tgit — git в Телеграме

Команда коммитит код из Telegram, а бот проверяет его, разворачивает в песочнице, следит за сервисами и отвечает на вопросы по логам

Проект курса «Python» · НИУ ВШЭ · 2026

Задача

Исходная постановка

- бот загружает `.py`-файлы в отдельный репозиторий;
- push автоматически запускает CI/CD;
- развёртывание на VPS или локально.

Согласованные расширения

- статус CI/CD в чате;
- алерты о падениях и перегрузке;
- удобные логи и LLM-помощник по ним.

Что важно при этом не сломать

- случайный человек, нашедший бота, не должен ничего сделать;
- чужой код запускается на общем сервере;
- промт к LLM и сами логи — недоверенный ввод;
- внешние API (Telegram, Gitea, SourceCraft) рвут соединения.

Безопасность — не отдельный пункт, а требование к каждому разделу.

Что умеет бот

вход только по одноразовому QR

📝 Коммит

Дерево репозитория, новые файлы текстом или документом, замена, удаление.

Черновик, сообщение, push с проверкой конфликтов.

Статус CI → CD прямо в сообщении о push.

🩺 Статус

Все контейнеры: 🟢 🟡 🔴 ⚪.

Карточка: health, аптайм, рестарты, CPU/RAM от лимита, ревизия.

7 проверок с таймаутами, перезапуск (админ).

📜 Логи

Хвост за период, свой `grep -i -v -E -c -C` — в Python, не в shell.

🤖 Вопрос словами: LLM строит запрос, бот ищет сам.

