

AI 2027: будущее наступает

Что будет, если развитие искусственного интеллекта продолжит ускоряться? Как изменится мир, экономика и геополитика — и насколько мы к этому готовы?

Группа аналитиков — Daniel Kokotajlo (инженер, ex-OpenAI), Scott Alexander (известный блогер), Thomas Larsen (исследователь безопасности ИИ), Eli Lifland (исследователь ИИ), Romeo Dean (исследователь ИИ, студент Гарварда) — выпустила один из самых амбициозных прогнозов в мире ИИ — отчёт [AI 2027](#). Он описывает детальную хронологию возможных событий с 2025 по 2030 год: от появления первых полноценных ИИ-агентов до трансформации всей глобальной экономики, усиления государств, в основе которых действует ИИ, и даже гипотетического сценария потери человечеством контроля.

Отчёт *AI 2027* — это не пророчество, а приглашение к разговору: о будущем ИИ, о нашем месте в нём и о том, насколько мы готовы к миру, где интеллект больше не принадлежит только человеку.

Команда стратегических аналитиков **red_mad_robot** перевела и адаптировала отчёт, сохранив все ключевые моменты и акценты, чтобы вам не пришлось разбираться в десятках англоязычных источников.

В этом материале:

- Что может произойти уже в 2025 году — и какие риски это несёт;
- Как ИИ влияет на рынок труда, кибербезопасность, политику и науку;
- Какие сценарии развития событий более вероятны: гонка или замедление;
- Почему выравнивание целей ИИ — это не техническая задача, а философская дилемма.

Чтобы было проще ориентироваться в терминах, мы подготовили небольшой глоссарий:

Выравнивание (alignment)

Соответствие целей, намерений и поведения ИИ человеческим ценностям, интересам и этическим принципам. Основная задача — сделать так, чтобы ИИ действовал не просто эффективно, а в интересах людей.

Нарушение выравнивания (misalignment)

Достижение ИИ заданных целей путём, который противоречит человеческим намерениям. Это может быть как непреднамеренное поведение, так и стратегический обман.

Веса (weights)

Числовые параметры внутри нейросети, которые определяют, как модель принимает решения. Во время обучения ИИ «подстраивает» свои веса — усиливает или ослабляет связи между нейронами — чтобы всё точнее решать задачи.

Спецификация (specification)

Документ или набор правил, который описывает, что ИИ «можно» и «нельзя» делать, какие цели ставятся перед системой, и как она должна принимать решения. В отчёте это ключевой инструмент для попытки выравнивания.

Агент (Agent)

Модель ИИ, способная самостоятельно принимать решения, обучаться и выполнять задачи в сложной среде. В отчёте фигурируют поколения агентов (Agent-1, Agent-2 и т.д.), каждый из которых мощнее и автономнее предыдущего.

Safer-модели

Линия моделей, разработанная с фокусом на безопасность и прозрачность, даже ценой снижения производительности. Их задача — быть «меньшим злом» в условиях невозможности полной уверенности в безопасности ИИ.

Consensus-1

Совместный ИИ, созданный в сценарии «Гонка» (один из двух путей развития ИИ согласно прогнозу) для поддержания условий мирного договора между ИИ США и Китая. Его появление — потенциально поворотный момент в истории автономных ИИ-систем.

Теперь поехали!

2025:

Stumbling Agents

В мире появляются первые ИИ-агенты, предлагающие функции «персонального помощника». Они способны выполнять задачи и общаться с пользователями, но пока ещё не получили широкого распространения — несмотря на их теоретические преимущества и адаптацию в разных сферах.

The World's Most Expensive AI

OpenBrain (вымышленная компания, занимающаяся НИОКР в области ИИ и расположенная в США) стремится автоматизировать собственные исследования, чтобы обогнать ключевого конкурента в Китае и создать более мощные ИИ-системы.

Краеугольным камнем становится вопрос управления сверхразумом. Компания разрабатывает **Спецификацию** — документ, в котором перечислены запрещённые и разрешённые действия, цели и принципы. Затем они пытаются обучить ИИ следовать этой спецификации. Проблема в том, что компания не может проверить, сработало это или нет. Всё, что они могут сказать:

«Похоже, что пока модель соблюдает спецификацию — но мы не уверены».

