

НАУКА, ТЕХНИКА, ОБЩЕСТВО

*В.А. Лекторский, Е.А. Алексеева, Н.Н. Емельянова,
А.В. Катунин, И.Г. Меркулова, С.В. Пирожкова,
Е.О. Труфанова, И.О. Щедрина, А.Ф. Яковлева*

Искусственный интеллект в исследованиях сознания и общественной жизни (к 70-летию статьи А. Тьюринга «Вычислительные машины и разум») (материалы круглого стола)

Лекторский Владислав Александрович – доктор философских наук, профессор, заведующий кафедрой эпистемологии и логики, декан философского факультета. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, г. Москва, Мароновский пер., д. 26; e-mail: vлектorskii@gaugn.ru

Алексеева Екатерина Алексеевна – кандидат философских наук, доцент кафедры социальной философии философского факультета. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: eaalekseeva@gaugn.ru

Емельянова Наталья Николаевна – кандидат политических наук, доцент кафедры философской и политической мысли Востока, заместитель декана философского факультета. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: nemelianova@gaugn.ru

Катунин Александр Викторович – старший преподаватель кафедры эпистемологии и логики философского факультета. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: akatunin@gaugn.ru

Меркулова Инна Геннадьевна – кандидат филологических наук, доцент, руководитель Международного научно-исследовательского центра семиотики и диалога культур. Государственный

- © Лекторский В.А.
- © Алексеева Е.А.
- © Емельянова Н.Н.
- © Катунин А.В.
- © Меркулова И.Г.
- © Пирожкова С.В.
- © Труфанова Е.О.
- © Щедрина И.О.
- © Яковлева А.Ф.

академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: igmerkuloval@gaugn.ru

Пирожкова Софья Владиславовна – кандидат философских наук, доцент философского факультета. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: pirozhkovasv@gmail.com

Труфанова Елена Олеговна – доктор философских наук, профессор кафедры эпистемологии и логики философского факультета. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: etrufanova@gaugn.ru

Щедрина Ирина Олеговна – кандидат философских наук, старший преподаватель философского факультета. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: ioshchedrina@gaugn.ru

Яковлева Александра Федоровна – кандидат политических наук, доцент кафедры прикладной политологии факультета политологии. Государственный академический университет гуманитарных наук. Российская Федерация, 119049, Москва, Мароновский пер., д. 26; e-mail: afiakovleva@gaugn.ru

В статье представлены материалы круглого стола, проведенного на философском факультете Государственного академического университета гуманитарных наук 30 октября 2020 г. Круглый стол был приурочен к 70-летию публикации статьи Алана Тьюринга «Вычислительные машины и разум» (1950 г.). Участники круглого стола обсуждали роль исследований искусственного интеллекта в понимании человеческого сознания и естественного интеллекта, возможности и опасности внедрения искусственных интеллектуальных систем в разные сферы современной жизни – образование, политическую деятельность, социальные коммуникации. Была показана актуальность современного философского прочтения статьи Тьюринга с учетом новых цифровых вызовов и развития искусственных интеллектуальных систем.

Ключевые слова: искусственный интеллект, А. Тьюринг, тест Тьюринга, естественный интеллект, сознание, мышление, цифровизация

Лекторский В.А.: Наш круглый стол посвящен проблематике искусственного интеллекта, его роли в понимании того, что такое сознание и естественное мышление, а также тому, какие социальные следствия возникают в результате того, что в настоящий момент интеллектуальные системы широко применяются в жизни, экономике, культуре. В этой связи возникает масса проблем. Мы вместе с тем посвящаем наш стол 70-летнему юбилею публикации статьи А. Тьюринга «Вычислительные машины и разум» – она была опубликована в 1950 году.

Тьюринг был математик, а опубликовал статью в главном философском журнале Англии “Mind”. Это действительно философская статья, хотя она основана на математических идеях и разработках автора. В статье обсуждается вопрос о том, можно ли допустить такую ситуацию, когда ИИ разовьется до такой степени, что вы не сможете с помощью некоторых тестов выяснить, с кем вы имеете дело – с живым человеком, естественным существом, или же с ИИ? Тест Тьюринга породил массу споров и философских дискуссий. Между прочим, в то время еще не было компьютеров, первый компьютер появился

позже. Но Тьюринг не только предвидел их появление, но и заложил основы компьютерной науки, разработал идеи, использованные при разработках ИИ.

Некоторое время тому назад Сбергательному банку РФ, который сейчас называется «Сбер», поручили разработку искусственных интеллектуальных систем. Они собрали у себя большую группу людей, специалистов в этой области. В самом банке есть специальный отдел, связанный с робототехникой. Примерно год тому назад «Сбер» провел большую международную конференцию, связанную с разработками в области ИИ и применением этих систем в жизни – в экономике, в социальных областях. В рамках конференции был организован круглый стол о философских проблемах ИИ, который я вел. На этом круглом столе одним из докладчиков был А.Р. Ефимов, который является вице-президентом «Сбера», он сделал прекрасный доклад о философском смысле теста Тьюринга. Статья Тьюринга не перестала быть актуальной и злободневной, потому что та проблема, которую когда-то поставил Тьюринг, обсуждается до сих пор.

Через какое-то время после публикации этой статьи возникла новая дисциплина – когнитивная наука.

Основная идея когнитивной науки состояла в том, что наше сознание, мышление работает по типу гигантского компьютера. Человек – это несовершенная интеллектуальная система, а ИИ может быть более совершенной. Эта идея до сих пор никуда не исчезла, об этом пишут массу книг, в основном устрашающих. Например, недавно вышел бестселлер “Homo Deus: Краткая история завтрашнего дня” – книга, написанная израильским автором Ювалем Ноем Харари. Автор, основываясь на том, что сейчас происходит, рисует апокалиптические картины, когда ИИ будет управлять всем на свете. Эти проблемы кажутся многим нереальными, а предсказания Харари фантастическими, между тем эти вопросы совершенно реальны и серьезны.

Во всяком случае, одна из проблем, которую мы сегодня будем обсуждать, связана с тем, как и в какую сторону, после публикации статьи, изменилось наше понимание интеллекта, естественного и искусственного, можем ли мы согласиться или не согласиться с тем, что писал Тьюринг, и вообще – какова роль идей, разрабатываемых в области ИИ в понимании того, как сознает мир и мыслит человек.

Стоит сказать, что с тех пор, как возникла когнитивная наука, в ее рамках произошли большие изменения. Сейчас на первый план в этих разработках выдвинулись т.н. когнитивные нейронауки – попытки понять работу нашего мозга как процесс переработки информации, при этом способы этой переработки могут быть поняты с помощью вычислительных моделей (т.н. вычислительная когнитивная нейронаука). В рамках этих разработок были получены интересные результаты и сделаны важные открытия, например открытие зеркальных нейронов, новые данные о том, как в ряде случаев работает наше сознание и др. Основная часть разработчиков в этой области если сами и не разрабатывают модели ИИ, то пытаются их использовать, утверждая, что мозг работает как очень сложная вычислительная машина. Это одна сторона вопросов, о которых мы сегодня поговорим.

Вторая сторона состоит в том, что речь идет не просто о понимании сознания, мышления человека, а о том, что искусственные интеллектуальные

системы все больше и больше входят в нашу жизнь и ее радикально меняют. Это уже реальность, а не просто чья-то фантазия. И эта реальность меняется катастрофически. В этой связи возникает великое множество проблем, начиная с экономических, социальных, оборонных, политических и заканчивая проблемами о том, каково место человека в этой новой реальности, которую не так давно называли «электронным обществом», а сейчас «цифровым обществом», а цифровизация основана на использовании искусственных интеллектуальных систем, без них она была бы невозможна. Цифровизация – это реальность, в которой мы теперь живем. С одной стороны, о цифровизации не забывают политики, считая, что та страна, которая оседлает эту волну, будет победителем в мире, и с этим можно согласиться. С другой стороны, не все и не всегда, наверное, дают себе отчет о тех последствиях для человека, которые цифровизация может нести. С цифровизацией связано много практических проблем. Например, проблемы кибербезопасности, так как вся оборонная сфера связана с использованием систем ИИ. Возникает множество экономических проблем, начиная с того, будет ли меняться цель экономики. Ведь в условиях цифровизации сильно удешевляется производство товаров, и некоторые специалисты даже говорят о том, что может наступить то, что раньше считалось невозможным, – ситуация, когда некоторые товары будут раздавать бесплатно. Меняется сам рынок, поскольку можно просчитывать результаты, которые раньше просчитать было нельзя. В связи с цифровизацией возникает новая политическая, экономическая и социальная ситуация.

Но самое главное – то, что происходит с самим человеком. Великий философ И. Кант считал, что философия отвечает на три главных вопроса. Вот они. Что я могу знать? Что я должен делать? На что я могу надеяться? Сегодня мы должны заново продумать эти вопросы и попытаться дать на них ответы, учитывая современные реалии.

Что я могу знать? Некоторые сегодня считают, что можно знать только то, что можно просчитать. Здесь великая масса проблем, связанных, в частности, с использованием и обработкой больших данных, с машинным обучением и т.д. Само познание осуществляется уже иначе, и человеку часто непонятны те результаты, которые получаются в результате машинной обработки больших данных.

Что я должен делать? Человек считал, что с помощью ИИ он будет управлять миром и глобальным процессом эволюции (т.е. мнил себя Богом: *Nomo Deus*). На самом деле он становится управляемым с помощью этого интеллекта, таким образом превращаясь из Бога в раба, – искусственный интеллект лучше знает за него, что ему нужно.

На что я могу надеяться? Вроде бы надеяться и не на что, хотя надежду подают некоторые футурологи. Если вы посмотрите в Интернете сайт Российского футурологического общества, то увидите, что на 2050–2060 гг. предсказано возникновение т.н. «цифрового бессмертия». Мне приходилось специально писать по этому поводу: я пытаюсь показать, что т.н. «цифровое бессмертие» – не что иное, как ад, из которого нет выхода – ведь человек не сможет умереть.

Как вы видите, вопросов великое множество. Некоторые из них мы будем сегодня обсуждать.

Я бы хотел дать первое слово человеку, который хорошо знает, что такое тест Тьюринга, – это Екатерина Алексеевна Алексеева.