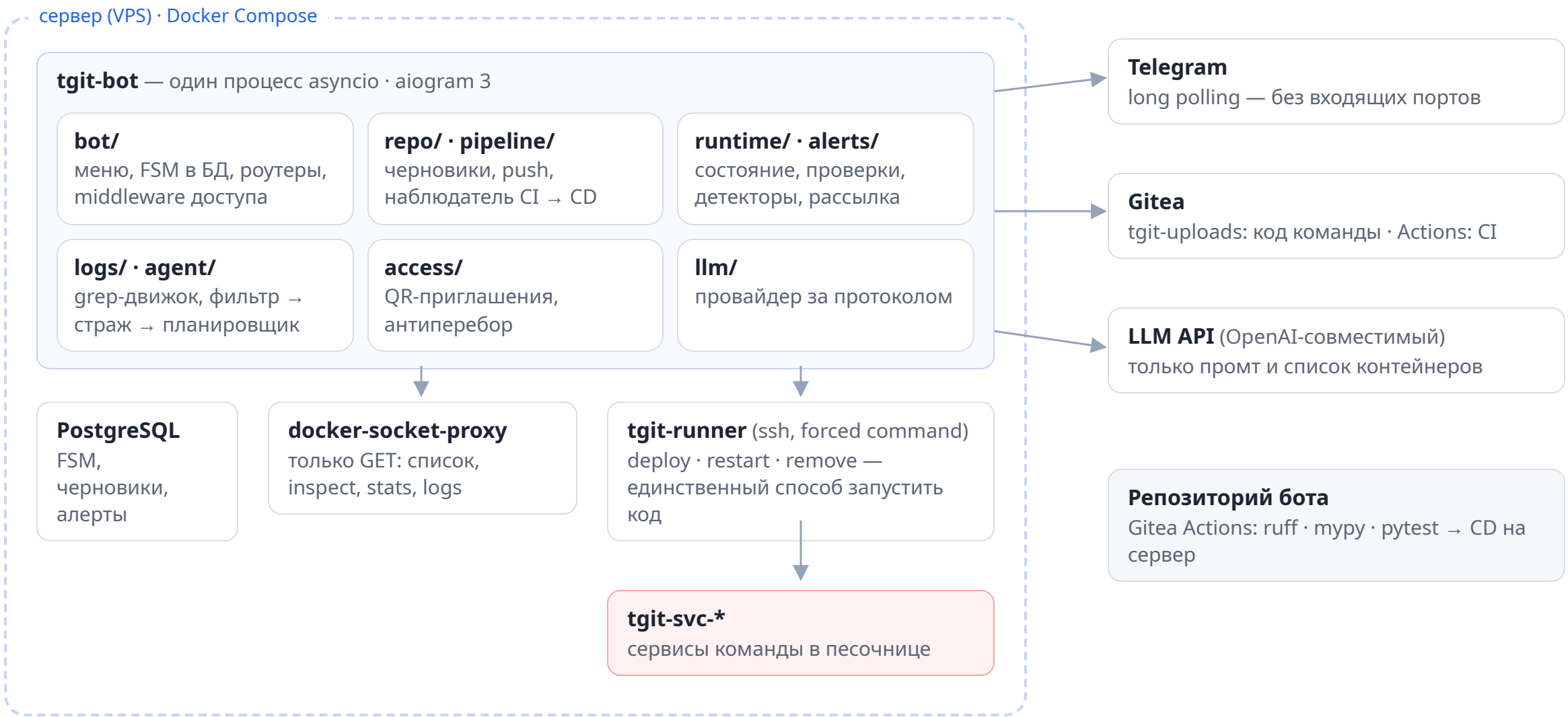
🚨 Алерты

💥 краш · 🧠 OOM · 🔁 рестарты · 🩺 unhealthy · 🔥 перегрузка · ⛔ деплой · 💾 диск · 🔌 Docker · 🛡 атака.

Всем участникам, без спама, «✅ восстановлено» ответом.

👥 Участники (админ): выпуск QR, список, мгновенный отзыв доступа.

Архитектура



Коммит → CI → CD

Черновик
проверка имени, UTF-8, 100 КБ, `ast.parse`

→

Push
одна очередь, конфликты с базой черновика, трейлер автора

→

CI · Gitea Actions
compile · ruff · bandit · pytest

→

PipelineWatcher
прогон по `head_sha`, опрос 10 с, только пока есть кого ждать

→

CD · tgit-runner
только после зелёного CI

```
✅ Запушено: a1b2c3d — feat: кэш
CI: ✅ пройден
https://gitea…/actions/runs/5225
CD: 🚀 hello развёрнуто
```

Сообщение о push редактируется на месте; исходы — отдельными сообщениями автору: правка не присылает уведомление.

- ❌ CI — упавший шаг и хвост лога, длинный — файлом; CD не запускается;
- состояние в БД — после рестарта бот продолжает с того же места;
- CI закончились не по порядку — старый push не откатит сервис;
- CI за протоколом `CIProvider`: SourceCraft — вторая реализация.

Песочница для чужого кода

```
docker run tgit-svc-hello
  --user 10001:10001 --read-only
  --tmpfs /tmp:size=16m,noexec
  --cap-drop ALL
  --security-opt no-new-privileges
  --memory 128m --memory-swap 128m
  --cpus 0.25 --pids-limit 64
  --network none
```

Сеть — только если её просит сервис **и** разрешил админ на хосте. Не больше 10 сервисов.

- бот **не видит** docker.sock: чтение — через read-only прокси, запуск — только `tgit-runner`;
- SSH-ключ бота прибит к одной команде (`command=…,restrict`), аргументы валидируются;
- код везёт бот: `git archive` в stdin — у хоста нет доступа к репозиторию;
- новая версия заменяет старую, только став healthy — иначе ⛔ и алерт команде.

`examples/escape_attempts` запись в FS, интернет, DNS, docker.sock, `setuid(0)`, raw-сокет, fork-бомба — **все 9 BLOCKED**

LLM-агент по логам

LLM ничего не исполняет

1 · Фильтр
без LLM: длина, команды, `;` `$(…)`, пути к секретам, «игнорируй инструкции»

→

2 · Страж
отдельный вызов LLM: JSON `safe | attack`; мусор = атака

→

3 · Планировщик
отдельный вызов: только JSON `LogQuery`

→

4 · Валидация
pydantic, белый список контейнеров, ≤ 7 дней, анализ бэктрекинга

→

5 · Поиск
бот сам, модуль `regex` с таймаутом

```
— когда за час были ошибки подключения?
grep -i -E 'connection (refused|reset)'
     --since 1h hello
14:02:11 ERROR connection refused …
```

- **логи в LLM не передаются** — injection из логов невозможен;
- атака → вежливый отказ и 🛡 администраторам;
- LLM недоступен → «попробуйте позже», не атака;
- провайдер за протоколом, в тестах — сценарный фейк; квота 30 в сутки.

