

ИНТЕНСИВ × ЖЕМАЛ ХАМИДУН

ТРАНСКРИПТ

Вебинар-интенсив (3 поток)

2025 · Интенсив · 3 поток

Расшифровка аудио: Deergram Nova-2, русский язык, уверенность распознавания 97.7%.

Абзацев: 245 · **Длительность:** 2 ч 5 мин

- 00:00:01** Тогда все, я все везде переставил по всем своим многочисленным мониторам. Сажу как оператор матрицы. Сейчас запущу презентацию. Так, теперь вам должно быть видно и прекрасный слайд. Ну что, давайте стартовать.
- 00:00:30** У нас сегодня с вами быстрый старт. Главное не фальш-старт, а быстрый. За 4 дня постараемся с вами погрузиться максимально глубоко, насколько это возможно вообще за 4 дня, в тему искусственного интеллекта. Дальше, естественно, путь может быть как бы самостоятельным. Можно много, на самом деле, в эту тему погружаться, и там есть различные области.
- 00:00:53** Но мы постараемся вам все их подсветить за эти 4 дня. Сегодня у нас первый день основа промптинга, потому что без этого, конечно, дальше двигаться максимально сложно. Нужно разобраться в том, как вообще взаимодействовать с нейросетями и дальше уже углубляться в различные другие практики, типы нейросетей и прочее. Ну и конечно, почему это важно? Потому что сейчас это уже не просто хайп, а временное явление.
- 00:01:21** Искусственный интеллект очень глубоко проник в нашу жизнь, в том числе даже вот в школах появился предмет такой издательство Просвещение относительно недавно тоже выпустило учебник по нейросетям для учеников с 5 по 9 класс. И в общем это говорит о том, что этот тренд с нами надолго, что это не временное явление, ну и видимо на разных уровнях, что сейчас про это все говорят, все пытаются это интегрировать в свою работу, в свою жизнь и так далее. В общем вы в правильном месте, в правильное время и на самом деле еще все только начинается. Поэтому точно не поздно вы опоздали. Почему я могу этим с вами поделиться?
- 00:02:02** Помимо того, что меня зовут Джемал, я Head of AI в Alpina Digital, отвечаю за внедрение инструментов искусственного интеллекта, с 2023 года курирую эту информацию в израильской группе Альпина. Мы начинали вообще с того, что пытались ускорить книжное производство, то есть применить нейросети к тому, чтобы быстрее работать на этом рынке, скажем так, без потери качества, безусловно. Вот и с этим мы, конечно, как раз и начали путь этой трансформации. А попутно много всего мы прошли. Сейчас мы находимся уже на уровне того, что мы внедряем в компании практики вайп-кодинга, используем и агентов внутри.
- 00:02:43** И в процессе этой трансформации мы, собственно, запустили Alpina GPT изначально как внутренний инструмент, который помогал сотрудникам

получить доступ к нейросетям, а потом уже вывели его на внешний рынок, потому что возникла гипотеза, что такая проблема может быть не только у нас, а еще и у других людей на рынке. Ну и вот мы там, наверное, с 2024 года активно этот инструмент развиваем, в том числе на внешнем рынке, и у вас в процессе интенсификации тоже будет к нему доступ. Соответственно, вы сможете воспользоваться всеми моделями, инструментами, которые есть внутри этого продукта. Ну и, конечно, меня тоже не миновала эта участь завела я свой телеграм-канал. Так что подписывайтесь, ставьте лайки.

00:03:23

Буду рад вас там видеть. Сегодня мы с вами первым днем поговорим. Вообще начнем с того, чтобы разобраться, откуда вообще все пошло, потому что чтобы понимать, где мы оказались и как мы туда вообще пришли, нужно понимать, откуда мы стартовали. Поговорим с вами, как работают современные нейросети, про некоторые риски, связанные с этим, какие-то галлюцинации и прочее, как их обходить, экосистему различных инструментов разберем, промтинг в разных его видах, посмотрим на некоторые практические кейсы и затронем касательно еще графические нейросети и нейросети для видео и аудио. По этим блокам у вас будут отдельные еще материалы, поэтому в общем сегодня мы с вами охватим максимум полезного.

00:04:10

Еще немножко такой вводной прелюдии. У нас будут кое-где практические блоки, соответственно я вам буду в чат кидать промты. Вы просто сможете повторять у себя, использовать Alpina GPT или Gemin Enterprise. Вопросы можете писать в чат заранее и запись вам будет доступна после. И здесь сразу хочется немножко с вами свериться, поскольку мы с вами в онлайн формате.

00:04:38

Да, эту презентацию конечно вам тоже пришлем обязательно. Ну и насчет Telegram. В общем, пытаемся быть там, но если какие-то другие каналы вам удобнее, то свяжемся с вами через другие каналы связи тоже. Хочется с вами свериться немножко. Напишите пожалуйста здесь у нас такой светофорчик.

00:05:00

Давайте условно возьмем цифры 123, где 1 регулярно использую, 2 иногда пробую и 3 еще нет. Напишите, пожалуйста, в чат циферку, используете ли вы искусственный интеллект в своей жизни, работе и так далее, чтобы мы тоже немножко выровнялись. Так, вижу 1232123213213. Супер, как у нас сегодня почти, ну не прям, конечно, поровну, наверное, да. Но есть и те, кто вообще не использует, и те, кто иногда пробует, и те, кто регулярно.

00:05:40

В общем, я думаю, что всем будет полезно с разных граней. Вот, конечно, особенно те, кто вообще еще не использует. Я думаю, что для вас сегодня

будет прям очень много полезной новой информации. Так что, пожалуйста, усаживайтесь поудобнее. И хочется еще последний такой момент еще уточнить у вас: а что бы вы хотели освоить в первую очередь?

00:06:05 Что наиболее важно в работе с нейросетями? Может быть это генерация писем, каких-то текстов, может быть исследование, может быть что-то еще, то, что здесь на слайде не уместилось, какие-то другие задачи. Так вижу 233232313. Супер, спасибо, что делитесь. Так правильно писать промты, да обязательно.

00:06:43 Окей, хорошо. Если вдруг что-то вас еще догонит, по ходу вы тоже пишете. Я в чат смотрю постоянно, поэтому буду стараться обращать на это все внимание. Ну и давайте мы начнем с вами с небольшой разминки, прежде чем мы пойдем с вами в серьезные темы. Сейчас я вам покажу 4 изображения, не вот эти 4, это так, затравка.

00:07:06 И наша с вами такая будет игра небольшая угадать, какие изображения сгенерированы искусственным интеллектом и какие являются объектами реального мира. И вот первое фото. Я держу небольшую паузу, чтобы вы могли его рассмотреть. Посмотрите внимательно на детали, что-то, на что вы, может быть, обратите внимание какие-то. Никаких здесь пока не буду раскрывать тайну.

00:07:32 Я вам ее скажу в конце после того, как вы посмотрите на все 4 изображения. И вы пока смотрите, как вы думаете, это настоящее изображение или ИИ. Так, следующее показываю. Так, да, вот я вижу по поводу первого уже написали фото. Так, теперь вот второе это фото или и.

00:07:56 Так, теперь выждал небольшую паузу. Теперь третье. Котик. Ну котик-то должен быть настоящим, да? Можно же что-то святое в этом мире остаться.

00:08:20 Так давайте посмотрим. Кто-то говорит, что все будут ИИ. А вдруг а вдруг нет но сейчас мы узнаем так смотрим на котика и сейчас пойдем дальше следующее И последняя четвертая. Тоже смотрим фото или и. Мнения разделились тут, я вижу, все вперемешку Фото и фото и фото и фото и фото и так Ага Так котик настоящий, остальное.

00:09:09 Ну котик должен быть настоящим. Обязательно. Так хорошо. Спасибо, что поделились. Так, ну, а тогда знаете еще хочется, а на какие, по каким признакам вы определили, что ИИ это ИИ?

00:09:27 То есть что-то вас натолкнуло на мысль, может быть. Или это просто была интуиция. Тут тоже зачитывается детали. Окей. Ну, вот на этой картинке я согласен.

- 00:09:41** Тут даже специально ее ее явно делают. Здесь надписи такие вот на каком-то непонятном языке несуществующем. Мужчина идет прямо в забор. Он, скорее всего, его смог бы обойти, наверное, маневрировать, но тем не менее. Все люди идут в 1 сторону.
- 00:09:57** Довольно странно. Вот поэтому да, вот в этой фотке я согласен. С котиком ну тут прям сложновато. Смотрите, даже видите на отсветах солнце светит в окно и вон видно видите такая значит что-то шерстяное висит, и из нее такая лосинка торчит. У котика усы загнулись, шерстка так в общем все неравномерно.
- 00:10:23** Ну в общем как будто бы сложно, да? С креветкой согласен. Хотя это может быть ресторанная рекламная съемка, поэтому действительно сложно. Вроде ну креветка и креветка. Понятно, что может быть есть какие-то признаки, если приглядываться.
- 00:10:43** Но и с женщиной тут тоже есть ряд признаков, я согласен. Например, вот 2 одинаковых практически мужчины сидят с одинаковыми сильными куртками. Ну и вот такие круги, хотя здесь вот на стакане практически нет вот этого эффекта, когда человек пьет обычно кофе. Здесь вот такая, ну всегда в общем какой-то след остается. Как будто выпита чашка, а следа нет.
- 00:11:11** Но вообще сложно, сложно. Не буду вас томить, это все изображения сгенерированные искусственным интеллектом, все 4 иишные. Использовалась здесь NanoBanana Pro. Это модель от компании Google или еще ее называют Geminite 3. И на самом деле очень сложно стало отличать изображения, которые сгенерированы искусственным интеллектом, от реальных фотографий, изображений реального мира.
- 00:11:37** Но не только изображений это касается. Давайте с вами посмотрим, ещё немножко в таком вот интерактиве с вами побудем. Сейчас я вам покажу видео. Видео со звуком, чтобы было видно. Первое Хамелеон.
- 00:12:00** Так, хамелеона посмотрели. Давайте еще разочек чтобы было Так, окей, следующее Ваше самое большое разочарование в детстве? Я хотел ручку с динозаврами. Папа сказал: Ты же не ребенок. Мне было 9.
- 00:12:21** Купили мне тогда унылую тетрадь в клетку. А я влюбился, где кот в кепке, папин бродяга. До сих пор думаю, может, был бы рэпером. Неделю выпрашивал ластик в форме лапки, там розовая такая была река, прям лапа лапочка. Но выбрали обычную, белую, чтоб не отвлекала.
- 00:12:36** Пучки единороги с подвесками. Я хотела только их. Но нельзя, сказала мама, слишком девчачья. Нет, это для меня. Просто карта лояльности она как

машина времени, в детство.

- 00:12:49** Они что, типа, впервые в метро попали? Да, впервые как дети. Индивидуальность начинается с ручки пенала и любой вещи, которая западает в сердце и не заканчивается с возрастом. Метра, готовы на отлично! Alouda card перемешивается на красной площади или где-то там.
- 00:13:20** Так. И последнее. Это нереально! Кофейно, карамельно, сочно и освежающе. Нереально летние напитки.
- 00:13:46** Только во Вкусно и Точка. На правах реклама нам за это не доплачивают, но тем не менее. Как вы думаете, 4 видео какие являются реальными, а какие сгенерированные искусственным интеллектом? Так, Все похоже на и. Все и.
- 00:14:11** Дети реальные. Последнее может быть реальное, все и. 2 реальное, и похоже. Хорошо, не буду вас тоже томить. Да, это все видео тоже сгенерированы искусственным интеллектом.
- 00:14:35** Вот такой вот подвох. В том числе коммерческие полностью сгенерированы голоса, видео, всё остальное. Некоторые видео взяты из канала 1 энтузиаста, который занимается видеогенерацией. Я сам тоже делал такие ролики, которые очень сложно отличить от настоящих, особенно в последних моделях. То есть эти ролики ещё некоторое время назад делались, поэтому чуть уже устарели.
- 00:15:00** Качество зум, к сожалению, подрезает. Да, поэтому понимаю вас. Но если вы даже посмотрите, там нет этих самих артефактов. Сами видео хорошо сгенерированы на коммерческом уровне и теперь уже можно генерировать вот такие ролики. Это делается относительно несложно.
- 00:15:15** Последняя модель, которая вышла от компании WhiteDance, модель называется Sidence 2, она вообще генерирует просто кинематографического уровня качества ролики и отличать их стало от настоящих. Всё сложнее и сложнее. В общем, есть такой тест интересный, чтобы вам показать прогресс. Когда-то давно с изображениями было точно так же. Если вы помните, были вот эти мемы ходили про шестипалые руки, которые генерировались.
- 00:15:43** Сейчас всего этого давно уже нет. Изображения вышли на уровень, что берешь просто 2 картинки и не можешь их отличить. Теперь это подходит постепенно к видео, и в 23 году был такой мем с Уиллом Смитом, который разошелся широко по соцсетям. Там Уилл Смит ел лапшу. Но это было сделано в первых видеомоделях и было ужасное качество, на это можно было только со смехом смотреть.

00:16:12

И потом, сам Уилл Смит тоже поддержал этот мем, и теперь каждый год есть специальный тест, как Уилл Смит ест лапшу для того, чтобы показать прогресс видеогенерации. Вот сейчас я покажу вам коротенькое видео, как раз таки и я В общем, к чему я все это? Прогресс продвигается так быстро, что на самом деле очень тяжело за всем этим уследить. И поэтому мне хотелось даже начать сегодня с такого вот особенного важного момента, который касается нас всех. Это то, что в принципе все, что мы с вами видим в интернете, нужно относиться максимально критически, критически подходить и оценивать то, что вы видите, потому что нейросети, даже те, которые есть в публичном доступе, подошли на такой уровень, что можно делать кинематографическое видео.

