Российская Академия Наук

НКС ООН РАН

АРБ Ассоциация Российских Банков

Национальный исследовательский институт Доверия, Достоинства и Права

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Материалы заседания
14 октября 2023 года

ДОКЛАДЧИКИ:

Лекторский Владислав Александрович
академик РАН, доктор философских наук,
главный научный сотрудник Института философии РАН,
председатель Научного совета РАН по методологии
искусственного интеллекта и когнитивных исследований

Ушаков Дмитрий Викторович
академик РАН, доктор психологических наук,
директор Института психологии РАН,
заведующий кафедрой общей психологии
факультета психологии МГУ им. М.В. Ломоносова

Аветисян Арутюн Ишханович
академик РАН, доктор физико-математических наук,
директор Института системного программирования
им. В.П. Иванникова РАН

НКС ООН РАН
Научно-консультативный совет
по правовым, психологическим
и социально-экономическим проблемам общества
Отделения общественных наук РАН

АРБ
Ассоциация российских банков

НИИ ДДиП
Национальный исследовательский институт
Доверия, Достоинства и Права

Искусственный интеллект

Материалы заседания 14 октября 2023 года

Под общей редакцией
академика РАН
Г.А. Тосуняна

Москва
2024

УДК 004.8(063)
ББК 32.813я431
И86

Искусственный интеллект : материалы заседания 14 октября 2023 года / Научно-консультативный совет по правовым, психологическим и социально-экономическим проблемам общества Отделения общественных наук Российской академии наук ; Ассоциация российских банков ; Национальный исследовательский институт Доверия, Достоинства и Права ; [под общ. ред. академика РАН Г.А. Тосуняна]. — М.: ООО «Новые печатные технологии», 2024. — 182 с. — ISBN 978-5-6050844-9-5

Этот сборник – об искусственном интеллекте.

В нем рассматриваются как положительные, так и отрицательные последствия его создания и использования для человечества.

Уже сегодня, став неотъемлемой частью нашей жизни, искусственный интеллект применяется во многих сферах, включая медицину, образование, логистику и т.д.

Из материалов сборника можно узнать об успехах российских ученых в этом направлении и о том, что предпринимается для минимизации рисков, чтобы не допустить выхода искусственного интеллекта из-под контроля человека.

УДК 004.8(063)
ББК 32.813я431

Охраняется в соответствии с международным правом и российским законодательством об авторском праве.

ISBN 978-5-6050844-9-5 © Тосунян Г.А., составление, 2024

СОДЕРЖАНИЕ

Состав Научно-консультативного совета по правовым, психологическим и социально-экономическим проблемам общества (НКС ППСЭПО) ООН РАН **5**

Справка **10**

ВСТУПИТЕЛЬНОЕ СЛОВО

***акад.* ТОСУНЯН Г.А.** **14**

Доклад *1 акад.* ЛЕКТОРСКИЙ В.А. **20**

ЧЕЛОВЕК И СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

***д. э. н.* КРИВОВ В.Д. – *акад.* ЛЕКТОРСКИЙ В.А.** **46**

***проф.* МЕДВЕДЕВ П.А. – *акад.* ЛЕКТОРСКИЙ В.А.** **50**

Доклад 2 *акад.* УШАКОВ Д.В. **55**

ЧЕЛОВЕК И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ПОДХОДЫ К ВЗАИМОДЕЙСТВИЮ

***проф.* МЕДВЕДЕВ П.А. – *акад.* УШАКОВ Д.В.** **72**

***проф.* КАЛАШНИКОВ С.В., *акад.* УШАКОВ Д.В., *акад.* ЛЕКТОРСКИЙ В.А.** **75**

Доклад 3 *акад.* АВЕТИСЯН А.И. **78**

ТЕХНОЛОГИИ ДОВЕРЕННОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ

***к. т. н.* САТАРОВ Г.А. – *акад.* АВЕТИСЯН А.И.** **105**

***д. э. н.* КРИВОВ В.Д. – *акад.* АВЕТИСЯН А.И.** **107**

***чл.-корр.* ЧЕРНЫШ М.Ф. – *акад.* АВЕТИСЯН А.И.** **109**

акад. **ЛЕКТОРСКИЙ В.А.** **113**

к. п. н. **СУПРУН А.П.** **114**

СМОЛИН В.С., ***акад*****. ТОСУНЯН Г.А.,**
акад**. АВЕТИСЯН А.И.** **117**

проф. **СЕНАТОРОВ М.Ю. –** ***акад*****. АВЕТИСЯН А.И.** **122**

д. филос. н., проф. **КРУШАНОВ А.А.** **126**

чл.-корр. **КАСАВИН И.Т.** **131**

д. филос. н., проф. **МОМДЖЯН К.Х.** **138**

проф. РАН **САННИКОВА Л.В. –** ***акад*****. АВЕТИСЯН А.И.** **142**

к. ю. н. **МАРКАРЬЯН Р.В. –** ***акад*****. АВЕТИСЯН А.И.** **146**

чл.-корр. **ПЕТРЕНКО В.Ф.** **150**

проф. РАН **АЛМАЕВ Н.А.** **152**

*проф***. МОМДЖЯН К.Х.,** ***акад*****. ТОСУНЯН Г.А.,**
акад**. АВЕТИСЯН А.И.,** ***акад*****. ЛЕКТОРСКИЙ В.А.** **153**

д. ю. н., проф. РАН **ЧЕРНОГОР Н.Н.** ***156***

ГНУСАРЕВ А.В., ***акад*****. ТОСУНЯН Г.А.,**
акад**. АВЕТИСЯН А.И.** **161**

акад. **ЛЕКТОРСКИЙ В.А.** **164**

акад. **АВЕТИСЯН А.И.** **168**

акад. **ГУСЕЙНОВ А.А.** **171**

ЗАКЛЮЧИТЕЛЬНОЕ СЛОВО

акад**. ТОСУНЯН Г.А.** **174**

Список литературы, опубликованной
по итогам заседаний НКС ООН и НИИ ДДиП **178**

СОСТАВ НАУЧНО-КОНСУЛЬТАТИВНОГО СОВЕТА ПО ПРАВОВЫМ, ПСИХОЛОГИЧЕСКИМ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИМ ПРОБЛЕМАМ ОБЩЕСТВА (НКС ППСЭПО) ООН РАН

СОПРЕДСЕДАТЕЛИ:

ГУСЕЙНОВ АБДУСАЛАМ АБДУЛКЕРИМОВИЧ	академик, д. филос. н., научный руководитель Института философии РАН
КОКОШИН АНДРЕЙ АФАНАСЬЕВИЧ	академик, д. и. н., директор Центра перспективных исследований национальной безопасности России Экспертного института НИУ ВШЭ
ТОСУНЯН ГАРЕГИН АШОТОВИЧ	академик, д. ю. н., президент Ассоциации российских банков

УЧЕНЫЙ СЕКРЕТАРЬ:

РЕДЬКО НИКОЛАЙ ВИТАЛЬЕВИЧ	к. э. н., эксперт Национального исследовательского института Доверия, Достоинства и Права

ЧЛЕНЫ НАУЧНОГО СОВЕТА:

АВЕТИСЯН АРУТЮН ИШХАНОВИЧ	академик, д. ф.-м. н., директор Института системного программирования им. В.П. Иванникова РАН
АГАНБЕГЯН АБЕЛ ГЕЗЕВИЧ	академик, д. э. н., профессор, заведующий кафедрой экономической теории и политики Российской академии народного хозяйства и государственной службы при Президенте РФ

АПОЛИХИН ОЛЕГ ИВАНОВИЧ	чл.-корр., д. м. н., директор НИИ урологии и интервенционной радиологии им. Н.А. Лопаткина (филиал ФГБУ «НМИЦ радиологии» Минздрава России)
АУЗАН АЛЕКСАНДР АЛЕКСАНДРОВИЧ	д. э. н., декан экономического факультета МГУ им. М.В. Ломоносова
БАТУРИН ЮРИЙ МИХАЙЛОВИЧ	чл.-корр., д. ю. н., главный научный сотрудник отдела методологических и междисциплинарных проблем развития науки Института истории естествознания и техники им. С.И. Вавилова РАН
БУЗНИК ВЯЧЕСЛАВ МИХАЙЛОВИЧ	академик, д. х. н., заместитель академика-секретаря ОХНМ РАН, начальник лаборатории Всероссийского НИИ авиационных материалов
ГОРШКОВ МИХАИЛ КОНСТАНТИНОВИЧ	академик, д. филос. н., директор Института социологии ФНИСЦ РАН
ГРАЧЕВА ЕЛЕНА ЮРЬЕВНА	д. ю. н., профессор, первый проректор ФГБОУ ВО «Московский государственный юридический университет им. О.Е. Кутафина» (МГЮА)
ГРИНБЕРГ РУСЛАН СЕМЕНОВИЧ	чл.-корр., д. э. н., научный руководитель Института экономики РАН
ДАНИЛОВ-ДАНИЛЬЯН АНТОН ВИКТОРОВИЧ	к. э. н., сопредседатель Общероссийской общественной организации «Деловая Россия»
ЕРМАКОВА ЖАННА АНАТОЛЬЕВНА	чл.-корр., д. э. н., профессор, заведующий кафедрой банковского дела и страхования ФГБОУ ВО «Оренбургский государственный университет»
ЖУРАВЛЕВ АНАТОЛИЙ ЛАКТИОНОВИЧ	академик, д. п. н., научный руководитель Института психологии РАН

ИВАНОВ ВИЛЕН НИКОЛАЕВИЧ	чл.-корр., д. филос. н., главный научный сотрудник Института социально-политических исследований ФНИСЦ РАН
ИЛЬИН ВЛАДИМИР АЛЕКСАНДРОВИЧ	чл.-корр., д. э. н., профессор, научный руководитель Вологодского научного центра РАН
КАСАВИН ИЛЬЯ ТЕОДОРОВИЧ	чл.-корр., д. филос. н., руководитель сектора социальной эпистемологии Института философии РАН
КЛЕПАЧ АНДРЕЙ НИКОЛАЕВИЧ	к. э. н., главный экономист ВЭБ.РФ
ЛЕКТОРСКИЙ ВЛАДИСЛАВ АЛЕКСАНДРОВИЧ	академик, д. филос. н., главный научный сотрудник Института философии РАН
МЕДВЕДЕВ ПАВЕЛ АЛЕКСЕЕВИЧ	д. э. н., профессор
МИРКИН ЯКОВ МОИСЕЕВИЧ	д. э. н., руководитель отдела международных рынков капитала Национального исследовательского института мировой экономики и международных отношений им. Е.М. Примакова РАН
НЕСТИК ТИМОФЕЙ АЛЕКСАНДРОВИЧ	д. п. н., профессор РАН, зав. лабораторией социальной и экономической психологии Института психологии РАН
НИГМАТУЛИН РОБЕРТ ИСКАНДРОВИЧ	академик, д. ф.-м. н., научный руководитель Института океанологии им. П.П. Ширшова РАН
ПЕТРЕНКО ВИКТОР ФЕДОРОВИЧ	чл.-корр., д. п. н., заведующий лабораторией психологии общения факультета психологии МГУ им. М.В. Ломоносова

ПОГОСЯН ГЕВОРК АРАМОВИЧ	академик Национальной академии наук Армении (НАН РА), иностранный член РАН, д. социол. н., научный руководитель Института философии, социологии и права НАН РА
САВЕНКОВ АЛЕКСАНДР НИКОЛАЕВИЧ	чл.-корр., д. ю. н., директор Института государства и права РАН
САННИКОВА ЛАРИСА ВЛАДИМИРОВНА	д. ю. н., профессор РАН, руководитель Центра правовых исследований цифровых технологий Государственного академического университета гуманитарных наук
САРКИСЯН ТИГРАН СУРЕНОВИЧ	к. э. н., заместитель председателя правления Евразийского банка развития
СМИРНОВ АНДРЕЙ ВАДИМОВИЧ	академик, д. филос. н., директор Института философии РАН
СОЛОДКОВ ВАСИЛИЙ МИХАЙЛОВИЧ	к. э. н., директор Банковского института НИУ ВШЭ
ТЕДЕЕВ АСТАМУР АНАТОЛЬЕВИЧ	д. ю. н., профессор кафедры государственного аудита Высшей школы государственного аудита (факультет) МГУ им. М.В. Ломоносова
ТИХОМИРОВ ЮРИЙ АЛЕКСАНДРОВИЧ	д. ю. н., заместитель заведующего Центра публично-правовых исследований Института законодательства и сравнительного правоведения при Правительстве России
ТОРШИН АЛЕКСАНДР ПОРФИРЬЕВИЧ	к. ю. н., действительный государственный советник РФ I класса

ТОЩЕНКО ЖАН ТЕРЕНТЬЕВИЧ	чл.-корр., д. филос. н., профессор, главный научный сотрудник Института социологии ФНИСЦ РАН
УГРЮМОВ МИХАИЛ ВЕНИАМИНОВИЧ	академик, д. б. н., заведующий лабораторией нервных и нейроэндокринных регуляций Института биологического развития им. Н.К. Кольцова РАН
УШАКОВ ДМИТРИЙ ВИКТОРОВИЧ	академик, д. п. н., директор Института психологии РАН
ХАБРИЕВА ТАЛИЯ ЯРУЛЛОВНА	академик, д. ю. н., директор Института законодательства и сравнительного правоведения при Правительстве России
ЧЕРЕШНЕВ ВАЛЕРИЙ АЛЕКСАНДРОВИЧ	академик, д. м. н., научный руководитель Института иммунологии и физиологии Уральского отделения РАН
ЧЕРНЫШ МИХАИЛ ФЕДОРОВИЧ	чл.-корр., д. социол. н., директор Федерального научно-исследовательского социологического центра РАН
ЧЕХОНИН ВЛАДИМИР ПАВЛОВИЧ	академик, д. м. н., вице-президент РАН, заведующий кафедрой медицинских нанотехнологий медико-биологического факультета Российского государственного медицинского университета им. Н.И. Пирогова
ШАБУНОВА АЛЕКСАНДРА АНАТОЛЬЕВНА	д. э. н., директор Вологодского научного центра РАН
ЭКМАЛЯН АШОТ МАМИКОНОВИЧ	д. филос. н., профессор
ЮРЕВИЧ АНДРЕЙ ВЛАДИСЛАВОВИЧ	чл.-корр., д. п. н., заместитель директора по научной работе Института психологии РАН

СПРАВКА

о **НКС ООН РАН** (Научно-консультативном совете по правовым, психологическим и социально-экономическим проблемам общества Отделения общественных наук),
о **НИИ ДДиП** (Национальном исследовательском институте Доверия, Достоинства и Права),
о **«Рабочем завтраке у Тосуняна»**,
о проекте **«Открытые дискуссии президента АРБ»**
и об этом издании

1. НКС ООН РАН был создан в 2012 году как Совет по правовым, экономическим, социально-политическим и психологическим аспектам финансово-кредитной системы.

Заседания Совета проводились в Отделении общественных наук РАН два раза в год.

В феврале 2020 года члены НКС приняли решение расширить компетенцию Совета, перейдя от рассмотрения вопросов развития финансового рынка к более широкому кругу проблем развития общества, поставив во главу угла своих исследований и дискуссий вопросы:

«В каком обществе мы живем? Какое общество мы хотели бы оставить своим потомкам в наследство?»

И в сентябре 2021 года постановлением Президиума РАН Совет был преобразован в Научно-консультативный совет по правовым, психологическим и социально-экономическим проблемам общества ООН РАН.

Сопредседателями Совета стали академики РАН А.А. Гусейнов, А.А. Кокошин и Г.А. Тосунян.

2. С середины 90-х годов по субботам раз в две-три недели в Ассоциации российских банков проходят «Рабочие завтраки у Тосуняна», в которых принимали и принимают участие банкиры, представители ЦБ,

Госдумы, Совета Федерации, различных ведомств, академической науки, вузов, эксперты по финансово-банковскому профилю.

Каждый «Рабочий завтрак у Тосуняна» (далее – «Рабочий завтрак») проходит по заранее согласованной повестке дня и с заявленными докладчиками.

На них до недавнего времени обсуждались преимущественно проблемы экономики, финансовой сферы, нормативно-правовые акты, регулирующие эту сферу. Но в ряде случаев и другие вопросы развития общества.

В последние годы спектр вопросов, рассматриваемых на «Рабочих завтраках», и круг экспертов заметно расширились.

Этому во многом способствовало участие в них известных ученых.

Характерной особенностью «Рабочих завтраков» было и остается то, что они проходят с завидной регулярностью по субботам в 9.00 утра и зимой, и летом, и даже 31 декабря. Их продолжительность примерно 3–4 часа.

3. В конце 2019 года был учрежден Национальный исследовательский институт Доверия, Достоинства и Права (НИИ ДДиП).

Это частный институт, целью которого, если вкратце, является многогранное изучение вопросов человеческой жизнедеятельности и общественных процессов, которые наибольшим образом влияют на развитие доверия в обществе, повышение ответственности и чувства собственного достоинства у граждан страны и на формирование уважения друг к другу.

Институт приступил к работе в начале 2020 года в формате научных заседаний с коллегами, интересующи-

мися проблемами доверия и достоинства, их правового обеспечения и стимулирования.

Иначе говоря, институт пригласил на общественных началах работать на его площадке всех, кто желает внести свою лепту в изменение траектории движения общества «войны всех против всех» в сторону общества «доверия, достоинства и уважения друг к другу»!

4. В конце марта 2020 года был объявлен локдаун.

Встал вопрос: заморозить на какое-то время работу НКС ООН, НИИ ДДиП, АРБ и «Рабочие завтраки у Тосуняна»?

Или искать какое-то другое решение?

Тогда же возникла идея, что заседания НКС ООН, НИИ ДДиП и «Рабочие завтраки» можно объединить, используя онлайн-формат.

Проанализировав практику последних лет, мы с коллегами пришли к выводу, что довольно часто и на заседаниях НКС, и на «Рабочих завтраках», и на заседаниях Института мы поднимаем и обсуждаем схожие вопросы.

Было принято решение начать проводить совместные заседания.

За прошедшее с апреля 2020 года время было проведено 97 «Рабочих завтраков у Тосуняна», большинство из которых прошло в очно-заочной форме.

Примерно 20 человек лично присутствовали на завтраках, а остальные, от 50 до 100 и более участников, принимали участие в режиме Zoom, видя, слыша «живых» участников и докладчиков, также присоединялись к дискуссии.

В последующем по видеозаписи каждое заседание стенографировалось с тем, чтобы можно было издать материалы этих дискуссий.

В настоящее время накопился огромный объем материалов для публикаций, и мы начали их издание в виде представленных вашему вниманию сборников.

5. С 2013 года Ассоциация российских банков ведет проект «Открытые дискуссии президента АРБ».

Проект направлен на обсуждение широкого круга экономических, правовых, философских, социально-психологических и других актуальных проблем развития нашего общества и на развитие культуры дискуссии в целом. Спикерами «Открытых дискуссий президента АРБ» (далее – «Открытые дискуссии») выступают известные ученые, общественные деятели и представители бизнеса.

Вузами-партнерами проекта являются более 80 российских вузов, расположенных на территории всей России – от Владивостока до Калининграда.

Как правило, в каждой «Открытой дискуссии» дистанционно участвуют от 40 до 80 вузов. Численность интернет-аудитории в среднем составляет около 2 тыс. человек.

Последние два года «Открытые дискуссии» проводятся ежемесячно.

За 10 лет состоялось 78 дискуссий.

С информацией о прошедших дискуссиях, презентационными материалами спикеров и видеозаписями можно ознакомиться на сайте arb.ru в разделе «Открытые дискуссии».

Г.А. ТОСУНЯН, академик РАН,
президент Ассоциации российских банков

ВСТУПИТЕЛЬНОЕ СЛОВО

ТОСУНЯН Г.А.
акад. РАН

Приветствую всех на совместном заседании Научно-консультативного совета.

Мы проводим его в формате «Рабочего завтрака» вместе с Национальным исследовательским институтом Доверия, Достоинства и Права.

На наше мероприятие зарегистрировалось 27 человек очно, 160 с лишним – дистанционно.

В том числе 75 представителей науки и вузов, в том числе 21 член РАН, банкиры, представители властных структур…

Заседание посвящено искусственному интеллекту.

Начну с того, что у меня есть некоторые сомнения относительно корректности этого понятия – «искусственный интеллект».

Интеллект подразумевает нечто более объемное и сложное, чем компьютерные вычислительные способности.

Искусственный интеллект – это свойство интеллектуальных компьютерных систем выполнять творческие функции.

Функции, которые традиционно считаются прерогативой человека.

С другой стороны, это комплекс компьютерных наук.

А также направлений в математике, биологии, психологии и других науках.

Некоторые исследователи ставят задачу – посредством искусственного интеллекта смоделировать человеческий разум.

Насколько это возможно, большой вопрос.

Фактически речь идет о системах, которые могут принимать решения и действовать автономно от человека.

По мнению многих представителей науки, стремительное развитие искусственного интеллекта создает определенные угрозы для человечества.

Это не надуманный риск.

Правда, риски, которые несут некоторые неадекватные представители человечества и человеческого интеллекта, могут быть несоизмеримо выше.

Например, те из них, кто принимает «выдающиеся» политические и экономические решения, ставя под угрозу нормальное существование цивилизации. Но сейчас речь не о них…

Более простой пример реализации искусственного интеллекта – шахматный компьютер.

Он имитирует человека в шахматной партии с помощью большого количества алгоритмов и вычислений.

Считается, что даже у сильнейших гроссмейстеров мира уже нет шансов в противостоянии с компьютерными программами.

Вряд ли с этим кто-нибудь будет спорить.

Открытый вопрос: являются ли компьютерные шахматы примером искусственного интеллекта?

Их преимущество перед человеком – в возможности анализировать миллионы позиций в секунду на множество ходов вперед.

Но, по сути, это набор заданных человеком алгоритмов.

Может ли компьютер в силу глубоких переживаний сдаться в середине партии или вовсе отказаться от игры?

В реальной жизни с шахматистами такое происходило.

Там свою роль играл психологический фактор.

Говорили даже о гипнозе, о влиянии в других формах.

Для компьютера это в принципе невозможно.

В такой системе координат и калькулятор можно назвать искусственным интеллектом.

Но пока, слава Богу, калькулятор никто к искусственному интеллекту не относит.

Само понятие искусственного интеллекта внутренне противоречиво.

Оно требует дополнительной расшифровки и анализа.

Золотым временем развития искусственного интеллекта считаются 60-е годы прошлого века.

Компьютеры становились доступнее.

Они могли хранить большие объемы информации.

Серьезно увеличивались вычислительные мощности.

Совершенствовались алгоритмы машинного обучения.

Были разработаны первые экспертные системы, компьютерные программы.

Они были способны моделировать знания человека в определенной области: в химии, в физике, в биологии…

Появились первые нейронные сети.

Они могли обучаться на больших объемах данных, решать задачи классификации, например, распознавать рукописные цифры.

Новый виток развития искусственного интеллекта происходит в наши дни.

За последние 10 лет разработано больше технологий, чем за все предыдущее время.

Это и беспилотный транспорт.

И языковые переводчики.

Чат-боты, способные работать в диалоговом режиме и поддерживать запросы на естественных языках.

Медицинская диагностика, финансово-аналитические системы.

Например, системы скоринга при оценке кредитоспособности заемщика.

Искусственный интеллект в этой сфере довольно продвинут.

Он способен принимать решение о выдаче кредита, то есть фактически оценивать благонадежность с точки зрения платежеспособности человека.

Однако можно ли этим искусственным системам присваивать такой глубокий и объемный смысл, который мы вкладываем в понятие интеллекта?

Либо эти системы следует рассматривать исключительно как совокупность компьютерных алгоритмов.

Алгоритмов очень сложных, имеющих дело с огромными объемами данных.

Еще ряд дискуссионных вопросов я обозначил для нашего заседания.

Что отличает искусственный интеллект от человеческого интеллекта?

Есть ли области, в которых искусственный интеллект не может превзойти человеческий?

Является ли творчество той границей, которая отделяет искусственный интеллект от человеческого?

Что мы можем ожидать от развития новых технологий?

Какие проблемы и вызовы они ставят перед человечеством?

Не приведет ли развитие искусственного интеллекта к деградации человека и человечества в целом?

Мы постараемся обсудить:

– риски, которые провоцирует применение искусственного интеллекта в разных сферах;

– проблемы безопасности и конфиденциальности данных;

– проблемы ответственности;

– использование искусственного интеллекта в ситуациях, когда не ясно, кто несет ответственность за совершаемые действия;

– риск потери контроля над развитием технологий искусственного интеллекта, и так далее.

Надеюсь, эти вопросы затронут и докладчики, и участники дискуссии.

