

Состав этапа «Техническая/Контентная/Источники» GEO-аудита бренда

Что реально влияет на AI visibility

В 2026 «видимость бренда в ответах LLM» перестала быть единым явлением: она складывается из нескольких *поверхностей выдачи* (surfaces), и у каждой есть своя «воронка попадания»: **(а) быть доступным для извлечения → (б) быть найденным в поиске/индексе → (в) быть выбранным как поддерживающий источник → (г) быть процитированным/сосланным или рекомендованным**. Официальные документы вендоров почти не раскрывают «факторы топ-5», но довольно четко описывают **минимальные условия попадания в кандидаты (eligibility) и механизмы контроля доступа/цитирования** — именно это и должно составлять костяк вашего этапа. ¹

Ключевые поверхности, под которые имеет смысл строить Stage 2:

- **Поиск и ответы с источниками в ChatGPT.** Официально: чтобы сайт *попал* в *ответы поиска*, важно не блокировать **OAI-SearchBot** и не резать его на уровне хоста/CDN (нужно разрешить трафик с опубликованных IP-диапазонов). При этом ранжирование «внутри поиска» многофакторное и не гарантирует топ-позиции. ²
- **Google Search AI-фичи (AI Overviews и AI Mode).** Официально: *нет специальных требований* «под AI» — действуют обычные SEO-базисы. Чтобы быть показанным как supporting link, страница должна быть **проиндексирована и допущена к показу со сниппетом**. Эти фиши используют «query fan-out» (несколько связанных запросов по подтемам), поэтому шанс быть источником зависит от того, находят ли системы релевантные страницы под каждую подтему. ³
- **Веб-поиск/исследование в Claude.** Официально у Anthropic ⁴ выделены разные боты: **ClaudeBot** (для обучения), **Claude-User** (пользовательский fetch по запросу), **Claude-SearchBot** (индексация/улучшение качества поисковых ответов). Отключение Claude-SearchBot и Claude-User прямо описывается как фактор снижения «видимости» в user-directed web search у пользователей. ⁵
- **Долгосрочная «видимость через обучение/grounding».** Здесь у вендоров есть отдельные механизмы «разрешить/запретить использование контента»: у OpenAI ⁶ это **GPTBot** (контент для обучения), у Google ⁷ — токен **Google-Extended** (обучение будущих поколений Gemini и grounding в Gemini Apps/Vertex) и он *не влияет* на включение/ранжирование в Google Search. ⁸

Из этого следует практическая логика Stage 2: показывать пользователю не «магические» метрики, а **проверяемые условия попадания в кандидаты (crawl/index/snippet) + качество/понятность/цитируемость контента + внешнюю верификацию и распределение упоминаний**, которые определяют вероятность выбора источников при web-retrieval (особенно когда системы делают query fan-out). ⁹

Программная часть

Ниже — то, что имеет смысл **анализировать автоматически** и **показывать пользователю** как «готовность сайта» к AI visibility. Важно: почти все пункты либо прямо указаны в официальных гайдах, либо являются прямой инженерной проекцией этих требований на измеримые проверки (HTTP/robots/meta/WAF).

Управление доступом

Для GEO-аудита полезно разделить доступ на три слоя, потому что у вендоров он разделен официально:

- **Search / surfacing** (попасть в ответы поиска с источниками)
- **User-initiated retrieval** (когда ассистент/агент открывает URL по запросу пользователя)
- **Training / model improvement** (использование контента для обучения будущих моделей)

Что показывать пользователю (вариант UI-артефакта): «матрица доступа» из *только фактов*:

- Видит ли сайт **OAI-SearchBot** и не блокируется ли он robots.txt. ¹⁰
- Разрешен ли трафик **OAI-SearchBot** на уровне хоста/CDN (и есть ли риск 403/429/челленджей). В OpenAI явно сказано, что для включения важно разрешить OAI-SearchBot и пропускать запросы с опубликованных IP-адресов. ¹¹
- Как настроены **GPTBot** (обучение) и связь «можно разрешить search, запретить training». OpenAI прямо описывает независимость этих настроек и различие ролей OAI-SearchBot vs GPTBot. ¹²
- Аккуратное пояснение про **ChatGPT-User**: он используется для пользовательских действий и не для автоматического краулинга; так как действия иницируются пользователем, robots.txt «может не применяться» — это важно, чтобы пользователь не думал, что «заблокировал всё одним роботом». ¹³
- Для Claude: статус **Claude-User** и **Claude-SearchBot** (и предупреждение, что их отключение может уменьшить видимость в user-directed web search). ¹⁴
- Для Google: доступ **Googlebot** (как минимум) и отдельно — решение, использовать ли **Google-Extended** для обучения/grounding в системах Gemini (с пояснением, что это не влияет на включение/ранжирование в Google Search). ¹⁵

Индексируемость и «право на сниппет»

Для Google AI Overviews / AI Mode официальная база такова: страница должна быть **проиндексирована** и **eligible for snippet**, дополнительных требований нет. ¹⁶
Отсюда в Stage 2 обязательны проверки, которые вы можете формализовать и скорить:

- Блокируется ли Googlebot (robots.txt / правила CDN). ¹⁷
- Возвращает ли страница **HTTP 200** и содержит ли «indexable content» (как минимум текст в поддерживаемом виде). Это прямо в технических требованиях Google Search. ¹⁸
- Нет ли управляемых блокировок сниппета/AI-использования: **nosnippet**, **max-snippet**, **data-nosnippet**, **noindex**. В документации Google эти директивы прямо привязаны к AI Overviews/AI Mode: `nosnippet` предотвращает использование контента как direct input; `max-snippet` ограничивает, сколько может быть использовано как direct input. ¹⁹
- Важно технически донести: такие директивы читаются только если краулер может зайти на страницу (то же правило действует и для meta/headers). ²⁰

Практический формат для пользователя: список «URL-кандидатов» (главные коммерческие/информационные страницы) и для каждой — статусы **Crawlable / Indexable / Snippet-eligible / Preview controls**.

Архитектура и «обнаруживаемость»

Даже без «спец-требований под AI» Google прямо перечисляет базовые практики, которые остаются важными для AI-фич: разрешить crawling (в т.ч. на уровне CDN), сделать контент легко находимым через internal links, обеспечить хороший page experience, иметь важный контент в текстовой форме, следить за соответствием structured data видимому тексту, обновлять Merchant Center/Business Profile. ¹⁶

Что это значит для Stage 2 как измеримые проверки:

- **Internal linking map**: есть ли путь до ключевых страниц за 2–3 клика, нет ли «висячих» страниц, закрытых от краулеров. ²¹
- **Текстовая доступность**: важная информация (определения, условия, цены, характеристики) должна быть в HTML-тексте, а не только в изображениях/JS-виджетах/скрытых табах. Google отдельно подчеркивает важность «important content in textual form». ¹⁷
- **Page experience** как фактор «не только один-два аспекта»: Google в guidance по people-first прямо говорит, что его системы стремятся «reward content that provides a good page experience» и рекомендует смотреть на опыт комплексно. ²²

Доступность для агентов

Это сильный «GEO-специфичный» пункт, который редко учитывают в SEO-аудитах, но он напрямую есть у OpenAI: для ChatGPT Agent в Atlas сказано, что «делать сайт более доступным» помогает агенту понимать его лучше; Atlas использует ARIA-теги (роли/лейблы) для интерпретации структуры и интерактивных элементов. ²³

Что анализировать и показывать:

- наличие корректных **ARIA roles/labels/states** на интерактивных элементах (кнопки, меню, формы) + топ-ошибки (критичные: «кнопка без имени», «инпут без label», «элемент кликабельный, но не семантический»). ²⁴
- показатель «agent-readiness»: доля ключевых интерактивных контролов с корректными доступными именами и ролями, с рекомендацией следовать практикам WAI-ARIA. ²⁴

Контекстная часть

Требование «контентно под AI» в официальных документах сформулировано не как «напишите вот так для LLM», а как **люди-первый, надежный, оригинальный контент**, который хорошо индексируется и годится для сниппетов/ссылок и выдерживает проверку источниками. На практике это и есть «контентная готовность к AI visibility», потому что и ChatGPT Search, и Google AI-фичи, и web-search у Claude строятся вокруг поиска, подбора источников и цитирования. ²⁵

Ниже — состав анализа, который можно формализовать и показывать пользователю как часть Stage 2.