00:18:01

Возможно, в лабораториях есть какие-то нейросети, которые еще, например, в публичный доступ не вышли. Долгое время SIDENTS был в ограниченном доступе. Вот сейчас его выпустили в открытый, пытаются как бы там ограничивать возможность использовать лица реальных людей. Но все равно есть способы обхода всех этих ограничений, поэтому там модели действительно продвинулись очень-очень сильно. Поэтому будьте, пожалуйста, внимательны.

00:18:25

Есть много случаев мошенничества, в том числе с искусственным интеллектом. Там 1 из наиболее известных, наверное, таких кейсов в 2024 году финансовый директор перевел 25 1000000 долларов мошенникам. Это был дипфейк видеозвонок, на котором все были искусственный интеллект, были специальные аватары. Был видеозвонок с SEO и коллегами. И вот есть много других тоже кейсов, не ударяясь в глубокие подробности.

00:18:53

Если интересно, можете почитать сами. Вот, но важно просто это иметь в виду. Поэтому в цифровой среде и контент стал очень сильно похож на реальный, дипфейки используются для различных манипуляций. Всегда проверяйте источники и критическое мышление это 1 из наших главных защит, скажем так. Зачем нужно вообще и мышление, и почему мы должны понимать, в том числе исходя из определенных правил безопасности, и это как калькулятор для такой интеллектуальной работы.

00:19:24

В общем-то, наверное, можно считать столбиком, можно вручную. Но зачем, если собственно есть калькулятор? Точно так же можно вручную писать каждое письмо, анализировать документы глазами, выгулять информацию по 10 вкладкам. В общем, зачем, если есть ИИ, он делает это в 10 раз быстрее и зачастую делает это качественнее. И второй аргумент, конечно, может быть жестковатый, но кто не использует ИИ, проигрывает в скорости и качестве.

00:19:48

Это, по сути, такое соревнование бегуна и велосипедиста. Бегун может сколько угодно быть талантлив, но велосипедист, скорее всего, все равно будет быстрее, особенно на длинной дистанции. Поэтому искусственный интеллект это ваш велосипед для интеллектуальной работы. Ну и третий это навыки промтинга. Это такое конкурентное преимущество сейчас в любой профессии: маркетолог, юрист, HR, менеджер, предприниматель.

00:20:10

Для всех это сейчас актуально. И через несколько лет использование искусственного интеллекта будет, наверное, таким же обязательным, как сейчас Excel или многие другие. Просто пользование компьютером. Для понимания статистики еще 1 цифру вам приведу интересную, тоже не так давно натолкнулся на исследование. На планете возьмем некое число около 8 1000000000 людей.

00:20:34

Из них 68 1000000000 еще не пользовались никакими инструментами искусственного интеллекта, вообще никакими, просто не соприкасались. 12 1000000000 людей соприкоснулись с бесплатными инструментами, то есть попробовали какой-нибудь бесплатный DeepSig, ChatGPT или что-то еще. Только 50 1000000 примерно по разным подсчетам человек на планете пользуются какими-то платными инструментами. То есть, если вы пользуетесь платным, вы уже входите плюс-минус 50 1000000 человек. Это меньше, чем пользуются подписки Netflix вообще в мире.

00:21:05

Там около 300 1000000 пользуются Netflix, смотрят кино. А вот здесь мы говорим с вами про инструменты для работы. И только по разным подсчетам тоже там от 3 до 5 1000000 человек строит какие-то продукты с искусственным интеллектом, то есть агентов делает, что-то еще и так далее. Поэтому если вдруг вокруг вас в соцсетях все стали вайб-кодерами, какими-то строителями агентов и прочим это просто соцсети так раздуты, такой вот пузырь информационный. Но если вот так широко взять широкую рамку, то в реальности все совсем не так.

00:21:37

Поэтому вы еще точно успеете в этой гонке, мы все успеваем за тем, чтобы как бы в этот поезд сесть, запрыгнуть. Но я думаю, что в течение двух-трех лет это уже будет условно у каждого в телефоне, у каждого в компьютере. Поэтому в общем расслабляться здесь тоже не стоит. Еще хочется вам быстренько показать матрицу, которую мы используем для задач и автоматизации, то есть прежде чем мы с вами вообще разберемся в техниках, пойдём вот в эту глубину, возникает вопрос вообще, что делать-то, что автоматизировать и зачем. Вот такой хороший, простой инструмент очень на самом деле.

- 00:22:15** Главная мысль какая, да? Нам нужно разобраться, какие задачи в первую очередь автоматизировать. И левый верхний квадрант это простые рутинные задачи, шаблонные письма, ежедневные какие-то отчеты, ответы на типы вопроса. Здесь может и заменять практически полностью. То есть вы 1 раз настроили промт или ассистента, дальше он работает за вас.
- 00:22:38** Правый верхний простые и творческие посты для соцсетей, идеи для контента, дизайнер баннеров и прочее. ИИ может ускорять очень сильно, в 5-10 раз. Левый нижний сложный и рутинный анализ договоров, отчетность, аудит данных ИИ помогает, но делает некую черновую работу. А вы проверяйте и дорабатывайте. То есть все-таки стоит относиться к этому как к черновику, и не отдавать совсем.
- 00:23:02** И правый нижний сложные творческие задачи: стратегия, продуктовые решения, переговоры и так далее. Здесь человек плюс ИИ. То есть машина вам дает данные варианты, а человек все-таки принимает итоговое финальное решение. Поэтому мой такой совет начать с левого верхнего квадранта, просто разложить свои задачи по этой матрице и начать с вот этих простых рутинных задач. Эффект будет мгновенным, а вы поверите инструмент и дальше будете расширять применение.
- 00:23:28** То есть очень важно на старте использования инструмента какой-то словить вау-эффект, что я это делал несколько дней и вдруг я сделал это за 10 минут. Вот тогда сразу же вы станете тем самым последователем, вот, как когда-то, там, у меня это было, и вам будет легче просто во все это пойти инвестировать в это время. 92% компаний на сегодняшний день из списка Fortune 500 уже используют генеративный ИИ в своих рабочих процессах. То есть это не эксперименты, а уже реальные данные. И на самом деле этот процент все только будет продолжать нарастать.
- 00:24:06** Но есть еще и обратная сторона это то, что большой процент из этих компаний реально попробовали только пилоты, а в работу на самом деле серьезно инструменты практически не интегрированы. И это во многом связано с сопротивлением самих людей. То есть люди в целом не очень-то спешат автоматизировать свою работу, потому что боятся. Ну очень много, целый комплекс причин. И если мы видим, что, например, в разработке можно больше 80 процентов задач решать с помощью искусственного интеллекта реально используется в процентах 20.
- 00:24:42** В общем, на эту тему тоже есть много разных исследований, но от цифры исследований давайте немножко переместимся в историю. И вообще-то, на самом деле, тема искусственного интеллекта стартовала ещё 70 лет назад. Нам важно туда заглянуть ещё, потому что важно понять, как мы оказались в

точке, в которой мы называем сейчас взрывным ростом возможностей, и каждый год по сути все это прогрессирует по экспоненте. А началось все это с простого вопроса может ли машина думать. В 1950-е годы, 1950 год Алан Тьюринг, гениальный британский математик, во время Второй мировой войны взломал немецкую шифровальную машину Энигма, он опубликовал статью с простым вопросом: может ли машина думать?

00:25:29

Это была не философская абстракция. Тьюринг предложил очень конкретный тест: если вы общаетесь с кем-то через текстовый чат и не можете определить, что человек это или машина, значит машина мыслит. И, собственно, это и был знаменитый тест Тьюринга. А через 6 лет, в 1956 году группа ученых собирается на конференцию в Дартмутском колледже. Именно там рождается этот термин *artificial intelligence*, то есть искусственный интеллект.

00:25:56

Они были уверены, что создадут эту думающую машину за 1 лето в 56 году. Конечно, прошло 70 лет, и мы только сейчас видим, что как-то это все подошло к той точке, откуда тогда они начинали. Тогда, естественно, не готовы были технологии совершенно, компьютеры были размером с дом, поэтому все это не работало. В 1966 году на тот момент появился первый чат-бот, назывался он Элиза. Создал его профессор MIT Джозеф Вайценбаум.

00:26:27

Элиза имитировала психотерапевта и задавала очень простые вопросы: Расскажите подробнее, а как вы себя чувствуете? И люди реально верили, что, разговаривая со настоящим врачом, секретарша Вайцинбауна просили его выйти из комнаты, чтобы поговорить с программой наедине. Это тогда показало очень важную вещь, что нам не нужен реальный интеллект, чтобы создать иллюзию интеллекта. И по сути, как бы сейчас мы с вами тоже легко можем попробовать, если на том конце сделать хороший правильный запромченный искусственный интеллект, то скорее всего вы с трудом распознаете, а может быть и не распознаете, что там не человек с вами общается. И потом, конечно, на тот момент все это было не готово, была так называемая зима искусственного интеллекта, которая как раз-таки породила создание другого подхода так называемых экспертных систем.

00:27:22

Это набор правил если то. Может быть слышали *if else*, и часто это в формулах Excel применяется. То есть, если температура пациента выше 38 и есть кашель, значит, вероятно, пневмония, условно. То есть, это работало, но не масштабировалось, потому что для каждой задачи нужно было вручную прописывать 1000 разных правил. Вот, но в целом как бы и переживал различные взлеты и падения, возвращался сильнее.

- 00:27:48** И в 97 году было очень важное тоже событие, которое привлекло еще больше к искусственному интеллекту внимания, попало во все газеты мира. Был создан специально спроектированный компьютер IBM Deep Blue, который обыграл Гарри Каспарова, действующего на тот момент чемпиона мира по шахматам со счета 35-25. Каспаров был в шоке, он обвинил IBM в мошенничестве, настолько хорошо играл в машинах. Но тут что важно понимать? Deep Blue в этом смысле не думал.
- 00:28:18** Он как бы не понимал шахматы, он просто перебирал 200 1000000 позиций в секунду и выбирал лучшую. Это примерно, как если вы, не знаю, все возможные комбинации замка попробуете, вы рано или поздно попадете в правильную. Никакого интеллекта, чисто вычислительная мощь. И это важно понимать, потому что искусственный интеллект принципиально отличается от того, как вот тогда работала IBM Deep Blue. Современные модели не перебирают варианты, они учатся паттернам, и об этом еще чуть дальше поговорим.
- 00:28:49** Но по сути в 1997 году такой вот первый тоже прорыв произошёл. 2012 год важное событие. Мало кто о нём знает за пределами индустрии, но оно многое изменило. Нейросеть AlexNet победила на конкурсе по распознаванию изображений ImageNet с большим отрывом. Компьютер по сути научился видеть, причем не просто видеть, а видеть как бы лучше человека.
- 00:29:13** Ошибка нейросети 35%, ошибка человека 5%. То есть впервые в истории на тот момент машина превзошла человека в задаче распознавания визуальных образов. Это стало возможным благодаря нескольким вещам. Первое это видеокарты с GPU, те самые, которые сейчас стоят в некоторых случаях как квартиры. Оказалось, что карточки, которые геймеры используют для игр, идеально подходят для обучения нейросетей.
- 00:29:37** Второе как бы да, но эти видеокарты считают очень быстро, быстрее обычных процессоров. Второе это большие данные. Интернет к тому моменту накопил уже 1000000 размеченных изображений, на которых можно учить сеть. И третье это архитектура глубокого обучения, то есть нейросеть со множеством слоев, каждый из которых выделяет все более сложные паттерны. И вот с этого момента и начал развиваться экспоненциально.
- 00:30:02** Каждый год был какой-то новый прорыв. Потом, в 2016 году, ещё 1 интересное событие произошло, на тему которого даже отдельный документальный фильм снят. Google DeepMind создал AlphaGo нейросеть, которая обыгрывает чемпиона мира по игре в го. И это важно, потому что игра го это был такой последний рубеж человечества с точки зрения игр.

Игра го считается игрой императоров, самая сложная стратегическая игра в мире.

00:30:32 Здесь как бы не то же самое, что в шахматах. В игре го, для понимания, количество возможных позиций превышает количество атомов во Вселенной. То есть перебор как у Деер Блю здесь просто невозможен в принципе. И AlphaGo по сути пришлось научиться ну как бы интуиции. Вот момент, который вошел в историю это ход No 37 во второй партии.

00:30:55 AlphaGo поставил камень в место, куда ни 1 профессиональный игрок в мире никогда бы не поставил его. Комментаторы решили, что на тот момент, что это какая-то ошибка. Ли Седоль, чемпион мира, встал из-за стола, ушел из зала на 15 минут, потому что не мог понять этот ход. И через 20 ходов стало ясно, что это был гениальный ход. Машина открыла стратегию, которую за 3000 лет не додумался ни 1 человек.

00:31:17 То есть это не перебор вариантов, а по сути машина как бы научилась творчеству. В 2017 году Google изобрел технологию трансформеров, 8 инженеров опубликовали статью Attention is All Your Need, то есть внимание, это все, что вам нужно. И простым языком если объяснить, то трансформеры нейросети они читали текст примерно как мы читаем книгу слово за словом, последовательно. Трансформер читает весь текст целиком одновременно. Представьте, когда вы смотрите на картину, вы же не рассматриваете пиксель за пикселем, вы видите всю картину сразу и понимаете весь контекст.

00:31:51 И вот трансформер делает то же самое с текстом. Это называется механизм внимания модель понимает, какие слова в предложении связаны друг с другом, даже если они далеко. И второе ключевое свойство трансформеры масштабируются. То есть чем больше данных вы им дадите, тем больше параметров и тем умнее они становятся. И, что важно, предсказуемо умнее.

00:32:13 И это свойство как раз и оказалось золотой жилой. То есть, все современные модели GPT, Клод, Джемини и прочие там дипсики они все построены на трансформерах. И по сути вот с этой статьи как раз и стартовала современная эпоха искусственного интеллекта. И в 2022 году произошла революция, OpenAI выкатил ChatGPT, цифры говорят сами за себя, он самый быстрорастущий продукт в истории набрал 1000000 пользователей за 5 дней. Для сравнения, Netflix набрал 1000000 пользователей 35 года.

00:32:48 И почему это важно? Потому что любой человек впервые в истории без технического образования, без программирования, специальных навыков мог сесть за компьютер, поговорить с искусственным интеллектом на

естественном языке, на любом, на русском, на английском, и получить в общем и целом осмысленный полезный ответ. Как гласит легенда, это вообще был незапланированный эксперимент, то есть технология GPT, как переводится Generative Pretrained Transformer, то есть генеративные предтренированные трансформеры. И идея была в том, что они просто ради эксперимента подключили чат-интерфейс к технологии GPT, и вот вдруг модель начала им осмысленно отвечать. И в общем, после этого произошел весь остальной бум в теме нейросетей, то есть остальные модели начали появляться и прочее-прочее.

00:33:40

Ну и как мы понимаем, что, конечно, это была не финальная точка на тот момент 25-26 год, и это уже новая реальность. Мы живём с вами в эпоху, когда ИИ стал массовым, GPT-5-4 уже есть, Клод последней версии 4.6. Говорят, у них уже есть новая модель Mythos, но они ее не выкладывают, потому что она стала слишком опасной, как они говорят. Я, кстати, не сгущаю краски именно так и сказали в лаборатории Антропик. Ее раздали крупнейшим компаниям сейчас в тестовый доступ, и они латают свои дыры в безопасности, потому что модель оказалась настолько сильной, что хакерские возможности ее таковы, что она нашла дыры в продуктах с 27-летней историей, то есть, которые 27 лет никто не видел.

00:34:26

Люди, которые создавали эти продукты, и разработчики не видели, а Мифус просто взял и сразу же нашел все уязвимости во всех системах. И поэтому эту модель ограничили к выпуску. И вот сейчас пока что она еще считается закрытой. Ее тестируют вот эти корпорации Google, Microsoft и все остальные. Есть также Gemini.

00:34:48

И все эти модели показывают уровень PhD профессорский по многим дисциплинам. То есть они не просто отвечают на вопросы, могут рассуждать, анализировать, находить закономерности. Уже появилась мультимодальность, то есть генерация изображения и видео в 1 экосистеме, появились модели, которые генерируют аватаров и многое-многое другое. В общем, сейчас мы живем с вами в эпоху мультимодальности, в эпоху, когда и агенты выполняют цепочки задач автономно, когда искусственный интеллект генерирует код и может делать по сути вам за вечер приложения, на которые раньше нужно было несколько разработчиков и несколько месяцев времени. Ну и конечно всё это используется всё глубже и глубже.

00:35:32

Что же ждёт нас в будущем? Некий такой overview из того, что явно видно, скажем так. Первый мейлстоун это сейчас, то, что мы с вами находимся в текущем моменте, чат-боты, генерация контента, помощники. Вот это такой базовый уровень. Вы задаете вопрос получаете ответ.

- 00:35:54** Это называется узкий искусственный интеллект. В ближайшие год-два произойдет переход к агентам. Они уже в целом работают. Мою презентацию сегодняшнюю тоже верстал агент. Я просто объяснил ему, что мне нужно, слайд за слайдом, согласовал с ним план.
- 00:36:10** Дальше он сверстал мне всю презентацию. Он сам придумал, какие стрелочки какого цвета должны быть, что размещено на слайдах и так далее. Вот у меня есть агент, который верстает мне презентации, потому что это частая задача конкретно в моей работе. Но у вас могут быть какие-то другие агенты. То есть они действуют.
- 00:36:29** Если обычный искусственный интеллект, чат GPT или что-то еще это как мозги, хорошие прокаченные мозги, то агенты это мозги с руками и ногами, назовем это так, которые могут куда-то сходить, что-то сделать, из CRM данные извлечь, документ сверстать. Все, что в общем вы можете сделать в цифровой среде, может сделать вам в принципе и агент. Дальше следующий этап горизонт на самом деле здесь разный. Ведущие лаборатории мира говорят уже, что мы подходим вплотную к этому этапу, к этапу AGI это общий искусственный интеллект. То есть для такого, наверное, если взять общепринятое понимание, то есть это система, которая может решить любую интеллектуальную задачу как человек.
- 00:37:15** То есть не узкоспециализированная, а универсальная, потому что мы ведь с вами многим вещам учимся, которых не знаем. То есть мы с вами не умеем, например, машину водить. Сели, разобрались немножко, посмотрели и научились. Может быть не сразу, не мгновенно, но научились. И вот общий искусственный интеллект будет в этом смысле точно таким же.
- 00:37:34** То есть он будет уметь все то же самое, что теоретически сможет уметь человек. До этапа робототехники, до того момента, как все это будет интегрировано в роботов, естественно, мы с вами будем пока говорить только про цифровую среду, но на перспективе относительно ближайшего горизонта событий это будут также и вещи, которые касаются физического мира. То есть общий искусственный интеллект будет связан с роботами, роботы смогут делать всю ту же самую работу, что могут делать люди в теории. Пока это некая концепция лабораторная, но, как показывают прогнозы, все, что было ранее спрогнозировано, агенты и прочее, и то, над чем смеялись, когда генерировали там эти шестипалые руки, вот сейчас, спустя 2 года, уже совсем вроде бы и не смешно. Все тоже то, что было там написано, оно сбылось в том или ином виде.
- 00:38:24** Возникает только вопрос, как бы что мы можем с этим делать. Гонка лаборатории запущена, так или иначе все в этой гонке будут бежать. AGI это

уже не теоретическая концепция, а только вопрос времени и вопрос как бы, когда это произойдет и как бы что будет происходить. Потому что в теории, если придет AGI, то следом, очень быстро после него придет ASI, то есть artificial super intelligence. Потому что если AGI сможет сам себя улучшать и таким образом себя дорабатывать, то очень быстро мы просто даже не успеем понять, как только AGI придет, сразу же практически наступит ASI.

00:39:01

Но это некая фантастическая концепция. Пока поживем увидим, как говорится. Пока будем с вами исследовать тему искусственного интеллекта и общаться с искусственным интеллектом вежливо, на всякий случай. Но это не точно. И что еще поможет искусственному интеллекту прийти на уровень AGI это тоже такая концепция, которую я исследовал на просторах разных статей.

00:39:31

По сути, нас сейчас ограничивает следующие вычислительные мощности. То есть мы для того, чтобы получить AGI, должны получить кратно больше мощности, чем есть сейчас. Для этого нам нужны квантовые компьютеры. Это не теоретическая область, а вполне себе практическая. Квантовые компьютеры существуют, активно развиваются, просто пока не дошли до стадии промышленной эксплуатации.

00:39:52

Также нам нужна энергия, много энергии столько, сколько сейчас все человечество не может производить, и на эту тему абсолютно серьезно рассуждает посмотрите. Есть интервью с Илоном Маском 2 с лишним часа. Он рассказывает о том, что нужно сделать для того, чтобы искусственный интеллект вышел на тот уровень, на который нужно. Он говорит о том, что проще делать искусственный интеллект в космосе, потому что если делать на земле, то мы не можем так вырабатывать электроэнергию, сколько нам нужно для того, чтобы в космосе. А если выведем в космос, то там мы, соответственно, развернем солнечные панели и сможем обеспечивать нужное количество электроэнергии для того, чтобы снабжать искусственный интеллект.

00:40:32

Ну и там же, соответственно, у него будут видеокарты летать. В реальности это или нет посмотрим. В ближайшие несколько лет он поставил себе задачу к 30-му году все это реализовать. Многие лаборатории сходятся в том, что к 30-му году так или иначе мы там окажемся. И последнее, чего не хватает искусственному интеллекту, помимо электроэнергии, и здесь может еще термоядерный синтез помогать, но это отдельная область, про нее сегодня не будем, Есть еще недостаток знаний о реальном мире.

00:41:00

То есть искусственный интеллект обучался на текстах. Что это значит? Он знает некие теоретические концепции, которые мы когда-либо описали с

вами в книгах, текстах, в интернете. То есть ему скормили весь интернет. Тогда возникает вопрос: а что же с физическим миром?

00:41:14

Ну вот если в пример взять Теслу, то там долгое время обучали автопилот как раз-таки на часах в автомобиле Тесла, как там камеры снимали, как ездит реальный водитель. И благодаря этому автопилот в целом стал неплохо работать. Но а как получить знания о всем вообще, что происходит в физическом мире? Для этого, первое, нужен IoT, то есть нужно встроить систему наблюдения во все устройства, все часы, носимые устройства, телефоны, смартфоны, камеры все соединить в единую сеть для того, чтобы дать источники данных о реальном мире. И вот как только это произойдет в сочетании с энергией, в сочетании с практически бесконечной датой о реальном мире и в сочетании с вычислениями квантовым компьютером мы получим с вами либо AGI, либо сразу практически ASI.

00:42:05

Ну и что произойдет дальше это вопрос, там разные есть концепции научной фантастики. Мы в эту тему философски ударяться не будем, но мне важно было вам подсветить, что происходит с точки зрения трендов. Таким образом Сэм Альтман говорит, что AGI к 27 году будет достигнут, Далио Амадей говорит, что мы уже на пороге. Ну и, соответственно, посмотрим, как это будет происходить. Прямо на этой линии можно увидеть по экспоненте, как вообще в целом выглядит рост ИИ.

00:42:38

Сейчас уже и агенты буквально за год продвинулись так, что они действуют практически автономно. Ну и если теперь немножко отойти от истории развития, от трендов, от того, вообще, куда все движется, теперь попробуем с вами разобраться, как же вообще все это работает. И я постарался простым образом, без занудства и формул, когда я рассказывал про ИИ, часто задается вопрос правда думает или нет? На самом деле, конечно, нейросеть не думает и не понимает, и не чувствует, она делает нечто другое, и очень хорошо так, что создается иллюзия мышления. Если помните вот эту Элизу, да, из 1966 года, то есть люди верили, что разговаривают с врачом.

00:43:19

А современные модели они в общем просто то же самое, только на принципиально другом уровне и аналогия в общем очень простая. Если вы помните T9 на старых телефонах, вы нажимаете какую-то кнопку, клавишу, и он вам прогнозирует, предлагает либо следующую букву, либо слово, которое вы могли часто писать или другие люди могли часто писать. И чем больше вы пишете, тем лучше T9 понимает ваши привычки. Если вы часто пишете слово привет, T9 начинает подсказывать там привет после первых 2 букв. Большие модели типа GPT, Клод, Джемина и в общем-то работают по тому же самому принципу.

00:43:56

Они предсказывают каждое следующее слово, только масштаб другой, конечно. T9 имел словарь в 10 1000 слов, учился на ваших сообщениях, а GPT-4 имеет 1000000000000 параметров, обучен на 10000000000 текстов из интернета книги, статьи, Википедия, форумы, научные работы, деловые документы и так далее. Говорят, даже книги тоже в некоторых случаях, они спиратели и на книгах тоже обучают. Microsoft вроде как даже купила 1 издательство не ради того, чтобы книги как-то развивать, а просто ради того, чтобы data set получить на основе книжек издательства, скормить их себе в нейронку и таким образом получить больше данных. И когда вы спрашиваете ChatGPT, какие документы там нужны, например, не знаю, для кредита, он как бы не знает ответ, он предсказывает, что после слова кредит с наибольшей вероятностью идут слова необходимы, следующие документы.

00:44:48

То есть это такая элегантная простая идея, но в масштабе 1000000000000 параметров она и порождает поведение, которое мы с вами воспринимаем как интеллект. Возможно, мы с вами работаем по тому же принципу, просто чуть более сложно мы обучались всю жизнь, и поэтому у нас там связи как бы как-то по-другому, наверное, это работает все-таки в биологии, но есть похожие паттерны. Токен еще 1 важная вещь, про которую стоит сказать прежде, чем перейдем с вами к практике промптинга. Это базовая единица, которой мыслит нейросеть. Вот пример, допустим, фраза привет, как дела?

00:45:28

Она разбивается на токены: pre, vet, запятая как дела? И вопросительный знак. То есть 1 токен это примерно 4 символа на английском, там на русском немножко по-другому считается, но тоже близко к этому. Важно это потому, что у каждой модели есть ограничение на количество токенов, так называемый контекст или контекстное окно. GPT-4O работает с контекстом 128 1000 токенов.

00:45:54

Клод обычно там от 200, но последняя модель 1000000 токенов контекстное окно. Чем больше контекст, тем более сложные задачи можно решить. Загрузить целый договор, большую таблицу, может быть, несколько документов сразу. И понимание токенов это понимание ограничений, возможностей ИИ, когда вы знаете, сколько информации примерно может помещаться в контекст, вы можете оптимально структурировать ваши запросы. Ну и, конечно, мы дальше подробнее еще разберем эту тему на практических упражнениях.

00:46:25

Еще, кстати, про токен, наверное, можно в такой аналогии сказать. Иногда возникает тоже вопрос, как все-таки считается, когда вы какое-то слово, например, можете целиком произнести, и оно не произносится в сокращенной форме. Какой пример? Допустим, слово кошка будет 1 токен, а

слово, условно, возьмем такую аналогию, а слово собака будет 2 токена. И возникает вопрос, а почему так?

00:46:55 А потому что бывает много собак без буквы А, а много кошек не бывает. По-другому можно как-то произнести, но слово кошке отдельно не существует. Поэтому кошка это 1 токен, как бы целая сущность. А собака 2, потому что собака и буква А на приставке. Это некий пример, аналогия, которая может тоже помочь понять вот эту токенизацию.

00:47:21 Вообще есть отдельные токены-калькуляторы, если эта тема интересна, тоже можно погуглить. Еще важная деталь конкретный пример, как нейросеть подбирает рекомендации. То есть представьте, что клиент обращается, например, и говорит: Мне нужно решение для автоматизации продаж. Что происходит внутри модели? Первый этап анализ запроса.

00:47:41 Модель разбирает контекст, автоматизация, значит, нужно убрать рутину. Продажи значит, CRM воронка лиды. Дальше второй этап сопоставление. То есть если модели дать информацию о каталоге ваших продуктов и услуг, она сопоставит параметры клиента с доступными решениями. Третий этап это ранжирование.

00:47:59 Модель выдает топ-3 подходящих варианта с обоснованием, почему именно они. Здесь важно понимать, и, конечно, не принимает решение за вас, он ускоряет работу, но вместо того, чтобы вручную перебирать весь каталог, вы получаете несколько релевантных вариантов за буквально 30 секунд. То есть по сути это такой вот способ, это постоянное выбрасывание случайных чисел, но благодаря тому, что нейросеть обучена на большом количестве данных, это совпадение иногда выглядит очень-очень реалистично, потому что загрузили в нее огромное количество данных на вход, правильно их разметили, что со всем сочетается, И теперь вот эта предсказательная машина, назовём это так, работает таким образом, что находит вот эти паттерны, предсказывает лучшие варианты ответов. Как это работает, например, с анализом документов? Ещё 1 такой пример.

00:48:53 Представьте, перед вами контракт с поставщиком, например, на 40 страниц. Раньше юрист или менеджер садился вручную, вычитывал каждую страницу, выписывал какие-то ключевые условия, искал подводные камни, и на 1 документ могло уходить от 30 минут до 2 часов. Что может сделать ИИ? На входе тот же самый PDF или Word и извлекает данные, стороны договора, суммы, сроки, порядок оплаты, штрафные санкции, условия расторжения, ограничения ответственности и так далее. А на выходе структурированная таблица ключевых условий плюс список красных флагов, то есть какие-то нестандартные эвильские пункты.

- 00:49:31** И, конечно, огромное количество времени на этом экономится. Но, естественно, тут важно понимать про риски, которые с этим есть. То есть, безусловно, мы можем это делать формате второго мнения или некоего черновика, потому что ИИ убирает рутину, но не убирает полностью необходимость проверки. И здесь важно понимать вот эти ограничения, критически важная тема, эти риски, и может очень уверенно врать. Это называется галлюцинациями.
- 00:50:03** То есть модель с абсолютной уверенностью выдает факты, которых не существует. Она может сослаться на несуществующий закон, привести выдуманную статистику, сочинить реферат на тему, которой вообще никогда не было, и сделать это убедительно со ссылками, с цифрами, экспертным тоном. И вот для бизнеса это, конечно, критический риск, потому что вы можете попросить у ИИ проверить контракт, а он выдумал пункт закона, которого не существует, или выдал неправильные цены конкурента. Ну и второй риск это, конечно, риск утечки данных. То есть, если вы загрузили конфиденциальные данные клиента в бесплатный ChatGPT, они могут использоваться для обучения модели.
- 00:50:40** Поэтому для работы с реальными данными какие-то нужны корпоративные инструменты с гарантиями безопасности. Третий это предвзятость. Ну, и обучен на данных из интернета, которые содержат стереотипы. То есть, иногда говорят, что там вот есть нейросети политически предвзятые, якобы их специально обучали. Но по факту получается так просто их учили на определенном сегменте интернета и поскольку много людей высказывали где-то какое-то мнение по какому-то вопросу, нейросети тоже вобрали это в свою обучающую выборку.
- 00:51:14** И, естественно, в каких-то странах есть определенные мнения и взгляды, которые могут не совпадать, например, с нашими. Нейросеть может уверенно их выдавать, Мы будем думать: ну как это так, нейросеть такая нехорошая. Но по факту это не связано вообще с самой нейросетью или с ее природой. Это просто данные, которые в нее загрузили, и полярность мнений в 1 сторону превесила. Четвертое это ошибки в расчетах.
- 00:51:46** ИИ может перепутать единицы измерения, неправильно посчитать процент. То есть, такое золотое правило, которое можно себе вывести. И ты как умный стажер очень умный, быстрый, эрудированный, работоспособный, но за ним все желательно проверять. В особенности, если мы говорим про какую-то, скажем так, высокорисковую информацию, тот же пример договора. На эту тему есть много кейсов, где люди получали большое количество проблем, особенно в судах, когда использовали искусственный интеллект для этих целей.

- 00:52:16 Поэтому, пожалуйста, тоже имейте это в виду. Как говорится, доверяй, но проверяй. Но в целом современные модели типа kloda или perplexity, допустим, как инструмент, они выдают вам ссылки на источники, с которых взяли информацию. Поэтому вот их можно использовать для того, чтобы делать некий фактчекинг. Экосистема искусственного интеллекта.
- 00:52:38 Давайте на несколько моделей с вами посмотрим. В целом много разных моделей, но условно можно разделить на несколько категорий. Первая это универсальные чат-боты, ChatGPT GPT, от Open AI, Cloud Антропика, Джемин от Гугла и так далее. Это как Microsoft Word, Excel и Google Doc условно, 3 лидера, каждый со своими сильными сторонами. Chat GPT, понятно, самый популярный, но далеко не самый лучший, на самом деле, сейчас уже.
- 00:53:03 Клод лучший, наверное, для рабочих задач. То есть я почти все свои задачи решаю с моделями Клода от компании Антропик, и они дают целую экосистему инструментов тоже. Но они, например, вообще не фокусируются на генерации видео, изображений, звука. То есть они работают только с текстом и кодом. Но в этом они, наверное, лучшие на сегодняшний день.
- 00:53:26 И есть экосистема Google Gemini, которая покрывает вообще все, что только можно придумать в области нейросетей. То есть она встроена в их экосистему Google документов, в их экосистему электронной почты, она встраивается также в ноутбук LM, В общем, везде, где только можно. Плюс внутри их экосистемы есть генерация видео с помощью Veo 3.1 и генерация изображения с помощью нанобанана. Недавно они еще создали инструмент для генерации музыки. То есть на самом деле там у них целый комбайн в 1 только экосистеме Гугла.
- 00:54:03 И, наверное, если перейти дальше, то это специализированный инструмент, perplexity 1 из тех, который может вам пригодиться. То есть это в некотором смысле тот агрегатор, но это изначально поисковик, то есть инструмент для того, чтобы в интернете искать с помощью нейросетей информацию. То есть он не только дает ссылки, но и сразу отвечает на вопрос со ссылками на источники. И ноутбук LM очень хороший инструмент, рекомендую вам попробовать. Вы загружаете ваши документы, а ИИ отвечает только по ним и может по ним сделать аудиоподкаст, может по ним сделать красивую презентацию, ворк-схему и многое-многое другое.
- 00:54:41 E-Sommet это специальный AI браузер, где еі видит все то же самое, что открыто у вас во вкладках и может даже за вас нажимать на кнопки. То есть вы можете дать ему задачу, например, забронировать вам авиабилет или что-то в маркетплейсе заказать, и он пойдет, натыкает все кнопочки нужные, положит в корзину, например, и дальше будет ждать вашего апрува, если вы

ему карточку не дали. А если карточку дали, то может и купить за вас. Но тут нужно быть осторожным. И следующая часть это агрегаторы, различные агрегаторы.

00:55:15

Ну вот, например, Alpina GPT тоже агрегаторы, есть разные другие. Есть в том числе, например, агрегатор видеонейросетей, так называемый казахстанский единорог, который, ну Heitzfield, он сфокусирован исключительно на генерации видео и изображений. И есть генерация контента Midjourney, Sorra, Eleven Labs и различные другие инструменты, которые позволяют вам генерировать разные типы контента. Также есть инструменты, которые позволяют получить к ним доступ абсолютно бесплатно. В принципе ChatGPT, понятное дело, можно тоже использовать в ограниченном режиме.

00:55:54

Cloud у них есть тоже бесплатный тариф. Google Gemini даёт то же самое. А в Gemini Enterprise вы можете получить даже целый месяц бесплатно. Perplexity поисковик, который вы можете использовать тоже бесплатно определенное количество запросов в день. Ноутбук LM позволяет вам тоже использовать его бесплатно.

00:56:11

И Perplexity Comet. В общем, на самом деле много инструментов. Я молчу еще про китайские модели, которые тоже есть. DeepSig, есть Quan и многие-многие другие. В общем, точно есть возможность получить доступ к инструментам.

00:56:25

Вопрос только в том, какой инструмент использовать. Ваш инструмент, который мы вам даем в процессе обучения, мы в него собрали все самые лучшие модели, которые есть на сегодняшний день, по крайней мере для работы с файлами и документами. Там есть встроенные ассистенты, есть специальные проекты с базами знаний, то есть вы можете их использовать тоже для своей работы. И вот этот инструмент вам будет доступен Alpina GPT прямо в процессе нашего интенсива. То есть вы сможете его использовать столько, сколько вам нужно.

00:56:56

И в этом месте как раз мы с вами сейчас перейдем к практике. Основы промптинга с вами разберем. Если у вас инструмент где-то там далеко, то вы можете потихоньку его тоже достать, чтобы дальше какие-то вещи повторять следом за мной. Почему промпт вообще важен, и почему он решает, все ли бесплатные версии доступны без VPN, тоже вижу вопросы. Я могу так сказать.

00:57:23

Большинство зарубежных моделей изначально в России недоступны базово, но они ограничены не с территории России, то есть Россия не запрещает

использование этих нейросетей, по крайней мере пока. А в обратную сторону сами провайдеры говорят, что как бы нельзя использовать их на территории России. Но, естественно, все понимают, как получить к ним доступ, также как и многим другим инструментам. Либо вы можете пользоваться инструментами, которые доступны из России. Например, Alpina GPT, собственно, мы для этого и делали, без всяких проблем получить доступ ко всем этим передовым моделям.

00:58:00

То же самое касается Perplexity. Perplexity доступен из России, и он даёт вам возможность использовать все эти зарубежные нейросети внутри Perplexity. Так что в целом вы можете получить без проблем эти доступы даже к этим передовым моделям. Собственно, если возвращаться к теме промптинга, это такой 1 из главных принципов работы с ИИ, то есть качество запроса определяет качество ответа. Если вы напишете промпт Напиши письмо клиенту, вы получите шаблонное безликое письмо, которое, в общем, никому не поможет.

00:58:33

А если вы напишете: Ты менеджер по ключевым клиентам, клиент ООО, Техносервис, оборот 500 1000000, сейчас на базовом тарифе РКО, например, афиндиректор Петрова, напиши коммерческое предложение на зарплатные проекты, эквайринг. То есть там 200 слов, деловой тон с конкретными цифрами, экономия, то вы получите совершенно другой ответ. Потому что вы загрузили нужную задачу, нужный контекст, и будет тот же самый и та же самая модель, но результат будет кардинально другой, потому что мы дали контекст, роль, задачу, формат и ограничение. Это по сути и есть промптинг, умение правильно формулировать, задавать ваш запрос ИИ. И на самом деле, есть еще побочный эффект, когда вы научитесь хорошо промтить, вы и с людьми начнете более структурно общаться.

00:59:19

То есть, если вы ставите задачу кому-то или кого-то чем-то просите, иногда мы просто там это делаем интуитивно. Но когда вот эта профессиональная деформация в хорошем смысле наступает, вы начинаете промтить людей, и они начинают гораздо лучше реагировать на те задачи, которые вы им ставите, или просьбы, которые к ним адресуете. Ну вот, собственно, здесь пример как раз, который я чуть-чуть заранее сказал. Слева как большинство людей часто общаются с нейросетью коротенькие, пишут промты, проанализируй данные, напиши письмо. Так можно, но вы получите среднестатистический результат.

00:59:56

Потому что еще раз помним, как работает ИИ. Это предсказательная машина. Чем больше контекстом мы ее как бы ограничиваем, тем больше мы отсекаем все лишнее. То есть фактически наш промпт это такой вот способ

сделать ответ модели очень узким специально под нашу задачу. Не выбирать из всего этого многообразия.

01:00:17

Самая такая, наверное, еще подходящая аналогия, вот я вам скажу сейчас, например, фразу, и вы попробуйте за мной ее мысленно продолжить: В лесу родилась и наверняка в голове прозвучала елочка, можно было еще интонацию добавить определенную, тогда бы точно угадали. Почему так происходит? Потому что наша с вами нейронка тоже была натренирована, Где-то, не знаю, в детстве мы эту песню изучали и так далее. И когда я произношу вот эти 2 слова, следует вероятность выпадения слова елочка высокая, она в нашем российском культурном контексте. Но если бы я то же самое сказал на англоязычную аудиторию, скорее всего, они бы вообще не поняли, что там такое в лесу родилось.

01:01:02

И вот примерно здесь точно так же работает, когда мы с вами пишем промпт, который убирает весь лишний контекст, то есть он убирает вот это облако смыслов, из которого нейросеть выхватывает какие-то слова, он его делает ближайшим. Потому что, если представить себе, попробовать вообразить довольно сложно, но давайте воображение включим. Вот это векторное пространство, в котором работает нейросеть. Все слова лежат вот так вот, в определенном связанном друг с другом такими разными веревочками, назовем это так. Представим себе визуализация, мастер визуализация.

01:01:41

И когда мы с вами какое-то слово говорим, протягивается вероятность вот эта к следующему слову, а потом к следующему, а потом к следующему, а потом к следующему. И когда ИИ выбирает, какое слово взять из облака вот этих смыслов и слов, оно берет исходя из того контекста, который вы ему задали. Поэтому, если вы дадите подробный контекст, то облако слов станет меньше, и оно будет тем, которое вам нужно. Но если вы этого не сделаете, то произойдет следующее: нейросеть будет выбирать что-то наиболее стандартное, то есть какие-то очень частые паттерны, которые встречаются у большинства людей, в большинстве случаев. И будет стандартное письмо, стандартный анализ данных и так далее.

01:02:25

Поэтому ваш промпт это по сути такой вот способ отсечь все лишнее из ответа нейросети. Формула эффективного промпта звучит следующим образом: здесь 5 компонентов, каждый из них увеличивает качество результата. Первый это роль кто отвечает. Ты бизнес-аналитик с десятилетним стажем, ты руководитель проекта, ты менеджер по ключевым клиентам. То есть, роль задает тон, глубину и экспертизу ответа.

01:02:50

Второй это контекст. Какая ситуация? То есть, чем больше деталей, тем точнее результат. Клиент, его параметры, история отношений, продукты и так

далее. Третий это задача, что конкретно нужно сделать.

01:03:03 Не просто проанализируй, а извлеки 5 ключевых условий и оцени 3 риска, например. Конкретика это четвертый формат. В каком виде вам нужен результат? Например, таблица, список, текст на 200 слов, может быть JSON нужен для системы, если вы разработчик и так далее. И пятый это ограничение.

01:03:26 Чего не делать нейросети? Не использовать юридический жаргон, максимум 300 слов, только по данным из документа, не додумывая и так далее. В общем, эта формула по сути универсальная, ее можно использовать как чек-лист каждый раз, когда вы пишете промт, но даже если вы заполните 2 из 5 контекст и задачу, этого тоже будет, скорее всего, достаточно, потому что это самые главные элементы. Роль вторична, ее не нужно каждый раз использовать в каждом промпте. Форматы и ограничения, в общем, наверное, тоже не обязательно.

01:03:56 Но контекст самое важное, что сейчас есть. И иногда я для контекста вообще загружаю целые письма, документы, какую-то предысторию вопроса, чтобы нейросеть понимала то же самое, что я понимаю вокруг этой задачи. Чем больше контекста вы дадите, тем лучше нейросеть сделает для вас задачу. AI для анализа документов. Теперь если немножко посмотреть, давайте эту формулу попробуем применить к конкретной задаче.

01:04:23 Анализ документов Это, наверное, 1 из самых частых универсальных сценариев использования искусственного интеллекта. Практически для любой профессии все мы работаем с документами разного вида: юристы с контрактами, HR с резюме и так далее, маркетологи с исследованиями, отчетами. И что умеет ИИ с документами очень хорошо делать? Он умеет извлекать ключевую информацию из PDF, вордов и так далее. Он умеет очень хорошо саморизовать, то есть превращать одностраничный документ в двухстраничную.

01:04:54 Допустим, 100 страниц вы можете превратить в 2 страницы саммари. Находить конкретные данные сроки, суммы, условия и так далее. Сравнить несколько документов между собой. Это, кстати, тоже бывает очень полезная штука. И структурировать хаотичные данные, когда у вас какой-то беспорядочный текст, а вам нужна, например, таблица.

01:05:17 То есть вы можете посмотреть на пример промта внизу слайда, и он как раз по нашей формуле настроен роль, контекст, задача, формат. Можете попробовать тоже прямо сейчас взять любой рабочий документ контракт, техническое задание, регламент, загрузить его в q1ot, например, в Alpina GPT

и попробовать вот такой результат. Ты юрист с опытом анализа контрактов, контекст договор с поставщиком, документ приложен, извлеки ключевое условие, найди риски, формат таблица, плюс красные флаги, плюс рекомендации. И можете посмотреть, что из этого получится. Далее, AI для анализа коммуникаций это тоже популярное применение Деловые письма.

01:06:00

Здесь AI отлично помогает с персонализацией, как адаптировать шаблон под конкретного клиента с подбором тона от формального до дружелюбного, со структурой вступление, суть, призыв к действию и так далее, с вычиткой, подготовкой ко встречам. То есть ИИ может подготовить вам материалы, сформулировать ключевые тезисы, предугадать вопросы, которые может клиент задать и составить план действий после встречи. Можно генерировать клиентские материалы, приветственные пакеты для новых клиентов и многое-многое другое. В общем, здесь ИИ тоже экономит огромное количество времени на всей этой рутине. И цепочка рассуждений еще 1 техника промтинга, которую мы можем с вами тоже забрать в работу.

01:06:43

Суть простая вы просите ИИ думать пошагово. Слева пример без цепочки рассуждений. То есть вы спрашиваете Оцени эффективность рекламной кампании. И вам отвечают: Эффективность высокая. Всё.

01:06:56

Откуда цифра? Вообще непонятно. Правильно ли она рассчитана. Нам нужно вручную перепроверять всё это. А справа с этой цепочкой рассуждений.

01:07:04

То есть вы говорите: Давай пошагово. Шаг первый оцени охват и бюджет. Шаг второй рассчитай конверсию по этапам. Шаг третий сравни бенчмарки в отрасли и так далее. Шаг четвертый, шаг пятый.

01:07:16

И вы получаете прозрачный ответ, где видно каждый шаг этого рассуждения и если где-то есть ошибка вы тоже видите на каком шаге или этапе она произошла. Вся shot learning это обучение на примерах. Идея по сути очень простая вы даете нейросети 2-3 примера того, что хотите получить, а она копирует этот паттерн для новых данных. Почему это работает? Потому что примеры это, по сути, самый эффективный способ объяснить формат.

01:07:43

Вместо того, чтобы все это описывать словами, как должен выглядеть результат, вы просто говорите вот так, сделай так же по аналогии. И это подходит для большого количества разных задач везде, где вам нужно повторить: тексты для постов в соцсети, документ, например, таблица определенного формата и так далее. Zero Shot Learning без примеров, по сути, только инструкция. Очень короткая, для простых задач, вообще идеально подходит. И разница между ViewShot, да, ViewShot вы показываете

примеры, например, вот жалоба, вот запрос, вот благодарность, теперь классифицирую новое обращение.

01:08:19 А Zero Shot вы просто говорите классифицирую обращение клиента по категориям жалобы, запрос информации, заявка на продукт, благотворительность и т. Д. То есть без примеров, только инструкция. Когда и что использовать? Zeroshot работает отлично для простых однозадачных задач, да, sorry за тавтологию.

01:08:38 То есть когда вам нужно, например, классифицировать обращение по четырем категориям это простая задача и исправиться без всяких примеров. Когда вам нужен пример, определенная таблица, определенная структура лучше использовать Viewshot. Ролевой промптинг. В целом мы частично его отчасти разобрали, то есть это назначение роли. Интуитивно очень простая техника.

01:08:59 Я только добавлю, что роль по сути не задает новых данных и новых знаний модели. Модель и так всеми этими знаниями обладает. Но, как вы помните, вероятность того, из каких слов модель выбирает. Если мы добавили роль в наш промпт, мы сразу же обозначили, из какого облака смыслов нейросеть будет выхватывать слова. И, по сути, роль просто меняет стилистику ответа и его глубину, но не добавляет новых знаний.

01:09:29 Нейросеть и так знала все, что вы у нее могли бы спросить, и до момента, как вы выбрали роль. Я выбираю роль обычно по следующему принципу. Я задаю себе вопрос: кто из людей лучше всего справился бы с этой информацией? И в зависимости от ответа, который я себе получаю, я выбираю роль для нейросети. И, наверное, самая важная техника контекст инжиниринг.

01:09:53 Пятая техника промптинга для бизнес-задач лучше всего подходит. То есть, чем полнее контекст вы даете для нейросети, тем качественнее получается результат. Контекст это не просто описание задачи. Здесь 4 компонента. То есть есть некий фон, бэкграунд: кто клиент, какая у него история отношений, например, с банком, если мы про банковские продукты, какие продукты он уже использует, какой оборот, отрасль, размер и так далее.

01:10:18 Дальше ограничения. Допустим, внутренние стандарты компании, бюджетные ограничения, KPI и прочее. Формат это таблица, текст, чек-лист, презентация, на сколько слов, для кого она может быть для руководства или для клиента и так далее. И как выглядит хороший результат? То есть покажите образец.

- 01:10:38** Контекст здесь будет очень важен, потому что, наверное, он важнее любой техники. Вы можете не использовать цепочку рассуждений, не задавать роль, давать примеров, но если вы дали полный контекст, результат все равно будет хорошим. Вот и, собственно, здесь пример слабого контекста. Напиши письмо клиенту, шаблонное письмо без деталей. Полный контекст вот здесь на примере дан тоже, да, клиент, ооо Технопром, оборот, текущий контракт, контакт, цель, дедлайн и так далее.
- 01:11:11** И безопасность. Прежде чем мы с вами перейдем в практические кейсы, последняя часть это безопасность. Очень важно, то есть что вы можете в целом использовать в общем таком поле общие вопросы по вашей сфере, шаблоны писем, документов, публичная информация о рынке, обезличенные примеры и так далее. Учебные задачи, конечно же, все. Что не стоит?
- 01:11:34** Слово нельзя, наверное, надо зачеркнуть, но это просто рекомендация. Персональные данные ФИО, паспорт, ИНН и так далее. Почему? Потому что, в принципе, все, что в облаке, грубо говоря, вы храните, теоретически можно получить к этому доступу. Ну, допустим, ваш аккаунт где-то даже в оригинальном чате GPT в теории можно взломать, получив доступ к содержимому этого аккаунта.
- 01:11:56** И все ваши переписки, которые вы там вели, станут тоже доступны. Alpina GPT немножко другая история, потому что мы все-таки в российском контуре находимся. Данные всех пользователей хранятся в России, не на зарубежных серверах. Соответственно, мы используем зарубежные нейросети только, чтобы подумать, а сами все чаты, пользовательские данные, то есть мы соблюдаем 152 ФЗ и прочее, поэтому в целом Alpina GPT чуть более защищенная, чем если использовать напрямую какой-то из этих инструментов, которые вы могли бы использовать. Ну и, конечно, там все, что касается коммерческой тайны, паролей, финансовых показателей и так далее лучше тоже не использовать.
- 01:12:37** Ну и давайте с вами в практические кейсы занырем. Первый практический кейс, который мы с вами попробуем анализ документов. 1 из самых востребованных сценариев. Сейчас я перейду в платформу Alpina GPT и покажу вам, как это будет выглядеть. Так, сейчас буквально секундочку.
- 01:12:55** Как раз я сделаю демонстрацию браузера. Вы пока тоже как раз можете открыть какой-то из инструментов, ну или Alpina GPT в частности, потому что принципе я сейчас на нем буду все показывать. И мы с вами переходим Так, секундочку, немножко подзависла. Так, шерю. Так, так так вот она.

- 01:13:33** Так, все за переместил. Так, вот, смотрите, небольшой навигатор по тому, что вы можете здесь видеть. Тоже, чтобы вам было чуть-чуть попроще. Так, сейчас я тут все перемещу куда-нибудь, это у меня тут чат, камеры, все остальное. Значит, смотрите, интерфейс очень простой, у вас есть поле ввода, куда вы можете ввести ваш промпт.
- 01:14:00** Здесь вы можете выбрать модели, какие модели вы тоже можете посмотреть. Каждая из моделей имеет свое описание, понятие стоимости и так далее. Вот вы также можете выбрать здесь компактный вид или не компактный. То есть можно выбрать вообще чтобы все модели были видны. Вот или модели с описанием или без описания.
- 01:14:22** Вот и здесь есть инструменты поиска. Он у вас по умолчанию скорее всего будет включенным, но если вы захотите вы можете его отключить. Инструменты чата. Здесь вы можете оставить включенным мышление, то есть, тогда вы будете видеть, как модель рассуждает. И тоже можно здесь выбирать, какое мышление будет среднее, высокое или низкое.
- 01:14:44** В зависимости от того, насколько вам важно, чтобы модель глубоко рассуждала. Также вы можете выбирать ассистентов. Про ассистентов и агентов у нас с вами будет отдельный материал, отдельный вебинар. Но, если что, имейте в виду, что вот здесь их можно создавать и выбирать готовых. Вот, то есть, есть какое-то количество преднастроенных, то вы можете создать также своего здесь.
- 01:15:06** Вот в таком случае, если вы нажимаете скрепочку, вы загрузите какой-то документ. Здесь сразу перечислены форматы, которые поддерживаются и, в зависимости от того, какую модель вы выбрали, количество форматов может быть разным. То есть, там какие-то модели больше форматов поддерживают, какие-то модели меньше. И, соответственно, вы можете посмотреть, если нужного вам формата с конкретной моделью нет, можете попробовать выбрать другую и там вполне может оказаться уже формат, который вам нужен. И следующая часть.
- 01:15:38** Ну, это понятно голосовой ввод. Я думаю здесь все очевидно. Есть еще 1 кнопочка интересная это шаблоны. То есть, если вы не знаете, как написать промпт, вы просто можете выбрать шаблоны, заранее предустановленные. Их здесь очень много по разным областям: обучение, реклама, продажи, HR.
- 01:15:57** Ну, в общем, тут их 300 с лишним штук. Все шаблоны составлены профессионально, в том числе многие шаблоны составлены автором книги 333 промта на каждый день. Это бестселлер от Альпины, написал его Дамир Халибов. Соответственно, вот некоторые из этих промтов они тоже здесь

находятся. Поэтому вы можете смотреть на эти промты как на некую эталонную штуку, которую можно дальше переиспользовать себе как паттерн.

01:16:26

Также, если вы обратите внимание, есть возможность чаты добавлять в избранное. То есть, вы можете добавить звездочку и чат у вас сохранится в избранном. И еще из интересного. У нас недавно появилась функция проектов. То есть, вы можете создавать проект.

01:16:41

В этот проект вы можете добавить различные файлы, поддерживаются до 100 файлов по 30 мегабайт каждый. Можно загрузить PDF и Docx и прочее. Эти файлы создадут контекст вашего проекта и, соответственно, внутри этого проекта вы сможете обсуждать какую-то тему. Не знаю, давайте житейскую какую-нибудь возьмем, представим, что мы создаем проект Здоровье, загрузили туда результаты всяких анализов, всего остального, что у нас есть Мы можем с нейросетью общаться по итогам этих анализов, насколько условно мы можем нейросеть спрашивать второе-третье мнение. Допустим может быть это кому-то покажется радикальным примером, но есть много кейсов, где нейросети давали очень полезные рекомендации.

01:17:26

Естественно, это не заменяет медицинского специалиста, но просто 1 из примеров. Или вы, например, загружаете все ваши HR-документы какие-то, практики, которые касаются внутренних регламентов компании, документы, и можете общаться в проекте onboarding, например, о том, как правильно в этой компании составляются заявления на отпуск, допустим, или что-то еще. То есть тут под вашу задачу вы можете создавать проект и еще важная деталь все чаты этого проекта обогащают общий контекст. Это значит, что когда вы в 1 проекте находитесь и обсуждаете, каждый ваш чат в этом проекте будет обогащать контекст предыдущих. То есть вы потом новый чат начнете, и в новом чате он будет и будет понимать, какие были обсуждения в предыдущих чатах этого проекта.

01:18:20

То есть, такой общий контекст наращивается. Вот, это очень тоже удобно для особенно работы с документами. И сейчас, помимо того, что здесь есть еще разные другие разделы, тут я сейчас ликбез такой очень в сокращенном формате даю, есть perplexity, есть генерация изображений. Perplexity это поисковик. То есть, выбирая его, вы можете соответственно использовать инструмент для поиска в интернете.

01:18:46

Также у вас есть генерация изображений. Здесь есть midjourney, flux и многое-многое другое. Сейчас в ближайшее время доедет нанобанана. Есть переводчик, профессиональный переводчик дипл для документов. То есть, вы можете переводить документы целиком с сохранением форматирования.

- 01:19:03** Есть аватары. Аватары для создания какого-то образовательного контента или рекламного контента. Очень хорошая штука. И есть транскрибация. То есть вы можете превратить аудиофайл или видеофайл в текст и в обратную сторону можете сделать озвучку.
- 01:19:21** Если вам нужно какой-то текст озвучить, выбираете голос в Eleven Labs и, соответственно, делаете озвучку и получаете нужную голосовую дорожку. Это вот такой очень простой краткий экскурс по функционалу. Мы с вами сейчас в основном в режиме генерации текста находимся, в основном разделе И сейчас загрузим сюда документ. Сейчас буквально секундочку я его тоже подгружу. Вот.
- 01:19:49** И собственно мы с вами попробуем ввести промт. Так, ты юрист-аналитик. Вот он у нас. Ты юрист-аналитик. Так, и подгружаем документ.
- 01:20:10** Так, сейчас секундочку. Вот сюда. Все, добавили. Все, очень просто все. Ты юрист-аналитик с 10-летним стажем для анализа договора.
- 01:20:26** Давайте вам тоже этот промс скину. И теперь отправляем наш запрос в нейросеть. Так, что-то пошло не так. Сейчас новый чат. Так, добавить.
- 01:20:48** Промпт. Так вот здесь вы можете увидеть рассуждение модели это то, что нейросеть сначала обдумывает до того, как выдаст вам ответ. Видите, это весь процесс ее рассуждений. И теперь он выдает мне анализ. Юридический анализ договора поставки.
- 01:21:20** И вот мы можем увидеть, как он составил эту таблицу: красные флаги, высокий размер предоплаты, жесткое ограничение ответственности поставщика, отсутствие условий расторжения. Естественно, важная деталь это все не заменяет юриста. Лучше всего, чтобы вообще это делал юрист изначально, но в некоторых случаях у вас может не быть времени, не быть юриста в конце концов под руками, а решение принять нужно, и это лучше, чем вообще отсутствие какого бы то ни было инструмента или проверки. Я могу сказать по себе: я в очень большом количестве реальных жизненных кейсов применял ИИ, в том числе для очень сложных вопросов налоговых, бухгалтерских, юридических, и зачастую получал результаты. Есть много кейсов, не только мои, где люди сэкономили 1000000 за счет того, что применяли искусственный интеллект, со страховщиками были кейсы.
- 01:22:23** В общем, если интересно, тоже можете погуглить применение и в юридических аспектах. Но были и негативные, когда и придумывал какие-то несуществующие законы по прецедентному праву, а потом люди получали проблемы, адвокаты, в частности, репутационные риски. Поэтому, в общем,

это палка о 2 концах. Просто нужно иметь ввиду, что есть риски, но как бы вы сами взвешиваете на чаше весов. Что перевешивает?

01:22:49 Риск или плюсы, которые вы благодаря этому имеете. Давайте пойдём дальше. Следующий кейс конкурентный анализ. Сейчас я тоже дам вам промт. Так, конкурентный анализ И вот такой промт отправляю вам.

01:23:12 Теперь смотрите, мы сейчас с вами попробуем. С 1 стороны мы можем попробовать инструмент perplexity, а можем попробовать Клода. Давайте мы сейчас в Клоде с инструментом web поиска сделаем то же самое. Значит, здесь мы можем написать в вашей отрасли. Какой отрасли?

01:23:27 Допустим AI или genAI. Сравнительная таблица. GenAI агрегатор. Агрегатор на рынке РФ. И контекст, задача, вроде бы все, кампания сравнительная таблица, да все окей.

01:23:55 Отправляем. И параллельно можно еще пойти в перплексии, но сейчас мы дождемся пока он подумает над нашей задачей, посмотрим на результат. Когда он ищет, естественно это какое-то время занимает. Вот. И вот пошли разные всякие инструменты.

01:24:40 Вот вы соответственно можете по своей как бы отрасли это все дело применить Соответственно, тоже получите хороший результат. В целом, очень хорошо он нашел. И дальше, если вот, не дожидаясь, сейчас я сделаю дублиаж вкладки, чтобы пока он там продолжает. Вот смотрите, переплекси делаем то же самое. А здесь, например, не знаю, возьмем какой-нибудь другой другой пример, какую отрасль бы мы хотели.

01:25:19 Например, книга издательская отрасль, издательская. Допустим, издательская на рынке РФ. И можем посмотреть, как работает perplexity. Здесь тоже есть отдельные настройки: глубокое исследование, академический режим, но нам не нужно сейчас такие. Тут можно также выбирать настройки, температура, все эти инструменты, их можно тоже в знаке опросика посмотреть, на что влияют эти параметры.

01:26:08 Вот так. Читаю. Ирина пишет: подскажите, пожалуйста, какие ей порекомендуете воспользоваться, если есть старые 200-летней давности планы и фасады зданий, но их необходимо привести в приличное состояние? Хороший, очень интересный вопрос. Смотрите, я бы что делал?

01:26:27 Я бы взял, наверное, NanoBanana. У нас, к сожалению, здесь ее нету, но она есть внутри экосистемы Google, а вы, соответственно, к ней можете получить доступ через, например, версию Gemini Enterprise. Сейчас я ее вам покажу есть просто Google Джемния а есть Энтерпрайз сейчас я ее Энтерпрайз

Джеминэй вот она значит так вот здесь есть сразу же. То есть, вы с обычным Google профилем можете туда зарегистрироваться. 30 дней вам дадут, как минимум, на использование.

01:27:04 И дальше внутри нее есть модель нанобанана. Значит, сейчас мы зайдем в обычный Gemini, и вот здесь у вас есть возможность не только проанализировать изображение, но и создать изображение. И тогда вы можете взять загрузить в какой-то план здания, фасада и прочего, и просите отреставрировать, или там условно визуально улучшить. Только вот важная деталь промтинг изображений он про конечное состояние объекта. То есть, там вам нужно не процесс указать, а вот итоговое состояние.

01:27:46 Что значит приличное состояние, вот ваша интерпретация. Это значит, чтобы он стал из черно-белого цветным или что-то подобное. И тогда вы получите уже более качественный результат, который вы ожидали. То есть, в принципе, он восстанавливает исходное фото, прямо 1 в 1 может делать, если вы правильно промт ведете. Еще 1 лайфхак, пока мы с вами не закончили сегодняшний день это метапромтинг.

01:28:13 Что это такое? Так, пока он тут перлексит у нас ищет. Здесь он у меня закончил. Что такое метапромтинг? Вы можете обратиться к ИИ как учителю и сказать ему: Я учусь, я Я учусь промтить, мне нужна твоя помощь в составлении промта для задачи.

01:28:44 И смотрите, берем задачу, которую написала Ирина, например, для задачи. Составь мне промт для Nano nanobanana pro. Все, окей. И смотрите, что он сейчас делает. То есть, нейросеть реально хорошо умеет составлять промты.

01:29:20 Поэтому, если вы не знаете, как составить правильный промт, вообще не беда. Я сам часто обращаюсь к нейросети для того, чтобы промты их формировать. То, что действительно они пишут это подробнее и быстрее, чем я бы сам написал. Вот промты для восстановления старых архитектурных планов и фасадов. Базовый вариант на английском, так как AI модели лучше понимают английский.

01:29:42 Да, абсолютно верно. И вот здесь прямо написано конкретно по то есть, можно перевести, посмотреть, но в целом все вроде бы правильно. Загрузите изображение вместе промтом, Итерируйте, если результат не идеален, уточняйте. Нано-бананопром может точно захватывать, переносить художественные стили и так далее. Вот пример.

01:30:07 Давайте я сейчас вам скину, возможно, Ирина вам будет полезна, и попробуйте это сделать. Пример промта. Вот есть более детальный вариант.

Тоже сейчас его скину. Соответственно, каждый из вас может сделать абсолютно то же самое в этом же инструменте.

01:30:22 Поэтому, если вы не знаете, как правильно промпт написать, то вот это такой лайфхак, который можно использовать дальше в своей практике постоянно. Деловая переписка. Сейчас еще быстренько 1 кейс покажу интересный. Давайте сейчас я вам промпт тоже скину. Промпт-маркетинг-аналитика выдает ошибку при формировании ответа.

01:30:46 Вы попробуйте новый чат создать или, может быть, модель попробуйте поменять. Иногда бывает такое, что модели перегружены. Особенно у Клода сейчас проблемы, потому что все резко начали на Клода переходить, поэтому как это безжалостно его используют. Поэтому, может быть, вам поможет. То есть, мы пользуемся с вами одинаковым инструментом.

01:31:05 Я в том же самом Alpine GPT, поэтому, если у меня работает, значит, возможно, это какая-то разовая ошибка может быть произошла. Так, отправляем, смотрим, что выдает нейронка. Вот, или, может быть, если у вас ошибка, посмотрите, может быть, дело в балансе. Иногда бывает, что там токены не зачислились, и тогда служба поддержки тоже поможет. То есть, если у вас так происходит со всеми нейросетями, и все не отвечают, значит, возможно, проблема в этом, и тогда служба поддержки вам поможет.

01:31:41 Вот смотрите, что я сейчас сделал. Я ввёл промт, роль менеджер по работе с клиентами, контекст, задача, формат, ограничения. Я хотел бы получить деловое письмо, деловое письмо, ответ. Коммерческое предложение, точнее, если быть вообще точным, расширение сотрудничества, решение, комплексное решение для вашего оборудования. Уважаемая Елена Викторовна бла-бла-бла.

01:32:07 В общем, здесь некоторое такое вот очень официальный текст. А теперь давайте представим себе, что нам нужно тот же самый текст, но для, не знаю, стартапа с молодым составом фаундеров. И я ему говорю: Перепиши текст для стартапа с молодым составом фаундеров. То есть, главная ошибка, которую часто бывает допускают, и это, кстати, исследование компании Антропик о том, как используют искусственный интеллект на большой выборке. И что они сказали в этом исследовании?

01:32:47 Они говорят, что большинство людей совершают какую ошибку? Они пишут вопрос, получают ответ и на этом все. А мы должны с ИИ общаться как с собеседником. То есть ИИ нам дал какой-то результат, мы говорим: отправь вот это, ищи, измени вот это, и вот так еще подправь. И в процессе с пятого, с шестого, например, раза, бывает, вы получаете нужный вам результат.

- 01:33:14** Как и с человеком. Если вы даете задачу человеку, какому-то сотруднику, то вероятность, что он не с первого раза выполнит идеальным образом, она присутствует. И если он приносит вам не то, вы же его не закрываете на этом. Хотя такие тоже кейсы бывают, но вы не говорите все, вы уволены. Вы продолжаете, объясняете ему, что он сделал не так, где он допустил ошибку.
- 01:33:39** Вот здесь то же самое, Если ИИ вам выдал что-то не то, что вы хотели, просто поправьте его. И вот здесь мы что видим с вами? Смотрите, насколько поменялся тон, если до этого была тема решения сотрудничества, комплексные решения для вашего оборудования. Уважаемая Елена Викторовна, благодарю вас и команду Технопром за 5 лет успешной совместной работы и вот так далее. Смотрите, как поменялся текст.
- 01:34:04** Апгрейд сотрудничества, сервис и расходники в 1 пакете. Привет, Елена! 5 лет вместе это уже не просто партнерство, а настоящий бизнес союз. Спасибо, что выбрали нас тогда и продолжаете работать с нами сейчас. Для нас Технопром 1 из тех клиентов, ради которых хочется делать больше.
- 01:34:20** Собственно, к делу предлагаем прокачать наше сотрудничество. Видите, какие слова используют уже: прокачать, сервис под ключ, через регулярное ТО, быстрое диагностика, ремонт с выездом 24 часа и так далее. В общем, здесь все очень по-другому написано, но суть та же самая. То есть, предложение точно такое же, а текст совершенно другой. Поэтому вот таким образом можно работать с текстом.
- 01:34:47** Что еще? Есть еще инструменты: Perplexity, NotebookLM, Gemini. Вот сейчас я вам быстренько покажу, как это внутри Perplexity устроено. Учитывая, что Perplexity это инструмент, который вы можете использовать, в том числе в бесплатном режиме, есть еще такая фишка у них называется пространство. На самом деле, это аналог наших проектов в Alpina GPT, но, например, вы можете тоже сюда загружать какие-то документы целой пачкой.
- 01:35:20** По сути, очень похоже: загружаете файлы или вставляете текст, добавляете какие-то ссылки и потом в этом пространстве спрашиваете все, что угодно у ИИ на тему, которую вы хотите, на тему тех документов, которые вы вставили. И то же самое есть у NotebookLM. Очень интересная особенность, как у них это выглядит, это то, что они еще сразу делают презентации. Сейчас посмотрим, я вам покажу, как выглядит ноутбук LM. Да что ж такое, никак не хочет.
- 01:36:06** О, наконец-то. А можно после ответа ИИ удалить загруженные файлы? Да, тут, смотрите, все зависит от того, где вы работаете, в какой платформе, то есть в облаке какой-нибудь ChatGPT. Они у вас остаются, по-моему, в

файлах и надо проверять удалились ли они оттуда. Потому что там есть такая особенность, что они там, даже если вы чат удалили, то файлы могут сохраняться и надо там отдельно их выискивать.

01:36:36

В Alpina gpt мы в целом файлы не храним, мы их превращаем сами файлы нигде не сохраняются, они превращаются в вектора. То есть мы их сейчас попытаюсь таким не очень заумным языком. Есть такое понятие чанки это кусочки текста, которые и очень хорошо понимает. То есть она легко может находить что-то в векторе, и каждый вектор содержит в себе чанк с определенным набором данных внутри. И когда вы что-то ищете, у вас по вектору находится определенный чанк, и он вам говорит, условно, Найдено на странице такой-то, но у него страницы нет.

01:37:17

Иишка не оперирует страницами, она оперирует вот этими чанками. Это в свою очередь означает, что ваш документ или неважно какой-то файл, он разбивается на фрагменты и уже ищет по этим фрагментам. Сам документ фактически он как бы в шредер попал такой. Представим себе, что есть виртуальный шредер, в который вы документ целиком засовываете, а вылезают оттуда маленькие такие вот пластиночки, кусочки. И вот в этом ковыряется в этом мусоре, такая вот аналогия вам, в этом мусоре ковыряется очень легко.

01:37:51

Ему там находить все, что угодно легко. Но человеку по этим чанкам потом что-то составить практически самому невозможно. Поэтому, если вы удалили проект, соответственно, у вас, по крайней мере из нашей, из альпины GPT, у вас удаляется и все содержимое чанков и всего остального, а документы больше нигде не хранятся. Вот так, это я постарался объяснить, насколько это получилось. И вот смотрите, пример у меня вот здесь просто сразу готовое покажу.

01:38:17

Вот я загружал конкретный это вот ноутбук LM инструмент от Гугла, И я здесь загружал специально смоделированные данные, и по ним ИИ-шка видите смотрите как искала, например, что важного по сотрудникам из документов, а он мне выдает здесь конкретные ссылки. И вот эта штука, она, видите, вам показывает сразу же, где все это ищется, в каком кусочке документа это было. В Alpina GPT примерно так же. Там тоже вам показывают ссылочки на информацию откуда это было. Единственное, что мы пока не умеем в Alpina GPT, как это умеет Gemini это вот просто супер фишка у них.

01:39:00

У них даже подкаст есть, представляете? Они могут по вашим данным сделать подкаст. Что это вообще такое? Сейчас я вам покажу, включу. Послушайте.

- 01:39:12** Так. Коротко по текучести кадров за первый квартал 2024 года и совершенно монументальные акты потерь, списания и задержек транспорта уже за первый квартал 2026 года. Да, массив данных впечатляющий. И если сложить все эти цифры вместе, они как бы перестают быть просто скучной корпоративной отчетностью. Точно!
- 01:39:34** Они превращаются в очень детальную карту того, как на самом деле работает вся эта огромная индустрия доставки. Наша миссия в этом детальном исследовании источников это буквально заглянуть за кулисы. За кулисы той самой заветной кнопки купить, которую мы все так любим нажимать в приложениях. Да-да, мы нажимаем и забываем. Вот именно!
- 01:39:58** А мы хотим вскрыть и понять саму анатомию этого механизма, увидеть, как банальная человеческая усталость, какие-то жёсткие корпоративные алгоритмы и чисто физические ограничения реального мира сталкиваются лбами. И превращаются в гигантские финансовые убытки. Да, это крайне важное погружение в тему, потому что обычно логистика кажется нам какой-то магией. Ну типа кликнул в телефоне и на следующий день веселый курьер уже у твоей двери. А наш сегодняшний анализ покажет реальную цену каждой такой доставленной посылки.
- 01:40:35** Мы разберем просто по винтикам как малейший сбой на 1 единственном складском стеллаже. В общем, я думаю, вы поняли 22 минуты вот этого подкаста автоматически сгенерированы на разные голоса, то есть это все автоматом сгенерировано, это вообще к людям не имеет никакого отношения, и они сгенерированы на основе вот этого всего датасета, то есть я загрузил сюда кучу документов, опять же, это все не настоящие документы, но я их специально смоделировал для целей обучения. И потом внутри вот он сгенерировал первое подкаст, второе он мне сгенерировал вот такую схему майнд мэп я могу ее развернуть и значит тут что-то посмотреть да по каждому блоку и вот он меня значит сразу сейчас перекинет на этот контекст потом он здесь сгенерировал мне презентацию, вот такую прессу, смотрите, автоматом. То есть, он взял весь контекст всех документов и под ним собрал вот такую красивую прессу. Вот он, раз слайд, 2 слайд, всё отрисовал сам.
- 01:41:42** Представьте, сколько у дизайнера вообще заняло бы вот это всё нарисовать. Потом он мне что сгенерировал, значит отчет большой красивый тоже, вот такой вот one pager одностраничник, и карточки, наверное, самое простое из всего этого, это просто некая такая, как бывает, знаете, в сторисах во всяких показывают. Итого, ноутбук LM прекрасный инструмент для того, чтобы глубоко погружаться в какие-то данные. Похожим образом можно использовать функционал в Alpina GPT как некий формат того, что вы можете

делать с проектами, с данными и так далее. Вот и последнее, что я хотел бы вам показать из такой вот прям прикладной работы с файлами.

01:42:38 Так, сейчас я найду, где тут у меня. Это анализ таблиц. Вот я сейчас простенькую табличку загружу, потому что это тоже у всех наверняка бывает. Покажу, как это работает. Вот и для исследования сейчас работа с таблицами.

01:42:58 Пример такого промта. Сразу же здесь тоже вставляю табличку. Табличка, добавляем ее и вставляем сюда промт. То есть сразу погрузил его роль Ты бизнес-аналитик, контекст файл CSV с данными о продажах, задача и так далее. Смотрим, что нейронка вырастет.

01:43:27 Смотрите, вот это очень важно, то, что сейчас происходит на экране. Он, видите, что делает? Сначала извлеку данные, то есть он заглядывает в файл и пишет как бы код к нему, обращение. Видите, вот этот код, который он пишет, и извлек всего 100 записей. Дальше мы видим, как он под капотом вообще приходит к этим заключениям, к этим расчетам.

01:43:57 То есть современные модели рассуждающие, они уже не просто вмещают к себе это в контекст, а они сразу под капотом производят расчет. Смотрите, как длинно он думает об этом. То есть, он не просто взял и выдал вам ответ, тот же сгаллюцинировал, а он прям рассуждает подробно очень, обстоятельно обдумывает наш вопрос. Конечно, он может ошибиться, как и человек, но вероятность этого с каждой новой моделью, с каждой новой версией все меньше и меньше. Особенно, если вы используете специализированные инструменты, которые используются для расчетов.

01:44:38 Мы некоторыми инструментами еще поделимся на следующем занятии, когда будет про агентов, ассистентов и вот это все. Там мы как раз рассмотрим некоторых готовых агентов, в том числе для работы с excel, для создания сайтов, для многого другого. То есть агентов уже есть целая куча разных предустановленных. То есть, можно использовать готовые, а можно, в общем-то, и своих создавать уже даже. Ну, смотрите, сколько он уже думает все, думает и думает.

01:45:09 Какой, какой умный Клод. Так, сейчас дождемся, пока он тут все это рассчитает. То есть, это все еще не его ответ, это только то, что он под капотом делает. Поэтому иногда, когда вы даете эту модель задание и ждете ее ответ, иногда это может раздражать, что так долго-то вообще, где уже ответ. А в принципе все это время он вот так считал, чтобы выдать нам ответ.

- 01:45:36** И, конечно, сразу скажу бесплатные и платные модели отличаются по качеству рассуждений, по качеству ответа, которые они выдают. Вот он посчитал ключевые показатели: аналитический отчет по продажам, топ-5 продуктов по выручке, рейтинг менеджеров и многое-многое другое. Видите, какие таблички все тут красиво нарисовал, все посчитал, все это дальше можно просто брать и упаковывать в некий отчет. Сейчас мы дождемся и как раз я вам покажу, что можно с этим делать дальше. Вот и смотрите, дальше у Вас есть кнопочка Скачать.
- 01:46:28** Представим, что Вам нужно было этот отчет кому-то дальше передать. Вы просто выбираете PDF, Markdown или TXT, впоследствии там добавится у нас что-то еще. Выбираете название файла, нажимаете скачать и у вас скачивается файл. Дальше тут же могу его открыть, вам показать. Вот такой файл у вас получился.
- 01:46:50** Вы можете его кому-нибудь переслать, как некий отчет из вашей системы. Ну и отдельный еще лайфхак можно делиться чатами, но это работает только внутри корпоративного контура. То есть, если вы сотрудник 1 компании, внутри 1 компании, то вы можете делиться друг с другом. У сотрудника, вашего коллеги будет появляться копия вашего чата. Так, собственно, смотрите, у нас с вами еще остается небольшая часть, связанная с генерацией изображений и прочим.
- 01:47:19** Мы с вами поступим следующим образом, чтобы все успеть и как бы ничего не пропустить, мы с вами подробно это разберем тогда в следующий раз, когда у нас с вами будет в конце занятия, посвященные агентам, ассистентам и прочим, мы как раз с вами разберем специальные инструменты, в том числе для генерации видеоизображений. Сегодня мы с вами разобрали 6 ключевых техник промтинга, разобрали истории моделей, разобрались в том, как модели вообще работают, куда вообще идёт в принципе искусственный интеллект и что нам в ближайшее время ждать, и попробовали несколько кейсов, разобрали конкурентный анализ, деловое письмо, исследование perplexity и некоторые другие в коде и анализ данных в Excel. Что делать сегодня? Попробовать промпты на реальной задаче. В первую очередь, конечно же.
- 01:48:16** На этой неделе обязательно у вас будет много еще материалов помимо сегодняшнего воркшопа. В конце недели вы сможете настроить себе несколько ассистентов под свои задачи и повторять их постоянно уже. Этот месяц может быть вы поделитесь кейсами, какими-то интересными моментами с командой, с вашими коллегами, с кем-то ещё, с кем вы вместе. Ну и дальше вы постоянно можете делиться друг с другом, то есть у вас

может формироваться комьюнити. Конечно же, вам рекомендую, в том числе внутри нашего интенсива делиться друг с другом, что у вас получается.

01:48:48

Пишите в чатик, общайтесь, потому что это не только про знания от спикеров или каких-то экспертов курса, но это и про то, как вы применяете и сами каждый день. То есть, вы, может быть, то, чем вы поделитесь друг с другом, может быть не менее ценным, чем то, что мы даем в рамках интенсива. Поэтому это прекрасный нетворкинг для того, чтобы дальше взаимодействовать друг с другом: Опробовал AI на задаче, получил результат, сохранил промпт, поделился с командой. И вот так формируется кейсбук, который прекрасно работает для того, чтобы еще больше вдохновляться примерами своих коллег. Итого работа без AI долго, рутинно, ошибки и прочее, значит, CI быстро, креативно и точно.

