



empenoso

4 часа назад

## „Просто используй Docker“ — мой путь от Whisper к WhisperX

Простой 6 мин 732

Open source\*, Настройка Linux\*, Python\*, Умный дом

Кейс

Несколько недель назад я опубликовал статью о том, [как превратить обычный диктофон в инструмент для расшифровки речи с помощью OpenAI Whisper](#). Идея была создать бесплатную и приватную систему ИИ диктофона, которая избавляет от необходимости переслушивать аудиозаписи лекций или выступлений. Тогда статья нашла своего читателя, собрав 140 закладок.

Название	Цена	Особенности	Дата
Plaud note pin	18 126 Р	Умный диктофон Plaud Note Pin (Cosmic gray) с чатом GPT - диктофон с искусственным...	18 сентября
SCHEMATEX KR-666	7 990 Р	Профессиональный цифровой диктофон KR 666 с 16 гб...	20 сентября
Диктофон профессиональный	3 179 Р (было 9 990 Р -68%)	61 шт осталось SCHEMATEX Диктофон профессиональный V65 / MP3-плеер / 32GB	20 сентября
ДИКТОФОН	19 301 Р (было 59 990 Р -67%)	70 шт осталось SCHEMATEX Диктофон PLAUD с искусственным интеллектом...	20 сентября

ИИ и обычные диктофоны

В процессе настройки я боролся с несовместимостью библиотек, подбирал нужные версии драйверов и вручную собирал рабочее окружение. В комментариях мне справедливо заметили: «Вместо всей этой возни можно было найти готовый Docker-контейнер и поднять всё одной командой». Звучало логично, и я с энтузиазмом принял этот совет. *Я ведь верю людям в интернете.*

Новая идея - не просто расшифровывать речь, а разделять её по голосам - как на совещании или встрече. Это называется диаризацией, и для неё существует продвинутая версия - WhisperX. Цель была проста - получить на выходе не сплошное полотно текста, а готовый протокол встречи, где понятно, кто и что сказал. Казалось, с Docker это будет легко.

Но я заблуждался. Путь «в одну команду» оказался полон сюрпризов - всё сыпалось одно за другим: то скрипт не видел мои файлы, то не мог получить к ним доступ, то просто зависал без объяснения причин. Внутри этой «волшебной упаковки» царил хаос, и мне приходилось разбираться, почему она не хочет работать.

Но когда я всё починил и заставил систему работать, результат превзошёл мои ожидания.

