

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>ПОДДЕРЖКА КАРТ ПАМЯТИ Micro SD</p> <p>МИКРОФОН ПЕЛЕНКА В КОНТАКТЕ</p> <p>Распродажа 10 60 ЧАСОВ ДЛИН АККУМУЛЯТОРА</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>150 Ч 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 СРЕДНЕЙ ЗАПИСИ</p> <p>64 Гб</p> <p>87 КОММЕНТОВ</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 воспроизведение без потери качества)</p> <p>чёткая запись (высокая детализация)</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 для те

Откройте счёт в ВТБ Мои Инвестиции

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 —

Откройте счёт в ВТБ Мои Инвестиции

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 с двумя высококачественными микрофонами. Features: RU, меню настроек на русском языке, OLED-дисплей, MP3, запись, 16 ГБ, поддержка карт памяти Micro SD, металлический корпус. Price: 2962 Р (7000 Р -57%). 220 шт осталось. Цифровой диктофон с дисплеем для записи голоса и разговора... 4.9 652 отзыва. Button: Послезавтра.
- Product 2:** ультра-компактный диктофон МИНИ. Features: 500ч записи, 15ч без зарядки, 8 гб памяти, наушники в подарок. Price: 1128 Р (3200 Р -64%). Оригиналы. Диктофон мини для записи звука, с датчиком звука. 4.7 2026 отзывов. Button: Завтра.
- 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. Цифровой диктофон с датчиком звука, встроенной памятью на 1... 4.8 1219 отзывов. Button: Послезавтра.
- 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. Цифровой диктофон с шумоподавлением. 4.7 763 отзыва. Button: Завтра.

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.

Откройте счёт в ВТБ Мои Инвестиции

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
```

```

1 # Установка окружения для Whisper
2 #
3 # 1. Обновление системы и установка базовых пакетов
4 sudo apt update
5 sudo apt install python3 python3-pip python3-venv ffmpeg
6
7 # 2. Проверка драйверов NVIDIA
8 nvidia-smi
9
10 # 3. Установка CUDA Toolkit
11 curl -O https://developer.nvidia.com/cuda-12.1-linux-archive-install
12 sudo sh cuda-12.1-linux-archive-install
13
14 # 4. Установка PyTorch с поддержкой CUDA
15 python3 -m pip install torch torchvision torchaudio
16
17 # 5. Проверка работы GPU
18 python3 -c "import torch; print(torch.cuda.is_available())"
19
20 # 6. Установка Whisper и других библиотек
21 python3 -m pip install openai-whisper
22
23 # 7. Запуск теста
24 python3 test.py
  
```

Запуск `./setup_whisper.sh`

Полный код `setup_whisper.sh` на Гитхабе.

Шаг 2. Запись и подготовка аудио

Откройте счёт в ВТБ Мои Инвестиции

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
```

