

Хабр



КАК СТАТЬ АВТОРОМ



Новые знания здесь



Войти



empenoso

2 сен в 05:23

Как локально и бесплатно распознать текст лекции или совещания и делать это регулярно

Простой

6 мин

8.4K

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

Кейс

ЦИФРОВОЙ ДИКТОФОН
С ДВУМЯ ВЫСОКОКАЧЕСТВЕННЫМИ МИКРОФОНОНАМИ

RU
МЕНЮ НАСТРОЕК НА РУССКОМ ЯЗЫКЕ

ОLED ДИСПЛЕЙ

MP3 РЕЖИМ ПЛЕЙБА

16 ГБ

ПОДДЕРЖКА КАРТ ПАМЯТИ Micro SD

МИКРОФОН ВЕЩАЮЩА В КОМПЛЕКТЕ

Распродажа 10 40 ЧАСОВ ЗАРЯДА АККУМУЛЯТОРА

2962 Р 7000 Р -57%

220 шт осталось

Цифровой диктофон с дисплеем для записи голоса и разговора...

★ 4.9 652 отзыва

Послезавтра

ДИКТОФОН УЛЬТРАТОНКИЙ

32гб ВСТРОЕННАЯ ПАМЯТЬ

150ч НЕПРЕРЫВНОЙ РАБОТЫ

ВЫСОКОЕ КАЧЕСТВО ЗАПИСИ

АКТИВАЦИЯ ГОЛОСОМ

Распродажа

4387 Р 11500 Р -61%

26 шт осталось

Диктофон супер тонкий 32 Гб 6 дней записи

★ 4.8 120 отзывов

2 сентября

ДИКТОФОН PLAUD NOTE
С ПОДДЕРЖКОЙ CHAT GPT

30 ЧАСОВ ЗАПИСИ

60 ДНЕЙ ЗАПИСИ

64 ГБ

07 КОЛОНКОВ

Распродажа

15749 Р 20500 Р -23%

13 шт осталось

Диктофон Plaud Note с поддержкой Chat GPT, черный

★ 4.9 54 отзыва

2 сентября

Максимум записи минимум забот

Ваш надежный помощник в любой ситуации

Формат WAV

Чистое воспроизведение без потери качества

Чёткая запись

Диктофон мини для записи разговора и прослушивания...

Распродажа

2434 Р 6750 Р -63%

102 шт осталось

Диктофон мини для записи разговора и прослушивания...

★ 4.8 410 отзывов

Завтра

На онлайн-площадках и ИИ-диктофоны и обычные вперемешку продаются

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

РЕКЛАМА

**Получи грант за код**

Конкурс open source проектов

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.



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

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](#).



Получи грант за код

Конкурс open source проектов

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](#), 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.



Получи грант за код

Конкурс open source проектов

Софт:

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

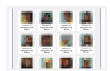
Практическая часть: пошаговая инструкция

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

Анализ лиц с домофона: как я победил несовместимости и собрал dlib+CUDA на U...

Каждый день мимо двери моего подъезда проходят десятки людей. Иногда это знакомые соседи, но чаще — ...

habr.com



Шаг 1. Подготовка окружения

Когда-то я собирал dlib с поддержкой CUDA для анализа лиц 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):



Получи грант за код

Конкурс open source проектов

- 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:



Получи грант за код

Конкурс open source проектов

- `.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 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:



Получи грант за код

Конкурс open source проектов

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.

**Получи грант за код**

Конкурс open source проектов

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.



Получи грант за код

Конкурс open source проектов

Автор: Михаил Шардин

 [Моя онлайн-визитка](#)

 [Telegram «Умный Дом Инвестора»](#)

2 сентября 2025

Теги: диктофон, whisper, Whisperx, openai, rtx 5060, ngc, rtx, cuda, расшифровка аудио

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

Редакторский дайджест



Присылаем лучшие статьи раз в месяц



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



212

63.3

Карма

Рейтинг

Михаил Шардин [@empenoso](#)

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

[Подписаться](#)



[Сайт](#) [Сайт](#) [GitHub](#)



Получи грант за код

Конкурс open source проектов

 Комментарии 12

Публикации

ЛУЧШИЕ ЗА СУТКИ

ПОХОЖИЕ

**Tirarex**

13 часов назад

Как я делал сеть на 2,5 гигабита с минимальным бюджетом — апгрейд, доступный каждому



Простой



9 мин



11K

Тutorial



+41



81



34

**Erwinmal**

17 часов назад

Сэндвич, сэр? История британских бутербродов от аристократических салонов до вокзальных буфетов



Простой



13 мин



3K

**Получи грант за код**

Конкурс open source проектов

 +41 20 2**oneastok**

18 часов назад

Умное зеркало на Raspberry Pi: пошаговое руководство



Простой



4 мин



4.4K

Обзор

Перевод

 +22 56 14**iLushkersky**

14 часов назад

Жизнь на Марсе? (снова)