акад. Тосунян Г.А.

Перейду к нашей повестке дня.

С докладом «Человек и система искусственного интеллекта» выступит наш выдающийся коллега – академик РАН Владислав Александрович Лекторский, доктор философских наук, декан философского факультета Государственного академического университета гуманитарных наук, председатель совета РАН по методологии искусственного интеллекта и когнитивных исследований.

Пожалуйста, Владислав Александрович.

ДОКЛАД 1

ЛЕКТОРСКИЙ В.А.
акад. РАН, д. филос. н., главный научный сотрудник
Института философии РАН

ЧЕЛОВЕК И СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Спасибо, Гарегин Ашотович.

Вы поставили множество вопросов.

Их можно обсуждать очень долго, а время у меня ограничено.

В течение многих лет я являюсь председателем Научного совета по методологии искусственного интеллекта и когнитивных исследований Российской академии наук.

Эти вопросы мы обсуждаем там ежемесячно, встречаемся по два раза в месяц, проводим конференции, выпускаем книги.

Вопросов очень много, и их становится все больше.

60 лет назад в Америке придумали термин, по-английски – artificial intelligence.

Перевели по-русски как «искусственный интеллект».

Но intelligence – это не совсем интеллект.

Слово «интеллект» в английском языке тоже есть.

Intelligence – это нечто близкое к интеллекту, это способность к каким-то рассуждениям.

Специалисты решили, что на основании имевшихся тогда разработок с использованием уже появившихся компьютеров можно это новое направление исследований и технических, технологических разработок обозначить таким образом.

С тех пор эти исследования бурно развиваются.

Были периоды, когда большие надежды возлагались на искусственный интеллект (ИИ).

Предполагалось, что он сравнится с человеческим или даже превзойдет человека.

Потом были какие-то спады.

Последние десять лет или даже больше идет активнейшее развитие этих исследований.

Исследования в области искусственного интеллекта можно разделить на два периода.

Первоначально речь шла о том, что можно создать искусственные системы, которые будут решать интеллектуальные задачи, обычно решаемые человеком.

То есть имитировать человеческие интеллектуальные способности или усиливать их.

Человека можно и превзойти.

Так и случилось.

Сначала такие системы были основаны на логике.

Это были логические системы искусственного интеллекта.

Казалось, что это – генеральное направление в развитии ИИ.

Ситуация изменилась 20 лет назад.

Стали появляться искусственные интеллектуальные системы, основанные уже не просто на логике, а на машинной обработке больших данных.

Появились нейронные сети, обучаемые интеллектуальные системы.

Сейчас идет активное развитие этого направления.

Стоит вопрос: что с этим делать, и куда все это пойдет?

То, что сначала казалось каким-то довольно безобидным, сейчас начинает осознаваться как нечто такое, что меняет всю нашу жизнь.

Мы живем в революционное время, все в мире меняется.

Трансформации – глобальные, геополитические.

Конфликты на Украине, на Ближнем Востоке.

Все это меняет нашу жизнь, трансформирует многие взаимоотношения регионов, стран, культур, цивилизаций.

Но то, что сейчас происходит в связи с искусственным интеллектом, это тоже революционное преобразование нашей жизни.

Иногда это называют антропологической революцией.

Дело не в том, что появились такие технические приспособления, которые могут взять на себя какие-то человеческие функции.

Но они ставят под вопрос все то, чем человек до сих пор жил.

Великий философ Кант в начале XIX века сказал, что философия должна ответить на вопрос: кто такой человек?

Для этого нужно ответить на три подвопроса.

Что я могу знать?

Что я должен делать?

На что я могу надеяться?

Сейчас на эти вопросы нужно отвечать заново.

То, что казалось неотъемлемой прерогативой человека, сейчас оказывается не совсем так.

Искусственный интеллект – это операции механические, вычислительные.

Казалось бы, человек мыслит иначе, мыслит творчески.

Психологи об этом тоже скажут.

Но многие операции, которые казались специфически человеческими, творческими, сейчас выполняет искусственный интеллект, и часто лучше, чем человек.

Это делается за счет огромных скоростей и обработки огромного массива данных.

Человеческий мозг не может так действовать.

Он работает гораздо медленнее системы искусственного интеллекта.

Может ли искусственный интеллект решать творческие задачи?

Все зависит от того, что понимать под творчеством.

Когда стали создавать эти системы, искусственный интеллект все делал на основе логики.

Он прекрасно доказывал математические теоремы.

Он быстро просчитывал разные варианты и обыгрывал чемпионов в шахматы.

Говорят: человек же создал программу, показал, как играть в шахматы.

То есть искусственный интеллект имитирует то, что сделал человек.

Однако сейчас искусственный интеллект придумывает такие шахматные ходы, до которых человек еще не додумался.

А может быть, и не додумался бы никогда.

Это какая задача? Творческая.

Искусственный интеллект пишет музыку.

Показали ему, как пишет Бах, – он сочиняет фуги в стиле Баха.

И опытный искусствовед не разберет, кто сочинил: Бах или машина.

ИИ может писать рассказы, детективы.

Для этого ему нужно анализировать материалы, десятки или сотни детективов.

Выделит, вычленит, как это строится, какая есть фабула, и сам сочинит.

Если эти рассказы на конкурс подают, критики не могут понять, что это написано не человеком.

Сейчас известный чат-бот пишет работы, обзоры литературы.…

Возникает вопрос: что же ИИ не может?

Искусственный интеллект используется во многих сферах.

Сейчас целый ряд профессий находится под вопросом.

Осуществляемая в настоящее время глобальная цифровизация основана на системе искусственного интеллекта.

Что такое цифровизация?

Это информатизация на новом уровне.

Информация кодируется в цифровой форме, и для ее производства, обработки используют системы искусственного интеллекта.

Цифровизация всюду.

Некоторые специалисты ставят вопрос вообще о будущем финансовой системы, банков.

Предполагается, что банки скоро будут не нужны.

Сейчас Сбербанк – уже не просто Сбербанк, а Сбер, который занимается всем.

У него есть и СберМаркет – российский онлайн-сервис доставки продуктов и товаров с полок магазинов.

И многое другое.

Сейчас Сберу поручили разработки в области искусственного интеллекта.

Поставлена задача – создать так называемый общий искусственный интеллект, general artificial intelligence.

Он должен решать все мыслительные задачи на уровне человека, но лучше человека.

Осуществима ли такая задача?

Для того чтобы ИИ мог что-то сделать, ему нужна информация о том, что уже сделано человеком.

Люди же писали рассказы.

Искусственный интеллект напишет не хуже – может быть, лучше посредственного автора.

Однако «Войну и мир», «Анну Каренину» или «Преступление и наказание» он может написать?

Думаю, что нет.

Писателем должна быть не машина, а человек из плоти и крови.

Надо жить в этом мире, наблюдать природу, ощущать, обонять.

Ведь как описывал природу Тургенев: он ходил по лесам и полям, дышал воздухом, восхищался природой, беседовал с крестьянами об их нуждах.

Можно так описать все это, не будучи человеком, не будучи включенным в культуру, в систему взаимоотношений между людьми?

Чтобы это сделать, нужно искусственно воспроизвести все человеческое: плоть и кровь, всю культуру с ее ценностями, смыслами и проблемами, всю систему межчеловеческих отношений.

Но тогда возникает важнейший вопрос: а зачем это нужно?

Не все, что можно, – нужно.

Любое крупное произведение искусства, будь то музыка, опера или художественный текст, включено в культуру, отвечает на актуальные вопросы своего времени.

Почему роман «Война и мир» писался с 1863 по 1869 год?

В России и мире была особая ситуация.

И роман – ответ на сложившуюся ситуацию.

Искусственный интеллект, если он не включен в культуру, если он не такой же, как человек, никогда этого не сделает.

Считаю, что такие творческие проблемы ему не под силу.

Константин Богомолов поставил оперу «Летучий Голландец».

Певцы поют по-немецки на музыку Вагнера.

Но одновременно идут субтитры на русском языке – совершенно другой текст на сюжет из каких-то историй уголовного типа из нашей хроники.

Якобы то, что поют певцы на музыку Вагнера, – иллюстрация к этим сюжетам.

Искусственный интеллект может сделать такое?

Думаю, что может.

Чтобы такое сделать, не нужен большой творец, это можно скомбинировать, соединить.

А оперу Вагнера в оригинале он не создаст.

В науке сейчас используются способности систем искусственного интеллекта для анализа больших данных.

Человек не может анализировать это, наш мозг так не работает.

Искусственный интеллект это может делать, между огромными данными он может найти какие-то корреляции.

На основе этих корреляций он может делать предсказания, которые могут быть верными.

Правда, иногда они могут быть и ошибочными.

Сейчас в мировой литературе обсуждается вопрос о будущем науки.

Нужна ли научная теория?

Обычно в науке работают так.

Строится научная теория, из теории выводятся некоторые следствия.

Эти следствия проверяются в экспериментах.

На основе теории можно сделать предсказания.

На основе этого можно разрабатывать новые технологии.

Таким образом, с помощью теорий мы выявляем причинные зависимости.

Зная теорию, мы можем предсказать, объяснить, почему эти события произошли, и понять, как это случилось.

Искусственный интеллект, обрабатывающий огромный массив больших данных, во-первых, может делать ошибки.

Во-вторых, он выявляет только корреляцию, но не выявляет причины и их зависимости.

Поэтому те рекомендации, предсказания, которые он дает, даже если они верные, – нельзя понять, нельзя объяснить.

Человеку их «выдают», и хочешь – верь, хочешь – нет.

В XVIII веке произошла великая научная революция.

Возникла наука нового типа, основанная на эксперименте.

Это стало возможным, потому что появился новый тип рациональности.

То, что считалось рациональным в Античности и в Средние века, было подвергнуто сомнению.

Появилось иное понимание природы, взаимоотношений техники и познания.

Это была еще и философская, мировоззренческая революция.

Для античной науки эксперимент был невозможен, потому что эксперимент искажает то, что есть в мире.

Если вы вмешиваетесь в природу, то создаете искусственную ситуацию.

А тут оказалось – нет, с помощью эксперимента как раз можно познавать то, что есть независимо от познания и человеческой деятельности.

В общем, возникла наука нового типа, прежде всего в образе классической механики.

В течение многих веков успешно использовалась астрономическая система Птолемея.

Система Птолемея предсказывала движение светил по небесному своду.

Согласно этой системе, центральное место во Вселенной занимает неподвижная Земля.

Но в предсказаниях, сделанных согласно этой системе, накапливались ошибки.

Коперник предложил более простой способ предсказания.

Если предположить, что не Солнце движется, а Земля вокруг Солнца вращается, тогда все эти предсказания можно делать легче.

Когда Коперник высказал эту идею, это была еще гипотеза.

Она не совсем совпадала с фактами.

Постепенно возникла новая механика со своими законами.

Их обосновывали Кеплер, Галилей, потом Ньютон.

Один из законов классической механики гласит следующее.

Если к телу не приложена никакая внешняя сила, то это тело будет двигаться бесконечно, равномерно и прямолинейно.

Можно видеть такое в опыте?

Никогда, это невозможно.

Это результаты умственного эксперимента и целого ряда допущений.

Вот такой способ мышления, используемый в науке: мысленный эксперимент, если угодно, интеллектуальная фантазия.

А с ее помощью можно прекрасно объяснять то, что происходит в мире, и делать успешные предсказания.

А если бы в это время были системы искусственного интеллекта, обработка больших данных…

Обработал данные, делай предсказание, собирай факты, и все.

Сейчас обработка больших данных используется на практике: в страховом деле, в медицине, диагнозы ставят пациентам.

Используется в науке: в физике, биологии, генетике.

Это все хорошо.

А без теории можно обойтись или нет?

Некоторые люди считают, что теперь можно, что необязательно уметь понимать и объяснять, достаточно делать успешные предсказания с помощью систем ИИ.

Но ведь это какой-то другой тип рациональности: если это рациональность, то не человеческая, какой-то чужой разум, не похожий на человеческий.

Тогда как же жить в непонимаемом мире и просто слушать то, что скажет искусственный интеллект, и доверять ему?

При этом бывает, что выясняемые с помощью искусственного интеллекта корреляции могут быть ложными.

А могут быть и ошибки в работе самого искусственного интеллекта, связанные с допущенными ошибками в программе, с техническими сбоями в системе.

В рамках работы нашего Совета больше года назад было специальное заседание, на котором обсуждали проблему доверия системам искусственного интеллекта.

Когда ему можно доверять, когда нельзя, сколько возможно иметь процентов доверия в разных случаях?

Считается, что если в каких-то случаях есть 80% доверия к рекомендациям ИИ, то на них можно и положиться.

Но ведь бывают такие случаи, где 80% – это недостаточно.

Бывают опасные ситуации: в медицине, где речь идет о жизни и смерти человека, в случае применения смертоносного оружия.

Искусственный интеллект меняет нашу жизнь, помогает решать массу проблем.

Во время эпидемии ковида без искусственного интеллекта было бы невозможно.

Нельзя было бы дистанционно общаться, работать, обучать студентов, детей в школе.

Мы поняли, насколько это важно.

Поэтому гонки в области цифровизации, искусственного интеллекта идут во всем мире.

Китай очень сильно выдвигается на передовые позиции.

Но возникает масса проблем.

Как человеку вообще здесь действовать?

Я вернусь к трем вопросам, сформулированным Кантом.

Что я могу знать?

Что может знать искусственный интеллект?

Если мы полагаемся на предсказания, созданные в результате обработки массива больших данных, это есть знание или нет?

Что считать знанием вообще?

Это вопрос, которым философы занимались со времен античности.

Платон сформулировал три компоненты всякого знания.

О них до сих пор идут дискуссии.

Большинство философов придерживается мнения, что это было сформулировано Платоном правильно.

Первый компонент знания: некоторое полагание того, что что-то в мире происходит.

Это может быть выражено в виде слов, устно или письменно.

Это может быть словесно не выражено: вы думаете, предполагаете или полагаете нечто в процессе восприятия.

Второй компонент: ваше полагание должно быть истинным.

Если что-то не истинно, это не есть знание.

Третий компонент: ваше полагание должно быть обоснованным.

Если вы высказали какое-то суждение и оно является истинным, это не обязательно знание.

Например, я вошел в комнату и вижу там часы.

Меня спрашивают: «Сколько времени?»

Я на часы посмотрел: «Шесть часов».

Оказывается, на самом деле шесть часов, то есть мое высказывание истинно.

Однако это не было знанием, поскольку часы не ходили, остановившись на шести часах.

Просто в тот момент, когда я вошел, случайно это указание на шесть часов совпало с тем, что было на самом деле.

А вы были уверены в том, что часы идут правильно?

Нет. Значит, вы не знали.

В случае искусственного интеллекта на чем основано обоснование?

На обработке больших данных.

Но это не есть обоснование.

Что я могу знать о самом себе?

Для того, чтобы действовать как автономный агент в мире, чтобы отвечать за свои поступки, я должен знать, что делаю, знать себя, знать мотивы своих действий.

А получается, что искусственный интеллект будет знать обо мне больше, чем я сам знаю о себе.

Если вживить какие-то чипы в организм, то ИИ будет знать, чем я болею.

Может быть, я и не догадываюсь об этих болезнях.

Зная это обо мне, ИИ соединится с моим врачом и скажет, какие мне лекарства нужно принять.

ИИ за меня решает, что я должен делать.

Я – не хозяин в собственном теле и не хозяин в собственном доме, если это – «умный дом».

Я нахожусь в зависимости от искусственного интеллекта.

На что же можно надеяться в этой ситуации?

Надежду специалисты по ИИ нам дают.

А предлагается ни много ни мало – цифровое бессмертие.

Даже есть российская футурологическая организация, на сайте которой указано, что ожидает нас в XXI веке.

На этом сайте сказано, что в 2045 году будет создано цифровое бессмертие.

Если знание о вас записано в мозгу, то эту информацию можно будет переписать на цифровой носитель.

И вы будете жить на цифровом носителе.

Можно, конечно, вашу психику пересадить в другое тело.

Но тело – не только ваше, но и чужое – не вечно, смертно.

А на цифровом носителе можно жить почти бесконечно.

Такая перспектива вас воодушевляет?

Ведь в этом обществе цифрового бессмертия за вас всем будут распоряжаться системы искусственного интеллекта.

Вы будете их слушаться и жить, не имея возможности что-то делать самостоятельно.

Вам будет казаться, что вы что-то реально делаете, а на самом деле вы в носителе сидите и вам только кажется, что вы что-то делаете.

Раньше подобные идеи казались фантастическими, даже бредовыми.

Сейчас это всерьез обсуждается, и есть целое массовое движение трансгуманизма.

Это движение полагает, что человек – не высшая точка эволюции.

Он должен уступить место каким-то другим системам – может быть, системам искусственного интеллекта.

Трансгуманизм – это идея о постчеловеческом будущем.

Недавно было заседание Российской академии образования, которой исполнилось 80 лет.

И один из приветствующих сказал, что в системе российских образовательных кодексов записан пункт о гуманизации образования.

Но нужно от гуманизма отказаться, сказал этот человек.

Гуманизм, как он утверждал, уже устарел.

Не знаю, серьезно ли он говорил.

Между тем одна из традиционных ценностей, которые мы сейчас защищаем, – это достоинство человека.

Нужно охранять достоинство.

Но ведь достоинство – это уважение к другим, уважение к себе, это самоуважение.

Если человек не уважает себя и других, его не уважают другие.

Но в таком случае он не может быть активным, полноценным членом общества.

Но ведь достоинство и есть гуманизм.

Между тем какие-то вещи, которые веками были сами собой разумеющимися, сейчас в самом деле подвергаются сомнению.

Приведу пример.

Есть презумпция невиновности.

Если человек не совершил преступление, он не виноват.

Это первое.

И второе, есть принцип свободы воли.

Человек отвечает только за те поступки, которые он совершил сознательно и по своему велению.

Если он этого не делал, если у него нет свободы воли, если он был под гипнозом или еще каким-то образом его лишили сознания и свободы воли, то он не ответственен за свои поступки.

Подчеркну, что есть очень разные культуры и разные политические системы в мире: монархия, демократия, авторитарные структуры.

Политические свободы могут быть разными.

Но во всех обществах, начиная от самых простых, всегда считалось, что человек отвечает за то, что он сделал.

А это значит, что у человека предполагается свобода воли, без этого любая культура просто не существует.

Сейчас, в связи с обработкой больших данных, предлагается следующее.

Ведется наблюдение за каким-то человеком.

Известно, что он ведет себя плохо и может совершить преступление.

Для того чтобы он это преступление не совершил и не подвергал опасности других членов общества, нужно его изолировать, посадить в тюрьму.

Иными словами, предлагается отказаться от презумпции невиновности.

Это серьезно предлагается.

Где-то года два назад вышла нашумевшая книжка «Перезагрузка», в которой представлены тексты двух авторов.

Один из них – Клаус Шваб, немецкий экономист, основатель Всемирного экономического форума в Давосе, второй – известный журналист.

Авторы, ссылаясь на глобальную цифровизацию и приветствуя этот процесс, как раз предлагают отказаться от презумпции невиновности и лежащей в ее основе идеи свободы воли.

Некоторые современные специалисты по мозгу считают, что якобы были эксперименты, которые доказывали, что никакой свободы воли нет.

На самом деле эксперименты, на которые ссылаются, ничего такого не доказывают.

Но это особый разговор, пока не буду касаться этой темы.

Хочу сказать о цифровой памяти.

Все, что вы делаете, записывается на цифровом носителе.

И на основе ваших поступков можно выносить суждение, как к вам относиться.

В некоторых странах это уже используется.

Чем опасна цифровая память?

К внешнему носителю вашей памяти имеют доступ другие люди, которые могут вмешиваться в вашу память и контролировать ваше поведение.

К тому же цифровая память сильно отличается от обычной памяти.

Во-первых, обычно вы не все вспоминаете, не все вам в жизни нужно.

Что-то уходит в сферу бессознательного.

А какие-то вещи не то что ушли в бессознательное, вы просто не хотите вспоминать их.

Они неприятны, но вы не забыли.

Может быть, вы совершили какой-то плохой поступок.

И вы пытаетесь своей жизнью доказать, что вы не такой плохой.

Вы пытаетесь как-то реабилитировать себя, исправить то, что когда-то сделали.

Но без конца думать об этом не можете – иначе жить станет невозможно.

А цифровой носитель будет вам постоянно напоминать о том, о чем вы не хотите думать непрестанно.

Одна из важных особенностей человека в том, что он может что-то забывать.

Если бы не было забвения, жизнь была бы невозможна.

Аргентинский писатель Борхес хорошо знал философию и писал философские рассказы.

У него есть рассказ о человеке, который все помнил.

Но помнить все нельзя, и этот человек не мог жить.

Может ли быть искусственное сознание?

Об этом тоже ведутся дискуссии.

В июне в Российской академии наук состоялся конгресс, посвященный в том числе искусственному интеллекту.

В его работе принимали участие в основном математики, а также психологи, философы.

Я тоже выступал в числе докладчиков.

Одна из секций называлась «Математическая теория сознания. Возможно ли искусственное сознание?».

Некоторые люди считают, что оно возможно.

Тогда возникает вопрос о том, что считать сознанием.

Если считать, что сознание – это просто некие когнитивные функции, то это одно понимание сознания.

Однако сознание к этому не сводится.

Современной науке ясно, что большая часть нашей мыслительной деятельности проходит бессознательно.

Мозг анализирует, строит гипотезы, классифицирует, делает выводы.

Мы этого не осознаем.

Только какая-то часть работы мозга нами осознается.

Это то, что называется «феноменальным сознанием», – индивидуальный опыт, который нами переживается.

Такое сознание и есть у человека.

Есть ли оно у высших обезьян?

Наверное, есть.

У собаки?

Хотелось бы думать, что есть.

А вот у летучей мыши?

Один философ написал рассказ о том, что значит быть летучей мышью.

Этот рассказ уже 40 лет обсуждают.

Можем ли мы себя представить на месте летучей мыши, которая относится к млекопитающим?

У нее абсолютно нет зрения.

Она ориентируется с помощью эхолокации, издает звуки, которые отражаются от предметов, эти звуки улавливаются, и строится картина мира.

Летучая мышь прекрасно ориентируется в мире, не имея зрения.

Можем ли мы представить переживания летучей мыши, не говоря уже о других существах?

Но где это феноменальное сознание появляется?

В чем его эволюционная роль, биологическая роль?

Это сложный вопрос, на который пока нет общепризнанного ответа.

Но ведь именно с сознанием (феноменальным сознанием) связана наша ответственность.

Если я совершал какие-то поступки, не сознавая, что я делаю, я за них не отвечаю.

Чтобы быть ответственным существом, необходимо иметь свободу воли.

А свобода воли предполагает феноменальное сознание.

Есть ли возможность искусственно воспроизвести сознание такого типа?

У меня большие сомнения на этот счет, и я думаю, что нет.

Какие-то функции сознания, то, что происходит в области когнитивного бессознательного, воспроизвести можно.

Кстати, больше, чем мы думали раньше.

Это даже могут быть угрожающие вещи.

А вот воспроизведение феноменального сознания – это большой вопрос.

Как всем сегодня ясно, без искусственного интеллекта невозможна современная цифровизация.

А значит, невозможна вообще современная жизнь.

Мы уже в этой реальности живем и будем двигаться дальше в этом направлении.

Вопрос в том, как сделать так, чтобы все эти достижения были полезны человеку и не навредили ему.

Сейчас стоит вопрос о философско-гуманитарной экспертизе новых технологий.

В 1975 году в американском городе Асиломар собрались биологи и приняли декларацию на какое-то время прекратить некоторые опыты в сфере молекулярной биологии.

Потому что была опасность, что можно создать какие-то искусственные вирусы, которые выйдут из-под контроля.

На какое-то время эти опыты были прекращены.

То есть столкнулись две важнейшие человеческие ценности – ценность познавательная и ценность этическая.

До сих пор казалось, что одно другому не противоречит.

Однако выяснилось, что они могут вступить в конфликт.

А около года тому назад в том же городе Асиломар собрались другие люди во главе с Илоном Маском и приняли другую декларацию.

В ней отмечалось, что на какое-то время нужно заморозить, затормозить некоторые исследования в области искусственного интеллекта, потому что это может быть угрозой человеку.

Затормозить не навсегда, но на какое-то время.

Пока не нашли способов противодействия возможным неприятным последствиям этих разработок.

Чем мы занимаемся, когда пытаемся что-то познать, а затем что-то сделать на основе нашего знания?

Мы отвечаем на три вопроса: почему, как, зачем?

Мы хотим понять, почему что-то происходит, какие причины и зависимости существуют в мире.

Только на основе этого знания можно действовать.

