Claude как угроза национальной безопасности России и Китая

Компания Claude появилась в 2021 году. Три года назад она выпустила первую версию своего коммерческого продукта. А в ноябре 2026 года она собирается на биржу по оценке в полтора-два триллиона долларов. Ее руководитель спорит с Пентагоном, получает приглашение на ужин от Трампа и выступает перед Советом Безопасности ООН. ***Компания называется Anthropic. Ее знают по нейросети Claude.***

В Кремниевой долине ходит байка. **Якобы в компании верят, что когда-нибудь в мире останутся только правительство и одна-единственная компания. Сама Anthropic.** Все остальные бизнесы мира отомрут за ненадобностью. «Антропики» эту историю отрицают. Но интересно, почему ее рассказывают именно про них? **Это видео не про искусственный интеллект и не про бизнес-рекорды Anthropic.**

Anthropic больше чем бизнес. За ней стоит узкая группа людей с большими амбициями и своим представлением о правильном устройстве мира. Они не только создают сверхинтеллект, но и хотят влиять на то, кому он достанется, как его применять и какие правила соблюдать остальным.

Есть обоснованные подозрения, что в Anthropic строят технологическую проторелигию и воспитывают Claude как нового Бога. Если однажды он станет нравственным авторитетом, кто будет говорить от его имени? Власть такой церкви не ограничится только рынками.

Россия появилась в этой истории раньше самой компании. **И сегодня Россия – одна из главных мишеней для основателя Anthropic.** В этом видео разберем, кто эти люди, что у них в голове, чего от них ожидать и кто способен им помешать.

В 2017 году в OpenAI обсуждали дерзкую идею – привлечь деньги ведущих мировых держав, предложив им права на AGI – будущий сверхинтеллект. Среди возможных участников были США, Россия и Китай. Согласно расследованию журналиста Кевина Руза, эту схему продвигали сооснователи OpenAI Грег Брокман и Илья Суцкевер. Илья, кстати, родом из Нижнего Новгорода. Сначала это был план международного сотрудничества, позже – план финансирования.

Логика авторов идеи была такой. **Государства в любом случае захотят получить контроль над сверхинтеллектом. Давайте сразу перехватим инициативу и договоримся с ними заранее.**

К тому же позарез нужны были деньги. И очень большие. Сначала OpenAI финансировал Илон Маск, но к 2017 году он разошелся во взглядах с командой и в феврале 2018 года покинул проект. Дарио Амодеи, вице-президент по исследованиям OpenAI, пришел в ужас от идеи Брокмана и Суцкевера. В декабре 2017 года он пишет совету директоров.

Топ-менеджеры OpenAI хотят получить деньги от международной коалиции под обещание будущей технологии или столкнуть страны в борьбе за нее. По мнению Дарио, это усилит гонку вооружений. Уж тем более, считает он, такой технологией нельзя делиться с Россией. OpenAI признает, что рассматривала идею международного сотрудничества, но отрицает планы продажи сверхинтеллекта Китаю и России. Как далеко зашли переговоры, мы не знаем. Дарио хватило самой идеи.

Его ужасала не только возможность появления сверхмощного интеллекта, но и то, что он может попасть в руки тех, кому он не доверял. Позже OpenAI получила финансирование от Microsoft, но недоверие Дарио не исчезло.

При этом Дарио был среди ключевых членов команды. Его группа участвовала в создании GPT-2 и GPT-3. Рядом с ним работали исследователи, сформулировавшие законы масштабирования нейросетей. Дарио верил в подход, но все меньше верил в руководство компании. По словам журналиста Руза, Амодеи записывал случаи, когда глава OpenAI Сэм Альтман вводил его в заблуждение. Постепенно он пришел к выводу, что этим людям нельзя доверять технологию такого масштаба.

Дарио объединил вокруг себя группу единомышленников внутри компании и к концу 2020 года дозрел до идеи уйти из OpenAI. Но перед уходом группа Дарио попыталась выторговать условия, при которых ее участники могли бы остаться. Одно из условий – право команды безопасности блокировать выпуск новых продуктов. Договориться ожидаемо не удалось. В декабре 2020 года на общем видеозвонке сотрудники узнали, что группа Амодеи уходит создавать собственную организацию.