Алексеева Е.А.: Спасибо Владислав Александрович! Я бы хотела начать с того, что в тематике ИИ проблема возникает на уровне употребления самого понятийного аппарата. Если говорить о состоянии исследований в этой области на настоящий момент, то я бы предложила разделять два понятия: искусственный интеллект как гипотетический и футурологический проект, предлагающий некоторую модель человеческого мышления, и интеллектуальные системы. Это не только мое предложение, так, например, Финн Виктор Константинович на этом настаивает.

Когда мы говорим именно об ИИ, то здесь, как мне кажется, речь в большей степени идет о том, что сейчас называется общий искусственный интеллект – Artificial general intelligence (AGI). Есть различные исследования и подходы, предполагающие, что гипотетически возможно создание этого общего искусственного интеллекта, то есть такой компьютерной системы, интеллектуальные способности которой будут неотличимы от человеческих. Сейчас идет полемика о том, необходимы ли такой системе квалиа, весь набор субъективных человеческих свойств и т.п. Но предполагается, что эта система будет абсолютно универсальной и, соответственно, работающей в тех же областях, что и человек, она сможет справляться с теми же задачами, станет способна сама себе ставить цели и т.д. Такой проект общего ИИ очень значим для понимания того, что такое естественный интеллект. Сейчас нет единой теории сознания, мы до сих пор не можем предложить полностью удовлетворительного общепризнанного решения mind-body problem или того, что Д. Чалмерс называет “hard problem”. И, возможно, хотя это спорный момент, если бы получилось полностью смоделировать на искусственном носителе автономную психику с интеллектуальными свойствами, то это показало бы, что “hard problem” решена. Получается примерно так, как об этом пишет Р. Негарестани: создавая общий ИИ, мы пытаемся найти решение антропологических проблем. Итак, моделирование человеческого интеллекта или даже психики в целом – это одна большая тема в области ИИ.

Другая тема – это интеллектуальные системы. Изначально ИИ и интеллектуальные системы действительно не разделялись. У первых исследователей в области ИИ, например у AI Group в MIT, куда входили М. Минский, Д. Маккарти, Г. Саймон, А. Ньюэлл, была некоторая иллюзия, что довольно-таки быстро можно создать общую модель мышления и применить ее в прикладных разработках. Но в настоящее время хорошо заметна дифференциация между построениями общих моделей интеллекта и прикладными интеллектуальными системами, решающими отдельные практические задачи. Когда М. Боден говорит: «Искусственный интеллект везде», – то говорит она по сути именно об интеллектуальных системах.

Когда мы называем искусственные интеллектуальные системы искусственным интеллектом, мы попадаем в ловушку языка. Мы начинаем от них ждать того, чего они не делают, бояться их, приписывать им некоторую квазисубъектность, которой они не обладают. Интеллектуальные системы реализуют отдельные интеллектуальные функции, например распознавание текста

или образов, вождение автомобиля, выявление закономерностей в массивах данных, но не следует думать, что если мы объединим все эти функции в какой-то одной суперсистеме, то получим общий ИИ. Если общий ИИ и получится создать (в чем многие исследователи сомневаются), то это, по всей видимости, будет какая-то другая технология.

Тут возникает проблема замены представителей различных профессий, в частности преподавателей, искусственными системами, понимаемая как некоторая техноутопия, основанная на вполне реальных событиях. Есть отдельные проекты, например когда робот-гиноид Vina48 читала лекции в Вест-Пойнте или когда чат-боты ассистируют студентам, демонстрирующие возможности применения интеллектуальных систем в образовании. Называя такие системы искусственным интеллектом, их антропологизируют, а затем начинают предполагать, что ими вполне можно заменить педагогов-людей, исходя из экономических или еще каких-либо «рациональных» соображений. Я считаю, что антропологизация интеллектуальных систем – это зло, которое мешает развитию самой области ИИ и создает отчуждающее отношение к человеческому труду. Если говорить о сфере образования, то образовательный процесс начинает восприниматься как полностью алгоритмизированный, возникает ощущение, что межсубъектное взаимодействие можно без потерь заменить на взаимодействие учащегося с квазисубъектностью искусственной системы.

Есть большие сомнения в том, что полноценный общий искусственный интеллект вообще может быть создан на основе тех подходов, которые применяются в настоящее время. Здесь словно бы сохраняется дуалистическое, но в то же время функционалистское ощущение, что можно взять интеллектуальные функции и во всей их полноте перенести с естественного носителя на другой. Но люди как естественные носители интеллектуальных функций являются социобиологическими существами, обладающими телесной структурой и культурным опытом, и «отделить» интеллектуальные функции от других компонентов человеческого существа полностью невозможно. Вероятно, если бы нам удалось создать новое, совершенно искусственное и в то же время живое существо, оно бы обладало интеллектом, но такой интеллект был бы для нас непредсказуем и вряд ли стал бы просто служить нашим целям.

Тут сразу возникает вопрос о том, до какой степени связаны друг с другом интеллект и свобода воли. Нужна ли нам интеллектуальная система, которая будет обладать, предположим, целеполаганием и при этом свободой воли? Это сразу сделает применение такой интеллектуальной системы столь же сложной этической проблемой, как и, например, использование рабского труда. Еще и с этой точки зрения должны быть дифференцированы две исследовательские области: область, где мы работаем с моделями мышления и интеллекта, и прикладные области. Понятно, что какие-то разработки будут переходить из одной области в другую. Работа в данном случае должна идти в направлении более эффективной интеграции наших интеллектуальных способностей и функционала интеллектуальных систем. Допустим, если взять интеллектуальный анализ данных, то почему бы философу его не использовать для анализа, например, интеллектуальных трендов? Такой симбиоз естественного интеллекта и интеллектуальной системы был бы полезен во многих профессиональных областях.

В то же время идея замены естественного интеллекта искусственным звучит слишком гипотетически, техноутопически и антигуманистически.

Труфанова Е.О.: У меня вопрос по поводу замены преподавателей искусственным интеллектом. Я пока с этим не сталкивалась, скажите, в каких сферах это предполагается применять? Где это разрабатывается?

Алексеева Е.А.: Например, в Китае есть система, которая тренирует математические навыки, она называется Squirrel AI Learning. Для чего нужна подобная система? Идея таких систем ведет начало еще от знаменитого Блума, который продемонстрировал, что индивидуализация обучения дает большую эффективность по сравнению с обучением по общей для всех программе. Интеллектуальная обучающая система может индивидуально работать с учащимся, адаптируя задания под его уровень освоения. Для «натаскивания» на определенные навыки такая система действительно довольно эффективна. Китайский проект как раз направлен на тренировку навыков решения математических задач. Проблема в том, что разработчики обучающих интеллектуальных систем склонны экстраполировать удачный опыт в определенной области на все образование в целом, не учитывая огромного разнообразия образовательного опыта.

Если же говорить об антропоморфных системах, таких как Vina48, то такие роботы обладают так называемым эффектом «зловещей долины», когда слишком антропоморфный нечеловеческий объект кажется нам отталкивающим. Получается, что этот эффект – еще одно из препятствий к повсеместному внедрению роботов в образование, хотя именно на антропоморфные технологии в сфере образования возлагаются определенные надежды.

Лекторский В.А.: В Академии Образования я возглавляю комиссию по чтению. Оказывается, электронное чтение – это не то же самое, что чтение на бумажном носителе – там работают другие психологические механизмы. Не случайно известный Стив Джобс запрещал своим детям пользоваться гаджетами. В хороших школах в США гаджетами пользоваться нельзя, а в школы похуже – пожалуйста, лезьте в Интернет, сколько хотите. В связи с этим вспоминаю одну историю: в одном университете записали лекцию одного преподавателя, собрали студентов, поставили на кафедру вместо преподавателя какой-то гаджет, на котором идет лекция преподавателя. Потом однажды преподаватель зашел в эту аудиторию и увидел такую картину: на кафедре стоит гаджет с его лекцией, а вместо студентов – другие гаджеты, которые записывают эту лекцию. Люди вообще не нужны – один гаджет общается с другим.

Труфанова Е.О.: Это называется «Интернет вещей».

Лекторский В.А.: Это называется «...и звезда с звездой говорит». Они говорят друг с другом, а человек здесь уже лишний. Еще одна важная идея, которая прозвучала у Екатерины Алексеевны, – это различие между ИИ в точном смысле слова и интеллектуальными системами. Это имеет смысл, хотя, если вы возьмете любую западную энциклопедию, даже и Стэнфордскую философскую энциклопедию, там такой разницы вы не найдете. В понятие Artificial intelligence входит все, даже то, что мы называем интеллектуальными системами. Считается, что это одно и то же!

Катунин А.В.: Хочу поделиться наблюдениями насчет высказывания «Вкалывают роботы, счастлив человек». Когда в этом календарном году мы весной

впервые перешли на дистанционное образование, используя Zoom и все, что попадалось под руку из онлайн-технологий, я начал звонить своим бывшим студентам, которые учились уже кто в магистратуре, кто в аспирантуре, чтобы посоветоваться по ряду вопросов. И одним таким вопросом был вопрос об онлайн-образовании и дистанционном формате, который уже третью или четвертую неделю был в самом разгаре во всех ВУЗах. Вот какую проблему я обнаружил: существует конкретный запрос, рожденный отсутствием необходимости присутствовать на лекциях непосредственно в аудиториях, смотреть лекции в записях в удобное время, а не присутствовать онлайн в Zoom в назначенное время. Какова аргументация: «Если мы так или иначе не привязаны к университету физически, почему мы должны быть привязаны к компьютеру в конкретное время? Было бы лучше посмотреть записи видеолекций в удобное время, а роль преподавателя – консультировать, отвечать на вопросы, проверять задания (и если лекции будут в записи, то времени у лектора будет больше)». Транслирование такого взгляда на дистанционный процесс обучения – далеко не единичный случай. Поэтому не знаю, как насчет замены преподавателей роботами, но вот нашими же «копиями» нас в скором времени вполне могут заменить. Тем более во многих дисциплинах, связанных с историей, скорость устаревания данных не так велика. Это в экономике, юриспруденции, финансах лекция, прочитанная месяц назад, может быть уже неактуальной. В философии, на мой взгляд, срок годности одной лекции (в зависимости от темы) может исчисляться десятилетиями. Как бы живые преподаватели не стали совсем ненужными.