РЕКЛАМА

**Получи грант за код**

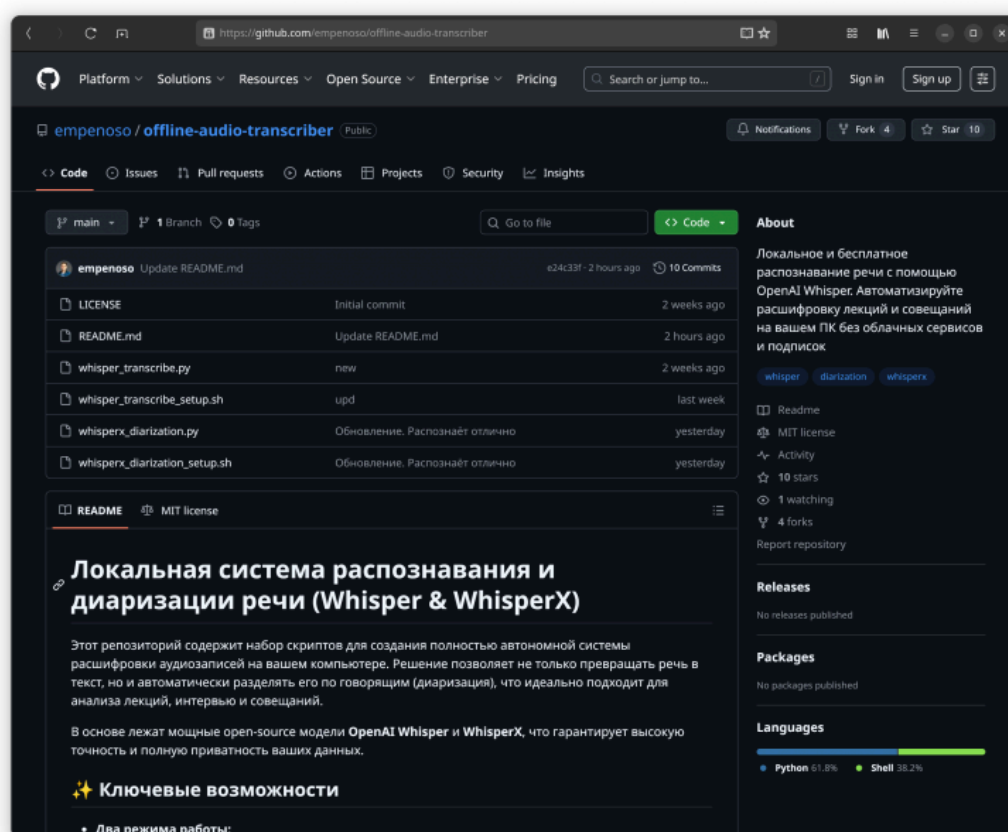
Конкурс open source проектов

Это был настолько лучший результат, что я смог передать его большой языковой модели (LLM) и получить глубокий анализ одной очень важной для меня личной ситуации - под таким углом, о котором я сам бы никогда не задумался.

Именно в этот момент мой скепсис в отношении «умных ИИ-диктофонов», которые я критиковал в первой статье, сильно пошатнулся. Скорее всего их сила не в тотальной записи, а в возможности превращать хаос в структурированные данные, готовые для анализа.

В этой статье я хочу поделиться своим опытом прохождения этого квеста, показать, как обойти все скрытые сложности, и дать вам готовые инструкции, чтобы вы тоже могли превращать свои записи в осмысленные диалоги.

Весь код выложен на Гитхаб.



Код на Гитхабе

## Docker как панацея и почему Linux

В комментариях меня критиковали за то, что я опять написал статью про Linux. Да, у меня на домашнем компьютере стоит Ubuntu в режиме двойной загрузки - и многим непонятно, почему я не сделал всё под Windows. Ответ прост: для задач с нейросетями Linux даёт меньше неожиданностей и больше контроля. Драйверы, контейнеры, права доступа - под Linux их проще исследовать и чинить, особенно когда начинаешь ковырять CUDA и системные зависимости.



Получи грант за код

Конкурс open source проектов

Для выбора загрузки какую ОС использовать выбрал rEFInd Boot Manager

Ещё меня критиковали за RTX 5060 Ti 16GB - мол, не у всех такие видеокарты. Согласен, это не смартфон в кармане. Но для работы с большими моделями и диаризацией нужна мощь GPU: я использую её как инструмент. К тому же подходы, которые я описываю, работают и на более скромных конфигурациях - просто медленнее.

А теперь начнём с самого начала - что такое Docker простыми словами? Представьте, что вместо того, чтобы настраивать компьютер под каждую программу, вы берёте готовую «коробку» и в ней уже есть всё: нужные версии Python, библиотеки, утилиты. Эта «коробка» запускается одинаково на любой машине - как виртуальная мини-кухня.

То есть мой план действий был такой:

1. Установить Docker.
2. Скачать готовый образ с WhisperX.
3. Запустить одну команду и получить готовый протокол встречи.

Так что могло пойти не так?

## Первое столкновение с реальностью

Уже на первом шаге начались сюрпризы:

### Секретный токен, который не дошёл до адресата

Чтобы запустить диаризацию, WhisperX использует модели от ruannote, а они требуют авторизации через токен Hugging Face. Я передал его как переменную окружения Docker ( -e

ера ожидала  
дель упорно



Получи грант за код

Конкурс open source проектов

## Война за права доступа

Следующая засада - `PermissionError` при попытке записи в системные папки `/.cache`.  
Контейнер как гость в доме: ему разрешили пользоваться кухонным столом, а он пошёл сверлить стены в гостиной. Разумеется, система его остановила. Решение оказалось простым - создать отдельную «полку» для кеша ( `~/.whisperx` ) и явно указать путь.

## Загадочное зависание

Запускаешь скрипт - и тишина. Ни ошибок, ни логов, будто процесс замёрз. На деле работа шла, просто механизм вывода в контейнере «затыкался». Решение - добавить индикатор прогресса.

Так что Docker - не магия, а всего лишь ещё один инструмент, который тоже нужно приручить.

## Решение: два скрипта

Я написал две утилиты - один раз подготовить систему, второй - управлять обработкой. Это простая, надёжная пара: установщик устраняет системные «подводные камни», оркестратор - закрывает все проблемы запуска (HF-token, кэш, права, прогресс).

