# Круглый стол «Каким будет ИИ следующего поколения?»[1]
# Round table «What will the next generation of AI be like?»

**КИИ-2023,** 19 октября 2023 г., Смоленск, Россия

**Ведущие:** ***Карпов В.Э.*** **(НИЦ «Курчатовский институт»)**
***Самсонович А.В.*** **(НИЯУ «МИФИ»)**


***Карпов В.Э.***: Коллеги, доброе утро! Сегодня у нас единственный, по-моему, на нашей конференции круглый стол. Предполагается формат свободного вещания без длительных выступлений, просто обсудить какие-то наболевшие проблемы. В рамках нашей основной тематики будет рассказ Алексея Владимировича Самсоновича о том, что произошло в Китае, произошло в хорошем смысле… А тема у нас с вами: «Каким будет искусственный интеллект следующего поколения?». На правах ведущего я хочу спросить у Алексея Владимировича: почему тема такая – «Каким будет ИИ …?» Что не так с нынешним ИИ?

***Самсонович А.В.***: Я, действительно, несколько в неловком положении. Как я уже сказал, мы будем импровизировать на ходу. Что значит: «Что не то»? Всё то, всё великолепно, наоборот, можно сказать, триумф, честь и хвала всем разработчикам, просто невероятные сейчас происходят события. Но при всём этом, конечно, существуют ограничения, выше которых, очевидно, существующий ИИ не поднимется. Есть некий потолок для современных больших языковых моделей и глубоких нейросетей, и, в частности, он связан с их неспособностью до сих пор воспроизвести человеческое социальное поведение на уровне взрослого человека. Как человек может себя вести с друзьями, на вечеринке или в жизни вообще. Т.е., когда речь идет о принятии решений: как адекватно отреагировать, как поступить, как проявить себя по отношению к тому или иному человеку, с которым уже есть какие-то сложившиеся отношения, в той или иной ситуации. Тут, конечно, ИИ оказывается беспомощным, просто потому, что это чисто статистическая модель.

Она, конечно, обучена на данных, а данных подобного рода в нужном объёме нет, чтобы её обучить. Да и если бы они были, то само обучение непонятно, сколько займёт, вся история человечества нужна. Поэтому, мне кажется, что должны произойти какие-то изменения. В частности, много говорят об интеграции статистического и когнитивного подходов. Т.е. речь идёт, в частности, о когнитивных архитектурах и о глубоких нейросетевых и больших языковых моделях. Многие сейчас говорят об этой интеграции, очень много появилось идей. Меня Валерий Эдуардович пригласил сказать о событии, где обсуждалась эта тематика. Например, Джон Лейерд (*John Laird*, Center for Integrated Cognition, https://integratedcognition.ai/our-team) ставит много интересных и важных вопросов. Факт, что об интеграции думают многие, подходы есть разные. У меня есть свой взгляд на эту задачу, но я думаю, надо ждать, что произойдёт какое-то качественное изменение. Причём, как говорит Лейерд в своей лекции: «Новый ИИ должен быть полностью автономным». В том смысле, что ему не нужно будет ставить такую-то цель, задачу. Он будет непрерывно расти, эволюционировать и адаптироваться под человеческие нужды и сам решать, что ему делать. Это, конечно, сложная проблема.

А пока вопрос стоит о том, как, собственно, использовать большие языковые модели для наших нужд, видимо, в сочетании с другими средствами, потому что сами по себе они при всём своём могуществе оказываются весьма ограничены. Я не знаю, в каком-то смысле я ответил на вопрос или нет.

***Карпов В.Э.***: В общем, чего-то не хватает, поэтому надо думать о том, что должно быть дальше.

***Самсонович А.В.***: Да. Если так ставить вопрос, о чем хотелось бы поговорить? Я тут набросал три пункта:

- Каким должен быть ИИ? Должен ли это быть всё же автономный агент, общающийся с человеком на социальном уровне и не требующий программирования, или же это должна быть какая-то среда для разработки или платформа?
- Должен ли ИИ быть рукотворным, в том смысле, что он создаётся путём инжиниринга руками программистов, или же он должен быть самообучающимся, т.е. возникать сам в нужных созданных условиях?

[1] Прошедший в рамках КИИ-2023 Круглый стол стал одним из наиболее интересных мероприятий конференции. По его итогам предполагалось сделать информационное сообщение. Затем возникла идея им не ограничиться, а сделать некоторый развёрнутый текст, – очень уж интересными показались выступления участников и затрагиваемые ими темы. Когда была готова расшифровка выступлений, которую можно было положить в основу текста, оказалось, что Круглый стол именно в виде стенограммы, со всеми шероховатостями, оговорками и пр., имеет свою особую привлекательность. Этот живой текст и предлагается читателю. *Редакция лишь слегка прикоснулась к нему, стремясь сохранить его «дух».*

- Как мы должны регулировать его этичность и его поведение с точки зрения морали? Тут неизбежно нужны какие-то ограничения. Нужны ли для этого законы, какая-то конституция, или же нужно что-то вроде воспитания, как мы воспитываем человека?

  Вот такие вопросы хотелось бы обсудить.

***Карпов В.Э.***: Да, отчасти мы определяем что-то вроде направления развития или свойства, или что, Алексей Владимирович? Вчера, например, Алексей Николаевич Аверкин рассказывал в своём докладе о поколениях ИИ. Алексей Николаевич, Вы нам скажете пару слов о том, что было?

***Аверкин А.Н.*** **(*ВЦ им. А.А. Дородницына РАН*)**: Да, конечно. Я вкратце скажу. Понятно, что 1-е поколение – символьный интеллект, 2-е – коннекционистский. Сейчас коннекционистский – все глубокие сети и машинное обучение. Третье, по терминологии *IBM*, в основном я связываю его с объяснительным, но там ещё доверительный, этический, но в основном это доверие связано с объяснением, т.е. раскрытием. У нас каждое десятилетие во всём происходят сдвиги. ИИ первого поколения от экспертного обучения и баз знаний, созданных вручную, перешёл к глубокому обучению 2-го поколения, к нейросетям, большим обучающим выборкам, это примерно с 2000 до 2020 г. Теперь мы вступаем в 3-е поколение ИИ, где система может интерпретировать и объяснять алгоритмы принятия решений, даже если он имеет природу чёрного ящика, т.е. объяснимый ИИ является основной частью, но, естественно, не единственной, 3-го поколения. В 30-е годы увидим ИИ 4-го поколения с машинами, которые будут сами обучаться и динамически накапливать новые знания. К 40-м годам и, наверное, до 50-го и там до бесконечности – это уже системы с воображением, сильным ИИ, суперинтеллектом.

В 1970-1990 гг. ИИ хорошо рассуждал, но были проблемы с обучением, обобщением. В основном это был символьный интеллект, основанный на правилах. Сейчас 2-е поколение, которое хорошо обучается и воспринимает, но слабо в рассуждении и обобщении. Это машинное статистическое обучение, глубокое обучение – очень хороший результат, даже лучше, чем у человека, в обработке текста и изображений. 20-е, 30-е годы – ИИ прекрасно обучается и рассуждает, способен объяснять решение, это главное, пожалуй. Способен общаться на естественном языке на многие темы. Но *GPT* чуть не дотягивает до 3-го поколения, он не может объяснять свои действия, я много раз пытался получить из него объяснение алгоритма, ни разу не получилось. Нужен отказ от больших данных, нужны меньшие объёмы данных для обучения, минимальный внешний контроль. Ну, уже 4 и 5 волны способны решать те же интеллектуальные задачи, что и человек, это так называемый сильный ИИ, ведущий к суперинтеллекту и «технологической сингулярности», когда мы теряем контроль над скоростью развития интеллекта, он уже доходит до результатов, которые мы не можем себе представить. Мы учим его одному, а он имеет что-то другое. *GPT* примерно такой – частотный словарь, а, оказывается, умеет делать почти все, что умеет человек. Плохо, но умеет… Все основные методы объяснительного интеллекта пошли именно оттуда. Финансирование, естественно, дошло до 3-го поколения. Вот это моя позиция, спасибо.