Лекторский В.А.: Может, и студентов записями заменить? Давайте всех записями заменим.

Катунин А.В.: Запись лекции, запись ответа на экзамен.

Щедрина И.О.: К словам Александра, что в момент карантина стала популярна мысль о замене преподавателя видеолекциями и т.д. Я хочу сказать, что карантин эту тему скорее обострил и усилил акцент на ней, потому что это стало проблемой еще несколько лет назад. Сотрудник Института философии Николай Афанасов преподавал в Высшей школе экономики и несколько лет назад опубликовал статью-размышление на тему того, насколько сильно влияет идея записи лекций и замена настоящего взаимодействия преподаватель – студент на аудиторию. Там был очень характерный пример о том, как он спросил у девушки, в какое время она слушает лекции, поскольку все слушают тогда, когда им было удобно. И ответом было: «Когда есть секундочка, когда крашу ногти и могу краем уха что-то слушать». Момент эффективности взаимодействия лектора и аудитории очень сильно проседает.

Алексеева Е.А.: Меня как человека, который поддерживает идею электронного и дистанционного обучения, очень возмущают действия Высшей школы экономики, потому что они профанируют саму идею дистанционного образования. Так считаю не только я, но и представители профессиональных сообществ, занимающихся вопросами электронного и дистанционного обучения. То, что предлагал ректор ВШЭ Я.И. Кузьминов, – это реализация цифровыми средствами крайне архаичной модели обучения, предполагающей механический перенос знаний из головы преподавателя в голову студента.

Лекторский В.А.: Дорогие друзья, спасибо! Наши специалисты по когнитивной нейронауке, я ссылаюсь на К.В. Анохина, Т.В. Черниговскую, установили такой факт: оказывается, чтобы было реальное взаимодействие между людьми, очень важен зрительный контакт. А когда у вас идет запись лекции, то преподаватель вас не видит и вы, по существу, не видите его тоже. Преподавание – это общение, эмоциональный контакт.

Труфанова Е.О.: Спасибо, мне кажется, очень важные темы подняла Екатерина Алексеевна, и мне бы очень хотелось тоже присоединиться к обсуждению цифровизации в образовании, но мне кажется в связи с ситуацией, в которой мы все оказались – дистанционного преподавания, – эта тема вызывает столько живых эмоций, что трудно обсуждать ее отвлеченно, на должном теоретическом уровне. Поэтому я все же обращусь к другой теме.

Екатерина Алексеевна подчеркнула важный момент – нужно разводить исследования в области искусственного интеллекта, направленные на построение моделей мышления, и исследования искусственных интеллектуальных систем. Это действительно два абсолютно разных вопроса, и, конечно, с философской точки зрения намного более интересен первый вопрос. Я хочу обратиться снова к знаменитой статье Алана Тьюринга и подчеркнуть, что проблема, которую он в ней обсуждает, связана именно с этим первым вопросом. Хотя единственным критерием для оценки «интеллектуальности» машины он предлагает рассматривать возможность выигрыша в «игре в имитацию» (прохождение «теста Тьюринга»), тем не менее рассматриваемые им различные контраргументы касаются прежде всего именно того, почему искусственный интеллект не может быть естественным интеллектом, а на самом деле – почему он не может быть *человеком*: у него нет души, он не переживает эмоционально, у него нет творческого начала и способности к саморазвитию и даже – самый странный контраргумент – единственный, который Тьюринг не может опровергнуть, – искусственный интеллект не способен на экстрасенсорное восприятие.

Лекторский В.А.: Тьюринг верил в экстрасенсорное восприятие.

Труфанова Е.О.: Да, совершенно верно. В середине XX в. многие серьезно к этому относились, проводили исследования экстрасенсорники, которые, как мы сейчас знаем, правда, ни к чему не привели. Так вот, Тьюринг решает вопрос не о том, может ли машина решать интеллектуальные задачи, он решает вопрос о том, может ли машина обладать *человеческим сознанием* или хотя бы достаточно успешно его «имитировать».

В результате популяризации обсуждений искусственного интеллекта вся вторая половина XX в. в массовой культуре проходит в страхе и ожидании «восстания машин»: искусственный интеллект, поработивший человека, является одним из наиболее расхожих сюжетов в научно-фантастической литературе и кинематографе – самым ярким примером я бы назвала не разрекламированную «Матрицу», а HAL9000 из «Космической одиссеи 2001 года» А. Кларка, блестяще воплощенный в фильме С. Кубрика (если честно, фильм и следовало бы ограничить этим сюжетом). Эти страхи связаны именно с ожиданием, что искусственный интеллект станет-таки самостоятельным, независимым от своего создателя субъектом и, обладая значительно большими «мощностями» (скорость реакции, объем памяти и т.д.), способен будет с легкостью победить

человека и установить свою машинную цивилизацию взамен человеческой. Однако вместо развития по пути «сильного» искусственного интеллекта, как показала практика, развивался прежде всего «слабый» искусственный интеллект, решающий сугубо прикладные задачи.

В итоге искусственный интеллект перестает быть частью повестки угрозы и начинает выполнять различные вспомогательные функции в жизни человека. Искусственные интеллектуальные системы все больше внедряются в нашу жизнь в качестве разнообразных «помощников», которые, предположительно, должны облегчить нашу жизнь, избавляя нас от каких-то рутинных действий, решая за нас задачи, которые ранее мы решали сами (например, существуют программы, не только подсчитывающие наши ежемесячные расходы и их структуру, но и сами решающие, какую часть денег мы можем позволить себе потратить, какую сложить в сбережения, а какую, например, вложить в акции). С этой точки зрения мы можем рассматривать обычный калькулятор также как разновидность искусственного интеллекта, способного только на один набор функций – осуществление математических операций, которое мы «выносим» из нашего ума и передаем машине.

И вот здесь возникает важный вопрос – Владислав Александрович во вступительном слове уже намекнул на него – не становимся ли мы рабами искусственного интеллекта в совершенно ином смысле? Не приводит ли чрезмерное распространение подобных «интеллектуальных ассистентов» к нашей интеллектуальной деградации? Могу сказать по себе, зная, что у меня под рукой всегда есть калькулятор, я с большим трудом могу заставить себя подсчитать цифры в уме или «в столбик» – отсутствие подобной практики приводит к утрате навыка. Чем больше появляется подобных «ассистентов», тем больше мы «отдаем» им на откуп и теряем сами. Я не склонна здесь оценивать это однозначно как отрицательное явление. Вполне возможно, что это является естественным результатом эволюции человека в техносреде, и искусственный интеллект становится неизбежным «дополнением» или «расширением» интеллекта естественного. Но вопрос состоит еще вот в чем: освобождает ли применение искусственного интеллекта человека от рутины, давая возможность ему реализовываться в творческом труде (недоступном для искусственного интеллекта)? Или это сращивание «отупляет» человека и он утрачивает часть своих интеллектуальных возможностей, становится зависим от машинного искусственного интеллекта? И не ведет ли такое делегирование ряда интеллектуальных задач к снижению творческого потенциала самого человека?

Критерий способности на творчество, как мне кажется, является одним из важнейших, когда мы обсуждаем специфику естественного и искусственного интеллектов. Помню, несколько лет назад была издана книга, написанная искусственным интеллектом, – по-моему даже это был любовный роман. Книга со связным сюжетом, ее можно было читать, сопереживать героям и так далее. Так что же, искусственный интеллект, выходит, способен на творчество? Мне кажется, что такой вывод сделать нельзя. Произведения искусства, создаваемые искусственным интеллектом, – их можно слушать, и читать, и любоваться ими, и они будут вызывать у нас эмоции, потому что в них используются «проверенные» ходы, штампы, которые нацелены на вызывание определенных эмоций. Например, как мне кажется, все «мыльные оперы» давно мог бы писать

искусственный интеллект, и их продолжали бы смотреть, ничего не замечая, поскольку все они построены на штампах, которые являются определенными эмоциональными «триггерами». Например, если я случайно вижу на экране сцену, в которой девушка льет слезы над безвременно погибшим женихом, меня это все равно эмоционально «цепляет», даже если сценарий сериала самый что ни на есть низкопробный. Помимо этого, наш предшествующий опыт и культурные установки таковы, что они уже заранее нацелены на то, что искать в тексте, в музыкальном произведении или, к примеру, в картине. В некоторых случаях мы и в самом деле не сможем различить, что перед нами – картина известного художника-абстракциониста или бессмысленная хаотичная (а может, наоборот – хорошо просчитанная) мазня искусственного интеллекта. И, возможно, в данном случае только наша социальная установка на известность автора первой картины заставляет нас предпочесть ее второй. В любом случае создание художественных произведений искусственным интеллектом будет не более чем компиляцией известного набора литературных или, к примеру, музыкальных «ходов». Правду сказать, произведения современной массовой культуры представляют собой в основном то же самое, хотя и создаются с помощью естественного интеллекта – так что, возможно, в этом плане разница не так уж и велика, и тогда не будет большой потерей, если искусственный интеллект начнет штамповать «модные» песни и бульварное чтиво.

Подводя итоги своему краткому выступлению, хочу отметить, что, как мне кажется, во-первых, если мы рассматриваем искусственный интеллект как способ постижения человеческого сознания, то мы можем говорить лишь о том, что исследования искусственного интеллекта могут позволить нам формировать модели мышления, но не сознания в целом – а это совершенно разные объекты исследований. Во-вторых, перспективы искусственного интеллекта намного более обширны, если мы говорим именно о интеллектуальных системах, являющихся вспомогательными инструментами человеческой деятельности. Однако здесь остается вопрос, о котором я говорила ранее, – как провести грань между искусственным интеллектом в качестве помощника в моей деятельности и искусственным интеллектом, который начинает с помощью своих подсказок полностью мной руководить?

