

# Отчёты по встречам

## Контур.Толка:

### contur\_transcribe

Для разработчиков. Состояние на 28.09.2026. Переменные, песочница и выкатка по шагам — [README](#) репозитория `Buharestskie/contur_transcribe`. На сервере проект развёрнут под именем `donu-la-kontur-reporter`.

## Коротко

`contur_transcribe` — демон без интерфейса и без входящих запросов. Раз в пять минут он забирает новые записи встреч из пространства Контур.Толка, получает расшифровку, делает управленческий разбор языковой моделью и раскладывает результат: короткий отчёт и файл расшифровки — в закрытую группу Telegram, структурированные поля — в рабочую таблицу Google Sheets отдела продаж. Задача — чтобы итоги тет-а-тетов и общих встреч РОПов доходили до таблицы без ручной работы.

## Как устроен один проход

1. **Список записей.** Один ключ пространства (`KTALK_SPACE_API_KEY`) читает записи всех РОПов за последние `LOOKBACK_DAYS` дней (по умолчанию 7). Уже опубликованные и тестовые (по шаблонам названий) пропускаются.
2. **Расшифровка.** Сначала берётся готовая расшифровка Толка. Если текста в ней нет, запись скачивается в качестве 240p, из неё вытаскивается звук (ffmpeg) и отправляется в сервис распознавания а5х; задача опрашивается до готовности (до часа). Если у Толка запись ещё не готова, встреча откладывается до следующего прохода — это не ошибка.
3. **Разбор.** Диалог уходит в языковую модель а5х (`A5X_MODEL`). Она возвращает короткий управленческий радар: план и факт отдельно по каждому сотруднику, одна строка по каждому участнику, до трёх главных проблем, все поставленные задачи с ответственными и сроками (до 12), где нужен руководитель, оценка полезности встречи 0–10. Отсутствующее помечается «Не зафиксировано». Длинные встречи разбираются через конспекты частей. Если модель недоступна,

разбор собирается из протокола Толка.

4. **Маршрутизация.** Если участников два-три, однозначно опознан один сотрудник и на встрече был его РОП — это тет-а-тет: заполняется строка сотрудника на листе этого РОПа. Всё остальное — на лист «Общие встречи». Полная расшифровка в любом случае пишется на скрытый лист «Расшифровки Контур.Толк».
5. **Telegram.** Отчёт одним сообщением (длинный — разбивается на части) и расшифровка отдельным `.txt`.
6. **Метрики.** В конце прохода счётчики `contur_transcribe_cycle_*` (найдено, опубликовано, отложено, упало, пропущено) уходят в Pushgateway мониторинга.

Каждый выполненный шаг сразу отмечается в `data/state.json`. Сбой посередине не теряет встречу и не публикует её дважды: следующий проход доделает только оставшееся.

## Кто есть кто

Сотрудники и РОПы описаны в двух справочниках вне кода: `config/rops.json` (РОП и название его листа) и `config/employees.json` (сотрудник, его РОП и признаки опознания: `emails`, `konturUserKeys`, `aliases`, `roomNames`, `titlePatterns`). Образцы — `*.example.json`. Заполнять руками не нужно: `python -m tools.sync_identities` показывает, что изменится, с `--write` — дописывает каждому найденному человеку ключ, почту и форму имени из справочника пространства.

## Мониторинг

Проба снаружи сервису бессмысленна: у него нет адреса. Вместо неё два правила в `prometheus/rules/contur_transcribe.yml` репозитория `infra`:

| Алерт                                | Когда                                                               | Обычная причина                       |
|--------------------------------------|---------------------------------------------------------------------|---------------------------------------|
| <code>ConturTranscribeStalled</code> | 30 минут подряд: встречи находятся, часть падает, опубликовано ноль | отказ распознавания или модели а5х    |
| <code>ConturTranscribeSilent</code>  | проходов не было больше 30 минут                                    | контейнер упал, завис или не поднялся |

Метрики уходят, только если задан `METRICS_PUSH_URL` (с ним — `METRICS_PUSH_USER`, `METRICS_PUSH_PASSWORD`, `METRICS_INSTANCE`); без адреса отправка молча выключена. Здоровое состояние без новых встреч выглядит как «найдено 59, пропущено 59, опубликовано 0» — пропущенные в основном уже опубликованы раньше.

# Где код и как запустить

- Репозиторий `Buharestskie/contur_transcribe`, Python 3.12, httpx, pydantic. Команды: `python -m app.cli check` — проверить все доступы, ничего не публикуя; `once` — один боевой проход (есть `--limit` и `--only`); `daemon` — бесконечный цикл, так работает контейнер.
- **Песочница:** `tools.create_sandbox` создаёт тестовую книгу, `tools.find_chat` находит чат тестового бота. Значения кладутся в `.env.sandbox` со своим `STATE_FILE`, команды запускаются с `--env-file .env.sandbox` — рабочая таблица и группа отдела не задеваются.
- Тесты — `pytest`, внешние сервисы замочаны.
- **Выкатка:** на сервер доставляются `app/`, `pyproject.toml`, `Dockerfile`, `docker-compose.yml`. Серверные `.env`, `secrets/`, `data/` и справочники не перезаписываются. Порядок: резервная копия `data/state.json` → `docker compose build reporter` → `check` и `once` разовым контейнером → `docker compose up -d reporter` → логи.

## На что наступают

- `data/state.json` **нельзя удалять или обнулять**. Пустой файл — повторная рассылка всех встреч за `LOOKBACK_DAYS` дней в группу и в таблицу. Перед любыми работами — копия.
  - **Не запускайте `once` поверх работающего демона**. Второй процесс упадёт с «Предыдущий запуск ещё работает»: состояние защищено блокировкой.
  - **Имена сотрудников в Толке и в справочнике расходятся**. Большинство заходит гостями и вписывает имя как придётся, иногда латиницей. Одно слово имени опознаёт человека, только если оно в отделе у одного. При двух Аннах «анна» бесполезна, и встреча уходит в общие. После любого изменения состава отдела прогоняйте `tools.sync_identities`.
  - `titlePatterns` **пишите по основе фамилии**, а не по полной форме: шаблон с «никитина» не совпадёт с «Никитиной».
  - **Листы РОПов сервис не создаёт**. Нет листа из `rops.json` — `check` сообщит об этом. «Общие встречи» и лист расшифровок создаются сами. Лист «Контроль РОПа+» считается формулами книги: после новых строк и месяцев проверьте, что диапазоны формул их охватывают.
  - **Файл для распознавания ограничен 110 МБ** (`A5X_MAX_UPLOAD_MB`, сервис a5x отказывает около 128 МБ). Без ffmpeg запись уходит видеофайлом и может не пройти.
  - **Сервер общий с другими боевыми контейнерами**, лимиты памяти и CPU в `docker-compose.yml` занижены намеренно.
  - **Три имени у одного проекта:** репозиторий `contur_transcribe`, на сервере `donula-kontur-reporter`, локальная папка часто `kontur-talk-reporter`. Ищите по всем трём.
-

Версия #1

Admin создал 2026-09-28 08:25:51 UTC

Admin обновил 2026-09-28 08:25:51 UTC