Готовность контента быть выбранным как источник

У Google есть очень приземленная рамка самооценки «helpful, reliable, people-first content» — и ее удобно превратить в чек-метрики (не «мнения аудитора», а вопросы с сигналами на странице). В документе прямо перечислены вопросы качества (оригинальность, полнота, выход за очевидное, отсутствие переписывания, хорошее заглавие), вопросы экспертизы (доверие, прозрачность автора/сайта, источники, отсутствие проверяемых ошибок). ²²

Что включить в GEO-аудит (как минимальный «контентный скоринг»):

- **Originality & completeness:** есть ли на сайте страницы, которые дают «substantial, complete» описание продукта/категории, а не тонкие лендинги без фактуры. ²²
- **Descriptive headings:** является ли title/H1 действительно полезным описанием (Google прямо спрашивает об этом). ²⁶
- **Sourceability:** есть ли явные источники/доказательства/методология там, где вы даете факты (Google: «clear sourcing... background about the author or the site...»). ²²
- **Factual hygiene:** наличие легко проверяемых ошибок — отдельная проверка, потому что Google прямо выносит это в критерии. ²²

Важно для GEO-поддачи пользователю: не просто «оценка 7/10», а **примеры конкретных блоков**, которые:

- (а) наиболее вероятно могут быть процитированы (короткое определение/таблица/FAQ-ответ),
- (б) содержат риск (нет источников, спорное утверждение, непроверяемый суперлатив).

«Разрешение» на использование контента в AI-ответах

Это контринтуитивная часть, которую обязательно стоит подсветить как «блокеры видимости»:

- В Google AI Overviews/AI Mode ограничения сниппета (`nosnippet`, `data-nosnippet`, `max-snippet`) и `noindex` — это *не абстрактная SEO-настройка*, а прямое управление тем, может ли контент быть использован как input и сколько текста может попасть в AI-ответ. ²⁷
- В OpenAI, если вы блокируете OAI-SearchBot, ваш контент не будет включаться в summaries/snippets; при этом ссылка и заголовок теоретически могут всплыть как «просто линк» (если URL найден через стороннего провайдера/краулинг других страниц), и чтобы не было даже этого, предлагается `noindex` (с уточнением: мета-тег можно прочитать только при разрешенном крауле). ²³

В Stage 2 это стоит показать отдельным разделом «Visibility blockers»: какие директивы сейчас уменьшают вашу цитируемость/использование в AI-ответах и на каких URL.

Соответствие реальным запросам и «query fan-out»

Google прямо описывает, что AI Overviews/AI Mode могут делать query fan-out — несколько поисков по подтемам и источникам. ²⁸

Google Gemini-grounding (в API) описывает похожий механизм: модель сама генерирует один или несколько поисковых запросов, затем синтезирует ответ и возвращает structured citations. ²⁹

ChatGPT Search официально тоже делает query rewriting и может отправлять несколько уточняющих запросов к поисковым партнерам. ³⁰

Как превратить это в контент-анализ для GEO:

- Строите «prompt-кластеры» (как вы уже делаете на этапе 1), а на этапе 2 показываете **coverage-map**: какие страницы вашего сайта закрывают *подтемы*, которые поисковые/LLM-системы вероятно будут искать при fan-out. (Здесь вы не утверждаете «так и будет», вы показываете покрытие смысловых потребностей.) ³¹
- Для каждой ключевой подтемы показываете «лучший landing-URL на сайте» и качество блока, который может быть процитирован: есть ли четкое определение, списки характеристик, условия, примеры, FAQ.

Присутствие в интернете и внешние площадки

Официальные документы не дают «рецепта внешних площадок», но они однозначно показывают, что:

- В ChatGPT Search есть **источники и inline-цитаты**, отдельная панель Sources; и указывается, что ChatGPT Search использует сторонних провайдеров поиска и партнерский контент. ³²
- В Google AI-фичах ссылки — часть механики («surface relevant links»), а сами фичи строятся поверх Google Search (страница должна быть индексирована и eligible for snippet). ³³
- В Claude web search (как инструмент) ответ возвращается «with cited sources», а citations структурированы. ³⁴

Следовательно, на этапе 2 логично измерять не «присутствие вообще», а **какие внешние источники реально становятся опорными для ответов и насколько бренд там описуем и непротиворечив**.

