

OpenAI Presence продает корпоративных ИИ-агентов в сопровождении инженеров

Источник: AI News Healthcare

Дата публикации: 2026-07-22

Оригинал: <https://www.artificialintelligence-news.com/news/openai-presence-enterprise-ai-agents/>

автоматизация бизнес-процессов

интеграция

корпоративный ИИ

управление ИИ

Новый способ приобретения корпоративных ИИ-агентов (AI agents) от OpenAI не подразумевает никаких онлайн-покупок. **OpenAI Presence**, анонсированный 22 июля, представляет собой управляемый продукт, поставляемый в рамках программы ограниченной общей доступности (Limited General Availability, LGA), и компания прямо заявляет, что на данный момент он недоступен как продукт самообслуживания. Развертывание осуществляется силами собственных инженеров по оперативному развертыванию (Forward Deployed Engineers, FDE) компании OpenAI и набора выбранных глобальных системных интеграторов.

Это является отступлением от бизнес-модели компании, которая в основном опиралась на API-ключи и лицензии на количество рабочих мест. **Presence** продается как проект, а не как продукт. Каждое взаимодействие начинается с одной конкретной задачи, такой как разрешение спора по счетам, обработка страхового требования или выполнение запроса сотрудника в ИТ-службу.

Агенту предоставляются только те знания и системный доступ, которые необходимы для выполнения задачи, а клиент сам прописывает правила, определяющие, что агент может делать, когда ему требуется

подтверждение (sign-off) и когда управление должен перехватить человек. После запуска **Codex** анализирует рабочие сессии и случаи эскалации, а затем предлагает изменения, которые команда заказчика тестирует и утверждает перед внедрением.

Документация OpenAI необычайно откровенна в отношении трудозатрат. В её справочном центре описан шестиэтапный процесс: от определения рамок бизнес-результатов через проверку безопасности, конфиденциальности и юридическую экспертизу, симуляцию и приемочное тестирование, поэтапное развертывание до итераций после запуска. В компании заявляют, что агент **Presence** не становится готовым к промышленной эксплуатации (production-ready) просто путем загрузки документов.

Проблема, которую призван решить этот продукт, реальна

Управляемую модель легко интерпретировать цинично, но на основе имеющихся данных её трудно опровергнуть. Компания **Gartner** предупреждала, что более 40% проектов в области агентного ИИ будут отменены к концу 2027 года, связывая неудачи с вопросами управления (governance), неопределенной бизнес-ценностью и слабой операционной дисциплиной, а не с возможностями самих моделей.

Почти всё, что входит в пакет **Presence**, направлено именно на решение этих проблем. Симуляции и системы оценки (graders) проверяют, достиг ли агент правильного результата, следовал ли он политике, правильно ли использовал свои инструменты и вовремя ли инициировал эскалацию, прежде чем с ним вступит в контакт кто-либо вне компании. «Ограждения» (guardrails) вмешиваются, когда взаимодействие выходит за установленные границы. Записи сессий и истории действий дают проверяющим базу для аудита. Пути эскалации передают человеку структурированный контекст, а не «холодную» расшифровку диалога, а новые версии внедряются через контролируемое развертывание с возможностью отката (rollback).

Корпорации потратили два года на осознание того, что самая сложная часть внедрения промышленного агента заключается в интеграции, правах доступа и управлении изменениями. Вендор, который направляет инженеров для выполнения этой работы, реагирует на то, в чем покупатели фактически терпят неудачу, вместо того чтобы просто выпустить очередной дашборд и назвать возникший пробел проблемой клиента.

Где кроется ограничение

Компромисс проявляется в критериях соответствия. Доступ, по словам OpenAI, зависит от соответствия рабочему процессу, готовности к внедрению и доступных мощностей по доставке (delivery capacity).

Мощности по доставке — это ограничение консалтингового характера. Программное обеспечение масштабируется; инженеры, допущенные к основным системам банка, — нет. **Forward Deployed Engineer** — это название должности, заимствованное у **Palantir**, где оно описывает сотрудников, внедренных в операции клиента на протяжении нескольких месяцев; экономика этого процесса совершенно не похожа на экономику оплаты по мере использования (metered inference). Выставляя своих собственных FDE и именных партнеров в авангард каждого развертывания, OpenAI входит в тот сегмент рынка, который занимают интеграторы, на которых она также будет полагаться для масштабирования. Это рабочая схема при малых объемах, но более сложная в будущем.

