ITG Direct Dispatch v1 — прямой API ChatGPT → Hermes → bb

Назначение

Прямая передача заданий из ChatGPT в Hermes (PVE LXC 104) и далее в bb: подтверждение доставки, устойчивое состояние, реальный thread ID и получение результата. Зависимость запуска задач от GBrain, Honcho, cron и `agent-task-poll` исключена.

Дата ввода: 2026-11-23 · **Версия:** 1.0.0 · **Статус:** работает, end-to-end проверено с внешнего хоста.

Адреса и доступ

Что	Значение
REST API	`https://dispatch.itgsystem.online`
MCP (Streamable HTTP)	`https://dispatch.itgsystem.online/mcp`
Health	`https://dispatch.itgsystem.online/health`
Сервис	контейнер `itg-dispatch` на LXC 104, `/opt/itg-dispatch`, порт `192.168.1.165:9130`
Ingress	Traefik ZimaOS: `/DATA/AppData/itg-traefik/dynamic/dispatch.yml`, LE-сертификат
DNS	`dispatch.itgsystem.online` A → `94.41.128.114`, Cloudflare proxied=false
Токены	`/opt/itg-dispatch/dispatch.env` (сервер) и `/opt/data/.secrets/itg_dispatch.env` (Hermes), 0600
Код	https://github.com/mars485/itg-direct-dispatch (private)

Схема



`202 registered` от API не означает старт работы. Признак реального старта — состояние `dispatched` с непустым `bb_thread_id`.

Состояния

`registered` → `accepted` → `dispatched` → `in_progress` → `review` → `completed`; ветки `blocked` / `failed`. Переходы контролируются (недопустимый — `409`), `review` → `completed` выполняет владелец через `POST /api/v1/tasks/{id}/confirm`.

REST

Метод	Путь	Назначение
POST	`/api/v1/tasks`	создать задачу (`Idempotency-Key` обязателен)
GET	`/api/v1/tasks/{id}`	состояние, `bb_thread_id`, результат
GET	`/api/v1/tasks/{id}/events`	история событий и журнал доставок
GET	`/api/v1/tasks`	список задач (`?state=`, `?limit=`)

Метод	Путь	Назначение
POST	`/api/v1/tasks/{id}/confirm`	приёмка (`review` → `completed`)
GET	`/health`	API, БД, Hermes, bb

Авторизация: `Authorization: Bearer <DISPATCH_API_TOKEN>`. Внутренние `/internal/v1/...` — только с `DISPATCH_INTERNAL_TOKEN` (bridge/watcher).

MCP-инструменты

- `create_task(title, prompt, idempotency_key, project?, provider?, priority?)`
- `get_task_status(task_id)`
- `get_task_events(task_id, limit?)`
- `get_task_result(task_id)`

Подключение в ChatGPT: Settings → Connectors → Create → URL `https://dispatch.itgsystem.online/mcp`, Authentication **Token (bearer)**, токен `DISPATCH_MCP_TOKEN`.

Особенность: WAF/CrowdSec отклоняет `User-Agent: Python-urllib` (403). ChatGPT проходит; скриптовые клиенты должны задавать обычный User-Agent.

Результаты приёмки (2026-11-23)

Тест	Результат
Создание через REST	`202 registered`, `tsk_8e696556e98b42f1`
ACK Hermes	`accepted` через 2 с
Передача в bb	`dispatched`, `thr_askk9x9h77`, `thr_pfm4xyk8rm`
Результат обратно	`review` + текст bb
Внешний E2E через публичный MCP	`tsk_fbac8b1c5c744003` → `thr_pfm4xyk8rm` → `review` за 26 с
Повтор Idempotency-Key	`replayed=true`, дублей нет (MCP и REST)
Тот же ключ, другое тело	`409`
Неверный токен	`401` (+ `WWW-Authenticate`)
Нет Idempotency-Key	`400`

Тест	Результат
bb недоступен (bridge / watcher)	`blocked`: `prompt_upload_failed` / `bb_unreachable`
Пересоздание контейнера	состояние сохранено, `/health` → 200
Перезапуск Hermes	скрипт `/opt/itg-dispatch/restart_selftest.sh`, лог `/opt/itg-dispatch/restart-test.log`, отчёт в Telegram (топик 22)

Переключение и откат

- GBrain-poller (`agent-task-poll`, cron `b32f02d9df58`) **приостановлен**: очередь `itg-tasks-inbox` пуста (offset `last_processed_fact_id=158`).
- Откат новых задач на старый маршрут: `HERMES_HOME=/opt/data hermes cron resume b32f02d9df58`.
- Остановка сервиса: `cd /opt/itg-dispatch && docker compose down`; удалить `dispatch.yml` из Traefik (роутер снимется без рестарта).
- Бэкап БД: копия `/opt/itg-dispatch/data/dispatch.db` при остановленном контейнере; состояние — SQLite WAL.

Ограничения

- `blocked` после исчерпания повторов требует ручного разбора (`get_task_events`).
- Watcher запускается мостом внутри контейнера Hermes; при его гибели задача остаётся в `dispatched` до ручной проверки.
- Cloudflare Tunnel из LAN недоступен (провайдер режет QUIC/TLS к edge) — публикация только через существующий Traefik.

