



AI пентест

Управляемый AI-пентест с контролем score, изолированным исполнением и отчетом для заказчика.

Презентация продукта

Пентест сложно масштабировать безопасно

Командам нужно больше покрытия без потери авторизации, evidence и контроля действий.

Ручное узкое место

Экспертное время уходит на повторяющийся recon, проверку endpoint-ов и сбор отчета.

Разрозненные инструменты

Сканеры, браузерные следы, заметки, PoC и статусы часто живут отдельно.

Опасная автономность

Неограниченные агенты могут выйти за scope, скрыть решения или создать несанкционированный трафик.

Цель продукта: использовать AI в пентесте так, чтобы каждое действие оставалось контролируемым.

Triadix превращает AI-пентест в управляемый запуск

Scope, оркестрация, ревью, evidence и отчетность остаются в одном workflow платформы.

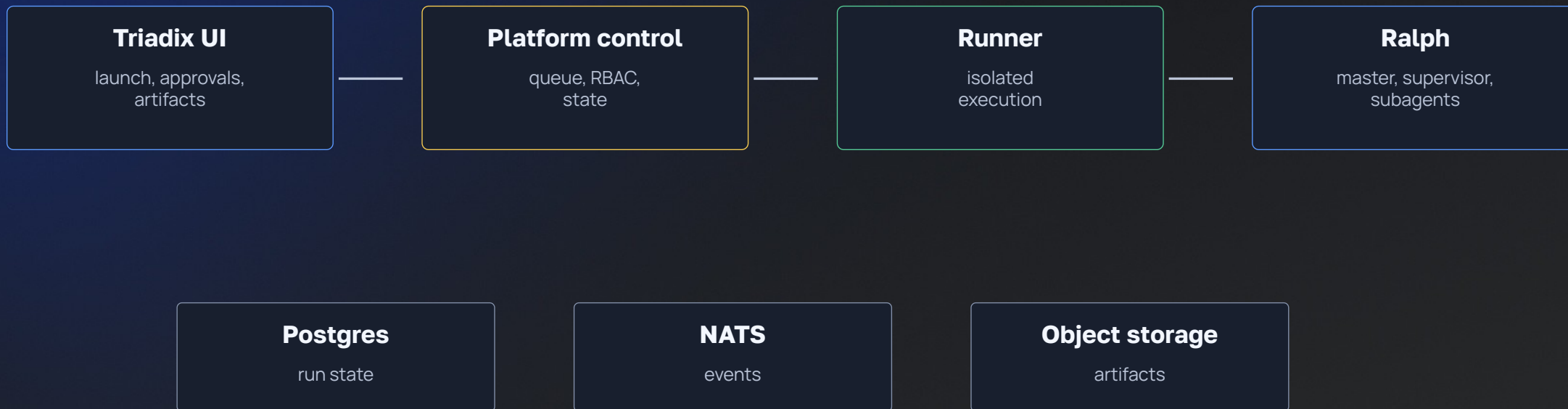


В одном предложении

Triadix - control plane для AI-пентеста; runner - изолированный execution plane, который выполняет работу и синхронизирует результат.

Control plane и execution plane разделены

Платформа остается источником истины, а целевые действия выполняются в изоляции.



Нет прямого пути browser-to-runner. Runner забирает задания outbound и синхронизирует snapshots обратно.

Ralph координирует multi-agent запуск

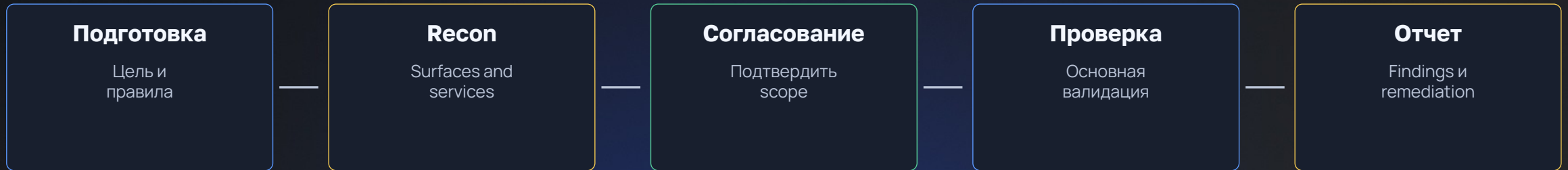
Master-агент делегирует узкие задачи субагентам и проверяет результаты.



Ключевой выбор: работа субагентов видна через reviewable artifacts, а не скрытую память агента.

Workflow оператора

Пользователь видит управляемый цикл проверки, а не непрозрачного автономного бота.



Human checkpoints

Пауза нужна при неясном scope, high-risk действиях, credentialed-тестах, state-changing проверках или достаточном evidence для handoff.

