

# **Лабораторная работа: Аннотирование данных. Основы аннотирования данных для обучения моделей искусственного интеллекта**

## **Цель работы:**

**Изучить основы аннотирования данных, понять роль аннотации в обучении моделей искусственного интеллекта (AI), а также ознакомиться с типами аннотаций и их применением в различных задачах AI.**

## **Теоретическая часть**

### **1. Понятие аннотирования данных**

**Аннотирование данных — это процесс маркировки или разметки данных (текстов, изображений, аудио, видео и т.д.) для создания обучающих наборов, которые используются для тренировки моделей искусственного интеллекта. Аннотации помогают алгоритмам AI понимать структуру и содержание данных, что позволяет им выполнять задачи классификации, распознавания, прогнозирования и других функций.**

### **2. Роль аннотации в обучении AI**

**Аннотированные данные играют ключевую роль в обучении моделей AI, особенно в задачах, связанных с машинным обучением и глубоким обучением. Без правильно размеченных данных модели не смогут корректно интерпретировать входные данные и выдавать точные результаты. Аннотации позволяют алгоритмам "увидеть" закономерности и связи в данных, что делает их обучение более эффективным.**

### **3. Примеры задач, требующих аннотированных данных**

- **Классификация текстов:** Определение тематики текста (например, спам или не спам).
- **Распознавание объектов на изображениях:** Выделение объектов на фотографиях (например, автомобили, пешеходы).
- **Сегментация изображений:** Разделение изображения на области, соответствующие разным объектам.
- **Распознавание речи:** Преобразование аудиозаписей в текст.
- **Анализ тональности:** Определение эмоциональной окраски текста (положительная, отрицательная, нейтральная).

### **4. Типы аннотаций**

- **Текстовые аннотации:** Разметка текста для задач классификации, извлечения сущностей, анализа тональности и т.д.
  - **Пример:** Выделение именованных сущностей (имена, даты, места) в тексте.
- **Изображения:** Разметка объектов на изображениях, сегментация, классификация.
  - **Пример:** Выделение границ объектов на фотографиях.
- **Аудиоаннотации:** Разметка аудиоданных для задач распознавания речи или классификации звуков.
  - **Пример:** Разметка фрагментов аудио с указанием, кто говорит и что именно.
- **Видеоаннотации:** Разметка видео для задач отслеживания объектов, классификации действий и т.д.
  - **Пример:** Выделение объектов в кадре и их перемещение по времени.

## **Практическая часть**

### **Задание 1: Аннотирование текстовых данных**

1. **Цель:** Ознакомиться с процессом аннотирования текстовых данных для задачи классификации.

## **2. Шаги:**

- a. Выберите небольшой текст (например, отрывок из статьи или новости).**
- b. Разметьте текст, выделив следующие категории:**
  - Имена собственные (имена людей, названия организаций).**
  - Даты и временные метки.**
  - Географические названия.**
- c. Используйте инструмент для аннотирования текста (например, Label Studio).**

## **3. Результат: Предоставьте размеченный текст с выделенными сущностями.**

### **Задание 2: Аннотирование изображений**

#### **1. Цель: Ознакомиться с процессом аннотирования изображений для задачи распознавания объектов.**

#### **2. Шаги:**

- a. Выберите изображение с несколькими объектами (например, фотография улицы с автомобилями и пешеходами).**
- b. Используйте инструмент для аннотирования изображений (например, Labelbox ).**
- c. Разметьте объекты на изображении, выделив их границы и присвоив им соответствующие метки (например, "автомобиль", "пешеход").**

#### **3. Результат: Предоставьте размеченное изображение с выделенными объектами.**

### **Задание 3: Анализ аннотированных данных**

#### **1. Цель: Проанализировать роль аннотированных данных в обучении моделей AI.**

#### **2. Шаги:**

- a. Выберите одну из задач (например, классификация текстов или распознавание объектов на изображениях).**
- b. Опишите, как аннотированные данные влияют на качество обучения модели.**

- с. Приведите примеры ошибок, которые могут возникнуть при недостаточном или некорректном аннотировании данных.
3. **Результат:** Краткий отчет с анализом роли аннотаций в выбранной задаче.

## **Отчет по лабораторной работе**

### **Структура отчета:**

1. **Введение:** Краткое описание целей и задач лабораторной работы.
2. **Теоретическая часть:** Описание понятия аннотирования данных, их роли в обучении AI, примеры задач и типы аннотаций.
3. **Практическая часть:**
  - а. Результаты выполнения заданий 1 и 2 (размеченные тексты и изображения).
  - б. Анализ роли аннотаций в обучении моделей AI (задание 3).
4. **Выводы:** Общие выводы по работе, описание трудностей, с которыми вы столкнулись, и предложения по улучшению процесса аннотирования.