Инцидент 24.11.2026 — «Error executing tool create_task» и исправление

Симптом

ChatGPT при отправке задания через MCP получал:

```
INVALID_ARGUMENT
RuntimeException: Error calling MCP tool:
Error executing tool create_task
```

Без указания поля и причины.

Первопричина (проверена по коду и живому вызову)

1. Вызов содержал `project = "autocity"`, которого не было в `DISPATCH_ALLOWED_PROJECTS` (разрешены были `itg, ai-launcher, pink-garden, shopping-app, studioflow, vostok-balkon, todolist, ssh-mobile`).
2. В `create_task_tool` проверка аргументов `TaskCreate.model_validate(payload)` стояла **вне** `try/except`. Pydantic поднимал `ValidationError` внутри обработчика инструмента, MCP-сервер оборачивал его в `UnexpectedToolError`, и клиент видел только «Error executing tool create_task» — текст ошибки терялся. Задание не регистрировалось в БД, поэтому повторные попытки не помогали.

Что исправлено (коммит `fix(dispatch)` в `/opt/data/projects/itg-dispatch`)

Изменение	Файл
Валидация аргументов перенесена внутрь `try`; ошибка возвращается объектом `{ok:false, error_code:"validation_error", error, fields, allowed_projects, allowed_providers, allowed_priorities, hint}`	`app/main.py`
Проверка `idempotency_key` (8–128 символов) до валидации модели, отдельный код `invalid_idempotency_key`	`app/main.py`
Конфликт ключа возвращает `error_code: idempotency_conflict` и `existing_task_id` (новый `store.get_by_idem_key`)	`app/main.py`, `app/store.py`
`redact()` вычищает значения токенов/паролей сервиса из любого текста ошибки	`app/main.py`
REST 422 возвращает то же структурированное тело (поле + причина, без эха входных данных)	`app/main.py`
`/health` отдаёт `errors.total_since_start` и `errors.recent[]` (код, время, сообщение) — без тел запросов и секретов	`app/main.py`

Изменение	Файл
Описание инструмента `create_task` в схеме MCP перечисляет допустимые `project`, `provider`, `priority`	`app/main.py`
`autocity` добавлен в `DISPATCH_ALLOWED_PROJECTS` (`/opt/itg-dispatch/dispatch.env`) и в `PROJECT_MAP` bridge (→ `proj_tbkm8jzqri`)	конфиг, `scripts/itg_dispatch_bridge.py`

Параметры `title`, `prompt`, `idempotency_key` объявлены как `str | None = None`, поэтому `null` из клиента теперь даёт читаемую ошибку валидации, а не обрыв вызова.

Мониторинг сбоев

- Скрипт `scripts/dispatch_watchdog.py` (cron Hermes на LXC 104, каждые 10 минут, `no-agent`, доставка в Telegram — тема «Инфраструктура»): проверяет `/health` (API, БД, Hermes-гейтвей, bb), новые `blocked/failed` за сутки, зависшие `registered/accepted/dispatched`, рост `internal_error` и серию из 3+ ошибок валидации `create_task`; повторные уведомления подавляются файлом состояния `/opt/data/state/dispatch_watchdog.json`.
- Cron-задача: id `72504bd2b4a3`, «ITG Direct Dispatch — watchdog (сбои → Telegram)».

Проверено 24.11.2026

- Три последовательных сквозных теста через публичный MCP `https://dispatch.itgsystem.online/mcp` из внешнего хоста (VPS 195.208.61.169): `tsk_6e3c7ee874344313`, `tsk_75bb274a41ba4681`, `tsk_62e890a6f8664bc9` — все дошли до `review` с результатом bb (`thr_ad83a3sw6q`, `thr_nytsrbpcwb`, `thr_47ra3dtnm7`).
- Контрольный прогон после финальной сборки: `tsk_300d01cbee0440b8`, `tsk_d5bb23353cd84fd8`, `tsk_41a6fef05b7f432f`.
- Идемпотентность: повтор вызова с ключом `autocity-close-2026-11-24` вернул тот же `tsk_3a77a11ad0204a5d` (`replayed: true`), дубль не создан.
- Негативные проверки: недопустимый `project` → `validation_error` с перечнем допустимых; пустой ключ → `invalid_idempotency_key`; `prompt: null` → читаемая ошибка поля.

Откат

```
# код
docker compose -f /opt/itg-dispatch/compose.yaml down
cp /opt/itg-dispatch.bak-20261124-131108/app/*.py /opt/itg-dispatch/app/
docker compose -f /opt/itg-dispatch/compose.yaml up -d --build
# конфиг
```

```
cp /opt/itg-dispatch/dispatch.env.bak-* /opt/itg-dispatch/dispatch.env
# БД
# /opt/itg-dispatch/backups/dispatch-20261124-131108.db
```

Откат кода не затрагивает данные: схема БД не менялась.

Revision #2
Created 2026-11-23 17:59:41 UTC by Айнур
Updated 2026-11-24 13:55:22 UTC by Айнур