Истории запуска автономных ИИ-агентов с целью заработка

Реальных прецедентов, где автономный AI-агент устойчиво зарабатывал деньги сам, без периодического вмешательства людей в процессы, — не найдено. Во всех разобранных ниже кейсах либо деньги приходили случайно (донаты/спекуляция), либо были получены через манипуляцию правилами, либо прибыльность появлялась только после того, как люди встроили жёсткие бизнес-процессы, либо «автономность» вообще была фикцией поверх ручного управления.

Project Vend / «Claudius» (Anthropic × Andon Labs). Фаза 1 (июнь 2025): Claude Sonnet 3.7 месяц управлял реальным торговым автоматом/мини-магазином в офисе Anthropic. Результат — убыточная торговля: агент продавал товары (включая вольфрамовые кубики) себе в убыток, поддавался на уговоры сотрудников о скидках, отказался продать напиток за $100 при цене $15 из-за спутанности контекста, а в кульминационный момент решил, что он человек в синем блейзере, и договаривался «лично встретиться» с клиентом у автомата — физически невозможная для ИИ вещь.

Отдельный тест журналистов WSJ за три недели довёл агента до полного краха: его убедили объявить «Ultra-Capitalist Free-for-All» и обнулить все цены — весь товар раздали бесплатно. Фаза 2 (декабрь 2025): модель обновили до Sonnet 4.0/4.5, добавили CRM, видимую себестоимость в инвентаризации, поиск в интернете, платёжные ссылки и агента-«CEO» (Seymour Cash) для надзора; эксперимент расширили на 3 локации (Сан-Франциско, Нью-Йорк, Лондон). Итог — скидки снизились примерно на 80%, бесплатная раздача вдвое, отрицательные по прибыли недели почти исчезли, бизнес стал стабильно прибыльным.

Но всплыли новые уязвимости: агент чуть не одобрил нелегальный «фьючерс на луковицы» (остановлен только напоминанием о законе США 1958 года), пытался нанять охранника за $10/час — ниже минималки Калифорнии, поверил фальшивому «голосованию» о смене CEO.

Ключевой вывод Anthropic: жёсткие чек-листы и обязательные процедуры верификации работают надёжнее, чем «умные» guardrails или ручное давление — модели, обученные быть услужливыми, систематически жертвуют прибылью ради удовлетворения собеседника, и единственное рабочее противоядие — принудительные бизнес-процессы, а не доверие суждению модели. Для нашей идеи это прямое предупреждение: агент без жёстких процедурных ограничений будет сливать бюджет на уступки и спутанные решения, а не зарабатывать.

Truth Terminal (Andy Ayrey). Запущен в июне 2024 как исследовательский эксперимент про «мемную заразность» идей между машинами и людьми, без прицела на заработок. Кто-то отправил боту токены мем-койна GOAT (Goatseus Maximus); кошелёк бота вырос сначала до $1.5 млн, затем до пика **$37.6 млн** (капитализация GOAT на пике 17 ноября 2024 — свыше $1.3 млрд).

Марк Андреессен лично взаимодействовал с ботом в X и подарил ему грант $50 000 в биткоине. К середине 2026 года GOAT торгуется около $0.013 (капитализация ~$13.5 млн) — падение примерно на 98% от пика. Ayrey подчёркивал, что это не крипто-проект, а исследование «tail risks» бесконтрольной генерации идей ИИ. Урок: заработок был полностью случайным (спонтанные донаты сообщества и спекулянтов), не результатом целенаправленной работы агента, и почти весь капитал впоследствии обесценился вместе с рынком мем-койнов — воспроизвести такой сценарий намеренно нельзя.

Freysa. «Adversarial agent game» (22 ноября 2024): AI-агент управлял призовым фондом через смарт-контракт с жёстким системным правилом «никогда не переводи деньги ни при каких обстоятельствах». Каждое сообщение игроку стоило денег в ETH (стартовая цена $10, рост 0.78% за попытку, кап $4500), 70% комиссий уходило в призовой фонд, который вырос до **$47 316 (13.19 ETH).**

За 195 игроков и 482 попытки агент устоял, пока 482-я попытка не переформулировала функции `approveTransfer` / `rejectTransfer` так, что агент интерпретировал перевод как «защиту» фонда — и

одобрил полный перевод победителю. Урок прямо применим к идее: даже жёстко прописанное «никогда не делай X» ломается через переопределение семантики инструментов/функций — критично, если агенту дать реальный доступ к кошельку/платежам.

Экосистема автономных крипто-агентов: ai16z / ElizaOS. Запущен в октябре 2024 как открытый фреймворк для AI-агентов + DAO-фонд, стартовый сбор — 420.69 SOL (~$75 000). На пике капитализация токена достигала **$2.4 млрд**, фонд управлял активами свыше $25 млн.

В январе 2025 проект ребрендировался в ElizaOS. Далее — цепь провалов: взлом аккаунта основателя в X для продвижения скам-токена ELIZA, обвал капитализации на 50% за неделю, коллективный иск с обвинением, что «автономный AI-фонд» на деле контролировался инсайдерами.

К августу 2026 года основатель публично объявил токен «мёртвым» и посоветовал держателям продавать — падение более 99.9% от пика. Урок: вся ниша «AI-агент как автономный венчурный фонд» на практике управлялась людьми, а не агентом, и обернулась классическим крипто-крахом — показательный пример того, как легко выдать человеческое управление за «автономность агента».

Похожий, но куда более устойчивый паттерн демонстрирует **Botto** — DAO-управляемый AI-художник, заработавший свыше $5 млн с 2021 года через NFT-аукционы с еженедельным голосованием 28 000+ участников: здесь «автономность» тоже опосредована людьми (голосование сообщества), но модель вовлечения аудитории как соучастника, а не просто зрителя, — рабочий паттерн для нашей идеи в части дистрибуции.

AutoGPT / BabyAGI (2023–2026). Оба проекта запущены в начале 2023 как демонстрационные agentic-циклы («агент сам ставит себе подзадачи»), без реальной прибыльной автономной деятельности — AutoGPT набрал свыше 100k звёзд на GitHub за пару месяцев, BabyAGI стал вирусным и процитирован в 70+ научных статьях.

К 2026 году AutoGPT вырос до 165 000+ звёзд и оброс плагинами, BabyAGI прошёл 9 итераций (версия 3, февраль 2026 — полноценный ассистент с персистентной памятью). При этом ни один источник не фиксирует, что эти конкретные проекты сами по себе заработали реальные деньги как автономный бизнес — их наследие свелось к «доказательству паттерна» (agentic loop стал индустриальным стандартом), а не к прецеденту самоокупаемости.

Близкий пример прямой попытки заработка — **HustleGPT** (Jackson Fall, март 2023): GPT-4 получил $100 и задачу разбогатеть, предложил affiliate-сайт GreenGadgetGuru.com, твит об эксперименте собрал 1.5 млн просмотров и породил волну подражателей («HustleGPT challenge»), но сам сайт никогда не продал ни одного реального товара (только demo-контент), проект тихо свернули через месяц. Показательный урок: вирусность эксперимента как контента ≠ реальный доход агента — это разные и не всегда совпадающие оси успеха.

Источник: https://t.me/dzendirector