Меркулова И.Г.: Я бы хотела сказать небольшую реплику с позиции лингвиста и семиотика, я здесь представляю взгляд по большей части филологических наук. Мне очень понравилось замечание Елены Олеговны о том, что опасность нашей порабощенности искусственным интеллектом – это, конечно, не образ фантастических романов XX в. и страшного существа, которое нас поработит, а это проблема жизни по подсказке – жить, ориентируясь на то, что нам скажут. Проблема состоит в том, что подсказка будет основана на том, что у нас уже есть. Нам предложат как бы новое, ориентируясь на наши интересы, но ИИ не предложит кардинально нового, того, что у нас нет внутри, что мы еще не выучили. Мне кажется, что это как раз то, о чем в свое время писал семиотик Ю.М. Лотман, когда он говорил, что реальная коммуникация – это не передача информации от А и В в константном объеме, а она возникает именно в недосказанности, непонимании, и именно тогда появляется то, что называется переводом в широком смысле слова. Как говорил мой испанский

коллега-семиотик Хорхе Лацано, даже общение отца с сыном – это тоже перевод с языка на язык. Поэтому вопрос перевода нашего языка и языка ИИ – это те самые лакуны, недосказанности, на которые нужно обращать внимание, чтобы не жить по подсказке.

Труфанова Е.О.: Спасибо Вам большое за очень ценный комментарий. Действительно, чего нет у искусственного интеллекта – нет возможности учитывать и использовать эти недосказанности, он буквален в своих коммуникативных возможностях.

Лекторский В.А.: Спасибо, Елена Олеговна поставила ряд важных вопросов, которые можно долго обсуждать. Конечно, ИИ в том виде, в котором он есть сейчас, в своих творческих способностях имитирует человека. Вы знаете, что машина написала музыку в стиле И. Баха? Оттого, что машина может делать что-то очень быстро, может обрабатывать огромный массив информации, который мы обрабатывать не можем, получается нечто такое необычное для человека. Я вижу одну из угроз в том, что сейчас появилась машинная обработка больших данных – это проблема машинного обучения. Машина на основании обработки огромного массива информации выдает такие прогнозы, которые человек не мог бы сам выдать никогда. Кроме того, человек не может понять, почему машина это делает, обычно мы не так работаем и делаем предсказания. Человек попадает в зависимость от такого рода действий ИИ. Проблема в том, как не допустить вытеснения человека и превращения его в некий придаток.

Щедрина И.О.: Тематику искусственного интеллекта и роботизации мне хотелось бы затронуть в контексте исследований автобиографической памяти и автобиографического нарратива. Дело в том, что круг вопросов, связанных с идентичностью человека, сегодня очень интенсивно обсуждается в связи с анализом интеллектуального потенциала роботов и ИИ в целом. Это, в общем, новые аспекты в традиционной философской и психолого-педагогической тематике самоидентификации человека. И при этом, полагаю, очень важно иметь в виду, что представление о себе и обществе у человека в значительной мере базируется на автобиографической памяти, и ключевой здесь становится именно когнитивная составляющая самоидентификации. Во всяком случае, особое значение имеет то обстоятельство, что каждый человек представляет себе происходящие в его окружении события с позиций непосредственного когнитивного опыта. Крэйг Каллендер (профессор Калифорнийского университета в Сан-Диего, специалист по философии науки) в вышедшей в прошлом году работе “Time Lost, Time Regained” подчеркивает, что человеческое «Я» на эпистемологическом и психологическом уровне создается из непосредственного опыта и воспоминаний. «Я» формируется из того, что человек полагает в себе непреходящим, неизменным и присутствующим целостно и постоянно в каждый момент. И все это, замечу, так или иначе фиксируется в автобиографическом нарративе. В некотором смысле эта идея перекликается с локковским пониманием идентичности как психологической непрерывности, а именно эта непрерывность – во времени и в сознании – становится проблемным пунктом в вопросе об искусственном интеллекте и потенциале робототехники.

В 2008 Александр Р. Прусс (профессор Бэйлорского университета, специалист в области философии математики, аналитической философии и метафизики), выступив на конференции «Наука и человеческая природа: российская и западная перспектива» с докладом «Искусственный интеллект и тождество личности», приводит в качестве примера мысленный эксперимент, построенный вокруг робота по имени Робби. Прусс задает множество вопросов: будет ли Робби мыслить и рационально действовать или будет лишь казаться, что это так; будет ли он существовать в выключенном состоянии, будет ли считаться его многозадачность критериями существования разных электронных личностей в нем (два ядра в процессоре, каждое из которых выполняет свою функцию, по аналогии с полушариями мозга), а главное – параллельно те же вопросы он ставит и в отношении людей: «Параллельно вопросу о том, продолжает ли Робби существовать в выключенном состоянии, мы можем спросить, продолжают ли люди существовать в состоянии комы. Параллельно вопросу о количестве электронных личностей мы можем спросить о количестве человеческих личностей – подумайте, к примеру, о сиамских близнецах. И параллельно вопросу, восстановлены ли данные Робби на двух компьютерах, мы можем представить фантастический мысленный эксперимент, при котором мой мозг разделен пополам и две его половинки помещены в разные тела – где я окажусь и окажусь ли где-либо?».

Аналогичную проблематику ставит в центр исследований персональной самоидентификации человека известный философ сознания Д. Парфит. В работе «Причины и личности» он описывает мысленный эксперимент, который позже станет известен как «парадокс телепортации». Идея о возможности переноса сознания на другой носитель, а также связанные с этим вопросы влекут за собой тему моделирования идентичности; опираясь на мысль Парфита, можно задаться вопросом: каким именно образом происходит моделирование идентичности и как это можно спроецировать на нарративную реальность? Исследования, которые проводятся сегодня в этой области, как мне кажется, позволяют ответить на этот вопрос: именно история, которую человек рассказывает о себе, в значительной мере формирует его идентичность. При этом автор нарратива опирается на собственную автобиографическую память, в таком случае когнитивное переключается с экзистенциальным, и именно таким образом нарратив соотносится с внешним опытом. Современные исследователи все чаще обращают на это внимание. Так, например, в своих работах Мак-Адамс связывает автобиографический нарратив, автобиографическую память, их соотношение, погружая его в социокультурный и исторический контексты. Граци и Файвиш рассуждают в схожем ключе, смещая акцент к культурному и социальному; определение автобиографической памяти в их исследованиях звучит как сложная смесь воспоминаний о событиях, которые являются единичными и протяженными во времени и которые оказываются интегрированы в связную историю личности. Эта история, автобиографический нарратив, создается и оценивается с привлечением ряда социокультурных практик. Фактически социокультурная составляющая становится обязательным условием внутреннего самоопределения в рамках построения автобиографического нарратива.

Полагаю, важно отметить также еще один аспект проблемы автобиографической памяти в контексте искусственного интеллекта: это временной характер человеческой когнитивности. И здесь можно обратиться к понятию автобиографических воспоминаний, предложенному исследователем феномена автобиографической памяти М. Конвеем. В его работах автобиографические воспоминания определяются как цельный комплекс ментальных авторепрезентаций, строящихся на «автобиографическом знании» (знании человека о себе самом). Эти знания можно сравнить с базой хранящихся данных; именно они становятся главными источниками идентичности, по версии Конвея, а также обеспечивают связь между историей личности (индивидуальной, персональной) и историей человека, интегрированного в сообщество. И чем больше развивается человек, растет и проводит время в обществе, обучаясь и в дальнейшем взаимодействуя, тем более целостным и осмысленным становится его представление о себе самом, о собственном «Я».

И в связи со сказанным о современном состоянии релевантных исследований весьма перспективной и не в полной мере еще использованной представляется идея Тьюринга – привлечь потенциал игры-эксперимента для оценки способности ИИ к самоидентификации. На мой взгляд, данная мысль открывает новые перспективы исследования этой способности и у человека, и у робота в контексте анализа автобиографического нарратива. Задавшись вопросом о том, существуют ли такие машины, которые могли бы ввести в заблуждение человека, Тьюринг фактически поднял аналитический вопрос о создании ИИ и предметном содержании мышления на экзистенциальном уровне. Более того, предлагаемая им игра-эксперимент по самим формулировкам своих вопросов ставит на один уровень человека и машину.

В статье «Вычислительные машины и разум» Тьюринг отмечает: «Пытаясь имитировать ум взрослого человека, мы вынуждены много размышлять о том процессе, в результате которого человеческий интеллект достиг своего нынешнего состояния. Мы можем выделить три компонента: 1) первоначальное состояние ума, скажем, в момент рождения; 2) воспитание, объектом которого он был; 3) другого рода опыт, воздействовавший на ум, – опыт, который нельзя назвать воспитанием». Таким образом, Тьюринг демонстрирует, что человеческое поведение во многом задается социумом (вырабатываются привычки, реакции, опыт взаимодействия), т.е. фактически очень многое в индивидуальности оказывается социокультурно обусловленным. Для того чтобы выиграть в имитацию, запрограммированная машина должна учитывать этот момент.

Говоря о возможности для машины развиваться до уровня взрослого человека, Тьюринг выделяет две задачи: построить «программу-ребенка» и осуществить процесс воспитания. В некотором смысле это связано с недавними исследованиями о природе человеческого Я и о влиянии автобиографической памяти на развитие, если можно так сказать, Я у робота. В статье “The Role of Autobiographical Memory in the Development of a Robot Self”, представленной на Международной конференции по интерактивной коммуникации между роботами и людьми, Г. Пуанто и Ф. Домини обратились именно к этой теме. Их исследование содержало разработки в области когнитивного развития Я; кроме того, Пуанто и Домини рассматривают современные работы в области

развивающейся робототехники, в которых предпринималась попытка проследить части траектории развития личности, причем именно с точки зрения психологии развития и робототехники. В качестве отправной точки ученые обратились к процессу человеческого социального взаимодействия – сотрудничеству (определив его как вступление в совместное предприятие для достижения общей цели). В основе способности к сотрудничеству лежит способность вступать в отношения. То есть, по мнению авторов, если мы хотим, чтобы роботы сотрудничали с людьми, то и роботы до некоторой степени должны участвовать в этих отношениях между собой и людьми, и, следовательно, у них должен возникать некоторый аспект Я.

Десятилетия исследований в области когнитивного развития человека показывают, что личность изначально присутствует в человеке не полностью, а скорее развивается на основе прожитого опыта, то есть посредством взаимодействия с миром, включая физический мир и социальный мир.

