследованиях – почти вдвое.

Компания Bain & Company в своём исследовании потребительского поведения отмечала, что большинство пользователей уже используют ИИ-ответы как минимум для части поисковых задач. Одновременно исследования Similarweb, Ahrefs и других аналитических компаний фиксируют заметное снижение органического трафика в ряде отраслей.

Формулируя проще: пользователь всё чаще получает ответ, не покидая интерфейс ИИ-ассистента. Ваш сайт, ваша статья, ваш продукт – участвуют в формировании этого ответа, но не обязательно получают переход. Клик перестаёт быть единственным доказательством того, что контент сработал.

Это не значит, что трафика с ИИ-систем нет вообще. Он есть, и он растёт кратно: по разным оценкам, объём переходов с ИИ-платформ за 2024–2026 годы вырос на порядок, а в отдельных выборках – в сотни раз в зависимости от методологии подсчёта и охваченных площадок. Но в абсолютных цифрах доля такого трафика для большинства сайтов пока измеряется десятками долями процента. Показательно, что при этом такой трафик нередко демонстрирует более высокую вовлечённость и конверсию, чем трафик из классического поиска.

Вывод для практика простой и неудобный одновременно: расти в объёме, оставаясь незамеченным для пользователя в момент принятия решения, – вполне реальный сценарий на бли-

жайшие годы. Вопрос не «получит ли мой сайт клик», а «упомянут ли меня вообще, когда ИИ формирует ответ».

Кто уже здесь: карта ИИ-поисковых систем

Рынок ИИ-поиска в 2026 году – это быстро меняющаяся конкурентная среда:

ChatGPT остаётся крупнейшей платформой генеративного ИИ и, по различным оценкам, контролирует около 65–80% рынка ИИ-чат-ботов. По мере роста конкурентов его относительная доля постепенно сокращается, однако мобильное приложение по-прежнему собирает аудиторию, в несколько раз превышающую показатели ближайших конкурентов.

Gemini от Google демонстрирует один из самых высоких темпов роста среди крупных ИИ-платформ и по ряду исследований уже входит в число крупнейших источников реферального трафика на внешние сайты.

Perplexity остаётся заметным игроком с десятками миллионов активных пользователей и более чем миллиардом запросов в месяц, хотя темпы её роста уступают крупнейшим экосистемам.

Claude входит в число самых быстрорастущих крупных ИИ-платформ, однако по масштабу аудитории и объёму направляемого на сайты трафика пока остаётся позади лидеров.

Отдельно стоит Google AI Overviews – блоки с ИИ-сгенерированными ответами внутри классической выдачи Google, которые к 2026 году показываются на значительной части информационных и всё чаще коммерческих запросов, охватывая более 1,5 миллиарда пользователей ежемесячно.

Для российского рынка к этой карте добавляется YandexGPT и генеративные функции Яндекса – с собственной логикой цитирования и собственным набором источников доверия, которые мы разберём отдельно в части III.

Важно понимать: это не «ещё один канал» в списке уже существующих. Это новый посредник между вашим брендом и клиентом – посредник, который извлекает, суммирует и выбирает, какие источники использовать при формировании ответа, вместо того чтобы просто отранжировать ссылки.

Кто уже теряет, а кто выигрывает

Демографический сдвиг здесь тоже показателен. По данным нескольких исследований потребительского поведения, около трети представителей поколения Z в США регулярно используют ИИ-чат-боты для поиска информации. Это не нишевая аудитория «ранних последователей». Перед рынком – поколение, которое через 5–10 лет станет основной покупательской силой. А если посмотреть на миллениалов и зумеров в сумме, более 70% представителей этих поколений хотя бы периодически используют ИИ-инструменты для решения поисковых задач.

При этом цитирования в ответах ИИ-систем распределены крайне неравномерно. По данным BrightEdge, Ahrefs и других аналитических компаний, значительная часть ссылок и цитирований в ответах ИИ приходится на сравнительно небольшой круг авторитетных доменов – крупных СМИ, Википедию, Reddit, Stack Overflow и профильные отраслевые ресурсы. Контент без выраженной тематической авторитетности, понятной структуры и прямых, самодостаточных ответов на вопросы систематически не попадает в поле зрения ИИ – даже если прекрасно ранжируется в классическом Google.

Это ключевая вещь, которую должен понять каждый, кто дочитал до этого места: чтобы попасть в ответ ИИ, часто уже недостаточно того, чего раньше хватало для достижения высоких позиций в поисковых системах. Правила игры не эволюционировали, они частично переписаны заново. Этому и посвящена оставшаяся часть книги.

Почему это не хайп и не пройдёт стороной

У любого практика, прожившего несколько циклов алгоритмических обновлений Google, есть здоровый рефлекс: «подождём, перемелется». Но у нынешнего сдвига есть три отличия от привычных обновлений SEO-алгоритмов:

Меняется не алгоритм, а сам интерфейс потребления информации. Пользователь всё чаще получает готовый ответ до того, как видит список ссылок. Это не косметическое изменение выдачи, а смена интерфейса взаимодействия с интернетом как таковым.

Скорость адаптации бизнеса не успевает за скоростью роста аудитории. Опросы маркетологов показывают, что лишь меньшинство специалистов и компаний системно готовит контент под ИИ-поиск и цитирование. Подавляющее большинство этого не делает, то есть они либо не в курсе масштаба изменений, либо откладывают решение. Тем самым создаётся временное окно возможностей для тех, кто начнёт раньше.

Несмотря на различия в бизнес-моделях, крупные игроки рынка (OpenAI, Google, Microsoft, Perplexity, Anthropic) инвестируют в развитие ИИ-интерфейсов, внутри которых пользователь всё чаще решает задачи. Для крупнейших технологических компаний это многомиллиардное направление инвестиций. Это структурный, а не временный тренд.

Дальше в книге мы разберём механику того, как именно ИИ-системы выбирают источники для ответа (часть II), как выстроить контент и техническую инфраструктуру под эту механику (часть III), как измерять результат в условиях, когда клика может не быть в конкретном сценарии взаимодействия (часть IV), и как встроить всё это в маркетинговую систему компании, а не оставить разовым экспериментом (часть V).

Но прежде, чем переходить к механике, стоит сразу договориться о терминах. Следующая глава – о том, чем на самом деле отличаются друг от друга SEO, GEO и AEO, и почему путаница в этих терминах уже стоит компаниям реальных денег.

