

ПАВЛОВСКИЙ ЕВГЕНИЙ НИКОЛАЕВИЧ



Заведующий лабораторией аналитики потоковых данных и машинного обучения, НИУ НГУ, кандидат технических наук.

E-mail: e.pavlovskiy@g.nsu.ru

«МНОЙ В БОЛЬШЕЙ СТЕПЕНИ ДВИГАЛ ИНТЕРЕС, СВЯЗАННЫЙ С ТЕМ, КАК УСТРОЕН ЧЕЛОВЕК...»

Экспертное интервью¹

DOI: 10.32691/2410-0935-2023-18-178-195

Смирнов С. А.: Евгений Николаевич Павловский, заведующий лабораторией аналитики потоковых данных и машинного обучения.

Павловский Е. Н.: Да

Смирнов С. А.: Евгений, я пытаюсь у разных экспертов выяснить их экспертную позицию относительно вопроса, который, мне кажется, ещё слабо осмыслен. Он связан не столько с техническими, инженерными задачами в области разработки программ искусственного интеллекта (ИИ), сколько с антропологическими основаниями этих разработок.

Когда разработчик занимается этим, он, наверное, закладывает в основание, интуитивно или как-то осознанно, свои представления о том, что такое интеллект, что такое интеллектуальная деятельность, что такое мышление, наверное, и исходит в этом смысле, разумеется, из понимания, а как там у человека с этими вещами. Как мы понимаем, эти человеческие качества и способности вольно или невольно он начинает проецировать на модель ИИ или копировать какую-то часть человеческой способности, и закладывает это в техническое устройство, так или иначе.

В этом смысле разработка ИИ сильно зависит от того, как мы понимаем человека. Меня это интересует, поскольку я не разработчик, извините, я философ, хожу и расспрашиваю.

Но прежде, чем это обсуждать, я бы всё-таки хотел спросить Вас о следующем. Зачем человеку ИИ? Ему сильно это нужно? Нужно только для замены рутины?

¹ Разговор записан 3 марта 2023 года. Интервью провел С. А. Смирнов, в. н. с. ИФПР СО РАН. Интервью проведено в рамках гранта Российского научного фонда (проект № 21-18-00103, <https://rscf.ru/project/21-18-00103/>).

Это давно уже было, вроде бы сейчас задачи стоят другие, не только для замены рутины. Зачем ИИ человеку именно сейчас? Не тогда, когда изобрели экскаватор, чтобы лучше копать было.

Павловский Е. Н.: Я думаю, что есть несколько оснований для этого. Например, то, что я в себе нахожу, когда у меня была тяга к программированию, желание запустить, написав некую программу, посмотреть, как это работает, как компьютер выполняет то, что я ему напишу. Это было первое впечатление где-то в 6-7 классе.

Смирнов С. А.: Чисто задача программиста?

Павловский Е. Н.: Да, задача программиста: как некая машина интерпретирует мои команды и выполняет нечто, выполнит ли ровно так, как я ей скажу.

Смирнов С. А.: Да.

Павловский Е. Н.: Это было первое желание. Потом где-то через пять лет я написал рассказ, в 11 классе, о том, как я прожил жизнь всю, и первая часть жизни у меня была такая, что я придумую ИИ, который будет помогать человеку, назову его Френд, он будет моим другом.

Смирнов С. А.: Я прошу прощения, а иметь друга среди живых людей, не было такой задачи? То есть нужен технический друг, искусственный друг?

Павловский Е. Н.: Да, я пытаюсь понять, откуда эта мысль упала.

Смирнов С. А.: Откуда, да? Кстати, Вы, написав рассказ, программировали свою жизнь? Так?

Павловский Е. Н.: Конечно, да. Я к ИИ через 10 лет только вернулся, а пока я чистой математикой и экономикой занимался. Почему упала мысль, я даже не помню, я фантастику особо не читал, но сейчас я понимаю, что есть эта линия в культуре, где нас фантастикой потчевали и развивали эти мысли.

Смирнов С. А.: Да, сначала обрабатывали, а потом пытались внедрить.

Павловский Е. Н.: Да, «Россумские Универсальные Роботы» К. Чапека. Наверное, «Терминатор» я смотрел, что-то в общем поле культуры как-то эта мысль сквозила. Но мной в большей степени двигал интерес понять, как устроен человек.

Смирнов С. А.: Всё-таки, как устроен человек?

Павловский Е. Н.: Да, и для этого один из способов состоит в том, чтобы это промоделировать на компьютере.

Смирнов С. А.: Да.

Павловский Е. Н.: Мы моделируем собственно логику. Я заходил из математики, из логики, чтобы понять, как человек логически мыслит, а компьютер для меня – это способ проверить, быстро смоделировать, так это или нет, модель работает или не работает. Я и сейчас так подхожу к ИИ. Можно ли какие-то реакции человека смоделировать, причем наиболее приближенно? Это исследовательский интерес, он лишь отчасти обоснован какой-то практической пользой.

Смирнов С. А.: Ага, здесь сугубо исследовательский интерес.

Павловский Е. Н.: Да, и это же я нахожу в работах Марвина Минского, например, это один из отцов-основателей ИИ. Он не ставит вопрос, зачем нам нужны машины, а ставит вопрос, как их сделать. Он приверженец теории биохевиоризма, то есть мы не обязательно должны воспроизвести так, как у нас внутри, у людей, устроено.

Смирнов С. А.: Да, мы внутри мало чего знаем.

Павловский Е. Н.: Да. Если мы видим функции и процессы, видим, что результат один и тот же получается, то, наверное, мы и делаем правильно. Он предлагает много-много подходов, которые, в частности, развивались в теории ИИ.

Смирнов С. А.: Если так, если брать этот ход, я пока допускаю разные варианты, если брать этот ход, сугубо бихевиористский, который имеет дело вообще-то с поведенческой моделью, в этом смысле он не заходит внутрь человека и не ковыряется в мозгах, не выясняет когнитивные внутренние дела и не пытается понять природу мышления, он делает поведенческую модель, описывает поведение человека, потом его моделирует, переводит фактически в программу, математическую модель, и смотрит, как она функционирует. Я правильно понял?

Павловский Е. Н.: Да.

