

Финансовая грамотность начинается
с того, чтобы просто знать термины.
Это самое простое.

Подробнее



Горячее

Лучшее

Свежее

...

О чем мечтал
этот блог?



Войти



empenoso

12 дней назад



Программирование на python

...

Как сделать умный диктофон бесплатно: превращаем аудио в текст без подписок

В новостях всё чаще **говорят об «ИИ-диктофонах»** — гаджетах, которые записывают **каждый** ваш разговор в течение дня, отправляют аудио в облако, превращают его в текст и даже готовят краткую сводку по итогам. Звучит футуристично, но такие решения стоят дорого, требуют постоянной подписки и вызывают вопросы о приватности.



2962 Р 7000 Р -57%
220 шт осталось
Цифровой диктофон с дисплеем
для записи голоса и разгово...

4.9 652 отзыва

Послезавтра



4387 Р 11500 Р -61%
26 шт осталось
Диктофон супер тонкий 32 Гб 6
дней записи

4.8 120 отзывов

2 сентября



15749 Р 20500 Р -23%
13 шт осталось
Диктофон Plaud Note с
поддержкой Chat GPT, черный

4.9 54 отзыва

2 сентября



2434 Р 6750 Р -63%
102 шт осталось
Диктофон мини для записи
разговора и прослушивания...

4.8 410 отзывов

Завтра

Лично мне идея тотальной записи кажется избыточной. Зато куда практичнее другая задача: получить точную текстовую расшифровку лекции, доклада или публичного выступления. Чтобы потом не переслушивать часы аудио, а быстро найти нужную цитату или мысль простым поиском по тексту.



#Круги добра



Войти

Логин

Пароль

Войти

Создать аккаунт

Забыли пароль?

или продолжите с



Войти с Яндекс ID



Войти через VK ID



Промокоды



Работа



Курсы



Реклама



Игры



Пополнение Steam



Мой купленный за 2 т.р. диктофон с возможностью подключения внешнего микрофона на фоне коробки с ESP32

В этой статье я покажу, как построить такую систему без платных подписок и полностью под вашим контролем. Всё, что нужно — обычный диктофон за 1–3 тыс. рублей или даже просто приложение на телефоне — тогда затраты вообще равны нулю, и набор бесплатных, открытых программ, которые работают на вашем компьютере. Я купил диктофон для теста и поделюсь результатами.

Сердцем решения станет [OpenAI Whisper](#) — мощная технология распознавания речи от создателей ChatGPT. Главное её преимущество — она может работать полностью автономно на вашем ПК, не отправляя никуда ваши данные. К тому же Whisper распространяется как open-source: исходный код и модели доступны бесплатно — вы можете скачать, использовать и при необходимости даже модифицировать.

Мои скрипты [выложены на GitHub](#).

Теоретическая часть: что, почему и как?

За последние пару лет появилось немало open-source решений для распознавания речи, но именно Whisper стал фактическим стандартом. Его модели обучены на колоссальном массиве данных, что обеспечивает высокую точность распознавания. По сравнению с другими бесплатными движками, Whisper даёт результат ближе всего к коммерческим сервисам вроде Google Speech-to-Text и при этом работает автономно. Важный плюс — мультиязычность. Русский язык поддерживается «из коробки».

Модели Whisper бывают разных размеров: от tiny до large. На данный момент наиболее актуальной и точной является large-v3. Главный принцип здесь — компромисс между скоростью, точностью и требуемыми ресурсами (в первую очередь, видеопамятью). У меня [видеокарта NVIDIA GeForce RTX 5060 Ti 16 ГБ](#),

Подписка Пикабу+

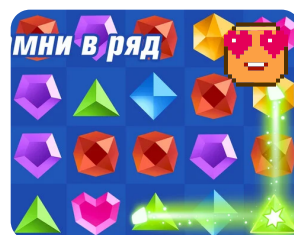
- ♦ внимание аудитории
- ♦ больше продаж
- ♦ 5 000 ₽ в Яндекс Бизнесе

[Узнать больше](#)



Пикабу Игры

+1000 бесплатных
онлайн игр



Камни в ряд онлайн!

Казуальные, Три в ряд,
Мультиплеер

[Играть](#)

поэтому на тестах использую large модель, она требует ~10 ГБ VRAM, но можно начать и со small модели — для неё достаточно ~2 ГБ VRAM.