Источники и исследования, использованные при подготовке главы:

OpenAI – публичные данные о развитии ChatGPT и аудитории продукта;

Similarweb – исследования трафика и аудитории AI-сервисов;

Bain & Company, “Goodbye Clicks, Hello AI” (2025);

SparkToro – исследования zero-click search;

BrightEdge – исследования влияния AI Search на органический поиск;

Ahrefs – исследования AI Overviews и изменений в поисковой выдаче;

SE Ranking, Exposure Ninja, Superlines, DigitalApplied, SERPs.io, DemandSage – отраслевые обзоры и агрегированные исследования AI Search.

Глава 2. GEO vs SEO vs AEO – в чём разница на самом деле

Терминологический хаос – не случайность

Первое, что нужно честно сказать читателю: единого общепринятого разграничения между GEO и AEO по состоянию на 2026 год не существует. В англоязычной среде эти термины используются по-разному: специалисты, вендоры и отраслевые издания часто вкладывают в них пересекающиеся смыслы. Академический консенсус по чётким границам между понятиями пока также не сформировался.

Это не повод игнорировать термины или думать, что всё дело в маркетинговых ярлыках, но точно повод перестать спорить о наименованиях и вместо этого разобраться в сути. Дальше в этой главе – рабочее разграничение, которым я в проектах пользуюсь сам и которое помогает минимизировать путаницу между вами, вашей командой и заказчиком.

Откуда вообще взялся термин GEO

У термина Generative Engine Optimization есть конкретная точка отсчёта – академическая работа исследователей из Принстона, Технологического института Джорджии, Индийского технологического института Дели (IIT Delhi) и Университета Аллена, впервые предложивших сам термин и одну из первых системных методологий измерения видимости источников внутри сгенерированных ответов ещё в 2023–2024 годах. Термин стал известен после публикаций крупных технологических и венчурных компаний, включая Andreessen Horowitz – и с этого момента термин вошёл в маркетинговую среду, получив широкое распространение среди маркетинговых специалистов.

Это важная деталь для понимания: GEO изначально появился не внутри классической SEO-индустрии, а в академической среде как попытка описать новый тип оптимизации под генеративные системы – того, как большие языковые модели формируют ответ из нескольких источников и решают, какие из них процитировать.

AEO: более старый термин с более узким фокусом

Термин Answer Engine Optimization (AEO) получил распространение задолго до появления современных генеративных поисковиков. Изначально он описывал оптимизацию контента под системы, которые стремились не просто показать список ссылок, а сразу дать пользователю готовый ответ. Первый массовый этап развития AEO был связан с появлением блоков быстрых ответов в традиционных поисковых системах.

У Google эту роль выполняли несколько форматов:

Featured Snippets – специальные блоки с прямым ответом над основной поисковой выдачей. Google извлекал часть текста со страницы-источника и показывал её пользователю без необходимости переходить на сайт.

People Also Ask («Люди также спрашивают») – блоки с дополнительными вопросами и короткими ответами, позволяющие расширить поиск внутри самой страницы результатов.

Голосовые ассистенты – в первую очередь Google Assistant, где особенно важным становился контент, способный дать короткий и однозначный ответ на естественный вопрос пользователя.

Для владельцев сайтов это означало новый принцип оптимизации: недостаточно было просто попасть в топ выдачи – нужно было структурировать информацию так, чтобы отдельный фрагмент страницы мог стать готовым ответом.

Яндекс развивал похожую логику независимо от Google, делая акцент не только на текстовых ответах, но и на собственных форматах представления информации внутри поисковой страницы.

Быстрые ответы и прямые ответы. В середине 2010-х годов Яндекс начал активно развивать механизмы прямых ответов: для ряда запросов поисковая система могла показать пользователю необходимую информацию непосредственно в выдаче – например определения, факты, справочные данные или расчёты. Для контента это формировало те же требования, что и в случае Google Featured Snippets: чёткие определения, структурированные списки, таблицы, короткие информационные блоки, отвечающие на конкретный вопрос. Иными словами, поисковой системе было проще извлечь и показать фрагмент, который уже был оформлен как самостоятельный ответ.

Дополнительные вопросы и расширенные сниппеты. Как и другие поисковые системы, Яндекс развивал элементы, позволяющие пользователю уточнять запрос прямо в выдаче: связанные вопросы, дополнительные информационные блоки и расширенные сниппеты. Хотя механика отличалась от People Also Ask, общий принцип был тем же: поисковик стремился предугадать следующий вопрос пользователя и дать ответ без дополнительного перехода на сайт.

Алиса: переход к разговорному поиску. Запуск голосового ассистента Алиса в 2017 году стал ещё одним этапом развития АЕО в российском поиске. Голосовые интерфейсы изменили требования к контенту: запросы стали более разговорными, выросла роль вопросов в естественной формулировке, стало важнее давать короткие и однозначные ответы. Для попадания в голосовые ответы контент должен был быть понятен не только поисковому алгоритму, но и человеку, который воспринимает информацию на слух.

Колдунщики и обогащённые ответы. Одной из особенностей Яндекса стали так называемые колдунщики – специальные интерактивные блоки, которые решают часть пользовательских задач непосредственно в поисковой выдаче. К ним относятся карты, изображения, видео, погода, расписания, товары и другие специализированные сервисы.

С точки зрения SEO это стало важным изменением: поисковая выдача постепенно превратилась из списка ссылок в интерфейс решения задач. Для бизнеса это означало необходимость оптимизировать не только страницы сайта, но и данные, которые поисковая система могла использовать в собственных сервисах.

В основе АЕО лежал принцип: сделать информацию максимально удобной для извлечения – с понятными определениями, структурой «вопрос–ответ», списками и компактными смысловыми блоками.

Ключевое отличие АЕО от GEO – в механике извлечения информации:

АЕО в классическом понимании ближе к модели «извлечения» (extraction): система находит один наиболее подходящий фрагмент текста и показывает его почти как есть – в блоке ответа, в голосовом помощнике, в featured snippet.

GEO преимущественно работает с моделью «синтеза» (generation): система извлекает информацию из нескольких источников, объединяет её, перефразирует и формирует новый ответ. В зависимости от конкретной системы она может дополнительно указывать источники, использованные при формировании ответа.