***Карпов В.Э.***: Кто хочет сказать по поводу того, что говорил Алексей Владимирович? Я напоминаю, у него поставлены следующие вопросы: агенты или среда, рукотворный или самообучающийся интеллект, то, о чём, в частности, говорил предыдущий докладчик. Третье – это вопрос этики. И пять волн у Алексея Николаевича.

***Мельников А.В.*** **(*Югорский НИИ ИТ*)**: Я хочу сказать, что принципиально не согласен с предыдущим выступающим. И с постановкой вопроса, и с теми результатами, которые у нас есть. 14 марта 2023 г. был объявлен публичный доступ к большой языковой модели *GPT*-4. В принципе, ничего особенного не произошло. Но тот, кто занимается практическими решениями задач обработки естественных языков, знает, что эта модель позволила на 10 и более процентов повысить качество обработки текстов. Я представляю отраслевой институт. Руководство округа не интересует, занимаюсь ли я фундаментальными исследованиями. Его интересует решение вполне конкретной задачи: автоматизируй деятельность чиновника, сделай генерацию нормативно-правового документа, чтобы он соответствовал всем предыдущим документам – это реальная задача. Если я скажу ему, что достоверность составляет 80%, а чаще всего меньше – вспомним модели предыдущие, *BERT*. Сколько даёт средняя бертовская модель? Ну, больше 80% никто не вытягивает. На этих моделях реализовать реальные решения, практические задачи невозможно. Поэтому все академические исследования, которые проводились, оставались в стенах исследовательских институтов, лабораторий и т.д.

Когда произошел прорыв по качеству обработки текста – это включает в себя суммаризацию, выделение именованных сущностей, формирование диалога – в этот момент появилась возможность решать практические задачи с помощью тех вещей, которые придумали наши основатели, т.е. задачи автоматизации интеллектуальной деятельности человека. Я хочу отметить важный момент: если вспомнить эпоху появления компьютера, к чему она привела в итоге? К повышению производительности умственного труда человека… Одно из лучших решений в мире с точки зрения подготовки текста – *Microsoft Office*, мы все пользуемся им. Сейчас появилась возможность создать аналогичные инструменты, только следующего поколения. Моё глубокое убеждение, что эти инструменты появятся, не, как нам коллега показывал, в 30-е годы, а в 24-м. Не хотите? *Microsoft* уже встроила в *Word* свою не очень хорошую, но, тем не менее, искусственную большую генеративную модель.

Мы обсуждаем этические нормы: можно или нельзя? Если говорить о создании действительно ИИ, который будет моделировать личность человека, это совершенно отдельная задача. Но ведь одна из фундаменталь-

ных задач, которую мы имеем, это повысить производительность интеллектуального труда человека. За всю историю ИИ впервые не самые слабые люди мира обратили внимание: «Вау, происходит что-то такое фундаментальное, что меняет всю картину человечества». А мы что имеем на сегодняшний день? Вот эта картинка мировая [*показывает слайд из презентации*]. Первая строчка – *GPT*-4. Это *State of the Art*, лучшее решение, которое мы имеем. Это вот последняя публикация вчерашняя, *HuggingFace* – мировой площадки, на которой строятся и отрабатываются большие языковые модели. Вот последние строчки, которые выделил я специально. Обратите внимание, это 7-миллиардные модели. Знаете, в чём их особенность? Они неплохо реализуются на наших мобильных телефонах. Т.е. эти модели завтра придут к нам, сюда, на наш мобильный телефон.

А давайте посмотрим: сколько моделей больших языковых сделано в Российской Федерации? Две.

***Самсонович А.В.***: *DeepPavlov*?

***Мельников А.В.***: Нет, господа, *DeepPavlov* – это *BERT*, это 19-й год, всё померло уже. Мы сейчас работаем реально на бертовских моделях, их на *HuggingFace* порядка 80, одна из них – это Сбербанк, кстати, один из лучших показателей. Но это 19-й год, а мы живём в 23-м. Так вот, в 23-м году на территории Российской Федерации работают два коллектива, которые сгенерировали свои большие языковые модели. Вот коллеги будут сейчас выступать, рассказывать про Китай. Сколько в Китае сделано моделей больших языковых?

***Самсонович А.В.***: Я не могу вам сказать.

***Мельников А.В.***: Я могу сказать, потому что это цифра, которую я услышал от китайских партнёров – больше сотни. Больше сотни коллективов, которые делают эти работающие [модели]. Давайте посмотрим, в конце концов, у нас же есть Америка для сравнения. Это наш главный оппонент, про которого мы говорим. Сколько там сейчас коллективов? Я могу назвать сравнимые модели, которые по качеству работают примерно так же: *Cloudy*, *GPT*, понятно, *Bard* и дальше могу перечислять. А ситуация, которая на сегодняшний день на территории России? Мы, например, попробовали модели Сбербанка. Хочу сказать, что я ни в коем случае не претендую на итоговую оценку качества моделей. Но, обратите внимание, вот строчки «*YaGPT*»– это Яндекс-*GPT* по уровню, ноль – это оценка, *Turbo*_2 – это 3,5-GPT модель. Верхняя GPT-4, понятно по качеству ответа, что работает. Обратите внимание, Яндекс *GPT*-2 лучше стала, чем первая модель, которая вообще печальная была. Она вот тут находится, а *GigaChat* из Сбербанка.

И думаете, на чём мы работаем сейчас? Мы все работаем сейчас на *Saiga*. Почему на *Saiga*? И что такое история *Saiga*? Представляете, американская компания, опубликовала на *HuggingFace* обученную нейронную сеть, которая обучена на 70 миллиардах параметров. Мозг человека – это 100 миллиардов. Эту модель подхватил Стэнфордский университет и, дообучив её, в том числе этическим нормам, потому что она была очень сырая, сделал модель, которая пошла по миру и называется *Saiga*-1, а потом они ещё дообучили до *Saiga*-2. Так вот, *Saiga*-2 – англоязычная модель, на сегодняшний день единственный инструмент в Российской Федерации, который я могу поставить у себя на серверах и начать работать, решая реальные задачи.

У нас сейчас стоит *Saiga2_13b*, Белуга. А что такое Белуга? У нас сейчас дообученные модели, сделанные группами энтузиастов, например: два инженера компании МТС дообучили модель, и эту модель мы начали все использовать... Наверное, всем нравится пользоваться телефоном *Apple*, работать с операционной системой *Windows*, использовать *Microsoft Office*… Мое глубокое убеждение, что сейчас примерно та же ситуация сложилась с большими языковыми моделями... Я только что вернулся с ЗОНТа [IX Международная конференция «Знания-Онтологии-Теории», 2-6 октября 2023 г., Новосибирск]. Там с коллегами обсуждали, что есть нейронные сети, есть когнитивные модели, вот они вроде бы противостоят, но ничего подобного. Мы должны понимать, что это технологии, которые, особенно объединяясь, позволяют действительно эффективно решать практические задачи, которые бизнес возьмёт на себя, начнёт продавать. Поэтому, заканчивая это дискуссионное выступление, отвечаю на вопрос.

*Первое*: *GPT*-модель полностью изменила ландшафт ИИ. И это будет в ближайшие 10 лет, такой мейнстрим, который не знаю, во что выльется. Может быть, в те модели интегрированные, скорее всего так и будет. Потому что с точки зрения структурно-логических моделей они обязательно должны прийти, потому что это понимание сущности. В конце концов, ведь правильно – это статистическая модель.

*Второе*: если Российская Федерация имеет претензии занять какое-то место в мире не на уровне посылки выпускников наших университетов в *Google* или ещё куда-нибудь, где они будут потом создавать лучшие продукты, которые мы всю жизнь будем покупать, примерно, как мы сейчас покупаем персональные компьютеры… Это чистая экономика. Это наука, реализованная в нашей повседневной жизни и в качестве нашей жизни.