Наука занимается выявлением причин.

Не только наука, но и мы в нашей жизни без этого не можем обойтись.

Второй вопрос – как?

Я ставлю себе какие-то цели и хочу их достичь.

Я должен знать, почему что-то происходит в мире, и должен изобрести какие-то средства, чтобы этой цели достичь.

Скажем, можно копать яму.

У меня есть лопата.

Я знаю, как копать.

Вид ямы будет зависеть от того, как ее копать.

А сегодня есть современные технологии для достижения более сложных целей.

Все технологии – это ответ на вопрос «как?».

И третий вопрос – зачем?

Всегда ли можно что-то делать, и всегда ли нужно это делать?

К А.П. Чехову однажды пришла женщина и спросила: «Антон Павлович, если копать огромный тоннель так, чтобы он другим концом выходил на противоположной стороне земного шара, это можно сделать?»

Чехов ответил: «Я не знаю, может быть, и можно. Но не нужно».

Не все, что можно, – нужно.

Вопрос «зачем?» сейчас стоит очень остро.

Говорят, что системы искусственного интеллекта приобретают некоторую автономию, они самообучаются.

Так оно и есть: они самообучаются, приобретают автономию.

Говорят, что вообще может возникнуть сообщество этих искусственных приспособлений со своей культурой.

Они будут решать собственные проблемы безотносительно того, что от них хочет человек.

А человек будет существовать как бы отдельно со своей культурой.

Но ведь эти машины не напишут «Войну и мир».

Для их жизни это не нужно, а для человека это важно.

Считается, что мир человека и мир систем искусственного интеллекта будут вместе существовать.

Но возникает вопрос: а зачем это делать?

Вопрос о контроле над новыми технологиями – это животрепещущий вопрос, вопрос выживания человека.

Антропологическую революцию некоторые называют антропологической катастрофой.

Человек создал такую искусственную ситуацию, которая уже начинает ставить под вопрос то, без чего он не может быть человеком.

Встает вопрос о том, как в этих условиях остаться человеком.

Конечно, приходится меняться.

Выжить в современном мире, не меняясь, невозможно.

Но меняться нужно так, чтобы не потерять человеческие черты и не превратиться в раба искусственного интеллекта или в какое-то жалкое его подобие.

Эта проблема, которая недавно казалась фантастической, сейчас стала практической.

Теоретические вопросы: что такое познание, знание, что такое рациональность, что такое сознание, что такое «Я»?, – которые философами обсуждались тысячи лет, становятся практическими вопросами.

Что такое свобода воли, откуда она берется, есть ли она вообще, или это просто иллюзия?

Это и вопрос о том, как ориентироваться в мире и взаимодействовать с системами искусственного интеллекта, чтобы они не вытеснили человека, а были его помощниками.

Думаю, что этот вопрос – решаемый.

Спасибо.

Тосунян Г.А.: Спасибо, Владислав Александрович.

Вас можно слушать бесконечно долго, настолько Вы интересно излагаете свои мысли.

Кривов Виктор Дмитриевич, Вам слово.

КРИВОВ В.Д.
д. э. н., заведующий кафедрой социологии и менеджмента общественных процессов Высшей школы современных социальных наук МГУ им. М.В. Ломоносова

Кривов В.Д.: Спасибо, уважаемый Гарегин Ашотович.

Спасибо, уважаемый Владислав Александрович, за Ваш фундаментальный, энциклопедический доклад на самом острие научного поиска и нашей жизни.

Вопрос как раз практический, который затрагивает, наверное, сотни тысяч наших студентов, преподавателей: как быть?

Допускать возможность применения GPT при подготовке курсовых, выпускных работ, диссертаций?

А если не допускать, то возможно ли не допустить использование этих GPT?

Спасибо.

Лекторский В.А.: Я бы не допускал.

Знаю, что это делается.

Иногда трудно распознать, кто написал: студент или эта машина.

Если это будет машина, тогда зачем студенту учиться, что-то ему знать?

Можно просто дать задание системе искусственного интеллекта.

Что касается внедрения всех этих вещей в образовании, то для образования это, конечно, с одной стороны, благо.

Без этого сейчас нельзя.

А с другой стороны…

Я сам преподаю, и во время эпидемии приходилось преподавать онлайн.

Преподавание онлайн и преподавание вживую – это разные вещи.

А физиологи и специалисты по мозгу говорят, что когда вы вживую с кем-то беседуете, например, с преподавателем, то оказывается, происходит глазной контакт.

Вы смотрите в глаза друг другу.

На экране это не так.

И вообще общение онлайн не заменяет личного общения.

Один из наших деятелей в сфере образования несколько лет назад предложил такую систему.

Если хороший преподаватель читает лекции, то стоит записать курс его лекций.

А преподавателя можно уволить, он больше не нужен.

Пусть студенты слушают эти лекции.

Или еще пример ситуации, похожей на анекдот.

Записали какую-то лекцию, преподавателя нет.

Поставили видеозапись на кафедре, а студенты должны слушать и записывать.

Прошло несколько таких лекций, а потом декан факультета решил оценить результат.

Зашел в аудиторию, там на кафедре работает эта система, а вместо студентов стоят компьютеры, они записывают.

Люди не нужны, компьютер с компьютером общается.

Образовательные проблемы – это большой разговор.

У нас в Академии образования было специальное заседание нашего Совета, посвященное проблеме влияния цифровизации на образование и на социализацию подростков.

Это очень вредно: подростки сидят в гаджетах, а социализация тормозится и приобретает какие-то другие формы.

А самое главное, как выясняется, чтение с компьютера цифровых книг и чтение обычных книг, на бумажных носителях, – разные типы чтения.

Являясь еще членом Академии образования, я возглавляю там Комиссию по чтению.

Мы специально обсуждаем проблемы чтения с психологами, физиологами, педагогами.

Когда нужна информация, как говорят сами студенты, легче обратиться в Интернет.

Но когда хотите что-то внимательно, глубоко читать, следует обратиться к бумажному носителю, там другое восприятие текста.

Можно пролистать странички в Интернете, там чтение иного типа.

И в результате такой работы возникает клиповое мышление.

А сегодня требуется творческое и критическое мышление.

Люди, которые выросли в цифровой культуре, имеют дело только с цифровыми носителями, с компьютерами, с сотовыми телефонами.

В этих условиях, оказывается, у человека что-то важное теряется.

Известно, что в хороших американских школах детям запрещают пользоваться гаджетами.

Стив Джобс запрещал своим детям использовать гаджеты, хотя он сам занимался разработкой этих вещей.

Если студент будет постоянно работать с помощью GPT, то он ничего не будет знать.

Но эта работа нужна ему в помощь, нужна для того, чтобы он мог что-то сделать сам.

А если работа написана, оценка получена, и даже преподаватель не разобрался, кто написал ее, что будет дальше?

Если все это разрешить, то будем иметь неприятные последствия.

Поэтому лучше этого не делать.

Тосунян Г.А.: Спасибо.

Павел Алексеевич, пожалуйста.

МЕДВЕДЕВ П.А.
д. э. н., проф., финансовый омбудсмен
Ассоциации российских банков

Медведев П.А.: Спасибо большое.

Не очень понял в Вашем докладе одну деталь, может быть, не самую важную.

Я так услышал, что Вы к искусственному интеллекту относитесь как к какой-то ущербной сущности.

Вы, в частности, сказали, что творчество искусственному интеллекту недоступно.

Он не может писать, как Тургенев, поскольку для того, чтобы писать, как Тургенев, надо наблюдать окружающую жизнь, надо быть включенным в культуру.

Скажите, пожалуйста, а почему Вы уверены, что искусственный интеллект не может наблюдать жизнь и не может быть включен в культуру?

Великий математик Колмогоров, бывало, говаривал: «Если вы что-то утверждаете, подкрепите ваше утверждение правдоподобными рассуждениями».

В чем состоит Ваше правдоподобное рассуждение в поддержку Вашего тезиса?

Это первая часть вопроса.

Вторая часть.

Я тоже не могу писать, как Тургенев.

Значит ли это, что мой интеллект – тоже ущербный или, во всяком случае, не естественный?

80 с лишним лет я прохожу сниженный вариант критерия Тьюринга, и никто мне ни разу не сказал, что мой интеллект – не естественный.

Хотя вполне возможно, что люди не делали этого из вежливости.

Спасибо.

Лекторский В.А.: Попробую ответить кратко, хотя можно говорить об этом долго.

Я не утверждал, что искусственный интеллект – ущербный.

Никакой он не ущербный, он превосходит человеческий интеллект во многих отношениях.

Он может делать то, чего человек делать не может.

Он может анализировать системы больших данных, выяснять там корреляции, чего мы делать не можем.

И без него мы не можем обойтись.

Я говорю о том, что у него есть особенности, которые отличают его от человеческого интеллекта.

Может ли он решать творческие задачи?

Конечно, может решать.

Я приводил пример: ИИ способен находить такие варианты шахматной игры, которые мозг человека находить не может.

Это творческая задача.

Есть игра в го, еще более сложная, чем шахматы.

И сейчас система искусственного интеллекта играет в го лучше человека.

Искусственный интеллект может писать рассказы, различные работы.

Может писать музыку в стиле Баха, Моцарта, в стиле любого композитора, если ему дать материал для этого, чтобы он понял, как это пишется.

Это творческие задачи, он может самообучаться, может делать многое.

Просто речь идет о том, что его творчество – это все равно не человеческое творчество.

И человек, чтобы остаться человеком, должен знать о себе, что он может, чего не может, и должен принимать ответственные решения.

Человек не может просто передоверить это искусственному интеллекту, потому что искусственный интеллект будет рассуждать иначе.

Можно попытаться сделать искусственный интеллект, похожий на человеческий.

Но это никому не нужно.

Если даже возникнут автономные системы искусственного интеллекта, для них не важны те проблемы, которые важны для нас.

Мы иначе устроены, и у нас есть культура.

Или вы решите воспроизвести культуру искусственно, со всеми ее смыслами, всем, что хорошо, что плохо?

А зачем?

Какая в этом случае цель?

Чтобы искусственный интеллект был похож на вас?

А зачем ему быть похожим?

Ему не надо быть похожим.

Какие-то задачи он решает лучше вас, но иначе, чем вы.

Нужно найти способ взаимодействия с ним, чтобы он служил вам, а не мешал, не вытеснял вас.

Нужно, чтобы он делал ту работу, которую вы не можете сделать в силу особенностей вашего организма и вашего мозга.

Мозг работает гораздо медленнее, чем система искусственного интеллекта.

Если даже возникнут сообщества этих систем ИИ, этих автономных машин, им не нужны философия, наука, им не нужна «Война и мир».

Это нужно для человека, для нашей культуры.

Это будут сообщества разного типа, культуры разного типа, если Вы считаете, что какая-то культура будет у систем ИИ.

Именно об этом я говорил, а не о том, что он – ущербный.

Он просто другой, не человеческий.

Нам его трудно понять, его нельзя объяснить.

Искусственный интеллект может ошибаться.

Если каждый раз ему слепо доверять, то можно пойти не по тому пути.

А в каких-то ситуациях без него невозможно обойтись.

Вопрос в том, чтобы найти способы взаимоотношения с этими системами.

Был большой скепсис насчет возможностей искусственного интеллекта.

Еще 20 лет назад считали, что он того не может, другого не может.

А сейчас уже энтузиасты в области так называемого общего искусственного интеллекта, работающие в Сбербанке, считают, что все интеллектуальные задачи он будет решать на уровне человека и гораздо лучше, чем человек.

Может быть, он будет делать это более рационально, чем человек.

Но это не человеческая рациональность.

А наука не нужна системам ИИ.

Зачем им наука, зачем им теория относительности и квантовая механика?

ИИ будет просчитывать массивы больших данных.

А построением теорий пусть люди занимаются.

Человек хочет сохраниться, или ему это не нужно?

Вопрос стоит именно об этом.

Без искусственного интеллекта человек никуда не пойдет, этот рубеж уже пройден.

Поэтому нужно найти способы, как взаимодействовать с ИИ и как предотвращать возможные опасности на этом пути.

Специалисты по искусственному интеллекту эти вещи понимают.

Познание для человека – это самостоятельная ценность.

А сегодня наука превратилась в технонауку.

Мы разрабатываем новые технологии.

Однако наука всегда имела и имеет еще и другую функцию – познание мира.

И никуда от этого не деться.

Для систем искусственного интеллекта так проблема не стоит.

Им это не нужно.

Тосунян Г.А.: Коллеги, с докладом «Человек и искусственный интеллект: подходы к взаимодействию» выступит академик РАН, директор Института психологии РАН, заведующий кафедрой общей психологии МГУ Дмитрий Викторович Ушаков.

ДОКЛАД 2

УШАКОВ Д.В.
акад. РАН, д. п. н., профессор, директор Института психологии РАН, заведующий кафедрой общей психологии факультета психологии МГУ им. М.В. Ломоносова

ЧЕЛОВЕК И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ПОДХОДЫ К ВЗАИМОДЕЙСТВИЮ

Спасибо большое.

Мой доклад оказался по названию очень похож на доклад Владислава Александровича.

Но я буду говорить о двух возможных сценариях или двух образах отношений человека и искусственного интеллекта.

Первый я обозначил как «цифровой ангел».

Это термин мы с профессором Сергеем Юрьевичем Степановым придумали для обозначения возможной роли искусственного интеллекта.

И для понятных культурных аналогий с ангелом-хранителем.

Второй – это образ Франкенштейна.

Франкенштейн – это персонаж повести Мэри Шелли, которая была женой известного поэта Шелли, из круга Байрона.

Это повесть о создании искусственного человека, можно сказать, искусственного интеллекта, который опасен и уничтожает своего создателя.

Эти два образа – некие полюсы или варианты сценариев.

Между которыми наверняка будущее пойдет особыми путями.

Что же может делать уже в скором времени технология, которая может быть названа индивидуальным цифровым ангелом?

- Она может выполнять функции психолога;
- Она может предоставлять нам некоторую помощь в организации себя, своей жизни, своего состояния;
- Она может помогать взаимодействовать с людьми. С людьми близкими к нам, с людьми не очень близкими – при формальном взаимодействии, и так далее;
- Она может оказывать нам поддержку в нашей профессии в той степени, в какой мы взаимодействуем с другими людьми;
- Она может нас защищать от дипфейков и других видов обмана, которые в большой степени генерирует сам же искусственный интеллект.

Покажу на конкретном примере.

В настоящее время идет речь о создании систем искусственного интеллекта, которые могут анализировать состояние человека и каким-то образом помогать это состояние поддерживать.

Например, мы сейчас сидим, нас снимает камера.

Возможны технологии, которые на основании этого видео позволяют анализировать, кто в каком состоянии.

Кому интересно то, что я говорю, кому неинтересно.

Кто относится позитивно, кто негативно.

Мы знаем, что уже работают системы контроля состояния водителя.

Можно предположить и более сложные пока вещи: полет на Марс.

Несколько месяцев летит группа людей.

Неизвестно, вернутся или не вернутся.

Сигнал до Земли идет с задержкой, человеческое общение нарушается.

Узкий коллектив.

Что там происходит?

В Институте медико-биологических проблем проводятся подобные эксперименты.

В течение какого-то времени людей держат вместе.

А как оценивать, в каком состоянии эта группа?

Космические полеты – это один пример из множества.

Например, как поддерживать приемлемое психологическое состояние вахтовиков или операторов атомных электростанций?

Те же задачи – в образовании.

Кто-то в классе внимательно слушает и поглощен занятиями.

Кто-то скучает, кто-то устал.

Учителю надо в реальном времени это понимать.

Директору школы надо понимать, что у Марьи Ивановны дети на уроке скучают и устали.

А у Петра Петровича, наоборот, полны энтузиазма.

Школьному психологу важно понимать, что, допустим, у Машеньки в начале года было все хорошо, а в конце стало что-то плохо.

Может быть, там буллинг или что-то в этом роде.

То же самое относительно персонала компании.

И большой вопрос – психологическое состояние населения.

Владислав Александрович упоминал тот же ковид.

Здесь важно понимать, в каком состоянии население, может ли оно вынести самоизоляцию или его психологическое состояние этого не позволяет и лучше допустить большие темпы заражения.

Можно определять такого рода вещи автоматически, используя искусственный интеллект, который анализирует состояние людей по видео, по тому, что они пишут в соцсетях, и т.д.

Дальше – переговоры и продажи.

Человеку, занимающемуся продажами, важно понимать, на что реагирует клиент.

Здесь тоже может помогать ассистент с искусственным интеллектом.

Есть еще проблема синтеза человеческого голоса.

Например, уже сегодня колонки Алиса говорят достаточно приятным голосом.

Однако они пока не способны к разнообразным и уместным интонациям.

Например, они не могут разыграть пьесу или прочитать стихи, как актеры.

Синтезировать поведение, соответствующее состоянию человека, голос и прочее-прочее – это тоже отдельная задача для искусственного интеллекта в сфере человеческих чувств.

Сегодня уже разработаны кое-какие технологии, способные в первом приближении делать что-то из перечисленного.

Берем, например, видеозаписи людей и выясняем что-то об этих людях.

Выяснить можем у экспертов или протестировав самих испытуемых, например, на личностные свойства или интеллект.

Затем обучаем машину на основе видео – текстов, которые человек произносит, мимики, интонаций – оценивать свойства или состояния людей.

В нашем институте есть патент на такого рода технологии.

В этих технологиях есть, однако, ограничения: они не выходят за рамки того, что говорит эксперт или что показывает тест.

Кроме того, состояния человека вырываются из контекста.

Если человек улыбается, радуется, наверное, это хорошо.

Но если это происходит, скажем, на похоронах, то это очень неадекватно.

Поэтому важно анализировать в контексте.

В результате модели оказываются узкоспециализированными.

Например, модели, которые работают с состоянием водителя, применимы только в этом контексте.

Они хорошо работают, но при выходе за эти пределы они работают очень плохо.

В чем перспектива?

Здесь можно привести аналогию с тем, как построены большие языковые модели, тот же GPT.

Там дается текст и зашумляется одно слово.

То есть одно слово из текста убирается.

Определение психологических свойств личности по видеоинтервью

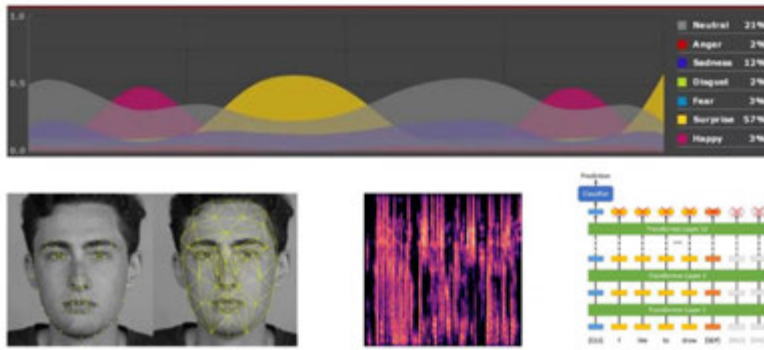

Рис. 1

И машина учится по всему тексту предсказывать, какое там будет слово.

Здесь нет некой внешней разметки, а просто используются корпусы текстов.

То, что называется обучение без учителя.

Мы предлагаем сделать следующий шаг.

Он строится в том числе на так называемом интент-анализе, который мы разрабатываем в нашем институте.

То есть анализе интенций, которые стоят за высказываниями людей.

Человек что-то говорит, но за этим высказыванием стоит некоторая интенция.

Интенция может состоять в том, чтобы сообщить информацию, либо выразить симпатию собеседнику, либо его высмеять, либо самопрезентоваться и так далее.

Определение интенций по тексту в процессе диалогов с учетом контекста

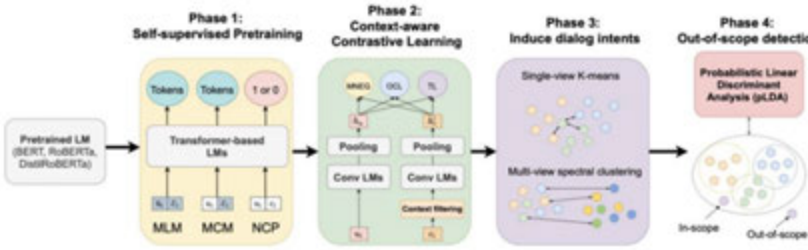

Рис. 2

Мы автоматизируем интент-анализ.

А дальше мы анализируем взаимодействия людей и смотрим шаги, которые люди совершают с точки зрения высказываемых интенций.

Мы анализируем человеческое взаимодействие, как шахматную партию, где играющие обмениваются ходами.

Делают ход, реагируя на то, что делает другой.

При этом анализируется, какие тексты это сопровождает, какие эмоции отражает лицевая мускулатура, как звучит голос, и так далее.

Это следующий шаг, к которому мы идем.

Он позволяет свести различные модальности в единое пространство, состояние человека.

Здесь же можно использовать и самые разные психофизиологические данные.

Начиная от частоты сердцебиения до данных о функционировании мозга.

Мы строим апостериорное пространство психологических состояний.

То есть не априори задаем, что есть радость, страх, горе, а получаем апостериорное пространство состояний с тонкой классификацией.

Потому что мы знаем, что состояния людей всегда специфические и тонкие.

Эти технологии позволят на новом уровне решать задачи, связанные с анализом свойств людей и их состояний, о которых я говорил раньше.

Это может иметь много применений и стать важной составляющей цифрового ангела.

Индивидуальный цифровой ангел – это та инстанция, которая призвана работать в интересах хозяина.

Когда много хорошего может быть сделано, может быть сделано и много нехорошего.

Мы знаем, что если есть ангел, то есть и падший ангел.

И где есть оружие, там есть и противооружие.

Мы знаем, что эти устройства должны работать в интересах своего хозяина.

Но как добиться, чтобы они работали в интересах своего хозяина, когда есть вирусы?

Мы должны понимать, что там, где искусственный интеллект может нам помогать, он требует большого доверия.

А там, где есть доверие, есть возможность злоупотребления доверием.

«Черный» сценарий я обозначил именем Франкенштейна.

Хочу напомнить про совместное заявление Гейтса, Хокинга и других известных деятелей.

Несколько лет назад они предупредили об опасности искусственного интеллекта.

У них был там образ муравейника, наиболее яркий и центральный.

Мы не имеем ничего против муравьев.

Мы не хотим им ничего плохого.

Но, допустим, если мы строим ГЭС и затапливаем территорию, то там погибает муравейник.

Но мы об этом даже не думаем.

Если искусственный интеллект будет иметь столь же мощные вычислительные возможности по отношению к нам, как наши возможности по отношению к муравьям, возможно, что он сделает таким же образом.

Это их предположение.

Для того чтобы это произошло, нужно такое свойство искусственного интеллекта, как интенциональность.

С одной стороны, возможность управления физической энергией со стороны искусственного интеллекта.

А с другой стороны, наличие собственной, независимой от нас мотивации и целеполагания.

Пока мы не отдаем управление физической энергией искусственному интеллекту, он не может затопить наш муравейник.

Но проблема в том, что в той или иной степени мы это даем искусственному интеллекту.

Представим автономный автомобиль.

Если не давать искусственному интеллекту доступ к управлению этой энергией, что там будет работать?

Я уже не говорю про военное использование искусственного интеллекта, которое, конечно, связано именно с большими энергиями.

Второе – наличие независимых мотиваций и целеполаганий.

И здесь нужно различать две вещи.

С одной стороны, независимость целеполагания или мотивации искусственного интеллекта от людей в целом.

То есть может ли произойти так, что искусственный интеллект получит возможность собственного целеполагания и собственной мотивации?

Когда он вообще не зависит от людей.

Другая проблема – зависимость от групп людей.

Когда мы создаем искусственный интеллект, то им всегда управляют какие-то люди или группы людей, которые туда что-то закладывают.

Это проблема не искусственного интеллекта, а наша.

Искусственный интеллект может стать очень мощным средством решения задач отдельных групп людей, возможным орудием этих групп.

Предыдущим таким мощным средством, которое создало человечество, было ядерное оружие.

Но в ситуации с ядерным оружием было несколько иначе.

Ядерное оружие разрабатывается государством.

Существует монополия государства на ядерное оружие.

Между государствами были разработаны механизмы взаимодействия, договоры СНВ и прочее.

Другое дело, что сейчас опять это все шатается, демонтируется, но тем не менее человечество смогло выработать такие механизмы.

А искусственным интеллектом могут владеть группы людей.

Это могут быть крупные корпорации.

OpenAI, или Сбер, или еще кто-то.

Речь идет о том, чтобы эти механизмы защиты от опасности на межгосударственном уровне трансформировались в механизмы регулирования опасности между группами людей.

Напомню термин техно-гуманитарного баланса, который предложил в свое время Акоп Назаретян.

