

**ПЛАТФОРМА ДЛЯ
ПОДГОТОВКИ ДАТАСЕТОВ
С ИСПОЛЬЗОВАНИЕМ ИИ**

ИНСТРУМЕНТ ДЛЯ ПОДГОТОВКИ ДАТАСЕТОВ



Платформа для создания датасетов для обучения нейросетей.

Единый инструмент для разметки всех основных типов данных:

изображения

аудио

видео

текст



Опытные разметчики выполняют любой объем работы с нужной вам скоростью из любой точки мира в режиме 24/7



Высокая точность разметки данных с помощью встроенных методов и инструментов контроля качества



Индивидуальные сроки и стоимость выполнения задач под ваш проект



Возможность создать полностью автоматизированные процессы разметки данных in-house

01.

Платформа обеспечивает весь цикл разметки: от загрузки сырых данных до выгрузки готового датасета

02.

Средства автоматизации процесса разметки позволяют увеличить скорость разметки на 40%

03.

Использование продукта не требует наличия технических специалистов

До 500

разметчиков могут работать
одновременно

Разметчик данных — это посредник между человеком и машиной. Им может быть как студент, которому нужна подработка, так и квалифицированный специалист (например, врач, размечающий снимки МРТ, или юрист, выделяющий разделы судебных документов).

Организация процесса по разметке данных:

01. Руководитель группы разметки

Организует работу группы разметчиков по конкретной задаче, контролирует процесс её выполнения.

02. Куратор

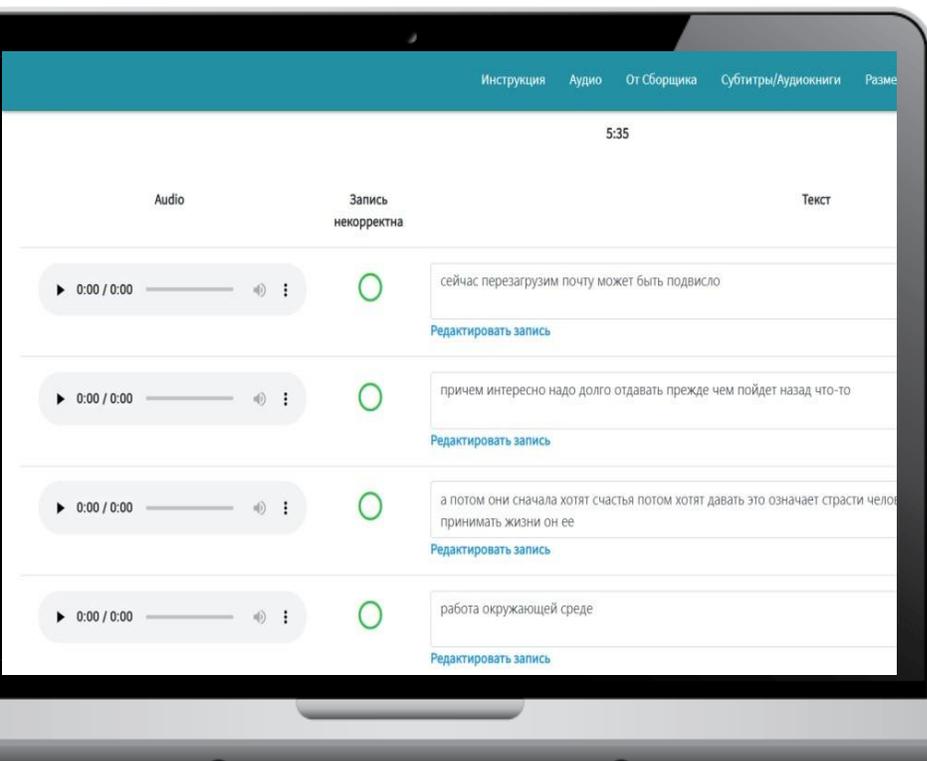
Пишет инструкции для новых заданий, рассчитывает ставки оплаты, формирует группы разметки, контролирует качество.

03. Эксперт

Опытный разметчик, проводит финальную проверку размеченных данных.

04. Разметчик

Собирает, чистит и размечает данные (тексты, аудио, изображения, ...) для машинного обучения.



- 01 Загрузка данных ▼
- 02 Выбор типа данных (текст, аудио, изображения или видео) ▼
- 03 Настройка параметров разметки ▼
- 04 Выбор исполнителей ▼
- 05 Запуск проекта ▼
- 06 Предварительная разметка с помощью нейросетей + ручная разметка ▼
- 07 Контроль качества ▼
- 08 Выгрузка готового датасета ▼



Сохранение данных в
базу данных платформы

Нейросеть «автовывделение объектов»



Нейросеть magic wand в разметке изображений

Применяется для быстрого выделения различных объектов на изображении.

Пользователь может делать клики двух видов: клик левой кнопкой мыши, который говорит о том, что данная точка является частью контура объекта, и клик правой кнопкой мыши, который говорит о том, что данная точка не является частью контура объекта.



