

Подготовка сайта и данных для ИИ-агентов

Сводный чек-лист: машиночитаемая разметка, технические данные и прямой доступ для автономных агентов и систем закупки. Объединённая редакция двух исходных чек-листов с дополнениями · 2026

Как читать

□ — пункт для проверки. Справа — приоритет: **КРИТИЧНО** — без этого агент не получит данные или получит неверные; **ВЫСОКИЙ** — заметно влияет на выбор и цитирование; **СРЕДНИЙ** — доводка. + **НОВОЕ** — пункт добавлен в этой редакции, в исходных чек-листах его не было.

01 Главные принципы

- **Единый источник правды.** JSON-LD, таблицы на странице, lms.txt, API-JSON и фиды генерируются из одного файла данных, а не правятся вручную по отдельности — иначе цифры начинают расходиться. + **НОВОЕ** **КРИТИЧНО**
- **Каждый факт — в отдельном поле.** Всё, что влияет на решение о закупке, не прячется в маркетинговом тексте. **КРИТИЧНО**
- **Никаких противоречий между каналами:** цена, наличие, характеристики одинаковы на сайте, в фидах, на маркетплейсах. **КРИТИЧНО**
- **Лучше меньше, но точно.** Неполная или противоречивая разметка хуже её отсутствия: агент воспроизводит ошибку. Не заполнять поля «для галочки» и не выдумывать значения.
В одном из источников приводилась цифра «штраф ~18% к цитированию» — источник не указан, как точное значение не использовать. **ВЫСОКИЙ**
- **Одна версия для людей и ботов.** Никакого клоакинга: агент и человек видят один и тот же контент. + **НОВОЕ** **ВЫСОКИЙ**

02 Сервер, доступность и индексация

Самая частая причина «невидимости» для агентов — не разметка, а то, что бот просто не получает страницу.

- HTTPS, ответ **HTTP 200**, не более одного редиректа; переезды — только 301. **КРИТИЧНО**
- **SSR:** весь контент и JSON-LD есть в сыром HTML первого ответа, без выполнения JavaScript. **КРИТИЧНО**
- Нет запрещающих X-Robots-Tag и meta robots="noindex". **КРИТИЧНО**
- **WAF и антибот-защита** (Cloudflare Bot Fight Mode, DDoS-Guard, Qrator и т.п.) не блокируют AI-краулеров. Проверяется отдельно: robots.txt может всё разрешать, а фаервол отдавать 403. + **НОВОЕ** **КРИТИЧНО**
- Честные статусы: 404/410 для удалённых позиций, никаких «мягких 404» (200 OK на пустой странице). **ВЫСОКИЙ**
- Корректные Content-Type и кодировка UTF-8 для HTML, JSON, Markdown, YAML. **ВЫСОКИЙ**
- Скорость: TTFB до ~600 мс, разумный вес HTML. Агенты обрывают медленные запросы по таймауту. + **НОВОЕ** **ВЫСОКИЙ**
- При перегрузке — ответ 429 с Retry-After, а не 403 или 5xx. + **НОВОЕ** **СРЕДНИЙ**