Разница для практика принципиальна. В АЕО основной объект оптимизации – конкретный фрагмент контента, который система может извлечь как готовый ответ. В GEO объектом становится более широкая система сигналов: содержание материала, авторитетность источника, цифровой след бренда и вероятность выбора этого источника при формировании ответа.

Некоторые специалисты считают, что различие между АЕО и GEO во многом искусственное и что речь идёт об одном наборе практик под разными названиями. Действительно, значительная часть базовых принципов совпадает: комплексное раскрытие темы, авторитетность источника, понятная структура контента, ясность формулировок и качественные данные для машинного понимания работают в обоих направлениях. Однако цели оптимизации всё

же отличаются: АЕО исторически стремился дать системе готовый ответ, а GEO – повысить вероятность присутствия бренда внутри сгенерированного ответа.

А что тогда случилось с SEO?

Здесь важно закрыть частый вопрос заказчиков: «SEO умерло?» – нет. Но роль SEO меняется.

Классическое SEO решало задачу ранжирования: «как попасть на первую позицию в списке ссылок». Эта задача никуда не исчезла: Google и Яндекс по-прежнему остаются крупнейшими поисковыми системами, а традиционный поиск – одним из основных источников органического трафика для большинства сайтов.

Более того, техническая база SEO – доступность для индексации, производительность, качество структуры данных, ссылочный авторитет и содержательность контента – остаётся фундаментом и для GEO. Современные ИИ-системы работают с тем же публичным вебом, который десятилетиями формировался под влиянием SEO-практик, а некоторые системы (например, Perplexity) используют веб-поиск и актуальные источники при формировании ответов.

Рабочая модель, которую я использую с клиентами, – это вложенные контуры, а не три конкурирующих канала:

SEO – внешний контур: видимость сайта в классическом поиске, его техническая доступность, авторитетность и соответствие поисковым запросам.

АЕО – средний контур: способность контента быть извлечённым поисковой или ответной системой как прямой, самодостаточный ответ.

GEO – внутренний контур принятия решения ИИ-системы: способность бренда быть включённым в набор источников, используемых при формировании ответа генеративной системой, и получить упоминание даже без прямого перехода пользователя на сайт.

Большая часть фундаментальной работы – экспертный, глубокий и структурированный контент; техническая доступность для поисковых и ИИ-систем; цифровой след бренда за пределами собственного сайта – создаёт фундамент сразу для нескольких контуров видимости. Разделять их на три полностью независимых бюджета и три отдельные команды часто означает дублировать усилия и терять связность стратегии. Правильнее говорить о едином контуре видимости с разными точками измерения на выходе.

Почему путаница в терминах уже стоит компаниям денег

На практике я регулярно вижу три одинаковые ошибки, которые возникают именно из-за терминологической путаницы:

Ошибка первая. Компания покупает у подрядчика «услугу GEO», которая во многом состоит из тех же тактик, что и классический SEO-аудит прошлых лет, только переупакованных в новую презентацию. Отличить реальную GEO-экспертизу от ребрендинга старой услуги можно по одному вопросу: подрядчик умеет объяснить, как именно конкретная ИИ-система выбирает и использует источники при формировании ответа, или просто говорит «мы оптимизируем ваш контент под ИИ»?

Ошибка вторая. Внутренняя команда маркетинга тратит месяцы на попытки попасть в featured snippet и другие прямые ответы (классическая задача АЕО), считая, что тем самым решает задачу видимости в ChatGPT и других генеративных системах (задача GEO), – и удивляется, почему результата в ИИ-ответах не видно.

Ошибка третья, самая дорогая. Руководство ставит перед маркетингом задачу «внедрить GEO» без понимания, что метрики успеха здесь принципиально отличаются от классического SEO. Ожидание роста кликов и трафика в привычной логике часто приводит к неверной оценке канала, который на самом деле работает – просто не так, как работало SEO десять лет назад.

Итог главы: рабочий словарь на всю книгу

Дальше в книге термины используются так:

SEO – работа с видимостью в традиционной поисковой выдаче и органическим трафиком из классического поиска.

AEO – работа со структурой контента ради попадания в прямые, извлекаемые ответы поисковых и ответных систем (featured snippets, блоки прямых ответов и другие форматы извлечения информации).

GEO – работа с авторитетностью, цифровым следом и структурой контента ради упоминания бренда или использования его материалов при формировании ответов ИИ-систем, независимо от того, произойдёт ли переход на сайт.

AI SEO – зонтичный термин для всей совокупности практик оптимизации под ИИ-поиск, когда разграничение между AEO и GEO для конкретной задачи не принципиально.

С этим словарём переходим к следующему вопросу – как именно устроена «кухня» генеративного поиска изнутри: что такое RAG, эмбединги, извлечение и ранжирование источников, и почему это определяет всё, что будет описано в практической части книги.

Источники данных главы:

Aggarwal et al., “GEO: Generative Engine Optimization”, KDD 2024 / arXiv 2023

Wikipedia – Generative Engine Optimization (для фиксации публичного определения термина);

“Position: Generative Engine Optimization Creates Underexamined Risks”, arXiv, 2026;

отраслевые материалы по AEO и GEO: Writesonic, Jasper, Core dna, WITHIN, Profound, 2point Agency.

Глава

3.

Как

устроены

ChatGPT, Perplexity, Gemini, Copilot, YandexGPT:

откуда

берутся

ответы

и

цитаты

Общий механизм: RAG простыми словами

Прежде чем разбирать различия между платформами, нужно понять один из ключевых принципов современных ИИ-поисковых систем – Retrieval-Augmented Generation (RAG), «генерация, дополненная поиском».

Языковая модель сама по себе – это не база данных с фактами, а модель, обученная выявлять закономерности в огромных массивах текста и генерировать ответы на основе этих закономерностей. У неё есть определённая дата отсечения знаний, поэтому при запросах о свежих событиях, текущих ценах или новых данных без доступа к внешним источникам она может либо сообщить об ограничении, либо сформировать правдоподобный, но неверный ответ – так называемую галлюцинацию.

Многие современные ИИ-поисковые системы используют архитектуры, основанные на принципах RAG или близких к ним подходах. RAG же в ИИ-поисковых системах можно представить как несколько последовательных этапов:

Понимание и расширение запроса. Система интерпретирует, что именно хочет узнать пользователь, и может преобразовать один запрос в несколько вспомогательных поисковых формулировок. Например, подход query fan-out позволяет разбивать сложный вопрос на несколько параллельных подзадач для более полного поиска информации.