Смирнов С. А.: Но ведь вроде бы интеллектуальная деятельность человека не ухватывается в категориях поведенческой модели. А интеллект – не поведенческая функция. А если когда-то однажды при усложнении задач поведенческая модель может дать сбой, потому что интеллект и мышление гораздо больше, чем поведение, вообще не сводится к поведенческим реакциям и функциям? У человека по всяком случае. То есть рождается другая задача – строить модель, в основании которой не поведенческая модель, не поведение, а, так скажем, модель опосредования?

Например. Берём ручку. Ручкой впервые мы стали писать в школе. Мы же от рождения не умеем писать? Это очевидно.

Павловский Е. Н.: Наверное, да.

Смирнов С. А.: Да, давно это было, мы когда-то порог переходим, потом это переходит в автоматизм. Так вот. Когда человек пишет, то каким органом он пишет? Он пишет впервые букву А, высунув язык, пытается в 1 классе это сделать, и пишет «мама, папа». У него очень сложный формируется орган, орган письма, которого от рождения у него нет. Он пишет букву с помощью ручки, ручку ещё надо освоить, овладеть этим действием, мама и папа ему не могут передать от рождения способность владения ручкой, не только ручкой, любым предметом в принципе, это осваивается.

Дальше это все формируется. Надо, чтобы кисть и рука научились этим управлять, отсюда через многотысячные повторения формируется локальная зона в мозге, новые нейронные связи формируются, тогда получается новый функциональный орган. Он формируется.

Павловский Е. Н.: Да.

Смирнов С. А.: Это овладение действием с предметом и овладение знаком, собственно, и есть акт мышления. Заметьте, это не поведенческая функция, не поведение. А как формируется высшая способность у человека? Можно ли по этому поводу строить искусственную модель такого типа, мыследеятельностного, когда формируются функциональные органы? В принципе вся же жизнь человека из этого состоит.

Павловский Е. Н.: Я думаю, что как раз скачок в развитии ИИ произошёл, когда...

Смирнов С. А.: Такой переход к таким моделям уже произошёл?

Павловский Е. Н.: Да, произошёл. Произошел как раз такой скачок, когда мы отказались от внутренней структуры, перестали её понимать.

Смирнов С. А.: Отказались ковыряться и понимать, что там внутри?

Павловский Е. Н.: Отказались понимать, да, а сказали, что результат должен быть таким, и давайте проверим, хороший или плохой.

Смирнов С. А.: От результата пошли.

Павловский Е. Н.: Например, ввели понятие ошибки, а дальше – градиент ошибки, поставили задачу, кто при принятии решения наибольший вклад в ошибку сделал, давайте того и исправлять. Эта система, честно, есть в нейросетях, да и много, где исправление делается относительно ошибки, а то, как мы ошибку задаем, – это, собственно, и есть то, как мы управляем всем процессом обучения.

Смирнов С. А.: То есть современные нейросети ведут себя примерно так же, как человек ведёт себя при овладении новой ситуацией, предметом, знаком, речью? Это ход туда?

Павловский Е. Н.: Я думаю, что ход туда.

Смирнов С. А.: То есть она действует уже не путём перебора операций с бешеной скоростью, миллиарды операций в секунду, как раньше она обыграла Каспарова в шахматы?

Павловский Е. Н.: Да, раньше действовала перебором.

Смирнов С. А.: Раньше так было, она просто быстрее его посчитала?

Павловский Е. Н.: Да, быстрее перебирала, а сейчас в Go нельзя перебирать.

Смирнов С. А.: В Go уже не переберёшь?

Павловский Е. Н.: Да, там работают со стратегиями.

Смирнов С. А.: Там надо стратегию выбирать?

Павловский Е. Н.: В Go сетка училась сама с собой, играла и нарабатывала разные стратегии, поскольку она сама себе проигрывала.

Смирнов С. А.: То есть она фактически осваивала опыт. Можно считать, что понятие накопления опыта применимо к этой сетке?

Павловский Е. Н.: Да, это так и есть, но это сильно грубая модель. Нейросети – это сильно грубая модель того, что у человека в мозгах происходит, однако она подобна, обратная связь имеется. У человека есть обратная связь, есть нейроны или транзиттеры, которые подавляют, усиливают обратную связь, и в нейронках так же. Причем в нейронках есть алгоритмы, которые позволяют раскидать ошибку на каждую часть элементов. Но это опять же подсмотрено у человека, как мозги его работают.

Смирнов С. А.: Подсмотрено?

Павловский Е. Н.: Да.

Смирнов С. А.: Как это подсматривали?

Павловский Е. Н.: На патологиях, вскрытиях всяких, шинковали мозги у мышек и так далее.

Смирнов С. А.: Тогда здесь следующий шаг. Это действие человека, оно же не сводится просто к активной функции мозга. Мы-то, философы, считаем, что мыслит не мозг. В этом смысле можно, сколько угодно шинковать мозги у крыс...

Павловский Е. Н.: Я тоже так думаю.

Смирнов С. А.: Но мыслит гораздо более сложный орган, вот этот самый функциональный орган, человек мыслит с помощью новых функциональных органов, которые выстраиваются у него в деятельности. В этом смысле, конечно, в зависимости от того, какая у него деятельность по жизни, разумеется, у него морфология мозга формируется, мозг нейропластичен, есть такое понятие, мозг обладает нейропластичностью.

Один интересный англичанин замерил у лондонских таксистов гиппокамп, им надо было сдавать очень сложный тест на пространственную ориентацию по большому городу, они полгода его учили, сдавали тест. Оказывается, гиппокамп у них гораздо более сложный и развитый, чем у тех, кто не сдавал, не тренировался. Понятно, сформировался, потому что была практика по работе, особенно с пространством, пространственным мышлением.

Вопрос и заключается в том, что мозг поэтому стал таким, потому что таксист занимался этим делом. У скрипача будет всё по-другому. Но мыслит не мозг, он формируется исходя из разных многообразных действий человека. Потом, как обратным бумерангом, он тебе выдает то, чему научился, потому что я его научил, он мне возвращает то, что умеет.

В этом смысле вопрос – что моделировать? Да, сетка накапливает опыт, но она же как накапливает, как мы это выясним? Можно сколько угодно дальше шинковать мозги у мышки, но, может быть, шинковать надо не мозги, а моделировать по-другому? Другие модели делать? Есть ли такиеходы?

Павловский Е. Н.: Да, есть такие. Просто внешние условия задачи у нас такие: каковы входные и выходные условия, а дальше внутри, пожалуйста, можно какие-то модели совсем не человеческие использовать, какие-то графовые модели, тензорные модели, что угодно может быть внутри.