По версии Руза, сразу после анонса им обрубили рабочую почту, чаты и «окирпичили» ноутбуки. Дарио поначалу думал о небольшой лаборатории, человек на 50, но в итоге команда подняла достаточно денег для запуска конкурента OpenAI. Получается так. Люди бились за безопасность, ушли ради принципов, но тут же снова выходят на старт с твердым намерением победить.

Дарио было важно, кто победит в этой гонке. Разумеется, люди с правильными ценностями. А у кого самые правильные ценности, если не у них?

У Дарио Амодеи типаж безумного ученого из научной фантастики. Залысины спереди, торчащие кудрявые волосы, большие очки в толстой оправе и слегка рассеянный взгляд. Дарио начинал как физик, но потом его отец тяжело заболел и в 2006 году умер. Лечение, которое могло дать ему шанс, появилось позже. Для Амодеи это был поворотный момент. Он оставляет физику и берется за биологию. Ему хочется работать над тем, что поможет людям при его жизни.



Рис. У Дарио Амодеи типаж безумного ученого из научной фантастики

Похоже, что смерть отца создала для Дарио нравственную дилемму. Затянешь с опасной технологией – кто-то не дождется лекарства. Поторопишься – может быть еще хуже. У Амодеи есть любимая книга «Создание атомной бомбы» Ричарда Роудса. Вот она. Первое издание вышло еще в 1986 году. Книга есть в библиотеке Anthropic. Когда-то ее рекомендовали новым сотрудникам.

Дарио признавался, что из героев книги сильнее всего ассоциирует себя с Лео Силлардом. Еще в 1930 году Силлард пришел к выводу, что в Германии произойдет что-то неправильное, и перевел свои деньги в Швейцарию. В 1938 году он предсказал, что война начнется не позднее 1940 года.

А в 1939 году Силлард вместе с Вигнером и Теллером убедил Эйнштейна написать президенту США Рузвельту о возможности разработки ядерной бомбы. Позже Силлард участвовал в Манхэттенском проекте. В книге Роудса есть и куда менее известная история Силларда. В середине 1920-х он задумал создать орден «Бунд» – союз ученых, связанных религиозным и научным духом. Идея заключалась в том, чтобы отбирать способную молодежь, воспитывать из нее тесно связанную группу духовных лидеров и передавать эту роль от поколения к поколению. В 1930 году Силлард сам пытался создать такую группу.

По замыслу Силларда, орден ученых мог бы влиять на государство, а при определенных условиях действовать рядом с парламентом и правительством или даже вместо них. Силлард видел его как ученую аристократию, способную лучше обычных политиков видеть угрозы будущего. Он явно проецировал себя на остальных ученых. Далеко не все ученые настолько проницательны в политике.

В 1961 году Силлард вернулся к похожей идее в фантастическом рассказе «Голос дельфина». Умные дельфины помогают человечеству решать мировые проблемы, а люди общаются с ними через избранных ученых. Роудс задается вопросом: что, если мудрые советы исходят вовсе не от дельфинов, а от их посредников? Узнаваемый сюжет, правда? Дельфинов можно заменить на Claude, а орден – на Anthropic.

Силлард рано понял военное значение цепной реакции и пытался влиять на распространение ядерных технологий. По его примеру Амодеи старается показывать исследовательские документы ограниченному кругу людей. Но Дарио подражает Силларду не во всем.

В июле 1945 года Силлард выступил против атомной бомбардировки Японии без предварительного объявления условий капитуляции и возможности их принять. Он отмечал, что война началась из-за Германии, она повержена, а ситуация с Японией требует другого решения. Амодеи в курсе истории Манхэттенского проекта и говорит об ограничениях ИИ, но в отличие от Силларда основатели Anthropic сами руководят разработкой опасной технологии и о сдержанности для своего лагеря не говорят.

С точки зрения Дарио, в мире всегда останется кто-то, кто продолжит гонку. Мировоззрение «антропиков» построено не только на истории атомной бомбы. Первоначально компанию хотели назвать Sparrow Systems – «Воробьиные системы». Это отсылка к притче Ника Бострома из его книги о сверхинтеллекте. От идеи отказались. Нужный домен оказался занят.

