🤖 Claude Managed Agents: готовые API для AI-продуктов

🚀 Введение в Claude Managed Agents

Ключевые тезисы:

- Claude Managed Agents — это набор готовых к продакшену API-эндпоинтов для создания продуктов на основе агентов.
- Платформа предоставляет примитивы (строительные блоки): агентов, окружения, сессии и события.
- Позволяет разработчикам сосредоточиться на продукте, а не на инфраструктуре (агентные циклы, сэндбоксы, контекст, аутентификация).
- Включает мощные функции: мульти-агенты, управление памятью, контроль разрешений, интеграция MCP и сквозная наблюдаемость.

🎯 Основные концепции и примитивы

Агент

Агент — это шаблон, определяющий поведение Claude в рамках продукта.

- **Системный промт**: Задает роль и инструкции для агента.
- **Набор инструментов (Tools)**: Можно выбирать, к каким инструментам (bash, веб-поиск, чтение файлов) у агента будет доступ. Можно ограничивать доступ в целях безопасности.
- **Интеграция MCP-серверов**: Агент может быть подключен к нужным MCP-серверам (например, Linear, Figma) для доступа к данным.
- **Контроль разрешений**: Можно настроить автоматическое выполнение одних инструментов (чтение файлов) и требовать подтверждения пользователя для других (выполнение bash-команд).

- **Версионность**: Агенты имеют версии, что позволяет откатиться к предыдущей конфигурации.

Окружение (Environment)

Окружение определяет шаблон для сэндбоксов, в которых работает Claude.

- **Сетевой доступ**: Можно разрешить или запретить доступ к сети.
- **Предустановленные пакеты**: Можно заранее установить необходимые пакеты из npm или pip.
- **Самостоятельный хостинг (Self-hosted)**: Можно использовать собственные контейнеры на инфраструктуре Cloudflare, Modal, Vercel или своей собственной, не выходя за пределы своего VPC.

Сессия (Session)

Сессия — это продолжающийся диалог с Claude (аналог нового чата в claude.ai).

- **Идентификаторы**: Для запуска нужны ID агента и окружения.
- **Ресурсы**: При создании можно подключить GitHub-репозитории или загрузить файлы, которые будут предустановлены в контейнере.
- **Поток событий**: Сессия представляет собой поток событий, куда клиент отправляет сообщения пользователя и получает ответы от Claude.

События (Events)

События — это способ взаимодействия с сессией и отслеживания её состояния.

Типы событий:

- **Пользовательские события (User Events)**: Сообщения, изображения, документы от пользователя или платформы.
- **Прерывания (Interrupts)**: Позволяют остановить Claude, если он выполняет нежелательные действия.
- **Результаты инструментов (Tool Results)**: Ответы от пользовательских инструментов, выполняемых на стороне платформы.

- **Подтверждения (Confirmations)**: Ответы пользователя для инструментов, требующих подтверждения (human-in-the-loop).
- **Цели (Outcomes)**: Позволяют отправить Claude файл или текст-спецификацию. Claude будет итеративно проверять свою работу против этой спецификации, пока не удовлетворит критериям. 💡 Мощный способ настройки агентов на успех.

События агента (Agent Events): Действия Claude: сообщения, "уплотнение" контекста, вызов инструментов (MCP или стандартных), координация мульти-агентов.

События сессии (Session Events): События жизненного цикла (начало, завершение, ошибки, повторные попытки, бездействие).

События промежутков (Span Events): Оповещения о начале и окончании долгих операций (например, генерации большого документа), чтобы клиент не думал, что процесс завис.

🛠️ Дополнительные возможности и интеграции

🔐 Хранилище учетных данных (Credential Vaults)

Позволяет безопасно хранить токены аутентификации для MCP-серверов (например, Linear) на стороне Anthropic. Токены инжектятся в сессию по необходимости и **никогда не попадают в контекстное окно Claude**, что повышает безопасность.

🧠 Хранилища памяти (Memory Stores)

Дают Claude доступ к постоянной памяти, в которую он может записывать информацию из одной сессии и читать её в последующих. Это позволяет агенту "учиться" и становиться лучше со временем. Разработчик может просматривать и редактировать эти воспоминания.

🌉 MCP-туннели (MCP Tunnels)

Позволяют безопасно подключаться к приватным MCP-серверам, расположенным внутри корпоративного периметра (без выхода в интернет). Claude получает доступ к данным через защищенный туннель.

👁️ Наблюдаемость и консоль разработчика

- **Живой мониторинг сессий**: Можно в реальном времени смотреть, что делает каждый агент в мульти-агентной системе, видеть вызовы инструментов и их результаты.
- **Просмотр хранилищ памяти**: Видеть, какую информацию Claude запоминает.
- **Быстрый старт (Quick Start)**: Интерактивный помощник на базе Claude, который помогает создавать агентов и сессии.
- **Готовые шаблоны**: Предопределенные конфигурации агентов для быстрого начала работы.

💡 Выводы и преимущества

- **Готовность к продакшену**: Claude Managed Agents избавляет от необходимости самостоятельно строить сложную инфраструктуру для агентов ИИ.
- **Композируемость и гибкость**: Можно выбирать только нужные примитивы и строить на их основе уникальные продукты.
- **Безопасность и контроль**: Гибкая настройка разрешений, безопасное управление учетными данными и приватными данными через самохостинг и туннели.
- **Мощные паттерны**: Встроенная поддержка мульти-агентов, итеративной работы по спецификациям (Outcomes) и долговременной памяти открывает возможности для создания сложных продуктов.
- **Ускорение разработки**: Использование Claude (например, через Claude Code с навыком Claude API) позволяет быстро прототипировать и строить интеграции, используя саму платформу для её же разработки.