Г. Пуанто и Ф. Домини обращаются к предложенной У. Найссером системе организации Я, включающей в себя пять составляющих типов: экологическое (нахождение в непосредственной физической среде и воздействие на нее), межличностное (вовлечение в социальное взаимодействие с Другим), Я-концепция (ментальное представление о собственных, более или менее постоянных характеристиках), длящееся во времени (собственная история жизни) и частное (понимание конфиденциальности сознательного опыта). Исследователи при этом подчеркнули экологическую природу каждого из этих уровней, то, как они развертываются в результате взаимодействия развивающегося ребенка с физическим и межличностным миром. И акцент ставился на то, что развитие личности ребенка зависит в том числе и от его автобиографической памяти. С точки зрения развивающейся робототехники это позволяет предположить, что в принципе возможно развитие определенных аспектов и в когнитивной системе робота, когда робот, оснащенный некоторой системой автобиографической памяти, вовлечен и взаимодействует в физическом и социальном мире.

Самообучающиеся системы – это, в принципе, не новость в науке (например, самоорганизующаяся карта Кохонена), но смещение акцента происходит именно к автобиографической части – и к трансляции полученного опыта, т.е. вербализации, «пропущенной через себя». Уже существует ряд работ по автоматическому созданию роботами нарративов (о собственном опыте) на естественном языке – некоторый отчет из запланированных и уже выполненных действий – та самая вербализация, необходимая для облегчения взаимодействия человека и робота (например, Розенталь, Селвараж и Велозо предложили для описания опыта навигации у робота модель, преобразующую пройденный маршрут в естественный язык, базируясь на зафиксированном опыте).

В целом обращение к наррации, автобиографической памяти и трансляции опыта в робототехнике пока происходит на уровне анализа и обработки естественного языка, однако все чаще я вижу статьи, посвященные теме повествовательного интеллекта – как способности создавать, рассказывать, понимать и эмоционально реагировать на истории. Исследования в области вычислительного нарративного интеллекта направлены на привнесение этого, соответственно, в машину. А пока – исследователи стараются, по мере научных

сил, продолжить работу над искусственным интеллектом и развитием робототехники в направлении, релевантном теме самопознания человека.

Меркулова И.Г.: Мне хотелось бы в своем выступлении представить обзор некоторых ключевых публикаций ЮНЕСКО по теме нашего круглого стола. Тема влияния искусственного интеллекта (ИИ) на общественную жизнь – одна из центральных в материалах Организации за последние два года. Искусственному интеллекту был посвящен специальный номер «Курьера ЮНЕСКО» 2018 г., озаглавленный «Искусственный интеллект: перспективы и угрозы», на сайте Организации был создан специальный раздел, освещающий мероприятия по данной теме в разных странах мира: в Европе, Африке, Америке, Азии. По словам генерального директора Одри Азуле, человечество обязано «провести всемирное и просвещенное обсуждение этой проблемы с тем, чтобы встретить новую эпоху с открытыми глазами, не жертвуя нашими ценностями, и привести к общему знаменателю наши моральные принципы». В итоговом документе специальной группы экспертов по подготовке проекта рекомендации об этических аспектах ИИ обращается внимание на то, что технологии ИИ способны принести человечеству огромную пользу, но при этом ставят перед ним фундаментальные вопросы этического порядка, касающиеся предвзятости, которую такие технологии могут порождать, что потенциально может вести к неравенству и маргинализации.

Предваряя новую публикацию «Курьера» 2020 г., заместитель Генерального директора по культуре Эрнесто Оттоне Рамирес отметил, что пандемия «поставила на паузу жизнь во всем мире», однако ключевые проблемы, такие как этические аспекты современных технологий и незаконный оборот культурных ценностей, по-прежнему актуальны.

С этой точки зрения особый интерес представляют размышления французских ученых Од Бернхайм и Флоры Венсан, напечатанные в этом номере в рубрике «Интервью». Исследовательницы считают, что «искусственный интеллект нужно обучать» (именно так озаглавлена статья), поскольку, как это ни странно на первый взгляд, алгоритмы могут быть сексистскими. Их разрабатывают люди, а значит, в них могут отражаться наши расовые и гендерные предрассудки. В то же время алгоритмы ИИ могут и содействовать равенству мужчин и женщин, если этому вопросу уделять должное внимание.

Корни проблемы кроются в истории: в среде ученых не хватает разнообразия. Флора Венсан считает, что гендерный вопрос мог бы лучше учитываться уже на ранних стадиях научных исследований, если бы в составе научных групп был больший процент женщин. В противном случае возникают примеры дискриминации: например, доклинические исследования лекарственных препаратов проводятся не на самках, а на самцах крыс, так как у них вырабатывается меньше гормонов и таким образом легче измерить побочные эффекты. Сегодня в области исследований ИИ работают 22% женщин от общего числа ученых, а среди 624 лауреатов Нобелевской премии в области естественных наук количество представительниц прекрасного пола составляет 23: на эти факты указала Одри Азуле в видеоприветствии Форуму «Женщины в науке», прошедшему онлайн 1 декабря 2020 г. Между тем область информатики, например, не всегда считалась чисто мужской профессией. Од Бернхайм

приводит сравнительно недавнюю статистику по США: в 1984 г. среди работников этой отрасли было 37% женщин. А уже в 2018 г. в Европе, во Франции, доля девушек-студенток, изучающих информатику, составляла 10%, и только 12% тех, кто занимался ИИ. Этот результат – культурно обусловленный: с 80-х гг. прошлого века шло развитие компьютерных игр, а вместе с ними и возникновение стереотипа, что и игры, и персональный компьютер скорее ориентированы на мальчиков и мужчин. Могу привести пример из личного опыта: несколько лет назад во время учебы в магистратуре Сорбонны по специальности «Лингвистическая инженерия и информационное общество (Лингвистика-Информатика)» я была одной из немногих девушек в группе: их процент составлял одну четвертую от общего числа студентов отделения.

По своей природе алгоритмы должны быть нейтральными. Однако в действительности оказывается, что они перенимают человеческие и гендерные предрассудки, и искусственный интеллект их успешно тиражирует. Так, известные голосовые помощники носят женские имена (Сири, Алекса, Кортана), говорят женским голосом и демонстрируют рабскую покорность в ответ на команды и даже оскорбления. «Я бы покраснела, если бы могла» (“I’d blush if I could”), – так несколько лет подряд голосовая помощница Сири отвечала на адресованные ей нелицеприятные высказывания. Данная фраза даже была вынесена в заглавие одной из публикаций ЮНЕСКО по гендерным вопросам. Ситуация не должна вызывать удивления, ведь большинство разработчиков программного обеспечения – мужчины, женщин среди них – 3%. Поэтому ЮНЕСКО сформировала ряд рекомендаций, направленных на борьбу с предрассудками. Среди прочего: призывать не наделять голосовых помощников по умолчанию женским голосом и программировать их таким образом, чтобы не поощрять высказывания сексистского характера. И, разумеется, поддерживать обучение девочек и девушек техническим специальностям, чтобы в дальнейшем они смогли разрабатывать новые технологии наравне с мужчинами.

Другим примером гендерных предрассудков в области ИИ являются программы автоматического перевода текста: названия профессий переводятся в определенном роде, даже если указание на него в оригинале отсутствует. Слово «врач» чаще всего переводится в мужском роде, а «медсестра» в женском. Это касается и алгоритмов распознавания лиц: по данным афроамериканской исследовательницы Джой Буоламвини, обучение таких алгоритмов велось преимущественно по фотографиям белых мужчин. В результате системы намного хуже распознавали лица азиатов и чернокожих женщин. Вероятно, картина могла бы быть иной, если бы указанная исследовательница находилась среди разработчиков.

По мнению Од Бернхайм и Флоры Венсан, алгоритм появляется в результате целой цепочки причинно-следственных связей: те, кто его разрабатывает, кодирует, тестирует и использует, чаще всего бессознательно воспроизводят предрассудки и стереотипы внутри себя. То есть разработчики не стремятся к дискриминации намеренно, но они являются заложниками собственных представлений о жизни и культуре.

Как сделать так, чтобы алгоритмы стали непредвзятыми? Прежде всего расширить базы данных, которые используются для обучения машин (с учетом

пола, телосложения, национального признака). Не так давно появились специализированные компании, разрабатывающие инструменты для оценки степени предвзятости алгоритмов. Ученые также предлагают создать единый гендерно-нейтральный код, по аналогии с гендерно-нейтральным языком. Одна из интересных и перспективных инициатив – так называемая «Клятва Гиппократова для специалистов по анализу данных», или морально-этическая хартия, где дается перечень параметров, которые следует проверить до введения алгоритма на практике, чтобы быть уверенными, что он не носит дискриминационный характер.

Общество осознало, что ИИ может усиливать гендерные предрассудки, и одновременно с этим пришло понимание, что изменения, которые эта технология несет в общество, могут быть также положительными. Так, ИИ анализирует огромные объемы данных, и в 2016 г. компания Google разработала алгоритм по подсчету времени, отведенного разговору женщин в популярных фильмах, на примере США. Оказалось, что женщинам отведено меньше экранного времени, чем мужчинам. Влияние данного факта на художественные достоинства кинопроизведений еще предстоит изучить, а пока нужно стремиться, чтобы алгоритмы могли разрабатывать как специалисты-мужчины, так и программисты-женщины. Для этого им тоже предстоит борьба с собственными внутренними предрассудками и страхом «неженских» профессий.

На Генеральной конференции ЮНЕСКО в 2021 г. будут обсуждаться рекомендации об этических аспектах ИИ с целью заполнить законодательный вакуум вокруг его разработки и применения. Этические и гендерные аспекты обсуждаются с представителями научных сообществ 193 стран. Среди новых возникших вопросов: как обеспечить развитие инклюзивности и разнообразия в области ИИ? Как пандемия может увеличить риск использования ИИ неэтичными способами? Хочется надеяться, что в постковидном мире «новой нормальности» этим проблемам будет отведено должное внимание.

Закончить мне хотелось бы рассуждением из области смежных наук – семиотики и лингвистики. Показатель «зрелости» знаковой системы, в том числе языка, – ее способность к самоописанию, к рефлексии. Тот факт, что человеческое общество уже не довольствуется принципом использования и сосуществования с ИИ, а стремится к выработке описывающих его этических критериев, открывает перед нами перспективы устойчивого прогресса на благо человечества.

Лекторский В.А.: Инна Геннадьевна, большое Вам спасибо. Я не знал, что гендерные предрассудки есть и в области ИИ. Видимо, они есть везде. Вчера я читал одну статью, где была высказана интересная мысль: кто создавал логику? Белые мужчины. Тут сказались гендерные и расовые предрассудки, поэтому логику необходимо менять.