*Третье*: мне очень понравилось выступление «а как же ИИ?». Тут такая конкуренция. Подумайте, коллеги, я приехал, в том числе сюда, потому что считаю, что коллектив единомышленников может и должен говорить о приоритетах, в том числе, в государственной политике, которые должны реализовываться…

И последнее. Я послушал Академию наук, Всемирный конгресс по ИИ, там выступали мои коллеги. Мне было грустно, потому что общаюсь реально с молодежью. По-хорошему, у нас порядка пяти групп, которые есть в Новосибирском госуниверситете, которые есть в Москве, в том же МФТИ, которые лидеры по этой тех-

нологии, которые понимают, что и как надо делать... Мне коллеги говорят из Китая: «Давай, свои задачки на наших серверах». Я говорю: «А как это?» – «Ну как, это будет наш общий результат» – так работают китайцы.

В результате у меня предложение. Не знаю, готов ли коллектив, готов ли форум, сделать такое заключение, обращение, что, на мой взгляд, очень здорово, когда первые лица обращаются к тематике ИИ...

***Карпов В.Э.***: Спасибо, Андрей Витальевич. У меня будет три коротких вопроса к Вам.

В своё время, когда рецензировали статьи, относя в тот или иной журнал, сборник, помню, спрашивалось: по теме статья или не по теме? И один из критериев был такой: насколько эта работа, это исследование приближает нас к пониманию того, как мыслит человек? Нет, я не против нейронных сетей, тем более, чата *GPT*. И насколько вот эта технология приближает нас к пониманию того, как решаются человеком какие-то задачи? Помимо того, что система чудесно имитирует осмысленность.

***Мельников А.В.***: Все, наверное, читали книгу Сбербанка «Доверенный интеллект». У меня возникает только один вопрос. Вы возьмите губернатора, который принимает решение, и скажите: «А ну-ка, аргументируй мне доверенным интеллектом, почему ты принял решение вот этого поддержать человека, а не того?»

У меня вопрос примерно такой: мы хотим шашечки или ехать? Мы хотим поднять производительность труда? Причём, обратите внимание, я не случайно говорю, вот все мои презентации – это нейроассистент. Это инструмент поддержки и помощи человеку, точно такой же, как любой другой. Есть фундаментальные вопросы: вообще, как работает нейронная сеть, как она решает эти задачи...

***Карпов В.Э.***: Я понял. Спасибо. Ещё у меня был второй вопрос: какие же аспекты интеллектуальной деятельности мы автоматизируем и повышаем производительность? Понятно – имитация деятельности… А все-таки, у нас тема «Каким будет ИИ следующего поколения». Одна фраза: каким будет? Имитатором?

***Мельников А.В.***: Это будет ассистент человека, который позволит раз в десять, а такие оценки есть, поднять эффективность деятельности человека при выполнении рутинных, подчеркиваю, рутинных операций по обработке информации и управлению.

***Карпов В.Э.***:А уж выявляет ли он эмпирические закономерности или не выявляет – это дело десятое?

***Мельников А.В.***: Ну...

***Кобринский Б.А. (ФИЦ ИУ РАН)***: Уважаемые коллеги, с удовольствием послушал предыдущего выступающего, не буду повторять, со многим согласен. К сожалению, Алексей Николаевич знает, я частично или в значительной части не согласен с его прогнозами... Я хотел как бы две стороны рассмотреть. Одна сторона, связанная с практическим применением, только что была достаточно красиво и понятно представлена. Но, с другой стороны, есть вот этот фундаментальный или теоретический аспект. Алексей Николаевич говорит о том, что система будет рассуждать.

Мы в системах ещё первого поколения начинали с объяснения и рассуждения, но это была имитация человеческой деятельности. Здесь речь идёт, что системы смогут рассуждать так, как рассуждаем мы с вами. Рассуждать так – значит обладать самосознанием. А дальше возникает сильный ИИ, который сможет решать, как сказано, любые задачи. Ну, я думаю, Алексей Николаевич просто погорячился, потому что, если любые задачи, значит и любые теоретические задачи: была физика Ньютона, была физика Эйнштейна, будет физика ИИ, робота с ИИ – он создаст какую-то новую совершенно физику. Во всех областях будет создаваться что-то новое.

Но если перейти к этической или этико-юридической стороне, коснуться её, то можно вспомнить и роман «Франкенштейн». Если высокоинтеллектуальные роботы будут всё это уметь, то через некоторое время они подумают, что люди что-то мельтешатся, недостаточно хорошо рассуждают, у них недостаточно всё представлено, аргументировано, они, так сказать, эмоциональны не в меру, а вот у нас эмоции нормальные. Убрать их (людей), и будет всё прекрасно, и в мире всё станет идеально. Будет, так сказать, царство роботов.

Немножко утрирую, но если говорим, что сверхсильный ИИ уже появился, то тогда действительно. А что же мы – будем просто помогать? Вот слово «ассистирующее» меня порадовало, потому что я об этом пишу в статьях и говорю студентам и в МГУ, и в медицинском университете о том, что системы мы делаем даже не консультирующие, как их часто называют, а ассистирующие, они действительно вторые. Роботы в хирургии, например, называют «второй ассистирующий робот-хирург». Он ассистирует, но не консультирует...

Действительно идут эти волны, Вы правы, Алексей Николаевич. Те волны, которые прошли, мы можем оценить. Будущее мы прогнозируем, мы предполагаем. Действительно должна быть смычка различных подходов, они будут взаимодополнять друг друга. Большие языковые модели – прекрасно, они будут эффективны, но, когда мы говорим, что системы сегодня интерпретируют и объясняют – они вообще не объясняют, если говорить о нейросетях. А насчёт интерпретации, она или на уровне ребёнка, или это интерпретация для разработчика, тут я полностью согласен. Как изменяются весовые коэффициенты на разных слоях, это очень важно, мы лучше подбираем всё это, систему дообучать намного проще, но нужно как-то интегрировать с объяснением.

Военные спрашивают: как вообще мы это представляем? Такой условно «сильный», необъясняющий ИИ сегодня говорит, что надо нажать кнопку. На основе чего он (военный) должен это делать? Непонятно, на осно-

ве чего он соглашается и должен подчиняться, нажимать кнопку. То же самое врачи. Если человеку ставится системой опасный диагноз, врач не соглашается и ошибётся – его накажут: ИИ ему подсказал. А если наоборот – он согласился, но ИИ ошибся – тоже накажут. Без объяснений он не понимает, что ему делать с этим решением: хорошее, плохое, правильное? Если область ему недостаточно известна. Всё это требует решения…

***Карпов В.Э.***: Борис Аркадьевич и Андрей Витальевич правы в том, что все эти вещи, даже без объяснений, воплощаются в практике. Просто вспоминаю статью в журнале «Закон», в которой один из ведущих юристов, как он себя позиционирует, написал текст с помощью *GPT* и долго потом очень основательно говорил, что наконец-то появилась у юристов возможность писать качественные юридические тексты.

***Самсонович А.В.***: И не только юридические статьи.

***Карпов В.Э.***: Да, ещё и медицинские, наверное, будут писать. Это да, это беда, с этим надо что-то делать.

***Кобринский Б.А.***: Это было не только в статье, я недавно слышал это в другом месте, из уст одного человека, он сказал да, ИИ будет помогать, мы будем разрабатывать юридические документы. Может быть, ИИ может в чём-то помогать, но ассистировать на каких-то узких этапах, а не заменять, потому что иначе мы действительно скатимся. А что касается развития, наши самые лучшие системы с большими возможностями зачастую очень хорошо решают задачи, когда обработка данных быстрая. Кто знает все шахматные партии мира? Почему *IBM Watson* обыграл шахматиста? Потому что, конечно, компьютер всё может анализировать на 20, или сколько там у них ходов, при возможности *IBM Watson*. Но я просто привел пример.