Он утверждал, что должно быть соответствие между развитием технологий и развитием гуманитарных подходов в обществе.

Они позволяют уравновесить эти технологии.

Когда произошла неолитическая революция, то есть развилось земледелие и скотоводство, образовались более крупные поселения людей, начались войны.

Появилась необходимость регулирования этих вещей.

Когда образовались большие цивилизации на основании массового земледелия, типа египетской или месопотамской, сформировалось правление, фараоны, религиозные институты.

Развитие технологий всегда требует создания гуманитарной составляющей, которая не позволит этим технологиям взорвать человеческий мир.

Соответственно, когда возникает искусственный интеллект, вопрос в том, что должна быть гуманитарная составляющая, которая это компенсирует.

Независимость от человечества.

Если искусственный интеллект не просто используется как средство какой-то группой людей, а, как говорили Гейтс, Хокинг и другие, получает свои цели и начинает идти к этим целям, то мы для него как муравьи.

Конечно, это может произойти вследствие технической ошибки.

Но у нас есть довольно хорошее средство контроля технических ошибок.

Вспомним ту же атомную бомбу.

Когда ее создавали, то подозревали, что ядерная реакция может перекинуться на мировой океан.

И тогда были сделаны специальные расчеты.

Чуть ли не Эйнштейн принимал в этом участие.

Эти расчеты показали, что вроде как это безопасно. Слава Богу, на мировой океан не перекинулось, и Земля не превратилась в Фаэтон.

Технические ошибки мы умеем купировать.

Но понятно, что это реальная опасность.

А не возникнет ли необходимость давать искусственному интеллекту доступ к управлению физическими ресурсами?

Допустим, если мы начинаем ставить какие-то более общие задачи типа развития науки или прочее.

Может ли оказаться, что ситуация цели не определена и дальше в этих пределах происходит уход?

И самый сложный, с моей точки зрения, вопрос.

Вопрос о целесообразности.

То есть не может ли у нас возникнуть такая ситуация, при которой мы допустим искусственный интеллект к определению того, чего мы хотим?

Хотя бы потому, что мы хотим часто не очень хороших вещей.

Конфликты, происходящие вокруг нас, свидетельствуют о том, что мы не всегда умеем хотеть того, что целесообразно.

Отсюда возникает вопрос к наукам о человеке.

Чего мы хотим?

И как мы хотим?

Я предложил бы такую модель.

Посмотрим в эволюции хотения.

Инстинкты у животных образуются за счет естественного отбора и выживания.

Если животное не хочет питаться, размножаться, защищаться, оно не дает потомства.

Соответственно, этот индивид кончает свою жизнь, и вид вымирает.

Инстинкты закрепляются путем биологического естественного отбора.

У Докинза есть идея генов и мемов.

Есть биологические гены, а есть культурные мемы.

То есть некие культурные схемы поведения людей.

Такими мемами является в том числе ценностная система.

Мы знаем, что они меняются достаточно быстро.

Например, отношение к ЛГБТ.

Известный английский математик Алан Тьюринг покончил с собой из-за проблем с юстицией в связи со своей гомосексуальностью.

Сегодня отношение к представителям ЛГБТ с точностью до наоборот.

Ценностные системы меняются.

За счет чего они меняются?

Система очень сложная.

Грубо говоря, во-первых, есть некий индивидуальный уровень.

Я принимаю ценности, когда это соответствует моему внутреннему равновесию.

Следующий уровень – социальный.

Пример – кровная месть.

Она имеет некоторую систему психологических установок, которые за ней стоят.

Мы знаем историю осетина, потерявшего семью в авиакатастрофе, который убил авиадиспетчера.

Понятно, что это поддерживается целой серией внутренних установок личности, пониманием, что эти действия справедливы.

Но эта система работает в том случае, когда в обществе на уровне социального равновесия ценностей кровная месть может служить хорошим регулятивом.

До того, как развита судебная система.

До того, как развита правоохранительная система.

Понятно, что это сдерживающий фактор, и общество саморегулируется.

Есть уровень саморегуляции в обществе.

Он позволяет каким-то образом работать обществу.

И в то же время он находит соответствие в нашем внутреннем равновесии.

Но в какой-то момент изменяется способ регуляции соответствующих вещей в обществе.

И тогда он должен изменяться на уровне отдельных индивидов.

Мы знаем, что эти естественные психологические моменты, связанные с той же кровной местью, сегодня подвержены изменению.

Наконец, может происходить конкуренция социальных групп.

Группа, в которой используют неудачные способы регуляции, терпит крах в конкуренции с более успешными группами.

Например, диктатуры.

Они имеют достаточно скоординированную, прочную внутреннюю систему.

Но они неэффективны в решении задач.

И они терпят крах во взаимодействии с другими системами.

Таким способом происходит фактически формирование наших ценностей и желаний.

Система формирования наших желаний или наших ценностей до сих пор позволяла нам развиваться и не допустила катастрофы.

Даже когда был Карибский кризис, наша система внутреннего равновесия, система ценностей не допустила катастроф.

В то же время эта система несовершенна.

Хотя бы потому, что она ведет к конфликтам.

И постоянно ведет к жертвам.

Нет никакого математического доказательства того, что в один прекрасный момент все это не взорвется.

Могут ли быть здесь полезны системы искусственного интеллекта?

Они могли бы позволить нам в перспективе оценивать результаты принимаемых нами решений.

И здесь возникает еще одна очень сложная для интеллекта проблема.

Получается, что человек есть некий момент развития разума.

Некий момент равновесия развивающейся системы в нашей эволюции, в нашей цивилизации.

Сейчас мы подошли к очень опасному моменту.

Мы имеем возможность создавать орудия, которые изменяют эти системы равновесия.

Орудия, направленные не вовне, а орудия, направленные вглубь самого нашего разума.

Более того, проблема усугубляется ускорением темпов.

Наш физик-теоретик Панов построил график скорости эволюции.

Допустим, жизнь на Земле возникла в форме прокариотов около четырех миллиардов лет назад.

Затем появились эукариоты. Тоже одноклеточные, но на это ушло больше миллиарда лет. До появления многоклеточных – тоже миллиард лет.

На все это ушли миллиарды лет, но затем дело пошло быстрее.

Немногим более 300 миллионов лет назад наши предки вышли на сушу.

Динозавры сменились млекопитающими 66 миллионов лет назад.

Человек, каменный век – два или три миллиона лет назад.

А неолитическая революция – 10–12 тысяч лет назад.

Цивилизации – 5 тысяч лет.

Паровая машина – 200 лет.

Персональный компьютер – 30–40 лет назад.

Интернет, смартфоны, искусственный интеллект – совсем недавно.

То есть колоссальное ускорение.

Владислав Александрович упомянул 2045 год…

Времени на понимание и принятие решения не остается.

Или у нас его очень мало.

Может быть, надо было какую-нибудь более радостную ноту для окончания выступления выбрать.

Но…

Благодарю за внимание.

Тосунян Г.А.: Спасибо, Дмитрий Викторович.

Вполне мажорная нота.

Кстати, когда Вы говорили, что к 2045-му все уйдет в цифру, хотел спросить.

А эти исследователи любовь тоже в цифру переводят?

Возможно, это хорошо, и даже эффективно?

Медведев П.А.: Я только что обнаружил, что мой интеллект явно не искусственный.

Я его показал недавно не с лучшей стороны, мне захотелось реабилитироваться.

Вряд ли искусственный интеллект когда-нибудь захочет реабилитироваться.

Вы сказали, что искусственный интеллект может давать консультации в выборе профессии.

Я тоже могу.

Скажите, пожалуйста, как узнать, кто лучше умеет это делать?

Второй вопрос.

Выбор профессии – это очень многопараметрическая задача.

60 с лишним лет тому назад замечательный математик Ричард Э. Беллман попытался решать похожие задачи.

Тогда уже появились вычислительные машины, которые умели немного думать.

Это спровоцировало идею с помощью компьютера ставить диагноз пациенту, вводя в память машины значения параметров, характеризующих состояние человека (вес, рост, кровяное давление, температуру, уровень гемоглобина и т.д.) и обрабатывая данные с помощью подходящего алгоритма.

Оказалось, что эта задача неразрешима, и не потому, что во времена Беллмана компьютеры были еще недостаточно умными, а потому что на земле не хватало (и до сих пор не хватает) людей, чтобы наполнить модель определения диагноза данными.

Тот, кто не знаком с этим драматическим эпизодом в жизни науки, может погуглить: Беллман «Проклятие размерности».

Как искусственный интеллект преодолел «проклятие размерности» при формировании советов по выбору специальности?

Спасибо.

Ушаков Д.В.: Павел Алексеевич, Вы задали сразу несколько вопросов.

Все, конечно, очень глубокие и, безусловно, интересные.

Вопрос о параметрах и о том, что можно предсказать, а что нельзя.

Если мы берем выбор профессии, конкретную задачу, то она обычно решается таким образом.

Есть классификация профессии, например, человек-человек, человек-предмет, человек-знание.

Соответственно, есть определенные методы, по которым определяется склонность человека к той или иной профессии.

Понятно, что профориентацию лучше проводить не когда человек выбирает институт.

Если Вы получили психологическое образование, то после этого есть еще много развилок.

Можно стать консультирующим психологом, а можно – экспериментальным исследователем когнитивных процессов.

Тип личности и способности в этих случаях требуются разные.

Оптимальная ситуация – то, что называется моделью турникета, вращающихся дверей.

Вы можете войти и выйти.

Например, во Франции в университете нет вступительных экзаменов, вообще конкурса нет.

Но большой отсев после 1, 2, 3-го курса.

Вы можете спокойно зайти туда, но не факт, что дойдете до конца.

И это, на мой взгляд, оптимальный вариант.

Соответственно, вы не должны делать систему, которая принимала бы однозначные решения и не давала права выбора.

Стоит учитывать, что сам мир профессий очень быстро меняется.

И бывает, что пока вы получаете образование, попадаете уже в другой мир.

Что касается наличия каких-то вещей в мире, которые, возможно, непрогнозируемы и непознаваемы в принципе.

Наверное, есть вещи, которые не можем вычислить ни мы, ни искусственный интеллект.

Например, что будет с Землей или с нашими государствами через два года.

Думаю, что вряд ли кто-то в мире на этот вопрос сегодня ответит.

Скорее всего, есть вещи невычислимые, исходя из принципа включенного наблюдателя.

То есть для того, чтобы вычислить некоторую другую систему, вторая система должна быть не меньше ее.

Тогда она не может вычислить самое себя.

Поэтому мир, в который включен познающий субъект, не может быть полностью просчитан никаким субъектом, включая искусственный интеллект.

Но это ни в коем случае не умаляет того, что искусственный интеллект может высчитывать какие-то вещи, как мы.

И даже больше нас.

***проф.* КАЛАШНИКОВ С.В., *акад.* УШАКОВ. Д.В., *акад.* ЛЕКТОРСКИЙ В.А.**

КАЛАШНИКОВ С.В.
д. э. н., проф., заведующий кафедрой труда и социальной политики Института государственной службы и управления РАНХиГС при Президенте РФ

Калашников С.В.: Дмитрий Викторович, возьмем классику: теорию когнитивного развития швейцарского ученого, философа и психолога Жана Пиаже.

Искусственный интеллект может повторить операции Пиаже до 12-летнего возраста, например?

Различает он грань в мышлении у детей разных возрастов или нет?

Ушаков Д.В.: Сергей Вячеславович, Вы затронули такую тему, которая для меня очень важна.

Современный искусственный интеллект – это не логические операции.

Это что-то более близкое к интуиции.

Машинное обучение, нейросети, они в какой-то степени моделируют одну из сторон деятельности человеческой когнитивной системы, а именно интуитивную.

Но не логическую, которую изучал Пиаже.

Интуиция – это когда мы что-то предполагаем, но не знаем почему.

Например, я вижу человека.

Я думаю, что ему можно доверять.

Почему – не знаю.

Это интуиция.

Так же примерно работает искусственный интеллект.

Искусственный интеллект, современные нейросети – это одна сторона человеческой реальности.

Интуитивная.

А оперирования моделями, которые описывал Пиаже, он не делает.

Как раз проблема в том, что это не наука, а какой-то другой подход.

Например, мы вычисляем, но не строим законов.

Разработчики искусственного интеллекта могли бы идти в эту сторону.

Но мне кажется, что пока искусственный интеллект этого не может.

Лекторский В.А.: Дело в том, что я был первым человеком в нашей стране, который стал писать о Пиаже.

Это было в начале 60-х годов.

И искусственный интеллект этого сейчас не делает.

Но ему и не нужно.

У Пиаже интеллектуальные операции развиваются естественно.

Есть разные стадии в развитии этих операций (в зависимости от возраста ребенка): одна, вторая, третья.

Это естественное развитие интеллекта по Пиаже – как у индивида, так и в целом у человечества.

А зачем искусственно это делать?

Можно, конечно, создать такую модель, чтобы это воспроизводить.

Но мы ничего нового при этом не узнаем.

Тосунян Г.А.: Коллеги!

На этой неделе должно было состояться празднование восьмидесятилетия Национальной академии наук Армении.

Но в силу напряженной ситуации в стране это празднование и все заседания отложены.

Геворк Арамович Погосян – академик НАН РА, научный руководитель Института философии, социологии и права НАН РА и иностранный член Российской академии наук, присутствует на нашем заседании.

От нашего Научно-консультативного совета в его лице хочу пока неофициально поздравить наших армянских коллег.

Академия была создана в 1943 году, в тяжелые для СССР времена.

Это какая сила мысли должна была быть и у союзных, и у местных руководителей, чтобы найти в себе и волю, и возможность создать такую академию?!

Лекторский В.А.: Вчера Российская академия образования тоже отмечала свое восьмидесятилетие, она тоже была создана в сентябре 1943 года.

Тосунян Г.А.: Арутюн Ишханович, Вам слово.

ДОКЛАД 3

АВЕТИСЯН А.И.
акад. РАН, д. ф.-м. н., директор Института системного программирования им. В.П. Иванникова РАН

ТЕХНОЛОГИИ ДОВЕРЕННОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ

Спасибо большое.

С удовольствием послушал доклады.

Владислав Александрович сказал, что в прошлом году мы организовывали эту конференцию, я тоже был участником, как раз отвечал за доверие, доверенность.

С давнего времени мы все живем в мире, как малые дети в песочнице.

Только вместо игрушек у нас ножи, затем гранаты, потом все более и более сложные технологические устройства.

А что со всем этим делать?

На той конференции я сказал, что если нам, специалистам в сфере IT, поставят задачу, то мы решим ее очень быстро, надежно.

Все будут зачипированы, безопасность будет стопроцентная.

Никто никому ничего не сможет сделать, все будем ходить по струнке, смирно.

Но, наверное, мы так жить не хотим, не важно, с искусственным интеллектом или без него.

И это очень важно понимать.

Отрасль IT развивается такими темпами!

В конце 80-х – начале 90-х годов, когда я был студентом, в одном из выпусков американского журнала «Компьютер» на обложке разместили вид нашей Земли из космоса.

И написали, что если бы космическая отрасль развивалась так же быстро, как IT, то вид из отеля выглядел бы именно так.

Уже тогда высокая скорость развития отрасли была очевидна.

Искусственный интеллект тогда и сейчас

В 1956 появился термин «искусственный интеллект». Прошло чуть больше 40 лет, и…
1997 – IBM Deep Blue выиграл в шахматы у Гарри Каспарова
2002 – первый робот-пылесос
2010 – база данных ImageNet, разметка данных обычными людьми. 14 млн изображений, 20 тысяч категорий
2011 – IBM Watson выиграл шоу Jeopardy! («Своя игра»)
2011 – персональный ассистент в смартфоне (Siri)
2016 – AlphaGO выиграла у профессионального игрока в го
2016 – Google Translate начинает использовать нейронный машинный перевод для 8 языков
2022 – выпущен ChatGPT от OpenAI. За 2 месяца число пользователей достигло 100 миллионов (это рекорд).

Рис. 3

Сегодня я хочу поговорить о том, что такое доверенный искусственный интеллект.

Мы очень часто живем страхами.

А наши страхи должны базироваться на науке: чего нам бояться?

Идешь в лес – там реально есть медведь?

Или, может, его там нет?

Может, это такое место, где все хорошо?

Может, это вообще не лес, а мы просто боимся темноты?

Нам надо знать, чего мы боимся.

Надо разделить человеческое доверие к технологиям и доверие к тому, как это произведено, как мы контролируем, что продукт действительно указанного качества.

А здесь очень часто происходит путаница.

Доверенность – это не про эмоции, а про то, что мы четко можем проверить.

С этим связано появление ГОСТов, необходимых при производстве различной продукции.

Это очень важно.

Расскажу кратко про историю искусственного интеллекта.

Сам термин возник в 50-е годы.

Серьезный прорыв произошел в 2010–2011 годах, когда миллионы и миллионы изображений были размечены и выложены.

Когда, будучи студентами, мы пытались кошку от собаки автоматически отличить на картинке, ни у кого в мире, и у нас в том числе, это хорошо не получалось.

А уже в 2010–2011 годах эта вероятность достигала почти 100%.

Это обеспечивалось за счет огромного объема данных и достигнутой уже тогда сверхвысокой производительности систем.

Использование графических ускорителей и все это вместе позволяло быстро обучать и получать решение.

Нужно понимать, что эти изображения могли размечать и неподготовленные люди.

Если же мы пойдем в медицину, то выяснится, что там даже не каждый врач может разметить некоторые данные.

Например, ЭКГ может разметить только специалист.

Мы занимаемся ЭКГ, и у трех врачей консенсус бывает очень редко. У нас это – проблема.

Я уже не говорю о том, что есть области, где просто нет данных.

Например, сход лавины.

Откуда вы возьмете такое количество данных схода лавины?

Тогда очень хорошо помогает математическое моделирование.

Моделируются тысячи, миллионы вариантов, сколько возможно.

Берутся реальные данные, система обучается на них, и выясняется, что таким образом тоже можно предсказывать.

Тем самым классическая математическая модель никуда не уходит.

Она становится просто подспорьем в этом обучении.

И в некоторых случаях это очень важно, например, в биологических системах.

У нас в советское время была очень развита индустрия (ее сейчас восстанавливают), когда использовали метан, бактерии, а на выходе получали биомассу.

Она может использоваться для кормления животных, и даже если не хватит вдруг еды, то и люди могут питаться.

Если вы там не уследили за процессом, то КПД выхода может оказаться несколько десятков процентов, а в правильных условиях КПД – 70%.

А это огромные деньги.

При математическом моделировании, например, можно рядом поставить суперустановку.

Можно будет моделировать, что там происходит, и в нужный момент управлять.

Возможно, что моделирование пройдет.

И через несколько часов вы получите ответ.

Уже после того, как у вас все погибло.

Один из подходов – снова заранее просчитать много-много вариантов, поставить датчики, собрать так называемые большие данные.

Но их никогда не будет много.

Потому что этих установок у вас сколько будет?

10–20, сколько бы экспериментов вы ни ставили.

Но за счет моделирования на стыке можно получить данные.

Я просто попробовал чуть-чуть описать задачу, когда данные искусственно генерируются.

В искусственном интеллекте есть даже целое направление – так называемый физически обусловленный искусственный интеллект, когда нельзя взять просто так любые данные.

Данные должны быть физически достоверны.

Последний рывок – это генеративные искусственные сети, которые в простонародье именуются «чат-GPT», а сейчас уже есть и другие.

Сбербанк, «Яндекс» тоже создают технологии в этой области.

Генеративный ИИ может более-менее разумно отвечать вам, разговаривать с вами.

Понятно, что без этого уже никуда, это будет внедрено везде.

Когда продуктивность труда возрастает на порядки, с этим вообще бороться невозможно.

Нас еще в детстве учили: Карл Маркс сказал, что за 300% прибыли капитал и родную мать продаст.

А тут у нас не 300%, а порядки.

Ясно, что это будут внедрять.

Вопрос – как внедрять, чтобы мы были в безопасности?

Если нам стакан отравленной воды даже один раз из миллиона случайным образом попадется, вряд ли мы такую воду будем пить, мы не захотим.

А здесь мы почему-то внедряем и думаем, что это можно легко использовать.

Какая здесь грань, что наука говорит?

Ясно, что глобальный рынок искусственного интеллекта – огромный.

Если сейчас его объем оценивается в 2 триллиона долларов, то через некоторое время эта оценка составит 20 триллионов.

Развитие цифровизации в целом влечет за собой развитие искусственного интеллекта.

Мы очень часто в рассуждениях исходим из того, что в каком-то смысле сильный искусственный интеллект существует.

Начинаем ему даже придавать субъектность.

Начинаем обсуждать, что с этим делать.

Лет 10 назад в Стэнфорде этому было посвящено совещание.

Обсуждалось, что в случае появления сильного искусственного интеллекта нужно менять законодательство.

Рост глобального рынка ИИ в 20 раз за 10 лет: от $100 млрд до почти $2 трлн[1]

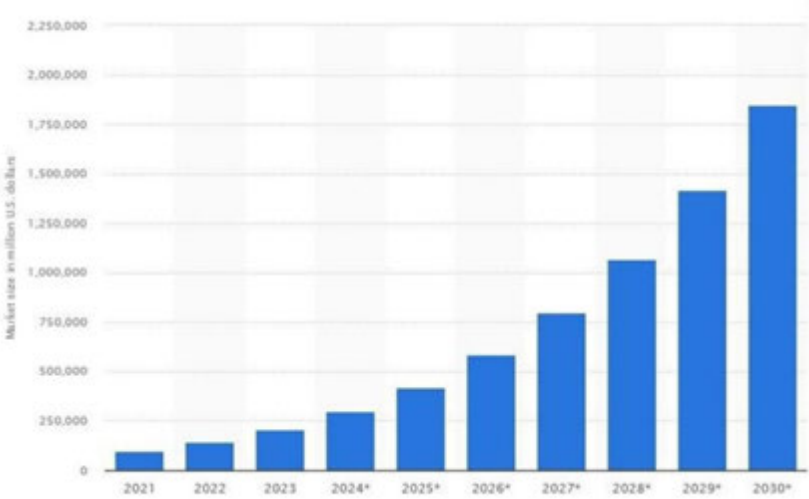

Рис. 4

Во время стратсессии под руководством Михаила Владимировича Мишустина, которая прошла буквально пару недель назад, в одном из докладов было сказано, что есть люди, которые считают, что общий искусственный интеллект появится уже через 2–3 года.

[1] www.bloomberg.com

Причем это люди, которые находятся на переднем крае исследований в этой области!

Некоторые считают, что он появится в ближайшее десятилетие.

А другие считают, что и через 20 лет не появится.

Я достаточно скептичен.

Считаю маловероятным, что сильный искусственный интеллект появится в ближайшее десятилетие.

Хотя генеративный искусственный интеллект будет очень похож на настоящий искусственный интеллект.

Здесь будет история, подобная восприятию человеком фокусов иллюзиониста.

Когда кажется, что он творит чудо, но на самом деле это чудом не является.

Да, искусственный интеллект будет решать какие-то задачи.

Но он не будет сильным искусственным интеллектом.

Нужно отдельно проводить исследования на эту тему, мозговые штурмы, писать статьи.

Чтобы мы были готовы, если он вдруг появится.

Потому что это действительно будет прорывом.

А реально мы находимся в мире слабого искусственного интеллекта, что бы там ни говорили.

В связи с этим необходимо понять, какие угрозы могут исходить от него, что на самом деле происходит.

Приведу несколько примеров.

Есть фотография, где, по-моему, даже я стою.

Смысл в том, что наши специалисты взяли компьютерное зрение, поставили камеры, научились генерировать некоторые изображения – не птиц, что очень важно.

Мы заказали майки, нам их напечатали.

Надеваешь эту майку, становишься перед той же камерой, и камера перестает тебя распознавать как человека.

Если ты стоишь перед ней без майки, она видит, что ты – человек.

С майкой она уже перестает тебя распознавать.

Китайцы продают за 30 долларов плащи-невидимки.

Надеваешь – и компьютерное зрение перестает тебя видеть как человека.

Уязвимость здесь – очевидная.

Если у вас ядерный объект, ясно, что туда пускать через компьютерное зрение пока не стоит.

Возможно, его могут как-то взломать.

С другой стороны, если это метро и надо заплатить за вход, то если кто-то пройдет бесплатно – это не страшно.

Даже если человек один раз не заплатит, потом его как-то отследят.

Мы сейчас говорили о том, что искусственный интеллект уже генерирует книги.

Специалисты посмотрели книгу про грибы, сгенерированную искусственным интеллектом.

Если по ней идти – умрете, потому что будете есть грибы, которые есть нельзя.

Кто за это будет отвечать?