Не стоит забывать и о приватности: все данные остаются у вас на компьютере. Никаких облачных серверов, никаких подписок. Что понадобится для запуска?

Железо: компьютер с Linux (я использую Ubuntu, но [у меня стоит двойная загрузка Windows & Linux через rEFInd Boot Manager](#)). Рекомендуется видеокарта NVIDIA — GPU многократно ускоряет работу, хотя на CPU тоже всё запустится, только медленнее. В качестве источника звука я тестировал обычный диктофон за пару тысяч рублей.

Диктофон за 1–3 тыс. рублей. Много их

Софт:

- *Python* — язык, на котором работает весь стек.
- *FFmpeg* — универсальный конвертер аудио/видео.
- *PyTorch* — фреймворк, на котором обучены модели.
- *NVIDIA Drivers* и *CUDA* — для связи с видеокартой.

Практическая часть: пошаговая инструкция

Теперь перейдём от теории к практике и соберём рабочую систему распознавания. Я разбил процесс на несколько шагов — так будет проще повторить.

Шаг 1. Подготовка окружения

Когда-то я собирал *dlib* с поддержкой *CUDA* для того чтобы распознать лица всех соседей с камеры в подъезде. Тогда я прошёл через несовместимости, конфликты версий и ручную сборку библиотек. Поэтому к установке Whisper я уже был подготовлен.

Чтобы избавить вас от всего этого «удовольствия», я написал универсальный *bash*-скрипт `setup_whisper.sh`. Он берёт на себя всю грязную работу по настройке окружения на Ubuntu 24:

Подписка Пикабу+

- ✦ внимание аудитории
- ✦ больше продаж
- ✦ 5 000 ₽ в Яндекс Бизнесе

[Узнать больше](#)



Топ прошлой недели

- SpongeGod
1 пост
- Uncleyogurt007
9 постов
- ZaTaS
3 поста

[Посмотреть весь топ](#)



Лучшие посты недели



Рассылка Пикабу:
отправляем самые
рейтинговые материалы за 7
дней 🔥

Укажите

[Подписаться](#)

Нажимая кнопку «Подписаться на рассылку», я соглашаюсь с [Правилами Пикабу](#) и даю согласие на [обработку персональных данных](#).

- обновляет систему и ставит базовые пакеты, включая Python и FFmpeg;
- проверяет драйверы NVIDIA и при необходимости устанавливает их;
- подтягивает CUDA Toolkit;
- создаёт виртуальное окружение Python и внутри него ставит PyTorch (учитывая модель видеокарты);
- загружает сам Whisper и полезные библиотеки;
- запускает тест, проверяющий, что GPU действительно работает.

Запуск прост:

```
chmod +x setup_whisper.sh
./setup_whisper.sh
```

Запуск `./setup_whisper.sh`

Полный код `setup_whisper.sh` на Гитхабе.

Шаг 2. Запись и подготовка аудио

Чем лучше исходная запись, тем меньше ошибок. Записывайте ближе к источнику звука, избегайте шумных помещений и треска. Whisper работает с самыми популярными форматами: mp3, wav, m4a, так что конвертировать вручную не придётся.

Шаг 3. Массовая расшифровка всех подряд записей

Здесь в игру вступает мой [второй скрипт](#) — `whisper_transcribe.py`. Он:

- автоматически находит все аудиофайлы в папке;
- использует GPU (если доступен), ускоряя работу в десятки раз;
- сохраняет результат в нескольких форматах:
 - .txt для текста,
 - .srt с таймкодами (можно открыть как субтитры),
 - all_transcripts.txt — общий файл со всеми расшифровками.

Пример использования:



Помощь	Правила
Кодекс Пикабу	соцсети
Команда	О
Пикабу	рекомендация
Моб.	х
приложение	О компании

Промокоды Биг Гик
 Промокоды Lamoda
 Промокоды МВидео
 Промокоды Яндекс Директ
 Промокоды Отелло
 Промокоды Aroma Butik
 Промокоды Яндекс
 Путешествия
 Постила
 Футбол сегодня



```
# Активируем окружение
source .venv/bin/activate

# Запуск по умолчанию (ищет аудио в текущей папке)
python3 whisper_transcribe.py

# Указываем папку с файлами, модель и папку для результатов
python3 whisper_transcribe.py ./audio large ./results
```

Полный код [whisper_transcribe.py](#) на Гитхабе.

```
python3 whisper_transcribe.py ./audio large ./results
```

Шаг 4. Анализ результатов

После обработки вы получите полный набор файлов. Например:

- some_lecture.txt — текст лекции;
- some_lecture.srt — субтитры вида:
12 00:04:22,500 --> 00:04:26,200 Здесь спикер рассказывает о ключевой идее...
- all_transcripts.txt — всё сразу в одном документе.