☐	Проверка подлинности ботов по reverse DNS или опубликованным IP-диапазонам — чтобы не раздавать ресурсы подделкам под GPTBot. + НОВОЕ	СРЕДНИЙ
03 Файлы уровня сайта		
☐	/robots.txt — явный Allow для GPTBot, OAI-SearchBot, ChatGPT-User, ClaudeBot, Claude-User, Claude-SearchBot, PerplexityBot, CCBot, Google-Extended, Applebot-Extended; нет блокировки по User-agent: *; строка Sitemap:.	КРИТИЧНО
☐	Для рунета в robots.txt — также YandexBot и YandexAdditional; сайт подтверждён в Яндекс.Вебмастере. + НОВОЕ	КРИТИЧНО
☐	/sitemap.xml — валидный XML, <lastmod> у каждой страницы с реальной датой изменения, а не датой генерации.	КРИТИЧНО
☐	/llms.txt — краткое описание сайта, правила использования контента, ссылки на ключевые разделы. llms.txt — соглашение сообщества, а не стандарт: крупные поисковики официально его не используют. Дешёвый пункт, но не замена JSON-LD.	ВЫСОКИЙ
☐	/llms-full.txt — полный структурированный Markdown-каталог: линейки, характеристики, рецептурные особенности, применение.	ВЫСОКИЙ
☐	Уведомление поисковиков об изменениях: IndexNow (Яндекс, Bing) и «Переобход страниц» в Вебмастере. + НОВОЕ	ВЫСОКИЙ
☐	/sitemap.md или /.well-known/sitemap.md — Markdown-карта сайта для агентов.	СРЕДНИЙ
☐	/AGENTS.md — инструкции для агентов: точки входа, как получить цену, единицы измерения, оговорки о достоверности.	СРЕДНИЙ
☐	/.well-known/ucrp — профиль Universal Commerce Protocol. Только если на сайте есть онлайн-продажа и checkout.	СРЕДНИЙ
04 HTML-разметка страницы		
☐	<html lang="ru">, уникальный <title> (до ~60 символов), <link rel="canonical"> на каждой странице.	КРИТИЧНО
☐	Характеристики — HTML-таблицей (<table>, <caption>, <th scope>), а не картинкой, баннером или скриншотом.	КРИТИЧНО
☐	meta description — не меньше 50 символов, оптимально 120–160.	ВЫСОКИЙ
☐	Open Graph: og:title, og:description, og:image, og:url, og:type, og:locale.	ВЫСОКИЙ
☐	Семантика: один h1, иерархия h2–h3 (не меньше 3 заголовков), теги main, header, nav, section, article, footer.	ВЫСОКИЙ
☐	Осмысленный alt у всех изображений: что изображено, а не «фото1». + НОВОЕ	ВЫСОКИЙ
☐	Числа всегда с единицами и в одном формате по всему сайту: «–40 °C», «5–100 кг», единицы СИ. + НОВОЕ	ВЫСОКИЙ
☐	Глоссарий терминов отрасли (плюс разметка DefinedTermSet) — чтобы агент понимал специфическую лексику.	ВЫСОКИЙ
☐	Доля текста в HTML (text-to-HTML ratio) выше 15% — страница содержит реальный контент, а не шаблонный код.	СРЕДНИЙ
☐	Видимые и размеченные даты публикации и обновления (datePublished, dateModified). + НОВОЕ	СРЕДНИЙ

05 Структурированные данные: Schema.org / JSON-LD

- ☐ Только JSON-LD — не Microdata и не RDFa; в `<script type="application/ld+json">` в `<head>`, в серверном HTML. **КРИТИЧНО**
- ☐ Связный граф: сущности ссылаются друг на друга через `@id`, а не дублируются (Product → brand → Organization; Article → author → Person → Organization). **КРИТИЧНО**
- ☐ Organization: `legalName`, `logo`, `address`, `contactPoint`, `sameAs` (соцсети, Wikidata, отраслевые реестры) и реквизиты юрлица через `identifier` (ИНН, ОГРН как `PropertyValue`). **КРИТИЧНО**

Тип	Зачем	Приоритет
Organization	Идентификация компании и бренда, реквизиты, <code>sameAs</code>	Критично
Product + Offer / AggregateOffer	Цена, валюта, наличие, SKU, ступени объёма	Критично
potentialAction	Что агент может сделать: <code>BuyAction</code> , <code>QuoteAction</code> , <code>ContactAction</code> , <code>SearchAction</code>	Критично
Article + Person	Авторство, даты, сигналы экспертизы (E-E-A-T) — для статей и новостей	Критично*
Service	Нематериальные услуги: что делает, кто оказывает, где доступно	Высокий
FAQPage	Прямое сопоставление вопрос → ответ	Высокий
HowTo	Пошаговые инструкции	Высокий
DefinedTermSet + новое	Машиночитаемый глоссарий	Высокий
MerchantReturnPolicy, OfferShippingDetails	Условия возврата и доставки	Высокий
ProductGroup + hasVariant + новое	Варианты (цвет, размер) как одна связанная группа	Высокий
IndividualProduct	Конкретный экземпляр с серийным номером (для уникальных изделий)	Средний
WebSite, WebPage, BreadcrumbList	Структура и иерархия сайта	Средний
Speakable	CSS-селекторы блоков для голосового ответа	Средний
ImageObject + новое	Изображения с подписью, автором и лицензией	Средний

* Article + Person критичны только при наличии статей, новостей, блога. Google с 2023 г. почти перестал показывать rich results для FAQ и HowTo в выдаче, но для LLM и агентов эта разметка по-прежнему полезна.

06 Продуктовые и технические данные

Для технических и сырьевых продуктов агенту нужны не описания, а параметры, пригодные для сравнения и расчёта.

