

TileScore AI

Руководство по подготовке и импорту данных
для ИИ-консультанта (Nanobot)

Версия 1.0 — Апрель 2026

Этот документ описывает форматы, структуру и требования к данным, которые принимает система ИИ-консультанта TileScore для обучения и работы с базой знаний клиента.

Содержание

1. Обзор архитектуры
2. Форматы импорта данных
3. Документы базы знаний (KB Documents)
4. Каталогные данные (KB Catalog)
5. Автозаполнение с сайта (URL Autofill)
6. Тренировочные Q&A;
7. Требования и лимиты
8. Процесс обработки данных
9. Рекомендации по подготовке
10. Шаблоны и примеры

1. Обзор архитектуры

ИИ-консультант TileScore (Nanobot) использует архитектуру **RAG** (Retrieval-Augmented Generation) для предоставления точных ответов на основе данных клиента. Система сочетает векторный поиск с полнотекстовым для максимальной релевантности.

Компоненты системы

- **Embedding Engine** — DashScope text-embedding-v3, размерность 1024
- **Векторный поиск** — pgvector (cosine similarity), порог 0.3
- **Полнотекстовый поиск (FTS)** — PostgreSQL tsvector с русской морфологией
- **Каскадный поиск** — текущий тайл → дочерние → родительские → гостевые
- **Чанкинг** — 500 символов на чанк, 100 символов перекрытие

Приоритеты поиска

Если RAG или каталожный поиск возвращают результаты, данные из Wiki исключаются из контекста. Скорректированные ответы (category=correction) имеют наивысший приоритет.

2. Форматы импорта данных

Система принимает данные через 4 канала. Каждый канал оптимизирован для определённого типа контента.

Канал	Форматы	Назначение	Лимит
KB Documents	PDF, TXT, MD, JSON	Тексты, инструкции, описания	10 МБ / файл, 100 файлов
KB Catalog	XLSX, XLS, CSV	Товары, услуги, прайсы	50 000 строк
URL Autofill	URL сайтов	Автопарсинг содержимого	5 сайтов, 20 подстраниц
Q&A; Training	Интерфейс	Вопрос-ответ пары	Без лимита

3. Документы базы знаний (KB Documents)

Текстовые документы — основной способ наполнения базы знаний. Система принимает неструктурированный текст и автоматически разбивает его на чанки для индексации.

Поддерживаемые форматы

Формат	Расширение	Особенности
Текст	.txt	Простой текст UTF-8. Рекомендуемый формат.
Markdown	.md	Markdown-разметка. Заголовки используются как разделители чанков.
PDF	.pdf	Текст извлекается клиент-стороной. Сканы НЕ поддерживаются.
JSON	.json	Валидируется и форматируется. Подходит для структурированных данных.

Требования к содержанию

- Минимум 10 символов, максимум 500 000 символов
- Кодировка UTF-8
- PDF: только текстовые (не сканы). Извлечение текста происходит на стороне клиента
- Избыточные пробелы и переносы строк нормализуются автоматически
- Встроенная проверка на спам

Пример TXT-документа

■ ■■■■■■■■ ■■■■■■■■ «■■■■■■■» ■■■■■■■■ ■ 2010 ■■■■. ■■
■■■■■■■■■■■■■■■■■■ ■■... ■■■■ ■■■■■■ 1. ■■■■■■■■■■■■■■■■ –
■■■■■■■■■■■■■■■■■■ ■■■■■■... 2. ■■■■ – ■■■■■■■■■■ ■■■■■■■■... 3.
■■■■■■■ – ■■■■■■■■■■■■ ■■■■■■■■... ■■■■■■■■: +7 (999) 123-45-67,
info@example.com

Пример JSON-документа

```
{ "company": "████████", "services": [ {"name": "████████████████████",  
"price": "██ 5000 ███"}, {"name": "██████", "price": "██ 15000 ███"} ],  
"contacts": {"phone": "+7 999 123-45-67"} }
```

4. Каталогные данные (KB Catalog)

Каталожный режим предназначен для структурированных данных: товары, услуги, прайс-листы. Данные загружаются в формате XLSX/CSV с маппингом колонок на поля системы.

Обязательные поля

Поле системы	Описание	Пример
name (Наименование)	Название товара/услуги	Перчатки нитриловые
code (Код/Артикул)	Уникальный идентификатор	PN-001

Опциональные поля

Поле системы	Описание	Пример
description (Описание)	Подробное описание	Нитриловые перчатки размер М...
category (Категория)	Группировка товаров	СИЗ / Средства защиты рук
price (Цена)	Цена в произвольном формате	450 руб / шт
unit (Единица)	Единица измерения	шт, пара, кг
metadata (Метаданные)	Доп. характеристики (JSON)	{"color":"синий","size":"М"}

Процесс загрузки

- 1. Подготовьте файл XLSX или CSV с данными
- 2. Загрузите файл через интерфейс «База знаний» → «Каталог»
- 3. Система покажет предпросмотр колонок
- 4. Сопоставьте колонки файла с полями системы (маппинг)
- 5. Подтвердите импорт

Пример структуры CSV

[illegible]

Важно: разделитель CSV — точка с запятой (;) или запятая (,). Система определяет разделитель автоматически. Кодировка UTF-8.