### **Критерии оценки:**

- **Качество аннотаций:** Точность и полнота разметки данных.
- **Анализ:** Глубина анализа роли аннотаций в обучении моделей AI.
- **Оформление отчета:** Логичность и структурированность отчета, наличие всех необходимых разделов.

### **Дополнительные материалы:**

- **Label Studio** — инструмент для аннотирования данных.
- **Labelbox** — платформа для аннотирования данных.

## **Пример отчета:**

### **Введение**

**В данной лабораторной работе мы изучили основы аннотирования данных и их роль в обучении моделей искусственного интеллекта. Были рассмотрены различные типы аннотаций и их применение в задачах классификации, распознавания объектов и анализа текстов.**

### **Теоретическая часть**

**Аннотирование данных — это процесс маркировки данных для создания обучающих наборов. Аннотации помогают моделям AI понимать структуру данных и выполнять задачи классификации, распознавания и прогнозирования. Примеры задач, требующих аннотированных данных, включают классификацию текстов, распознавание объектов на изображениях и анализ тональности.**

### **Практическая часть**

- 1. Аннотирование текстовых данных: Были размечены имена собственные, даты и географические названия в тексте.**
- 2. Аннотирование изображений: На изображении были выделены объекты (автомобили, пешеходы) с помощью инструмента CVAT.**
- 3. Анализ роли аннотаций: Было установлено, что качество аннотаций напрямую влияет на точность моделей AI. Некорректная разметка может привести к ошибкам в классификации и распознавании.**

### **Выводы**

**Аннотирование данных является важным этапом в обучении моделей AI. Качество аннотаций напрямую влияет на результаты работы моделей. Для улучшения процесса аннотирования рекомендуется использовать специализированные инструменты и проводить проверку качества разметки.**

Памятка для выполнения:

Для начала зайти на сайт

The screenshot shows the Label Studio documentation website. The browser address bar displays 'labelstud.io'. The page title is 'Label Studio Documentation — Quick start guide for Label Studio'. A navigation sidebar on the left lists various sections: DISCOVER & LEARN, INSTALL & SETUP (with 'Quick start' highlighted), CREATE & MANAGE PROJECTS, IMPORT & EXPORT, MACHINE LEARNING, MANAGE YOUR ORGANIZATION, and INTEGRATE & EXTEND. The main content area is titled 'Quick start' and contains a numbered list of steps: 1. Install Label Studio (with a terminal snippet 'pip install label-studio'), 2. Start Label Studio (with a terminal snippet 'label-studio start'), 3. Open Label Studio at 'http://localhost:8080', 4. Sign up with an email address and password, 5. Click 'Create' to create a project and start labeling data, 6. Name the project and optionally enter a description and select a color, 7. Click 'Data Import' and upload the data files, 8. Click 'Labeling Setup' and choose a template, 9. Click 'Save' to save your project. A tip box suggests a quickstart tutorial. At the bottom, there are sections for feedback ('WAS THIS USEFUL?'), contributing to the docs ('CONTRIBUTE TO THE DOCS'), and finding help ('CAN'T FIND WHAT YOU'RE LOOKING FOR?').

Label Studio Documentation — Quick start guide for Label Studio

SEE FASTER LABELING WITH PROMPTS-NOW GENERALLY AVAILABLE IN SAAS

Search docs

Learn Integrations Enterprise 27,445 Quick start

Discover & Learn

- About Label Studio
- Terminology
- Compare editions

Install & Setup

- Quick start**
- Installation
- Set up the database
- Set up persistent storage
- Start Label Studio
- Upgrade Label Studio
- Secure Label Studio

Create & Manage Projects

- Create project
- Configure labeling interface
- Manage projects
- Labeling guide

Import & Export

- Add project storage
- Import data
- Export annotations

Machine Learning

- Machine learning integration
- Write your own ML backend
- ML Examples and Tutorials

Manage Your Organization

- Add users
- Manage user accounts
- Account and settings

Integrate & Extend

- API
- Python SDK
- Webhooks

## Quick start

1. Install Label Studio:

```
pip install label-studio
```
2. Start Label Studio:

```
label-studio start
```
3. Open Label Studio at `http://localhost:8080`.
4. Sign up with an email address and password that you create.
5. Click **Create** to create a project and start labeling data.
6. Name the project and optionally enter a description and select a color.
7. Click **Data Import** and upload the data files that you want to use. If you want to use data from a local directory, cloud storage bucket, or database, skip this step for now.
8. Click **Labeling Setup** and choose a template and customize the label names for your use case.
9. Click **Save** to save your project.

You're ready to start [labeling and annotating your data!](#)

**Tip**

For a quickstart tutorial that includes demo data, see [Zero to One: Getting Started with Label Studio](#).

WAS THIS USEFUL?