Для полного качественного автовывделения объекта обычно достаточно 1-3 кликов мышки, что ускоряет процесс разметки примерно в 240 раз (0,5 секунды вместо 2 минут)

Нейросеть в разметке изображений

Нейросеть one-shot

Нужна для быстрого выделения объектов одного типа (например, люди, машины, животные и т.д.).

Для этого пользователь делает одну эталонную разметку, после чего one-shot автоматически размечает все подобные объекты, которые встретятся в датасете.



Заменяя интересующий объект на новый, можно быстро разметить большое количество различных объектов. В связке с нейросетью «автовыведение объектов», ускорение разметки получается ещё более значительным.

АВТОМАТИЧЕСКАЯ ПРЕДРАЗМЕТКА

На платформе есть возможность автоматической нейросетевой предразметки. Она опирается на сделанные ранее разметки и упрощает разметку нового материала.

В процессе разметки датасета вышеописанными сетями — автоматически обучается и дообучается на размечаемых данных отдельная нейронная сеть (в зависимости от задачи разметки — детекция, сегментация, распознавание и т.д.).

В итоге можно быстро доразметить оставшуюся большую часть датасета обученной нейросетью, сократив таким образом общие затраты к минимуму.

Автоматическая предразметка аудио, текстов, видео и изображений.

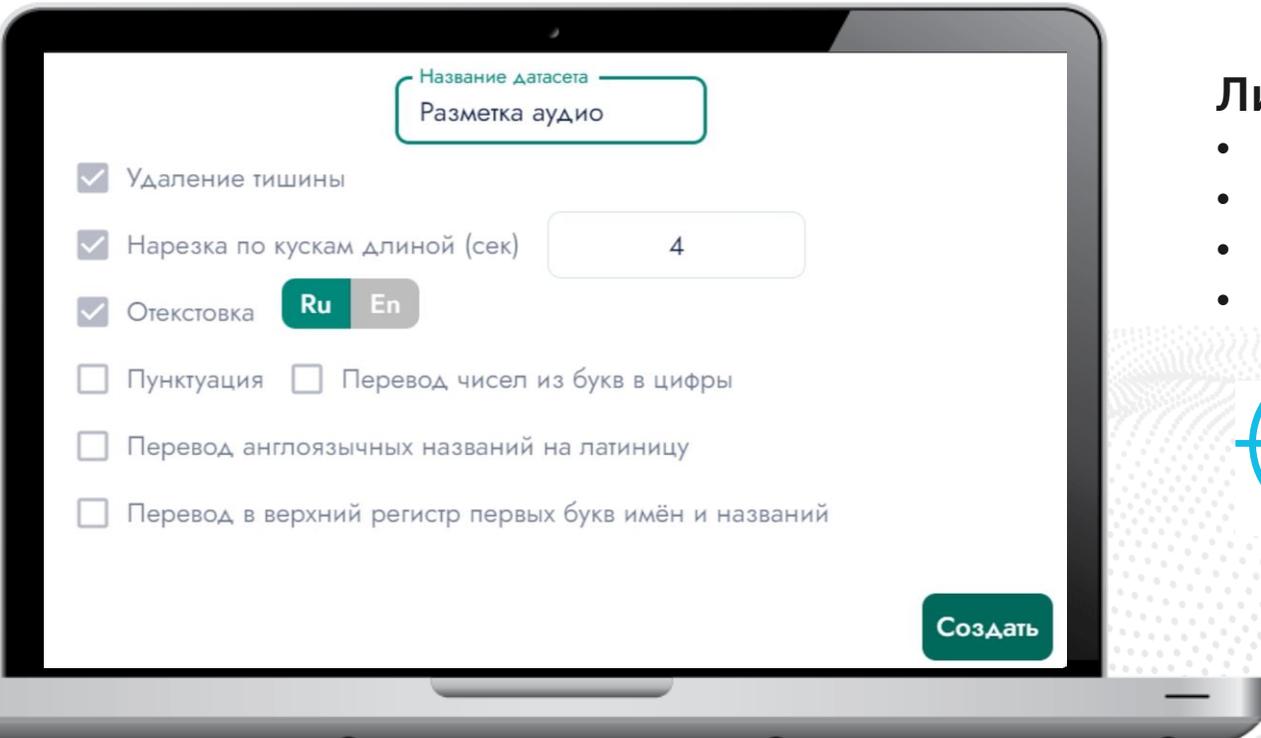
На 30 %

ускорение разметки данных



По итогу, платформа позволяет быстро разметить большой датасет даже команде, состоящей всего из одного разметчика.

Опция автоматической нейросетевой разметки аудио



АВТОМАТИЧЕСКАЯ ПРЕДРАЗМЕТКА

Предразметка аудио с использованием нейросетей

На платформе доступна опция автоматической нейросетевой предразметки аудио.

Предразметка дает возможность:

- Удалить тишину из аудио;
- Нарезать аудио;
- Удалить тишину из аудио.

Лингвистический модуль позволяет:

- Восстановить пунктуацию;
- Нормализовать текст (перевод чисел из букв в цифры);
- Работать с англоязычными названиями;
- Работать с регистром.