- ☐ Идентификаторы: `sku`, `mpn`, `brand`; `gtin` — если он реально присвоен. Не выдумывать GTIN, если его нет. **КРИТИЧНО**
- ☐ Параметры — массив `additionalProperty` из `PropertyValue`: `name` — полное название без аббревиатур («Вязкость по Брукфильду»), `value` — строго число (4500), `unitCode` — код UN/CEFACT, `unitText` — человекочитаемая единица. **КРИТИЧНО**
- ☐ Диапазоны — через `minValue` / `maxValue`, а не строкой «5–100». **КРИТИЧНО**

☐	Документация: TDS (техническое описание), SDS/паспорт безопасности (ГОСТ 30333), сертификаты, ТУ — текстовые PDF с выделяемым слоем, стабильные URL, дата и номер версии.	КРИТИЧНО
☐	Метод и условия испытания для каждого значения: стандарт (ISO, ГОСТ), температура, состояние образца — через measurementTechnique или description. Без этого число неинтерпретируемо. + НОВОЕ	ВЫСОКИЙ
☐	Условия переработки и совместимость: совместимые инициаторы/отвердители, температура формования, время гелеобразования, жизнеспособность, срок и условия хранения.	ВЫСОКИЙ
☐	Варианты (цвет, размер, исполнение) — отдельные атрибуты с понятными названиями, без сокращений.	ВЫСОКИЙ
☐	Таксономия отражает суть продукта, без категорий «Другое» и «Разное».	ВЫСОКИЙ
☐	Применение, ограничения, предупреждения, типовые ошибки — отдельными размеченными полями (FAQ, HowTo), а не только в тексте.	ВЫСОКИЙ
☐	Атрибуты для сравнения: аналоги, замены, совместимые комплектующие и аксессуары.	СРЕДНИЙ

07 Коммерческие и логистические параметры

☐	Наличие через ItemAvailability: InStock, PreOrder, OutOfStock, MadeToOrder, в реальном времени.	КРИТИЧНО
☐	Если публичной цены нет (B2B, под T3): не ставить price = 0 и не выдумывать цену; объявить QuoteAction / ContactAction с целевым URL и явно написать об этом в lms.txt и AGENTS.md. + НОВОЕ Google Merchant Center без цены товар не примет — это нормальное ограничение для продуктов под заказ.	КРИТИЧНО
☐	Цена: price + priceCurrency; оптовые ступени (от 1 шт., от 100 кг, от контейнера) — AggregateOffer / UnitPriceSpecification с eligibleQuantity.	ВЫСОКИЙ
☐	Минимальная партия и срок изготовления (eligibleQuantity, deliveryLeadTime). + НОВОЕ	ВЫСОКИЙ
☐	Логистика: масса брутто/нетто, габариты тарного места, объём упаковки (м³), класс опасности ADR, OfferShippingDetails.	ВЫСОКИЙ
☐	Связка SKU ↔ variant ID ↔ checkout ID ↔ order ID без разрывов (для e-commerce).	ВЫСОКИЙ
☐	priceValidUntil — срок действия цены. + НОВОЕ	СРЕДНИЙ
☐	MerchantReturnPolicy — условия возврата (для B2C обязательно).	СРЕДНИЙ

08 API и машинные интерфейсы

☐	Открытый эндпоинт /api/v1/products/{id}: чистый JSON со всеми характеристиками, без капчи и авторизации, с открытым CORS.	ВЫСОКИЙ
☐	Детерминированные URL: /catalog/{category}/{series}/{sku}, без сессионных ID и дублей.	ВЫСОКИЙ
☐	Форма заявки, пригодная для агента: документированный POST-эндпоинт с описанием полей; honeypot вместо капчи; e-mail как запасной канал. + НОВОЕ	ВЫСОКИЙ
☐	OpenAPI / Swagger-схема, связанная со страницами документации и AGENTS.md.	СРЕДНИЙ
☐	Markdown-зеркала страниц с <link rel="alternate" type="text/markdown">.	СРЕДНИЙ

☐ Content negotiation: по заголовку Accept: text/markdown сервер отдаёт Markdown-версию.	СРЕДНИЙ
☐ Атрибуция агентских обращений на сервере: логирование User-Agent и источника, поле «источник» в заявке — без опоры на cookies и пиксели.	СРЕДНИЙ

09 Правовое и доверие

Раздел полностью добавлен в этой редакции: для агента это сигналы надёжности поставщика, для компании — обязательные требования.

