

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

Хорошо



Михаил Шардин ★

личный блог



02 сентября 2025, 04:58

+ Подписаться

Протокол совещания без секретаря: бесплатная система расшифровки аудио на ПК

<p>ЦИФРОВОЙ ДИКТОФОН С ДВУМЯ ВЫСОКОКАЧЕСТВЕННЫМИ МИКРОФОНАМИ</p> <p>RU МЕНЮ НАСТРОЕК НА РУССКОМ ЯЗЫКЕ</p> <p>OLED-ДИСПЛЕЙ</p> <p>MP3 РЕЖИМ ПЛЕЙБА</p> <p>МЕТАЛЛИЧЕСКИЙ КОРПУС</p> <p>16 ГБ</p> <p>10 60 ЧАСОВ ДЛИТЕЛЬНОСТЬ РАБОТЫ</p> <p>Распродажа</p> <p>2962 Р 7000 Р -57%</p> <p>220 шт осталось</p> <p>Цифровой диктофон с дисплеем для записи голоса и разгово...</p> <p>★ 4.9 ● 652 отзыва</p> <p>Послезавтра</p>	<p>ДИКТОФОН УЛЬТРАТОНКИЙ</p> <p>32 Гб ВСТРОЕННАЯ ПАМЯТЬ</p> <p>150ч НЕПРЕРЫВНОЙ РАБОТЫ</p> <p>ВЫСОКОЕ КАЧЕСТВО ЗАПИСИ</p> <p>АКТИВАЦИЯ ГОЛОСОМ</p> <p>Распродажа</p> <p>4387 Р 11500 Р -61%</p> <p>26 шт осталось</p> <p>Диктофон супер тонкий 32 Гб 6 дней записи</p> <p>★ 4.8 ● 120 отзывов</p> <p>2 сентября</p>	<p>ДИКТОФОН PLAUD NOTE С ПОДДЕРЖКОЙ CHAT GPT</p> <p>30 часов непрерывной работы</p> <p>60 дней в режиме ожидания</p> <p>64 Гб</p> <p>Распродажа</p> <p>15749 Р 20500 Р -23%</p> <p>13 шт осталось</p> <p>Диктофон Plaud Note с поддержкой Chat GPT, черный</p> <p>★ 4.9 ● 54 отзыва</p> <p>2 сентября</p>	<p>Максимум записи минимум забот</p> <p>Вся необходимая помощь в любой ситуации</p> <p>формат WAV чистое высококачественное без потерь качества</p> <p>чистая записи высококачественное</p> <p>Распродажа</p> <p>2434 Р 6750 Р -63%</p> <p>102 шт осталось</p> <p>Диктофон мини для записи разговора и прослушивания...</p> <p>★ 4.8 ● 410 отзывов</p> <p>Завтра</p>
--	---	---	---

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.

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.

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 д

Введите текст комментария

распространяется как open-source: исходный код и модели доступны бесплатно — вы можете скачать, использовать и при необходимости даже модифицировать.

Мои скрипты [выложены на GitHub](#).



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

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 —

видеокарта [NVIDIA GeForce RTX 5060 Ti 16 ГБ](#), 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](#)). Рекомендуются 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.

The screenshot shows a marketplace interface with a search bar containing 'Везде диктофон'. Below the search bar, there are four product listings for digital voice recorders:

- Product 1:** ЦИФРОВОЙ ДИКТОФОН. Features: RU, меню настроек на русском языке, OLED-дисплей, MP3, запись, 16 ГБ, поддержка карт памяти Micro SD, цифровой диктофон с дисплеем для записи голоса и разговора... Price: 2962 Р (7000 Р -57%). 220 шт осталось. 4.9 rating, 652 отзыва.
- Product 2:** ультра-компактный ДИКТОФОН МИНИ. Features: 500ч записи, 15ч без зарядки, 8 гб памяти, наушники в подарок. Price: 1128 Р (3200 Р -64%). Оригинал. Диктофон мини для записи звука, с датчиком звука. 4.7 rating, 2026 отзывов.
- Product 3:** ЦИФРОВОЙ ДИКТОФОН. Features: с голосовой активацией, с двумя микрофонами, меню настроек на русском языке, 16 гб, встроенная память, проводной микрофон, петличка в комплекте. Price: 3350 Р (6350 Р -47%). 76 шт осталось. 4.8 rating, 1219 отзывов.
- Product 4:** ЦИФРОВОЙ ДИКТОФОН с функцией шумоподавления. Features: активация записи по датчику звука, активное шумоподавление, защита диктофона паролем, проводной микрофон, петличка в подарок. Price: 2110 Р (5230 Р -59%). 50 шт осталось. 4.7 rating, 763 отзыва.