Поиск по каталогу

Система использует трёхуровневый поиск по каталогу:

- **Уровень 1** — точное совпадение кода/артикула (ранг 100)
- **Уровень 2** — полнотекстовый поиск (FTS) с русской морфологией
- **Уровень 3** — ILIKE-фоллбэк по ключевым словам (≥ 4 символов)

5. Автозаполнение с сайта (URL Autofill)

Система может автоматически извлечь контент с вашего сайта и добавить его в базу знаний. Это удобно для быстрого старта.

Параметры

- До **5 сайтов** (доменов)
- До **20 подстраниц** с каждого сайта
- Извлекается текстовый контент (HTML-теги удаляются)
- Навигация, футеры и служебные блоки фильтруются

Как использовать

- 1. Откройте вкладку «AI Training» в настройках тайла
- 2. Выберите «Автозаполнение с сайта»
- 3. Введите URL главной страницы сайта
- 4. Система просканирует страницы и создаст документы в KB

6. Тренировочные Q&A;

Структурированные пары «вопрос — ответ» позволяют точно обучить консультанта на типовые запросы клиентов.

Категории вопросов

Категория	Назначение	Примеры
Паспорт	Базовая информация о компании	Что такое ваша компания? Чем вы занимаетесь?
Q&A;	Частые вопросы клиентов	Какие способы оплаты? Сроки доставки?
Funnel	Вопросы воронки продаж	Как оформить заказ? Есть ли скидки?
Correction	Исправления ответов ИИ	Если ИИ ответил неверно — добавьте коррекцию

Совет: Коррекции (correction) имеют наивысший приоритет в поиске. Используйте их для исправления систематических ошибок ИИ.

7. Требования и лимиты

Параметр	Значение
Макс. размер файла (KB Document)	10 МБ

Параметр	Значение
Макс. количество документов на тайл	100
Мин. длина контента	10 символов
Макс. длина контента	500 000 символов
Кодировка файлов	UTF-8
Форматы документов	PDF, TXT, MD, JSON
Форматы каталогов	XLSX, XLS, CSV
Макс. строк каталога	50 000
URL Autofill: макс. сайтов	5
URL Autofill: макс. подстраниц	20 на сайт
PDF: сканы (OCR)	Не поддерживается

8. Процесс обработки данных

Этап 1: Загрузка и валидация

- Проверка формата, размера и кодировки файла
- Нормализация пробелов и переносов строк
- Проверка на спам и дубликаты
- Проверка лимитов (количество документов, размер)

Этап 2: Чанкинг (разбиение на фрагменты)

- Размер чанка: **500 символов**
- Перекрытие (overlap): **100 символов**
- Разбиение по границам предложений (при возможности)
- Каждый чанк сохраняется с метаданными: порядковый номер, ID документа

Этап 3: Векторизация (Embedding)

- Модель: **DashScope text-embedding-v3**
- Размерность вектора: **1024**
- Каждый чанк преобразуется в числовой вектор
- Вектор сохраняется в pgvector для быстрого поиска

Этап 4: Индексация

- Создание полнотекстового индекса (tsvector) для FTS
- Создание векторного индекса для cosine similarity
- Для каталогов — дополнительная индексация по коду и категории

Схема обработки

■■■■ → ■■■■■■■■■■ → ■■■■■■■■ (500/100) → Embedding (1024d) → pgvector + FTS ■■■■■■

9. Рекомендации по подготовке данных

Документы (KB)

- Используйте **ТХТ** или **MD** — это самые надёжные форматы
- Структурируйте текст заголовками и абзацами
- Один документ = одна тема (услуга, продукт, раздел)

- Избегайте таблиц в текстовых файлах — используйте каталог
- Указывайте контактные данные явно

Каталоги

- Первая строка файла — заголовки колонок
- Код (артикул) должен быть уникальным
- Используйте категории для группировки
- Описание — чем подробнее, тем точнее поиск
- Цены указывайте в одном формате

Q&A;

- Формулируйте вопросы так, как их задают клиенты
- Ответы — краткие и конкретные
- Используйте категорию «Correction» для исправления ошибок ИИ

10. Шаблоны и примеры

Шаблон XLSX/CSV для каталога

Код	Наименование	Категория	Описание	Цена	Ед.
ART-001	Товар А	Категория 1	Описание товара А	1500	шт
ART-002	Товар В	Категория 1	Описание товара В	2300	шт
SRV-001	Услуга X	Услуги	Описание услуги X	5000	час

Шаблон ТХТ для базы знаний

[illegible]

Шаблон Q&A;

Категория	Вопрос	Ответ
Паспорт	Чем занимается ваша компания?	[Краткое описание]
Q&A;	Какие способы оплаты вы принимаете?	Наличные, карта, перевод
Q&A;	Есть ли доставка?	Да, доставка по городу бесплатно от 5000 руб
Funnel	Как оформить заказ?	Позвоните или оставьте заявку на сайте