☐ Политика обработки персональных данных и согласие в форме заявки (152-ФЗ для сайтов в РФ). + НОВОЕ	КРИТИЧНО
☐ Реквизиты юрлица в открытом виде на странице и в разметке: полное наименование, ИНН, ОГРН, юридический адрес. + НОВОЕ	ВЫСОКИЙ
☐ Контакты, которые реально отвечают: e-mail и телефон совпадают на сайте, в разметке, llms.txt и справочниках. + НОВОЕ	ВЫСОКИЙ
☐ Правила цитирования контента агентами (в llms.txt / AGENTS.md) и оговорка о типовом характере значений. + НОВОЕ	СРЕДНИЙ

10 Тестирование и валидация

☐ Schema Markup Validator (validator.schema.org), Google Rich Results Test, валидатор микроразметки Яндекс.Вебмастера.	КРИТИЧНО
☐ Проверка сырого HTML: curl -s https://site.ru/ grep application/ld+json. Если разметка появляется только после JS — переделать.	КРИТИЧНО
☐ Проверка глазами бота: curl -A "GPTBot" -I https://site.ru/ (и ClaudeBot) — ответ 200, а не 403 от файрвола.	КРИТИЧНО
☐ Live-тест: задать Claude / ChatGPT / Perplexity с поиском конкретные вопросы о продукте. Если ИИ догадывается, а не знает — данных не хватает.	ВЫСОКИЙ
☐ Сверка консистентности: одни и те же цифры в таблице на странице, JSON-LD, llms.txt, API и фидах. + НОВОЕ	ВЫСОКИЙ
☐ Проверка Markdown-выдачи: curl -H "Accept: text/markdown" https://site.ru/. + НОВОЕ	СРЕДНИЙ
☐ Ежемесячный разбор логов: какие AI-боты заходят, какие страницы и коды ответа получают. + НОВОЕ	СРЕДНИЙ

11 Чего избегать

Ошибка	Как правильно
Microdata / RDFa	Только JSON-LD
Разметка, вставляемая клиентским JS	Только server-side
Дублирование сущностей	Ссылки через @id
Разные цены и наличие на сайте и в фидах	Единый источник данных
Критичные факты только в тексте или на картинке	Отдельные поля и HTML-таблицы
Категории «Другое», «Разное»	Осмысленная таксономия
Выдуманные GTIN, рейтинги, отзывы (aggregateRating без реальных отзывов) + НОВОЕ	Заполнять только реальными данными
Разный контент для ботов и людей (клоакинг) + НОВОЕ	Одна версия страницы

Ошибка	Как правильно
Сканы вместо текстовых PDF для TDS/SDS + новое	PDF с текстовым слоем
Файрвол, молча режущий AI-ботов + новое	Отдельная проверка по User-Agent

12 Порядок внедрения

Добавлено в этой редакции: последовательность, в которой пункты дают наибольший эффект.

Этап	Что сделать
1. База (1-2 дня)	Доступность для ботов (robots.txt, WAF, SSR, коды ответа), canonical/lang/title, sitemap.xml, JSON-LD Organization + Product/Service, характеристики HTML-таблицей, правовые документы.
2. Данные (1 неделя)	additionalProperty с единицами UN/CEFACT и методами испытаний, FAQ и глоссарий с разметкой, llms.txt и llms-full.txt, TDS/SDS в текстовых PDF, QuoteAction или Offer.
3. Машинный доступ	JSON-API и OpenAPI, Markdown-зеркала и content negotiation, AGENTS.md, IndexNow, форма с POST-эндпоинтом.
4. Контроль (постоянно)	Валидаторы, live-тесты в ИИ, сверка консистентности, разбор логов ботов, обновление lastmod и dateModified при правках.

13 Приложение: коды единиц UN/CEFACT (Rec. 20)

Самые частые для технических и сырьевых продуктов. Полный перечень — UN/CEFACT Recommendation 20.

Величина	Единица	unitCode
Масса	кг / г	KGM / GRM
Длина	м / см / мм	MTR / CMT / MMT
Площадь / объём	м ² / м ³	MTK / MTQ
Объём жидкости	л	LTR
Температура	°C	CEL
Доля	%	P1
Давление, прочность	МПа	MPA
Динамическая вязкость	мПа·с (= сПз)	C7
Плотность	кг/м ³ / г/см ³	KMQ / 23
Сила	Н	NEW
Время	с / мин / ч	SEC / MIN / HUR