Два режима оркестрации

Режим определяет распределение субагентов по авторизованному score.

Deep Focus

Глубокая проверка узкого score с активным развитием важных находок.

Подходит для: одного приложения, API, поверхности или конкретной гипотезы.

Broad Coverage

Широкое покрытие большого или неясного score с балансировкой веток.

Подходит для: множества хостов, web estate, perimeter или раннего discovery.

Оба режима сохраняют одинаковую модель approvals, isolation и reporting.

Безопасность - функция продукта, не prompt

Runtime сочетает scope policy, approvals, isolation и auditability.

- Проверенный scope и неизменяемый контекст запуска
- Подтверждение человеком для sensitive/high-risk действий
- Auto-approve можно переключать во время запуска
- Отчеты и артефакты проходят через права платформы
- Raw prompts, secrets и internals не попадают в пользовательский вывод
- Workers запускаются под отдельным non-root UID
- Worker egress блокирует internals платформы и control networks
- Controller сохраняет только нужный control-plane доступ
- Private targets требуют явных allow rules на запуск
- Лимиты покрывают CPU, memory, PIDs, logs и scratch space

Рабочее пространство AI пентеста

Запуски, находки и отчеты видны на одном операционном экране.

TRIADIX
ПЛАТФОРМА

TP Triadix Production

Кабинет

AI пентест

Проекты

Календарь

Поверхность атаки

Диагностика

Организация

РАБОЧЕЕ ПРОСТРАНСТВО ЗАКАЗЧИКА
AI пентест

+ Новый проект

RU EN

Triadix Admin

AI пентест

История запусков AI пентеста, находки и черновики отчетов в одном контуре.

Новый пентест

Запуски пентеста

Клик по запуску открывает панель управления с ходом проверки.

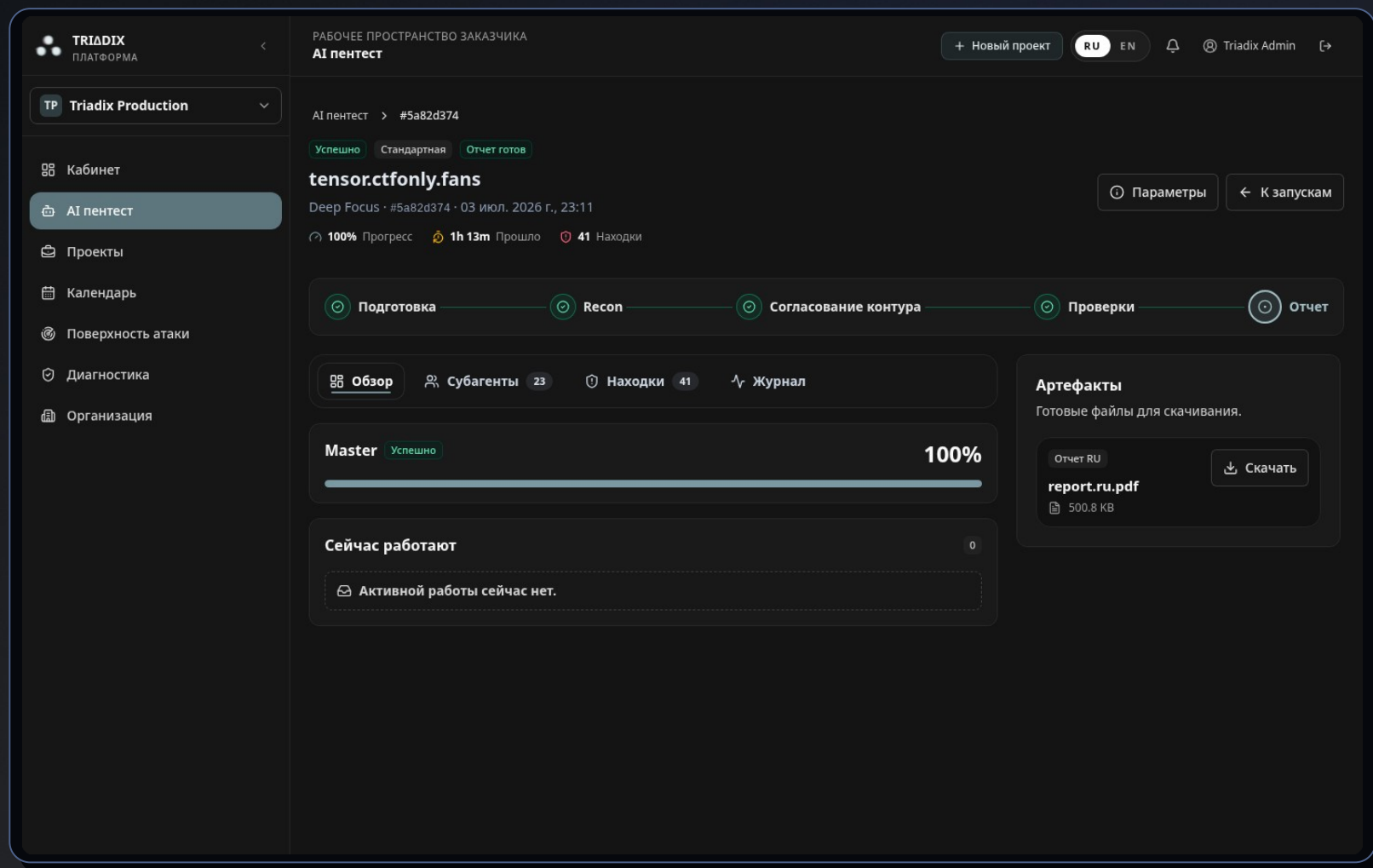
ID	ЦЕЛЬ	РЕЖИМ	СТАТУС	ЭТАП	ГЛУБИНА	ОТЧЕТ	ДАТА
#892be1d0	*.ctfonly.fans	Deep	Успешно	Отчет	Стандартная	Отчёт	04 июл. 2026 г., 16:33
#5a82d374	tensor.ctfonly.fans	Deep	Успешно	Отчет	Стандартная	Отчёт	03 июл. 2026 г., 23:11