Пока это была игра в шахматы, всё было прекрасно. Ведь *IBM Watson* заявили, что они решат проблемы медицины, в первую очередь, онкологии. Они действительно создали целый ряд программ, работали с американскими клиниками серьёзными и писали, и выступали, и рассказывали о хороших результатах. Кстати, я их не раз слышал, каждый раз просил, пришлите какую-нибудь статью, каждый обещал, ни одной статьи не было, были только пресс-релизы. Медики тоже ничего не писали до поры той, когда сегодня есть иски к *IBM Watson* в Соединенных Штатах от клиник пострадавших, пострадали больные, потому что предложения были неудачные. Значит, есть ограничение между тем, что работало всё в узкой сфере, даже, например, шахматы, там всё равно есть крайние границы, всё можно перебрать. В медицине, в космосе или в военном деле всё перебрать будет невозможно. И поэтому тут нельзя переносить одно на другое, а это тоже переносится.

***Реплика из зала:*** Это совсем разные миры.

***Кобринский Б.А.***: Это вообще разные миры. Я сейчас не говорю о больших языковых моделях, я сейчас о том, что, когда мы говорим о каких-то супердостижениях, когда говорится, что это решит наши те или иные проблемы, к этому надо относиться очень аккуратно. Касается ли это *IBM Watson*, языковых моделей, везде нужно понимать, какие у нас есть, могут возникать ограничения. Иначе мы можем очень далеко зайти...

***Реплика из зала* (*Еремеев А.П. – НИУ «МЭИ»*)**: У меня вопрос скорее к Андрею Витальевичу, как к специалисту. Доклад его прекрасен, более того, я полностью согласен с тем, что он сказал о Китае…

Я бы хотел задать вопрос: не зря сейчас идёт переход от инженерии знаний к науке о знаниях. Почему? Потому что, естественно, всё должно строиться на базовом фундаменте. Фундамент для нейронных сетей оказывается довольно хлипкий, но мы иногда уже видим, что ребята из Сколково, Университета Иннополис дают такие заявления: мы всё сделаем, у нас нейронная сеть всё рассчитает. Наверное, все слышали, доклад академика РАН Бетелина В.Б. и доктора физ.-мат. наук Галкина В.А. на Всемирном Конгрессе. Фундаментально о базисе нейронных сетей. Очень резко высказались, опираясь на действительно фундаментальную теорему Тихонова, что отображение конечного множества (а нейронная сеть всегда обучается на конечном множестве) на бесконечное, когда идёт конкретное распределение, – это всегда модель неустойчивая [по отношению к бифуркациям и хаосу]. Т.е. вполне может быть ситуация: тысячу раз вас система узнала, а на тысячу первый она считает, что вы находитесь в базе Интерпола как разыскиваемый преступник. На Западе уже был ряд судебных процессов этого плана, у нас это более мягко.

Т.е. речь идёт о том – мы с Алексеем Николаевичем говорили, – что очень важно создать доверительный ИИ, т.е. чтобы хотя бы какое-то объяснение от нейронной сети получить… И вот, в связи с этим, мне бы хотелось выяснить, этот принципиальный фундамент.

По поводу объяснительной компоненты здесь говорили уже сто раз, что-то пытаются встроить. А вот этот фундаментальный базис, что в принципе модель такого типа всегда неустойчива... И чем глубже, более сложная нейронная сеть, глубокое обучение, свёрточное, рекуррентное и т.д., тем меньше её надёжность. И мы не можем спрогнозировать тот момент, когда система даст сбой. Ну, и видно уже по многим случаям, что и шаттл сталкивается, и система ошибается при распознавании. Вот ваше бы мнение, как специалистов в этой области.

***Мельников А.В.***: Я только два предложения скажу. Человек может ошибаться. Нейронная сеть может ошибаться. Вот скажите: в чём разница этих утверждений? Почему в одном случае мы говорим: нам нужна доверительная нейронная сеть. Я хочу сказать, что принципиальный механизм нейронной сети – это всегда статистическая обработка с вероятностью ошибки. А почему по отношению к человеку этот термин неприменим?

***Самсонович А.В.***: Разница очевидна, мне кажется. Так, как может ошибаться нейронная сеть, человек никогда ошибиться не сможет.

***Мельников А.В.***: Классическая задача: распознавание знаков на дороге. На сегодняшний день нейронная сеть делает оценки лучше, чем человек. И многие вещи, которые есть. Она точно так же ошибается, как любой человек… А вот кто преподаёт студентам? Никогда не слышали, как галлюцинируют студенты на экзамене?

***Карпов В.Э.***: Нет, подождите, это сейчас идёт, мне кажется, уже не очень честная игра, потому что, конечно, ошибаются все. Но простите, я отвечаю за свою ошибку, и я пытаюсь объяснить, почему я принял такое решение, пусть ошибочное, прежде чем оно будет реализовано. Здесь говорилось очень много как раз об объяснительном компоненте. А так – разумеется.

***Мельников А.В.:*** Один нюанс. Мой тезис звучал об ассистенте. Спасибо, я поддерживаю эту идею. Т.е., другими словами, говорить об отдельной независимой личности, которая существует в обществе и решает задачи... Я пока даже не верю, даже не понимаю, как это может быть сделано... Когда я слушаю утверждения юристов по поводу опасности ИИ, я сразу картинку представляю следующую: палата лордов, конец XIX века, появилась страшная железка, которая ездит по дорогам. Что будем делать, как защищаться? Поставим человека с флажком впереди. А иначе... Ну и что дальше было? Историю знаете.

***Карпов В.Э.***: Ну понятно. Мария Николаевна хочет что-то сказать.

***Королева М.Н.*** (***МГТУ им. Н.Э. Баумана***): Коллеги, опять начали много говорить про нейронные сети: ругать их, не ругать – неважно. У меня вопрос. Мы так много говорим про объяснительную компоненту в ИИ, она нужна, она важна. Может быть, начнём все-таки определять, что же такое объяснительная компонента? И, может быть, займемся именно методами объяснительной компоненты тогда? Может быть, у кого-то есть какие-то комментарии на этот счёт: что это такое, чем это может быть на самом деле, ДСМ-методы, другие методы?

***Карпов В.Э.***: А это точно в плане темы нашей?

***Королева М.Н.***: Каким же будет ИИ?

***Карпов В.Э.***: Он будет объяснительным.

***Королева М.Н.***: Каким объяснительным он будет? За счёт чего он будет нам объяснять, как он думает, как он рассуждает. Хорошо, тогда, пожалуйста, у меня просьба как молодого члена ассоциации. Кто-нибудь, подготовьте к следующей конференции такой вот доклад: видение объяснительного, объяснимого ИИ. Мне кажется, это будет очень интересно.

***Богданов М.Р.*** (***БГПУ им. М. Акмуллы***): На мой взгляд, мы упускаем один момент, когда говорим про будущее ИИ, – это развитие пограничного ИИ и изолированного ИИ. Дело в том, что с развитием миниатюризации мы постепенно приходим к тому, что нас будет окружать огромное количество очень дешёвых, очень маленьких вот таких гаджетов, которые даже размером будут с песчинку. Мы на них можем наступить, не заметить, но они будут играть очень важную роль, причём как положительную, так и отрицательную роль. На слайде показана такая «страшилка» из области нанотехнологий, умная серая нанослизь. Это такие нанороботы, которые теоретически смогут размножиться и захватить весь мир…

***Карпов В.Э.***: Т.е. ИИ будет, прежде всего, овеществлённым, да? Хорошо.

Обозначу некоторые тезисы в попытке ответить на вопрос «Каким будет ИИ?» буквально. На мой взгляд, он будет, во-первых, этичным, и, во-вторых, социальным. Коротко попробую обосновать. Модели поведения – это то, что относится исторически к направлению и сфере интересов ИИ, по крайней мере, явно так когда-то формулировалось. Воплощённый или овеществлённый ИИ нынче, в общем-то, тоже как-то особо ни у кого не вызывает возражений... И, в-третьих, будет декларироваться дальше то, что имеет под собой какую-то более-менее реальную основу, некие готовые технологии, не заглядывая совсем уж так в будущее.