Yes No

CONTRIBUTE TO THE DOCS

Our docs are open source. See something that could be improved? Submit a pull request.

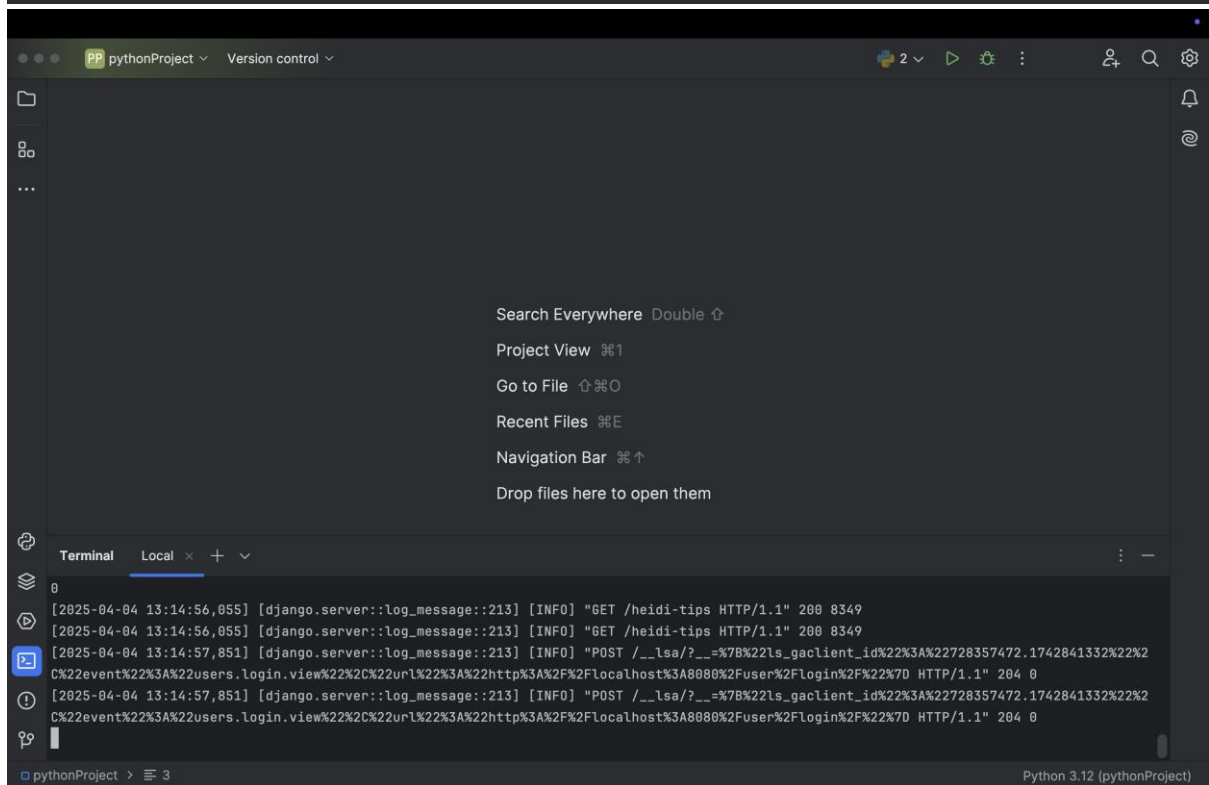
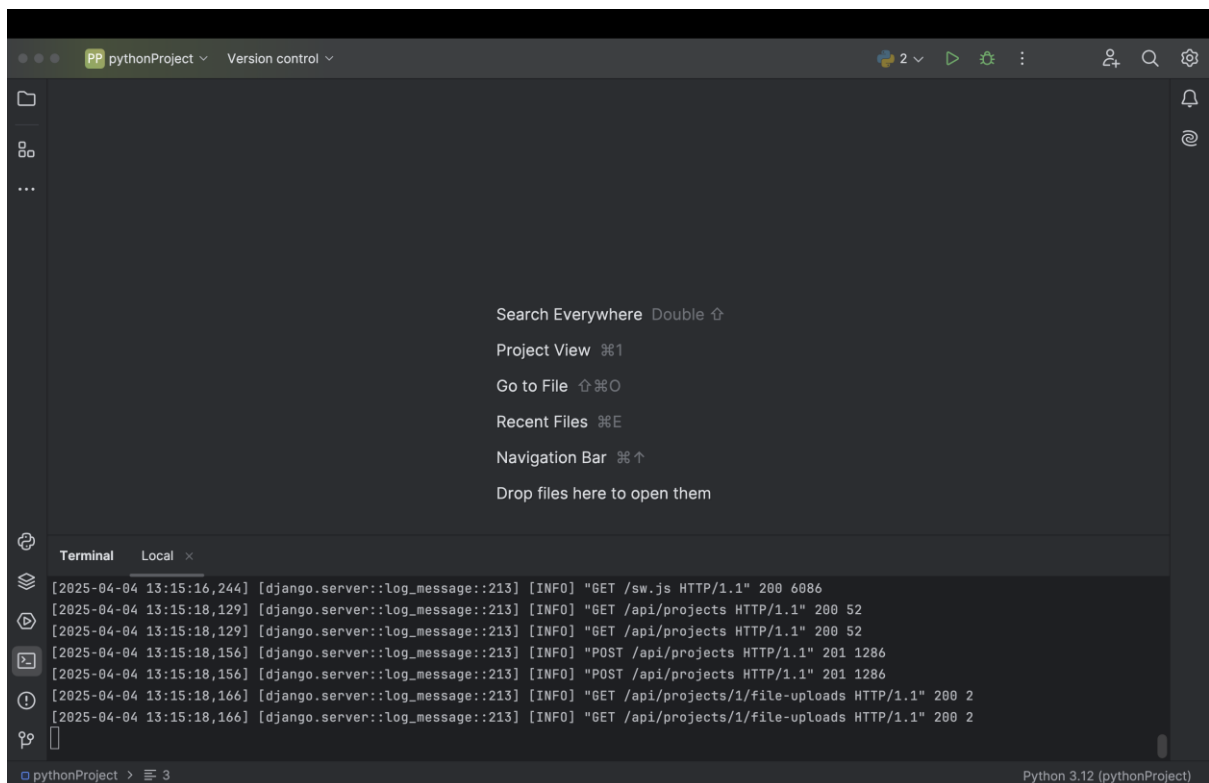
[Make a contribution](#)