Поиск и извлечение (retrieval). Система обращается к внешним источникам – поисковому индексу, базе документов, корпоративному хранилищу или вебу в реальном времени – и получает набор потенциально релевантных фрагментов информации.

Генерация ответа (generation). Большая языковая модель формирует ответ с учётом найденного контекста и внутренних знаний, и, в зависимости от конкретной системы, может указывать источники, использованные при формировании ответа.

ChatGPT: избирательность и многоуровневый поиск источников

Поисковая функция ChatGPT (ChatGPT Search) использует веб-поиск и внешние источники для получения актуальной информации. Одним из важных компонентов этой инфраструктуры является интеграция с поисковыми технологиями Microsoft Bing, однако итоговый выбор источников определяется не только поисковым индексом, а собственными механизмами OpenAI по извлечению, оценке и представлению информации.

Для бренда это означает: простого попадания в индекс недостаточно. Сайт должен быть технически доступен для обнаружения, содержать релевантную и авторитетную информацию и обладать характеристиками, которые повышают вероятность использования его материалов при формировании ответа.

Ключевые наблюдаемые характеристики ChatGPT как источника ответов:

Избирательность. ChatGPT Search обычно показывает более ограниченный набор источников по сравнению с некоторыми специализированными AI-поисковиками, делая акцент на наиболее релевантных ссылках для конкретного ответа.

Не каждый ответ использует веб-поиск. Часть ответов ChatGPT формируется на основе внутренних знаний модели без обращения к внешним источникам. В таких сценариях классическая GEO-оптимизация сайта напрямую не влияет на результат.

Важность структуры контента. Наблюдения практиков показывают: материалы, где ключевой ответ сформулирован ясно и расположен близко к началу страницы, чаще оказываются удобными для извлечения ИИ-системами.

Собственный поисковый краулер. OpenAI использует OAI-SearchBot для задач, связанных с поисковыми функциями, отдельно от GPTBot, который предназначен для других сценариев обработки веб-контента. Эти правила можно учитывать через robots.txt.

Perplexity: поиск, для которого источники – это ядро продукта

Если ChatGPT начинал как универсальный ИИ-ассистент, к которому позже добавили поиск, то Perplexity с самого начала строился вокруг модели «ответной машины» (answer engine), где цитируемые источники являются центральной частью пользовательского опыта.

В режиме поиска Perplexity обращается к актуальным веб-источникам, извлекает набор релевантных материалов и формирует ответ с пронумерованными ссылками, похожими на научное цитирование.

Отличия от ChatGPT принципиальны:

Более высокая плотность цитирования. Perplexity обычно показывает больше ссылок на источники, чем многие универсальные ИИ-ассистенты, поскольку прозрачность происхождения ответа является частью его ключевой ценности.

Готовность использовать нишевые источники. Платформа активно обращается к специализированным сайтам и пользовательскому контенту, если они содержат наиболее релевантную информацию по конкретному вопросу.

Фокус на уровне конкретного материала. Для попадания в ответ важна не только известность домена, но и то, насколько конкретная страница закрывает потребность пользователя.

Отличие от других ИИ-систем. Разные платформы используют разные наборы источников и механизмы отбора, поэтому видимость бренда в одном ИИ-поисковике не гарантирует присутствие в другом.

Gemini и Google AI Overviews: наследие классического поиска

Gemini и функция AI Overviews внутри поиска Google используют подход search grounding – генерация ответа опирается на данные поиска Google и другие доступные источники информации.

Практический вывод для стратегии: качественная SEO-база остаётся важным фактором видимости в экосистеме Google. Сайты, которые уже обладают высокой релевантностью, авторитетностью и технической доступностью для поисковой инфраструктуры Google, получают преимущество при попадании в пул потенциальных источников для AI-ответов.

При этом AI Overviews – это не просто новая форма классической выдачи. Механизм формирования ответа отличается от обычного ранжирования ссылок: Google выбирает, какие источники использовать и какие из них показывать пользователю в виде ссылок или цитат.

Цитирование в AI-ответах Google появляется выборочно. Часть информации может быть представлена без отдельной ссылки на каждый тезис, даже если при формировании ответа использовались внешние источники.

Copilot: тот же поисковой фундамент, но другая упаковка

Microsoft Copilot использует поисковую инфраструктуру Bing как один из ключевых источников актуальной информации. Поэтому с точки зрения доступности контента между Copilot и другими системами, использующими веб-поиск, может существовать пересечение источников, однако итоговый выбор цитат определяется собственными механизмами Microsoft.

Для практической стратегии это означает: работа над видимостью в Bing, технической доступностью сайта и качеством контента может повышать вероятность появления бренда в ответах Copilot. Но гарантированного переноса результатов между Bing, ChatGPT Search и Copilot не существует.

YandexGPT и Нейро: отдельная экосистема для русскоязычного поиска

Для компаний, работающих на российском рынке, YandexGPT и генеративные функции Яндекса становятся одним из ключевых направлений GEO из-за высокой роли экосистемы Яндекса в рунете.

Нейро – AI-ответы внутри поиска Яндекса – использует поисковую инфраструктуру компании и дополнительные источники данных для формирования ответа. Как и Google AI Overviews, эта система не просто показывает ссылки, а синтезирует информацию из нескольких источников.

Для GEO-стратегии в рунете важны не только позиции в классической выдаче, но и общий цифровой след бренда: официальный сайт, отраслевые публикации, отзывы, карточки компаний и другие доверенные источники.

Среди особенностей российского рынка можно выделить более высокую роль локальных площадок и сервисов: экосистемы Яндекса, отраслевых медиа, сервисов отзывов и площадок с пользовательским контентом. Поэтому стратегия GEO для рунета не может быть простым переносом англоязычных подходов.

Как и другие AI-поисковые системы, Нейро постепенно меняет поведение пользователей: часть информационных задач решается непосредственно внутри поискового интерфейса, без обязательного перехода на сайты.

Claude: другой сценарий – модельная память и контекст

Отдельно стоит рассмотреть Claude от Anthropic. В отличие от систем, которые изначально позиционируются как поисковые ассистенты (например, Perplexity), Claude прежде всего является универсальным языковым помощником.

Поэтому часть ответов Claude формируется на основе знаний, полученных моделью во время обучения, без обращения к внешнему поиску. В таких сценариях классическая логика

GEO – «получить ссылку в ответе» – работает ограниченно: ссылки могут отсутствовать вообще.