### Шаг 1. Фундамент: `whisperx_diarization_setup.sh`

Назначение: однократно подготовить Ubuntu - поставить Docker, NVIDIA toolkit, скачать образ



**Получи грант за код**

Конкурс open source проектов

- проверяет дистрибутив и наличие GPU ( `nvidia-smi` );
- устанавливает Docker и добавляет пользователя в группу `docker`;
- ставит NVIDIA Container Toolkit и настраивает runtime;
- подтягивает образ `ghcr.io/jim60105/whisperx:latest` ;
- создаёт `./audio` , `./results` и `~/whisperx` , выставляет права и генерирует `config.env` .

Пример:

```
# создаём директорию и конфиг
mkdir -p ./audio ./results "$HOME/whisperx"
chmod -R 777 ./audio ./results "$HOME/whisperx"
cat > ./config.env <<'EOF'
HF_TOKEN=your_token_here
WHISPER_MODEL=large-v3
DEVICE=cuda
...
EOF
# загрузка образа
sudo docker pull ghcr.io/jim60105/whisperx:latest
```

## Шаг 2. Пульт управления: `whisperx_diarization.py`

Роль: оркестратор - перебирает файлы, формирует корректную команду `docker run` и решает описанные проблемы. Как он их решает:

- HF\_TOKEN передаётся и как `-e HF_TOKEN=...` , и в аргументах `--hf_token` при запуске `whisperx`;
- глобальная папка кеша `~/whisperx` монтируется в контейнер и назначается `HOME=/models` , `XDG_CACHE_HOME=/models/.cache` - проблем с `PermissionError` нет;
- прогресс-бар: чтение `stderr` контейнера и человеко-понятные статусы (VAD → транскрибация → выравнивание → диаризация);
- проверка готовности: `--check` тестирует Docker, образ и права записи.

Пример:

```
# проверка системы
python3 whisperx_diarization.py --check
# обработать всю папку
python3 whisperx_diarization.py
```

Подробная инструкция и актуальные скрипты - в репозитории:

👉 <https://github.com/empenoso/offline-audio-transcriber>



Получи грант за код

Конкурс open source проектов

Когда все технические баталии были позади, я наконец смог оценить, стоила ли игра свеч. Результат был отличный.

В первой статье обычный Whisper выдавал сплошное текстовое полотно. Информативно, но безжизненно. Вы не знали, где заканчивается мысль одного человека и начинается реплика другого.

#### Было (обычный Whisper):

...да, я согласен с этим подходом но нужно учесть риски которые мы не обсудили например финансовую сторону вопроса и как это повлияет на сроки я думаю нам стоит вернуться к этому на следующей неделе...

#### Стало (WhisperX с диаризацией):

[00:01:15.520 --> 00:01:19.880] SPEAKER\_01: Да, я согласен с этим подходом, но нужно учесть риски, которые мы не обсудили.  
[00:01:20.100 --> 00:01:22.740] SPEAKER\_02: Например, финансовую сторону вопроса и как это повлияет на сроки?  
[00:01:23.020 --> 00:01:25.900] SPEAKER\_01: Именно. Я думаю, нам стоит вернуться к этому на следующей неделе.

WhisperX с диаризацией превращает этот монолит в сценарий пьесы. Каждый спикер получает свой идентификатор, а его реплики - точные временные метки. Разница колоссальная. Теперь это не просто расшифровка, а полноценный протокол.

#### Мой личный кейс

Но настоящая магия началась, когда я решил пойти дальше. Я взял расшифровку одного личного разговора, сохранённую в таком структурированном виде, и загрузил её в нейросеть Gemini 2.5 Pro с простым запросом: «Действуй как аналитик. Проанализируй этот диалог».

Именно из-за структуры Gemini смогла отследить, кто инициировал темы, кто чаще соглашался или перебивал, как менялась тональность и динамика беседы. В итоге я получил анализ скрытых паттернов в общении, о которых сам никогда бы не задумался. Это был взгляд на ситуацию с абсолютно неожиданной стороны, который помог мне лучше понять и себя, и собеседника.



**Получи грант за код**

Конкурс open source проектов



**Получи грант за код**  
Конкурс open source проектов

Даже бесплатное приложение в телефоне может служить источником

Я понял, что их главная ценность «ИИ-диктофонов» - не в способности записывать каждый ваш шаг, а в умении превращать хаос человеческого общения в структурированные, машиночитаемые данные. Это открывает возможности: от создания кратких сводок по итогам встреч до глубокого анализа коммуникаций, который раньше был невозможен.