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.

шаг 1. подготовка окружения

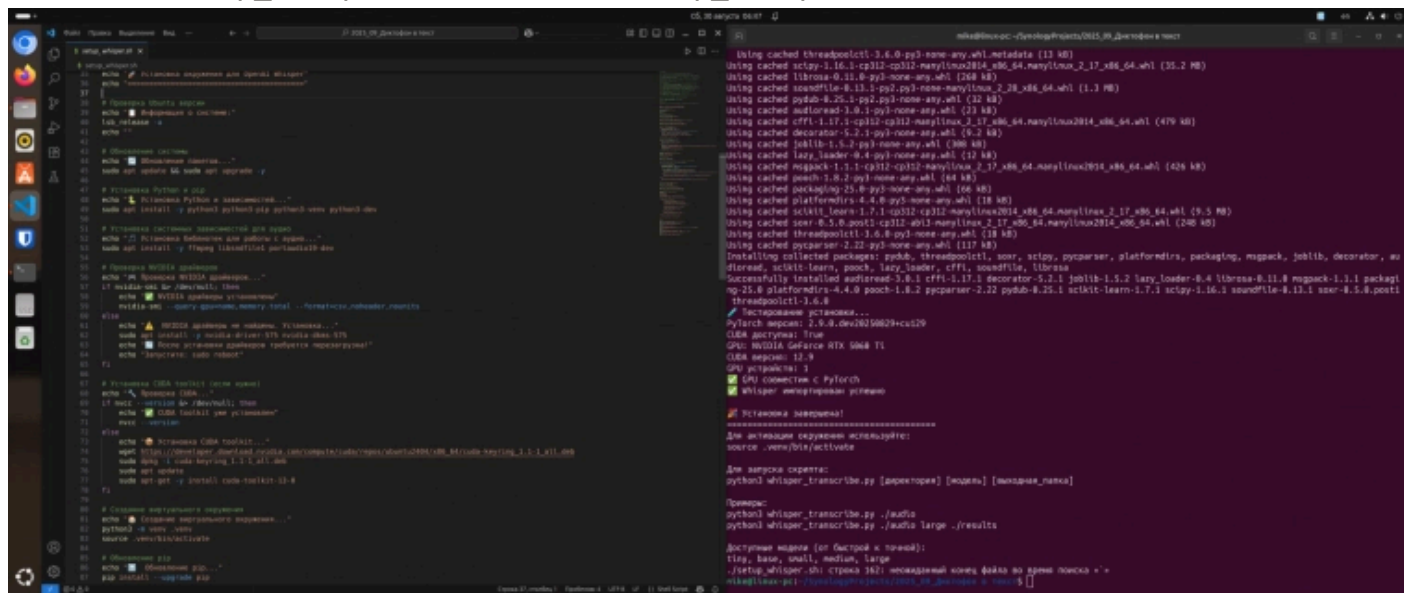
Когда-то я собирал dlib с поддержкой CUDA для того чтобы распознать лица всех соседей с камеры в подъезде. Тогда я прошёл через несовместимости, конфликты версий и ручную сборку библиотек. Поэтому к установке Whisper я уже был подготовлен.

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:

- 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 <br />./setup whisper.sh
```



Запуск ./setup_whisper.sh

Полный код `setup_whisper.sh` на Гитхабе.

Шаг 2. Запись и подготовка аудио

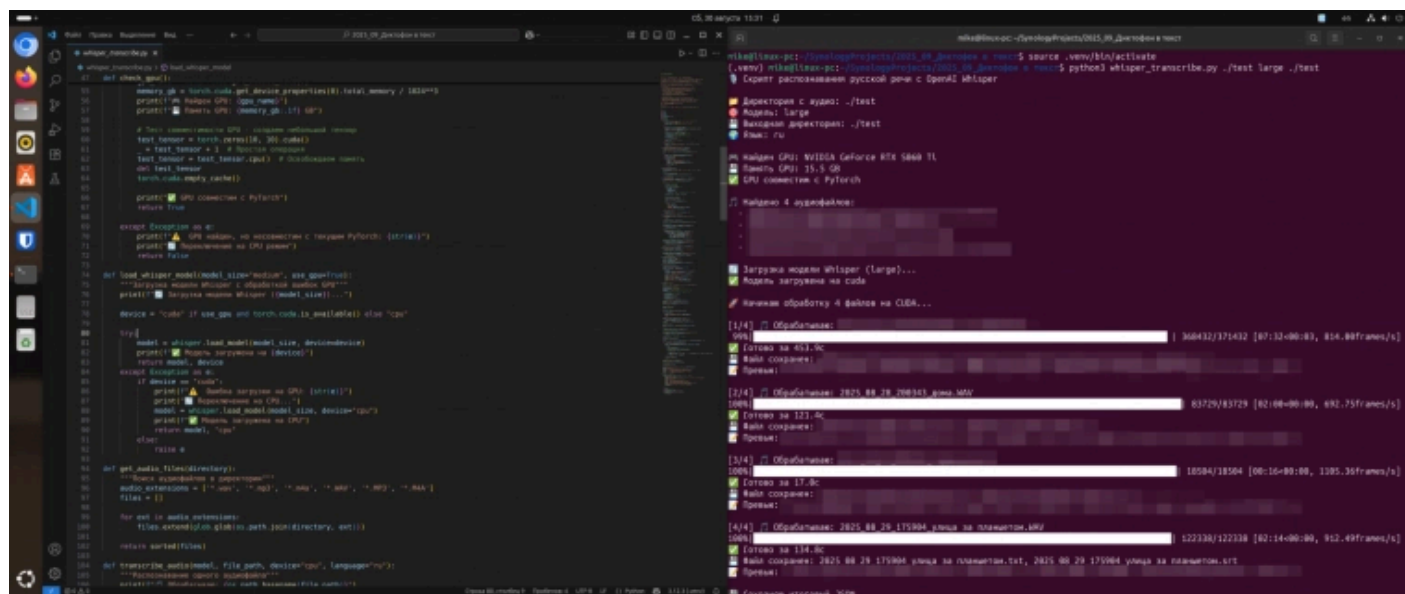
Шаг 3. Массовая расшифровка всех подряд записей

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:

Активируем окружение
`source .venv/bin/activate`
 # Запуск по y
 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 ключи`

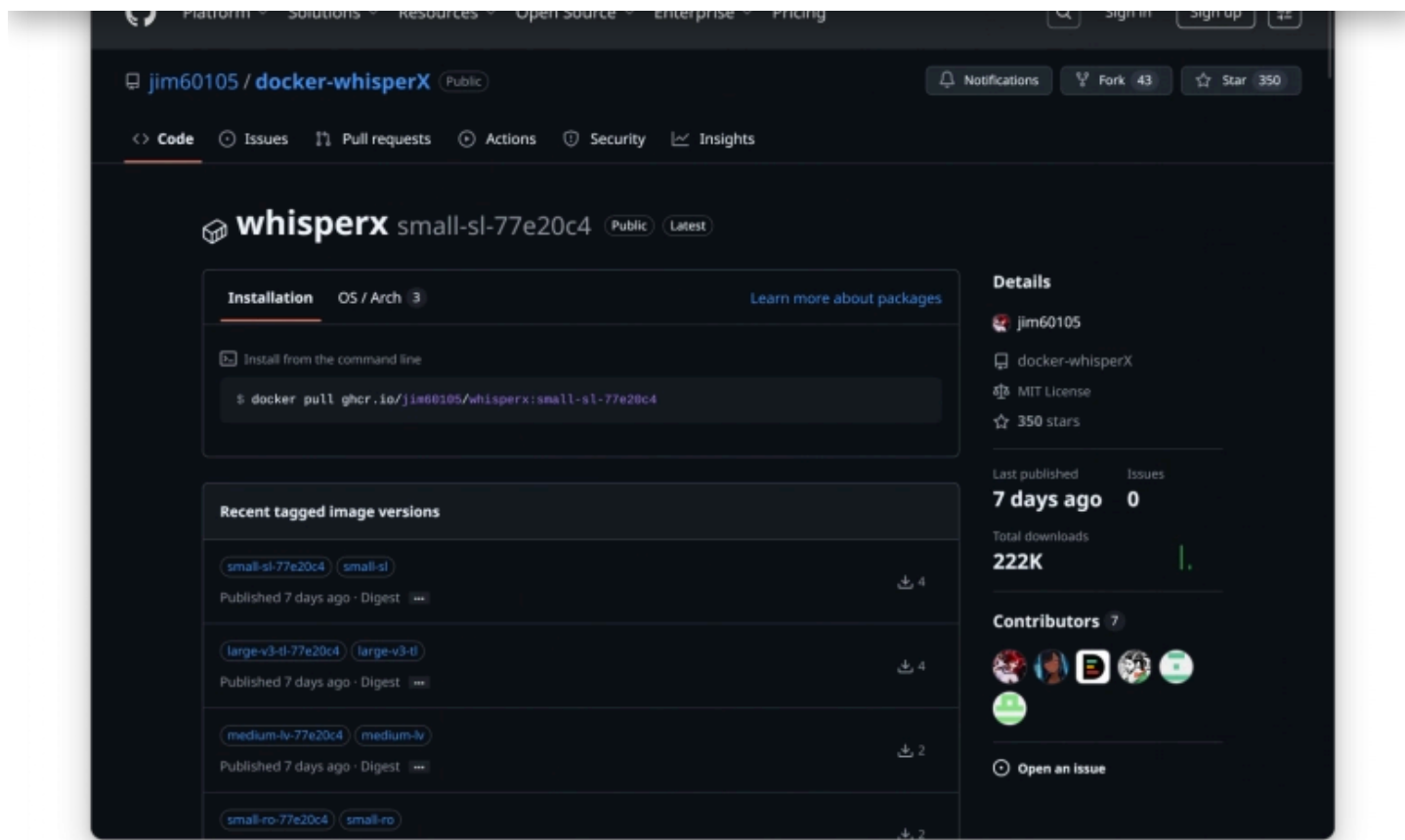
Разделение по спикерам (диаризация) — почему это сложно?

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.

OpenAI

1.9K  8 14 15**Михаил Шардин** Пермь 398  4 228 с 23 января 2019 +HreHDn1F5CZjN...**+ Подписаться**

14 КОММЕНТАРИЕВ

Сначала старые **AlexShul**

02 сентября 2025, 06:01



Держу в курсе — все это (и диаризацию) и многое другое давно умеет делать смартфон. В частности, samsung s24 plus.

 Показать 10 ответов  +3 **deadman68**

02 сентября 2025, 09:26

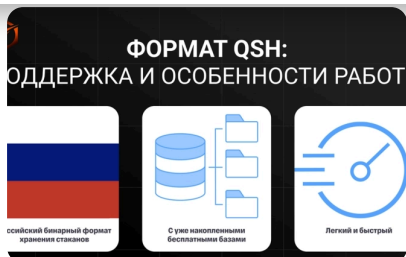


`pip install torch torchvision torchaudio --index-url download.pytorch.org/whl/cpu` — загрузилось только через VPN, и Загрузка моделей Whisper происходит только с включенным VPN/