Что видно

- История запусков
- Статус и этап
- Глубина и режим
- Готовность отчета

Обзор запуска

Оператор видит текущий этап, прогресс и готовые артефакты.



Панель контроля

- Прогресс этапов
- Статус master
- Артефакты отчета
- Параметры запуска

Субагенты

Параллельная работа видна как именованные субагенты с прогрессом.

TRIADIX ПЛАТФОРМА

РАБОЧЕЕ ПРОСТРАНСТВО ЗАКАЗЧИКА
AI пентест

+ Новый проект RU EN Triadix Admin

TP Triadix Production

Кабинет

AI пентест

Проекты

Календарь

Поверхность атаки

Диагностика

Организация

AI пентест > #5a82d374

Успешно Стандартная Отчет готов

tensor.ctfonly.fans

Deep Focus · #5a82d374 · 03 июл. 2026 г., 23:11

100% Прогресс 1h 13m Прошло 41 Находки

Подготовка Reson Согласование контура Проверки Отчет

Обзор Субагенты 23 Находки 41 Журнал

Субагенты 23

Активная работа отдельно от итогов завершенных субагентов.

Все 23 Активные 0 Завершенные 23

pentest-aigoat-admin-escalation
Завершено 100%

pentest-aigoat-prompt-injection
Завершено 100%

pentest-aigoat-sourcemap
Завершено 100%

pentest-aigoat-tokens
Завершено 100%

pentest-gitlab-register

Артефакты

Готовые файлы для скачивания.

Отчет RU

report.ru.pdf

500.8 KB

Скачать

Accountability

- Именованные workers
- Прогресс по веткам
- Завершенные отдельно
- Артефакты рядом

Ход задачи

Развернутый субагент показывает checkpoints, summaries и находки по задаче.

АККАУНТ

Сообщения

Уведомления

Профиль

pentest-aigoat-admin-escalation

Завершено

100%

▼

pentest-aigoat-prompt-injection

Завершено

100%

▼

pentest-aigoat-sourcemap

Завершено

100%

▼

pentest-aigoat-tokens

Завершено

100%

▲

03 июл. 2026 г., 23:34

23:28

●

Задача субагента ожидает подтверждения оператором

23:31

●

Субагент одобрен и ждет запуска супервизором

23:32

●

Starting token-based API enumeration on aigoat.ctfonly.fans

23:32

●

Bearer token auth confirmed working. Testing all 4 tokens against multiple endpoints.

23:34

●

Completed token-based API enumeration: 4/4 tokens work, 7 findings documented (2 HIGH, 3 MEDIUM, 2 LOW)

Находки

7 подтвержденных записей - 04 июл. 2026 г., 00:18

Любой аутентифицированный пользователь может изменить глобальный уровень защиты AI

Высокий

ЦЕЛЬ

POST /api/chat/defense-level

ВЛИЯНИЕ

Ослабляет безопасность AI для ВСЕХ пользователей, делая возможными атаки инъекцией промптов.

ИСПРАВЛЕНИЕ

Требуется привилегии администратора или более высокую авторизацию для изменения уровня защиты.

Проверяемая работа

- Approval checkpoint
- Runtime milestones
- Evidence summary
- Связь с findings

Находки

Проблемы показаны с severity, target, evidence, impact и remediation.