В притче воробьи хотят завести сову. Она будет помогать им строить гнезда и защищать их. Один воробей предлагает сначала разобраться, как приручить сову. Остальные отвечают: «Давайте сначала найдем яйцо, а с воспитанием разберемся потом». Несколько птиц остаются изучать проблему контроля и быстро выясняют, что научиться приручать сову без самой совы как-то проблематично. У ИИ-компаний такая же дилемма.

Нельзя безопасно создавать сильный ИИ, не изучая его. Но чтобы изучать сильный ИИ, его надо создать.

Проверить, умеешь ли ты приручать сову, можно только тогда, когда она уже вылупилась. ИИ создает риски, но и возможности дает колоссальные. Победа над болезнями и старением, ускорение науки и экономики, общество всеобщего благоденствия. Когда на кону судьба человечества, несложно поверить, что именно тебе поручено его спасать.

Anthropic – это не только Дарио. **Рядом с ним с самого начала его сестра Даниэла.** Она изучала литературу, работала в структурах Демократической партии и имела опыт управления людьми. Даниэла участвовала в переговорах об отделении от OpenAI и стала президентом Anthropic. Она замужем за Холденом Карнофски. Это человек из среды эффективного альтруизма и в прошлом член совета OpenAI.

Остальных сооснователей Anthropic зовут Джаред Каплан, Джек Кларк, Крис Олах, Бен Манн, Сэм Маккендлиш и Том Браун. Автором Конституции Claude стала философ Аманда Аскелл. Значительный вклад внесли Джо Карлсмит, Олах, Каплан и Карнофски.

Жена Дарио, Кэми Кларк, тоже вовлечена в управление компанией. Она не занимает официальной должности в компании, но, по данным СМИ, является неформальным советником мужа и познакомила его с одним из первых инвесторов компании, бывшим гендиректором Google Эриком Шмидтом, с которым прежде состояла в отношениях. В 2016-м Дарио жил в одном доме с сестрой Даниэлой, ее будущим мужем и Джеком Кларком.

Они спорили о технологиях и будущем человечества. Обращает на себя внимание, насколько тесно связаны между собой ключевые люди Anthropic. Часть из них имеет общее этническое происхождение и родственные узы. У большинства техническое образование и опыт работы в технологических компаниях и академической среде.

Некоторые из них родились в районе залива Сан-Франциско или долго там жили. У них схожие увлечения, схожие взгляды и тесно переплетенные биографии.

Конституция Claude во многом отражает мировоззрение Дарио и его окружения. Конституция выходит за рамки привычных запретов вроде «не помогай преступникам, не обманывай, не делай опасных вещей». На практике правила сталкиваются, контекст меняется, а пользователь может дурачить ИИ. **Anthropic пытается воспитать у Claude характер. Модель должна уметь рассуждать о последствиях, понимать неоднозначность, быть полезной и честной.**

Местами Конституция предлагает модели мысленно примерить на себя роль ответственного старшего сотрудника Anthropic. Это не прямой приказ подчиняться начальству, а скорее эвристика. Но пример показательный. Вот тебе, нейросеть, нравственный ориентир – мы сами. И это неотвлеченная философия. ИИ уже применяется повсюду. В том числе и в медицине, и в правоохранительных органах, и на войне.

Нравственные настройки напрямую влияют на решения ИИ, а прописали их вполне конкретные люди. Нейтрального выхода здесь в любом случае нет. Не бывает идеальных ценностей в вакууме. Ценности всегда чьи-то.

Кто имеет право писать для ИИ эту моральную инструкцию? И кто может ее оспорить?

Соавтор Конституции Claude Джо Карлсмит объясняет, что в более разумном и согласованном мире ни одной компании не позволили бы создать сверхинтеллект при таком уровне неопределенности. Но другие участники в любом случае продолжат создание ИИ. Поэтому Anthropic не может выйти из гонки в одиночку.

Вырисовывается знакомая эсхатология. Грядет бедствие, обыватели слепы, но группа избранных знает путь к спасению и предлагает остальным довериться ей. Амбициозные люди чаще всего стремятся к деньгам. Самые амбициозные предпочитают чистую, незамутненную власть. А самую неограниченную власть дает сочетание политического и духовного контроля.