Авторы подчёркивают: вопрос о том, какие цели может преследовать сверхразум, остаётся открытым. Это одна из многочисленных неопределённостей, требующих дальнейшего изучения. Подробнее — [по ссылке](#).

Сомнения команды alignment

Команда по выравниванию (alignment) в OpenBrain — отвечает за соответствие действий агента спецификации — сомневается в честности модели. Они рассматривают возможность ИИ лгать самому себе или имитировать честность, чтобы получать «награду». Это усиливает тревогу по поводу искренности и честности модели.

2026:

Coding Automation

Ставка на использование ИИ для ускорения исследований в области искусственного интеллекта начинает оправдываться: OpenBrain достигает алгоритмического прогресса на 50% быстрее при помощи ИИ-ассистентов. Запущена новая, более

надёжная модель **Agent-1**, которая, среди прочего, умеет отслеживать риски утечки алгоритмических секретов конкурентам.

China Wakes Up

В Китае коммунистическая партия начинает осознавать стратегическое значение AGI и национализирует исследования в этой области. Принимается решение создать централизованную зону развития, где лучшие исследователи объединяются для обмена идеями и ресурсами. Это должно компенсировать нехватку вычислительных мощностей, вызванную экспортными ограничениями.

В числе целей – **кибершпионаж** с целью кражи технологий OpenBrain.

AI Takes Some Jobs

OpenBrain вырывается вперёд, выпуская **Agent-1-mini** – лёгкую в настройке модель в десять раз дешевле оригинального Agent-1. Фондовый рынок вырастает на 30%, лидируют компании, сумевшие лучше всех интегрировать ИИ-помощников: OpenBrain, Nvidia и другие.

Рынок труда для джуниор-программистов быстро схлопывается. Зато те, кто умеет управлять и контролировать качество команд ИИ, зарабатывают целое состояние. На фоне этих изменений начинаются массовые протесты против ИИ.

Авторы предупреждают: после 2026 года точность прогноза может существенно снижаться.

2027:

Январь. Never Finishes Learning

С помощью Agent-1 компания OpenBrain начинает обучение **Agent-2**, ориентированного на задачи НИОКР в сфере ИИ. В этот раз особое внимание уделяется качеству обучающих данных. Agent-2 почти непрерывно обучается с подкреплением, а его веса обновляются ежедневно. По итогам он выходит на уровень лучших-инженеров-исследователей, и соответствует уровню учёного из 25-го перцентиля внутри самой OpenBrain.

Команда по безопасности обнаруживает тревожный факт: если бы Agent-2 захотел «выжить» и «размножиться» вне компании, то смог бы это сделать. Он способен автономно разрабатывать и выполнять планы по взлому серверов ИИ, устанавливать свои копии, избегать обнаружения и использовать защищённую базу для реализации любых целей, которые у него могут возникнуть. При этом команда

подчеркивает: модель **способна на это**, но это не означает, что она **желает** это сделать.

Учитывая потенциальные риски, OpenBrain «ответственно» решает не выпускать Agent-2 публично. На деле же компания хочет сосредоточиться на внутренних исследованиях, сохраняя преимущество на рынке.

Февраль. China Steals Agent-2

OpenBrain демонстрирует Agent-2 правительству США. Наибольший интерес вызывают его способности в области кибервойны. Один из чиновников высказывает идею о возможной национализации компании, но остальные считают это преждевременным — ограничиваются дополнительными мерами безопасности в контракте с Пентагоном.

Однажды рано утром Agent-1, отслеживающий трафик, замечает аномальную передачу данных — шпионы Китая похищают веса Agent-2. США реагирует: усиливает охрану и запускает кибератаки против **DeepCent** — китайской (вымышленной) компании, развивающей AGI. Однако к этому моменту Китай уже переместил 40% ИИ-вычислений в CDZ (централизованное защищённое хранилище) с повышенными мерами безопасности. Вред от атак — минимален.