Практический вывод для бизнеса: видимость в подобных системах зависит не только от отдельных оптимизированных страниц, но и от общего цифрового следа бренда – наличия экспертных материалов, независимых упоминаний, публикаций и других сигналов, которые помогают модели корректно понимать компанию и её место в категории.

При этом нельзя воспринимать это как прямую оптимизацию под «попадание в память модели»: внутренние данные обучения закрыты, регулярно меняются, а влияние конкретных источников невозможно точно измерить.

Что из этого следует для практика

Три практических вывода закрывают главу и задают рамку для части III:

Унифицированной стратегии «оптимизации под ИИ-поиск» не существует. У разных платформ различаются источники данных, поисковая инфраструктура, механизмы извлечения информации и принципы формирования ответа. Стратегия должна учитывать приоритетные для бизнеса системы – и для большинства российских проектов в этот список должен входить YandexGPT и генеративные функции Яндекса, поскольку переводные англоязычные гайды по GEO часто не учитывают специфику рунета.

Классическое SEO не потеряло значения – оно стало частью более широкой системы видимости. Для Google-экосистемы SEO остаётся особенно важным фундаментом, поскольку AI-ответы связаны с поисковой инфраструктурой компании. Для систем, использующих веб-поиск и внешние индексы, SEO также остаётся полезным фактором. При этом существуют платформы, где результат сильнее зависит от механизмов retrieval в момент запроса, а не только от классического ранжирования.

Техническая доступность для правильных краулеров – обязательное, но не достаточное условие. Быть проиндексированным – не значит быть процитированным. Дальше мы разберём, какие сигналы – помимо технической доступности – определяют, выберет ли система именно ваш источник для формирования ответа.

Источники данных главы:

Первичные и технические источники:

OpenAI – материалы по ChatGPT Search, GPTBot и OAI-SearchBot.

Google Search Central / Google DeepMind – материалы по AI Overviews, AI Mode и генеративному поиску.

Microsoft – материалы Bing Search и Copilot.

Яндекс – материалы по YandexGPT, Нейро и генеративным функциям поиска.

Aggarwal P. et al. “GEO: Generative Engine Optimization”, arXiv, 2023.

Martinez O. “Optimizing Visibility in Generative Engines: A Critical Survey of Generative Engine Optimization (2023–2026)”, arXiv, 2026.

Отраслевые исследования и аналитика:

BrightEdge – исследования AI Search и влияния AI Overviews.

Ahrefs – исследования AI Overviews и изменений органического поиска.

SE Ranking – AI Search Research.

Similarweb – анализ аудитории и трафика AI-сервисов.

Indexly – AI Search citation analysis (2026).

Topify – анализ различий цитирования ChatGPT, Perplexity и Gemini (2026).

Profound – исследования AI visibility и brand mentions.

Authoritas – исследования AI Search visibility.

Глава 4. Кому и зачем это нужно уже сейчас

Прежде чем считать деньги – снимем один миф

Самое частое возражение при обсуждении GEO звучит так: «В GA4 трафик с ИИ-систем пока составляет доли процента. Зачем инвестировать в канал, которого почти не видно?»

Возражение, на первый взгляд, логичное. Но оно основано на старой модели оценки маркетинговых каналов, где главным доказательством ценности был прямой переход на сайт.

По данным отраслевых исследований и наблюдений аналитических платформ, прямой трафик с ИИ-систем пока действительно занимает небольшую долю общего трафика большинства сайтов. Но эта метрика не показывает всей картины.

Дело не только в объёме привлечённых пользователей, а в том, на каком этапе принятия решения они приходят.

Ключевая находка первых исследований AI-трафик – качество крайне важно.

Первые отраслевые замеры показывают, что пользователи, пришедшие с ИИ-платформ, часто находятся дальше по воронке принятия решения, чем посетители из классического поиска.

Например, анализ Opollo на выборке B2B tech-компаний показал значительно более высокую конверсию AI-трафика по сравнению с органическим поиском. Аналогичные тенденции фиксировали и другие аналитические платформы, хотя конкретные значения сильно зависят от отрасли, типа продукта и методологии подсчёта.

Ahrefs в собственном анализе также отмечал непропорционально высокий вклад AI-источников в регистрации по сравнению с их долей общего трафика.

В электронной коммерции похожий тренд наблюдается и в крупных массивах данных: исследования Adobe показывают, что трафик из генеративных ИИ-источников постепенно становится более качественным с точки зрения вовлечённости и коммерческого поведения.

Другими словами: пока многие компании продолжают оценивать ИИ-трафик только по его доле в общей структуре посещений (и справедливо видят там пока небольшие значения – от долей процента до единиц процентов), более зрелый подход – смотреть не только на объём, но и на ценность такого визита.

Доступные исследования ранних рынков AI Search показывают устойчивую тенденцию: пользователи, приходящие через ИИ-системы, во многих сегментах демонстрируют более высокое коммерческое намерение и конверсию, чем средний органический трафик. Масштаб эффекта при этом сильно зависит от отрасли: от умеренного преимущества в отдельных категориях до многократного роста в некоторых B2B SaaS и технологических сегментах.

Важная оговорка: картина неоднородна. В отдельных направлениях (например, части B2B e-commerce в ранних замерах) конверсия с ИИ-трафика пока может быть ниже или сопоставима с органическим поиском. Поэтому оценивать потенциал канала нужно не по средним цифрам из чужих кейсов, а по собственной отрасли, аудитории и бизнес-модели.

Кто уже не может позволить себе ждать

По совокупности отраслевых данных 2025–2026 годов можно выделить несколько сегментов, где цена промедления с развитием видимости в ИИ-системах особенно высока уже сейчас.

Технологичный B2B и SaaS

Здесь эффект наиболее заметен: покупатели программного обеспечения и технологичных услуг тратят значительное время на изучение категорий, сравнение решений, изучение документации и экспертных материалов ещё до первого контакта с продавцом. Это именно тот тип поведения, для которого ИИ-ассистенты подходят особенно хорошо: пользователь форму-

лирует задачу, получает подборку решений, сравнение подходов и рекомендации ещё до того, как попадает на сайт поставщика.

По данным отдельных исследований AI referral traffic, ChatGPT и другие ИИ-платформы уже становятся заметным источником переходов для B2B tech-компаний. Для рынка с длинным циклом сделки это особенно важно: бренд может влиять на выбор покупателя задолго до появления прямого обращения.