Простой



3 мин



2.3K

 +17 5 10**TrexSelectel**

15 часов назад

Nintendo Virtual Boy: неожиданное возрождение виртуальной реальности из 90-х



5 мин



1.1K

 +14 3 3**mio_anni**

18 часов назад

От мини-ЭВМ и перфокарт к IDE и фреймворкам. Как поменялось программирование за 50 лет — взгляд изнутри



12 мин



1.9K

 +12 15 35**Получи грант за код**

Конкурс open source проектов

Станислав Петров: «Ключевые отличия РЕД ОС М от Android – вовсе не в интерфейсе»

 Средний  8 мин  6.5K

Интервью

 +10

 10

 43



Albert_Wesker

18 часов назад

Миф о быстром и медленном пути выполнения программы

 Средний  11 мин  1.5K

Обзор

Перевод

 +9

 16

 0



kilokanat

5 часов назад

Механическая клавиатура LARKeyboard

 Простой  5 мин  481

Тutorial

 +8

 4

 2



beeline_cloud

9 часов назад

Научный «дипфейк»? Как галлюцинации нейросетей — и другие проблемы — просачиваются в академические статьи

 Простой  8 мин  872

Аналитика

 +8

 11

 2



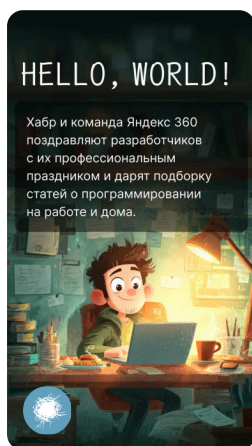
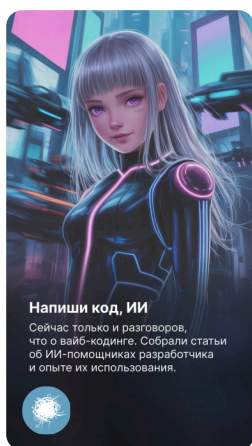
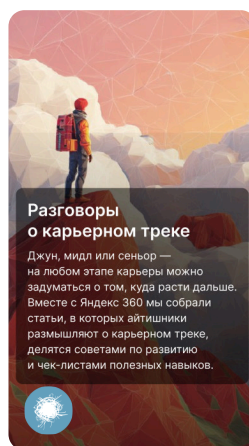
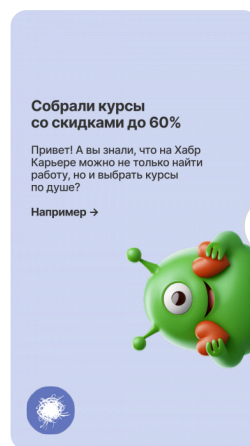
Получи грант за код

Конкурс open source проектов

Турбо

Показать еще

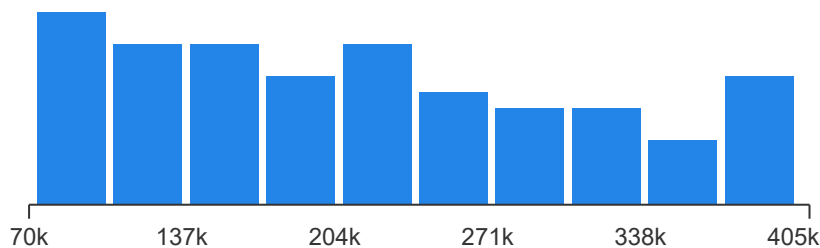
ИСТОРИИ

**Чай, торт и код: с Днём программиста!****Made in AI****Чего хотят лиды в бигтехе?****Как расти в IT: советы, гайды и опыт сеньоров****Курсы со скидками до 60%**

СРЕДНЯЯ ЗАРПЛАТА В IT

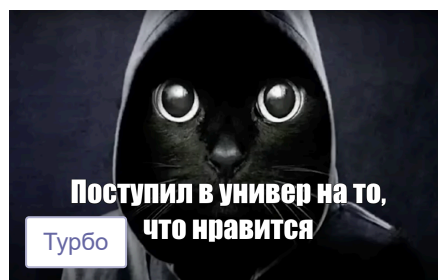
214 724 ₽/мес.

— средняя зарплата во всех IT-специализациях по данным из 27 443 анкет, за 2-ое пол. 2025 года. Проверьте «в рынке» ли ваша зарплата или нет!