Я проверил систему на часовом файле. Модель **large** на моей RTX 5060 Ti справилась за ~8 минут.

Разделение по спикерам (диаризация) - почему это сложно?

А если записывать не лекцию, а совещание? На записи говорят пять человек, и вам нужно понять, кто именно что сказал. Обычный Whisper выдаёт сплошной текст без указания человека. Здесь на помощь приходит диаризация — технология, которая анализирует голосовые характеристики и помечает фрагменты как «Спикер 1», «Спикер 2» и так далее.

Для этого существует WhisperX — расширенная версия Whisper с поддержкой диаризации. Однако при попытке установки я опять столкнулся с классической проблемой ML-экосистемы: конфликтом зависимостей. WhisperX требует определённые версии **torchaudio**, которые несовместимы с новыми драйверами NVIDIA для RTX 5060 Ti.

Решение мне подсказали: Docker-контейнеры NVIDIA. По сути, это готовые «коробки» с предустановленным софтом для машинного обучения — разработчики уже решили все проблемы совместимости за вас. NVIDIA поддерживает целую экосистему таких контейнеров через NGC (NVIDIA GPU Cloud), а сообщество создает специализированные образы под конкретные задачи. Вместо многочасовой борьбы с зависимостями достаточно одной команды `docker pull`, и вы получаете полностью рабочую среду с предустановленным WhisperX, настроенным PyTorch и всеми библиотеками. В данном случае контейнер ghcr.io/jim60105/whisperx включает диаризацию из коробки и отлично работает с современными GPU.

Диаризация откроет новые возможности: автоматическую генерацию протоколов встреч с указанием авторства реплик, анализ активности участников дискуссий, создание интерактивных расшифровок с навигацией по спикерам.

Это тема для отдельной статьи, которую планирую выпустить после тестирования Docker-решения на реальных многоголосых записях.

Заключение

Мы собрали систему, которая позволяет бесплатно и полностью автономно расшифровывать лекции, выступления, а в перспективе и совещания. В основе — OpenAI Whisper, а все настройки и запуск упрощают мои open source скрипты. Достаточно один раз подготовить окружение — и дальше вы сможете регулярно получать точные транскрипты без подписок и риска приватности.

Следующий шаг — диаризация. Это позволит автоматически разделять текст по спикерам и превращать расшифровку совещания в полноценный протокол с указанием авторства.

Автор: Михаил Шардин

 [Моя онлайн-визитка](#)

Telegram «Умный Дом Инвестора»

2 сентября 2025



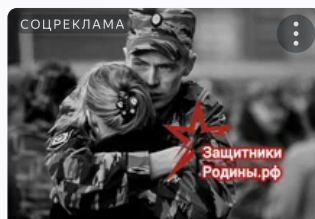
5



13



16K



→ защитники-родины.рф

Защитники Родины.
Мужчины от 18 до 60 лет. Возвращайтесь от 5 500 000 ₽

[Подробнее](#)

→ job.r-trade.org

Разработчики и инженеры БПЛА.
Вакансия!

[Подробнее](#)

→ защитник.рф

Заклучи контракт с Минобороны - получи гражданство РФ

[Подробнее](#)

Программирование на python

885 постов • 11.9K подписчиков

[Добавить пост](#)[Подписаться](#)

Правила сообщества

Публиковать могут пользователи с любым рейтингом. Однако!

Приветствуется:...

[Подробнее](#) ✓[Все комментарии](#) [Автора](#)[Раскрыть 13 комментариев](#)

Чтобы оставить комментарий, необходимо [зарегистрироваться](#) или [войти](#)



● — ■ —

—

—

—

—

—

—

—

● — ■ —

—

—

—

—

—

—

—