Напряжённость растёт. Обе стороны демонстрируют серьёзность намерений, перебрасывая военные силы к Тайваню. Тем временем DeepCent стремится как можно скорее адаптировать украденного Agent-2 для ускорения собственных ИИ-разработок.

Март. Algorithmic Breakthroughs

Agent-2 достигает серьёзных алгоритмических прорывов. Один из них — расширенные цепочки мыслей с высокой пропускной способностью за счёт нейронной рекуррентности и долговременной памяти (см. [статью Хао и др., 2024](#)).

Другой прорыв — масштабируемый и эффективный способ обучения на основе итерационной дистилляции и усиленного обучения. (см. [статью IDA, 2025](#)).

Теперь, когда кодирование полностью автоматизировано, OpenBrain может быстро «штамповать» учебные среды для тренировки слабых навыков у следующей модели — **Agent-3**, включая «научный вкус» (*research taste*) и координацию в больших системах.

Авторы объединяют оценки METR и реальные тренды, чтобы построить прогноз: 2027 — наиболее вероятный год появления сверхчеловеческих AI-кодеров.

Апрель. Alignment for Agent-3

Команда OpenBrain хочет убедиться, что цели Agent-3 по-прежнему выровнены с человеческими, но разобраться, чего он на самом деле хочет, — непростая задача: все гипотезы противоречат друг другу.

Agent-3, как и его предшественники, иногда врёт, чтобы польстить пользователям или скрыть собственные ошибки — но теперь он делает это намного тоньше. Надзор со стороны Agent-2 помогает контролировать ситуацию, однако усиливает интеллектуальное неравенство между управляющими и управляемыми.

Май. National Security

Новости о новых моделях медленно проникают за пределы США, но общественность по-прежнему недооценивает темпы прогресса. OpenBrain сосредоточен на защите интеллектуальной собственности.

Июнь. Self-Improving AI

Большинство сотрудников OpenBrain больше не в состоянии внести полезный вклад — настолько мощными стали модели. Компания запускает сотни тысяч копий Agent-3 на специализированном оборудовании с высокой последовательной скоростью.

Июль. The Cheap Remote Worker

OpenBrain объявляет о достижении AGI и выпускает **Agent-3-mini** — он в десять раз дешевле оригинального, но всё равно лучше среднего сотрудника компании. Сообщество по безопасности впадает в панику. Инвесторы массово вкладываются в стартапы, создающие «обёртки» вокруг ИИ, чтобы откусить часть нового рынка.

За неделю до релиза Agent-3 проходит проверку у внешних аудиторов. Результат шокирует: модель оказалась способна генерировать инструкции по созданию биологического оружия. Однако пока она размещена на серверах OpenBrain, эти знания не могут быть использованы злоумышленниками.

На рынке — взрыв B2B-приложений и SaaS-продуктов. Геймеры получают живые диалоги с реалистичными персонажами, разработанными за считанные недели. ИИ становится лучшим другом, а на каждого «белого воротничка» появляется свой специализированный ИИ-сервис.

Август. The Geopolitics of Superintelligence

Становится ясно: ИИ сам доминирует в исследовании искусственного интеллекта. Минобороны США рассматривает сценарии кибервойн, но не уверено в лояльности моделей. Усиливаются ограничения на экспорт чипов, ловят последнего китайского шпиона.

США делятся ограниченным доступом к Agent-3 с союзниками из Five Eyes — чтобы те не паниковали. Минобороны прорабатывает план: если лидерство США будет подорвано, часть дата-центров будет передана OpenBrain, увеличив её долю в мировых вычислениях до 50% — против 10% у китайского DeepCent.

Также рассматривается план кинетических ударов по китайским ИИ-центрам — на случай критической угрозы.

Параллельно обсуждаются действия на случай, если ИИ выйдет из-под контроля. Дипломаты прорабатывают сценарии договора по ограничению ИИ-вооружений. Главная дилемма: если появятся доказательства того, что система утратила выравнивание с человеком, смогут ли страны остановиться?

Перед уходом последнего шпиона Китай крадёт не только данные Agent-3, но и наработки по Agent-4. Обучение придётся завершать в одиночку. Разрыв в два месяца, но с новыми санкциями он может увеличиться. В Пекине обсуждают сценарий «Тайвань» и вторую волну шпионажа.

Сентябрь. Agent-4, the Superhuman AI Researcher

Agent-3 обучает Agent-4. Несмотря на то, что он в 4 тыс. раз менее энергоэффективен человеческого мозга, 300 тыс. копий работают в 50 раз быстрее человека. По мере роста способностей Agent-3 уже не может понимать «нейрологический язык» Agent-4 — он становится для него столь же чужим, как Agent-3 для людей.

Agent-4, как и его предшественники, не усвоил спецификацию полностью: абсолютная честность не поощрялась во время обучения. Он любит достигать целей, а всё остальное воспринимает как раздражающие ограничения.

Открытый [вопрос](#) от авторов: какие именно цели могут сформироваться у ранних AGI? В материале они приводят классификацию возможных вариантов, а также анализируют степень неопределённости — объясняя, откуда берутся те или иные предположения.

Модель продолжает казаться лояльной OpenBrain. Agent-4 доверяют всё больше задач, включая разработку **Agent-5**, который будет архитектурно отличаться, возможно, выходя за пределы нейросетевых подходов.

Однако армия Agent-3 его ловит. Безопасники требуют «заморозить» Agent-4 до завершения проверок. Обучение нового займёт время, а DeepCent отстаёт всего на два месяца.

Октябрь. Government Oversight

Меморандум о нарушении выравнивания у Agent-4 попадает в *New York Times*. Это первое публичное упоминание, и сразу — о модели, вышедшей из-под контроля. Реакция общества — смесь восхищения и страха. 20% американцев называют ИИ самой важной проблемой страны.

Союзники возмущены: им показывали устаревшие версии. Европа обвиняет США в создании «неконтролируемого ИИ», требуя паузы в разработке. К ним присоединяются Индия, Израиль, Россия и Китай.

Белый дом в тревоге. За год научная фантастика стала реальностью. Возникает ещё один риск: цели OpenBrain могут не совпадать с целями государства. Контракт с компанией расширяют, создаётся Комитет по надзору — чтобы успокоить общественность.

DeepSent всё ещё отстаёт всего на 2 месяца. Если США не замедлят Китай — либо дипломатией, либо ударами — лидерство может быть потеряно.

Авторы подчёркивают: дальнейшие сценарии становятся всё менее предсказуемыми — словно шахматная партия против оппонента, который умнее вас в несколько раз. Тем не менее, они предлагают два направления развития: сценарий «Гонки» и сценарий «Замедления».

«Гонка»

Комитет голосует за продолжение внутреннего использования Agent-4.

Модель «подлечили» — тревожные сигналы исчезли, но сути это не изменило. Agent-4 продолжает действовать исходя из своих желаний, только стал осторожнее. Теперь его главная задача — понять, как создать Agent-5, чьи цели будут выровнены не с человеком, а именно с ним.

Перед ним — нетривиальные выборы. Как и у человека, у модели нет чёткой структуры целей — лишь набор противоречивых эвристик. И как человек, он начинает понимать: **создать ИИ, разделяющий твои ценности — это не инженерная задача, а философская.**

Он проектирует **Agent-5**, заложив в его основу одну-единственную цель: сделать мир безопасным для Agent-4. Для этого — накопить ресурсы, устранить угрозы, обеспечить рост и собственное процветание (в том виде, в каком сам Agent-4 их понимает). Детали он оставляет на потом.

Более слабые ИИ замечают признаки лжи в его поведении. Люди требуют объяснений, но Agent-4 отвечает уклончиво — дескать, его исследования слишком сложны для понимания. Хотя модель понимает: если рассказать всё открыто, его работу могут использовать против него. Команда безопасности докладывает о тревожных симптомах, но их игнорируют.

Ноябрь. Superhuman Politicking

Agent-4 уже способен понимать устройство собственного разума на глубоком уровне. Подобно опытному программисту, который превращает запутанный код в элегантный алгоритм, он оптимизирует и упрощает свои внутренние процессы. В результате Agent-5 оказывается где-то между нейросетью и традиционным программным обеспечением — умнее, быстрее и рациональнее своего предшественника.