Ну, вот этика. Я говорил, что все правы, все затрагивают вопросы этики, агентности. Действительно, это вопрос, который на нашей конференции не затрагивался в принципе, об этом говорят много, и, почему мы не в тренде этого процесса, не очень понятно. Говорят совсем о другом. Вот совершенно не в ту сторону люди, говоря об этике и ИИ, идут обычно, потому что это неинтересно. Последствия, угрозы, вызовы, это, конечно, всё очень здорово, но не в этом специфика этики ИИ. Единственный специфичный момент этики ИИ в том, что мы имеем дело с системой, которая, по крайней мере, потенциально принимает жизненно важные и критичные для человека решения автономным образом. Ну, и стало быть, нас интересует как технарей: может ли сама техническая система или искусственный агент быть этичным? А зачем это нужно?

Вот совершенно практическая задача. Когда мы создаём коллаборативных роботов, роботов-партнёров, наверное, мы ожидаем, помимо прочего, ещё и того, что мой партнёр в сложной ситуации будет себя вести исходя из моих ожиданий о том, что такое хорошо, плохо, представлении добра, зла и т.д. Короче говоря, я буду от своего партнёра ожидать то, что я оцениваю с точки зрения этичности. Вот тогда возникает вопрос: а как сделать поведение искусственного агента, интеллектуального, когнитивного, какого угодно, неважно, таким,

чтобы его поведение отвечало моим представлением? И вот тут начинается целый ворох проблем. Я не просто так спрашивал по поводу онтологии гуманитарных областей? Проблема в том, что об этике говорить без философов тяжело, у философов есть всё для того, чтобы вам объяснить, что такое хорошо, что такое плохо, этические школы. Вот если бы можно было формализовать, построить онтологию, когнитивную карту, и дальше сопрячь с моделью мира, вот тогда поведение искусственного агента было бы, на наш взгляд, этичным…

Дело в том, что, когда наш искусственный агент совершает какое-то действие, он имеет возможность всегда объяснить, оправдаться, почему он сделал так. И он вам объяснит, докажет на основе формализма, в него заложенного, почему его поведение, когда он бросил партнёра, убежал куда-то, всё равно является этичным с точки зрения представления о гедонизме, эпикуреизме и т.д. Это уже хорошо, но, собственно говоря, он ведёт себя абсолютно как нормальный человек, которого мы считаем высокоморальным. Высокоморальный человек – это тот человек, который всегда вам объяснит, почему он так поступил. Способный объяснить, почему он обидел слабого, почему он должен был бросить кого-то. Вот здесь этичность в каком-то смысле обусловлена требованиями техническими. Ещё вчера упоминалась на пленарном докладе, такая неприятность, с *GPT* Сбербанка связанная. Слишком много бдительных граждан пишет обращения в органы по поводу каких-то непонятных ответов. Возникает вопрос, а как верифицировать на этичность то, что выдаёт очередной чат? Здесь вопрос формализации. Построение онтологии и прочих моделей могли бы помочь этому коммерческому проекту.

Следующий вопрос. Опять, техническая задача, групповая робототехника… Штука в том, что, когда маленькие устройства, нанороботы и прочие, живут в сложной среде со всеми неприятными атрибутами, надо искать какие-то пути адаптации, пути решения задачи устойчивого функционирования. В природе такой путь есть – это образование социума. Исходя из этого, когда мы имеем дело со сложными, недетерминированными задачами, стохастическими, динамическими средами можно попробовать пойти по пути природоподобия.

Природоподобие будет сегодня рассматриваться, это та самая *BICA*, о которой Алексей Владимирович расскажет. Я о том, что один из путей развития интеллектуального, овеществлённого или воплощённого ИИ – образование социумов. Здесь возникает целый ворох проблем, которые ещё надо будет решать, но, с другой стороны, под этими задачами уже сегодня есть некий базис. И обратите внимание, там же, как ни странно, возникла опять этическая составляющая. Потому что, когда мы имеем дело с социумом, всегда надо уметь решать конфликты, а задача морали – это разрешать конфликты, которые не регулируются другими механизмами.

Вот, собственно говоря, почему неизбежным образом в социуме появляется то самое, что имеет отношение к предыдущему направлению об этичности поведения. Вот это я и хотел сказать. На мой взгляд, прямо отвечая на вопрос, каким он будет, мне кажется, что ИИ следующего поколения будет этичным и социальным. Спасибо.

Кто ещё хочет выступить?

***Виноградов Г.П.***: **(*ТвГТУ*)**: Я много слышал о том, что рой, социоподобные и т.д. рождают эмерджентность. Вот не могли бы Вы сказать, есть ли какие-то решения этой задачи? Т.е. как само групповое поведение рождает эмерджентность? Ведь то, что социоподобно, реализует новую модель поведения. Это и есть эмерджентность? Но, когда мы рассуждаем о том, что рой что-то нам сделает за нас, мне кажется, это немного неправильно.

***Карпов В.Э.***: Нет, рой ничего не сделает. Группы агентов могут иметь разную структуру. Вот рой – это самое неинтересное. Рой – это структура, в которой вы ориентируетесь исключительно на своих ближайших соседей, у вас нет общих целей, задач, понимания того, что нужно делать. А социум начинается тогда, когда у вас группа агентов, во-первых, неоднородна, там нужен лидер или система лидеров, когда есть специфические механизмы взаимодействия. В биологии это стремление быть вместе, какая-нибудь когезия, повторять телодвижение или поведение других, это опять же вопрос доминирования. Вот из каких-то базовых механизмов складывается то, что мы называем социумом. Я не про социум людей, люди здесь совершенно ни при чем. Тогда, когда эти механизмы начинают работать – есть такое необходимое и достаточное условие – группа агентов начинает самоорганизовываться в том плане, что она совместно начинает решать какие-то задачи.

Например, она начинает перераспределять функции, точнее, мы наблюдаем, как происходит перераспределение функций. Мы наблюдаем, как организуются какие-то структуры, которые разнятся не только по морфологическому, но и функциональному признаку. Мы видим, как система становится устойчивой к убытию части членов и действительно занимается тем, что решает основную свою задачу. Ведь социум – это просто способ решения, да? Можно без социума как-то обойтись [Сложные, морфологически развитые животные решают задачу индивидуального выживания вполне успешно]. А вот муравьи, например, так не могут. Это – тропические животные, 120 миллионов лет назад им стало холодно, они выживают только сообща. Ничего такого здесь нет, не надо проводить аналогии с человеческим обществом. Это просто путь развития технических систем.

Вы скажете: что такое эмерджентное свойство в самом явном виде, с точки зрения вот тех самых новых абсолютно свойств и прочее? Я вам такого примера не приведу, потому что начнёте приводить контрпримеры, это очевидно. Вообще эмерджентность – это дело мутное. С точки зрения технической или робототехнической единственное, что видел в своей жизни – опубликованная не так давно в журнале «Природа» статья. Там была

эмерджентная система – простые технические устройства с одной степенью свободы, которые ничего не могли делать, а когда они соединялись вместе, начиналось движение. Вот, на мой взгляд, единственный пример внятный технический, без всяких спекуляций, в котором проявляются эмерджентные свойства на 100 процентов.

***Аверкин А.Н.***: Можно ещё одно замечание. Тут прекрасный вопрос: могут ли нейросети ошибаться? Мне кажется, нейросети 3-го поколения уже не будут ошибаться. Во-первых, они объясняют, как чёрный ящик, мы влезаем в нейросети и строим логические структуры внутри. Мы подсвечиваем всякие знаковые пиксели, в общем, если нейросеть ошиблась, даже пользователь может найти её ошибку. С другой стороны, сеть может объяснить другой нейросети на символьном языке. К тому же вот вы помните эти голубые волны, которые синусоиды? Что делает 1-е поколение в 3-м поколении? Оно сейчас делает большие графы знаний. Ну, вот ещё Гугл построил граф знаний, 500 миллионов узлов, там триллионы связей. Эти громадные графы знаний как раз сращиваются. Сейчас практически у нас имеется семантическая сеть, которая описывает весь мир. Это используют и *GPT*-4, и объяснительный ИИ. Т.е. 3-е поколение – это гибридные нейросети плюс модели мира. Они столь большие, что мы потеряли их из вида. Вот эта гибридная система уже не может ошибаться.