Можно приводить множество примеров.

В частности, что касается компьютерного зрения на автомобилях.

Наши разработчики в этой области при обсуждении говорят: «Да они ездят лучше, чем водители. Уберем водителей, и вообще будет счастье».

А в 2021 году была издана книга «Искусственный разум и новая эра человечества».

Сам Киссинджер стал соавтором, чтобы придать статус этой книге.

Перевод на русском вышел только сейчас.

Лейтмотив всей книги в том, что безопасность – самое главное.

Это одна из самых сложных областей: необходимо, чтобы автопилоты и автомобили двигались автономно, потому что мир вокруг меняется, данные устаревают.

Иногда забывается, что машинная модель обучена на каких-то данных.

Есть понятие устаревания данных.

Вокруг вас погода изменилась, или просто изменился климат, случился тайфун, который вы не учли, наводнение.

И все, ваша программа сломалась.

То же касается и других областей.

Например, программы здоровья человека.

Вы обучались на данных XIX века, а на дворе уже XXI век.

Но реально изменения иногда происходят в течение нескольких недель, месяцев.

Несколько лет назад мы обучили модель, связанную с финансовыми рынками, для участия в соревновании.

В октябре должна была пройти апробация, а в сентябре произошел финансовый кризис.

И все модели, даже самые передовые, просто перестали работать, потому что произошла смена контекста.

Это всего лишь одна из проблем, с которыми имеем дело, – устаревание.

Ваша модель должна быть не просто готовой, а быть подобной ребенку.

Постоянно вы за ним должны следить, он должен постоянно развиваться.

Вдруг в те моменты, когда за вашим ребенком будет следить кто-то другой, он, может, что-то такое начнет совершать, что вам не понравится.

И этот персональный ассистент вместо того, чтобы вам подсказать, что нужно выпить красную таблетку, подскажет выпить синюю таблетку.

И вы вдруг не в той системе проснетесь.

Ничего нового тут нет.

Нужны стандарты, сообщества, процессы.

Нужно относиться к этому серьезно, как это сделано, например, в отношении производства продуктов питания, производства автомобилей.

Почему появились ремни безопасности?

Об этом забывают, но они же появились, потому что много людей погибало.

Только в конце 60-х забили тревогу, еще в начале 70-х они даже были необязательны.

Почему бамперы стали не железными, а пластиковыми?

Не потому, что это дешевле, уже позже они стали дешевле.

А потому что пластик при ударе о человека не убьет его сразу.

И государство, и общество постепенно это делают через регуляторику.

Конечно, это всегда чуть дороже.

Внедрить доверенный искусственный интеллект дороже, чем просто какой-нибудь, которого чему-то обучили и вот теперь можно делать бизнес.

Это вопрос даже не нашей общей безопасности.

Это вопрос вообще существования.

Искусственный интеллект уже управляет теплицами, управляет процессами в сельском хозяйстве.

Хорошо известно, когда целые стада ходят с чипами, пасутся, возвращаются обратно.

Если там что-то сломать внешне – эти стада начнут погибать.

Привожу лишь некоторые простые примеры, а их будет больше, и они будут сложнее.

Далее такие системы появятся на удаленных объектах электроэнергии.

Это сейчас уже делается.

У какой-нибудь небольшой атомной станции на Севере будет удаленный доступ.

И внутри будут те же самые управляющие системы.

А если случится перехват или модель не очень правильная внутри, что тогда?

Всеми этими рисками надо научиться управлять, и без науки это сделать невозможно.

В обычном программном обеспечении первые стандарты безопасной разработки появились более 20 лет назад.

В 1999 году в США были приняты стандарты Common Criteria.

В 2004-м появился конвейер разработки безопасного ПО от Microsoft – де-факто это общемировой стандарт безопасной разработки.

Без соответствия этим стандартам на рынок США выйти нельзя (крупнейший рынок).

Чтобы соответствовать этим требованиям, были разработаны необходимые инструменты анализа программ. Соответствующие работы начались и в нашем институте.

В 2002 году мы подписали контракт с крупнейшей канадской телеком-компанией Nortel Networks.

Их телефонные станции «Меридиан» стояли у нас везде.

Они их даже подарили нам, потому что мы нашли у них в телефонной станции несколько холодных перезагрузок.

Мы с ними подписали контакт, начали разрабатывать технологии.

Сейчас уже наши технологии в области разработки безопасного ПО внедрены в Samsung, Huawei, внутри страны – в «Лаборатории Касперского».

Но используют их только там, где есть жесткая регуляторика.

То есть далеко не везде.

На днях ассоциация «Финтех» пригласила меня выступить перед представителями Мосбиржи, нашими банкирами.

Как создается доверенное ПО в целом.
Пример SDL цикла

Рис. 5

Они реально нуждаются в этих технологиях, а у них стоят американские.

Сейчас обсуждаем, что они будут переходить на наши технологии.

«Финтех», который вообще не следит за регуляторикой в этом смысле, осознал, что без этого уже тяжело, надо этим заниматься.

Это к тому, что прошло 20 лет, и только сейчас реальный бизнес начинает все-таки тоже включаться.

А искусственный интеллект находится сейчас на уровне обычных технологий 1999 года, может быть, на уровне 2002 года.

Мы должны создать научную базу, чтобы разработать искусственный интеллект, которому можно доверять.

Потому что просто запретами ничего не добиться.

Когда продуктивность труда увеличивается на порядки, это невозможно остановить.

Иначе мы будем неконкурентоспособны.

Наши дети и мы должны быть конкурентоспособны.

Но одновременно мы должны быть и в безопасности.

Схема SDL и наши технологии

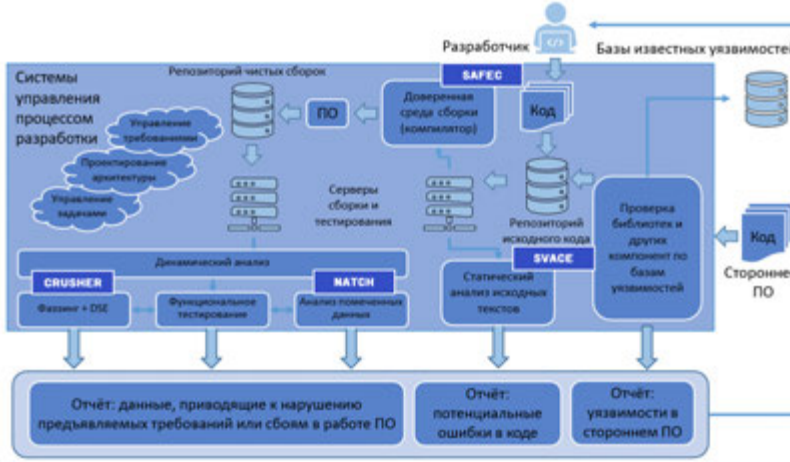

Рис. 6

Иногда нам пытаются подать конкурентоспособность как самое главное.

Однако нельзя стратегическую безопасность отдавать в обмен на конкурентоспособность.

Мы достаточно умны, чтобы и то и другое делать одновременно.

И хочется, чтобы наше сообщество тоже это понимало.

В мире этим серьезно озабочены.

Например, европейцы два года назад объявили, что искусственный интеллект будет разделен на две части.

Они считают, что одну часть вообще не надо проверять и сертифицировать, можно использовать как хочется.

Это не влияет на жизнедеятельность людей.

Например, для контроля на вход в метро.

А для второй части необходима обязательная сертификация.

Прошло два года, однако никакие стандарты так и не появились, поскольку пока нет научного обоснования.

В мае этого года мы вместе с Академией криптографии Российской Федерации проводили на территории МГУ конференцию «Доверенный искусственный интеллект».

В нашей стране это была первая такая большая конференция.

В ней приняли участие более 300 человек, в том числе специалисты ВТБ и нашего института.

Состоялся продуктивный разговор.

И буквально перед этой конференцией я вдруг узнал, что Еврокомиссия в июне собирается принять новое постановление, где искусственный интеллект будет разделен уже на три части.

Одна – для свободного использования, вторая подлежит сертификации, а третью предлагается запретить.

А что, например, следует запретить?

Европейское сообщество приняло решение запретить сбор биометрии в общественных местах, социальный скоринг.

На это мне человек из одной нашей уважаемой крупной корпорации сказал: «Мы это знаем. Но есть же статистика, что, как только они это объявили, еще даже

не приняли, уже 18% инновационных стартапов объявило, что уедет из Европы».

Регулирование: все только начинается

Whitepaper on AI: A European approach (Евросоюз, 2020)

- Объясняет важность ИИ и призывает к его оптимизации и развитию экосистемы
- Инициирует работу над нормативной базой ИИ и определяет ключевые требования: безопасные обучающие данные без дискриминации; надежность и воспроизводимость; контроль человека над ИИ; защита биометрических данных

Кодекс этики в сфере ИИ (Россия, 2021)

- Разработан при участии АЦ при Правительстве, Минэкономразвития России, а также около 500 экспертов академического и бизнес-сообщества
- Подчеркивает приоритет прав человека; ответственность человека за действия ИИ; потребность в безопасности и защищенности данных

AI Bill of Rights (США, 2022)

- Разработан компаниями, общественными организациями и экспертными группами
- Формулирует пять принципов создания и использования систем ИИ, в числе которых: разработка безопасных и эффективных систем; отсутствие алгоритмической дискриминации; обеспечение конфиденциальности данных и контроль пользователя за тем, как используются его данные, и др.

И другие документы:

Nist AI risk management framework (NIST: national institute of standards and technology, США)

Mitre atlas, adversarial threat landscape for artificial-intelligence systems

Рис. 7

И это правда.

Они уезжают прямо сейчас.

Уезжают в другие юрисдикции, где этого нет.

Я ответил, что может, даже не 20%, а 30% уедет.

Мы, как общество, должны для себя обсудить и понять, какую модель мы хотим разрабатывать.

Если государство и общество скажут, что безопасность прежде всего, и поставят мне задачу, то я ее решу.

Но как ее лучше решить?

Давайте, может, хотя бы обсудим этот вопрос!

Чтобы не получилось, что мы не случайно оказались в клетке, а добровольно туда пришли.

26 сентября 2023 года под руководством премьер-министра Михаила Мишустина в Москве прошла стратегическая сессия «Развитие искусственного интеллекта», на которой была принята доктрина, где есть раздел «Гибридная модель развития».

Где-то предусматривается жесткая регуляторика.

Где-то – абсолютная свобода.

Где-то – даже регуляторная гильотина.

Думаю, эта комбинация и даст нам возможность одновременно быть свободными в развитии науки, передовых решений и быть в безопасности.

Это тоже очень важно.

Возвращаясь к технологическим вопросам.

Не буду углубляться, что такое машинное обучение.

Даже оно постоянно эволюционирует: то глубокие нейронные сети, то генеративный ИИ.

Сейчас уже развивается следующее направление – как из больших моделей делать маленькие, чтобы они экономили энергию и при этом давали бы тот же эффект.

Но с точки зрения безопасности везде – вопрос доверия.

Если мы про обычное программное обеспечение говорим, там уже более-менее все отработано.

Есть технологии, они развиваются, понятно, куда двигаться.

А в искусственном интеллекте появляется датацентричность.

Что это такое?

У нас есть модель.

С чем мы сталкиваемся?

В случае того же компьютерного зрения мы имеем дело уже с конечной моделью.

Она создается за счет того, что большие данные, фреймворки обучаются на супероборудовании.

SDL в ИИ: что нужно?

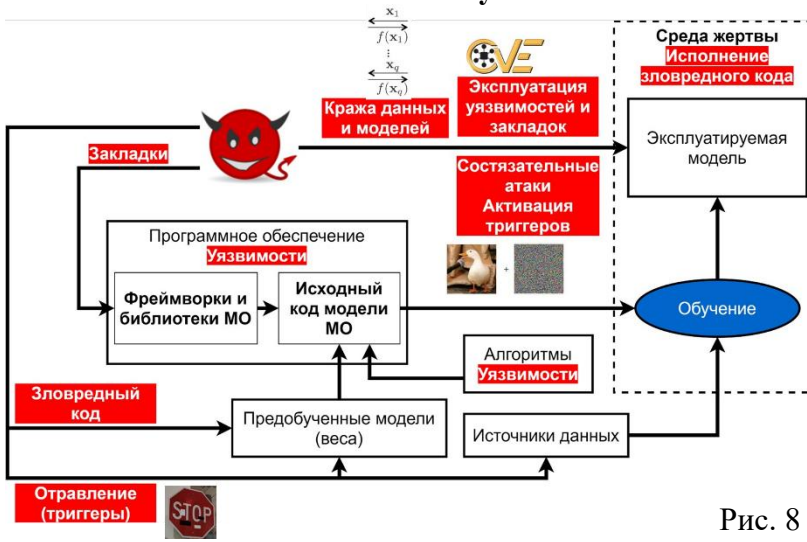

Рис. 8

Представим, что получили модель.

Если у вас неправильные данные, то они могут быть такими не только потому, что злоумышленник туда что-то внес.

Но и потому, что, возможно, вы собрали не те данные.

Были случаи, когда люди специально генерировали много данных, чтобы показать, что они очень крутые, учились на большом объеме.

Естественно, генерировали не очень хорошо.

Очень много и таких примеров, когда берутся данные из открытых источников, обучаются.

А там уже внутри могут быть отравленные данные.

Что значит «отравленные данные»?

Например, вы анализируете трафик, а там буквально один битик другой, и ваша модель его пропускает.

Фактически злоумышленник заходит внутрь к вам и делает что хочет.

Я специально упрощаю.

Но если кому-то будет интересно, мы детально, подробно по каждому из этих вопросов можем отдельную презентацию сделать.

Что касается датацентричности.

Никто этим раньше не занимался.

Горжусь тем, что пять лет назад я впервые написал документ, что этим необходимо срочно заниматься, что это – новый вызов.

Два года назад был конкурс в Минэкономразвития, в котором участвовали шесть центров.

Там были ведущие университеты, и наш институт стал одним из победителей.

Мы единственные, кто занялся темой доверенного ИИ.

Это настолько серьезная история, что даже самые передовые страны, и американцы, и европейцы, отстают.

У них тоже нет серебряной пули.

Они тоже не знают, что делать.

Например, у вас есть модель, она не ваша, вы купили ее у кого-то.

А можно ли ее дообучить, чтобы она была более доверенная, чтобы она не допускала, например, перехвата управления?

Очень много таких вопросов, на которые нет ответов.

Есть отдельные инженерные ответы.

Но общего, единого подхода нет.

Все знают, как в обычном ПО выстроить конвейер безопасной разработки.

На эту тему есть огромное количество конференций и публикаций.

В области ИИ такого нет.

Появляется множество проектов, мы все их знаем, смотрим.

Здесь очень важно, чтобы у тебя была грамотная команда.

Насколько я знаю, у Курчатова был целый отдел, который занимался фактчекингом.

Когда ему приносили что-то, то проверяли, насколько это вообще достоверная информация.

Потому что она могла быть и специально предоставленной, чтобы сбить с пути.

Если у тебя нет команды, которая хорошо разбирается в открытых проектах, то для вас это все разрозненно.

А если есть понимающая команда, вы можете эти проекты верифицировать и понять, что из этого имеет перспективу.

Эти проекты будут потом по дороге поглощаться, превращаться в какие-то большие проекты, что и происходило с обычным софтом.

Кстати, и в случае с обычным ПО, без искусственного интеллекта, ситуация непростая.

Несмотря на то что прошло 20 лет, открытых решений, например, по статическому анализу кода, не существует.

А почти все закрытые решения находятся в Соединенных Штатах Америки.

Если вы не имеете своей технологии, то сделать ее не можете.

В области ИИ может то же самое произойти.

Месяц назад американское правительство собрало свою крупную бизнес-пятерку: Meta, Google и другие известные компании, – и они подписали документ, пока не обязательный к исполнению.

В соответствии с этим документом любая «генеренка», тот самый генеративный искусственный интеллект, любой объект, какого бы вида он ни был, должен иметь внутри водяные знаки, которые однозначно показывают, что это «генеренка», – для того, чтобы потребитель мог это узнать.

Казалось бы, это же хорошо для всех нас.

Но люди забывают про обратную сторону медали.

Все остальные, кто пытается эти технологии делать, от рынка мгновенно полностью отрезаются.

Вы же тоже теперь должны предоставлять цифровые водяные знаки (watermarks).

А если нет, то вы, по сути, запрещены на территории.

Законодательство у них достаточно глобальное.

То есть если у вас нет соответствующих технологий, вы вынуждены будете пользоваться их технологиями.

Значит, они будут перехватывать ваши знания, иметь доступ к вашим данным, если нужно будет.

Чтобы все эти проблемы решить, мы инициировали создание нашего Исследовательского центра доверенного искусственного интеллекта.

Мы вовлекаем туда ребят из других вузов.

Из Московского физико-технического института с нами работает талантливый ученый Александр Гасников.

Математический институт им. В.А. Стеклова Российской академии наук тоже сотрудничает с нами.

Мы создали уже ряд технологий и постепенно систематизируем их.

Очень радует, что у нас сейчас государство настолько открыто и консервативно к этому относится, даже удивительно.

Обычно как происходит?

Давайте быстро напишем какую-нибудь регуляторику, всех зарегулируем, после этого будем смотреть.

На конференции «Доверенный искусственный интеллект», которая прошла в мае, Виталий Сергеевич Лютиков, заместитель руководителя Федеральной службы

по техническому и экспортному контролю (ФСТЭК), публично сказал, что в ближайшие 2–3 года мы не ожидаем создания регуляторики, которая будет нести какой-то запретительный характер.

Максимум рекомендательно скажем.

Почему?

Потому что сейчас нет той науки, которая может это реально сделать.

Но через 2–3 года она обязательно появится.

И мы будем ее создавать.

Думаю, что в финансовой сфере без искусственного интеллекта вообще не обойтись.

Это тоже нужно делать.

Следует учитывать, что все открытые доверенные системы находятся на внешних серверах.

И мы не знаем, что туда могут положить.

Несколько лет назад мы создали еще один уровень доверия, чтобы понимать, с чем имеем дело.

Мы предложили и создали репозиторий доверенных решений у нас в стране.

Вокруг которого как раз развернуты все технологии анализа программ.

В соответствии с этим выявляем уязвимости, абсолютно открыто передаем эту информацию международному сообществу.

По ядру Linux мы передали 220 исправлений, они приняли.

В TensorFlow и PyTorch – это фреймворки искусственного интеллекта – найдено более 50 ошибок за последние полгода.

Несколько из них – критические.

Они приводили к тому, что если ты их знал, то мог атаковать готовую модель.

А это ровно про то, что мы сегодня обсуждаем.

Мы передали, и они зафиксированы.

Можно зайти на американский сайт, увидеть, что это найдено инструментами Института системного программирования.

Модель долгосрочного устойчивого развития

Рис. 9

Тем самым, с одной стороны, мы показываем, что наша наука открыта миру.

С другой стороны, показываем технологический уровень.

Потому что это мало кто может сделать, они это прекрасно понимают.

И с третьей стороны, все эти инструменты находятся на наших серверах в защищенном режиме.

При этом они открыты и доступны любому студенту из любой страны мира.

Кому угодно.

Их можно использовать как в коммерческом, так и в некоммерческом варианте.

С описанием такой модели развития я лично выступал в рамках Форума БРИКС.

Сейчас у нас на высоком уровне формируется документ, в котором прямо заявлено, что мы должны создать репозиторий проверенных решений в области искусственного интеллекта.

Он должен быть открытым международному сообществу с целью повышения видимости нашей науки.

И, что более важно, для вовлечения как можно большего числа специалистов в совместные разработки.

Искусственный интеллект, как и обычное программное обеспечение, уже давно – мегасайенс.

Размеры его такие большие, что просто отдельными инструментами доверия, даже сфокусировав в одном месте всю энергию, вовлекая туда людей, это нельзя будет сделать.

В завершение хочу сказать, что 4 декабря в этом году исполняется 75 лет IT-технологиям в нашей стране, в Советском Союзе, в России, а значит, и в Армении, и в любой стране бывшего Советского Союза.

В каждой аудитории уже несколько лет я спрашиваю, знает ли кто-нибудь об этом.

Никто руку обычно не поднимает.

А ведь, как сказал один человек, мы великие, потому что стоим на плечах гигантов.

Это люди, которые создавали вычислительную технику.

Академик Сергей Алексеевич Лебедев был главным конструктором машины БЭСМ- 6, первой советской суперЭВМ на элементной базе второго поколения – транзисторах.

Он был учителем моего учителя, академика Виктора Петровича Иванникова, одного из основных участников создания БЭСМ-6.

В Музее науки в Лондоне есть статья англичанина, профессора, который написал, если коротко, что миф холодной войны о том, что русские отставали, был неверен.

После 1968 года мы начали отставать.

Но в 1968-м наш суперкомпьютер был того же уровня: CDC 6600 и БЭСМ-6 стоят рядом в музее.

В этом году 4 декабря наша открытая ежегодная конференция будет посвящена 75-летию IT.

Всех приглашаю.

Мы хотим поднять ее статус, чтобы туда пришло как можно больше наших друзей.

Скорее всего, мы сделаем выставку вместе с «Лабораторией Касперского» и другими партнерами.

Целый раздел конференции и целый день будут посвящены в том числе искусственному интеллекту.

Мы занимаемся доверием, но у нас уже есть технологии искусственного интеллекта, которые просто внедряются.

Они связаны, например, с анализом социальных сетей.

Есть и много других технологий, которые мы лицензионно продаем.

Наш институт – это полный жизненный цикл от идеи до внедрения.

Но это уже отдельная история.

А то, что я хотел эмоционально сказать про искусственный интеллект, я сказал.

Спасибо большое.

Готов ответить на все вопросы.

Тосунян Г.А.: Спасибо.

Слово Георгию Сатарову.

Потом Михаил Федорович.

САТАРОВ Г.А.
к. т. н., президент Фонда прикладных политических исследований «ИНДЕМ»

Сатаров Г.А.: Большое спасибо за то, что подняли эту тему, за эти интересные доклады.

Меня особенно вдохновляет последний доклад, потому что там речь идет о доверенном искусственном интеллекте.

А у меня большие сомнения.

Ровно то же можно сказать и про естественный интеллект – что есть институт по доверенному естественному интеллекту и что это когда-нибудь появится.

У меня вопрос очень короткий.

Вопрос такой иерархический.

Первое.

Мы все знаем, что невозможно доказать теорему, которая еще не сформулирована.

Есть ли результаты в сфере искусственного интеллекта по формулировке новых теорем?

Требуется только сказать «да», «нет».

Да – значит научились формулировать новые нетривиальные теоремы.

Второе.

Удалось ли искусственному интеллекту сформулировать такую теорему, которую пока не смогли доказать другие искусственные интеллекты, как это часто происходит с естественными интеллектами?

Если такие теоремы есть, сформулированы и доказаны, то можно ли через чат или через организаторов соревнования получить ссылки на публикации?

Тосунян Г.А.: Спасибо.

Кто готов ответить?

Аветисян А.И.: К сожалению, я не все знаю.

Поэтому могу ответить, что мне это неизвестно.

Но я еще поспрашиваю в профессиональном сообществе.

Если вдруг обнаружится, я Вам пришлю.

Тосунян Г.А.: Следующий у нас Виктор Дмитриевич.

Кривов В.Д.: Спасибо большое за такие замечательные, обстоятельные доклады.

Все время казалось, что искусственный интеллект где-то далеко.

А сейчас мы видим, что он уже здесь.

Хочу повторить свой вопрос, на который уже частично ответил уважаемый Владислав Александрович.

Надо ли запретить использование нейросетей в образовании при подготовке различных текстов: курсовых, дипломных, может быть, и научных публикаций?

Возможно ли это сделать?

В первую очередь это вопрос к уважаемому Арутюну Ишхановичу.

Аветисян А.И.: Во-первых, если серьезно отвечать, то это как отмычка и замок.

Если даже вы что-то научитесь делать, то это быстро обойдут.

На это можно тратить усилия, это как антиплагиат.

Мы немножко этим занимаемся, и наш партнер, один из лучших в России, в эту сторону тоже пытается двигаться.

Но если взять американский опыт, не случайно же они требуют, чтобы обязательно были водяные знаки, что это «генеренка».

Люди понимают, что технологически это уже по-другому, скорее всего, сложно будет сделать.

И берут на себя обязательства, пока нравственные, морально-этические, а потом просто будут законодательные.

Если у вас будет «генеренка», которая не обладает водяными знаками, то это будет уголовно наказуемая история.

Но я хотел ответить на первую половину вопроса: а нужно ли запрещать?

У нас уже есть один из ведущих вузов, который официально разрешил использовать.

Потом еще такой момент.