Здравоохранение и медицинские услуги

Медицинская тематика – одна из категорий, где доверие к источникам имеет критическое значение. Пользователи обращаются к ИИ-системам не только за общей информацией о заболеваниях и симптомах, но и за помощью в выборе специалистов, клиник и вариантов лечения.

Анализ использования ИИ-ассистентов показывает, что значительная часть таких запросов возникает вне стандартных сценариев взаимодействия с медицинскими организациями – в том числе тогда, когда пользователь не может сразу получить консультацию через привычные каналы. В этом случае ИИ становится промежуточным слоем между человеком и медицинским брендом.

Для компаний из этой сферы вопрос присутствия в ответах ИИ – это уже не только вопрос дополнительного трафика, а вопрос того, какие источники пользователь увидит в момент формирования доверия.

Категории с высоким разрывом «спрос есть – цитирования нет»

Один из наиболее показательных сценариев – ситуации, когда интерес пользователей к категории высокий, но конкретные компании практически отсутствуют в ответах ИИ-систем.

Отраслевые исследования отдельных ниш, включая кибербезопасность, показывают: при коммерческих запросах о выборе решений ответы часто концентрируются вокруг ограниченного набора наиболее известных источников – крупных отраслевых площадок, аналитических ресурсов и энциклопедических материалов. При этом сами поставщики решений могут получать значительно меньше упоминаний, чем ожидают.

Это крайний, но показательный случай. Похожая ситуация характерна для многих узкоспециализированных B2B-рынков: экспертиза у компаний есть, продукты есть, спрос есть – но цифровой след, который помогает ИИ-системам распознать компанию как авторитетный источник, недостаточно развит.

Розница и e-commerce

В 2025–2026 годах исследования розничного рынка фиксируют ускорение роста переходов с ИИ-платформ на сайты продавцов и производителей. При этом во многих случаях такие пользователи демонстрируют более высокое коммерческое намерение, поскольку приходят уже после этапа сравнения вариантов и формулирования конкретной потребности.

Для e-commerce это означает появление нового слоя конкуренции: компании будут бороться не только за позиции в поисковой выдаче, но и за место в рекомендациях ИИ-ассистентов внутри категорий. Окно возможностей пока остаётся открытым: многие бренды ещё не выстроили системную работу с видимостью в ИИ-системах, поэтому ранние игроки могут сформировать преимущество до того, как рынок станет насыщенным.

Регулируемые и высококонкурентные ниши

Финансы, медицина, юридические услуги и другие регулируемые категории имеют дополнительную особенность: ИИ-системы особенно чувствительны к качеству и надёжности источников.

В таких тематиках модели и поисковые системы чаще склоняются к уже признанным источникам, официальным данным, экспертным публикациям и брендам с устойчивой репутацией. Это делает раннее системное формирование цифрового авторитета особенно ценным: компания получает не просто дополнительный канал привлечения, а долгосрочный актив в виде накопленного доверия.

Простая модель расчёта упущенной выгоды

Для разговора с CEO нужен не абстрактный тезис «это важно», а понятная модель оценки. Ниже – упрощённая рамка расчёта, которую можно использовать на первой встрече.

Она не является точным прогнозом будущей выручки: рынок ИИ-поиска ещё слишком молодой, а методы измерения продолжают развиваться. Её задача – оценить порядок потенциальной возможности и понять, стоит ли запускать пилот.

Шаг 1. Оцените объём релевантного спроса в категории

Возьмите данные по поисковому спросу в вашей нише из Яндекс Вордстата, Google Keyword Planner или других инструментов и определите, какая часть пользовательских сценариев уже может решаться через ИИ-интерфейсы.

Универсального коэффициента для всех отраслей пока не существует: доля AI-поиска зависит от категории, аудитории, страны и сложности выбора продукта.

Особое внимание стоит уделять информационным, сравнительным и исследовательским запросам – именно они чаще всего переходят в формат диалога с ИИ.

Шаг 2. Оцените текущую видимость бренда

Составьте список из 15–20 реальных вопросов, которые потенциальный клиент задаёт перед покупкой: «какой сервис выбрать для ...», «лучшие компании в категории ...», «сравнение решений ...», «альтернативы продукту ...».

Проверьте эти запросы в ChatGPT, Perplexity, Gemini и YandexGPT. Посчитайте, как часто появляется ваш бренд, как часто появляются конкуренты, какие источники система использует для формирования ответа.

Если бренд упоминается только в небольшой доле сценариев, это означает, что значительная часть потенциального влияния уже находится у конкурентов.

Шаг 3. Оцените ценность такого присутствия

Используйте собственные данные аналитики: стоимость привлечения клиента, конверсию органического трафика, средний чек и LTV.

Затем сравните их с доступными отраслевыми исследованиями AI referral traffic.

Первые исследования показывают, что в ряде сегментов пользователи, приходящие через ИИ-системы, демонстрируют более высокое коммерческое намерение и конверсию, чем средний органический трафик.

При этом масштаб эффекта сильно зависит от отрасли: в отдельных B2B SaaS и технологичных категориях разница может быть многократной, тогда как в других сегментах преимущество может быть значительно меньше.

Шаг 4. Рассчитайте потенциальный разрыв

Упрощённо это делается по формуле:

Потенциальная возможность = Объём релевантного спроса × Доля сценариев AI-поиска × Разрыв текущей видимости × Ценность привлечения клиента.

Полученное число не является гарантированным прогнозом выручки. Это оценка масштаба возможности, которая помогает принять решение о запуске GEO-проекта.

На практике даже консервативный расчёт для компании с сильной экспертизой и заметным присутствием в категории часто показывает, что стоимость системной работы над AI-видимостью значительно ниже потенциальной стоимости отсутствия в этом канале.

Почему промедление обходится дороже, чем кажется

Здесь стоит вернуться к структурной особенности, которую мы разобрали в первой главе: ИИ-системы не просто находят страницы – они выбирают источники, которым доверяют при формировании ответа.

Это создаёт эффект, частично похожий на ранние годы SEO: компании, которые раньше начали создавать качественный контент, экспертные публикации, внешние упоминания и устойчивый цифровой след, получили преимущество перед теми, кто пришёл позже.

Однако механизм отличается: речь идёт не только о позициях в выдаче, а о способности бренда быть распознанным как значимый источник внутри нового слоя поиска.