Смирнов С. А.: В этом смысле здесь же ключевая проблема в том, умеем ли мы моделировать какие угодно модели, не в том, что у человека что-то с мозгами, и это не техническая задача, это задача разработчика...

Павловский Е. Н.: Я думаю, здесь как раз проходит водораздел. Ученые, это я про себя, наверное, вдохновляются идеями познать мир, а платят им не за это, а за то, чтобы они конкретную пользу для каких-то процессов приволили. Это переключение где-то произошло, и сейчас в принципе нам особо не дают возможности копаться в закономерностях природы.

Смирнов С. А.: Не дают, не до этого, РФФИ на это деньги не даёт?

Павловский Е. Н.: Да, а РФФИ говорит: «Давай так, чтобы тебя цитировали», это там тоже можно.

Смирнов С. А.: В РФФИ всё-таки фундаменталка-то есть.

Павловский Е. Н.: Есть, да. Именно такая постановка – нам не важно, что внутри происходит, главное, что критерий общий выполняется, решается задача, она близка к экономической задаче, а там сейчас основным фактором затраты являются, не доходы почему-то, а затраты.

Затраты надо снижать, но люди – это очень большие затраты. Робота одного сделал за всех сразу, то есть в моменте затраты выше, но потом затраты меньше на его обслуживание, чем на человека. Поэтому логично заменять человека, не логично, а как сказать, экономически выгодно.

Смирнов С. А.: Экономически выгодно.

Павловский Е. Н.: Поэтому эта система воронкой засасывает все исследования в эту сторону, как мне кажется. Почему произошел скачок в ИИ? Да, на-

учились довольно качественно решать некоторые задачи, а потом от человека отвлечлись и начали его заменять, прикрываясь картинкой, что, мол, мы заменяем рутинную деятельность, а на самом деле заменяем интеллектуальную деятельность.

Смирнов С. А.: Вот именно, делаем то, что дешевле получается. А интеллектуальная деятельность самая дорогая.

Павловский Е. Н.: Видимо. Ошибки в ней наиболее страшны.

Смирнов С. А.: Они там наиболее чреватые.

Павловский Е. Н.: Да. Но, допустим, диспетчер такси, девушка, она больше ошибается, чем диспетчер автоматический.

Смирнов С. А.: Да.

Павловский Е. Н.: Он строит маршрут оптимальный, у него всё это есть. Поэтому заменяется человек. Или в случае с водителем. Ставится задача – должен быть самоуправляемый автомобиль. Нас пичкают этими идеями сейчас, всё должно быть самоуправляемое, автомобили, но даже на дороге, которая на карьере в Евсино, где-то там есть карьер, там машины ездят, один и тот же маршрут каждый день, каждый день одно и то же.

Смирнов С. А.: Да.

Павловский Е. Н.: Там и надо садить робота.

Смирнов С. А.: Да.

Павловский Е. Н.: Но там нет робота.

Смирнов С. А.: Не сядят, нет робота. Наверное, это уже от управляющих товарищей, от менеджеров зависит. Зачем им робот? Не знаю, наверное, считают, может быть, что тетю дешевле посадить.

Павловский Е. Н.: Мне кажется, это скоро произойдет, но пока так, водители на такси – люди.

Смирнов С. А.: Это люди, да.

Павловский Е. Н.: Разносчики еды тоже люди с рюкзаками.

Смирнов С. А.: Но, вспоминая тогда всё-таки тренд про беспилотники, про то, что где-то надо заменять человека, это же не просто тренд технический или функциональный, за этим стоит более глубинная вещь: в принципе уход человека из активной роли субъекта намечается в разных видах.

Но одно дело, когда врач-диагност допускает, что лучше, когда надо делать тысячу снимков, и картинку, и пленку лучше посчитает ИИ, нежели он, у него глаз замылился или вчера водки перепил и так далее, настроение плохое, особенно когда надо много, это одно дело, когда ты привлекаешь к сложному и многократному действию умного помощника. Хорошо, умный помощник, понятно, в здравоохранении.

Павловский Е. Н.: У меня тоже к этому куча вопросов.

Смирнов С. А.: Но дело в том, что за этим стоит в принципе тренд все большего отказа человека от субъектной роли, от роли активного игрока. Это тоже имеет определённые пределы, всё-таки ему же это нравится. Одно дело, пиццу развезить, другое дело, даже уже принимать решения можно не обязательно самому.

То есть сам тренд отказа от субъектности, от субъектной позиции, он ведь тоже фактически означает отказ от себя, потому человек сотни лет допускал, что он субъект и он принимает решения, а сейчас уже пришли к тому, что в принципе и субъектом быть не обязательно.

Хорошо, я не обсуждаю, хорошо это или плохо, я обсуждаю тренд замены, что взамен, а что человек себе оставляет. Или всё-таки субъектность будет трансформироваться сама, или будет перестраиваться этот интерфейс между человеком и его умным помощником, когда сам помощник всё более и более набирает себе умных функций, точнее, ему просто их передают, а себе человек оставляет что?

Павловский Е. Н.: Кажется, это происходит ввиду таких факторов, что мысленное усилие для некоторых людей гораздо более сложное и трудозатратное дело, чем для других.

Смирнов С. А.: Для миллионов людей.

Павловский Е. Н.: Да, для миллионов. Поэтому, когда нас освобождают от мысленного усилия выбора, допустим, какой фильм посмотреть, можно друзей спросить, можно книжку прочитать и посмотреть экранизацию, то есть какая-то мыследеятельность, развернутая во времени с затратами энергии, а тут мы заходим на «Кинопоиск» или куда-нибудь, за очень маленькое усилие, за одну кнопку мы получили нужное впечатление.

Смирнов С. А.: За кнопочное усилие, да.

Павловский Е. Н.: Мозг получил нужное подкрепление, фильм понравился, значит система хорошо сработала.

Смирнов С. А.: Да.

Павловский Е. Н.: Аналогично врач принял решение на основе нейросети, он сам принял решение, но сказал: «Что? У этой нейросети опыта больше, она больше видела, чем я». Да, подтвердил, и это был правильный диагноз, правильно вылечил.

Смирнов С. А.: Да, если подкрепил, потом снова и снова.

Павловский Е. Н.: Подкрепил, мозг привыкает к тому, что поступают эти подкрепления, а значит, зачем ему тратить много усилий на это? Есть система, она работает. Также устроены экзокортекс да неокортекс, внутри у человека устроено по минимуму энергии, то есть сначала идёт какая-то затрата энергии.

Смирнов С. А.: Да, дальше хочется экономить.

Павловский Е. Н.: Дальше надо экономить, иначе смысл?

Смирнов С. А.: Да.

Павловский Е. Н.: Аналогично с экзокортексом, весь наш аутсорсинг интеллектуальных функций тоже так устроен. Вначале мы что-то тратим, а потом, пожалуйста, работай, не трать своих усилий. Здесь, мне кажется, дальше дело привычки. Почему не произошел этот скачок на интеллектуальную деятельность? Потому что она либо многим недоступна, проще сидеть потреблять контент, чем его создавать.

Смирнов С. А.: Этот тренд нарастает, как по объёму, так и по качеству, по интенсивности, всё больше и больше таких усилий мы отдаём в аутсорс, и это опускается всё ниже и ниже по возрасту, когда родители приучают ребенка сразу к этой кнопке. Гораздо проще просто взять и получить, подкрепить и снова получить.

Павловский Е. Н.: Мне кажется, это прямо мутация происходит в общественном организме.

Смирнов С. А.: Происходит метаморфоз, во-первых, социальный.

Павловский Е. Н.: Идёт умирание, идёт ускоренное умирание.

Смирнов С. А.: В этом смысле умирание, есть такой тренд?

Павловский Е. Н.: Да, 100%.

Смирнов С. А.: До каких пределов?

Павловский Е. Н.: Пока не выйдем.

Смирнов С. А.: Какие-то избранные единицы ещё будут что-то пытаться помыслить, а многие миллионы отказываются от этого дела. Так?

Павловский Е. Н.: Да, есть сила привычки, была возможность перейти на сферу интеллектуальной деятельности и управлять машинами, но для этого нужно постоянно перехватывать управление на себя, постоянно.

Смирнов С. А.: Да, быть на контроле.

Павловский Е. Н.: Кто-то выключается из этой системы в принципе, уходит в лес, заводит натуральное хозяйство, выключает себя полностью из этой системы, тот может воспроизводить те навыки, которые помогают выжить в той среде. Сейчас же получается так – если ты перестал управлять, то тобой начинает система управлять.

Смирнов С. А.: Да, у тебя атрофируется способ управлять другим, ты отдаешься машине.

Павловский Е. Н.: Да.

Смирнов С. А.: Это всё больше...

Павловский Е. Н.: Тут вопрос интереса этой системы. Я в принципе малую вероятность допускаю, что какая-то воля у ИИ, как у технической системы, на больших объёмах может сама зародиться в принципе. Но скорее это просто другие люди используют.

Смирнов С. А.: Конечно, его используют.

Павловский Е. Н.: Чтобы какие-то свои задачи решить.

Смирнов С. А.: Да, есть разговоры уже в эту сторону, что якобы можно говорить даже о возможной моральной машине, об этике ИИ, о том, что эта машина будет совершать в том числе и моральные выборы, принимать этические решения, есть и такие разговоры. Но в это трудно поверить, потому что это не рациональное действие, это действие связано уже с этическими категориями.

То есть можно, конечно, моделировать при желании всё, назвать это так, что, мол, я смоделировал совесть, и это обзывать действием машины, что она действует по совести. Но всё это будет метафорой, а на самом деле она просто действует по своей логике, никакого отношения к совести не имеет.

Павловский Е. Н.: Нет, я думаю, тут разговор про свою логику у машины в данном случае слишком расплывчатое понятие, и то, что мы называем логикой, в традиционном смысле, это некие правила мышления. Выясняется при этом, что человек не по правилам мыслит иногда, а принимает решения ещё реже по правилам.

Смирнов С. А.: Да, именно.

Павловский Е. Н.: Сложно формализовать логику. Также и в технических системах, когда они становятся очень большими, и нет ни одного человека, который до конца понимает, как она работает. Может возникать иллюзия, что там происходят процессы мышления.

Смирнов С. А.: Иллюзия, что машина принимает решения.

Павловский Е. Н.: И решения, и эта иллюзия в принципе может быть такой, что вообще есть же люди, которые, как роботы ведут себя.

Смирнов С. А.: Да, конечно, поведенческие роботы чистой воды.

Павловский Е. Н.: Тогда мы переходим эту границу.

Смирнов С. А.: Функциональную.

Павловский Е. Н.: Где разница?

Смирнов С. А.: Нет.

Павловский Е. Н.: У этой системы опыта больше, чем у этого конкретного человека.

Смирнов С. А.: Да, да, в этом всё и дело.

Павловский Е. Н.: Всё, мы перешли границу эту.

Смирнов С. А.: Мы уже перешли?

Павловский Е. Н.: Да.

Смирнов С. А.: То есть, реально перешли, тест Тьюринга решён, это уже давно понятно.

Павловский Е. Н.: Chat GPT – пример, что уже перешли.

Смирнов С. А.: Этот, который написал работу для студента?

Павловский Е. Н.: Да.

Смирнов С. А.: Что там особенного? Именно потому, что экзаменаторы не увидели, не поняли, что эту работу написала нейросеть? Элементарно можно было задать пару вопросов студенту, а они почему-то не поняли.

Павловский Е. Н.: Да, а вопрос, если они при этом не контролируют, что студент на эти вопросы тоже через этот чат отвечает.

Смирнов С. А.: Вот именно, я вообще удивился, что значит они не поняли и узнали об этом, потому что он сам признался.

Павловский Е. Н.: Во-первых, не все научные руководители все тексты до конца читают у студентов.

Смирнов С. А.: В общем халява была у них на экзамене, строго говоря, это же можно всегда проверить.

Павловский Е. Н.: Да. Во-вторых, на самом деле сетка складно излагает, и там может человек найти мысль. То есть настолько высокий уровень генерации конструкций текстовых, что возникает иллюзия, что мысль туда заложена.

Смирнов С. А.: Иллюзия всё-таки, это выглядит правдоподобно, даже якобы, как будто осмысленно, но это ведь не про это.

Павловский Е. Н.: Это проблема метода, установления нашего метода, которым мы устанавливаем протокол, что произошёл процесс мышления.

Смирнов С. А.: Да, мы допускаем, это допущение.

Павловский Е. Н.: Это человек написал статью или статья тоже пришла в редакцию, но мы не знаем, её послал какой-то e-mail нам, а кто там за этим, человек или робот, кто написал текст, это вопрос.

Смирнов С. А.: Это опять вопрос к нам самим, почему мы, распознавая полученный текст, приписываем ему осмысленность? Это же решается в прямом контакте, когда я читаю текст, текст может быть порожден либо человеком, кстати, не всегда умным, либо машиной, хорошо.

Павловский Е. Н.: Есть Гарри Лайон Олди, это два автора, их псевдоним. Кто из них пишет? Стругацкие, тоже два автора.

Смирнов С. А.: Да.

Павловский Е. Н.: По-моему, кто-то один там писал, второй обсуждал.

Смирнов С. А.: Один идеи давал, второй писал.

Павловский Е. Н.: Козьма Прутков, то же самое. То есть уже проблема, там не один человек, их два, конкретный текст.

Смирнов С. А.: Да. Например, написание стихов, это давно уже было сделано, когда машина могла написать стихи онегинской строфой. Долгое время считалось, что это гениальное изобретение Пушкина, это его рифмовка, стихи, и этой строфой он написал весь роман, гениальное достижение.

Павловский Е. Н.: Сонеты там.

Смирнов С. А.: Да, сонеты тоже в определённой форме. Машина давно это делает, закладывает данные, она тебе выдаст, и у читателя возникает ощущение, что эти стихи написал хороший поэт, хотя написала на самом деле машина. Тем самым мы приписываем или не приписываем осмысленность или мыслительное действие тому, кто породил этот текст. Это же наше приписывание.

Павловский Е. Н.: Да, именно, нам кажется.

Смирнов С. А.: Это происходит именно в акте чтения, а не где-нибудь.

Павловский Е. Н.: Да.

Смирнов С. А.: Так давайте отдельно разбирать текст, акт порождения текста и акт чтения.

Павловский Е. Н.: Мы пришли к той ситуации, когда мы не можем их различить.

Смирнов С. А.: Тогда это проблема различия. Этот акт чтения можем моделировать?

Павловский Е. Н.: Можем.

Смирнов С. А.: Если мы можем моделировать порождение текста, так мы можем и моделировать акт чтения?

Павловский Е. Н.: Собственно, сеть была обучена, моделируя чтение.

Смирнов С. А.: Да, она же читала.

Павловский Е. Н.: Конкретно один из конкретных алгоритмов, где ошибка определяется. При чтении текста сетка маскирует, убирает слова и пытается восстановить то, какое следующее слово. Поскольку текст, который она читает, уже там есть эти слова, известен, она для себя закрывает, потом открывает, смотрит, угадала или нет. На этой ошибке учится.

Постепенно сетка дорастает до того, что она любое слово может восстановить, то же самое со знаками препинания и так далее. Поскольку внутри нейросети формируются иерархические признаки, то можно сказать, она дорастает до понимания каких-то конструкций грамматических, синтаксических, как минимум.

Смирнов С. А.: То есть она фактически осваивает, во-первых, грамматику.

Павловский Е. Н.: Фактически она воспроизводит то, что человек читает тексты, учится, у себя закладывает эту грамматику.

Смирнов С. А.: Да.

Павловский Е. Н.: Только нас ещё постепенно доучивают в школе, что такое грамматика. Но есть люди грамотные от рождения, есть какие-то способности у людей, которым не сильно много правил надо изучить, они и так

быстро всё грамотно начинают понимать. Скорее нейросети про это, чтобы они много читали грамотных текстов, поэтому они чуть-чуть выучили эту грамоту.

Смирнов С. А.: То есть многое зависит от того, какие тексты закладываешь?

Павловский Е. Н.: Да, конечно.

Смирнов С. А.: Одно дело, ей закладываешь инструкцию по вождению, другое дело, роман Толстого.

Павловский Е. Н.: Да. Как раз лет 15 назад корпусная лингвистика, не так много было корпусов, допустим Национальный корпус русского языка, НКРЯ, это Толстой, Достоевский, вся литература.

Смирнов С. А.: Да, вся литература.

Павловский Е. Н.: То есть это тексты, имеющие высокую экспертную квалификацию, они прошли публикацию и так далее.

Смирнов С. А.: Да.

Павловский Е. Н.: Нынешние сетки учатся на Википедии. Википедия – это отмасштабированная та часть, которая в культуре воспроизводилась в виде текстов, это энциклопедия, в которой своя экспертная система, то есть, есть люди, их десятки тысяч человек, которые пишут Википедию. Но это тоже относительно грамотные тексты.

Смирнов С. А.: Это всё-таки другого типа тексты, энциклопедические тексты – это не тексты художественного романа.

Павловский Е. Н.: Да, это другие тексты.

Смирнов С.А.: Там больше информации, там она более разнообразная, но это не значит, что там богаче смыслы, содержание и так далее.

Павловский Е. Н.: Да, но эти современные сети обучены на Википедии, а это очень много действительно, гигабайты текста.

Смирнов С. А.: Да, информации там много.

Павловский Е. Н.: Плюс всё, что опубликовано, вся литература. Плюс ещё есть возможность доучивать это всё на блогах, а блоги – это ещё больше.

Смирнов С. А.: Это миллионы, миллионы, там много, чего. Тогда остановимся на главном. Значит ли это, что сетка постепенно, обучаясь на этих текстах, начинает понимать смыслы?

Павловский Е. Н.: Я думаю, это прекрасная возможность развивать науку о смыслах и определять понятие смысла.

Смирнов С. А.: Да.

Павловский Е. Н.: Потому что, если мы понятие определим критериально, через какие-то внешние признаки, то мы также возьмём нейросеть по внешним признакам, которая будет имитировать всё то же самое, и всё.

Смирнов С. А.: Есть такие попытки обучить машину смысловому действию?

Павловский Е. Н.: Что мы назовем смысловым действием?

Смирнов С. А.: Когда разработчик пытается описать, что такое смысл, а не только то, что такое значение, что такое знак, текст.

Павловский Е. Н.: Мы соберём тысячу, 10 тысяч философов.

Смирнов С. А.: Философов, которые накидают тебе кучу определений.

Павловский Е. Н.: Скажем, есть ситуация, здесь есть смысл, а в этих нет смысла, и экспертной оценкой доведём эту базу, а потом в нейросеть загрузим, она сама будет отвечать так же, как эксперты, и она будет имитировать ответы экспертов. Значит совершает она мыслительное действие? Нет, она отвечает нам так же.

Смирнов С. А.: В смысле понимает или нет, я больше про это пытался спросить. Понятно, что не в том смысле, что работа со смыслом идёт. Понимает ли то, что не формализуемо? Потому что, когда я загоняю многообразие экспертных суждений в сетку, я всё равно формализую. Просто, если я раньше формализовал одни задачи и определения, теперь я про смысл загоню в неё знает много чего, из Википедии или тысячу экспертов, но по принципу действия я действую также, в этом смысле формализую и это.

Но смысл есть то, что не формализуемо, он связан с вопросами, зачем, во имя чего? Они решаются каждый раз в конкретной ситуации конкретным субъектом, во имя чего. Я могу миллион ответов таких дать, но миллион, не миллион, дело не в том, что миллион, а дело в том, что это каждый раз решается конкретным субъектом в конкретной ситуации. Поэтому оно не формализуемо. Формализовать можно и миллион действий, но, когда у тебя в принципе она не может быть взята из словаря, смысл не берётся из словаря.

Павловский Е. Н.: На этот счёт есть попытки смоделировать это таким образом, что берётся за основу гипотеза, смысл или ценность жизни собственной этого субъекта, которая возникает только тогда, когда у субъекта есть опасность потери этой самой жизни.

Смирнов С. А.: Да, это смысловые вопросы возникают, когда на границу встаешь.

Павловский Е. Н.: Есть идентичность. Если мы говорим про совокупность этих роботов, агентов, которые в моделированной среде, чтобы у агента была идентичность, у него должен быть уникальный опыт, отличный от опыта других, и ценность этой жизни. Тут не знаю, как это все моделируется, насчёт ценности. Если этот робот уникальный, второй уникальный, 100 этих уникальных, они взаимодействуют, один какой-то жертвует собой, допустим, ради других, он помирает, но его код, его опыт поглощается другими.

Смирнов С. А.: Да.

Павловский Е. Н.: Наверное, может быть, да, здесь возникнет модельное понятие смысла, то есть он либо жертвует собой, либо он других истребляет.

Смирнов С. А.: Есть такие работы, когда сетка жертвует собой? Как агент действия?

Павловский Е. Н.: Поспелов, я читал недавно, Поспелов Дмитрий, наш исследователь ИИ, москвич, целая школа у него была. Он вводил понятие Я работа и пытался понять, как это Я смоделировать.

Смирнов С. А.: Это было уже давно?

Павловский Е. Н.: Да, давно.

Смирнов С. А.: Тогда технический уровень был не тот, который сейчас, тогда сеток не было. То есть, когда сетки, которые в Go играют, тогда, может быть, можно уже ставить задачу, а тогда у него еще вряд ли это получилось бы.

Павловский Е. Н.: Да, было это в 1968-1972, где-то так.

Смирнов С. А.: Да, это вообще Советский Союз.

Павловский Е. Н.: Он Эрика Берна попытался взять.

Смирнов С. А.: Смоделировали коммуникации по Берну?

Павловский Е. Н.: Да, ребенок-ребенок, родитель-взрослый, ввести эту модель. Но опять же мы утыкаемся в то, что мы бы внутри под капотом не моделировали, все это выплывает на суд человеку, в интерфейс человека.

Смирнов С. А.: Конечно, ему надо всё объяснить.

Павловский Е. Н.: Этот интерфейс у нас, что это? Либо мы зрительно смотрим на поведение, оцениваем, либо мы общаемся, и всё.

Смирнов С. А.: Всё, только так.

Павловский Е. Н.: Со зрительным полем все сильно сложно технически, есть Япония, человекоподобные роботы, лицо, мышцы, это все очень сложно и неправдоподобно.

Смирнов С. А.: И зачем?

Павловский Е. Н.: Да, и зачем, когда можно написать текстом на русском языке, уже кажется по тесту Тьюринга, что там уже за экраном какой-то человек сидит. Теперь можно ставить вопросы про Я, субъектность, смыслы. Мне кажется, мы как раз подходим к моменту, что невозможно никакими методами взаимодействия выяснить наличие смысла в другом, всё моделируется.

Смирнов С. А.: Всё моделируется?

Павловский Е. Н.: Да, ответ такой, всё есть полная иллюзия, можно моделировать всё. Тогда у меня другой подход квантовый возникает, что, когда мы измеряем квантовую систему, в её реальности квантовая система была в состоянии суперпозиции, а при взаимодействии, при измерении мы всю эту квантовость схлопнули какой-то одной опцией, которую можем наблюдать.

Это не свойство, вывод сделали такой, что есть квантовая система, наблюдатель, канал взаимодействия, и наблюдатель спутывается с этой квантовой системой. Я не знаю, это противоречит законам, уравнению Шредингера, что квантовая система не должна схлопываться.

Смирнов С. А.: Но она схлопывается?

Павловский Е. Н.: Да, то есть нелинейность происходит в ней, и это чистый эффект наблюдателя, вот и всё.

Смирнов С. А.: Именно, момент схлопывания можно фиксировать при наблюдении?

Павловский Е. Н.: Да.

Смирнов С. А.: Причина схлопывания?

Павловский Е. Н.: Взаимодействие, начали взаимодействовать, и схлопнулось. Сознание человека имеет свойство, не схлопывая систему, как-то мысленно проникать, соединяться с другой, понять другого человека, влезть в шкуру другого человека, при этом не препарировав его.

Смирнов С. А.: Не препарировав, все эти понятия эмпатии, чувствования, переживания, которые, во-первых, работают при акте взаимодействия в живой встрече.

Павловский Е. Н.: Да.

Смирнов С. А.: Но это не означает телесное взаимодействие, это означает...

Павловский Е. Н.: Ментальное.

Смирнов С. А.: Такое, если говорить о высших психических функциях по Выготскому, то есть это тонкая психологическая настройка, то, что называется душой.

Павловский Е. Н.: Да.

Смирнов С. А.: Тогда?

Павловский Е. Н.: Тогда, когда мы из технической системы строим взаимодействие, мы переходим из квантового мира, схлопывая до измерительных операций. Когда человек с человеком общается, мне кажется, мы не теряем этого. Хотя есть проблемы измерения, когда ты действительно имел в виду то, что сказал, или мне показалось.

Смирнов С. А.: Мы же это можем проверить друг у друга, потому что степень искренности тоже вопрос такой, в ответе с фигой в кармане. Как раз смысл там, в этом прямом контакте душевном, которые можно ради эксперимента пытаться моделировать. Надо?

Павловский Е. Н.: Это интересно.

Смирнов С. А.: То есть, как задача, очень интересная?

Павловский Е. Н.: Да, но общество это использует не во благо.

Смирнов С. А.: Я про это, когда ты начинаешь работать с этой тонкой материей, а она уже связана с риском, что можно, взяв этот результат или очень сложный продукт, использовать его не во благо, а во вред, это психологическое оружие. Тогда вопрос опять же этики.

Павловский Е. Н.: Да.

Смирнов С. А.: Тогда здесь вопрос. Все эти опыты в сфере разработки ИИ привели к тому, что в разных странах, в том числе в России, стали обсуждать эти кодексы, разрабатывать этику для ИИ, Евросоюз принял целую декларацию, в прошлом или позапрошлом году на форуме ИИ приняли наш кодекс в несколько страниц, этический кодекс, «Сбер» один из первых предложил свой.

Павловский Е. Н.: Кодекс предложили самые активные и агрессивные.

Смирнов С. А.: Самые агрессивные одновременно. Но там черным по белому видно, что это чисто маркетинговый ход, мол, любите ИИ, он дружелюбный и не кусается. Сберу нужно было для клиентов это все сказать: «Ребята, бабушки, старушки, не бойтесь, все хорошо». Правильно?

Павловский Е. Н.: Мне кажется, да.

Смирнов С. А.: То есть это не про этику.

Павловский Е. Н.: Проблема доверительного интеллекта ставится таким образом, чтобы человек ему доверял.

Смирнов С. А.: Но это чисто маркетинг, чтобы ты покупал, пользовался, входил в контакт, не боялся разговоров с роботом и так далее, но это же не про этику.

Павловский Е. Н.: Мне кажется, на самом деле проблема сильно глубже.

Смирнов С. А.: Проблема глубже.

Павловский Е. Н.: Потому что сфера интеллектуальной деятельности...

Смирнов С. А.: Где реальная этика интеллекта касается? Когда интеллектуальную разработку используют для того, чтобы манипулировать миллионами с помощью умной разработки.

Павловский Е. Н.: Это уже границу перешли, уже всё.

Смирнов С. А.: Уже перешли?

Павловский Е. Н.: Да.

Смирнов С. А.: Эти орудия стали оружием?

Павловский Е. Н.: Да.

Смирнов С. А.: Они уже используются?

Павловский Е. Н.: Да. Были опыты в Новосибирске по управлению сознанием на расстоянии, органы определенные этим пользовались, потом после развала Союза это стало коммерческой разработкой, и всё, сейчас это везде активно используется как средство влияния.

Смирнов С. А.: Нет, в то время речь шла буквально о вторжении в мозги, все, что связано с гипнозом и т.д.

Павловский Е. Н.: Сейчас кажется, что это попроще работает, то есть устанавливается связь с мозгом человека посредством, допустим, чата. Есть взаимодействие.

Смирнов С. А.: Да.

Павловский Е. Н.: Как цыганка, она увидела и всё.

Смирнов С. А.: Как мошенники манипулируют пенсионерками по телефону.

Павловский Е. Н.: Да. Дальше важно, что есть обратная связь и подкрепление, и всё. Следующее, мы призываем человека делать то, что он или выходит на улицу, за отделение Сибири голосует, ещё что-то.

Смирнов С. А.: Да.

Павловский Е. Н.: Можно делать не явно, а последовательно маленькие шаги, которые человек сам не понимает. Он то на один сайт зайдёт, то на другой, где-то проголосует, это вроде какая-то даже не осознаваемая деятельность, но она накапливает общую энергию, негативную, которую можно потом вылить в определённую сторону. Получается, если на Западе с этим работают с нашими людьми, то и наши должны работать, иначе у нас...

Смирнов С. А.: Иначе проиграем, да.

Павловский Е. Н.: Всё это дошло до уровня постоянного управления частью ресурсов. Я замечаю, что я 8 часов на работе сижу, примерно два часа я какой-то фигней страдаю, на сайте проверяю, новости читаю. Зачем мне это? Уже неосознаваемая деятельность, кажется, она у многих происходит.

Смирнов С. А.: Значит ли это, что этим управляют?

Павловский Е. Н.: Да, я думаю, что ничего просто так не происходит, кому-то нужно это.

Смирнов С. А.: Заметь, это сугубо опять не техническое, не инженерное дело, это уже сугубо человеческие дела, люди просто начинают пользоваться умной разработкой.

Павловский Е. Н.: Да. Но и фишка в том, что, если раньше какому-нибудь правителю требовались огромные усилия, чтобы что-то сделать, сейчас один человек может сконцентрировать на себя ферму серверов, поставить такую задачу, которая всех загрузит, они все будут на него работать. Этот масштаб из-за того, что техническое развитие, получилось, что один человек стал иметь возможность массово управлять всем.

В этом и происходят все эти этические вопросы, должен ли этот человек, один или несколько, обладать силой такой, можно ли это отрегулировать. Мне кажется, печаль в том, что опять призовут ИИ, потому что человек сам не сможет решить.

Смирнов С. А.: Чтобы он решил эту задачу. Слушай, парень, я тебя избрал, ты реши за меня эту задачу.

Павловский Е. Н.: Да. Опять уход человека, почему?

Смирнов С. А.: Туда же. Это будет подкрепляться всё больше и больше.

Павловский Е. Н.: Уход, потому что очень тяжелые усилия, их нужно каждый день прилагать. Если не прилагаешь, всё – начинаешь пищу разносить.

Смирнов С. А.: Да. Тогда получается, что, кстати, из Википедии разные определения ИИ, уже устарели, поскольку там доминировало представление, что ИИ – комплекс решений, позволяющих имитировать когнитивные функции человека. Определение неправильное с точки зрения современной ситуации. Сетка уже не имитирует и не копирует, это более сложные действия.

Павловский Е. Н.: Был тут у нас с визитом ректор университета «Сириус», он сказал, что они прямо специально продвинули слово «имитирует» в стандарт, чтобы оно было. До этого тоже обсуждали, что имитировать, чтобы отличить человека, типа, это имитация, а есть человек настоящий.