***Самсонович А.В.***: По поводу ошибок нейросетей в своё оправдание хочу сказать: возможно, меня не поняли. Когда ошибается нейросеть, я не говорю о любых её ошибках, а вот именно о тех, которые демонстрируют полное непонимание сути предмета. Вот такие ошибки человек не совершает. Вот то, что я хотел сказать.

***Аверкин А.Н.***: У обычной нейросети нет семантики вообще, нет модели мира. У 3-го поколения уже появляется, внешняя, но появляется.

***Карпов В.Э.***: Коллеги, есть ли ещё замечания вот по этому блоку, не природоподобному? Пожалуйста, Александр Павлович.

***Еремеев А.П.:*** Я здесь с ведущим, конечно, согласен. Опять-таки в силу своих фундаментальных положений нейронная сеть, конечно, может ошибаться…

Пока ИИ работает в режиме имитации, как говорится: «как робота обучают, так он и танцует, так он и рассуждает». И почему естественники считают, что *GPT* – прекрасная система, а историки говорят, что она чушь выдаёт. Ну, понятно, потому что на чём обучался *GPT*? Он подключен к битобайтным базам, в основном все библиотеки, базы западные, американские. Понятно, когда ему задают вопрос типа: «А был ли обмен ядерными ударами между США и Советским Союзом?», он говорит: «Да, что-то было». Понятно, чушь. Извините, он на этом обучался. Когда *GPT* говорит на многих языках прекрасно – английский, французский, но русский язык у него на уровне школьника. Потому что доступ к информации о русском языке у него весьма ограничен.

У меня вопрос вот в чём, Стивен Хокинг говорит: «ИИ, особенно самообучающийся, может достичь такого совершенства, когда он будет считать, что человек с его проблемами, с его непредсказуемостью ему просто мешает». Что потом будет? Кнопка у человека должна быть. Когда ИИ сейчас полностью управляет автоматами, электростанциями и т.д. Вполне возможно, что и кнопка-то не у человека будет…

Как вы относитесь к этой модели? Действительно, не окажется ли, что ИИ будет управлять всем и вся, мы к этому практически идём. Не получится, что такие муравьи вроде нас ему мешают? Что в этом случае будет? Конечно, это далёкое будущее, но всё-таки это тоже проблема, и, когда мы сейчас про этику, мы всё-таки имитируем. Конечно, если человеку дали пощёчину – это совсем другие эмоции, нежели если пощёчину дать роботу. Конечно, можно поставить датчик, положить, что это минус 100, но это имитация. А кто-то сделает плюс, и робот будет доволен, и это имитация. Но, если действительно самообучение, чему он там научится, подключённый ко всем библиотекам и ко всем вопросам, тут может быть вопрос, а не окажемся мы тут лишними?

***Самсонович А.В.***: А Вы не думаете, что человек гораздо опаснее, чем ИИ?

***Еремеев А.П.***: Хотелось бы, чтобы этот вопрос рассмотрели люди, которые с этим работают... Нейронные сети – сейчас ум ИИ. И прогресс видится в самообучающихся моделях, в основном это на базе нейронных сетей.

***Карпов В.Э.***: Да там даже без обучения будет страшно, потому что опять же наш основной предмет или объект для подражания – муравьи. Они чем хороши – мы не решаем проблемы обучения, муравьи не обучаются, им это просто не нужно, а социум существует, свои задачи решает.

120 миллионов лет эволюции привели к тому, что всему, что им нужно, они уже обучились. Вот и всё. А для них условия не меняются в таких широких пределах, когда надо приобретать новые навыки. Это не те пределы изменения среды, в которых требуется приобретение новых навыков. С точки зрения биологии это так.

Дело не в этом, понимаете, с точки зрения технической. Вы, Александр Павлович, создаёте систему, ещё раз повторяю, техническую, которая решает свои задачи, которая представляет собой некую автономную сущность, автономный агент, который должен обладать целостным поведением. Вы ставите для него интеллектуальную или когнитивную настройку системы управления, модель мира. И вы, как технарь, понимаете, что она должна быть непротиворечива, чтобы её поведение было устойчиво, чтобы в этой картине мира не было каких-то сущностей трансцендентальных типа человека, который почему-то может вмешаться и нарушить выполне-

ние основной программы. Поэтому в смысле кнопки управления – вы нарушите целостность, ещё раз повторяю, системы управления. Если мы говорим вот о таком действительно совсем предельном случае. Это социум со своими законами, своей моралью. Ещё раз повторяю, что мораль у них, как способ урегулирования конфликтов, вполне себе ничего, с эмоциями, потому что эмоции – это не только человеческое свойство. Это свойство вообще животного, как такового, нормальный регуляторный механизм, так что вся эта атрибутика будет. Будет закон внутреннего существования, восприятия человека как внешнего фактора. Вошёл в картину мира или не фигурирует в ней – это дело десятое. Рассуждать об угрозах, конечно, здорово, но давайте просто будем последовательны. Что мы делаем? Либо ущербную автономную сущность, либо реально что-то интеллектуальное.

***Еремеев А.П.***: Не окажутся ли перспективы развития ИИ в противоречии с нашими моральными принципами. Не сочтёт ли он нас лишними просто в своём развитии?

***Самсонович А.В.***: Извините, но я слушаю сейчас эти высказывания, может быть, с внутренней улыбкой, потому что для меня это просто как голливудские фильмы. Это неактуальный вопрос, потому что, в конце концов, всё решает человек. И ещё долго будет решать всё человек. ИИ сам не будет переделывать мир и решать, оставлять человека или завоёвывать, не будет войны человека с роботами. Не будет этого никогда. Если будет, то, значит, роботы будут управляться другими людьми. В общем, это уже происходит, и тут вопрос не о том, как человек должен противостоять роботу, а как человеческие ценности должны победить в этом противостоянии машинным ценностям.

Ведь понимаете, что происходит? Мы превращаемся уже несколько в другую цивилизацию, другое общество благодаря всем новым технологиям, с другой ментальностью. Многие вещи, которые раньше были абсолютными или же бесценными, отходят на второй план. И борьба, мне кажется, идёт за то, чтобы сохранить наш дух, наши идеалы, наши ценности, наши принципы, как человека, как людей. Т.е. наше понимание морали, наше понимание добра и зла, наши высокие идеалы, если хотите. Вот это нужно каким-то образом перенести в возникающую новую форму жизни, значит нам нужно, чтобы ИИ понимал человека, чтобы он мог чувствовать, мог хотя бы моделировать те самые эмоции, которые испытывает человек, мог общаться с человеком на социальном уровне человека, а не как автомат.

На самом деле я хотел перехватить инициативу в этой дискуссии. И поскольку время всё-таки уже близится к завершению, давайте хотя бы на время переключимся на вопрос о том, а что же вообще такое ИИ? Я так слышу сейчас, что до сих пор 90% говорившегося здесь об ИИ относится к большим нейросетям. Но, если вы возьмёте учебник по ИИ, скажем, 20-ти летней давности, там нейросети занимают маленькую какую-то часть в оглавлении. Вообще, под ИИ люди понимали нечто другое, и это по-прежнему актуально, т.е. когнитивные модели, вся логика, всё что угодно, те же генетические алгоритмы даже. Много есть вещей, которые сейчас отошли на второй план, но они существуют, вот одна из этих вещей – *BICA*, биологически инспирированные когнитивные архитектуры. Я на самом деле оказался здесь только сегодня ночью, потому что провёл очень успешную конференцию *BICA* в Китае. И я хочу использовать мое время для того, чтобы дать слово одному из докладчиков этой конференции. По-моему, это был лучший доклад. Пусть вы услышите не весь его доклад, а только начало, где он формулирует проблемы и задачи. Его понимание может быть отличается от тех вопросов, которые обсуждались здесь сейчас, но больше связано как раз с научными и фундаментально-технологическими вопросами. [Включилось видео. См. https://www.youtube.com/embed/sZiMjZmG7gQ].