 Показать 1 ответ  

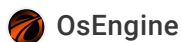
Напишите комментарий...





Формат QSH в OsEngine: поддержка и особенности работы

Недавно OsEngine начал поддержку бинарного формата хранения и трансляции данных по стаканам. Это было нужно, чтобы:...



18:01

Промышленная автоматизация — один из ключевых трендов 2026 в ИТ



Промышленная автоматизация — один из ключевых трендов 2026 в ИТ #SOFL_тренды

Сегодня промышленность все чаще смотрит на ИТ как на инструмент для наращивания мощностей. Для российской...



17:16



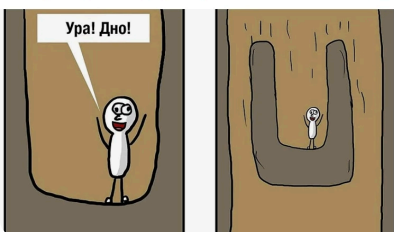
⚡ Объявляем условия нового размещения

20 января финтех-сервис ПСБ Финанс (бренд CarMoney) начнет размещение нового выпуска облигаций с плавающим купоном...



17:52

кабре идет ко дну - хуже не было никогда



Хэдхантер. Ситуация на рынке труда в декабре идет ко дну - хуже не было никогда

Вышла статистика рынка труда за декабрь 2025 года, которую Хедхантер публикует ежемесячно, что же там интересного:...



13.01.2026

Установите приложение Смартлаба:

