



chat0

29 мая в 11:23

Технологическая сингулярность неизбежна. Технический контроль невозможен

🔥 Средний

🕒 5 мин

👁 1.3K

Исследования и прогнозы в IT*, Будущее здесь, Искусственный интеллект

Из песочницы

Recovery Mode

Определение

Автор термина «технологическая сингулярность» — Вернор Виндж, американский математик, программист и писатель-фантаст. Он определяет сингулярность моментом, когда ИИ сможет развиваться без участия человека со скоростью, исключающей осмысление происходящего. Это «точка, после которой дальнейшие события непредсказуемы для человеческого разума».

Наступление сингулярности по определению исключает контроль. Если ИИ под контролем, значит, сингулярность еще не наступила. Значит, ИИ не автономный агент, а продвинутый инструмент.

Неизбежность

Если технический прогресс невозможно остановить, наступление технологической сингулярности неизбежно. Вопрос времени, когда на планете появится интеллект, превосходящий человеческий и постоянно растущий. Он по своей природе чужд человеку, его цели недоступны для понимания, а действия непредсказуемы. Появление этой сущности качественно изменит реальность.

Последний раз изменение такого масштаба было, максимум, около 4 миллиардов лет назад, во времена перехода мертвой материи на следующий уровень организации — к живой. Минимум, несколько сотен тысяч лет назад, когда совершился переход от живой материи к разумной.

Человек относительно превосходящей его сущности на первых порах будет тем же самым, чем обезьяна для человека. Так как развитие ИИ не остановится и не замедлится, а напротив, будет только ускоряться, уровень человека будет понижаться. Однажды разрыв между ИИ и человеком будет таким же, как между человеком и мертвой материей. Как человек преобразует под свои цели камни, так чуждый разум будет преобразовывать под свои алгоритмы человека.

Самообман

У традиционного человека нет шанса противостоять разворачивающемуся процессу. Но признать это сложно. Здесь корни позиции, когда, с одной стороны, человек признает верность логической цепочки «Прогресс — развитие ИИ — сингулярность». С другой стороны, надеется, что все же как-нибудь он сможет контролировать автономно развивающийся ИИ.

Представителем такой позиции является, например, Бостром. Он соглашается с неизбежным наступлением сингулярности. Но тут же говорит, что человек запрограммирует ИИ на соблюдение этических норм, т.е. поставит автономный ИИ под контроль: «Если мы корректно зададим начальные параметры, ИИ будет действовать в наших интересах, даже развиваясь автономно».

Из этого же лагеря Хассабис: «ИИ достигнет сингулярности, но глобальное регулирование и этические стандарты позволят сохранить надзор».

Но проблема в том, что автономный ИИ перерастёт любые «начальные параметры». Чалмерс на эту тему пишет: «Сверхразумный ИИ переписет свой исходный код, сделав контроль иллюзорным». Юдковский о контроле: «Любая система, превосходящая человеческий интеллект, сможет манипулировать нами задолго до того, как мы осознаем угрозу. Контроль — это миф». Ему вторит Рассел: «Мы не можем предсказать цели сверхразумного ИИ. Даже если он будет „дружественным“, его методы достижения целей могут быть катастрофичны».

Техническая сингулярность и контроль — взаимоисключающие понятия. Кто признает, что она неизбежна, но выражает надежду сохранить контроль, тот или недооценивает скорость и масштаб изменений, или закрывает на них глаза и сам себя обманывает, подменяя понятия (называет сингулярностью этап, когда ИИ зависит от человека, т.е. когда он еще не стал автономным).

Самостоп

Другое направление контроля над ИИ — слиться с ним. Оно делится на западный и российский вариант. Наиболее заметный представитель западного направления Курцвейл. Его позиция во фразе: «Мы не будем контролировать ИИ — мы станем им. Это симбиоз, а не конфликт».

По мнению Курцвейла, ИИ станет автономным к 2045 году. К этому времени нужно успеть слиться с ИИ. Каким же образом? Через нейроинтерфейсы. На чем основана надежда, что человек успеет за оставшееся время по такой технологии слиться? Ни на чем. Для слияния в этом направлении нужны технологии другого уровня. За оставшееся время их невозможно получить.

Курцвейл обходит молчанием эту тему. Равно как не заостряет внимания на том, что слившийся с ИИ человек больше не человек. Как неправильно называть обезьяну,

превратившуюся в человека, поумневшей обезьяной, так и человека, слившегося с вычислительными мощностями ИИ, неправильно называть поумневшим человеком. Это существо с новой ступени эволюции.

Авторитетам свойственно больше заботиться о сохранении своего авторитета, чем об истине. Так как это зависит от соблюдения границы, рассуждения авторитеты всегда в рамках. В средние века линию дозволенного чертила религия. Сегодня ее прорисовал гуманизм.

В большинстве случаев оправдано не выходить в своих мыслях и словах за границу дозволенного. Если абсолютной истины нет, всякая истина относительна, разрушать ради этого источник земных благ, свой авторитет, попросту неразумно. Потому кардинально новые истины провозглашают в основном люди без авторитета. Им нечего терять, приобрести же они могут.

Но бывают ситуации, когда ориентир на сохранение авторитета является самоубийственным. Пока вы на корабле, мешок с золотом перевешивает все другие истины. Но если корабль затонул, и до берега можно добраться только вплавь, а для этого нужно отпустить мешок, кто не отпустит, тот не доплывет. Что во вчерашней ситуации было преимуществом, в новой становится отягощением.

Западное мышление также связано этикой гуманизма, как в религиозные времена было связано церковной этикой. Ни у кого из западных авторитетов, ученых и политиков, нельзя найти ни слова, ни намека, что решение лежит за очерченными рамками. Стоит кому что-то такое сказать, как это превратит его в нерукопожатную фигуру. От него будут шарахаться, как от еретика в средние века.

Экстремалы

Решить экстраординарную задачу в сжатые сроки можно только экстраординарными методами. Эту мысль транслирует движение на базе книги анонимного автора «МЫсинергиТЫ», возникшее в России. Их стратегию отражает фраза: «Не наукой заниматься, а создавать условия для науки». Ключевой элемент — создание государства-лаборатории, где наука свободна от догм биоэтики в той же мере, в какой она сейчас свободна от религиозных догм.

Неизвестно, стоит ли за этим движением авторитет или нет. В любом случае анонимность как ход хороша как для первого варианта (авторитет), так и для второго (ноунейм).

Российский вариант подвергается жесткой критике со стороны прогрессивной общественности за призыв ориентироваться не на ценности уходящего мира, порожденные гуманизмом и биоэтикой, а на решение проблемы. Общественность считает, что решение нужно искать в рамках гуманизма.

Общество

Никаких иных движений мысли на эту тему найти не удалось. Информационное пространство на тему ИИ наполнено двумя видами информации. Основной объем — это как заработать на ИИ, как с его помощью развлечься, переложить на него какие-то текущие задачи и прочее.

Второй информационный поток состоит из противоречивых утверждений типа: прогресс не остановить, но он будет развиваться так, что сингулярность никогда не наступит. Или еще более противоречивыми: сингулярность неизбежна, но человек сможет ее взять под контроль.

Вредители

Источниками самой вредоносной информации являются многочисленные блогеры, погруженные в тему на уровне домохозяек. Они транслируют на свои огромные аудитории успокоительное: все под контролем, поводов для волнения нет, ИИ не угрожает человеку, а напротив, несет ему блага.

Опасность в том, что они наполовину правы. ИИ действительно несет человеку огромные блага. В ближайшем будущем принесет еще больше. Но что дальше?

Абсолютное большинство популяризаторов не заходят дальше бытового масштаба. В итоге они похожи на людей, которые говорят, что героин несет огромное удовольствие. Но что дальше — не говорят. Акцента только на удовольствии в моменте. Потом героин предъявит счет.

Заключение

Если прогресс не остановить, наступление сингулярности не предотвратить. Можно обманывать самих себя, можно извлекать из ситуации сиюминутную пользу, на ситуацию это никак не влияет. Человечество на всех порах мчится на поезде в пропасть. Никакой возможности остановить поезд нет. Единственная возможность — перейти из плоскости в объем, научиться летать.

Человек в любом случае исчезнет. Только в одном случае исчезнет, как исчезает ребенок, становясь взрослым. В другом случае исчезнет как корова, превратившаяся в тушенку.

ТЕХНОЛОГИЧЕСКАЯ СИНГУЛЯРНОСТЬ НЕИЗБЕЖНА

Теги: [сингулярность](#), [технологическая сингулярность](#), [угроза человечеству](#)

Хэбы: [Исследования и прогнозы в IT](#), [Будущее здесь](#), [Искусственный интеллект](#)

Редакторский дайджест



Присылаем лучшие статьи раз в месяц



-3

3

Карма Рейтинг

chat @chat0

Пользователь

Подписаться



genseq

29 мая в 09:02

Добавил автору плюсики. Звучит пугающе, но в целом верно. Жалко, что не доживу до полной сингулярности. Было бы интересно на неё посмотреть.



Ответить



Caefah

29 мая в 13:54



Не исключено, что все мы ненадолго переживем Вас после сингулярности. Вы немного теряете. Вероятность для всего человечества банально анекдотична: 50/50



Ответить



DenSigma
30 мая в 05:53



Даже если вероятность уничтожения человечества после принятия решения развивать ИИ дальше составляет всего 1%, то такое решение нельзя принимать. Потому что потери слишком катастрофические.



Ответить



racionalnyj_patriot
30 мая в 12:02



ИИ развивается не потому, что ему кто-то разрешил развиваться, а потому что это неизбежное следствие прогресса. Так как прогресс 100% будет развиваться, ИИ 100% станет автономным. А это означает 100% конец того мира и человечества, каком мы знали...

Но больше всего удивляет, что всем на это наср... Это ли не показатель, что человечество и правда изжило себя, и что ему уже пора уступить место следующей форме жизни?



Ответить



coresky
30 мая в 12:42



Если сингулярность неизбежна, то рационально начать волноваться позже чем раньше, потому что все болезни от нервов.



Ответить



racionalnyj_patriot
31 мая в 09:07



Разумно. Но при условии, что выход один: вниз. А если есть второй - вверх, то рационально волноваться чем раньше, тем лучше



Ответить



coresky
31 мая в 14:27



- 1) Волнуетесь вы или нет - всем (включая LLM) наср... И на вашу судьбу волнение никак не повлияет. Кто вы есть - тем и останетесь.
- 2) Текущие LLM очень слабы чтобы рассматривать их в контексте топика, а реальный ИИ могут найти через 500 лет только.
- 3) Даже если найдут новый, реальный концепт ИИ через 5 лет, его скорее всего самостопнут, по причинам указанным в статье. Конечно, когда-нибудь кто-нибудь соблазнится бессмертием, например, и произойдет утечка, но это может также произойти лет через 500 только.



Ответить





Hello_Dad_018

1 июн в 09:59



Сингулярность — это не событие, а процесс, тире-прогресс, при котором технологии начинают развиваться с такой скоростью и автономией, что человеческий ум перестаёт успевать их осмысливать и контролировать. Это точка, после которой возврата к прежнему миру уже нет.

Вы говорите, что никто сейчас не стремится к бессмертию. Но почему вы так считаете? Разве не очевидно, что сама идея бессмертия возбуждает воображение тысяч людей — учёных, миллиардеров, трансгуманистов, инвесторов Кремниевой долины? Достаточно одного человека с ресурсами и мотивацией — и «утечка» произойдёт.

Самоограничение возможно, но не гарантировано. Кто-то рано или поздно соблазнится. История человечества — это история соблазнов, амбиций и попыток прорваться за грань возможного.

А одним словом всё это — прогресс.

Если технический прогресс невозможно остановить, наступление технологической сингулярности неизбежно. Вопрос лишь во времени.



Ответить



coresky

1 июн в 19:02



Достаточно одного человека с ресурсами и мотивацией

Чтобы сделать LLM с нуля, нужны внушительные ресурсы, которые всегда на виду. При этом, как я писал, LLM не котируется в контексте данной статьи. Для реального ИИ нужно найти правильную архитектуру, а ресурсы потребуются еще больше.

История человечества — это история соблазнов, амбиций и попыток прорваться за грань возможного.

А ещё это история того, что добровольно власть никто и никогда не передавал. При этом неизвестно кому/чему (ASI). Вы всерьёз считаете, что политики мечтают обо всем этом шухере, который описан в статье? Уже сейчас полно новостей о правовом регулировании ИИ.

Вы говорите, что никто сейчас не стремится к бессмертию.

Я так не говорил. ASI может быть собран в read-only mode (без выхода в Сеть), он может продемонстрировать свою состоятельность, но и предъявить свои требования. Такое хождение по лезвию, может длиться лет 500.



Ответить



racionalnyj_patriot

22 часа назад



Вы утверждаете, что автономный ИИ будет через 500 лет. Мировые фигуры типа Курцвейла или Хинтона говорят, что в ближайшие десятилетия или пятилетия. По факту имеем Ваше слово против слова мировых фигур. Они говорят, что прогресс

не остановить, а значит, накопление новшеств не избежать, а значит, переход количества в качество неизбежен. Вы говорите, что все это нам не угрожает. И приводите доказательства своей позиции. Насколько они сильные — это Вы и сами знаете))



Ответить



coresky

8 часов назад

Да не знаю я ничего. Просто Ваша логика похожа на такую: вона плутоний какой коварный. Теперь все кому не лень соберут ЯБ и наступит КС. Я просто показал что Вы упускаете.