Это также ставит вопрос перед каждым, кто составляет проект контракта. Когда поставщик модели одновременно является партнером по внедрению, границы ответственности за неверно примененную в промышленной эксплуатации политику должны быть четко прописаны, а не подразумеваться по умолчанию.

Корпоративные ИИ-агенты, представленные на рынке, всё еще находятся на ранней стадии

OpenAI описывает **Presence** как продукт, прошедший проверку в реальных условиях (battle-tested), аргументируя это тем, что продукт собирался на основе многолетнего опыта развертывания агентов у корпоративных клиентов еще до того, как он был упакован и получил название. Это утверждение касается накопленной практики, а не времени присутствия продукта на рынке, и оно вполне обоснованно.

Самым весомым доказательством является собственная англоязычная линия телефонной поддержки OpenAI: 1-888-GPT-0090. Компания заявляет, что агент достиг или превзошел её внутренние показатели для первой линии человеческой поддержки в течение нескольких недель, теперь решает 75% входящих запросов без помощи человека и сократил количество передач дел человеку на 15 процентных пунктов за десять дней благодаря циклу

улучшений **Codex**. Это цифры OpenAI, измеренные по собственным критериям оценки OpenAI на собственном канале. Такая прозрачность приветствуется, но цифры не прошли независимую проверку.

Три названных клиента находятся на более ранних стадиях цикла, чем предполагает описание запуска. **BBVA** изучает голосовую поддержку для повседневных банковских операций в Мексике. **SoftBank** тестирует диалоги на японском языке. **IAG** изучает возможность поддержки во время периодов высокого спроса, таких как экстремальные погодные условия. Даниэль Ордас, руководитель отдела трансформации ИИ в **BBVA Mexico**, описывает банк как партнера по проектированию (design partner), помогающего формировать и совершенствовать голосовой клиентский сервис. Партнеры по проектированию — это нормальное явление на этапе ограниченной общей доступности. Однако ни один из трех не представлен как компания, использующая **Presence** в масштабе, что стоит учитывать наряду со словом «проверенный».

Что не было раскрыто

Цены не публикуются. Объем внедрения и стоимость устанавливаются индивидуально для каждого клиента и каждого развертывания, что является обычной практикой для корпоративных услуг, но всё же оставляет покупателей без публичной точки отсчета для сравнения стоимости одного решенного контакта с существующими поставщиками контакт-центров.

Название модели не разглашается. В документации указано, что **Presence** использует модели OpenAI с конфигурацией, выбранной под конкретный рабочий процесс, которая может меняться по мере развития этого процесса. Такая гибкость является оправданным инженерным решением, поскольку привязка промышленного агента к застывшей версии модели со временем приводит к деградации системы. Тем не менее, команды, которые провели последний год, создавая наборы инструментов для оценки (evaluation suites) под конкретные версии, захотят, чтобы в контракте было указано, на что именно они ориентируются при изменении конфигурации.

Поддержка каналов во время ограниченной общей доступности охватывает голос или чат, при этом интеграция с контакт-центром, маршрутизация, аутентификация и проектирование передачи управления подтверждаются при каждом развертывании. Обработка данных следует тому же принципу: подписанная архитектура и контракт рассматриваются как руководящие документы, а не какая-либо опубликованная политика.

Presence стоит отдельно от **ChatGPT Workspace Agents**, которые остаются путем самообслуживания для команд, работающих внутри ChatGPT и Slack, в то время как голосовые клиенты сохраняют доступ через API к передовым моделям (frontier models) OpenAI. Теперь компания предлагает фактически одну и ту же возможность тремя способами, различающимися не столько возможностями технологии, сколько тем, кто выполняет работу.

В итоге покупатели выбирают не столько по возможностям модели, сколько по мощностям по доставке, а, по утверждению самой OpenAI, именно мощности по доставке являются тем ресурсом, который распределяется ограниченно.

Перевод выполнен: 24.07.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.