## Заключение

В итоге путь от «просто используй Docker» к рабочей связке WhisperX показал очевидную вещь: контейнеры - удобный инструмент, но не магия.

Подготовка системы и правильная оркестровка запуска - это то, что превращает хаос в рабочий процесс. Если вы готовы потерпеть небольшие сложности ради удобства в дальнейшем - результат оправдает усилия: структурированные протоколы и возможность глубокого анализа бесед.

**Автор:** Михаил Шардин

 [Моя онлайн-визитка](#)

 [Telegram «Умный Дом Инвестора»](#)

23 сентября 2025

**Теги:** [диктофон](#), [whisper](#), [whisperx](#), [rtx 5060](#), [cuda](#), [расшифровка аудио](#)

**Хабы:** [Open source](#), [Настройка Linux](#), [Python](#), [Умный дом](#)

## Редакторский дайджест

Присылаем лучшие статьи раз в месяц



Оставляя свою почту, я принимаю [Политику конфиденциальности](#) и даю согласие на получение рассылок



214 51.5



Получи грант за код

Конкурс open source проектов

[Автоматизация](#) / [Данные](#) / [Финансы](#) / [Умные дома](#)



[Подписаться](#)[Сайт](#) [Сайт](#) [GitHub](#)

Комментарии 4

## Публикации

[ЛУЧШИЕ ЗА СУТКИ](#)[ПОХОЖИЕ](#)**sobolevn**

18 часов назад

### Список докладов с PythoNN в рамках ITGorky

Средний

3 мин

1.9K

[Дайджест](#)

+50

45

8

**alizar**

19 часов назад

### Коллизии хэш-функций и парадокс дней рождения

**Получи грант за код**

Конкурс open source проектов

+42

32

9



techno\_mot

22 часа назад

## От наскальных рисунков до S3: эволюция хранения данных

12 мин 2.9K

Обзор

+39

23

2



RationalAnswer

13 часов назад

## Криптовзятка Трампу на \$2'000'000'000, о которой никто не говорит

3 мин 5.7K

+30

11

11



dudukalo

18 часов назад

## Операторы rest и spread в JavaScript

4 мин 2.2K

Обзор

+30

15

4



alexeybashuk

22 часа назад

## «Сверхразум запатентовал обычный лоток для яиц и теперь судится буквально со всеми» — вот что произошло на самом деле

Простой 3 мин 8K

Аналитика

+30

12

18



Antxak

20 часов назад

## Telegram: голосуй и проиграешь

Простой 7 мин 7.2K

Мнение

+28

22

20



interpres

16 часов назад

## Актуальны ли спустя 40 лет советы из «Жемчужин программирования»?



Получи грант за код

Конкурс open source проектов

+26

35

12



RationalAnswer

23 часа назад

## Цукерберг надел умнеющие очки, а также вызовы на допрос за банковскую карту из Казахстана

9 мин

11K

Дайджест

+24

10

22



BiktorSergeev

22 часа назад

## M5Stack Cardputer: кастомизируемый компьютер размером с кредитку за 30 \$. Что за девайс?

6 мин

2.4K

+23

13

3

## Цифровые двойники и Лаборатория ИИ — мир ИТ-прома, каким вы его не знали

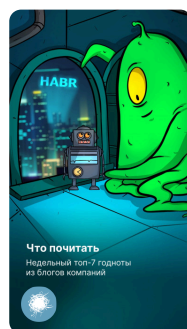
Турбо

Показать еще

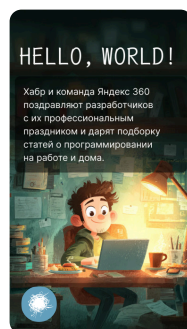
### ИСТОРИИ



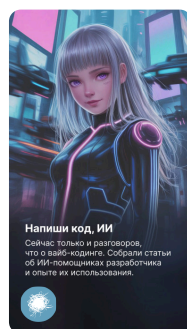
Прогулка по IT Elements



Годнота из блогов компаний



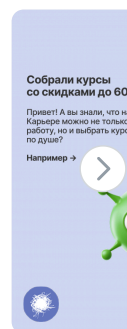
Чай, торт и код: с Днём программиста!