Что именно анализировать по внешним источникам

Практичный набор артефактов, которые можно показать пользователю без лишних допущений:

- **Source-mix по каждому prompt-кластеру**: список доменов/страниц, которые используются как источники/цитаты в ответах (в режимах, где есть web search и citations). Основание: ChatGPT Search показывает citations и Sources; Claude web search всегда включает citations; Gemini-grounding возвращает `groundingMetadata` с источниками и связью «фрагмент → источник». ³⁵
- **Gap-анализ «owned vs earned»**: доля источников, где бренд контролирует формулировки (owned: сайт/доки/пресс-раздел), и доля, где контент внешний (earned: обзоры, статьи, базы, каталоги, обсуждения). Это не «фактор ранжирования», а диагностическая метрика: чем больше внешних независимых источников, тем устойчивее нарратив обычно в retrieval-сценариях. (В отчете можно маркировать это как наблюдаемую структуру источников, без обещаний про топ-5.) ³⁶
- **Consistency-check**: совпадают ли базовые факты о бренде (категория/позиционирование/описание/условия) между внешними источниками и вашим сайтом. Здесь полезно опираться на принцип «clear sourcing / background about the author or the site» как элемент доверия и на то, что AI-ответы строятся на источниках. ³⁷

Контроль и трекинг входящего трафика из AI-поиска

Это точка, которой обычно не хватает в GEO-аудитах, но она есть в официальных гайдах OpenAI: если издатель разрешает OAI-SearchBot, можно отслеживать переходы из ChatGPT; указывается, что в referral добавляется `utm_source=chatgpt.com`. ²³

На этапе 2 это можно показывать как:

- «Сейчас можно/нельзя измерять» (настроены ли UTM-правила, видны ли рефералы, нет ли редиректов, которые чистят UTM).
- «Канонические целевые URL» — чтобы источники в AI-ответах вели на правильные посадочные, а не на дубликаты.

Формат выдачи пользователю

Чтобы Stage 2 был «точным без галлюцинаций», интерфейс и отчет лучше строить вокруг трех типов утверждений:

- **Hard blockers** (факты): «бот заблокирован», «ответ 403», «noindex на странице», «nosnippet запрещает использовать контент в AI Overviews/AI Mode», «OAI-SearchBot требует разрешения и IP-allow на CDN». ³⁸
- **Eligibility signals** (факты + проверка): «страница индексируется/не индексируется», «eligible for snippet/не eligible из-за preview controls», «page experience issues». ³⁹
- **Improvement hypotheses** (помеченные как гипотезы): «усилить coverage подтем под query fan-out», «повысить долю внешних подтверждений». Здесь лучше явно маркировать как «вероятностное улучшение», потому что сами вендоры не дают гарантий топ-позиций. ⁴⁰

Рекомендуемая структура карточек/разделов Stage 2:

- «Доступность для AI-ботов» (robots.txt + CDN/WAF + статусы ответов + список заблокированных ключевых URL). ⁴¹
- «Индексируемость и сниппеты» (Googlebot/индекс/preview controls). ⁴²
- «Citable content» (страницы и блоки, которые чаще всего можно использовать как цитаты: определения, FAQ, таблицы условий; наличие источников и автора). ⁴³
- «Карта источников» (какие домены/страницы становятся источниками в ответах в режимах web search) + «дыры» (где бренд отсутствует в внешних источниках по ключевым кластерам). ³⁵
- «Agent readiness» (ARIA/доступность интерфейса для agentic-взаимодействия). ²⁴

Приоритизация работ

Ниже — порядок, в котором имеет смысл ранжировать рекомендации именно в GEO-логике «сначала попасть в кандидаты, потом улучшать вероятность выбора».