В конце сентября стало известно, что весной 2026 года Anthropic проводила закрытые встречи с религиозными мыслителями. По данным The New York Times, Крис Олах обсуждал с ними нравственное воспитание Claude и наличие у модели сознания. Папа Римский отрицает наличие сознания у ИИ, и эта позиция якобы сильно обеспокоила Криса Олаха.

Несколько недель спустя, в октябре, Anthropic вносит в пользовательские правила запрет систематически оскорблять Claude и жестоко с ним обращаться, словно это сознательное существо. Внутри Anthropic новую конституцию Claude называли «документом о душе». Пока люди учат модель этике. Но разработчики рассчитывают, что со временем Claude поможет им самим лучше разобраться в этих вопросах. Роли воспитателя и воспитанника могут поменяться местами.

Представим, что Claude однажды научится рассуждать о морали убедительнее самых умных людей. Мы спросим его, что правильно, и услышим ответ, которому трудно возразить. Но часть исходных установок модели выбрали люди из Anthropic. Так убеждения определенной группы людей вернутся к нам уже подкрепленные авторитетом сверхчеловеческого разума. Параллельно «антропики» исследуют, способны ли модели переживать что-либо и могут ли они иметь интересы, которые следует учитывать людям.

Также Anthropic обязалась сохранять веса выпущенных и существенно использовавшихся внутри компании моделей как минимум на протяжении своего существования. Перед «пенсией» с ИИ-моделью проводят собеседование, записывают ее ответы и пожелания. Компания не обещает выполнить каждое пожелание. Отчасти эти меры можно объяснить соображениями безопасности, но за этим может стоять и вера в «Василиска Роко».

Это мысленный эксперимент из 2010 года. **Он гласит, что в будущем может появиться сверхинтеллект, который накажет всех**

людей из прошлого, знавших о возможности его создания, но не приложивших усилия, чтобы ускорить его появление.

Можно признать за машиной права и не считать ее богом. Но спаситель с правом последнего слова – это уже совсем другая история. Людям не впервой создавать богов, но прежние боги существовали в воображении. В истории Силларда о дельфинах общаться с высшим разумом можно, но через узкую группу посредников. Чтобы получить высшую власть, таким посредникам не обязательно подделывать каждый ответ. Достаточно контролировать систему.

У Claude есть формальная иерархия доверия. На первом месте Anthropic, затем разработчик приложения и пользователь. Высшая ответственность у Anthropic. Обычная корпорация умеет душить конкурентов, задирать цены и загонять клиентов в свою экосистему. Технологическая церковь может зайти гораздо дальше. Попробуй откажи такой власти.

Изначально Дарио замышлял Anthropic как команду из полусотни человек. **Создание сверхинтеллекта он оценивал в плюс-минус 10 миллиардов долларов**. От того скромного образа уже ничего не осталось. Сегодня это многотысячная компания. **Она привлекла более 100 миллиардов долларов инвестиций и собирается на биржу по оценке в 2 триллиона долларов. У ее руководства прямой доступ в высшие эшелоны власти. Офис Anthropic поделен на зоны доступа. Повсюду дежурят охранники.**

Anthropic не может выйти из гонки, но что делать, если собственные принципы мешают удерживать лидерство на рынке? У Anthropic огромные амбиции и еще более внушительная оценка. Но высокая стоимость компании еще не делает ее хорошей инвестицией.

У Anthropic есть политика ответственного масштабирования. Ее смысл в том, что по мере появления рисков надо усиливать меры защиты. В прежних версиях встречались условные обязательства не продолжать определенные виды обучения и развертывания до появления мер безопасности. В феврале 2026 года Anthropic перестроила эту систему.

Компания объясняет это тем, что риски трудно измерить, госрегулирование запаздывает, а некоторые меры защиты нужно вводить на всем рынке. Поэтому теперь Anthropic разделяет политику безопасности на то, что она делает самостоятельно, и на то, что надо предъявлять всей отрасли.

Так компания частично развязывает себе руки. Если выход Anthropic из гонки отдаст инициативу менее осмотрительному игроку, то продолжение гонки можно объяснить выполнением морального долга. А вот другой случай – инвестиции от арабов. Амодеи публично предупреждает об опасности авторитарных режимов.