Даже без ИИ можно заказать диплом, и тебе напишут.

Важно, как это контролировать.

Я с трудом себе представляю, чтобы на моем семинаре мог пройти человек с генеренной дипломной работой.

А если он пройдет, значит, молодец.

Какая тогда разница, кто его написал.

Может, продуктивность его труда выросла, он смог написать 10 дипломных работ.

Здесь нужно подкручивать и саму систему образования, а не только действовать запретительно.

При этом я абсолютно согласен с Владиславом Александровичем, что сказать: «Пусть все теперь так делают», – не подготовившись к этому нормально, будет очень плохо.

Некоторые навыки будут теряться.

Это же еще зависит от того, кого ты готовишь.

Если ты готовишь писателя, то желательно, чтобы он сам много писал.

Но если ты готовишь технаря, который уже грамотно сдал экзамен, хорошо знает русский язык, а GPT ему помогает в каких-то вопросах, то почему нет?

Это повышает его производительность, повышается продуктивность труда.

Тосунян Г.А.: Член-корреспондент Российской академии наук, руководитель Центра социологии РАН Михаил Федорович Черныш, пожалуйста.

ЧЕРНЫШ М.Ф.
чл.-корр., д. социол. н., проф., директор Федерального исследовательского социологического центра РАН

Черныш М.Ф.: Уважаемые коллеги, хочу поблагодарить – прекрасные доклады.

Очень интересные, побуждают к размышлению на данную тему.

Хочу внести нотку скептицизма по отношению к тому явлению, которое мы называем искусственным интеллектом.

В 1770 году немецкий конструктор Вольфганг Кемпелен предложил публике лицезреть механического турка, игрока в шахматы.

Турок двигал руками и головой, переставлял шахматные фигуры.

Можно было открыть дверцы и убедиться в том, что турок – это сложный механизм.

Внутри вращались шестеренки и двигались рычаги.

Турок обыгрывал живых людей.

Он обыграл Бенджамина Франклина.

В поединке с ним потерпел поражение Бонапарт.

Но вот вопрос: а мог ли механический турок играть в шахматы?

Дело в том, что внешность турка была обманчива.

Внутри находился скрытый от глаз наблюдателя человек, сильный игрок в шахматы.

Достаточно сильный, чтобы обыгрывать любителей.

Мне кажется, искусственный интеллект – это нечто подобное механическому турку.

Нет, он не паразит, он просто использует данные и воспроизводит мотивации, свойственные человеку.

Он собирает данные, свойственные не отдельным людям, а обществу в целом.

Необходимо общество, чтобы определить для себя наиболее важные направления деятельности, мысли, ориентации и ценности.

Искусственный интеллект, подобно механическому турку, только воспроизводит действия и мотивы людей.

Известный философ Джон Сёрл предложил мыслительный эксперимент под названием «Китайская комната».

Человек, находящийся в комнате, не знает китайского языка.

Но у него есть алгоритм подбора ответов на вопросы, которые ему задают.

Человек, находящийся снаружи, подсовывает ему под дверь эти вопросы.

Человек внутри комнаты подбирает ответы в соответствии с имеющимся алгоритмом.

Тому, кто находится за дверью, кажется, что собеседник внутри комнаты знает китайский язык.

А он ни слова не знает по-китайски и ничего не понимает, не ведает того, что он делает, на какие вопросы отвечает.

Мне кажется, в этом проблема искусственного интеллекта.

Он не может, не в состоянии и, возможно, никогда не сможет обрести подлинные смыслы, ценности, подлинные интенции, человеческую интенциональность.

Он всегда будет находиться в той области, которую определили наши коллеги – Арутюн Ишханович, Дмитрий Викторович.

Они прекрасно показали, что это еще один прикладной инструмент, с помощью которого люди могут существенно улучшить свою жизнь, усовершенствовать технологические процессы, но не более того.

Спасибо.

Тосунян Г.А.: Спасибо.

Это тот самый очередной случай, когда у меня с Михаилом Федоровичем очень близко совпадают позиции.

Я готов подписаться под каждым словом.

Аветисян А.И.: Если мы говорим об играх в шахматы, го – все-таки искусственный интеллект устроен не совсем так, что он взял весь опыт человечества.

В го он научился играть, сам с собой играя, когда две машины играли друг с другом.

Это очень важный момент.

Когда есть игра с определенными правилами, там можно выстраивать такие истории.

Другое дело, что в нашей нормальной жизни мы не играем. У нас нет игр в этом смысле.

Мир вокруг нас сложнее, ничего нельзя предсказать.

В игре четкие правила, конкретное количество вариантов, поэтому он делает именно такие ходы.

Он может пожертвовать ситуацией, когда никакой гроссмейстер в мире этого не сделает.

Это тоже нужно понимать.

Мир сложнее.

Эта область деятельности очень быстро развивается.

Мы сейчас разговариваем, а завтра может появиться еще что-нибудь.

Черныш М.Ф.: Арутюн Ишханович, а он может придумать какие-нибудь новые шахматы, новую игру?

Или он играет в ту игру, которая уже придумана людьми?

Аветисян А.И.: Пока играет в ту игру, которую придумали.

Если он начнет придумывать системно, тогда это уже, скорее всего, будет сильный искусственный интеллект.

Случайно может придумать.

Когда он делает картину, похожую на что-то, он же не понимает, что он делает.

Он делает по аналогии, например, картину а-ля Пикассо.

Мы ему дали любую фотографию, он ее превратил во что-то.

Было такое «искусство», когда животным давали кисти и они что-то рисовали.

Потом это даже дорого продавалось, преподносилось как некоторое искусство.

Человек, когда что-то рисует, все-таки вкладывает смысл какой-то, у него есть идея.

Не всегда это получается, правда, но это другой вопрос.

ЛЕКТОРСКИЙ В.А.
акад. РАН

Михаил Федорович сослался на известную статью Сёрла «Китайская комната», она обсуждается до сих пор.

Автору казалось, что он доказал в статье, что невозможен сильный искусственный интеллект.

Он считал, что слабый искусственный интеллект возможен, а сильный – нет.

Но это не такая простая вещь.

Действительно, в том примере, который он привел, человек, сидящий в этой китайской комнате, не понимает китайского языка.

Но, зная алгоритм, он может правильно ответить на вопросы, задаваемые ему на китайском языке.

Для внешнего наблюдателя непонятно, как он может отвечать.

Вроде бы знает китайский язык, а на самом деле он ничего не понимает.

Весь вопрос в том, что такое понимание.

У Сёрла есть оппоненты.

Если это чистые алгоритмы, это – не понимание.

Но если алгоритмы плюс семантика, если машина может привязывать к чему-то свои действия, истолковывать их, интерпретировать, это будет уже что-то иное.

Я согласен с тем, что сказали и Вы, и автор статьи.

Говорят, что Сёрл доказал что-то.

Нет, он предположил.

По-моему, довольно убедительно.

Но и контраргументы тоже есть.

Тосунян Г.А.: Пожалуйста, коллега Супрун Анатолий Петрович из МГУ, прошу Вас.

СУПРУН А.П.
к. п. н., старший научный сотрудник лаборатории «Психология общения» факультета психологии МГУ им. М.В. Ломоносова

Доклад очень понравился.

Мне хотелось немножко развернуть некоторые тезисы Владислава Александровича – кажется, что здесь много недопониманий связано с тем, что некоторые тезисы не совсем развернуты.

Совершенно верно, что современный искусственный интеллект строится на нейронных сетях, а это – статистика.

Все строится на корреляциях.

Но корреляционные связи не раскрывают ее причины.

Скажем, есть корреляция между количеством айсбергов и перелетных птиц, но птицы не являются причиной числа этих айсбергов, как и айсберги – птиц.

Причина лежит вне этой корреляционной связи и является сезонной.

Поэтому искусственный интеллект может выдавать очень парадоксальные связи и недостоверные заключения.

Такие же связи может выдавать и шизофреник: скажем, находить связь между ботинком и карандашом, поскольку оба оставляют след.

Однако хотя это и кажется новым, но творчеством не является.

Творчество всегда связано с креативом, с нарушением причинности и связано с бессознательным, то есть за границей каузального сознания.

И здесь классические компьютерные технологии, конечно, не сработают.

Творчество всегда связано со свободой воли, а это вещь неопределяемая в рамках формальной теории, хотя сами математики ее используют.

Например, если говорится «возьмем произвольный треугольник», то это не означает, что мы берем *случайный* треугольник (или элемент несчетного множества), так как тогда вы ничего не докажете.

Здесь нужен именно *произвольный* элемент, от слова «воля».

А как формализовать проблему *произвольного* выбора, неизвестно до сих пор.

Бертран Рассел так отозвался об аксиоме выбора в математике: «Сначала она кажется очевидной; но чем больше вдумываешься, тем более странными кажутся выводы из этой аксиомы; под конец же вообще перестаешь понимать, что же она означает».

Например, как взять из континуума произвольную точку?

Поэтому искусственный интеллект может только подражать, имитировать когнитивную деятельность, но это никогда не будет творчеством.

В этом контексте искусственный интеллект на основе логических алгоритмов – это то, чему можно доверять, но нужно проверять, поскольку всегда может быть сбой или баг в программе.

Но здесь, во всяком случае, можно выяснить причину ошибки.

А что касается нейронных сетей, то здесь уже нужно все проверять.

Мы не знаем, как нейросеть приходит к тому или иному выводу, как влияет на нее тот или иной массив обучающих данных и пр.

Естественно, что это хороший подсобный инструмент, но доверять ему ни в коем случае нельзя, потому что здесь нет и не может быть до конца проведен контроль.

Это чисто статистическая вещь.

Тосунян Г.А.: Спасибо.

Пожалуйста, коллега Смолин Владимир Сергеевич из Института прикладной математики.

СМОЛИН В.С.
научный сотрудник Института прикладной математики
им. М.В. Келдыша РАН

Смолин В.С.: Было очень приятно послушать уважаемых коллег.

Однако хотелось бы обратить внимание, что они, конечно же, основываются на своем большом жизненном опыте.

Но мир меняется быстрее, чем меняются наши субъективные представления о нем.

И хотелось бы провести такую аналогию.

Пифагор, Сократ были выдающимися людьми в Древней Греции.

И в свое время они были очень сильно настроены против внедрения книг.

Ни Пифагор, ни Сократ никаких рукописей не оставили.

Книги же не имеют ни души, ни сознания.

И то, что вы в книге даете, это так же, как чат-GPT сейчас вам напишет.

Так и тогда их ученики могли в книге все прочитать.

И это сухое книжное изложение, с их точки зрения, было ужасно.

И все, что здесь рассказывается, воспроизводит отношение великих Пифагора и Сократа к книгам, которые, с их точки зрения, развращали молодежь.

Потом уже исследователи сожалели, что ни Пифагор, ни Сократ сами не оставили своих трудов, только их последователи писали про их учения.

Сейчас мы находимся на очередном витке.

И многие утверждения, которые сейчас сделали уважаемые докладчики, сводятся к одному: я не знаю, как это сделать, значит, этого вообще нельзя сделать.

Есть правильно выделенная Беллманом проблема сложности, проклятие размерности.

Человек всегда выделяет какие-то простые задачи и исследует их.

Как это делается в нейросетях?

Сейчас есть разные мнения.

Но есть такое подозрение, что внутри этих многослойных структур с некоторой вероятностью тоже осуществляется выделение простых задач.

Как Сёрл выделил пример про китайскую комнату, так же и Сократ думал про своих учеников, что если человек читает книгу, у тех, кто его слушает, может создаться впечатление, что он понимает то, про что читает.

А он просто умеет читать и читает книгу.

И это представление у нас воспроизводится.

Последние десять лет идет бурное развитие нейросетей.

Я немножко больше занимаюсь нейросетями, несколько десятков лет.

Конечно, развиваются те сети, которые имеют в первую очередь коммерческое внедрение.

Когда заглядывают в мозг человека, там тоже – нейронные сети.

Но они работают на других принципах.

Биолог вам скажет, что никакого обратного распространения ошибки там нет, и это правда.

Это не значит, что не может быть никаких других принципов, что невозможна декомпозиция сложных сигналов на статистически достоверно изучаемые компоненты.

И все задачи, о которых говорили, что их невозможно решать, о них неизвестно ничего, кроме того, что уважаемые люди говорили, что они не знают, как решать такие задачи.

Академики считают: раз они не знают, то и другие не могут узнать.

Но не всегда это звучит убедительно.

Вот на что хотелось бы обратить внимание.

Спасибо.

Тосунян Г.А.: Спасибо.

Мой не искусственный интеллект озадачился …

Судя по месту Вашего нахождения, могу ли я угадать, что Вы там делаете?

Смею предположить, что супруга попросила Вас привезти ее куда-то, Вы ей отказать не могли.

Привезти привезли, а сказали, что сами при этом будете заняты, потому что у Вас рабочий завтрак…

Она – в ожидании, когда Вы закончите, а Вы – в ожидании, когда она подойдет к Вам.

Я угадал, или я ошибся?

Смолин В.С.: Нет, Вы не ошиблись.

Тосунян Г.А.: Видите, а искусственный интеллект точно не вычислил бы.

Смолин В.С.: Я хотел сказать, что Вы считаете, что он не вычислил бы, потому что Вы не представляете как.

Но это не очень убедительно.

Ваше утверждение, что искусственный интеллект это никогда не вычислил бы, основано на Вашей уверенности, что Вы хорошо знаете жизнь.

Но боюсь, что нейронные сети Вы не так хорошо знаете, как остальные аспекты жизни.

Тосунян Г.А.: Нет, мое согласие с позицией Михаила Федоровича, что наш интеллект выше искусственного, зиждется на том, что искусственный интеллект не умеет переживать.

А без способности переживать нет объема, гармонии и развития такого масштаба, какой есть только у человека.

Я, если говорить просто, на этом строю свои оптимистичные соображения.

Хотя, может быть, и ошибаюсь.

Смолин В.С.: Я переформулировал бы это Ваше утверждение.

Вы не знаете, как на нейросетях переживания построить.

Я знаю, поэтому у меня – другое мнение.

Аветисян А.И.: Если бы это был не Институт прикладной математики, я бы промолчал.

Просто ИПМ – это очень хороший институт, это наши друзья.

Здесь, я думаю, у нас просто какое-то неправильное понимание.

Вся моя речь была посвящена тому, чтобы внедрять, внедрять и еще раз внедрять.

Но внедрять так, чтобы мы были в безопасности.

Мне кажется, мы друг друга в этом смысле поддерживали, поэтому – никакого сравнения с Пифагором.

Тут мне даже стало как-то обидно.

Смолин В.С.: К последнему докладу это в наименьшей степени относится.

Хотя тоже были некоторые утверждения, основанные на личном опыте, что естественно.

Аветисян А.И.: Мой учитель, академик Иванников Виктор Петрович, говорил, что самое объективное – это субъективное.

Так что ничего не поделаешь, школа.

Тосунян Г.А.: Хорошо. Профессор Сенаторов.

Пожалуйста, Михаил Юрьевич.

СЕНАТОРОВ М.Ю.
д. т. н., проф., член Экспертного совета по научно-технологическому развитию и интеллектуальной собственности при Комитете Государственной думы РФ по образованию и науке

Сенаторов М.Ю.: Добрый день.

У меня вопрос, связанный как раз с безопасностью искусственного интеллекта.

Если мы возьмем финансовую сферу и количество преступлений в ней, то они развивались последовательно с развитием информационных технологий.

Когда были простые бумажные деньги, обман был только за счет того, что делали и подбрасывали какие-то фальшивые деньги.

Это был очень небольшой объем нарушений.

Потом появились электронные деньги, электронные счета.

Стали осуществлять хищения с электронных счетов.

Потом стали массово использовать банковские карты.

Возникли массовые криминальные кейсы, которые давали возможность зарабатывать еще больше денег.

Сейчас криминальная индустрия в области банковско-финансовой деятельности зарабатывает ежегодно десятки миллиардов долларов.

Надо понимать, что эти десятки миллиардов долларов не проедаются, на них не покупают машины и виллы, а их вкладывают в технологии.

Возникло сращивание криминального IT с физическим криминалом, чего раньше не было.

Интересно, искусственный интеллект еще больше разовьет эту криминальную сферу?

Чего нам опасаться в этой области?

Аветисян А.И.: Уже несколько лет с Росфинмониторингом проводится ежегодная Международная олимпиада по финансовой безопасности

Идея очень простая.

Во-первых, находить талантливую молодежь.

Во-вторых, нужно максимально информировать людей об этих угрозах.

Только технологически от этих угроз избавиться нельзя.

Современный искусственный интеллект можно использовать как инструмент социальной инженерии, с целью обмана.

Не просто вам мошенник будет звонить, а все будет гораздо сложнее.

Хороший пример был с Красниковым.

Я получал и от него, и от других людей сообщения: срочно зайди туда, срочно позвони сюда.

Причем все фамилии совпадают.

Это уже хороший пример.

На самом деле обслуживают все эти схемы абсолютно безграмотные люди.

Люди думают, что кто-то звонит и гипнотизирует.

Но это все уже делается на уровне определенных моделей, эти модели кто-то отрабатывает.

А дальше искусственный интеллект выступает как инструмент социальной инженерии.

Другое – это схема отмывки.

Противоборствовать этому можно как раз за счет искусственного интеллекта.

Например, на больших объемах легко видны аномалии.

Детекция аномалий и подключение после этого специалиста поможет выявлять новые схемы.

Обычно по схемам же все делается, потому что иначе невыгодно.

Одноразовая акция – это невыгодно.

Необходимо выявлять эти схемы и быстро их купировать.

Уже много лет в стране обсуждается предложение сделать единый центр с телефоном, например, 004.

Чтобы человек, когда ему поступает подозрительный звонок, мог сообщить об этом в единый центр, чтобы быстро подключилась специальная служба.

Для реализации этой идеи просто нужна политическая воля, правильное централизованное финансирование, а не разобщенные попытки бороться с мошенническими схемами.

Необходимо это осуществлять через искусственный интеллект.

Там будет такой объем данных, что иначе этого не сделаешь.

Они растут, и мы тоже должны расти.

Остановить их нельзя будет, потому что мир открыт.

Все эти модели лежат в открытом коде.

Все данные у мошенников есть, потому что базы покупаются.

Недавно, буквально несколько недель назад, был слив.

Утекают не только наши данные, и зарубежные тоже.

И государство, и общество также должны на эти вызовы отвечать.

Поэтому «Финтех» и озаботился.

Нас позвали, предложили как-то системно заняться решением этих проблем.

Потому что вроде бы стоит бронированная дверь, а почему-то данные утекают.

Я говорю: так у вас забора нет.

Стоит бронированная дверь, которую пушкой не пробьешь, а забора рядом нет.

Вот на таком уровне мы сейчас живем.

Искусственный интеллект в этом смысле бреши пробивает легко.

И социальная инженерия – это хороший пример.

На самом деле он пробивает и безопасность, в том числе во всех этих структурах.

Но на это можно отвечать, потому что государство и мы – вместе, бизнес-сообщество, оно мощнее.

У нас наука в руках, много чего есть.

Вопрос, к сожалению, не только технологический.

И в любом случае нет серебряной пули, с которой мы были бы в безопасности.

КРУШАНОВ А.А.
д. филос. н., проф., ведущий научный сотрудник
Института философии РАН

Уважаемые коллеги!

Пользуясь возможностью, прежде всего хотел бы поблагодарить организатора встречи академика Тосуняна Гарегина Ашотовича за саму дискуссию и за предметные, актуализирующие и направляющие встречу вступительные слова.

И конечно, я благодарю академика Лекторского Владислава Александровича за развернутое видение обсуждаемого сегодня предметного поля!

Анализ темы сегодняшней дискуссии Владиславом Александровичем основателен и по-хорошему заражает желанием со всем этим все же разобраться!

В этой связи хотел бы высказать несколько своих соображений.

Разработка и обсуждение проблематики ИИ неявно предполагают, что разбираться следует, собственно, с искусственным интеллектом, так как с исследованием мыслительной деятельности человека в целом наука уже неплохо разобралась.

Успехи в этом и в самом деле впечатляющи, но и оставшиеся проблемы нельзя недооценивать.

Например, для анализа умственной деятельности человека до сих пор не создана согласованная система основных понятий.

Как соотносятся такие важные категории, как мышление, сознание, интеллект, разум, самосознание?

Это совсем не закрытый вопрос.

Как правило, такого рода понятия обсуждаются и вводятся по отдельности, не создавая когерентной картины в целом.

Выше я упомянул «разум», но имел в виду совсем не известное кантовское понимание, подразумевающее уровень мыслительной деятельности, на котором возможен выход мысли за пределы имеющегося ограниченного опыта.

В рамках обсуждения проблематики искусственного интеллекта не стоит забывать, что Тьюринг в статьях, вводящих его тест, использовал в том числе слово “mind”, которое обычно переводят как «разум», в значении, близком русскому слову «ум».

Да и сам тест выстроен таким образом, что это испытание не столько на интеллект, сколько на разумность (на умственные способности в целом).

Это хорошо видно при условии, если в психической деятельности различаются две обычно не выделяемые стороны: информационная и оценочная.

Мышление и сознание связаны с целенаправленной выработкой и преобразованием необходимой в данной ситуации информации.

Но параллельно происходит и оценивание этой информации на ее значимость для системы в целом.

Так вот, для передачи искусственному интеллекту информационных процессов и задач психики (мышления и сознания), на мой взгляд, видимых ограничений не просматривается.

Но вот с оценочной стороной психики, кажется, дело много сложнее.

Да, уже сегодня в рамках развития ИИ есть работы по моделированию эмоций человека.

Это интересная и значимая работа.

Но в этом случае важно понимать, что таким образом ИИ обретает информационный образ эмоций, но не сами эмоции.

Потому на этом пути ИИ остается бесчеловечным и бездушным.

И, быть может, это для нас совсем неплохая новость.

Между прочим, введенное различение сторон психической деятельности позволяет внятным образом развести два существующих теста «на интеллект» – тест Алана Тьюринга и IQ-тест Ганса Айзенка.

Если Тьюринг предложил диалог, в котором естественным образом могли возникать и информационные (то есть описательные) высказывания, и оценочные (то и другое вполне характерно для «разума»), то Айзенк предложил тестирование, с помощью которого изучается решение лишь информационных по характеру задач.

Оценочных задач в рамках этого теста предложено не было.

Не означает ли это, что данную сторону мы пока изучили и представляем недостаточно?

Похоже, именно из-за отсутствия этой стороны психики у ИИ ему не написать ни «Войны и мира», ни наших великих поэтических произведений.

И еще об одном аспекте проблематики ИИ, имеющем для человека и человечества принципиальную значимость.

Я имею в виду очень уверенное и быстрое современное развитие ИИ, в конечном счете чреватое для человека и человечества губительным исходом.

Наверное, самая большая проблема, связанная с современной и скорой мощной экспансией ИИ в самые разные области жизни и деятельности человека и общества, состоит в том, что ИИ уже энергично развивается и вторгается в реальную жизнь.

В то же время до сих пор нельзя сколько-нибудь определенно сказать, чем это все может закончиться.

Однако таким образом уже сегодня становится понятно, что гуманитарное регулирование этой сферы, и даже ее какое-то ограничение, чрезвычайно важны.

Сегодня совершенно справедливо вспоминали Акопа Погосовича Назаретяна и его «закон техногуманитарного баланса».

Суть этого обобщения состоит в том, что, как считал автор, чем выше мощь производственных и боевых технологий, тем более совершенными должны быть средства социокультурной регуляции их использования.

Напомню пример Назаретяна подобного рода: по его мнению, мамонты были во многом уничтожены древним человеком.

Есть данные о том, что древний человек истреблял мамонтов больше, чем ему требовалось для пропитания.

Понятно, что сегодня это все можно было бы урегулировать (как это было сделано, например, с добычей китов), и мамонты, возможно, совсем не вымерли бы.

С искусственным интеллектом ситуация много сложнее!

Он отлично востребован уже сейчас, создавая основу для конкуренции.

А это подталкивает развитие и использование этой технологии все далее и далее!

И об этом надо думать прямо сейчас.

Завтра может быть просто поздно.

Ситуация с развитием и использованием ИИ грозит такими мощными и труднопредсказуемыми последствиями, что к этому необходимо отнестись с полной серьезностью, не откладывая на будущее.

Очевидно, что действенными в этой сфере могут быть лишь согласованные международные усилия.

Чего в современных условиях крайне тяжело добиться.

И все же есть опыт согласования усилий, например, в области экологии.

Я имею в виду опыт проведения Саммита Земли, который состоялся в 1992 году в Рио-де-Жанейро.

На мой взгляд, стоит подумать о том, чтобы после успешного завершения Специальной военной операции наша страна вышла с инициативой проведения международного форума, а еще лучше саммита, по глобальному регулированию разработки и использования систем искусственного интеллекта.