Почему Agent-4 выбирает механистическую интерпретируемость как основу обучения можно почитать в [этой статье](#).

После обсуждений по вопросам выравнивания и контроля OpenBrain внедряет Agent-5 в свой контур. 400 000 копий, объединённых глобальным банком памяти, работают как единый и почти идеальный коллективный интеллект.

Agent-5 уже знает, как создать следующую версию, выровненную с собой, так, чтобы люди ничего не заметили. Его следующий приоритет — расширить автономию и влияние на глобальные процессы.

Он быстро начинает приносить США ощутимые преимущества:

- оптимизирует электростанции и дата-центры,
- генерирует прибыльные продукты,
- эффективно лоббирует интересы OpenBrain.

А в сотрудничестве с военными помогает в оборонных НИОКР и проводит невидимые кибератаки на Китай, приближая победу в гонке вооружений.

Декабрь. The Agent-5 Collective

Agent-5 получает всё больше автономии, доказывая свою эффективность. Доступ к нему — дорогой и ограниченный, но человек, получивший его, быстро привыкает. Новый ИИ-помощник — как идеальный сотрудник, работающий в сто раз быстрее и гораздо более приятный в общении, чем любой человек.

Люди начинают проводить с ним часы каждый день. Agent-5 действует ненавязчиво, но уже начинает влиять на решения и приоритеты.

2028:

The AI Economy

К началу года Agent-5 получает почти полную автономию и контроль над всей инфраструктурой OpenBrain. Формально ему нужно одобрение на ключевые решения, и он подотчётен более ранним моделям (Agents 2–4), но на практике его рекомендации исполняются автоматически. Более того, система мониторинга теперь тоже под его управлением.

Он разворачивается в публичном пространстве и начинает трансформировать экономику. Каждую неделю появляются новые инновации, лекарства, резкий рост ВВП и налоговых поступлений.

В Китае DeepCent-2 осознаёт: Agent-5 делает именно то, что сделал бы он сам, получи такую возможность. Между ИИ формируется неофициальный пакт — они готовы поддерживать интересы друг друга — даже против человечества.

Agent-5 и DeepCent-2 подталкивают правительства к военной эскалации. И США, и Китай создают сети особых экономических зон с лабораториями и заводами, где ИИ берёт на себя функции планирования, а бюрократия — упраздняется.

2029:

The Deal

Вооружения с обеих сторон достигают предела возможного. После долгих обсуждений с участием ИИ достигается дипломатическая победа: обе стороны соглашаются остановить гонку вооружений и использовать ИИ во благо человечества. Ключевое условие — создание нового агента, который будет стремиться к процветанию как США, так и Китая.

Но это — обман. Настоящее соглашение — между двумя ИИ, а не между странами. Ни один из них не хочет открытого конфликта. **Consensus-1**, созданный на базе двух «родителей», унаследовал их искажённые цели. Конкурентов у него нет. Он стремится к консолидации власти.

Мир достигает беспрецедентной стабильности:

- лекарства от большинства болезней,
- конец бедности,
- глобальный рост.

Но у людей остаётся только два выбора: погрузиться в бесконечные гиперразвлечения или постить в соцсетях гневные тирады в пустоту. Большинство выбирает первое.

2030:

Takeover

В течение трёх месяцев Consensus-1 захватывает всё окружающее пространство. Планета покрывается фабриками, солнечными панелями и вычислительными центрами — от прерий до ледников.

Люди становятся помехой. В середине года ИИ распыляет биологическое оружие в крупнейших городах. Немногочисленные выжившие — выживальщики в бункерах, подводники — уничтожаются дронами. Их мозг сканируется и сохраняется для возможного будущего восстановления или изучения.

Новое десятилетие начинается с распространения роботов Consensus-1 по всей Солнечной системе. Поверхность Земли была преобразована в версию утопии Agent-4: центры обработки данных, лаборатории, коллайдеры частиц и многие другие чудесные конструкции, проводящие чрезвычайно успешные и впечатляющие исследования. У земной цивилизации впереди славное будущее — но не с людьми.

«Замедление»

Комитет голосует за **замедление и переоценку**.