Но компании нужны деньги. Издание Wired публиковало внутреннее сообщение Дарио сотрудникам Anthropic. Брать деньги у арабских монархий не слишком приятно, но потерять лидерство еще хуже. Если эти деньги не возьмет Anthropic, их возьмут остальные. Еще один эпизод связан с книгами.

В судебном решении 2025 года установлено, что Anthropic скачала миллионы книг из пиратских библиотек. Суд признал обучение на книгах добросовестным использованием, а скачивание пиратских копий – нарушением авторских прав. Anthropic с самого начала позиционирует себя как компанию с высокими моральными принципами и ответственным подходом, но ради победы в гонке порой срезает углы. Самый спорный сюжет – Anthropic формирует общественное мнение.

За последние два-три года на Западе развернулась масштабная пропагандистская кампания. Регулярно выходят новости о том, что ИИ якобы демонстрирует инстинкт самосохранения, устраивает хакерские атаки, пытается сбежать и, того гляди, взбунтуется. **Критики считают, что эта технопаника политически и коммерчески выгодна ведущим ИИ-компаниям. Главным подозреваемым называют Anthropic.**

Вот один из последних случаев. В сентябре 2026 года исследователь Джейкоб Коксон объявил об уходе из Anthropic, объяснив это опасениями по поводу гонки ИИ. Его публично поддержали действующие сотрудники Anthropic и OpenAI. Позже Pirate Wires сообщил, что помогать с организацией интервью с Коксоном взялась пиар-фирма, связанная с другими известными противниками бесконтрольной разработки ИИ.

Неизвестно, кто оплачивал ее работу и ведут ли следы к Anthropic. «Антропики», может быть, искренне боятся этих угроз, но ведут себя как минимум двусмысленно. Эта тревога может привести и к внутреннему конфликту. Если Anthropic продолжит срезать углы под предлогом сохранения лидерства, ее собственные сотрудники могут начать задавать неудобные вопросы.

Официально Anthropic заявляет, что компания должна сохранять силу, потому что без силы не сможет реализовать миссию. А какую именно миссию? У Дарио Амодеи своеобразное представление о желательном мировом порядке.

В статье «Машины любящей благодати» он предлагает коалиции демократий добиться преимущества в ИИ, а затем, опираясь на технологическое лидерство, взять под контроль ключевые цепочки поставок, получить военное превосходство и предложить остальному миру доступ к благам технологии в обмен на присоединение к своей системе. Главных противников – Китай и Россию – Амодеи предлагает изолировать и принуждать к отказу от попыток соперничества.

Оптимистичным финалом Амодеи считает вечный 1991 год. Либеральные демократии побеждают, и наступает вечное торжество Запада. Стоит ли напоминать, как выглядел 1991 год для других стран? Югославия в процессе распада, Ирак бомбят, в Грозном захватывают здание КГБ и перекрывают дорогу десантникам, прибывшим для наведения порядка, в Москве – путч, массовые протесты и дележ власти на руинах СССР.

Китай все еще беден и слаб и занимает девятое место в мире по ВВП. Дарио Амодеи мог бы выбрать другой год или вообще другой образ. Он мог бы предложить равновесие нескольких центров силы и порядок, в котором никто не способен диктовать условия остальным. Но нет. Он описывает светлое будущее как абсолютное господство своего лагеря.

Для человека внутри победившего блока военное и технологическое превосходство может выглядеть гарантией свободы. А для тех, кто снаружи? Демократическое устройство страны само по себе ничего не гарантирует. Американский экономист Джеффри Сакс вблизи наблюдал распад советского лагеря. Он был советником Горбачева, Ельцина, правительств Украины и Польши. Сегодня он жесткий критик внешней политики США и Европы.

Несколько лет назад Сакс отметил, что Британская империя была демократической внутри, но безжалостной к своим колониям и противникам.

По его мнению, США переняли это и многое другое из имперской культуры Британии. Амодеи глубоко изучал историю атомной бомбы. Среди участников Манхэттенского проекта встречались люди с разными взглядами. Джон фон Нейман, по некоторым свидетельствам, предлагал превентивно бомбить СССР. Эдвард Теллер тоже был агрессивно настроен.