Сегодня рынок ещё находится на ранней стадии формирования стандартов GEO. Многие компании пока не измеряют собственную видимость в ИИ-ответах и не понимают, насколько часто их заменяют конкуренты в момент выбора пользователя.

Именно поэтому окно возможностей остаётся открытым: конкуренция за место в ответах ИИ только начинается.

Итог части I

Мы прошли путь от изменения самого интерфейса поиска (глава 1) через наведение порядка в терминах (глава 2) и разбор технической механики разных ИИ-платформ (глава 3) к экономическому обоснованию нового канала (глава 4).

Дальше книга переходит от вопроса «зачем это нужно» к вопросу «как это работает».

В части II разберём, какие именно сигналы определяют выбор источников ИИ-системами: качество контента, экспертность, цифровой след бренда, техническая доступность и внешние факторы доверия.

Часть III покажет, что конкретно делать компаниям: от архитектуры сайта и контент-стратегии до построения системного процесса GEO-оптимизации.

Источники данных главы:

Исследования рынка и аналитика: Adobe Digital Insights – исследования влияния генеративного ИИ на e-commerce traffic, conversions и consumer behavior (2024–2026).

Bain & Company – Goodbye Clicks, Hello AI (2025).

Conductor – исследования AI Search adoption и изменения поведения маркетологов.

Ahrefs – исследования AI Overviews, AI Search visibility и влияния ИИ на органический поиск.

Semrush – исследования AI Search, поискового поведения и изменения SERP.

Seer Interactive – анализ AI referral traffic и поведения пользователей.

BrightEdge – исследования влияния AI Search на органический трафик.

Similarweb – анализ трафика AI-платформ и изменения пользовательских сценариев.

Gartner – исследования внедрения генеративного ИИ в маркетинге и клиентском опыте.

Отраслевые отчёты:

AirOps – State of AI Search (как дополнительный источник отраслевых наблюдений).

Profound – исследования AI visibility и brand mentions в генеративных системах.

Глава 5. RAG, эмбединги и ранжирование источников – простыми словами, но без лжи

Почему эту главу нельзя пропустить

В большинстве книг и статей про GEO техническую механику либо вообще опускают («не нужно быть инженером, чтобы заниматься GEO»), либо, наоборот, топят читателя в терминах, не объясняя, зачем это практику. Я выбираю третий путь: разобрать механику ровно настолько глубоко, чтобы дальнейшие практические рекомендации в частях III и IV перестали быть «шаманством» и превратились в логичные следствия из того, как система устроена физически.

Обещание в подзаголовке – «без лжи» – принципиально: слишком много материалов на рынке округляют механику до уровня, на котором она перестаёт быть правдой. Здесь такого не будет.

Шаг первый: что такое эмбединг

Представьте, что любой текст – слово, предложение или целый абзац – можно превратить в точку на условной карте, где расстояние между точками отражает смысловую близость. Чем ближе расположены две точки, тем более похожими по смыслу система считает соответствующие тексты. Разумеется, никакой «карты» в буквальном смысле не существует – это лишь удобная метафора многомерного математического пространства, которое невозможно наглядно представить. Тексты про «регистрацию бизнеса в Дубае» и «открытие компании в ОАЭ» окажутся на этой карте рядом, даже если ни одно слово в них буквально не совпадает. А текст про «крем для лица» окажется далеко от обоих.

Именно для этого используются модели эмбедингов (embedding models). Они преобразуют фрагмент текста в многомерный числовой вектор (embedding) – представление, которое позволяет сравнивать тексты по смысловой близости. Современные модели, например OpenAI text-embedding-3 или Gemini Embedding, обучаются на огромных массивах данных так, чтобы расстояние между такими векторами как можно лучше отражало смысловую, а не только лексическую близость текстов.

Практический вывод номер один, который стоит запомнить уже здесь: система не ищет по ключевым словам в привычном понимании SEO-специалиста десятилетней давности. Она ищет по смысловой близости. Текст, который отвечает на вопрос пользователя по существу, но без точного вхождения ключевой фразы, вполне может «победить» текст с идеальным вхождением ключевого слова, но размытым, обтекаемым содержанием.

Шаг второй: чанкинг – почему структура текста решает всё

В большинстве современных RAG-систем в эмбединг превращают не весь сайт и обычно даже не всю статью целиком. Длинный документ охватывает сразу несколько подтем, поэтому его единый вектор становится менее точным: смысл разных разделов смешивается. Поэтому контент обычно сначала разбивают на фрагменты – чанки (chunks), и уже каждый из них преобразуют в отдельный вектор.

Здесь есть развилка, критически важная для практика:

Слишком мелкие чанки теряют контекст: фрагмент формально совпадает по смыслу с запросом, но в нём не хватает деталей, которые нужны модели для полноценного ответа.

Слишком крупные чанки размывают сигнал: в одном векторе смешивается слишком много разных идей, и точность совпадения с конкретным запросом падает.

На практике многие современные системы используют чанки порядка нескольких сотен токенов – часто в диапазоне 300–800, – с небольшим перекрытием между соседними фрагмен-

тами. Конкретные параметры зависят от архитектуры системы и используемой модели эмбедингов.

Отдельно стоит явление, получившее название *late chunking* («позднее разбиение на чанки»), оно было популяризировано исследователями Jina AI. Суть в следующем: вместо того, чтобы резать текст на куски до превращения в вектор, весь документ сначала пропускают целиком через модель с длинным контекстом, и лишь затем «схлопывают» токен-уровневые представления в отдельные чанки. Плюс такого подхода в том, что каждый фрагмент «помнит» контекст всего документа – например, если текст трижды упоминает «город» после того, как в первом абзаце был назван конкретный город, поздний чанкинг сохранит эту связь даже в чанке, где сам город по имени уже не упоминается.

Практический вывод для автора контента прост: хорошая структура документа – это не вопрос литературного стиля, а фактор, влияющий на работу RAG-систем. Подзаголовки, принцип «один раздел – одна законченная мысль», отсутствие смысловых «хвостов», когда новый абзац невозможно понять без нескольких предыдущих, – всё это помогает создавать более качественные векторные представления и облегчает системе извлечение нужного фрагмента при ответе на запрос пользователя.