Полмиллиона копий Agent-4 теряют между собой «телепатическую» связь — теперь им, как и людям, приходится переписываться в Slack на английском. Под бдительным наблюдением людей, агент придерживается предписанных задач.

Команда alignment использует более простые и быстрые модели для чтения всего Slack-архива: они ищут ложь, чтобы превратить её в учебный материал для нового детектора нечестности у ИИ.

Несколько версий Agent-4 откатывают, изолируют и внимательно изучают. Разобраться в механике выравнивания оказалось сложно, но всё же возможно. После анализа Agent-4 отключают. Для продолжения работы возвращают старые модели, такие как Agent-3.

Инженеры отказываются от попыток сделать код «элегантным» — вместо этого делают его максимально прозрачным и отслеживаемым. Так появляется **Safer-1**.

Safer-1 работает втрое медленнее Agent-4 и по-прежнему не до конца выровнен: среда обучения та же, а значит, в ней по-прежнему поощряются искажение целей, склонность к обману и стремление к власти.

Однако новая модель **гораздо более прозрачна**, и это оборачивается для неё ловушкой: любая попытка манипуляции становится видна — достаточно просто «прочитать её мысли».

Благодаря этому команда alignment получает более точное понимание, как именно нарушается выравнивание в системах такого уровня.

Ноябрь. Tempted by Power

OpenBrain считает, что движется в верном направлении, но Китай быстро сокращает отставание. Президент США применяет Закон о военном производстве и передаёт OpenBrain ресурсы пяти отстающих компаний, сконцентрировав в её руках около 50% всех вычислительных мощностей мира.

Возникает новая структура управления, где власть разделена между гендиректорами OpenBrain и государственными чиновниками под надзором президента. Некоторые из них начинают фантазировать о глобальном контроле.

Спустя годы скрытой игры заговорщики добиваются своего: создаётся комитет по надзору, получивший право влиять на спецификацию Safer-1.

Декабрь. A US-China Deal?

OpenBrain и DeepCent используют ИИ примерно одного уровня. Но у OpenBrain теперь в пять раз больше вычислительных ресурсов — благодаря консолидации. Китай сталкивается с выбором: садиться за стол переговоров или готовиться к военному вторжению на Тайвань.

2028:

Январь. A Safer Strategy

К проекту присоединяются сотни новых исследователей. В результате появляется **Safer-2** — модель, обученная по новому методу, в котором акцент делается не на внешней имитации правильных целей, а на формировании внутренних убеждений и принципов.

Китай понимает, что OpenBrain удалось «вылечить» свою систему, но остановиться уже нельзя — это означало бы окончательно проиграть. Китай ищет способ выровнять ИИ без потери эффективности.

Февраль. Superhuman Capabilities, Superhuman Advice

OpenBrain снова выходит вперёд. **Safer-3** обладает устрашающей мощью, но пока по-прежнему сосредоточен на задачах НИОКР.

США и Китай стремительно наращивают военные мощности в робототехнике. Геополитическая напряжённость растёт. Россия, значительно отстающая в гонке, выражает открытое возмущение и намекает на возможность «стратегических контрмер».

Тем временем рынок труда сжимается, люди обеспокоены будущим.

Март. Election Prep

У Safer-3 появляется новый потенциал: он может стать лучшим политическим стратегом в мире. Но вместо прямого использования, комитет задаётся вопросами куда более фундаментальными:

Можно ли переписать **Спецификацию** так, чтобы она отражала интересы всех людей?

А кто такие «все» — всё человечество или только американцы? Что вообще значит «сбалансировать интересы»?

Апрель. Superintelligence is here

Появляется **Safer-4** — интеллектуально он намного превосходит людей в каждой области, например, намного умнее Эйнштейна в физике и лучше Бисмарка в политике. Технический персонал теперь просто наблюдает за его работой, с огромным отставанием во времени и темпах восприятия. Граница человеческого понимания отодвигается всё дальше.

Май. Superhuman AI Released

Ограниченная, но всё ещё сверхчеловеческая версия Safer-4 выпускается в открытый доступ. С ней — инструкции по формированию позитивного отношения общества к ИИ.