01:49:34

Ну такое, Денис, что именно такое? Что делиться друг с другом не хочется. Ну, иногда может быть и не нужно, если это какой-то суперсекретный ингредиент, вы нашли, можно, наверное, и не делиться. Но я сторонник того, что делиться друг с другом все же это хорошо. Итого, готов поотвечать на вопросы.

01:49:54

Если вы мне их позадаете, специально тут 10 минуток заложил на QA сессию, поэтому, пожалуйста, делитесь, может быть, просто фидбэком, может быть, какие-то вопросы у вас есть. Некоторые заранее вопросы я тут написал. Вот нам. Так. Да, здорово, что можно скачать док.

01:50:19

А у Deep бесплатного DeepSI только копировать можно результат, файл скачать нельзя. Приходится потом править его. Да, все так. Потому что мы тоже старались опираться на те недостатки, которые мы видели в целом по разным инструментам. Ну и теперь вот стараемся сделать инструмент, который будет всем удобен и закрывать разные проблемы, которые есть в других.

01:50:46

Я выжидаю небольшую паузу. У У нас, конечно же, сегодня только отправная точка. Впереди дальше будет еще много материалов и в записи, и в текстовом виде, и еще у нас с вами будет GO workshop, так что мы сегодня у нас только начало, чтобы вас не перегрузить слишком сильно. Да, спасибо. Резидент скинул ссылочку на мой канал.

01:51:10

Я, кстати, тоже регулярно пишу про нейросети, так что подписывайтесь, ставьте лайки. Буду рад вас там видеть. Мы там готовим яичницу, используя хорошие рецепты. Так, Сергей. Вижу, Сергей поднял руку.

- 01:51:25** Сергей. Здрасьте. Мне может быть просто будет проще голосом. Да, конечно. Спасибо за то, что были сегодня с камерой.
- 01:51:38** Я устал на вас много сегодня. Сегодня я много с вами общался. Хорошо. Взаимно. У меня, наверное, такой вопрос, по крайней мере, по своему опыту, я, наверное, с учетом GPT уже, ну, не знаю, года полтора, наверное, уже так или иначе что-то там делаю, веду какие-то беседы и какие-то рабочие задачи, но там по-разному.
- 01:52:00** С чем сталкивался? Наверное, с тем, что иногда бывает такое, что ну, бывает такая штука, когда, ну по крайней мере так было, когда диалог вообще уходит как бы в некий цикл, когда пытаешься что-то сделать, он делает это, а ты говоришь нет по-другому, он делает то же самое, ты говорит вот опять то же самое и все. И после этого он иногда даже сам говорит: лучше сделай новый диалог, тут похоже все пошло лесом. Такое было раньше, сейчас это было, по крайней мере, в генерации изображений. Довольно часто встречалось, когда он прямо входил в некий цикл и просто падал сюда.
- 01:52:45** Это вот, наверное, первая штука, не то, что как бороться, но пока у меня единственный вариант это просто открывать новый диалог, потому что там память будет новая. Абсолютно верно, просто контекст обнуляется и это помогает, потому что в конкретном чате модель просто вот она в определенном контексте находится, опять же она уже что-то взяла из определенных паттернов и из этого ей выйти уже тяжело. А если вы новый чат начинаете, то это спасает. Вопрос про токены, про вот это вот как раз про расходование и так далее. То есть, он как бы такой, то есть есть токены, которые `twalpin gpt`, то есть я так понимаю, что это как бы расход, он идет как бы на каждый запрос или это на чат.
- 01:53:33** Да, понял. Смотрите, вообще как выглядит, когда вы пользуетесь какой-то подпиской, и кажется, вроде бы она безлимитная за 20 долларов. На самом деле разработчики умные ребята, они вас ограничивают в другом. То есть они вам не дают превысить эти 20 долларов в виде затрат на токены. И по средней медиане большинство людей не превышает 5 долларов, кто пользуется.
- 01:53:56** Но если вы, например, слишком много запросов сделаете в единицу времени, они вам скажут: а теперь 3 часа, подожди следующего запроса. Когда вы пользуетесь какими-то инструментами, где этого ограничения нет, в теории у вас и затраты не лимитированы ничем, кроме вашего кошелька, который привязан. Тогда токенизация начинает иметь значение, потому что если мы просто пользуемся, то нас ограничат автоматически, и мы все равно 20 долларов превысим. Но в среднем Alpina GPT, мы знаем по среднему

потреблению, пользователи не тратят больше 300-500 рублей в месяц на все модели. То есть, потратить больше это надо прямо постараться, если вы особенно текстовые инструменты используете.

01:54:41

И имеет значение весь чат. Что я имею ввиду? Это значит, что у вас контекстное окно общее, то есть когда я захожу в Alpina GPT, я могу в рамках 1 чата переключать разные модели. Мне модель Cloud могла дать ответ, я тут же переключился и продолжаю диалог в GPT в этом же контексте этого же чата. Как мы этого достигаем?

01:55:03

Возникает вопрос. А мы держим контекст чата у себя, он не в модели находится, мы сохраняем все чаты на своей стороне, и каждый ваш следующий запрос передает весь контекст предыдущий в модель. То есть он ваш вопрос передает и все, что было выше в этом чате. Это с 1 стороны плюс, потому что вы всегда в модель подгружаете актуальный контекст каждый ваш запрос. С другой стороны минус, потому что вы таким образом, если чат очень длинный, вы как бы каждый запрос, весь предыдущий чат тоже закидываете.

01:55:36

Это токены съедает. Поэтому рекомендация такая, особенно в Alpina GPT или в любых других агрегаторах периодически, ну, то есть, если вопрос исчерпался, вы его решили, начинайте новый чат. Не пытайтесь в 1 чате обсудить абсолютно все темы, потому что это просто станет дороже каждый ваш запрос. Вот. Такая история.

01:55:59

Про контекстное окно, которое в токенах. Я сейчас не имею вопрос именно тарификации, а хранение вот этого контекста, то есть это получается как раз внутри 1 чата, да то есть я не знаю там чат GPT 5-2 или когда-то там O3 O4 да вот оно продвинутое рассуждение и у него контекст большой. Если мы берем Fast Basic, он будет как бы поменьше, он будет побыстрее давать результат, но, соответственно, он просто меньше хранит вот это вот данных. Да, да. Сейчас почти все модели приближаются к контекстному окну в 1000000 токенов Джеминей держит 1000000 токенов, Клод держит 1000000 токенов последний, gpt-шка тоже к 1000000 подошла.

01:56:46

Соответственно, в принципе, 1000000 токенов это прям очень много. Потратить 1000000 токенов в 1 окне это надо книжку целиком, например, загрузить. Единственная проблема как только вы подходите к 1000000 токенов, у вас модель начинает терять, может потерять начало или середину чата, то есть она начинает плавать уже. Это чем-то похоже, как если вы, например, книгу прочитали, вы, если там еще, не знаю, последнюю страницу можете почти целиком воспроизвести как цитату, то что там было в начале книги, ну вы примерно расскажете, но вероятность ошибиться довольно

высокая. И вот здесь примерно такая же история, поэтому лучше не перегружать контекст, не дожидаться его исчерпания.

01:57:29

Ну да, я примерно Но тут вот лично у меня вот в этой задаче, у меня тогда вопрос а что тогда делать? То есть, например, там у меня есть творческая задача делать там рилсы в определенной структуре, да, вот в определенном как бы сценарии, да, то есть там несколько как бы диалогов я добился того, вот чтобы вот как бы логика была вот такая и такой О, вот это классный пример, дальше делай вот так. И у меня есть диалог, в котором он понимает как бы чего я от него хочу, да я говорю давай сделаем на вот эту тему, он такой ну вот такие, я говорю вот берем вот этот, да мы его там как-то допилили, сделали все получилось. Но постепенно этот диалог он как бы расширяется и он начинает или где-то повторяться или сбивается или еще что-то. И мне с 1 стороны нужно чтобы он понимал чего я хочу, да то есть как бы чтобы у него эта память была, но она как будто он уже начинает в ней вот двигаться.

01:58:26

И вот как тогда получается сделать новый диалог и перетащить ему какой-то пример, и как бы чтобы там это пошло, ну грубо говоря из более пустого контекста или все-таки продолжать в этом, но каждый раз его там периодически корректировать, потому что он будет уже там петлять. Ну, потому что это там, ну не знаю, у меня уже там сценариев может быть 10, да. И я планирую продолжать их делать, а оно же будет расти. Да, я понял. Смотрите, лучше всего для этого подходит функционал проектов, который кстати есть и в Клоде, и в чате GPT, поскольку каждый чат он догружает общий контекст и, соответственно, вы можете начинать новый, у вас контекстное окно обнулилось, но общий контекст проекта остался.

01:59:14

И поскольку в проект можно загружать также файлы, вы можете здесь, например, в виде варда ваши предыдущие посты, сценарии или что-то еще. И поскольку вот здесь происходит векторизация, то есть документы, которые вы загружаете, переводятся в векторную базу, так называемые раги еще есть такое понятие. И вот создается маленький раг. Про раги мы поговорим еще отдельно, когда до ассистентов, агентов доберемся. Короче, маленькая векторная база специальная, по которой ИИшки очень легко искать.

01:59:47

Поэтому вы можете сюда все-все догрузить, и каждый ваш новый вопрос ИИ, он не будет перегружать контекст модели, но модель будет знать, что у нее есть база, туда нужно сходить и там что-то посмотреть, и исходя из этого дать вам ответ. И этот же функционал есть не только в Alpido GPT, он есть и в оригинальном, вот давайте я покажу, в чате GPT. Да-да-да. Я просто думал, ну, окей, там да, есть у проектов может быть общая память, может быть обособленная внутри проекта, но мне почему-то казалось, что он эту память

проекта как будто заполняет сам, то есть чем больше я этих чатов веду, он сам что-то туда как бы вкидывает. Можете дать ему.

02:00:29

Можно это сделать и как бы более адресно просто как бы. Хорошо. Да, то есть вы как бы берете, когда новый проект создаете, допустим там сейчас 123 и здесь видите то же самое есть чаты, а есть источники. То есть, вы загружаете источники, привязываете всю информацию сюда и делаете более глубокий контекст вашего проекта. То же самое есть у Клода, если зайти в Клод, там тоже есть функционал проектов.

02:01:04

Есть также функционал проектов у вот этой платформы Perplexity, там тоже вот это все есть. И здесь, сейчас я покажу проекты. Соответственно, все просто так же здесь создаете какое-то имя и сюда подгружаете. Вот то же самое инструкция есть, есть файлы, которые сформируют общий датасет, и вот примерно таким образом. То есть, для таких вот задач как раз специально этот функционал был создан.

02:01:34

Но есть еще ассистенты, как функционал, но ассистенты, каждый чат с ассистентом будет изолирован от предыдущих. Отличаются они от проектов только тем, что контекст чатов в проекте общий. Как-то так. Хорошо. Ладно, ну вот ассистентов я пока не касался, но вот здесь да, это хорошее решение, спасибо.

02:01:55

И последний вопрос, он такой наверное более абстрактный, но как бы тем не менее. Я 1 раз на это натолкнулся и был немало удивлен. Это как это проблема 10 10, как бы, с часами, то есть когда, как это, поскольку большинство фотографий часов, наручных или каких-то еще, имеют стрелки на 10 часов 10 минут, потому что они просто красиво выглядят, да. Нейросеть обучилась, что все стрелочные часы выглядят вот так и попросить его как бы нарисовать там 4 часа 45, а не в состоянии, потому что никогда не видела такого, да. А логически ему это ну вот объяснить тоже практически ну бывает сложно.

02:02:34

То есть я, наверное, я больше всего попадал это, когда я пытался отредактировать какие-то картинки, что я говорю, вот здесь вот эта линия не туда, вот это вот убери, там что угодно может происходить, но только не то, о чем бросим. Да, но это как бы проблемы генеративных нейросетей, они же все-таки генерируют, ну как сказать, все время определенные случайные числа или случайные пиксели. Понятно, что они натренированы на большом объеме данных, например, есть еще 1 такой же кейс, и тяжело нарисовать беззубую акулу, потому что беззубых акул вроде как не бывает. Он все-таки синтезирует все то, что подобрал в себя, И поэтому с такими кейсами иногда можно столкнуться. В общем, и не волшебная такая все-таки пиллюля, он

некоторые вещи не может, но какие-то специализированные инструменты, иногда надстройки, они помогают убраться то, что изначально модель не умеет.

02:03:30

Под капотом какой-нибудь хитрый промт будет, выдаваться этот инструмент будет как за уникальное ноу-хау, но под капотом может быть обычная gpt, просто специальным образом докрученная. Поэтому, вот когда мы до агентов, ассистентов доберемся, там как раз мы разберемся, как можно было бы сделать какой-то инструмент, который какие-то специальные задачи решает, например, часы рисует со стрелками. Хорошо. Спасибо. Спасибо большое за вопросы.

02:04:04

Может быть еще кто-то тоже хотел бы поделиться или какой-то вопрос задать в лучших ресторанах до последнего клиента работаем. Так, окей, хорошо. Ну, тогда, если вопросов нет, то спасибо вам большое сегодня за активное участие, за готовность вечером в понедельник постигать всю эту тему. Вот будем тогда с вами на связи. И если вопросы появляются пишите в чат, мы с командой там тоже будем за всем этим приглядывать.

02:04:46

Будем стараться вам помочь максимально за эти 4 дня погрузиться глубоко настолько, насколько это за 4 дня вообще возможно. Спасибо большое. Тогда на этом сегодня завершаем.