`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<br />00:04:22,500 --> 00:04:26,200<br />Здесь спикер рассказывает о ключ
```

Откройте счёт в ВТБ Мои Инвестиции

Разделение по спикерам (диаризация) — почему это сложно?

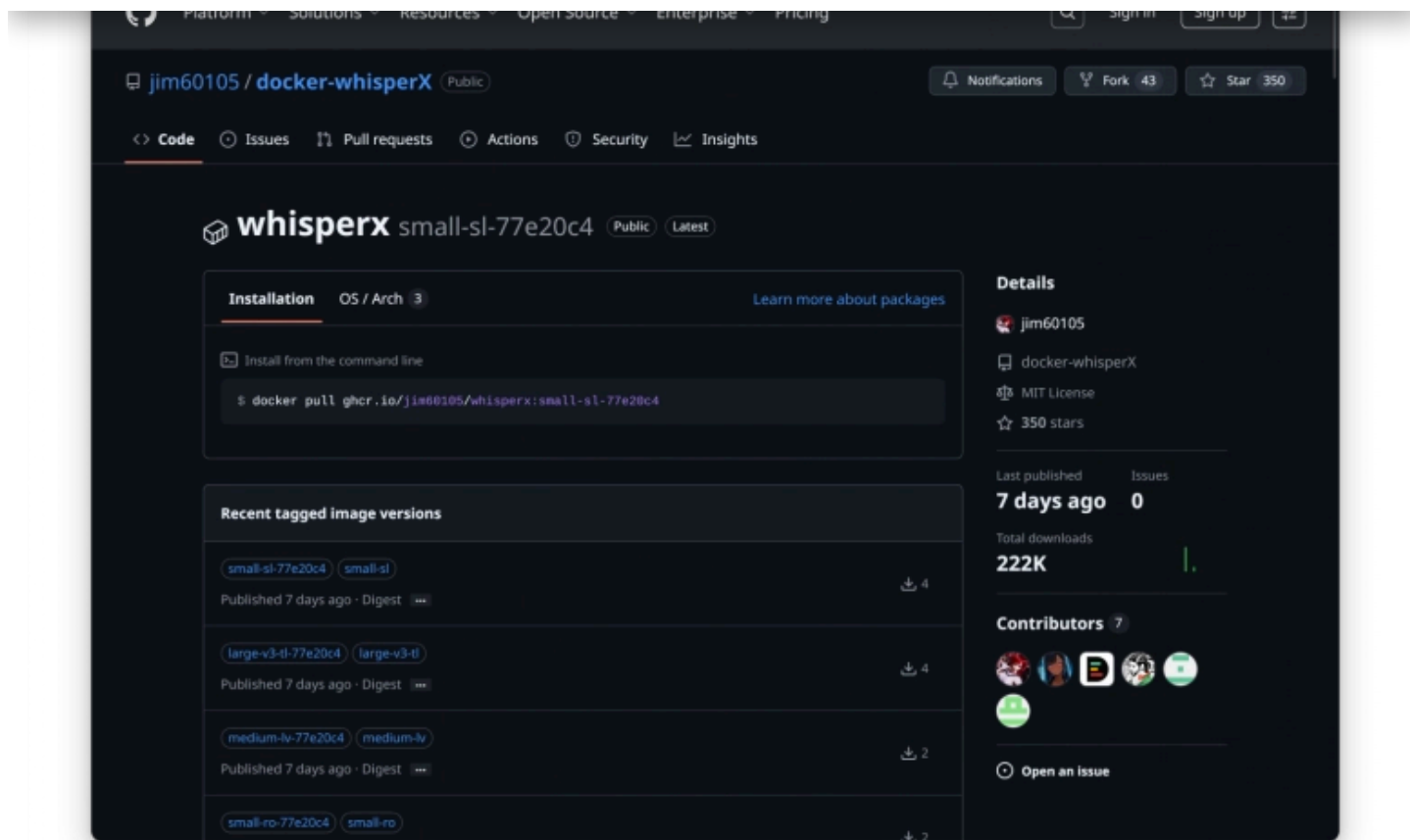
А если записывать не лекцию, а совещание? На записи говорят пять человек, и вам нужно понять, кто именно что сказал. Обычный Whisper выдаёт сплошной текст без указания человека. Здесь на помощь приходит диаризация — технология, которая анализирует голосовые характеристики и помечает фрагменты как «Спикер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

Откройте счёт в ВТБ Мои Инвестиции



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.

Откройте счёт в ВТБ Мои Инвестиции

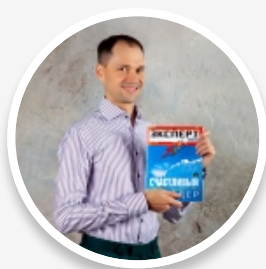
OpenAI

1.3K

7

14

14

**Михаил Шардин**

Пермь

257 2 502

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

+2

**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...



Откройте счёт в ВТБ Мои Инвестиции

ЗА ПОЛУГОДИЕ, ЗАКОНЧИВШЕЕСЯ 30 ИЮНЯ 2025 ГОДА (НЕАУДИРОВАННЫЙ)
(в миллионах российских рублей, за исключением прибыли на акции)

Примечание	За полугодие, закончившееся	
	30 июня 2025 года	30 июня 2024 года
ВЫРУЧКА	16	171 233
СЕБЕСТОИМОСТЬ РЕАЛИЗАЦИИ	17	(144 244)
ВАЛЮТНАЯ ПРИВЛЕЧЕНОСТЬ	26 989	41 827
Коммерческие, общезастольные и административные расходы	18	(15 722)
Прочие операционные доходы	19	1 383
Прочие операционные расходы	(137)	(115)