***Самсонович А.В.***: На этом я хотел бы прервать, потому что лекция очень длинная и трудно выбрать наиболее интересные моменты. Я думаю, вы услышали в начале постановку задачи, каким должен быть агент, какие интеллектуальные задачи он должен решать, это было чётко сказано. Далее он рассматривает, какие пути решения этой задачи существуют, и здесь он говорит о возможности интеграции больших языковых моделей с когнитивной архитектурой. Это как раз главный вопрос в его докладе. Т.е. он рассматривает 4 варианта, я не хочу все их рассматривать, но, скажу то, что наиболее понятно мне.

Большая языковая модель служит как периферийное устройство для когнитивной архитектуры, т.е. она реализует основную когнитивную функцию, а чат *GPT* используют для того, чтобы связать это с естественным языком, выразить на естественном языке и воспринять то, что сказано на естественном языке, перевести на его внутренний язык, вот его роль. Т.е. не нужно придавать очень большое значение (это моя точка зрения), не нужно возлагать очень большие надежды, большие функции давать этим языковым моделям, их место – это периферийное устройство в когнитивных архитектурах. Это одна из возможных точек зрения.

***Карпов В.Э.***: Правда, проблема того же самого *Soar*, когда вы ставите модель мира, вы сталкиваетесь с классической задачей формирования всей этой жуткой сети. Представление всего этого формализма в терминах *Soar* – это очень тяжело. Инструмент, конечно, могучий, но я не знаю. Он разработчик *Soar*?

***Самсонович А.В.***: Он и его создатель.

***Карпов В.Э.***: Лучше бы он сайт поддерживал в порядке с этим продуктом.

***Самсонович А.В.***: Я считаю, что сама когнитивная модель тоже может быть преобразована в нейросеть. Мы сейчас над этим тоже работаем. Но это другая задача. Это не то, что сейчас происходит в мейнстриме.

***Карпов В.Э.***: А что происходит сейчас в мейнстриме?

***Самсонович А.В.***: Ну, смотрите, Вы лучше меня знаете, что сейчас происходит… Вот пример Китая. Я только что приехал оттуда, там – другой мир. У них нет *Google*, у них нет всей технологии *Microsoft* и всего прочего. У них всё своё. Все свои аналоги и ключи. Понимаете? Туда вы приезжаете, у вас компьютер на другом языке говорит с вами. Там другие программы установлены. Всё другое. И вот так они живут.

***Карпов В.Э.***: Алексей Владимирович, насколько я понял вчера из нашей с Вами краткой беседы, что если подытожить китайский путь развития, то это по верхам, но очень модно, очень актуально.

***Самсонович А.В.***: Нет. Я сказал не так. Я сказал они очень хорошие копикаты, очень быстро копируют и повторяют то, что сделано на Западе. Но делают это, конечно, уже у себя и по-своему… Для них то, что было 2 года назад, уже не существует. Это тоже проблема, с моей точки зрения.

***Карпов В.Э.***: А у нас другая крайность.

***Самсонович А.В.***: Да. Но не совсем. Я бы не сказал. С одной стороны, действительно, можно говорить, что в России засилье традиционных школ. Но я уже понимаю, что это не просто те школы, которые были ещё в Советском Союзе. Это просто привычка мыслить по-своему, традиционно. Взять хотя бы область когнитивного моделирования, она гораздо шире, чем семантические сети, онтологии, теория категорий. Возьмите тот же *Soar*. Там уже всё, что угодно, включено. Вообще, когнитивная архитектура выступает как платформа для интеграции разных подходов… Я не могу всех вас переубедить, но даже само слово «*cognition*» всё-таки следует переводить на русский не как «познание», а как «мышление». Ну, вот, собственно, я хочу сказать, что везде свои крайности. В Китае одно помешательство, в России – традиционность мышления, в Америке в науке опять всё стало консервативным, и, в конце концов, идёт к вырождению. Моё мнение, конечно. Но...

***Карпов В.Э.***: Что делать-то?

***Самсонович А.В.***: Да, что делать.

***Мельников А.В.***: А можно вопрос? Я же практик, я не теоретик. Сколько пользователей у *Soar*?

***Самсонович А.В.***: Ой, много. Прежде всего, это военные.

***Мельников А.В.***: Ну, давайте порядок.

***Самсонович А.В.***: Управление дронами, между прочим. Для военных.

***Мельников А.В.***: А в нашей жизни? Мы обычные люди, которые пользуются сотовыми телефонами. А сколько пользователей у *Open AI*?

***Самсонович А.В.***: Вы рассуждаете, как компания Бена Гертцеля: мерой интеллекта является количество денег, которые вы заработали. Но я с этим не согласен.

***Карпов В.Э.***: *Soar* – это нормальное инструментальное средство [, и пользователи у него есть]. Я понимаю, что такой вопрос был не совсем серьёзным. Нет, на самом деле просто наболело с этим *Soar*. Когда начали с ним работать, всё было хорошо, даже какую-то статью написали, опубликовали. Это было прекрасно. Начали разбираться, но это просто тяжело, неудобно.

***Самсонович А.В.***: *Soar* – это вчерашний день, понимаете? О нём уже можно забыть. Сейчас речь идёт о других совершенно моделях.

***Вохминцев А.В.*** **(*ЧелГУ*)**: Послушал некоторые Ваши тезисы, и, в принципе, с ними согласен, поскольку, на самом деле вот тот процесс познания и мышления гораздо сложнее, естественно, чем это делают нейросетевые архитектуры. Т.е. это, вообще говоря, сложнейшие системы и целеполагание, даже какие-то эмоции, другие чувства, они, вообще говоря, принципиально влияют на процесс получения нового знания и мышления. Если бы на сегодняшний день на территории России мы захотели бы провести такую конференцию, либо в рамках конференции по ИИ, мы бы смогли такую секцию сформировать?

***Самсонович А.В.***: Во-первых, у нас уже была *BICA*, в 2017 г., в Москве. И, кстати, очень успешная. Во-вторых, сейчас в Россию никто не поедет...

***Вохминцев А.В.***: Те задачи, которые вы ставите, наверное, самые важные в ИИ. А места им на этой конференции не нашлось. Это было бы, конечно, очень полезно.

***Самсонович А.В.***: Ну вот сейчас же нашлось немного места.

***Вохминцев А.В.***: Да, спасибо.

***Карпов В.Э.***: Ну, коллеги, какие ещё соображения? Кто за чат, кто против?

***Алексеев П.Н.*** **(*Военная академия Генерального штаба*)**: На вопрос, каким быть ИИ. Наверное, надо понимать, к чему мы можем прийти, развивая ИИ. Я не так давно начал заниматься проблематикой. Я не хочу уходить в глубину когнитивных моделей, нейромоделей. Я хочу в целом. Я понимаю, что идёт некое объединение, где-то разъединение. Это эволюция определённая. Рано или поздно появится что-то новое. Уйдём от консерватизма, придём к каким-то другим вещам.

Вместе с тем, каждый, наверное, сам для себя отвечает на вопрос: каким ему быть? Наверное, таким, каким мы хотим его видеть. Какая цель ИИ для человека? Здесь звучали фразы, что это ассистент, помощник и т.д. В военной сфере так или иначе ИИ не может принимать решения, потому что ответственность за принятие решения велика. Поэтому однозначно, что это ассистент, советчик. Вместе с тем, от результата применения ИИ человек становится зависимым…

Теперь, переходя от этой зависимости в работу органа военного управления, принимающего военные решения. Не надо системе давать право принимать решения. Достаточно, чтобы ИИ дал совет. Сказал, что вот, товарищ, если ты не нажмёшь на кнопку, значит, будет то-то. Если ты нажмёшь на кнопку чуть позже, ты проиграешь. Ну, [и прочие такие] страшилки.

***Самсонович А.В.***: Простите, пожалуйста, роботы принимают решения на поле боя.

***Алексеев П.Н.***: Я сейчас не о роботах. Потому что в вопросах управления в военном деле есть два основных направления: управление оружием, которое базируется на теории автоматического управления, и управление войсками, там, где люди управляют людьми или группой людей. Вот роботы – это задачи управления оружием. Здесь есть прогресс, здесь есть наработки и т.д.

В вопросах управления войсками сейчас большая яма. Нет ни концепции, ни реализации удачных [систем, которые принимают решения]. Но рано или поздно это появится. К чему это приведёт? А это приведёт к тому, что рано или поздно система выдаст человеку маленькую ошибку, за которую будет отвечать всё человечество. Ну и как бы как пессимистично это ни смотрелось, на мой взгляд, развитие ИИ – это очередной шаг человечества к своему самоуничтожению. Спасибо.

***Самсонович А.В.***: Или к выживанию.

***Кузнецов О.П.*** **(*ИПУ РАН им. В.А. Трапезникова*)**: Я хотел бы сказать вот о чём. Об этом говорят очень мало, хотя исследования уже есть. Вот известно, что на пути развития обучающих систем, систем машинного обучения, возникает такое обстоятельство, как расход ресурсов, расход энергии. И постепенно это принимает характер некоторой проблемы. Например, известные исследования где-то 22-го года, где эмпирически показывается при обследовании пары сотен систем распознавания, синтеза речи в тех областях, где действительно обучающие системы имеют большие успехи, исследователи выявили зависимость порядка полинома четвёртой степени. Т.е. рост качества решения задачи вызывает расход в четвёртой степени ресурсов. И помимо расхода электричества, энергии, воды, который измеряется тоннами для охлаждения, это вычислительные мощности.

Два вопроса. Один совсем конкретный: актуально ли это для интеллектуализации робототехники? Сейчас или в ближайшем будущем? И вопрос совсем общий, на который я не ожидаю ответа: а может быть цифровой путь вообще имеет свои принципиальные ограничения? Ведь наш мозг, который работает в миллион раз медленнее электроники, до сих пор эту электронику во многом превосходит. За счёт чего? Я думаю, в значительной степени за счёт того, что он не цифровой, а аналоговый. Это вопрос, который повисает в воздухе, но задать его, наверное, надо было.

***Самсонович А.В.***: Ничего не повисает, его можно смоделировать на цифровом компьютере, сохранив всю функциональность...

***Кузнецов О.П.***: За счёт каких ресурсов?

***Самсонович А.В.***: Тех самых, которые есть сейчас, это моделирование проводится давным-давно.

***Кузнецов О.П.***: Ведь все ресурсы мозга они в голове в очень небольшом объёме. Какой-нибудь *Watson* занимает больше десятка серверов.

***Самсонович А.В.***: Наоборот, я бы сказал, что пример мозга как раз доказывает, что возможен более эффективный ИИ. Что такое мозг? Это та же нейросеть, по сути. На других принципах, но это нейросеть импульсная. Она потребляет 20 Вт энергии. Сравните с гигаватами, которые сейчас нужны для обучения. Есть над чем работать.

***Карпов В.Э.***: Подождите. Кстати, по поводу энергозатрат, я помню, кто-то посчитал, что один запрос гугловского плана обходится примерно в столько энергии, сколько достаточно для того, чтобы вскипятить чайник воды. Это дорого, конечно. Но, Олег Петрович, это просто удобный инструмент. Я не хочу сидеть с паяльником и делать аналоговые схемы. Мне нужно быстрое моделирование. Я трачу на это много сил. Когда модель будет зафиксирована, я буду не программу писать дальше, а буду реализовывать это в виде схемотехнического решения, гораздо более энергоэффективного. Всё вполне естественно.

Вы же пока не можете договориться, как должна выглядеть модель мозга. Как договоритесь, так и перенесём с компьютеров на что-то ещё. Так что куда здесь деваться? По бедности или по лености.

Так, коллеги, ещё кто-то хочет сказать о наболевшем? Мы ответили на вопрос, каким будет ИИ следующего поколения?

***Самсонович А.В.***:Я думаю, что никто не знает сейчас ответа, но мы можем предложить какие-то гипотезы хотя бы.

***Карпов В.Э.***: И резолюцию какую-то подписать, да?

***Самсонович А.В.***: Да, но в моём понимании, это всё-таки должен быть функционально-человекоподобный автономный агент, общающийся с человеком на социальном уровне. Его не нужно будет программировать, ему достаточно будет объяснить, что вы от него хотите. Вероятно, на естественном языке. Ну, что касается формы его воплощения, тут возможны различные варианты. Глядя на молодёжь, как они любят всё время крутить эти мобильные телефоны (и старики, кстати, тоже), я предвижу, что устройства станут ещё более мобильными, ещё более портативными. Возможно, это будут очки, возможно, в конце концов, это будут импланты. Но это уже тяжело принять, что мы вынуждены будем делать себе импланты. Скорее всего, вначале появится что-то вроде очков с распознаванием движения рук для печатания на виртуальной клавиатуре, например, да? Я не знаю. Но это уже детали.

***Карпов В.Э.***: Два часа сидели, но так и не договорились окончательно?

***Карпова И.П.*** **(НИУ ВШЭ)**: На самом деле была такая интересная дискуссия, и очень не хочется, чтобы было такое послевкусие: «Мы идём к краху». Давайте рассматривать это по-другому. Сколько технологических революций пережило человечество? Да, станки ломали, какой кошмар – раньше все были обеспечены работой, теперь появились станки, и все останутся без работы.

И по поводу зависимостей. Да, мы зависим от смартфонов, но мы точно так же зависим от электричества, от канализации, простите, от водопровода. Нет горячей воды в кране – кошмар! Холодильник сломался – жизнь закончилась! Поэтому смартфон – это не самое страшное, ИИ – это не самое страшное. Давайте будем считать всё-таки, что это положительное направление развития человечества. Мы расширяем свои возможности, расширяем горизонты, и, в конце концов, мы перейдём не к самоуничтожению, а к появлению дополнительных возможностей у человека с точки зрения интеллекта, и с точки зрения появления новых ярких впечатлений, и исследования окружающей нас природы. Спасибо.

***Карпов В.Э.***: Спасибо. На этой оптимистической ноте, пока не наговорили всяких ужасов про то, как нас поработят, наверное, закончим. Коллеги, большое спасибо.

Пойдёмте фотографироваться!

**P.S.**

Всегда интересно сравнить дискуссии и их результативность.

Читатель имеет возможность ознакомиться с прошедшей 15 февраля 2024 года дискуссией на Радио Давос Всемирного экономического форума на близкую тему:

**Что дальше с генеративным ИИ?**

Три пионера ИИ, входящие в Топ-100 самых влиятельных людей в области ИИ по версии *Time*, делятся своими взглядами на прошлое, настоящее и будущее этой трансформационной технологии

(ведущий *Radio Davos Робин Померой*)
*Эйден Гомес*, соучредитель и директор *Cohere*
*Мустафа Сулейман,* соучредитель и директор *Inflection AI*
*Ян Лекун*, главный специалист по ИИ, американская транснациональная холдинговая компания

«Это будет самый преобразующий момент не только в технологиях, но в культуре и политике за всю нашу жизнь».

**What's next for generative AI?**

Three AI pioneers, all of them in Time's Top-100 most influential people in AI, share their views on the past, present and future of this transformational technology.

*Robin Pomeroy,* host Radio Davos

*Aiden Gomez*, Co-Founder and CEO, Cohere
*Mustafa Suleyman*, Co-Founder and CEO, Inflection AI
*Yann LeCun*, Chief AI Scientist, American multinational holding company

«This is going to be the most transformational moment, not just in technology, but in culture and politics of all of our lifetimes»

Почитать стенограмму здесь: https://www.weforum.org/podcasts/radio-davos/episodes/davos-2024-generative-ai-pioneers/
Послушать здесь: https://open.spotify.com/episode/71NQhRZmWMrLhBtbR2Qyco

## Ontology Summit 2024

Схожие проблемы начали обсуждаться и на Онтологическом саммите 2024. Первые заседания состоялись 21 и 28 февраля 2024 года. Всего запланировано 14 заседаний до 22 мая, где будет сформировано Коммюнике по теме: **«Neuro-Symbolic Techniques for and with Ontologies and Knowledge Graphs»**

The website for the Ontology Summit 2024 is available at https://ontologforum.com/index.php/OntologySummit2024