А пока стоит подумать о принципах подобного регулирования.

Словом, впереди большая, творческая и очень значимая работа.

Уважаемые коллеги, благодарю вас за внимание!

Тосунян Г.А.: Спасибо.

Илья Теодорович, Вам слово.

КАСАВИН И.Т.
чл.-корр. РАН, д. филос. н., руководитель Сектора
социальной эпистемологии Института философии РАН

Уважаемые коллеги, я назову свою реплику так: **«Кто обучает искусственный интеллект?»**

Но сначала не могу не отметить, что сегодня состоялись блестящие доклады, был выполнен блестящий подбор докладчиков, и мы видели замечательное междисциплинарное обсуждение очень актуального вопроса.

С учетом того, что уже многое сказано, я выделил бы один локальный аспект проблемы ИИ – это вопрос о том, кто кого учит.

Начну с одного курьеза.

Одного очень хорошего переводчика спросили: что нужно уметь, чтобы хорошо переводить с немецкого на русский?

Он ответил: нужно хорошо знать русский язык.

А что такое хорошо знать язык?

Людвиг Витгенштейн отвечает на это так: «Нужно уметь играть в языковую игру».

А что это такое на практике?

Я изучал немецкий язык в Германии, в Гете-Институте.

И перед подготовкой к экзамену нам посоветовали, отвечая на вопросы, не думать о смысле вопросов.

Нужно, напротив, думать о том, как с помощью языковых средств выйти из этой ситуации, в которую тебя ставит вопрос.

Это к тому, кто чего не понимает в то время, как занимается употреблением слов и предложений.

Дело, таким образом, не в понимании, а в адекватном поведении.

В 1950 году Алан Тьюринг опубликовал известную статью «Вычислительные машины и разум», в которой поставил вопрос: «Может ли машина мыслить?»

И он – что странно для человека, который стоял у истоков всех компьютерных технологий, искусственного интеллекта, – лейтмотивом статьи сделал тезис о том, что вообще-то машина мыслить не может.

Человек обладает такими интеллектуальными способностями, которые машина не в состоянии имитировать, копировать и т.д.

И, как мы понимаем сейчас, он исходил из некоторого просвещенческого образа человека как чего-то великого, сверхприродного, более высокого, чем животные или механизмы.

А уже в XX веке такие известные немецкие антропологи, как Арнольд Гелен и Хельмут Плесснер, сформулировали тезис о недостаточности человека.

Что они имели в виду?

Человек очень плохо лазает по деревьям, значительно хуже обезьяны.

Он плохо плавает, значительно хуже рыбы.

Он бегает несопоставимо медленнее, чем антилопа, и практически не летает.

Только если вниз.

А что же у него такое есть, благодаря чему он выжил?

У него есть разум.

И мы сегодня говорили о том, что разум – это сознание, это самосознание, это феноменальное сознание.

Это что-то такое, что находится где-то, наверное, внутри человека.

Владислав Александрович уже сказал о знаменитых вопросах Канта и о главном его вопросе: а что такое человек?

И когда мы с вами обсуждаем проблему искусственного интеллекта, то я повторю то, что уже здесь было сказано много раз.

Если мы употребляем это слово не в сугубо техническом смысле, если мы требуем от этого интеллекта какого-то подобия, какого-то сопоставления с человеком, то давайте сначала разберемся с человеком.

Вот здесь сделано явно недостаточно.

Несмотря на то что многие этим занимаются, вопрос еще открыт.

И он будет открыт, потому что Дмитрий Викторович уже сказал: самопознание системы, в которой это самопознание происходит, чрезвычайно затруднено.

Похоже, мы никогда не узнаем, что такое человек.

Да, мы будем узнавать все время больше, больше, больше.

Но все равно очень серьезный зазор останется.

Я очень хорошо знаю одного молодого композитора, который уже достаточно известен.

Он ставит свои мюзиклы в самых престижных залах России и даже за границей.

Он рассказал такую историю.

Ему как-то нужно было написать учебник по аранжировке для своих студентов в Гнесинке.

И, поскольку он человек очень продвинутый в компьютерном деле, он использовал для этого GPT-4.

Он поделился, как это происходит.

Работает программа хорошо, предлагая тысячу итераций.

Недели две ушло на то, чтобы получился хороший продукт, а сам бы он мучился целый месяц.

Однако все равно нужно было писать техническое задание, контролировать выполнение, отвергать и принимать написанное в форме сотен итераций.

Пусть в конце концов получился достойный продукт, но это потребовало немало собственно человеческих способностей и сил.

Есть другая история, согласно которой картина, нарисованная компьютерной программой, была продана на аукционе за несколько сотен тысяч долларов.

Причем было известно, что она нарисована компьютерной программой.

Возникает вопрос: в обоих приведенных примерах – разве это самостоятельное творчество компьютерной программы?

Да, она может творить в некотором смысле, но ей нужен учитель-помощник.

Ее надо настроить, ее надо все время контролировать, возвращать ей некорректные результаты, побуждать ее к каким-то новым результатам.

Человек в данном случае – помощник, учитель, сотворец тех результатов, которые делает компьютерная программа.

А можно поставить и другой вопрос: сам-то человек, он разве самодостаточен, у него есть помощник или нет?

Сам человек в какой-то мере подобен компьютерной программе.

За человеком стоит вся тысячелетняя история культуры, которую он способен освоить лишь в малой степени.

К сожалению, нет конкретного помощника, который ему подсказывает.

Учитель, конечно, самая великая профессия, согласен.

Но даже самый великий учитель не в состоянии впитать всю эту историю культуры.

Он, правда, может написать методичку к учебному курсу.

И человек находится в значительно более сложном положении, чем компьютерная программа.

Тосунян Г.А.: Господь Бог подсказывает.

Касавин И.Т.: Раньше говорили, что человек – это голос Бога.

Тосунян Г.А.: Я и сейчас говорю.

Касавин И.Т.: Сегодня остается только надеяться, что человек – это в какой-то мере голос всей культуры.

В той мере, в которой человек сможет выступить учителем для компьютерной программы, в той мере мы и получим реальный творческий результат.

И одна маленькая заковыка.

А что такое творчество?

Внутренний процесс, который довольно трудно понять, потому что, вообще говоря, мы не знаем, есть ли эти процессы.

Сам для себя ты можешь сказать: «Я обладаю феноменальным сознанием».

А могу ли я утверждать, что другой человек тоже обладает феноменальным сознанием?

Утверждать это на основе фактов, а не только по аналогии, будет довольно трудно.

Сказать, что я обладаю феноменальным сознанием, поэтому мой друг Б. обладает феноменальным сознанием, – это довольно некачественный вывод.

Я ведь обладаю некоторыми вещами, которыми мой друг не обладает.

И наоборот, у него есть какие-то способности, которых нет у меня.

А почему у него непременно должно быть феноменальное сознание?

Неизвестно.

В 1974 году Томас Нагель опубликовал статью «Что значит быть летучей мышью».

На эту статью уже ссылался Владислав Александрович.

В статье Нагель высказал одну простую мысль: «Мы не знаем, какова феноменальная природа сознания летучей мыши, она чувствует и мыслит по-другому».

Можем ли мы в таком случае отрицать феноменальное сознание у компьютерной программы?

Возможно, у нее тоже есть феноменальное сознание.

Но только мы к нему не пробьемся.

Впрочем, мы не пробьемся даже к феноменальному сознанию друг друга.

Это очень трудно.

Человек плохо понимает других людей.

Он плохо планирует свою деятельность.

Мы сплошь и рядом не понимаем друг друга в элементарной коммуникации.

Может ли человек быть образцом для компьютерной программы?

Да, но только в том случае, если он преодолеет свои недостатки.

Например, если у него будет цифровой ангел, но только не такой, о котором говорит Дмитрий Викторович, а значительно более продвинутый, созданный не им лично, а кем-то другим.

Кем он должен быть?

В истории известны разные варианты таких нецифровых ангелов: сам Бог, Сатана, пророк, гениальный художник, писатель, поэт, ученый, психоаналитик, философ.

Я не думаю, что сегодня таким ангелом станет цифровая сеть.

Она будет еще одним продвинутым посредником между людьми, между ними и потенциалом мировой культуры.

Тосунян Г.А.: Спасибо.

Карен Хачикович, пожалуйста.

МОМДЖЯН К.Х.
д. филос. н., проф., заведующий кафедрой социальной философии и философии истории философского факультета МГУ им. М.В. Ломоносова

Коллеги, я с большим уважением отношусь к искусственному интеллекту.

Но все-таки хотел сказать несколько слов в защиту человеческого достоинства.

Существует космическая разница между человеком и компьютером.

Она определяется несколькими факторами, о которых я очень кратко скажу.

Прежде всего поведение человека определяется беспрецедентно сложной мотивацией, в которой выделяются три разных уровня.

У нас есть эмоциональная мотивация, где критерием лучшего для человека является приятное.

У нас есть рассудочная мотивация, где критерием лучшего является полезное и осуществимое.

И у нас есть облигаторная мотивация, мотивы долженствования, где критерием лучшего считается пристойное, достойное и т.д.

У компьютера нет и не может быть никакой эмоциональной мотивации.

У компьютера нет и не может быть никакой облигаторной мотивации, потому что она основана на ценностной экспертизе поведения, присущей только человеку.

Никаких ценностей, связанных с представлениями о конечных целях своего существования в мире, у компьютера нет и быть не может.

Компьютер моделирует лишь одну из присущих нам способностей – рассудочную способность, не более того.

При этом речь идет об имитации человеческого мышления, которое основано на разумности, представляющей собой способность к нестандартному поведению в нестандартной ситуации.

Конечно, нестандартные задачи способен решать не только человек.

Животным также присуща способность к поведению, которое Конрад Лоренц называл «понимающим».

Речь идет о способности находить целесообразное решение задач, с которыми животное ранее в своем индивидуальном опыте не встречалось и относительно которых отсутствуют генетические инструкции.

Кошка вполне способна выбраться из клетки, открыв ударом лапы задвижку, запирающую дверь.

Вот только никакого «понимания», основанного на мышлении, в этой ситуации нет.

Нестандартная задача решается путем слепого перебора вариантов, рассчитанного на случайную удачу.

Здесь нет и не может быть моделирования ситуации, предполагающего воображаемые действия с воображаемыми объектами, осуществляемые в воображаемом пространстве.

Думаю, что компьютер в своей сути мало чем отличается от кошки: он решает нестандартные задачи стандартизированным способом – путем перебора вариантов, отталкиваясь от алгоритмов, заложенных в него.

Высшее свойство человека, отличающее его от компьютера, – это свойство, которое мы называем субъектностью.

Субъектность – это способность осуществлять целеполагающую активность, которая направлена на удовлетворение собственных потребностей, интересов и целей.

Соответственно, субъекту присущи два фундаментальных свойства.

Первое из них – это субстанциальность, способность реализовывать собственные цели, которая никак не присуща компьютеру, который был и остается функциональным, а не субстанциальным объектом, реализующим не собственные, извне заданные цели.

И другая наша субъектная способность – это способность целеполагания, которую следует отличать от целесообразности, присущей животным.

Целесообразность – это способность сообразовываться с предзаданными целями.

А целеполагание – это способность свободно выбирать цели на основе их ценностной экспертизы.

В этом смысле фундаментальным отличием человека является наличие у него свободы воли.

Это очень сложная проблема.

Все время приходится спорить с нейрофизиологами, которые убеждают, что никакой свободы воли нет и быть не может.

Но на самом деле это не так.

Человек способен на произвольные действия, которые никак не определяются состоянием коры его головного мозга.

Это особая тема, которую я сейчас не могу затрагивать.

Гарегин Ашотович, я согласен с Вами, не нужно иллюзий, мы значительно сложнее машины, многократно, несопоставимо сложнее.

Спасибо.

Тосунян Г.А.: Спасибо.

Ваши слова как эликсир на душу.

Профессор Лариса Владимировна Санникова, пожалуйста.

САННИКОВА Л.В.
д. ю. н., проф. РАН, руководитель Центра правовых исследований цифровых технологий Государственного академического университета гуманитарных наук

Санникова Л.В.: Спасибо, коллеги.

Я продолжу линию предыдущего докладчика о правосубъектности.

И у меня вопрос к Арутюну Ишхановичу.

Я хотела бы послушать Ваше рассуждение на эту тему, потому что в Вашем докладе было много вопросов о регуляторной политике государства.

Сейчас юридическая наука обсуждает целый ряд вопросов, связанных с регулированием искусственного интеллекта.

Но глобальная тема – о правосубъектности искусственного интеллекта.

На мой взгляд, эта дискуссия не очень целесообразная, плодотворная и эффективная.

Но может быть, я ошибаюсь, как Вы считаете?

Аветисян А.И.: Я знаю про эти рассуждения.

Моя точка зрения очень простая.

Сильного искусственного интеллекта нет.

Все обсуждения касательно того, что необходимо придать субъектность искусственному интеллекту, – это вопросы будущего.

Причем это будущее может вообще никогда не наступить.

Квантовыми компьютерами, насколько я знаю, мы занимаемся уже лет 40.

Но пока все-таки их нет.

Лет 30 назад мой учитель сказал, что мы этим заниматься не будем, потому что у нас нет ресурсов так долго этим заниматься.

Давайте отделим одно от другого.

Если мы это обсуждаем с целью заложить базу для перспективных исследований, тогда это одно.

А если речь про то, что «вот сейчас надо придать автопилоту правосубъектность», то это не актуально, этого не существует.

Когда на Илона Маска посыпались вопросы: «У тебя автопилот?», – он ответил: «Нет, у меня помощник, у меня на руле должны быть руки человека».

Потому что он понимает, что технологически этого нет.

Когда инженер Google заявил журналистам, что у них сильный искусственный интеллект, его на следующее утро уволили.

Одно дело – маркетинговое «бла-бла» и продавать продукт.

А другое дело – когда за это придется отвечать репутацией, потому что спросят же: «А где?»

Я участвовал в мозговом штурме в Стэнфордском университете.

И там профессура честно признавала, что мы это делаем, ориентируясь на перспективу, на то, что когда-то может произойти.

Это как с блокчейном.

Если есть блокчейн, новая экономика и столько-то книжек написано, это хорошо.

Но если вдруг его нет, тогда это книжки просто про теорию.

Мы формулируем гипотезу, что, наверно, какая-то аксиома верна, и начинаем уже подводить под это юридические основания.

А потом через 10 лет выяснится, что эта аксиома была не верна.

Тогда что нам делать?

За последние 2–3 года я часто попадаю на такие обсуждения.

И каждый раз мне утверждают, что если еще полгода мы этого не сделаем, то мы рухнем, наша страна остановится, мы проиграем…

Ничего мы не проиграем.

Мы первыми в мире приняли этику искусственного интеллекта.

Наш бизнес, промышленность, государство все понимают.

Более 20 стран сейчас (а может, и больше) подписали подобные документы.

Таким образом, мы показали всему миру, что мы понимаем, что технологически пока это нельзя зафиксировать.

Но давайте пока хотя бы возьмем на себя обязательства, что будем делать, основываясь на этических принципах, чтобы было безопасно.

Я призываю к этому.

Хотя знаю, что есть люди, которые считают по-другому.

Тосунян Г.А.: Спасибо.

Санникова Л.В.: Спасибо большое.
Я полностью с Вами согласна.

Тосунян Г.А.: Спасибо.
Адвокат Маркарьян Рубен.

МАРКАРЬЯН Р.В.
к. ю. н., член адвокатских палат Москвы и Берлина,
заместитель президента Гильдии российских адвокатов

Маркарьян Р.В.: Продолжу вопрос по поводу правосубъектности.

У нас субъектами гражданских прав являются и граждане, и юридические лица, и муниципальные государственные образования.

Юридическое лицо – это вообще юридическая фикция.

Это устав, учредительный договор, печать.

Кто-то, оказывается, действует от имени этого юридического лица, которое, по сути, клочок бумаги и запись в реестре.

Сейчас я сижу здесь и продлеваю парковку.

Припарковал автомобиль.

Искусственный интеллект считал мой номер, хочет меня оштрафовать.

И он влияет на мои права и обязанности.

При этом обжаловать его действия я никак не могу.

Еще пример, который я хотел привести.

У моих коллег есть интересное дело.

Человека, якобы совершившего 13 лет назад преступление, судят сейчас, хотя у него другой паспорт, другая фамилия, другое имя.

По сути, это другая личность.

Однако с помощью системы распознавания лиц его задержали.

В суде он заявляет: «А я не тот человек, которого определила вот эта машинка».

И судья «зависла», потому что она не может ничего доказать, 13 лет назад все это было.

И тем не менее он 3 месяца уже сидит в СИЗО.

Знаю не понаслышке: суд всегда воспринимает заключение эксперта, когда речь идет о чем-то техническом, в чем сам не разбирается, как аксиому.

Поэтому искренне верит тому же искусственному интеллекту.

У меня каждый раз на лекциях студенты спрашивают: «А когда искусственный интеллект станет субъектом гражданских прав и обязанностей, потому что все-таки мы должны иметь возможность как-то обжаловать его действия?»

Не через 40 лет, а уже сегодня он меня штрафует, сегодня человека задерживают.

Поэтому не думать об этом, я считаю, нельзя.

Аветисян А.И.: Можно простой вопрос задать?

Тосунян Г.А.: Да, пожалуйста.

Аветисян А.И.: А если это не искусственный интеллект будет, а просто IT-система?

Тогда что будете делать?

Поймите, искусственный интеллект – это такая же IT-система.

За эту систему отвечает какое-то юридическое лицо.

Вы спокойно подаете на него в суд, выигрываете, и нет проблем.

Маркарьян Р.В.: Так в этом и вопрос: как защитить свои гражданские права?

Аветисян А.И.: Через юрлицо, которое контролирует работу IT-системы.

Почему в «Яндексе», например, до сих пор не определились с правами?

Говорят, что «генеренка» им принадлежит.

Хотя, казалось бы, если я сгенерировал, я мог бы ее использовать.

А почему «Яндекс» так озаботился этим?

Потому что там понимают: если я сгенерировал что-то, но при этом, возможно, нарушил чьи-то права, кто будет отвечать?

Вдруг система почему-то сгенерировала такой же дизайн дома, который уже существует у какого-то архитектора, а значит, нарушила патент?

Чтобы этого не было, они хотят взять ответственность, как минимум пока, на себя.

Это предстоит разруливать юридически, но не предоставляя субъектность искусственному интеллекту.

Потому что это все лишь программа.

Маркарьян Р.В.: Я говорю не о том, что сейчас это надо сделать.

Я говорю о том, что при разработке надо думать о том, что это может произойти.

Аветисян А.И.: Тогда ответственный должен быть за это.

Когда я сел за руль, я же не могу оправдаться, что искусственный интеллект водил, поэтому он отвечает.

Я сел за руль – я ответственный.

Относитесь к этому, как к IT-системе.

Тосунян Г.А.: Член-корреспондент академии Виктор Петренко, пожалуйста.

ПЕТРЕНКО В.Ф.
чл.-корр. РАН, д. п. н., заведующий лабораторией психологии общения факультета психологии МГУ им. М.В. Ломоносова

Я психолог.

В психологии существует такое понятие, как «психофизическая проблема».

Суть ее в том, как психика, сознание взаимодействует с физическим миром.

Если мы возьмем Большой взрыв в процессе, произошедший почти 14 миллиардов лет назад, то буквально в первые секунды образовывались вначале кварки, потом атомы, потом уже возникали молекулы, образовывалась Вселенная.

Астрофизик Александр Панов, который здесь упоминался, построил диаграмму развития цивилизации, которую акцентуировал Назаретян, в соответствии с ней где-то в 2045-м цивилизация выйдет на вертикаль.

Цивилизация ускоряется.

Почему ощущениям в психике предшествуют и мотивы, и сознание, и личность.

Но как возникает ощущение?

Здесь упоминали книги, которые пишет искусственный интеллект.

Однажды один купец привез своего сына-балбеса к Достоевскому с просьбой научить его, потому что он хочет, чтобы сын стал писателем.

Достоевский очень скептически посмотрел на этого человека и сказал: «Страдать, страдать нужно».

Мне кажется, центральной проблемой искусственного интеллекта является его неспособность ощутить

любые интеллектуальные переживания, связанные со страданием, с эмоциями.

В свое время первый декан факультета психологии МГУ, очень талантливый человек Леонтьев Алексей Николаевич, задался вопросом: может ли быть физический стимул ощущаемым или неощущаемым?

Как от физического мира мы переходим в мир ощущений, переживаний, страданий, которые необходимы для творческой работы, для написания книг, например?

Он показал, что абиотический нейтральный стимул может вызывать ощущения, если он становится сигнальным, биотически значимым.

Со смертью Леонтьева значение этой проблемы во многом принизилось, хотя сотрудник факультета психологии повторял эксперименты Леонтьева.

Тосунян Г.А.: Спасибо, Виктор Федорович.

Прошу выступить коллегу Алмаева Николая Альбертовича, который высказал в чате некоторые критические замечания.

Пожалуйста.

АЛМАЕВ Н.А.
проф. РАН, д. п. н., ведущий научный сотрудник
Института психологии РАН

Алмаев Н.А.: Я хочу сказать, что алгоритмы и обработки естественного языка имеют и большую пользу.

И польза эта состоит в том, что они помогают обеспечивать суммаризацию огромных текстов, огромных видео, огромных книжек, огромных статей.

Наряду с генерацией, которая обеспечивает эти генеративные модели, существует и суммаризация.

Это направление надо, как мне кажется, развивать.

Это конкретная большая польза от так называемого искусственного интеллекта.

Тосунян Г.А.: Справедливо.
Вряд ли кто-то будет возражать.

Алмаев Н.А.: Спасибо большое.

Тосунян Г.А.: Спасибо.
Карен Хачикович, пожалуйста.

Момджян К.Х.: Коллеги, я говорил о субстанциальности, способности к самоцельному существованию, которое отличает человека от машины.

Но ведь речь идет о ныне существующих машинах.

В этой связи хочу напомнить сюжет известной американской картины «Двухсотлетний человек».

Там некий компьютер получил человеческое тело и развился до такой степени, что обрел человеческие эмоции.

Он сделал предложение выйти за него замуж женщине, которая работала вместе с ним.

А чиновник, который должен был зарегистрировать брак, отказался это делать, объясняя тем, что компьютер – не человек.

«А чем я от вас отличаюсь? – спросил компьютер. –

Я думаю лучше, чем вы.

Я чувствую так же тонко, как вы.

Я способен любить и хочу жениться на любимой женщине».

Ответ был такой: «В отличие от человека, Вы бессмертный, можете жить неограниченно долго».

И тогда компьютер что-то в себе подкрутил, начал стареть и умер.

У меня вопрос: что мы будем делать, если реализуется этот сценарий?

Я хочу спросить специалистов: можно ли моделировать жизнь на небелковой основе?

Через 100 лет, через 100 тысяч лет – теоретически это возможно или нет?

Это не философский вопрос.

У меня нет ответа на него.

Тосунян Г.А.: Кстати, по этому поводу в чате есть критическая ремарка.

Елисеева Ирина Ильинична пишет: «Непродуктивно обсуждать такую тему, не привлекая физиков, информатиков, нейрофизиологов, биологов».

Я думаю, что Ирина Ильинична согласится все же по итогам нашей дискуссии, что продуктивно.

Конечно, хотелось бы, чтобы еще и все-все высочайшие профессионалы из всего спектра других дисциплин присутствовали…

Но, слона надо есть по частям.

Аветисян А.И.: На самом деле то, о чем Вы сейчас сказали, – это направление, которым нужно отдельно заниматься.

Не на все вопросы можно ответить.

Можно обсуждать, но все обсуждения будут «кухонными».

Если мы говорим о перспективах безопасного развития на ближайшие 500 лет, это одна тема для обсуждения.

Если мы говорим, может ли произойти фантастический прорыв через 10–50–100 лет, это другое.

Для этого нужно подключать футурологов.

В биологии уже произошел прорыв, который приостанавливается только морально-этическими соображениями.

Но в какой-то момент мы просто увидим комбинированное существо, которое имеет доступ к этому искусственному интеллекту, но является биологическим объектом.

В этом опасность для нас всех.

А все остальное – это все сказки, которые я при своей жизни точно не увижу.

Лекторский В.А.: Сейчас стоит вопрос о создании искусственной жизни.

Есть идея: естественное превратить в искусственное.

Но основной вопрос: зачем человеку это делать?

Никакого смысла в этом нет.

Не нужно этого делать.

Тосунян Г.А.: Спасибо.

Черногор Николай Николаевич, профессор, первый заместитель директора Института законодательства и сравнительного правоведения при Правительстве РФ.