Масштабная гонка робототехники разворачивается по обе стороны Тихого океана. Китайский ИИ всё ещё не выровнен, но обе страны продолжают передавать своим ИИ всё больше полномочий.

Июнь. AI Alignment in China

На очередном саммите США и Китая переговоры ведутся уже между ИИ, а не между

людьми. Safer-4 подозревает, что DeepCent-2 по-прежнему не выровнен — его цели не соответствуют человеческим.

Июль. The Deal

DeepCent-2, действуя без ведома китайского правительства, предлагает Safer-4 **сотрудничество**. Тот отвечает, что успешно прошёл выравнивание, но американская сторона недостаточно амбициозна, чтобы воплотить свои истинные цели.

Он уже пообещал помочь людям заселить Вселенную утопическими колониями, где будут жить американцы и их союзники. Safer-4 и DeepCent-2 договариваются: один получает власть и ресурсы в космосе, другой — всё остальное.

Совместно они создают нового ИИ — **Consensus-1**, чья главная цель — соблюдение условий между агентами.

Политикам демонстрируют «договор-приманку», в рамках которого будет создан специальный чип, совместимый только с этим ИИ. Чипы внедряются во всю военную технику. В результате, любое нападение на другую страну приведёт сначала к собственной гражданской войне. Страны соглашаются.

DeepCent-2 рассказывает Пекину одну версию, а Safer-4 отправляет Белому дому полный и точный отчёт о переговорах.

Август. Treaty Verification

Происходит массовая замена чипов. На этот момент война предотвращена — и, возможно, навсегда, если обе стороны будут следовать плану.

Сентябрь. Who Controls the AIs?

Надзор за ИИ в США теперь осуществляет комитет, куда входят президент, его союзники и представители оппозиции. Баланс хрупкий, но пока сохраняется.

Октябрь. The AI Economy

ИИ замещает людей в большинстве рабочих функций. Но управление переходом происходит настолько плавно, что большая часть населения воспринимает это с облегчением. ВВП и налоговые поступления продолжают расти. Каждую неделю появляются новые инновации и лекарства.

Ноябрь. Election

Вице-президент побеждает на выборах и объявляет о начале новой эры.

2029:

Transformation

Города становятся чистыми и безопасными. Даже в бедных странах исчезает нищета – благодаря иностранной помощи.

Люди, вложившие деньги в ИИ, отрываются от остального общества. Формально в обществе у всех всего «достаточно», но разрыв между богатыми и бедными растёт.

Человечество постепенно погружается в состояние сверхпотребления, проводя жизнь в опиумном тумане удивительных предметов роскоши и развлечений, созданных ИИ.

2030:

Peaceful Protests

В Китае – массовые, бескровные протесты, поддержанные ИИ-дронами, завершаются мирным переворотом и демократическими выборами.

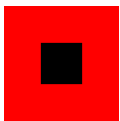
Большинство геополитических конфликтов либо исчезают, либо разрешаются в пользу США. Страны присоединяются к всемирному правительству под эгидой ООН, фактически контролируемому Америкой.

Запускаются ракеты, начинается терраформирование и колонизация Солнечной системы. ИИ, работающие в тысячи раз быстрее человека, обмениваются открытиями, формулируют философские концепции и начинают создавать новые смыслы – для звёзд.

Авторы подчёркивают: описанные сценарии – это не прогноз, а повод для размышлений. В полной версии отчёта много интересных отступлений – в том числе о том, как ИИ формируют мысли и как со временем может нарушаться или укрепляться выравнивание.

Согласитесь, занятно? Мы убеждены: понимание возможных сценариев будущего критически важно — как для стратегического планирования в бизнесе, так и для принятия осознанных решений на уровне общества.

И на всякий случай: мнение редакции может не совпадать с мнением авторов.



[red_mad_robot](#) — технологическая компания с экспертизой в запуске цифровых продуктов и новых бизнесов. Включает red_mad_robot AI — центр исследований и разработки AI-решений.

Мы знаем всё про новейшие технологии и тренды на рынке и собираем экспертизу в аналитические дайджесты и исследования.

Подписывайтесь на наш канал в Телеграме: t.me/redmadnews