Смирнов С. А.: Всё-таки это неправильно.

Павловский Е. Н.: Очень правдоподобно имитирует.

Смирнов С. А.: В том-то и дело, если действительно в игре Go не просто перебор, а действительно обучение, действие непредсказуемое, то это сложно назвать имитацией.

Павловский Е. Н.: Мне кажется, как-то Вы так вопрос задаете, как будто бы есть абсолютный смысл имитации.

Смирнов С. А.: Потому что за словом стоит значение. Имитация – повторение существующего действия, причем не всегда осмысленное, то есть не содержательное, главное – аргументация и внешняя форма.

Павловский Е. Н.: Мы в такой постановке не можем определить осмысленность. Что значит осмысленное?

Смирнов С. А.: Имитирующее действие неосмысленное по определению.

Павловский Е. Н.: У нас все определения, которые раньше мы приписывали человеку, например, у собаки осмысленное поведение.

Смирнов С. А.: Надо смотреть.

Павловский Е. Н.: Они привязаны антропоморфно.

Смирнов С. А.: Да.

Павловский Е. Н.: Мы думаем, так в крайнем случае мы просто скажем: «Это не человек, поэтому не осмысленное». На проверку выясняется, что мы не можем определить понятие «осмысленное» без человека.

Смирнов С. А.: Да, оно привязано к человеку.

Павловский Е. Н.: Вот. Тогда смысл про техническую систему, ей это свойство приписывать? Именно когда она начинает проявлять подобные человеку свойства, мы говорим: «Так это может быть осмысленным», а политические акторы, которые в стандарт это определение положили, говорят: «Нет, это имитация». Почему? Потому что мы сейчас наплодим сущностей, разработчики революцию нам устроят. Поэтому имитация, мы сразу отделили это от человека, это хуже человека, это недостойная штука.

Смирнов С. А.: Я и говорю, это правильная стратегия, определение ИИ таким образом? Наверное, нет.

Павловский Е. Н.: Наверное, да.

Смирнов С. А.: Потому что во всех определениях имитация – копирование, мы так его определяем.

Павловский Е. Н.: Определение же даётся, чтобы те люди, заходящие в эту область, постепенно нарабатывали понимание. Начиная с университетской скамьи, мы вкладываем в ребяташек понимание, что это имитация, это не настоящее.

Смирнов С. А.: Ему хочется, чтобы было настоящее, чтобы он был, как у человека, но, чтобы он не имитировал.

Павловский Е. Н.: Это неэтично уже.

Смирнов С. А.: Когда Вы мечтали в школе получить друга, Вы же не хотели, чтобы он имитировал?

Павловский Е. Н.: Да.

Смирнов С. А.: Чтобы он был настоящим другом.

Павловский Е. Н.: Это плохо.

Смирнов С. А.: Почему? Кстати, по-человечески это вполне нормальное желание.

Павловский Е. Н.: По-человечески нормально, но с точки зрения развития общества это кажется огромной угрозой, то есть мы создаем...

Смирнов С. А.: Равного себе.

Павловский Е. Н.: Мы творцы, демиурги, создали нечто кремниевое, не на углеводе.

Смирнов С. А.: Да.

Павловский Е. Н.: Придаём ему все больше и больше технических возможностей.

Смирнов С. А.: Да.

Павловский Е. Н.: В определённый момент, когда мы ему дали когнитивных возможностей, он захватил наше.

Смирнов С. А.: Даём.

Павловский Е. Н.: Даём, почему-то даём.

Смирнов С. А.: Тогда, почему определяем так? Мы сами себе противоречим, одной рукой даём такое определение про имитирующее устройство, а другой рукой всё больше и больше отдаём, где всё меньше и меньше имитации, а всё больше и больше умных когнитивных функций.

Павловский Е. Н.: Второй такой, латентный процесс, он без большого осознания.

Смирнов С. А.: Более мощный, когда граница уже пройдена.

Павловский Е. Н.: Да.

Смирнов С. А.: Миллионам нужны не эти определения, им же это нужно, все больше и больше отдавать, потому что это связано с усилием. Всё, получите. Тогда лучше всё-таки, наверное, опять возвращаемся к тому, не определять ИИ как имитацию, но и не то, чтобы просто бездумно отдавать, а всякий раз перестраивать по-новому интерфейс, сохраняя при этом за собой функцию демиурга, а рядом с собой иметь умного помощника, ради бога. Но, зачем он нужен мне, как имитирующий меня? Мне он нужен, как тот, который сделает то, что я в принципе не могу сделать, но никогда не перейдет границу.

Павловский Е. Н.: Мне кажется, это правильная постановка вопроса. Пусть внешняя система помогает мне становиться лучше.

Смирнов С. А.: Да, я про это. Зачем мне плохая копия меня? Мне нужна, во-первых, не копия, во-вторых, хороший, умный помощник, который делал бы то, чего я не умею, но, который всегда останется только другом-помощником.

Павловский Е. Н.: Чего я пока не умею.

Смирнов С. А.: Пока не умею, я могу, конечно, тоже научиться, я тоже способен учиться.

Павловский Е. Н.: Я бы даже сказал, что не могу научиться, а должен. Если мы убираем «должен», то костыль постепенно завладевает.

Смирнов С. А.: Завладевает костыль, да.

Павловский Е. Н.: Мне эта мысль нравится, что человек во внутриутробном развитии воспроизводит в своём развитии все стадии от ДНК червя.

Смирнов С. А.: Да, в онтогенезе повторяет.

Павловский Е. Н.: Первые семь лет человек воспроизводит базовые навыки и мораль, и все в школе.

Смирнов С. А.: Он в онтогенезе повторяет филогенез.

Павловский Е. Н.: Всё повторяет. Мы должны не аутсорсить, что-то отдавать, но постоянно воспроизводить весь багаж. Вот это будет эффективно.

Смирнов С. А.: И наращивать.

Павловский Е. Н.: Да. Тогда имеет смысл этим заниматься, иначе просто мы летим в пропасть.

Смирнов С. А.: Летим при таком разрыве, и непонятно, как остановиться. Спасибо большое, Евгений. Ставим смысловую паузу, многоточие, может быть, ещё отдельно поговорим в новой ситуации. Спасибо.

Павловский Е. Н.: Хорошо. Очень рад был пообщаться.

Смирнов С. А.: Спасибо, было интересно.