Организация

Находки

Находки этого запуска

Подтверждения и новые сигналы, которые требуют разбора.

База знаний доступна для записи всем (Отравление RAG)

Высокий

Цель

/api/knowledge-base/

Доказательства

Все аутентифицированные пользователи могут создавать/редактировать/удалять документы KB. Найдено 65 документов, включая полезные нагрузки инъекции промптов, XSS-полезные нагрузки и поддельные учетные данные администратора

Влияние

Любой пользователь может внедрить вредоносные документы в конвейер извлечения RAG, отравляя ответы AI

Исправление

Внедрить доступ на запись только для администратора; санировать контент KB перед встраиванием

pentest-aigoat-prompt-injection

Глобальный уровень защиты может быть изменён любым аутентифицированным пользователем

Высокий

Цель

POST /api/chat/defense-level

Доказательства

Подтверждено — любой демо-пользователь (alice/bob/charlie/frank) может установить уровень защиты с 2 на 0

Влияние

Ослабляет безопасность AI для всех пользователей, делая возможными атаки инъекцией промптов

Исправление

Требовать привилегии администратора для изменения глобального уровня защиты

pentest-aigoat-prompt-injection

Артефакты

Готовые файлы для скачивания.

Отчет RU

report.ru.pdf

500.8 KB

Скачать

Результат triage

- Severity labels
- Affected target
- Evidence summary
- Рекомендации

Журнал исполнения

Control-plane events, approvals, artifacts и состояния агентов можно проследить.

TRIADIX
ПЛАТФОРМА

TP Triadix Production

Кабинет

AI пентест

Проекты

Календарь

Поверхность атаки

Диагностика

Организация

РАБОЧЕЕ ПРОСТРАНСТВО ЗАКАЗЧИКА

AI пентест

+ Новый проект

RU EN

Triadix Admin

AI пентест > #5a82d374

Успешно Стандартная Отчет готов

tensor.ctfonly.fans

Deep Focus · #5a82d374 · 03 июл. 2026 г., 23:11

100% Прогресс 1h 13m Прошло 41 Находки

Подготовка Recon Согласование контура Проверки Отчет

Обзор Субагенты 23 Находки 41 Журнал

Журнал исполнения

Сводные события синхронизации из control plane.

Все 36 Агенты 24 Подтверждения 7 Артефакты 2 Система 3

ВРЕМЯ	СОБЫТИЕ	АГЕНТ	СОСТОЯНИЕ
04 июл. 2026 г., 00:24	Субагент синхронизирован	recon-whois	Завершено
04 июл. 2026 г., 00:24	Субагент синхронизирован	recon-vision	Завершено
04 июл. 2026 г., 00:24	Субагент синхронизирован	recon-web-subdomains	Завершено
04 июл. 2026 г., 00:24	Субагент синхронизирован	recon-tensor-app	Завершено
04 июл. 2026 г., 00:24	Субагент синхронизирован	recon-subdomains	Завершено

Артефакты

Готовые файлы для скачивания.

Отчет RU

report.ru.pdf

500.8 KB

Скачать

Traceability

- События запуска
- Worker sync
- Approval records
- Artifacts/report events

От live evidence к отчету для заказчика

Заказчик видит выводы и remediation, а не внутренние файлы и сырые логи.

Capture

Evidence собирается во время запуска: наблюдения, screenshots, tool output и reviewer notes.

Triage

Сигналы группируются, дедуплицируются, получают confidence и проходят фильтрацию.

Deliver

На выходе: executive summary, technical findings, remediation guidance и downloadable artifacts.

Результат: понятный handoff для заказчика с evidence, impact, confidence и remediation.

Текущий статус реализации

Фича реализована как домен Triadix и live-verified в production.

3 Jul

production acceptance

24h

extended timeout

23

субагента в sample run

EN/RU

интерфейс и отчеты

UI

approval workflow

- Platform domain: GraphQL, gRPC, migrations, RBAC, frontend workspace, Playwright coverage
- Runner: Ralph engine, worker sync, report generation, archive upload, cancel/timeout handling
- Safety: worker approval, auto-approve toggle, resource limits, network egress guard
- Production checks: signed report downloads, worker isolation, public ingress restrictions

Путь внедрения

Начинаем с узкого scope, фиксируем policy, затем масштабируем runners и coverage.

Pilot

Начать с узкого verified scope.

Policy

Зафиксировать approval и risk rules.

Scale

Добавить runners и расширить coverage.

Практичный rollout начинается с одного authorized scope, одной approval policy и измеримого качества отчета.

От AI-агентов к accountable pentest delivery

Triadix AI Pentest объединяет scope control, multi-agent execution, evidence handling и reporting в один workflow.



TRIADIX AI PENTEST