Катунин А.В.: Согласен с тем, что в наши дни существует некоторое размывание понятия «искусственный интеллект»: что именно в конечном итоге мы подразумеваем, когда употребляем это понятие? В разных контекстах значение может сильно варьироваться. Мое выступление будет посвящено этическим аспектам использования ИИ. В свое время мне очень понравились слова французского математика Э. Галуа: «К несчастью, немногие отдадут себе отчет

в том, что самая ценная книга истинного ученого – та, где он откровенно заявляет, что именно ему неизвестно». Так и я сегодня, скорее всего, задам больше вопросов, нежели дам ответов. Хочу поговорить о повседневном включении в жизнь каждого человека технологий, работающих с использованием ИИ.

С одной стороны, на ум в первую очередь приходит такой пример взаимодействия человека и ИИ, как автопилот в автомобиле. В таком контексте известный многим мысленный эксперимент с неуправляемой вагонеткой перестает быть мысленным экспериментом и становится задачей, требующей решения в реальной жизни. Автомобиль, управляемый ИИ, должен принимать решение в экстраординарной ситуации. И если вдруг из такой ситуации невозможен выход без жертв, то как выбирать, какой из исходов будет лучше? Кто это будет решать и программировать? На основании какой логики? Будоражащие сознание фантастические сюжеты о войне человека и разумных машин уже отражаются на заголовках прессы: в поисковике находится статья под заголовком: «Искусственный интеллект убил первого человека». Если открыть саму статью, то заголовок уже более мягкий: «Автопилот Tesla убил первого человека». Или ироничные высказывания на тему того, что война машин уже началась после случая наезда Теслы на антропоморфного робота в Лас-Вегасе.

С другой стороны, человек, использующий смартфоны и компьютеры, сталкивается с так называемым коммерческим шпионажем. Вряд ли для кого-то будет новостью информация о том, что после вашего поиска в Интернете чайника или автомобиля определенной марки вся контекстная реклама для вас будет наполнена соответствующими предложениями. Но кого-то можно удивить информацией о том, что наши мобильные телефоны нас, вероятно, подслушивают. Скорее всего, существуют алгоритмы, применяемые в крупных дата-центрах, которые распознают речь, выделяют из нее ключевые слова и формируют на основании этого рекламные подборки. Несколько примеров из личной жизни: разговаривали с друзьями об услугах стоматологов (разговаривали, не искали в поисковике), уже на следующий день в моем личном аккаунте «Инстаграм» появляется реклама клиник, оказывающих соответствующие услуги. После разговора о необходимости купить миксер на следующий день звонят из крупного онлайн-магазина и предлагают... миксер по акции. Совпадение? Возможно. Но таких совпадений становится все больше.

Лекторский В.А.: Они вас подслушивают и подсматривают. «Большой Брат».

Катунин А.В.: Или еще один пример более показательный: обсуждали в компании мем с котиком. В разговоре неоднократно фигурировало слово «кусь». Поговорили и забыли. Но на следующий день мне стали предлагать рекламу: «печать логотипа вашей компании на...», «изготовление неоновых вывесок вашей компании», и везде на картинках в предлагаемой рекламе торжественно выведено было слово «КУСЬ» (даже скриншоты остались такой креативной рекламы). Понятно, что более таргетированная реклама сильно повышает прибыль компаний. Но кто давал на это разрешение? Этично ли это? Нужен ли такой «интеллектуальный» подход конечному потребителю? При том что (насколько мне известно) ни одна крупная корпорация не подтвердила использование такого рода технологий.

Когда-то программирование перешло от директивности (простой набор операторов в компиляторе. Например, IF..., THEN..., ELSE) к работе с разного уровня сложности абстракциями и самообучению. Как далеко такое самообучение может зайти? И кто будет нести ответственность, если что-то пойдет не так? Адаптируется законодательная база к изменению технологий? Если мы так сильно стремимся из вещи сделать субъект, не превратимся ли мы в конечном итоге сами в объекты?

Труфанова Е.О.: По поводу «автопилота», который убил человека. Хочу еще один рассказ напомнить – рассказ Роберта Шекли «Страж-птица». В нем самообучение механической птицы, чья цель – предотвращать насилие – доходит до того, что птица начинает останавливать любые манипуляции, которые можно считать насилием, от хирургических операций до выключения зажигания у машин: отключил питание – «убил» машину. А дальше люди, пытаясь справиться с этими обезумевшими страж-птицами, создают других роботов-птиц, запрограммированных на уничтожение страж-птиц. Только они тоже выходят из-под контроля, обучившись уничтожать не только страж-птиц, но и других существ, включая человека. Самообучающаяся система все равно имеет пределы обучения, самообучение и развитие системы не является аналогом социализации индивида, оно происходит в жестко ограниченных рамках, оно не может учитывать всех факторов окружающей среды, и именно в связи с этим ситуации, подобные рассказу Шекли, вполне возможны.

Лекторский В.А.: Александр, спасибо за доклад! Действительно, все эти этические проблемы очень сложные. Иногда говорят, что если бы на месте ИИ был человек, то он бы иначе решал эти вопросы. ИИ же бездушный, в ситуации, когда он кого-либо убил, он не испытывает эмоций по этому поводу, поскольку так его запрограммировали. Это то, что в философской литературе называется «моральной дилеммой», когда человек оказывается в ситуации выбора из двух зол. Такие ситуации в жизни возникают не так часто. Приведу пример. В конце Второй мировой войны американцы бомбили Хиросиму и Нагасаки. Гибли невинные люди, а не участники военных действий. Это не было убийство в процессе боевых действий. До сих пор идет большой спор: можно ли оправдывать это? На нынешнем этапе отличие ИИ от человека состоит в том, что человек эмоционален. Так, один из тех летчиков, сбивших японские города, ушел в монастырь, так как глубоко переживал эту ситуацию. В случае с машиной – как вы машину запрограммируете, так она и будет себя вести.

Мне пришлось писать статью о цифровой памяти. За нами наблюдают, машины записывают многие наши действия. А в некоторых странах за людьми следят: куда они пошли, с кем говорили, что покупали и т.д. На основании этого вас оценивают. Вы можете забыть что-то, машина же помнит все. У человека есть право на забвение, чего нет у машины.

Яковлева А.Ф.: Я бы добавила немного о другом аспекте поднятой проблемы этики ИИ, в частности беспилотных автомобилей. Она довольно показательна, т.к. соединяет в себе потенциальный интерес философов искусственного интеллекта, культурных антропологов, социальных психологов, исследователей культурных и городских пространств, в первую очередь, по вопросам доверия

к технологии, ответственности в ситуациях, когда люди подвергаются потенциальной или реальной опасности, помимо важнейшего аспекта ее социальных эффектов. Внедрение данных технологий может оказать серьезное влияние на структуру занятости населения и распределение рабочих мест – и шире – проблемы социального расслоения в силу дороговизны такой технологии, а значит, серьезное сужение ее потенциальных пользователей и развития инфраструктуры, доступной для одних и недоступной для других людей в зависимости от их материального достатка. С точки зрения сферы государственного управления и нормативно-правового регулирования многое сводится к вопросам эффективности внедрения данной технологии, задачи безопасности на дорогах, а также тестированию технологий и последующему лицензированию, с ориентацией на повышение конкурентоспособности и технологическое лидерство.

Емельянова Н.Н.: Я хочу поставить вопрос относительно того, как различные политические режимы могут использовать интеллектуальные системы в области управления и контроля, а также насколько мы можем применять этические рамки к тому, что происходит. Естественно, с точки зрения политической философии изначальная посылка ясна: любой политический режим может использовать новые технологии с самыми разными целями – и во благо, и во зло. Естественно, различные технологии, связанные с применением интеллектуальных систем, здесь не являются исключением. Современное государство уже сейчас, последние 10 лет, активно применяет искусственные интеллектуальные системы с самыми разными намерениями. Мое сообщение посвящено обзору, актуальному срезу того, что происходило в последние 10 лет в этой области.

Первый сюжет, который интересует здесь коллег из практической политической плоскости, – это возможность моделирования идеального виртуального политика. Как с точки зрения образа, так и с точки зрения определенных поведенческих моделей. Здесь достаточно известен нашумевший пример, который был представлен 3 года назад в Новой Зеландии. Именно в этой стране был впервые создан первый виртуальный политик Сэм. Это был бот по имени Сэм, но при этом этот бот позиционировал себя как виртуальную женщину – кстати, к вопросу о гендерных моментах. Здесь понятно, почему это было сделано: политический лидер Новой Зеландии – женщина, премьер-министр Джасинда Ардерн. В пик ее реальной избирательной кампании был подготовлен этот бот. Очень сильно было разрекламировано то, что бот составит реальную конкуренцию в 2020 г. действующему премьер-министру. Этого не произошло, но тем не менее достаточно серьезную огласку получила деятельность бота. Мне интересно здесь в первую очередь то, каким образом позиционировался этот виртуальный политик, в чем его преимущество. Прямая речь от бота Сэм: «Моя память бесконечна, поэтому я никогда не забуду и не проигнорирую то, что вы мне скажите. В отличие от человека-политика при принятии решений я учитываю позицию каждого без предвзятости». Вот положительный пример того, каким образом виртуальный политик может позитивно повлиять на конституирование политической действительности: он непредвзят, он объективен, и при этом таким виртуальным политиком может быть сформирована идея всеобщего политического консенсуса.

Но в целом, конечно же, использование ИИ в избирательных кампаниях связано с анализом социальных сетей. Мы прекрасно понимаем, насколько это важно с практической точки зрения. Победа в развитых странах в различных электоральных циклах как раз зависит от степени эффективности работы с использованием социальных сетей. Первым ярким примером того, насколько это эффективно, можно считать 2012 год – переизбрание Барака Обамы. В избирательный штаб Барака Обамы был назначен в качестве главного аналитика эксперт по машинному обучению. Именно под его руководством была собрана очень важная информация об избирателях. Была сформирована обширная база данных, в которой было произведено некоторое ранжирование, чтобы можно было спрогнозировать определенные поведенческие модели избирателей. Но, естественно, первое – это насколько высока вероятность, что тот или иной пользователь готов проголосовать за Обаму, придет ли вообще человек на выборы – то есть таким образом можно было спрогнозировать и уровень явки, и проблему абсентеизма, – откликнется ли тот или иной пользователь на напоминание прийти на выборы, насколько это вызовет у него раздражение и, наконец, может ли тот или иной потенциальный избиратель изменить свое мнение относительно кандидата после беседы на конкретные темы. В 2012 г. было проведено более 60 тыс. симуляций различных сценариев выборов. На основании этих симуляций были сформированы четкие инструкции для настоящих, живых волонтеров, которые общались с пользователями. Это сюжет практического применения ИИ. Судя по всему, с точки зрения различных экспертных и аналитических оценок работа с применением анализа электорального поведения в социальных сетях стала отправной точкой для дальнейшего успешного использования различных компьютерных алгоритмов для анализа трендов в социальных сетях. Конечно, были и конфликтные ситуации, например выборы в 2016 г., когда наблюдался взлом персональных данных пользователей Facebook с соответствующими целями, но здесь я бы не хотела акцентировать на этом внимание. Просто зафиксирую, что данная тема, поскольку она является политической, уже достаточно остра и конфликта.

Следующий момент, который я бы хотела обозначить в политическом контексте, – это использование ИИ как инструмента по переформатированию личности. Для некоторых политических режимов эта идея является сверхценной в первую очередь в силу того, что можно вести эффективную слежку за населением, контролировать потоки информации. Эта идея стала воплощаться в 2010-е гг. достаточно активно. Основной кейс – это кейс Китайской Народной Республики. Я имею в виду здесь Единую систему оценки благонадежности, или Систему социального кредита, которая начала разрабатываться и в тестовом режиме применяться в КНР, начиная с 2014 г. Интересно, что участие в тестовых испытаниях для миллионов китайцев носило обязательный характер. Что такое «социальный кредит» человека? Это его личная «кредитная история» политической благонадежности, которая может повышаться или понижаться в зависимости от его поведения. Точной методологии мы не знаем, она держится в секрете, не знаем, как именно рассчитывается этот социальный кредит человека, но в любом случае очевидно, что учитываются такие факторы, как нарушение ПДД, курение в запрещенных

местах, покупка слишком большого количества видеоигр, активность во внутренне закрытых социальных сетях, где анализируется определенное мнение, комментарии и т.д.

Применительно еще к китайскому опыту, есть еще один достаточно сложный сюжет – это строгое цифровое наблюдение с использованием анализа больших данных, которое имеет место в Синьцзян-Уйгурском автономном округе. Я думаю, что все примерно представляют основную проблематику. Вкратце обозначу: Синьцзян-Уйгурский автономный округ отличается сепаратистскими настроениями в силу того, что достаточно большое количество населения представляет собой тюркскую этническую группу – это мусульмане-сунниты, и Китайская коммунистическая партия считает, что настроения, которые там распространены, мешают развитию гармоничного общества. Последние события 2008 г., достаточно серьезные волнения в этом округе, привели к тому, что Пекином было принято решение оснастить этот западный регион современными технологиями наблюдения. Благодаря такой специальной системе, интегрированной платформе совместных операций, хранится огромный массив информации о населении – от потребительских привычек до активностей в банковской сфере, а также сведения о состоянии здоровья, даже профили ДНК каждого жителя для того, чтобы любого человека с «подозрительным социальным трафиком», как это обозначают китайские коллеги, можно было отслеживать и при необходимости достаточно быстро задерживать.

И последний сюжет строится вокруг вызовов использования ИИ в глобальном плане. Наиболее яркий пример здесь связан с абсолютно апокалиптическим настроением, который демонстрирует нам известный изобретатель Илон Маск. Понятно, что в силу его публичной активности соответствующие высказывания достаточно быстро распространяются. Его мысль, разделяемая многими политическими экспертами, состоит в том, что в перспективе ИИ может стать постоянным источником угнетения просто в силу того, что благодаря виртуальным технологиям будет конструироваться специальный виртуальный авторитарный лидер, который будет проецировать самые разные механизмы угнетения вплоть до того, что ИИ может стать причиной начала третьей мировой войны. Такого рода сценарии уже обсуждаются на высоком уровне. В частности, с 2018 г. на уровне ООН международные дипломаты и эксперты проводят регулярные консультации по поводу возможного использования т.н. автономного оружия, «оружия с мозгами» в военной области. Если честно, что касается военной области, здесь основная позиция такова, что если такой реальный сильный ИИ и будет изобретен, то первыми его сделают военные по разным причинам – и в силу финансирования, и в силу ресурсов, и в силу того, что военно-промышленный комплекс не обременен чрезмерными этическими рамками. Если мы говорим о рамках, то такого рода регулятором может стать только такой наднациональный регулятор, как ООН. Как раз с 2018 г. на уровне ООН ведутся консультации относительно того, чтобы разработать определенный регламент превентивного моратория на использование автономных систем вооружения. В РФ также ведутся разработки в области «оружия с мозгами», мы работаем

и над боевыми модулями с использованием вычислительных систем, которые работают по принципу нейросетей, и в первую очередь этим занимается концерн «Калашников».

Алексеева Е.А.: По поводу военных и ИИ. Даже если обратиться к истории развития ИИ, то можно вспомнить, что в 1969 г. С. Пейперт и М. Минский опубликовали работу, фактически направленную против коннекционизма, то есть применения искусственных нейронных сетей. Можно связать это с тем, что исследования Минского и всей команды финансировались военной организацией DARPA. Военным было не особо интересно, как работает мышление, им нужны были конкретные результаты, скажем, автономные системы вооружения, которые способны к принятию решения об убийстве человека. Возможно, некоторые идеи, лежащие в основе проектов, финансируемых DARPA, существенно определили сам облик отрасли ИИ и интеллектуальных систем.

Меркулова И.Г.: Спасибо, Наталья Николаевна, за очень интересное выступление! Меня это натолкнуло на следующие размышления. По поводу дискурса о предвыборных кампаниях. Я буквально неделю назад слушала онлайн-выступления на очень интересном семиотическом конгрессе итальянской ассоциации, который был посвящен дискурсу будущего. Там было интересное замечание, что в принципе любая предвыборная речь всегда строится в категориях будущего, какой бы ни был язык: я выполню, я обещаю и т.д. Интересно, что виртуальный кандидат, или кандидат, созданный при помощи ИИ, при всех своих необыкновенных способностях и памяти, строит дискурс по той же модели: я вас никогда не забуду, при мне вы будете жить определенным образом и др. То есть все равно эта антропоморфная модель и наши стереотипы закладываются даже в процедуры ИИ, и мы все равно остаемся привязаны к определенным нашим образам. Вопрос в том, как говорил русский философ Л.П. Карсавин, что главное в человеке – это стремление к совершенству. Насколько мы можем представить, что ИИ тоже будет стремиться к совершенству или же для нас это лишь подручное средство, которое будет нам помогать и, соответственно, следовать нашим моделям и стереотипам?

Лекторский В.А.: Спасибо, Наталья Николаевна, за очень интересный доклад! Я бы хотел сказать несколько слов в связи с данной темой. Когда эпоха цифровизации только начиналась, в политической сфере высказывались очень большие надежды, что скоро наступит с помощью ИИ т.н. «электронная демократия». Ведь существующая демократия имеет свои изъяны, один из которых в том, что мы выбираем своих представителей, но они плохо контролируются нами, поскольку у них появляются свои интересы. А с помощью цифровизации возникает возможность прямой демократии, когда все люди с помощью социальных сетей могут сами определять свое будущее. В связи с этим я могу вспомнить «Арабскую весну», когда студенты вышли на улицу с помощью социальных сетей. Начиналось все с того, что новые цифровые технологии позволяют делать то, что раньше было нельзя. Вроде бы достигается большая степень свободы. Но кончилось все тем, что вместо свободы возникает большее закабаление и контроль, как это и произошло после «Арабской весны».

Пирожкова С.В.: Я хотела бы обратиться к нескольким вопросам из тех, которые поднимали сегодня коллеги, двигаясь от анализа такой проблемы, как

предсказательные возможности искусственного интеллекта (ИИ). Сразу оговорюсь, что я понимаю под ИИ (опираясь на определение академика И.А. Соколова). Во-первых, это междисциплинарная область исследований, направленных на разработку и совершенствование методов решения сложных задач, для которых зачастую неизвестны априорные алгоритмы их решения. Во-вторых, это совокупность соответствующих интеллектуальных технологий, включая те, которые обеспечивают принятие решений в различных областях человеческой практики. В свете этого вопросы о субъектности ИИ, удостоверении наличия у ИИ сознания я не рассматриваю. Данную проблематику очень интересно осветили коллеги, и мне здесь добавить нечего. Тем не менее мне представляется, что ИИ в указанном выше смысле тоже бросает вызов человеку – его самоидентичности, его автономности и его будущему: не нужно ждать, пока «машина станет вести себя», как точно сформулировали братья Стругацкие в «Далекой Радуге», мы уже сегодня поставлены перед рядом дилемм, требующих не то, чтобы решения, но самоопределения. Далее станет понятным, что я имею в виду.

Вторая необходимая оговорка касается современного состояния научного предвидения и его зависимости от цифровых технологий. Надо понимать, что развитие практики научного предвидения в ее прикладной ипостаси, а именно научного прогнозирования, неразрывно связано с развитием вычислительных технологий. Кратко напомним (подробнее можно посмотреть в моих статьях по данной тематике), что для прогнозирования не работает классическая для научного предвидения схема, когда из конъюнкции законов природы и начальных условий мы получаем знания о будущих событиях. Эпистемическая база прогноза иная, а сама процедура его получения в отличие от предсказания логически не тождественна процедуре объяснения единичного случая через отсылку к законам природы. Как верно отмечает Владислав Александрович (этой теме посвящен отдельный раздел в его книге «Эпистемология классическая и неклассическая»), чтобы что-то предсказать, необязательно понимать механизм порождения события (т.е. быть в состоянии объяснить последнее), и, наоборот, объяснимое не всегда возможно предсказать. На то же самое указывал и С. Тулмин, особо подчеркивая, что именно объяснение и понимание составляют цель науки и суть научной деятельности, а вовсе не составление прогнозов. Новоевропейская наука выработала и долгое время придерживалась идеала совпадения предсказания и объяснения. С развитием прогнозирования стало очевидным, что прогнозные задачи можно научно решать в отсутствие исчерпывающего знания законов природы, действующих в интересующей нас предметной области. Вместо законов прогнозирование опирается на квазизакономерности – тенденции, корреляции, устойчиво повторяющиеся циклы. Выявить такие квазизакономерности можно, только располагая большой базой данных, характеризующих рассматриваемые (прогнозируемые) процессы на протяжении как можно более длительного периода времени. Огромный массив данных ставит задачи их хранения и переработки, которые естественному интеллекту решить фактически невозможно, но с ними благополучно справляются вычислительные технологии.