Made in AI



Закрывать вкладки — открыть окно



Курсы со скидками до 60%

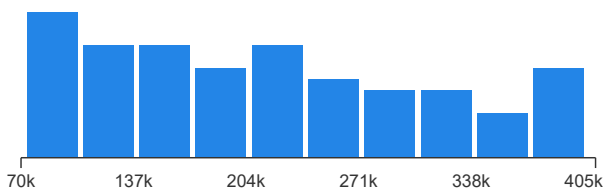
### СРЕДНЯЯ ЗАРПЛАТА В IT

**214 536** ₽/мес.

— средняя зарплата во всех IT-специализациях по данным из 30 033 анкет, за 2-ое пол. 2025 года. Проверьте «в рынке» ли ваша зарплата или нет!

**Получи грант за код**

Конкурс open source проектов



Приветствие от организаторов

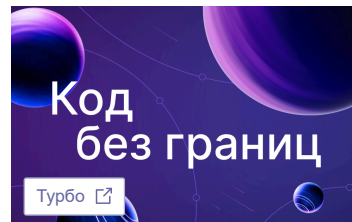


Взорвите рынок в безумной игре ко Дню программиста



Турбо

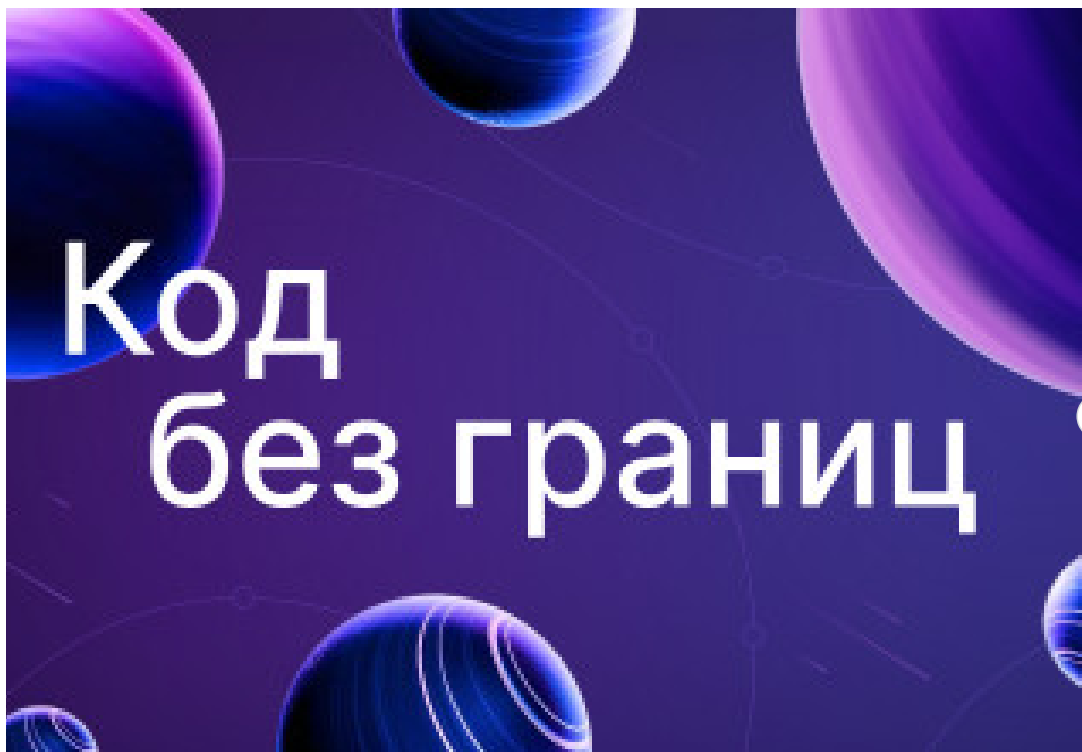
Цифровизация на максималках: чем живёт IT-пром



Турбо

Отправь свой open source проект на конкурс и выиграй грант

#### БЛИЖАЙШИЕ СОБЫТИЯ



3 сентября — 31 октября

**Программа грантов для развития open source проектов «Код без границ»**

Онлайн

Разработка

Больше событий в календаре



**Получи грант за код**

Конкурс open source проектов

Хабр



🌐 Настройка языка

Техническая поддержка

© 2006–2025, Habr



**Получи грант за код**

Конкурс open source проектов