CAN'T FIND WHAT YOU'RE LOOKING FOR?

[Submit a GitHub issue](#)

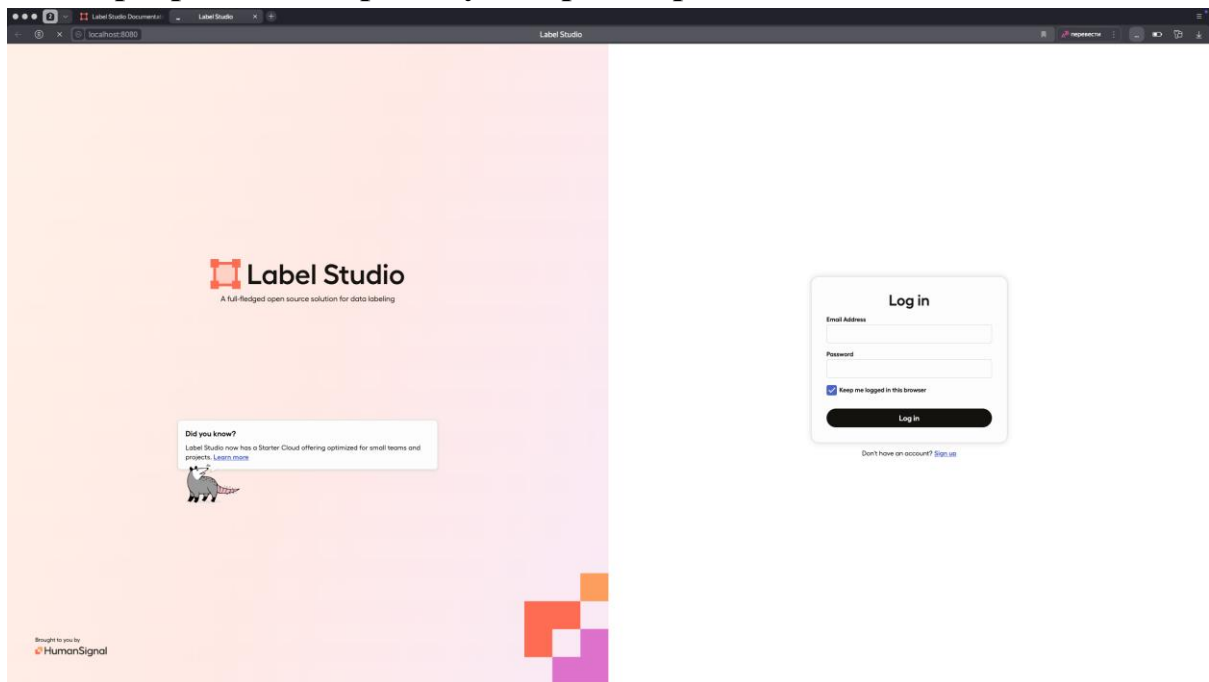
[Join Slack Community](#)

**Затем скопировать команды и ввести в командой строке(например в пайчарме)**





## Вас перебросит на страницу для регистрации



**После регистрации можно создавать проект и приступать к выполнению**