Языки: русский, английский.
По желанию заказчика мы можем оперативно добавить любой язык в модуль предразметки аудио

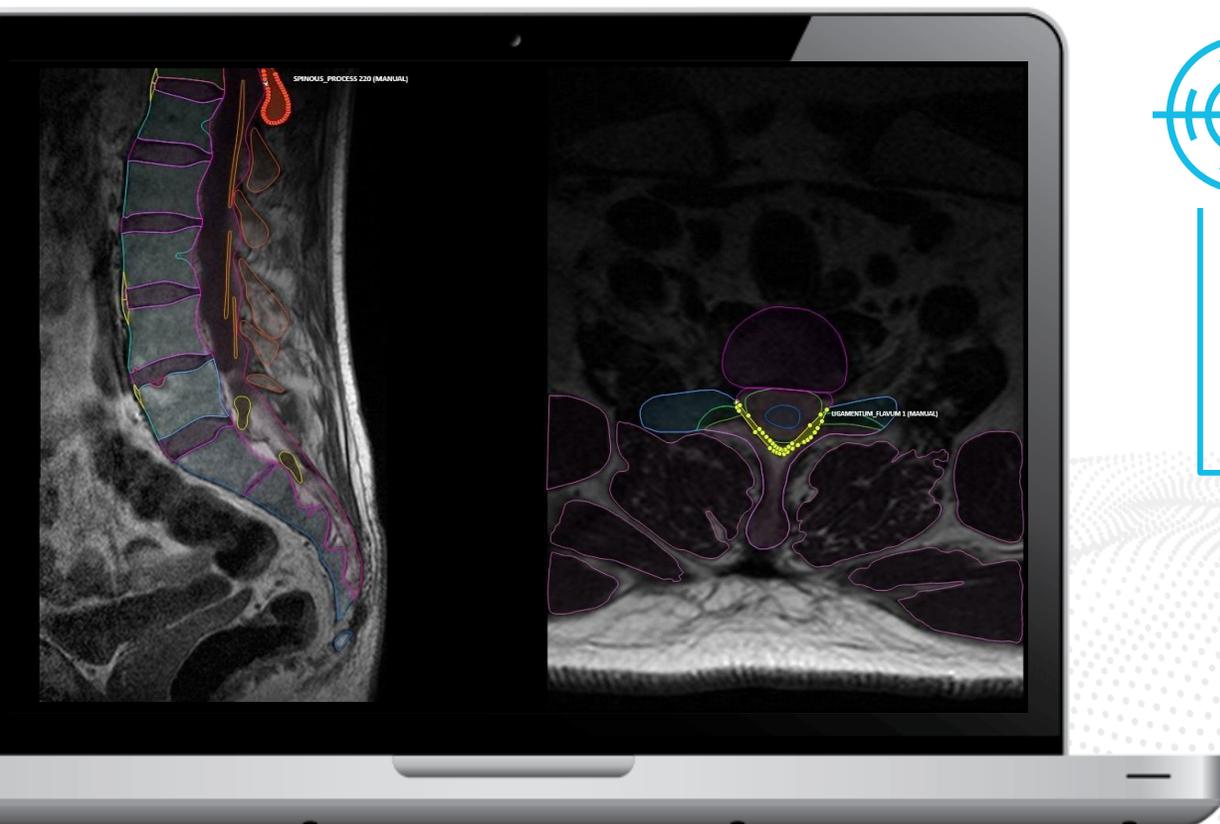
Медицинский формат

На платформу можно загружать и размечать файлы с медицинским отраслевым стандартом DICOM без дополнительной конвертации



НА 20%

снижение стоимости медицинских проектов за счет экономии времени на самостоятельную конвертацию в необходимый формат



Прочие типы данных



Разметка видео:

- Встроенный видеоплеер
- Управление скоростью видео
- Категоризация

Разметка текста:

- Категоризация
- Разметка сущностей
- Редактирование текста

СОЗДАНИЕ ДАТАСЕТА ДЛЯ СИНТЕЗА РЕЧИ



Функционал позволяющий диктору прочитать текст и далее записать его на микрофон.

При этом есть возможность делать запись неограниченное количество раз до достижения приемлемого результата.

Например, диктор первый раз запнулся, а со второго раза записал всё правильно.

Опция «запись аудио»



Ханипоты, или примеры- ловушки

Мы заранее знаем ответ на них и смотрим по системе – как ответил разметчик. Такой способ позволяет сразу отсеивать невнимательных исполнителей.



Таймер

Программа не даст разметчику закрыть страницу с заданиями ранее минимального заданного времени, которое рассчитывается по результатам тестовой разметки.



Система рейтинга

Рейтинг разметчика привязан к его ответам на ханипоты. К заданиям повышенной сложности допускаются разметчики только с высоким рейтингом. Разметчики с низким рейтингом блокируются.

Планы развития:

Биллинг

- Расчет зарплаты для разметчиков.
- Выплата зарплаты на карты сотрудников.
- Аналитика расходов по проектам, разметчикам, дате.

Журнал операций

- Возможность просмотра журнала операций разметчиков.
- Привязка журнала операций к рейтингу, разметчиков и метрикам по качеству разметки.

**БЛАГОДАРИМ
ЗА ВНИМАНИЕ!**



фцприи.рф