

Политика в отношении источников данных и интеллектуальной собственности

Версия 1.0 · Действует с 1 ноября 2026 · Владелец: Александр Сенкевич, katelabs, Белград, Сербия · katelabs.io

1. Цель и область применения

Настоящая политика описывает, откуда берётся датасет katelabs, что в нём хранится, какими правами на него обладает katelabs и на какие права katelabs не претендует. Она распространяется на каждую запись, предоставленную в составе образца, пробного доступа, лицензии или по API-ключу, а также на исследовательские заметки, публикуемые на katelabs.io.

2. Что представляет собой датасет

Датасет представляет собой свод заголовков англоязычных новостей о криптовалютах, непрерывно собираемых с 4 октября 2024 года. Каждая запись содержит заголовок, краткое резюме, предоставленное издателем (если оно есть), канонический URL, название и домен издателя, время публикации, время записи в базу, активы, к которым относится заголовок, и два значения, формируемые katelabs: оценку ожидаемого влияния на цену от -1 до 1 и тип (проактивный или реактивный). Оценки и типы присваиваются однократно, при записи в базу, автоматизированным конвейером на основе языковой модели и впоследствии никогда не пересматриваются. Время присвоения оценки хранится вместе с каждой записью.

3. Источники

- 1. Только публичные материалы.** katelabs собирает заголовки и резюме, которые издатели делают общедоступными: RSS- и Atom-ленты, новостные карты сайта (sitemap), публичные JSON-эндпоинты и лицензированные сервисы агрегации новостей. Ничто не собирается из источников, требующих входа в систему, платной подписки, членства в закрытой группе или канале в мессенджере, и скрейпинг текстов статей не производится.
- 2. Условия издателей.** Сбор осуществляется с соблюдением директив robots и условий использования лент каждого издателя. Издатель, попросивший исключить его, удаляется из сбора в течение пяти рабочих дней; записи, уже переданные лицензиатам, не отзываются, поскольку датасет является point-in-time (на момент времени), однако издатель исключается из всех последующих поставок.
- 3. Отсутствие персональных данных заложено в архитектуру.** Записи не содержат данных о подписчиках, читателях или пользователях. Имена встречаются только там, где издатель поместил их в заголовок. katelabs не обогащает записи персональными данными из каких-либо иных источников.

4. Прозрачность. Каждая запись содержит домен и URL издателя, поэтому лицензиат всегда может проследить строку до её публичного источника. Список доменов, входящих в датасет, предоставляется любому лицензиату по запросу.

katelabs · Политика в отношении источников данных и интеллектуальной собственности · v1.0 katelabs.io

4. Что katelabs хранит, а что — нет

Хранится: текст заголовка, резюме, которое издатель прилагает к элементу ленты (как правило, от одного до трёх предложений), URL, издатель, временные метки, теги активов, оценка, тип. Не хранится и никогда не распространяется: тексты статей, изображения, профили авторов, комментарии и любой иной контент за пределами самого элемента ленты. Параметры отслеживания удаляются из URL перед сохранением.

5. Интеллектуальная собственность

- 1. Контент издателей.** Заголовки и резюме остаются собственностью их издателей. katelabs не претендует на авторские права на них и воспроизводит их в виде коротких выдержек с указанием источника в целях индексации, классификации и количественных исследований. Права лицензиата на эти материалы ограничены внутренним анализом; лицензиат не вправе переиздавать текст заголовков в качестве новостного продукта.
- 2. Производные данные katelabs.** Оценки, типы, теги активов, временные метки записи в базу и присвоения оценки, новостной индекс по каждому активу, структура датасета и принципы отбора, конвейер сбора, промпты, методология оценки качества и все исследовательские заметки являются интеллектуальной собственностью katelabs. Они предоставляются по лицензии и остаются в собственности katelabs.
- 3. Результаты лицензиата.** Модели, сигналы, стратегии, отчёты и иные результаты, которые лицензиат создаёт на основе данных, принадлежат лицензиату. Исходные записи и производные поля не могут извлекаться из этих результатов и передаваться третьим лицам в качестве данных.
- 4. Открытый фрагмент.** Трёхмесячный фрагмент датасета опубликован на Hugging Face на условиях, указанных в его карточке датасета; эти условия распространяются только на данный фрагмент.

6. Гарантия point-in-time

После записи оценка, тип и временные метки записи не изменяются. Исправления метаданных издателя (например, исправленный URL) вносятся в виде новых полей, никогда — путём перезаписи оригинала, и документируются в журнале изменений датасета. Известные дефекты конвейера, включая пробелы в покрытии, раскрыты в исследовательской заметке 4 и в словаре данных, прилагаемом к каждой поставке.

7. Удаление по запросу и вопросы

Издатель, автор или правообладатель, считающий, что материал собирается или воспроизводится ненадлежащим образом, может написать на **aler.senkevich3301@gmail.com**, указав соответствующий URL. katelabs подтверждает

получение в течение двух рабочих дней и разрешает вопрос в течение десяти. Лицензиаты получают уведомление о любом удалении источника, затрагивающем их поставки.

katelabs · Политика в отношении источников данных и интеллектуальной собственности · v1.0

katelabs.io

8. Пересмотр

Настоящая политика пересматривается не реже одного раза в год, а также при каждом существенном изменении метода сбора. Актуальная версия всегда доступна по адресу katelabs.io/docs.