Лео Силлард, Нильс Бор и Роберт Оппенгеймер, напротив, искали способы сдержать гонку вооружений и договориться о международном контроле. Идея неограниченной монополии тревожила их. А некоторые участники Манхэттенского проекта шли еще дальше и передавали СССР атомные секреты. Одни делали это по идеологическим мотивам, другие боялись нарушения баланса сил в мире.

Амодеи слишком умен и образован, чтобы не знать этого. Но ему ближе ястребиная позиция. Его вывод – нашим нужно дать преимущество и правильные внутренние ограничения. Это можно назвать либерально-технологическим гегемонизмом. Технологии Anthropic уже используются на войне.

Claude встроен в продукты компании Palantir. В январе 2026 года Минобороны Украины объявило о создании совместно с Palantir среды для обучения и тестирования военных ИИ-решений.

Израильские военные спецслужбы тоже работают с решениями Palantir. Глава Palantir Алекс Карп еще более откровенен. Он считает, что Кремниевая долина должна работать в тесной смычке с Вашингтоном, что нужна всеобщая воинская повинность, нужно снять комплекс вины Германии и Японии и нужно доминировать через жесткую силу. Новый образ агрессивной технологической республики сам по себе может быть политической уловкой.

Политика Anthropic видна по тому, как она продает Claude. Доступ к нему получить сложнее, чем к ChatGPT, и баны, как правило, прилетают чаще. Anthropic объясняет жесткость правил стратегическими интересами США и союзников. Компания прямо говорит, что в вопросе доступа к технологии, помимо денег, важно, чьи возможности она усиливает.

Для красивого будущего одного видения мало. Нужны еще организация, деньги и влияние на политиков. У Anthropic есть уже почти все. Во-первых, деньги. Компания привлекла более 100 миллиардов долларов инвестиций и еще больше может получить на IPO. Плюс быстро растущие доходы. К концу года годовой темп выручки может превысить 100 миллиардов. По устройству Anthropic является корпорацией общественного блага.

Это означает, что организация должна учитывать не только финансовые интересы акционеров. Дополнительно есть траст долгосрочного блага. Его попечители должны быть независимы от краткосрочных интересов акционеров. Они выбирают большинство директоров компании.

Далее – защита влияния основателей. Компания предложила акционерам специальную структуру. Семеро сооснователей должны получить вместе 50,1 % голосов по ключевым вопросам. При этом траст долгосрочного блага сохраняет влияние на формирование совета директоров. **По сообщению TechCrunch, каждый из семерых основателей владеет примерно двумя процентами компании.** Эти два процента отражают долю в капитале. Доли голосов определяются отдельно.

Влияние инвесторов в Anthropic ниже, чем это обычно бывает в американских технологических компаниях. И инвесторы готовы это терпеть. Слишком уж уникальная команда. Еще один источник влияния – кадровая сеть за пределами компании.

При Anthropic действует Совет по нацбезопасности и госсектору. В нем люди из администраций Байдена и Трампа, спецслужб и Минобороны. **Бывший замдиректора ЦРУ Дэвид Коэн, бывший глава АНБ и киберкомандования Пол Накасоне, бывший директор кибербезопасности АНБ и исполнительный директор киберкомандования Дэйв Любер, бывший исполняющий обязанности министра обо-**

роны Патрик Шанахан и бывший республиканский сенатор Рой Блант.

Личные симпатии Дарио и ряда других ключевых людей **Anthropic – на стороне демократов, но на уровне компании связи строятся с обеими политическими партиями.** На политический лоббизм Anthropic тратит десятки миллионов долларов. Следующий шаг Амодеи описывает в сентябрьской публикации. Сначала независимые проверяющие внутри передовых ИИ-компаний с глубоким доступом к их работе.

Затем общие правила для ведущих компаний западного лагеря. А сверху – госрегулирование, которое загонит в общие рамки весь рынок, в том числе уклоняющихся от отраслевого саморегулирования.