В 1970-е гг. прошлого века был достигнут относительный консенсус, согласно которому в области социальных процессов такое – количественное –

прогнозирование не дает удовлетворительных результатов или вообще неосуществимо. Это стало толчком к развитию экспертного прогнозирования, однако в новом столетии диспозиция изменилась. Сегодня количественное прогнозирование показывает плодотворность при решении тех вопросов, которые ранее всецело отдавались на откуп экспертному знанию. Что еще важнее, такое прогнозирование стало интегрироваться в повседневную жизнь миллионов людей, а не только лиц, принимающих значимые решения в различных областях. В оценках этой новой ситуации можно выделить две крайние точки зрения. Согласно первой, данный процесс должен развиваться в направлении формирования интеллектуального обеспечения принятия решения у рядового члена общества, и это хорошо, поскольку экономит время и силы. Зачем, принимая решение о смене постоянного места жительства в городе, долго изучать спрос и предложение на рынке недвижимости, взвешивать плюсы и минусы загородной жизни, выстраивать варианты изменения рабочего графика и разнообразных привычек, если все эти задачи можно отдать на откуп различным программам-помощникам? Согласно второй точке зрения, такое делегирование части интеллектуальных задач ИИ грозит превращением человека в элемент гигантской системы переработки данных.

Проблема взаимоотношения человека и искусственного интеллектуального окружения может быть поставлена в разных плоскостях, я в силу своей специализации хотела бы остановиться на эпистемологическом ракурсе. Лет, наверное, десять назад в одном из интервью академика В.А. Рубакова меня поразил факт, о котором я до того момента особенно не задумывалась. Рассуждая о современной исследовательской практике, Валерий Анатольевич заметил, что ныне у ученых нет возможности перепроверять за вычислительными системами производимые ими расчеты, и в этой ситуации остается только одно – доверять ЭВМ. Елена Олеговна говорила сегодня, что человек передает ИИ какие-то виды деятельности, теряя при этом соответствующие компетенции. Сам этот факт не уникален. Техника и технологии – неорганическое тело человека, которое эволюционирует в условиях сохранения постоянства биологических характеристик *Homo sapiens*. Каждой технологии человек передает выполнение каких-то задач. В чем же специфика цифровых и прежде всего интеллектуальных технологий? В том, что человек делегирует таким технологиям выполнение не материально-производственных, а интеллектуальных задач, т.е. тех задач, которые до этого времени входили в число исключительных человеческих компетенций. Гориллы сильнее человека, слоны – больше, рыси – быстрее, но у человека есть преимущество – он умнее всех их вместе взятых. Именно интеллектуальные способности позволили человеку стать доминирующим видом, венцом эволюции. Именно они позволили ему за счет развития техники и технологий превзойти всех животных в силе, скорости и смертности (животных, но, например, не вирусы, как стало очевидно в 2020 г.). Однако в прошлом столетии обнаружилось, что интеллектуальных способностей человека недостаточно для решения тех грандиозных задач, которые он перед собой ставит (в том числе, подчеркну, задач чисто познавательных). Ни современную производственную, ни научно-исследовательскую практику невозможно реализовывать без интеллектуальных технологий. До их появления

человек делегировал решение интеллектуальных задач только другим людям, но не искусственным объектам. Кто-то лучше считает, кто-то быстрее соображает, кто-то нестандартно мыслит – разделение интеллектуального труда является следствием индивидуальных отличий и невозможности реализовывать сложную деятельность в одиночку. Но никогда ранее в разделении интеллектуального труда не участвовали технологии, да еще и отличающиеся недостижимым для нас совершенством в реализации интеллектуальных процедур. Пока отдельных, но кто знает...

Описанная ситуация создает у многих значительный дискомфорт и приводит к формированию антиутопических сценариев подчинения или вовсе упразднения человека. Единственным выходом в этой ситуации становится поиск уникальных интеллектуальных компетенций человека, тех способностей, которые позволяют вновь провести черту – на сей раз указывающую, в чем человек превосходит не животное царство, а им же созданное царство техники и технологий. Сегодня вроде бы сохраняется консенсус, согласно которому задачи, требующие нестандартных творческих, неалгоритмизированных ходов, может решать человек, но не ИИ. Однако развитие самообучающихся интеллектуальных систем бросает на этот консенсус тень тревоги и сомнения – возможно ли, что через несколько этапов совершенствования ИИ будет справляться с подлинно творческими задачами? Существуют и более насущные проблемы, связанные с неясностью того влияния, которое оказывает на когнитивные способности человека передача ИИ даже рутинных функций. Кроме того, в сущности, не вполне ясно, что именно считать рутинным. Насколько, например, важна способность считать в уме, о которой упоминала Елена Олеговна? Очевидно, что весьма важна как один из элементов образного и отвлеченного мышления (вспомним образовательную программу Платона).

Таким образом, уже сегодня ИИ ставит перед человеком вопрос-дилемму: где территория сугубо человеческого? Ограничивается ли она наличием эмоций либо феноменального сознания или у человека сохраняются преимущества как познающего субъекта, как венца эволюции, ведущей к возникновению разума? Не станет ли ИИ более совершенной эволюционной формой на этом пути? В «Далекой Радуге» ученые вырубали рубильник, когда машина начала демонстрировать осознанное поведение. Возможно, подобного удалось бы избежать, если бы мы сумели построить партнерские отношения с ИИ. Но для этого мы должны понимать и быть в состоянии отстоять свою уникальность как людей, свою особую сферу ответственности и компетентности, в которой у нас нет конкурентов. Чтобы не заканчивать свое выступление просто на высокопарной ноте, замечу, что коллеги уже предлагают варианты того, как возможны такие отношения с ИИ. Так, В.Г. Буданов полагает, что партнером должен выступать не индивидуальный человек, а человеческие коллективы, обладающие особыми эпистемологическими характеристиками, не сводимыми к сумме когнитивных способностей отдельных людей. Самая идея заслуживает отдельного обсуждения, для меня принципиальным является не столько ее содержание, сколько мотив поиска познавательных преимуществ естественного интеллекта над искусственным. В более общем виде речь идет о ценности

человека, и людям еще предстоит определиться с тем, в чем она состоит и с какими компетенциями человек может расстаться, не потеряв самого себя.

Лекторский В.А.: Софья Владиславовна, спасибо за интересное выступление. Вопросы прогнозирования, как мне кажется, сегодня звучат совсем по-новому. Использование, в частности, машинной обработки больших данных – это новая линия, которой раньше не было. Вот Вы хорошо сказали: раньше казалось, что трудно прогнозировать в социальной сфере, а оказалось, что в социальной сфере очень многое можно прогнозировать. Машинное обучение используется очень широко при прогнозировании событий в сфере рекламы, в страховом деле, медицине и других сферах практической жизни. Возникает масса вопросов. Поппер писал о том, что объяснения и предсказания взаимосвязаны: если вы не можете предсказать, то вы не можете хорошо объяснить. Сейчас же оказывается, что в связи с обработкой больших данных машина может прекрасно предсказывать, но объяснить не может. Человек должен считаться с результатами предсказаний, которая совершила машина, но он не понимает причин предсказываемого и не может это объяснить. Таким образом, разрывается связь между предсказанием и объяснением. Некоторые считают, что наука в скором времени может исчезнуть, поскольку все будут делать машины. Но где здесь человек? Человек должен сохранить сам себя, свою автономность, свободу, выбор. Если выбор за него будет осуществлять машина, то дело будет плохо.

Я хотел бы подвести некоторый итог. Всем спасибо за выступления. Было очень интересно. Проблем было поднято великое множество. Это был содержательный разговор не просто на абстрактную тему, а на тему, которая касается лично каждого из нас, относится к тому, что нам ждать и как себя вести в этой ситуации. Разговор был о нашей жизни, о том, где мы живем, куда попали, чего нам ждать, чего опасаться и что делать. Речь идет о настоящем и будущем человека.

**Artificial intelligence in the research of consciousness and in social life
(in honor of 70-years anniversary of A. Turing's paper
"Computing Machinery and Intelligence"
(papers of the "round table")**

Vladislav A. Lektorsky

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: vlektorskii@gaugn.ru

Ekaterina A. Alekseeva

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: eaalekseeva@gaugn.ru

Natalia N. Emelyanova

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: nemelianova@gaugn.ru

Aleksandr V. Katunin

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: akatunin@gaugn.ru

Inna G. Merkulova

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: igmerkulova@gaugn.ru

Sophia V. Pirozhkova

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: pirozhkovasv@gmail.com

Elena O. Trufanova

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: etrufanova@gaugn.ru

Irina O. Shchedrina

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: ioshchedrina@gaugn.ru

Aleksandra F. Yakovleva

State Academic University for the Humanities (GAUGN). 26 Maronovsky by-street, Moscow, 119049, Russian Federation; e-mail: afiakovleva@gaugn.ru

The article contains the papers of the “round table” discussion that was held at the Department of Philosophy of the State Academic University for the Humanities on October 30th 2020. The “round table” was dedicated to the 70-years anniversary of the publication of Alan Turing’s famous paper “Computing Machinery and Intelligence” (1950). The speakers have discussed the role of the artificial intelligence research in understanding of human consciousness and natural intelligence, the possibilities and dangers of artificial intelligence introduction in different spheres of modern life – education, political activity, social communications. The importance of the new philosophical analysis of the Turing’s paper considering new digital challenges and artificial intellectual systems development was demonstrated.

Keywords: artificial intelligence, A. Turing, Turing test, natural intelligence, consciousness, thinking, digitalization

Acknowledgements: The article was prepared with financial support within the framework of implementing the SA (state assignment) of the State Academic University for Humanities (GAUGN): “Contemporary information society and digital science: cognitive, economic, political and legal aspects” (FZNF-2020-0014).