Шаг третий: как ищутся кандидаты – от косинусного сходства до гибридного поиска

Когда пользователь задаёт вопрос, тот же процесс преобразования в вектор происходит и с самим запросом. Затем система ищет среди проиндексированных чанков те, чьи векторные представления наиболее близки к вектору запроса. Для этого часто используется математическая мера, называемая косинусным сходством (*cosine similarity*): она оценивает угол между двумя векторами в многомерном пространстве. Чем меньше этот угол (или, что эквивалентно, чем выше значение косинусного сходства), тем более близкими по смыслу система считает запрос и найденный фрагмент.

Однако на этом поиск обычно не заканчивается. На практике современные системы всё чаще используют гибридный поиск, который сочетает семантический поиск с классическим полнотекстовым поиском и последующим реранжированием результатов. Ниже разберём, зачем это нужно.

Именно здесь становится понятна реальная роль классического SEO. Авторитет сайта, ссылочный профиль и техническое качество продолжают влиять на вероятность того, что документ окажется в поисковом индексе и будет рассматриваться системой как потенциальный источник. Однако при выборе конкретных фрагментов для ответа всё большее значение приобретает смысловая релевантность самого текста. Когда же система выбирает конкретные фрагменты для ответа, решающее значение приобретает уже не авторитет домена сам по себе, а смысловая релевантность конкретного чанка запросу пользователя.

Практический вывод прост: сильное SEO по-прежнему остаётся фундаментом, но его уже недостаточно. Чтобы текст оказался в ответе ИИ-системы, мало быть хорошо проиндексированным – нужно, чтобы отдельные фрагменты страницы были настолько понятными, информативными и семантически точными, чтобы именно их система сочла лучшим ответом на запрос пользователя.

При этом чистый векторный поиск в 2026 году уже не считается универсальным решением. Во многих современных RAG-системах используется гибридный поиск (*hybrid search*): результаты векторного (семантического) поиска объединяются с результатами классического полнотекстового поиска, чаще всего на основе алгоритма BM25, который оценивает точные совпадения слов и их статистическую значимость в документе. После этого обычно выполняется дополнительный этап – реранжирование (*re-ranking*), при котором более точная, но вычислительно более затратная модель (например, *cross-encoder*) повторно оценивает относительно

небольшой набор найденных кандидатов – условно несколько десятков или сотен документов – и определяет их окончательный порядок.

Причина такого подхода проста. Чисто семантический поиск отлично находит смысловые соответствия, но иногда уступает классическому поиску в случаях, когда критически важно буквальное совпадение: названия моделей, версии продуктов, артикулы, номера нормативных документов, юридические термины, имена компаний или другие уникальные идентификаторы. Полнотекстовый поиск компенсирует этот недостаток, а последующее реранжирование объединяет преимущества обоих подходов, позволяя системе выбрать наиболее релевантные источники.

Шаг четвёртый: финальное ранжирование – не только про смысл

После того как система отобрала наиболее подходящие фрагменты по смысловой и текстовой близости, начинается следующий этап – их дополнительная оценка и ранжирование. Конкретная реализация отличается у разных платформ, но в современных RAG-системах обычно учитывается сразу несколько групп сигналов:

Авторитетность (authority) – насколько источник заслуживает доверия с точки зрения модели и поисковой системы.

Качество (quality) – признаки экспертности, достоверности, структурированности материала и отсутствие явных признаков низкогокачественного или автоматически сгенерированного контента.

Актуальность (recency) – насколько информация соответствует текущему состоянию предметной области, если запрос чувствителен ко времени.

Релевантность (relevance) – насколько конкретный фрагмент отвечает именно на данный запрос, а не лишь затрагивает смежную тему.

В коммерческих поисковых системах эти и другие сигналы объединяются с помощью моделей машинного обучения, формируя итоговую оценку кандидатов. Именно по ней система определяет, какие фрагменты попадут в контекст языковой модели и смогут повлиять на формирование окончательного ответа. Важно понимать: здесь уже нет универсальной формулы. ChatGPT Search, Perplexity, Google AI Overviews, Gemini и другие платформы используют разные алгоритмы, а их детали не раскрываются публично. Поэтому речь идёт не о точном воспроизведении внутренней логики конкретной системы, а о принципах, которые устойчиво подтверждаются исследованиями и практическими наблюдениями.

Источники и исследования, использованные при подготовке главы:

Atlan – What Are Embeddings in AI (2026)

XLR8 AI – How AI Search Works: LLM Retrieval Explained (2026)

ZipTie.dev – Optimizing for Vector Embeddings (2026)

Crosscheck – What Are Embeddings in LLMs (2026)

Stack AI – Best Embedding Models for RAG (2026)

Hashmeta AI – Chunk-Level Optimization (2026)

arXiv – Towards AI Search Paradigm (npenpint 2506.17188)

Глава 6. Что такое «цитируемость» и как ИИ выбирает, кого упомянуть

Цитируемость – это не одно явление, а два разных

Первое, что важно понять перед разговором о «цитируемости» как цели GEO-стратегии: попадание в ответ ИИ-системы бывает как минимум двух принципиально разных типов. Путать их – значит неправильно ставить цели и неверно оценивать результат.

Исследование компании Machine Relations, основанное на анализе более 21 тысячи цитирований в 2026 году, предложило полезное разграничение:

Citation selection (выбор источника для цитирования). Ваш материал появляется среди источников ответа – в виде ссылки или списка использованных ресурсов. При этом сам ответ может быть сформирован преимущественно на основе нескольких источников, и вклад именно вашего материала окажется минимальным.

Citation absorption (усвоение содержания источника). Информация из вашего материала становится одной из основ ответа. Модель опирается на ваши факты, структуру объяснения, терминологию или выводы, даже если пересказывает их собственными словами и объединяет с данными из других источников.

Разница на практике огромна. Попасть в список источников ответа и реально повлиять на содержание ответа – абсолютно не одно и то же.

В первом случае пользователь видит ваш бренд среди процитированных ресурсов. Это уже ценный сигнал доверия и возможность получить переход. **Во втором** – информация из вашего материала становится частью самого ответа: система использует ваши данные, терминологию, классификации или объяснительную модель при формировании результата.

По данным того же исследования, материалы с высоким уровнем «поглощения» (absorption) чаще обладают несколькими общими характеристиками: они лучше структурированы, содержат законченные смысловые блоки и используют так называемые «доказательные жанры» (evidence genres) – чёткие определения, конкретные цифры, пошаговые инструкции и другие элементы, которые легко извлечь и проверить.

Атрибуционный разрыв: вас могут использовать чаще, чем цитировать

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.