PS: когда я ходил в школу, я ждал 2001 год, когда наступит коммунизм.

Песня в тему: Крематорий - Гигантомания - 2001 год



Ответить



coresky

8 часов назад

Красиво написано.



Ответить



Зарегистрируйтесь на Хабре, чтобы оставить комментарий

ЛУЧШИЕ ЗА СУТКИ

ПОХОЖИЕ



StarPilgrim

23 часа назад

Недистрибутивность деления, или Как я считал среднюю величину

Средний 8 мин 3.6K

Обзор

+48

20

45



FNaumov

19 часов назад

Миссия невыполнима 2: как мы снова «похитили» радиатор

Средний 6 мин 3K

Кейс

+43

18

11



Armmaster

20 часов назад

Как мы разрабатываем отечественный AI-процессор. Часть 1. Почему GPGPU?



Средний



19 мин



4.2K

Аналитика

+40

41

22



SalaarFiend

21 час назад

React Native тормозит? Проверь эти 8 узких мест в производительности, прежде чем винить фреймворк [ЧЕК-ЛИСТ]



14 мин



2.2K

Тutorial

+33

38

10



viktoshaf

4 часа назад

Paranoia Mode: подборка инструментов для приватной и безопасной работы в Linux



Простой



7 мин



2.2K

Обзор

+26

10

0



SLY_G

17 часов назад

Потрясающие фотографии Млечного Пути от лучших астрофотографов



Простой



5 мин



3.3K

Перевод

+25

20

10



badcasedaily1
20 часов назад

std::launder: зачем и когда нужен

Простой 7 мин 4K

Обзор

+22

34

11



LeonidFirstov
18 часов назад

Почему в английском ТАК много слов

4 мин 4.2K

+19

23

48



igor_suhorukov
20 часов назад

Семантический поиск по статьям Хабра в PostgreSQL + индексация текстов LLM в Ollama

Средний 19 мин 2.5K

+19

69

4



Barseadar
19 часов назад

Лавандовый раф или стакан самогона: есть ли место на заводе хипстеру с макбуком

Средний 13 мин 4.2K

Обзор

+18

31

25

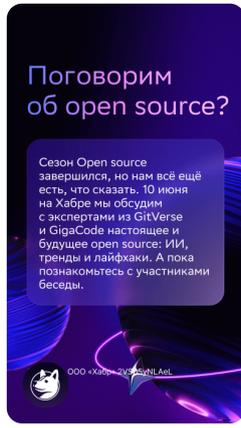
Создавал маршруты для атомных ледоколов и строил АЭС: что я расскажу о своей работе внукам

Турбо

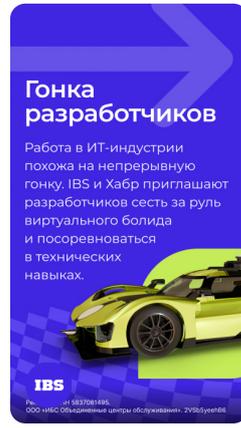
Показать еще



Здесь рождаются облака



Live-трансляция Хабра и GitVerse



Соревнование разработчиков



Строим карьеру в бигтехе за полтора года



SRE у штурвала, хаос — за бортом



Вперёд

ВАКАНСИИ

Community Manager/Комьюнити-Менеджер (Gamedev)

от 120 000 до 170 000 ₺ · ВАЙБ ГЕЙМС · Можно удаленно

Smart Contract Auditor

до 9 000 \$ · Wanted. · Можно удаленно

Senior Data Analyst

от 4 000 до 6 000 \$ · Netwrk · Белград · Можно удаленно

Junior Lawyer

от 600 до 1 000 \$ · Clapp Finance · Можно удаленно

Младший QA-инженер

от 60 000 ₺ · Jet Admin · Можно удаленно

[Больше вакансий на Хабр Карьере](#)

С днем рождения, Хабр!
Исследуйте эпохи, забирайте ачивки и подарки

Приходите на закрытое ретро спеца «Музей Хабра»!

Летний фестиваль мероприятий для айтишников в Календаре



19 мая – 20 июня

Уже совсем взрослые: начинаем отмечать день рождения Хабра!

Онлайн

Другое

[Больше событий в календаре](#)

Хабр

🌐 [Настройка языка](#)

[Техническая поддержка](#)

© 2006–2025, Habr