Изменение справедливой стоимости инвестиционной недвижимости		128
ОПЕРАЦИОННЫЙ УБЫТОК (ПРИВЛЕЧЕНОСТЬ)	(7 599)	9 770
Финансовые доходы	20	545
Финансовые расходы	20	(26 632)
УБЫТОК ДО НАЛОГА НА ПРИБАВЛЕНИЕ	(33 686)	(12 364)
Доходы по налогу на прибыль	8 527	2 036
Чистый налоговый убыток и чистый налоговый убыток на капитал	(25 159)	(10 328)

М.Видео - есть ли в мультивселенной компания, где миноритариям что-то достанется?

М.Видео меняет формат допэмиссии. Ранее акционеры одобрили допэмиссию до 500 млн новых акций по закрытой подписке — о...



Mozgovik

03:51

ДАЙДЖЕСТ ПО РЕЙТИНГОВЫМ ДЕЙСТВИЯМ В ВЫСОКОДОХОДНОМ СЕКТОРЕ, ПОРТФЕЛЕ PROBONDS ВДО И РОЗНИЧНЫХ ИНВЕСТИЦИОННЫХ ОБЛИГАЦИЙ ЗА...

● ПАО «ЕВРОТРАНС» АКРА подтвердило кредитный рейтинг на уровне A-(RU) ПАО «ЕвроТранс» — один из крупнейших независимых топливных операторов на рынке Московского региона...



Иволга Капитал

13.09.2025

Банк России снизил ключевую ставку до 17% — что делать? Мнение аналитиков

Сегодня на заседании Банк России снизил ключевую ставку на 1 п.п., до 17%. Теоретически, для рынков акций и облигаций — это определенно благоприятное событие, но на практике рынок...



БКС Мир инвестиций

13.09.2025

Установите приложение Смартлаба:



RuStore



AppGallery



App Store



О смартлабе

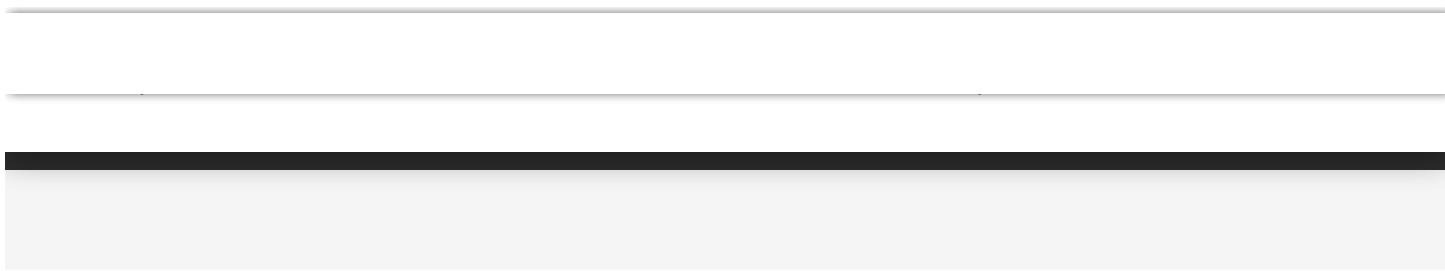
Реклама

Полная версия



Московская Биржа является спонсором ресурса smart-lab.ru
Источник: ПАО Московская Биржа

Откройте счёт в ВТБ Мои Инвестиции



Откройте счёт в ВТБ Мои Инвестиции