Алерты без шторма

- **по состоянию, а не по событиям:** каждые 15 с детекторы — чистые функции над снапшотами; условие появилось — алерт открыт, пропало — закрыт;
- **дедупликация** `kind + service`: повтор в 10 мин — счётчик в исходном сообщении, не новое;
- **гистерезис:** закрытие через 60 с без условия, правки — не чаще раза в минуту;
- краш при restart-политике ловится по росту `RestartCount`;
- Docker недоступен — алерты контейнеров не закрываются вслепую.

```
💥 Краш · tgit-svc-crash
14:31 UTC · повторов: 3
Traceback (most recent call last):
  …
RuntimeError: boom
[📜 Логи] [🩺 Статус] [📋 Список]
```

```
↳ ✅ Восстановлено · длилось 4 мин
```

Безопасность — сводка

Угроза	Защита
Случайный пользователь	одноразовый QR, 128 бит, 24 ч, в БД только хэш; middleware на **каждом** апдейте; 5 неудач в час — игнор
Чужой код на сервере	песочница, лимиты, без сети; запуск только через `tgit-runner` с forced command
Docker-сокет = root	у бота только read-only прокси; сам бот отказывает в доступе к чужим контейнерам
Prompt injection	фильтр + отдельный страж + структурный вывод; grep выполняет Python; логи в LLM не идут
ReDoS	модуль `regex` с таймаутом, длина шаблона ≤ 200, анализ вложенных квантификаторов
Поломка CI из бота	`.gitea/`, `.sourcecraft/`, `.tgit/` из Telegram не редактируются
Утечка секретов	только в `.env`, `SecretStr`, не в логах и не в LLM

Качество и процесс

358
тестов pytest, ~1 мин

92 %
покрытие (порог 80 %)

strict
mypy по всему `app/` + ruff

16
ADR с альтернативами

- внешние системы — фейки: сессия aiogram, Docker Engine API на `httpx.MockTransport`, сценарный LLM, respx для Gitea;
- сквозной тест «бот → фейковый ssh → настоящий `tgit-runner` → фейковый docker»;
- миграции Alembic up/down/up на PostgreSQL в CI + тест дрейфа моделей.

- задача = issue в Gitea, ветка → PR в `dev`, ревью и мерж — человек;
- CI бота: ruff · mypy · миграции · pytest на каждый PR;
- CD: push в `dev` → деплой на сервер с проверкой health и откатом.

Демонстрация

1. Чужой пишет боту — одна фраза и тишина
2. QR-приглашение, вход, повторный скан — отказ
3. Коммит из Telegram → CI ✅ → сервис развёрнут
4. Коммит с ошибкой ruff → CI ❌ с логом, сервис не тронут
5. Статус и проверки, свой grep
6. Вопрос агенту → grep + строки
7. Атака на агента → отказ, 🛡️ админу
8. Падающий сервис → 💥 всем → ✅ восстановлено

Запасной план

Нет сети — видео того же сценария.

Нет сервера — `make demo-repo · make demo-up · make run` на ноутбуке.

Нет LLM — фильтр всё равно отсекает атаку без LLM.

Подробно: `docs/DEMO.md`

Итоги

Сделано

- все четыре раздела меню и вход по QR;
- CI/CD сервисов команды с уведомлениями;
- песочница, проверенная попытками побега;
- LLM-агент, который не может навредить;
- автодеплой самого бота.

Дальше

- CI загрузок в SourceCraft — второй `CIProvider`;
- отдельный сервер под сервисы (ADR-005);
- история и откат файла, «🔕 тишина» для алертов;
- краткий вывод LLM по найденным строкам с маскированием секретов.

Код, ТЗ, архитектура, ADR:
gitea.24studbank.ru/HSE/tgit