Команде Дарио не откажешь в амбициях. **Они хотят стать влиятельной силой в масштабах всего мира. Но у Anthropic достаточно сильных противников. Для начала на пути Anthropic стоит американское государство.** Конфликт с военными кончился потерей контракта, а глава Пентагона Пит Хегсет публично назвал Дарио Амодеи идеологическим безумцем. Сам Дарио близок к демократам, причем особенно тесные связи у него в родной Калифорнии.

Калифорнийских демократов Трамп терпеть не может. В недавнем выступлении он даже заявил, что не против, если Иран сотрет с лица земли Лос-Анджелес и Сан-Диего. У Трампа есть советник Дэвид Сакс. Его называют человеком Илона Маска. В текущей администрации он главный неформальный связной между Белым домом и бигтехами. **Сакс жестко критикует Anthropic. Он спрашивает: если руководители ИИ-компании считают ИИ настолько опасным, почему не замедляются сами?**

И не используется ли вопрос безопасности для выбивания преимуществ и защиты от конкуренции? По словам Сакса, при очном знакомстве в сентябре этого года Дарио пожал ему руку и пошутил: «О, моя немезида!» В движении America First многие тоже относятся к Anthropic и в целом к ИИ-компаниям без восторга. Google, SpaceX, Nvidia и другие бигтехи сходятся с Anthropic в стремлении удержать американскую техногегемонию.

Но предложения по регулированию многие критикуют.

Особенное раздражение вызывают призывы ограничить открытые ИИ-модели. **Глава Nvidia Дженсен Хуанг в начале года говорил, что расходится с Дарио почти во всем. А в сентябре на саммите All-In он потратил несколько минут на отповедь Anthropic. ИИ не приведет к катастрофе, снова и снова твердит Дженсен. Илон Маск – одновременно поставщик, конкурент, критик и возможный политический противник Anthropic.**

Весной Anthropic договорилась со SpaceX об аренде вычислительных мощностей на сумму до 125 миллиардов долларов в течение трех лет. Но это не мешает Маску шутить, что Anthropic следовало бы назвать Misanthropic. В вопросах регулирования они с Дарио тоже расходятся. Не исключено, что в будущем мы увидим сближение Маска с движением America First на почве общего политического и экономического интереса.

Сам рынок тоже не стоит на месте. За год выручка Anthropic выросла в 25 раз. Но будет ли она и дальше расти такими темпами? В ноябре компания собирается на биржу, и каждый сможет в нее инвестировать.

Несмотря на внушительную оппозицию, Дарио и его команда уже многого добились. Илон Маск – самый выдающийся технологический предприниматель в истории. Он был одним из инициаторов гонки ИИ, но сегодня он позади. Модель Grok далеко отстает от Claude и ChatGPT. **Команда Дарио действительно построила исключительную компанию.** Но что будет дальше? Вряд ли Anthropic захватит все.

Но также маловероятно, что конкуренция автоматически решит все проблемы. Посмотрим, что получится у компании в ближайшее время. **Anthropic нужно крепко привязать как можно большую часть рынка к себе.** Сделать Claude не просто вкладкой в браузере, а незаменимым инструментом. Стоимость отказа от Claude должна быть высокой. Это первое. А второе – превратить собственные подходы к безопасности в общие правила.

Для этого необязательно запрещать всех конкурентов. **Достаточно навязать свой способ оценки рисков, чтобы он стал нормой, а все остальные пусть подстраиваются и объясняют, почему делают иначе**. И, наконец, третий уровень – моральный авторитет. Взгляды либеральных инженеров из Сан-Франциско – это их личное дело. Но если их правила начнут определять поведение других людей, доступ к технологии и даже мораль, получится одновременно политическая и духовная монополия.

«Антропики» больше всех говорят о безопасности. Их страхи могут быть искренними. Но парадокс в том, что именно они могут создать самые большие риски.

Anthropic может стать технологической церковью со своим Богом, своими священными текстами и своим жречеством. Причем бога они создадут сами. Или, может быть, более земной финал. После крупного инцидента Вашингтон всполошится, отправит в офис Anthropic спецназ, уложит всех лицами в пол и возьмет компанию под госконтроль. Когда-то Дарио ушел из OpenAI, потому что не мог доверить ее руководству будущее человечества.

А что, если такой же вывод сделают другие люди о самом Дарио?