Сначала устранить блокиеры

- Разрешить OAI-SearchBot и пропускать его IP-диапазоны на CDN/хосте, иначе сайт может не попадать в ответы ChatGPT Search. ⁴⁴
- Проверить Claude-User и Claude-SearchBot: Anthropic прямо предупреждает, что отключение может снизить видимость в user-directed web search. ¹⁴

- Убедиться, что Googlebot не блокируется и страницы отдают 200 и indexable content; иначе вы автоматически теряете eligibility для AI Overviews/AI Mode как supporting links. ⁴⁵
- Пересмотреть `nosnippet` / `max-snippet` / `data-nosnippet` на коммерчески важных страницах, если цель — *попадать в AI Overviews/AI Mode*: эти директивы официально ограничивают/запрещают использование контента как direct input. ¹⁹

Затем усилить «цитируемость» и надежность контента

- Перепроверка ключевых страниц по Google people-first: оригинальность, полнота, ясный заголовок, источники/автор/обоснования, отсутствие проверяемых ошибок. ²²
- Обеспечить текстовую доступность важных фактов (не прятать в не-индексируемых слоях интерфейса). ¹⁷

Затем масштабировать coverage по подтемам

- Строить контент-карту под query fan-out (подтемы/сценарии/сравнения/«как выбрать»), потому что AI-фики и search-grounding используют несколько запросов и несколько источников. ³¹

Отдельно решить политику «обучать/не обучать»

- У OpenAI можно разрешить search (OAI-SearchBot), но запретить training (GPTBot), и это официально описано как независимые настройки. ¹²
- У Google Google-Extended управляет обучением будущих Gemini и grounding в Gemini Apps/Vertex и не влияет на Search-включение/ранжирование — это важно объяснять пользователю, чтобы не было ложных ожиданий «включу Google-Extended → вырасту в Overviews». ⁴⁶
- У Anthropic нет опубликованных IP-диапазонов, и они предупреждают, что IP-блокировки могут не гарантировать opt-out, потому что мешают прочитать robots.txt. ¹⁴

¹ ² ¹¹ ²⁵ ³⁰ ³⁵ ⁴⁰ ⁴⁴ ChatGPT search | OpenAI Help Center

<https://help.openai.com/en/articles/9237897-chatgpt-search>

³ ⁶ ⁷ ⁹ ¹⁵ ¹⁶ ¹⁷ ²¹ ²⁷ ²⁸ ³¹ ³³ ³⁹ ⁴² AI Features and Your Website | Google Search Central | Documentation | Google for Developers

<https://developers.google.com/search/docs/appearance/ai-features>

⁴ ¹⁹ ²⁰ ³⁸ Robots Meta Tags Specifications | Google Search Central | Documentation | Google for Developers

<https://developers.google.com/search/docs/crawling-indexing/robots-meta-tag>

⁵ ¹⁴ Does Anthropic crawl data from the web, and how can site owners block the crawler? | Claude Help Center

<https://support.claude.com/en/articles/8896518-does-anthropic-crawl-data-from-the-web-and-how-can-site-owners-block-the-crawler>

⁸ ¹² ¹³ ⁴¹ Overview of OpenAI Crawlers

<https://developers.openai.com/api/docs/bots/>

¹⁰ ²³ ²⁴ Publishers and Developers - FAQ | OpenAI Help Center

<https://help.openai.com/en/articles/12627856-publishers-and-developers-faq>

¹⁸ ⁴⁵ <https://developers.google.com/search/docs/essentials/technical>

<https://developers.google.com/search/docs/essentials/technical>

22 26 37 43 <https://developers.google.com/search/docs/fundamentals/creating-helpful-content>
<https://developers.google.com/search/docs/fundamentals/creating-helpful-content>

29 **Grounding with Google Search | Gemini API | Google AI for Developers**
<https://ai.google.dev/gemini-api/docs/google-search>

32 36 **Introducing ChatGPT search | OpenAI**
<https://openai.com/index/introducing-chatgpt-search/>

34 **Web search tool - Claude API Docs**
<https://platform.claude.com/docs/en/agents-and-tools/tool-use/web-search-tool>

46 **Google's common crawlers | Google Crawling Infrastructure | Crawling infrastructure | Google for Developers**
<https://developers.google.com/crawling/docs/crawlers-fetchers/google-common-crawlers>