[Проверить свою зарплату](#)**Получи грант за код**

Конкурс open source проектов

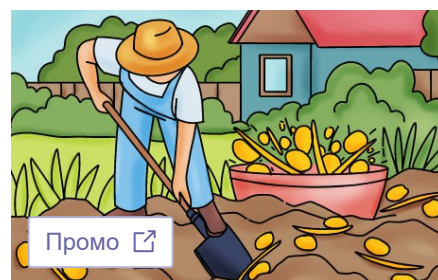
МИНУТОЧКУ ВНИМАНИЯ



Где обучают вести кибертурниры и разрабатывать игры

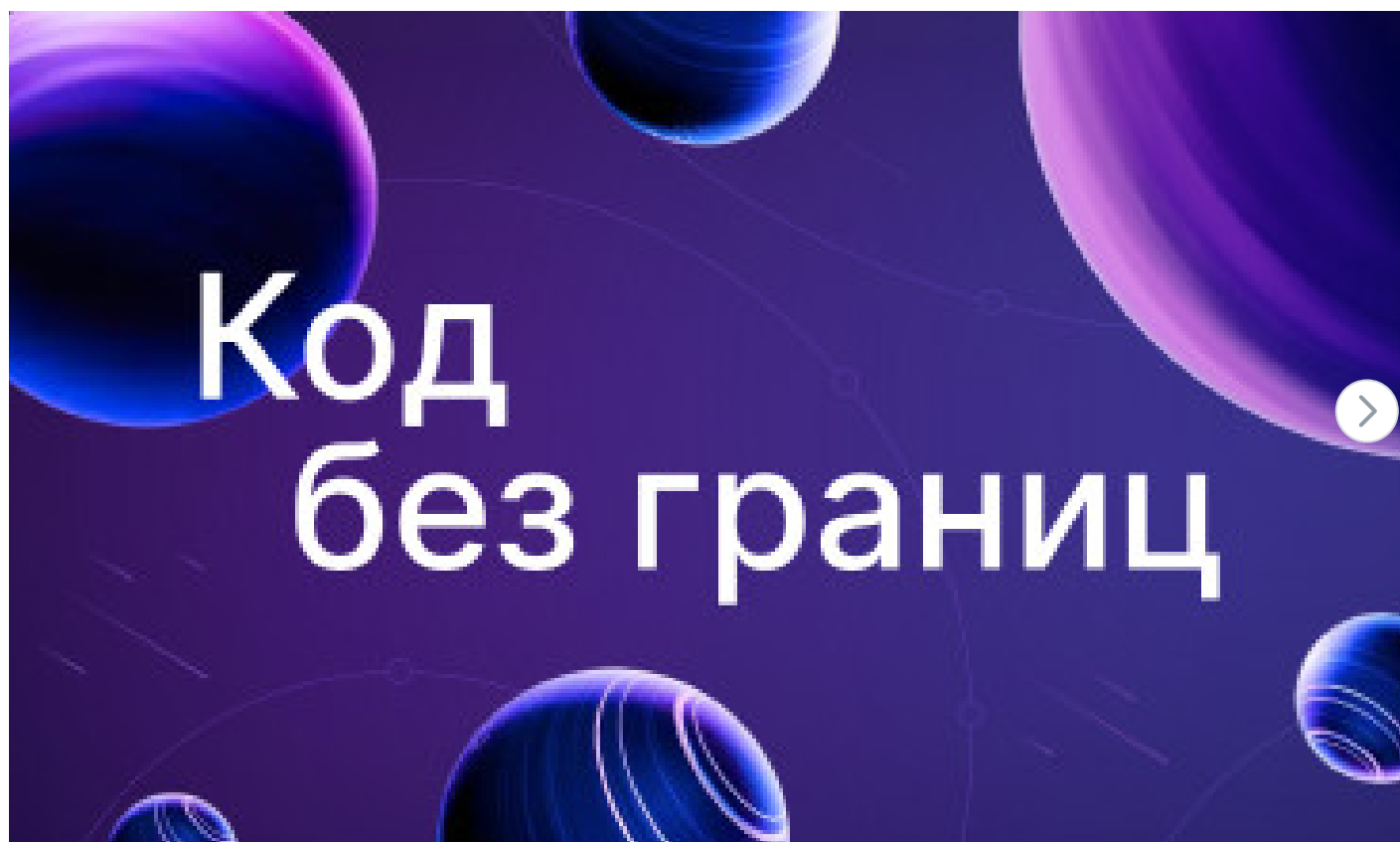


Цифровизация на максималках: чем живёт ИТ-пром



Не нужно копать картошку — хватит скидок в Промокодусе

БЛИЖАЙШИЕ СОБЫТИЯ



3 сентября – 31 октября

Программа грантов для развития open source проектов «Код без грани



Получи грант за код

Конкурс open source проектов

[Больше событий в календаре](#)

Хабр



 [Настройка языка](#)

[Техническая поддержка](#)

© 2006–2025, Habr



Получи грант за код

Конкурс open source проектов