Agent QA Toolkit
Portable Evidence Packs для tool-using AI агентов
- Offline, self-hosted, без утечки данных
- CI gate decisions + security сигналы
- ЦА: platform/QA инженеры, продакшн-агенты

Проблема
- Хэндовер инцидентов = скриншоты + частичные логи
- Передать run без доступа к UI сложно
- Tool I/O и retrieval контекст разрознены
- Секреты/PII делают шаринг рискованным

Решение
- Один run → один Evidence Pack
- report.html (человеку) + compare-report.json (CI контракт)
- assets/ + manifest.json с sha256 целостностью
- Strict verify: portability + integrity checks

Как работает
- Runner запускает кейсы
- Evaluator сравнивает baseline vs new
- Выдаёт Evidence Pack (offline, shareable)
- CI читает gate decision per-case

Что даёт
- Regression diffs по кейсам
- Root cause attribution (RCA)
- Security scanners (PII, injection, exfiltration, risk)
- Self-contained, portable report directory

Почему не Promptfoo / DeepEval / LangSmith
- Они про model quality и hosted traces
- Мы про agent behavior + offline handoff
- Evidence Pack можно вложить в тикет
- Без SaaS зависимости

Деплой
- Только self-hosted
- Работает в air-gapped
- Нет аккаунтов и внешнего хранения
- Вы контролируете ретеншн и шаринг

Цены
- Free: open-source core
- Services-first: onboarding + CI setup + custom rules
- Pro (позже): rules library + trending
- Enterprise: advanced redaction + support SLA

Use Cases
- Release gates перед деплоем новой версии агента
- Support bundles для клиентов/вендоров
- Evidence packs для compliance команд

Call to Action
- Пилоты: 5–10 команд
- Подключаем агента за 1 день
- Начните с demo bundle и CI gate