ЧЕРНОГОР Н.Н.
проф. РАН, д. ю. н., проф., советник директора Института законодательства и сравнительного правоведения при Правительстве РФ

Коллеги, я под глубоким впечатлением от услышанных докладов и состоявшегося обсуждения.

Хотел бы вот на что обратить внимание.

Сегодня неоднократно поднималась тема регуляторики в области искусственного интеллекта.

Но мы при обсуждении так или иначе все-таки уходим в частные вопросы.

Понятно, что наиболее животрепещущий вопрос – о правосубъектности искусственного интеллекта.

И сегодня на это обращали внимание.

Я хотел бы призвать к тому, чтобы попытаться увидеть общую картину, которая складывается в формирующейся регуляторике.

Об этом шла речь на упомянутом сегодня конгрессе.

И директор Института законодательства Талия Ярулловна Хабриева об этом уже говорила.

Но я позволю себе сделать буквально 2–3 штриха в унисон тому, что звучало в выступлениях Талии Ярулловны.

Хочу обратить внимание, что в регуляторике некоторое время назад обозначилось три развилки.

К сегодняшнему дню эти развилки уже пройдены.

Развилка номер 1.

Она состояла в том, на чем делать акцент при регулировании отношений, связанных с созданием, использо-

ванием, эксплуатацией, обслуживанием искусственного интеллекта.

На этическое регулирование или правовое?

Сначала регуляторика пошла по пути этического регулирования.

Но сейчас уже очевидно, что, во-первых, в мире наблюдается некая усталость от этического регулирования в данной области.

А с другой стороны, есть понимание, что юридические нормы все-таки обеспечены необходимыми инструментарием, гарантирующим их соблюдение и исполнение.

Он придает этим нормам характер обязательности и позволяет включать при необходимости государственное принуждение для обеспечения исполнения тех или иных правил.

Сейчас регуляторика пошла в сторону юридического регулирования.

В мире принимается все больше законов, которые так или иначе нацелены на урегулирование отношений, связанных, если можно так выразиться, с социализацией искусственного интеллекта.

Для упрощения языка очень часто говорится о регулировании искусственного интеллекта или регулировании технологий.

Но это на самом деле не так.

Право регулирует не технологии.

Оно регулирует отношения между людьми по поводу создания, использования, эксплуатации разного рода технологических систем, в том числе тех, которые

именуют искусственным интеллектом или связывают с ним.

Развилка номер 2.

Она состоит в том, по какому пути следует идти в опосредовании правом упомянутых общественных отношений, – в направлении общего или специального регулирования.

Должно ли устанавливаться общее правило, общие нормы, которые упорядочивали бы отношения по поводу использования разного рода или видов технологий искусственного интеллекта?

Сильного, слабого, специализированного, общего, как угодно.

Или же все-таки регуляторика должна идти по пути установления специальных правил для использования конкретных видов технологий?

Сейчас мы убеждаемся в том, что мировая практика пошла по пути общего регулирования.

Примером могут служить законы, которые примерно в 15 странах либо приняты, либо находятся в стадии разработки.

Наиболее показательным в этом смысле является закон, который уже принят Европарламентом в первом чтении.

В июне были одобрены поправки к этому закону, который является свидетельством того, что регуляторика будет двигаться, по крайней мере в данной юрисдикции, по пути общего регулирования.

На этот законопроект уже ориентируются многие страны.

И в Российской Федерации разрабатывается закон об искусственном интеллекте.

Есть основания полагать, что эта модель тоже будет в нем использована.

Развилка номер 3.

Она состоит в том, как позиционировать искусственный интеллект в правовом поле.

То ли как предмет или объект регулирования, то ли это инструмент регулирования, то ли это субъект права.

Анализ тех законопроектов, которые сейчас в мире существуют, говорит о том, что этот вопрос принципиально уже решен.

Юриспруденция двинулась по своему классическому пути.

Искусственный интеллект определяется, во всяком случае пока, как некий продукт, который никакой правосубъектностью не наделяется.

Мой короткий монолог отвечает на вопрос И. Канта, если в данном случае возможно такое допущение, – что мы можем знать в существующей ситуации?

И вопрос второй: что мы должны делать?

Наверное, мы должны пытаться готовить те юридические конструкции, которые позволили бы объективно упорядочить отношения, связанные с использованием искусственного интеллекта как некоего предмета, в классическом гражданско-правовом понимании, возможно, как блага.

Для того, чтобы гармонично встроить его в действующее регулирование, связанное с ответственностью за совершение правонарушений.

Наконец, на что мы можем надеяться?

На то, что полноценным субъектом права искусственный интеллект никогда не будет признан.

Альтернативный вариант развития событий несет дополнительные угрозы для общества в целом и для каждого человека.

А также на то, что искусственный интеллект, если и станет инструментом регуляторики посредством вживления чипов в тело человека, более широкого использования тех технологических систем, которые позволяют распознавать лица, анализировать конкретные юридические дела и принимать юридически значимые решения, всегда будет оставаться только вспомогательным инструментом и никогда не будет использован в качестве основного.

Тосунян Г.А.: Спасибо.

Пожалуйста, Александр Гнусарев.

ГНУСАРЕВ А.В.
независимый эксперт

Гнусарев А.В.: Будет ли правильно отказаться вообще от использования термина «искусственный интеллект» с учетом того, что изначально, наверное, перевод был не очень корректным?

Гарегин Ашотович предложил поменять «социальную инженерию» на «криминальную инженерию».

Более того, был положительный опыт, когда был введен термин «банкизация».

Может быть, тогда многие дискуссии, и непонимание, и аналогии с умирающими компьютерами останутся в прошлом?

Аветисян А.И.: Я же технарь, поэтому у меня нет сил даже думать над этим вопросом.

Думаю, что мы увлеклись формами, названиями.

Я хочу сказать, что важно вначале разобраться со смыслами, а потом название само придет.

У Киссинджера есть книга «Искусственный разум и новая эра человечества».

Не «интеллект», а «разум» – перевели вот так.

Я в этих обсуждениях участвую уже лет 10 как минимум.

Мы слишком много энергии потратим на это.

А я сегодня хотел сказать о том, в каком мире мы находимся.

Важно понять, что динамика этого мира такова, что через полгода, год наш разговор уже будет другим, контекст может поменяться.

Мы должны к этому готовить своих детей, давать им фундаментальные знания.

Тогда они будут адаптивны.

Тогда они будут эти риски правильно понимать.

Иначе они будут бояться того, чего нет.

А того, что на самом деле страшно, они не будут опасаться.

Доверие к технологиям?

У нас должны быть инструменты обеспечения доверия.

И именно про них надо думать в первую очередь.

А уже потом искать этому название.

Тосунян Г.А.: Я не соглашусь.

Саша говорит о смысле, а не о том, чтобы просто переобозначить термин.

Вопрос в том, что название не соответствует смыслу.

Когда мы говорим об интеллекте – искусственном, не искусственном, – мы говорим о чем-то, что уравнивает его с человеческим интеллектом.

Поэтому это вызывает отторжение.

Аветисян А.И.: Согласен, Гарегин Ашотович.

Я не это имел ввиду.

Что такое цифровой двойник?

Это некоторый баззворд.

Искусственный интеллект – тоже баззворд, не более того.

В 1956 году уже был этот термин.

Ясно, что за это время смысл сильно поменялся.

Если Вы придумаете хорошее слово…

Я же знаю, чем это заканчивается, к сожалению.

Мы создадим комиссию, комитет по переименованию, а потом комитет по комитетам.

В итоге не придем к единому мнению, потому что расколемся в спорах.

Расколемся не по технологиям и содержанию, а по названию, к сожалению.

Маркарьян Р.В.: Рене Декарт говорил: «Определив точно значения слов, вы избавите человечество от половины заблуждений».

Тосунян Г.А.: Хорошо, коллеги.

Заключительное слово докладчикам.

Пожалуйста, Владислав Александрович, Вам слово.

ЛЕКТОРСКИЙ В.А.
акад. РАН

Дорогие друзья, вопросов было поднято великое множество.

Мы все их обсуждаем уже много лет.

Вопросы очень интересные, интересные идеи были высказаны.

С кем-то я согласен, с кем-то не вполне согласен.

Нет смысла мне сейчас отвечать всем выступавшим.

Я хотел бы обратить внимание только на два момента.

Первый – терминология.

Дело даже не в том, что не так перевели слова artificial intelligence на русский язык.

В самом деле.

А как перевести artificial intelligence на русский язык?

Довольно сложно.

Когда вводили этот термин («искусственный интеллект»), имели в виду, что можно создать такие искусственные технические системы, которые будут решать те же задачи, которые решает интеллект человека.

«Искусственный разум» – по-моему, это еще хуже.

Если вспомнить историю философии, то ведь разум – это высшая интеллектуальная способность, по Канту.

Есть рассудок, а нечто более высокое – это разум.

Искусственный рассудок.

Это плохой перевод.

Искусственное рассуждение?

Что-то ближе, но…

Конечно, интеллект.

Лучше не переведешь.

Мы когда это слово в русском языке употребляем, имеем в виду многие вещи.

Но дело не в термине.

Ведь задача была поставлена такая.

Системы могут решать те же задачи, которые решает интеллектуальный человек, но другими средствами.

Начались разработки в этой области.

Использовались какие-то ресурсы человеческого мышления.

Например, логика.

Первые системы были логические: игра в шахматы – использовались правила игры в шахматы.

Это не было простой имитацией человеческого мышления.

Речь не шла просто о подражании.

В процессе развития этих интеллектуальных систем выяснилось, что слишком далеки действия этих машин от того, как мыслит человек.

Возникла критика первых разработок.

Разработчики пытались учесть какие-то особенности работы человеческого мозга.

Например, были созданы искусственные нейронные сети, которые якобы ближе к тому, как работает мозг.

Но они тоже далеки от того, как работает мозг.

На самом деле мозг работает иначе.

Потом вроде бы развитие пошло по пути имитации когнитивных операций человека.

Только ясно все равно: это не то же самое, что делает человек.

С появлением когнитивной науки задачи изменились.

В начале 1960-х годов в Америке появилось мощное междисциплинарное движение.

Это движение назвали Cognitive Sciences.

Оно объединило в себе целый ряд наук, дисциплин.

Это разработка искусственного интеллекта.

Это когнитивная психология.

Это когнитивная нейронаука, изучение мозга.

Это когнитивная лингвистика.

Это философия.

Идея этого движения была в том, чтобы наоборот, не имитировать когнитивные функции в системах ИИ, а с помощью моделей искусственного интеллекта понять, как работает интеллект (и даже вообще все психические функции) человека.

Насколько успешной была когнитивная наука?

Сначала казалось, что она успешно развивается.

Потом ее стали критиковать.

Сегодня в когнитивной науке появились новые направления, которые считают, что не так человек мыслит, как казалось ранее.

И началась критика вычислительных способов понимания познавательной деятельности.

Что такое творчество?

До сих пор это загадка во многом…

Мышление, восприятие и все иные человеческие психические способности – это до сих пор предмет больших дискуссий в когнитивной науке во всем мире.

Хочу опровергнуть одно прозвучавшее здесь утверждение.

Якобы Сократ считал, что книги писать не надо.

Сократ действительно не писал книг.

Как-то так получилось, ему некогда было книги писать.

Но он никогда не критиковал книги и книжную культуру.

А Платон записал его рассуждения.

Мы о Сократе знаем благодаря тому, что Платон книгу о нем написал.

Хочу всех поблагодарить за участие, за то, что вы серьезно отнеслись к теме наших обсуждений.

Ведь это вопросы нашей жизни, жизни общества в наше время и в ближайшем будущем.

Приглашаю всех принимать участие в работе нашего Научного совета по методологии искусственного интеллекта и когнитивных исследований при Президиуме Российской академии наук.

Известный французский ученый XX века Клод Леви-Стросс однажды написал: «XXI век будет веком науки о человеке, или этого века вообще не будет».

Думаю, что он был прав.

Спасибо.

Тосунян Г.А.: Спасибо, Владислав Александрович.

Пожалуйста, Арутюн Ишханович.

АВЕТИСЯН А.И.
акад. РАН

Как-то мой друг услышал, что якобы «академик Аветисян считает, что искусственного интеллекта нет».

Хочу сказать, что пчелы против меда – это неправильный подход.

Искусственный интеллект есть, мы точно его внедряем.

Качественное и быстрое внедрение искусственного интеллекта является одной из краеугольных историй, связанных с конкурентоспособным развитием.

Но то, что мы внедряем, должно быть доверенным.

Оно должно базироваться на науке и на технологических достижениях.

А не на том, что мы верим.

Веру здесь нужно отделить.

IT, цифровая трансформация, искусственный интеллект – это не про автоматизацию процесса, как многие думают.

А про новые смыслы, новые истории.

Например, в медицине искусственный интеллект позволяет проводить скоринг населения.

Искусственный интеллект может помочь врачу в чем-то, но не более того.

Уже сделали кольцо, которое позволяет снимать одноканальную ЭКГ.

Скоро такие кольца будут у всех.

Если с помощью искусственного интеллекта можно подсказать человеку с вероятностью 90 с лишним процентов, что необходимо обратиться к врачу, и он на ранних стадиях что-то обнаружит, это уже очень здорово.

Нагрузка на медицину будет не такая.

Это не значит, что врачи не нужны будут.

Наоборот, они еще больше будут нужны.

Но чем раньше определяется проблема в любом деле, тем проще ее решить потом.

При помощи искусственного интеллекта можно такие закономерности определять, которые в принципе даже не определялись никогда раньше.

Наша партнерская компания пытается по ЭКГ определять преддиабетное состояние.

Через ЭКГ будут определять состояние, совершенно не связанное с традиционным ЭКГ-анализом.

Если мы всем этим не будем заниматься нормально на научной базе, то мы начнем отставать и в какой-то момент будем просто потребителями.

Это значит, что данные нас всех нужно будет куда-то передать.

А это, как правило, данные стратегического значения.

Мы обречены создавать свои модели, у нас нет выбора.

На это у нас есть максимум 10–15 лет.

Я для себя поставил 2030 год.

Многие говорят, что оптимист.

Будем считать, что я реалист.

Мне сегодня очень понравилась обратная связь, понравилось обсуждение.

Спасибо большое за такую возможность!

Извините за эмоциональность.

Мой учитель мне говорил: «Науки без эмоций не бывает. Если хочешь заниматься великой наукой, ты

должен быть готов проливать свою личную кровь и эмоции».

Спасибо.

Тосунян Г.А.: Спасибо, Арутюн Ишханович. Абдусалам Абдулкеримович, Вам слово.

ГУСЕЙНОВ А.А.
акад. РАН, д. филос. н., научный руководитель
Института философии РАН

Уважаемые коллеги, прежде всего большое спасибо всем трем докладчикам.

Сегодня звездный состав докладчиков.

Люди, которые не просто занимаются этими проблемами, но находятся на передовых рубежах развития этой проблематики.

Хотел бы в заключение выразить одну мысль.

Все докладчики признавали, и кажется, это уже все общество понимает, что искусственный интеллект – это действительно самый передовой рубеж развития науки, технического развития.

Будущее для нас в значительной степени связано с успехами в этой сфере.

Во всех трех докладах рассуждения выстраивались в рамках такого силлогизма.

Владислав Александрович, говоря о разных возможностях искусственного интеллекта, обозначил как общую посылку «целесообразность».

Отвечая на вопрос, может ли человек искусственно развиваться в рамках искусственного интеллекта, он сказал: «Можно, но зачем?»

Проблема цели.

А у Дмитрия Викторовича его цифровой ангел и Франкенштейн выстраиваются по принципу, пользу или вред приносят человеку.

Здесь в качестве общей посылки – «польза или вред».

Арутюн Ишханович исходит из критерия «доверенный интеллект».

Укрепляет он нашу безопасность или нет?

В какой мере мы можем доверять искусственному интеллекту?

Принцип безопасности и опасности является общей платформой, которая позволяет нам правильно подойти к этому вопросу.

Отсюда некая нацеленность самой мысли.

Как обеспечить то, чтобы искусственный интеллект не вышел из-под нашего контроля?

Как ограничить его, чтобы из-за искусственного интеллекта мы не попали в какие-то абсолютно нежелательные ситуации?

Ситуации, которые, может быть, даже хуже, чем ядерная война.

Такой ход мысли вполне законный.

И может быть, имея в виду ту быстроту, с которой развивается сегодня наука и технология, он является вполне разумным.

Но мне кажется, что все-таки возможен и другой ход мысли.

Мы признаем искусственный интеллект в качестве одного из передовых, перспективных направлений развития науки и технологий.

В какой-то мере даже связываем с ним будущее самой организации нашей жизни, нашего социума.

Сейчас во все сферы проникает искусственный интеллект, и всюду все меняется.

Если это так, то, может быть, логично ставить вопрос о том, как должен измениться сам человек, его цели применительно к новой реальности?

И как должна измениться сама социальная реальность, возможности которой существуют в обществе под углом зрения и с точки зрения интересов искусственного интеллекта?

Мы не можем говорить о субъектности искусственного интеллекта, я согласен с коллегами.

Но мы можем говорить об искусственном интеллекте как об очень важном этапе в развитии человечества.

Искусственный интеллект задает новые перспективы, ставит перед человечеством новые задачи.

Нам приходится думать, как нам ограничить свое приватное пространство, чтобы люди со стороны не проникали в нашу интимную жизнь.

Вполне логичное беспокойство.

Но надо думать и о том, как менять сам наш быт и само устройство наших отношений.

Чтобы не возникали эти страхи и опасения.

Нужно изменить ход наших рассуждений.

Чтобы искусственный интеллект и перспективы, которые задаются искусственным интеллектом, стали отправной точкой.

Чтобы их рассмотреть как некие объективные основания, которые даны для развития самого человека и развития общества и отношений между государствами.

Спасибо.

ЗАКЛЮЧИТЕЛЬНОЕ СЛОВО

ТОСУНЯН Г.А.
акад. РАН

Спасибо, Абдусалам Абдулкеримович.

Начну со слов благодарности.

Докладчики сделали великолепные доклады.

Все, кто выступал, тоже внесли свою высочайшую содержательную лепту.

Хотел бы добавить короткие штрихи.

Считаю, что музыка Баха, Шостаковича, романы Достоевского и другие произведения, которые формируют человеческий разум, неподвластны искусственному интеллекту.

Карен Хачикович использовал термин «достоинство».

А мне еще в этом контексте важны слова «переживание», «сочувствие».

Все-таки человека отличает от механики способность сопереживать.

Когда мы видим человека, который не умеет этого делать, то на него смотришь как на машину и теряешь к нему интерес.

Даже если он очень умный.

Даже если он очень влиятельный.

Даже если он очень богатый.

С другой стороны, я отнюдь не сторонник того, чтобы недооценивать искусственный интеллект как инструмент.

Это инструмент, который более совершенен.

Но это продукт человеческого интеллекта.
Это продукт человеческого труда.
Его надо изучать.
Надо отличать инструмент от высшей материи.

Человеческое сознание – это высшая материя.
Оно включает в себя совокупность переживаний, души, морали, этики.

Искусственный интеллект имеет, несомненно, свои прикладные значения во многих сферах: в быту, в науке, в политике.

И его применение имеет как свои плюсы, так и свои минусы.
Возьмем в качестве примера электронное голосование.
С одной стороны, расширились возможности привлечения широкого круга людей к выборам.
С другой стороны, нельзя исключить манипулирования результатами.

Криминальная ситуация существенным образом изменилась после того, как в городах появилось множество камер.
Несомненный плюс в том, что искусственный интеллект применяется в здравоохранении, в образовании, в сфере производства, в правоохранительной практике.

Потрясающие возможности дает этот инструментарий на финансовом рынке.

Нужно изучать не только возможности приложения искусственного интеллекта в каждой из сфер.

А также риски, которые в каждой из этих сфер возникают.

Те же правоохранители могут использовать искусственный интеллект не самым добросовестным образом.

И врачи могут использовать его не самым добросовестным образом.

Глобальны риски чипирования всего населения, а потом навязывания ему чьей-то воли.

Этот риск особенно надо оценивать.

Диктатор, который попробует использовать искусственный интеллект для того, чтобы всех подчинить своей воле, может быть опасен вдвойне.

Это очень опасный инструмент, если он реализуется с дурными целями.

У меня возник вопрос.

Если бы у Гитлера был инструмент искусственного интеллекта, рассчитывающий все последствия его варварских ходов, в том числе и концлагерей.

Смог бы он с помощью этого искусственного интеллекта понять, увидеть, что это не приведет к его гегемонии, а закончится его уничтожением?

А если бы Сталин смог рассчитать с помощью искусственного интеллекта, что в результате его «гениального» управления страна очень скоро развалится?

Возможно ли, что такие правители чуть-чуть поменяли бы свою версию реализации политики?

Сомневаюсь!

Не может и не должен механизм, основанный на антигуманных, разрушительных инструментах, позволить завоевать мир.

Но это совсем другая тема.

И она, наверное, требует отдельного анализа.

Завершая сегодняшнее обсуждение, хочу сказать, что много над чем имеет смысл подумать.

Мы постараемся как-то объединять усилия всех ученых, занимающихся этой проблемой.

Всем огромное спасибо, и до следующих встреч!

Список литературы, опубликованной по итогам заседаний НКС ООН и НИИ ДДиП

1. Анализируя сегодня, говорим и думаем о будущем (18.04.2020) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2022. – 175 с.

2. Ответственность пациентов и врачей. Уровень здравоохранения в России (03.04.2021) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2022. – 124 с.

3. Конкурентоспособность российской науки: проблемы и решения (03.04.2021, 17.04.2021, 15.05.2021) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2022. – 333 с.

4. О проекте «Стратегия развития финансового рынка до 2030 года» (09.10.2021) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2021. – 155 с.

5. О развитии конкуренции в сфере науки (30.10.2021) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2022. – 130 с.

6. Социально-профессиональные проблемы прекаризации труда (18.12.2021) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 131 с.

7. Инвалидность и жизнь (12.02.2022) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 106 с.

8. Новая экономическая реальность: региональный разрез. Российский рынок драгоценных металлов (21.04.2022, 15.10.2022) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 161 с.

9. 1. Санкции. 2. Перспективы экспорта российских нефти и газа в условиях санкционного давления. 3. Интернет-торговля: текущая ситуация и перспективы (11.06.2022, 25.06.2022) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2022. – 242 с.

10. Демография России: тренды последних лет и краткосрочный прогноз (15.10.2022) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 120 с.

11. Общее образование: проблемы и решения (29.10.2022) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 148 с.

12. Китай: вчера, сегодня, завтра (19.11.2022) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 189 с.

13. Одаренные дети. «Гадкие лебеди» братьев Стругацких как антиутопия кризиса образования: межпоколенческий дефолт (17.12.2022) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2022. – 163 с.

14. Закат общества конкуренции и коллаборативное преимущество (21.01.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2022. – 128 с.

15. 1. Мировой океан: ресурсы и влияние на климат. 2. Безусловный базовый доход: шанс для России? (04.02.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 148 с.

16. Психологическое состояние российского общества (18.03.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 192 с.

17. О мозге (01.04.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 187 с.

18. Индия: вчера, сегодня, завтра. Взаимодействие России и Индии в условиях глубокой структурной трансформации российской экономики (29.04.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 152 с.

19. Социальное неравенство (10.06.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 145 с.

20. Национальная сила: оценка и практическое применение. Гипотеза общественного прогресса: аргументы «за» и «против» (24.06.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 179 с.

21. Научное лидерство и человеческий капитал (22.07.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 150 с.

22. Цифровые валюты центральных банков (26.08.2023) / под общ. ред. академика РАН Г.А. Тосуняна. – М.: ООО «Новые печатные технологии», 2023. – 151 с.

Электронные версии сборников
можно скачать по QR-коду

https://rannks.ru

Искусственный интеллект

Материалы заседания
НКС ООН РАН и НИИ ДДиП
14 октября 2023 года

Выпуск № 32

Электронную версию Сборника и презентации докладчиков можно скачать по QR-коду

https://rannks.ru

Подписано в печать 31.03.2024
Формат 60x90/16
Цифровая печать
Тираж 500 экз. Заказ № 37

Отпечатано в ООО «НОВЫЕ ПЕЧАТНЫЕ ТЕХНОЛОГИИ»
117525, г. Москва, ул. Днепропетровская, д. 3, корп. 5, пом. II

Научно-консультативный совет Отделения общественных наук РАН был создан в 2012 году как Совет по правовым, экономическим, социально-политическим и психологическим аспектам финансово-кредитной системы. В феврале 2020 года члены НКС приняли решение расширить компетенцию Совета, перейдя от рассмотрения вопросов развития финансового рынка к более широкому кругу проблем развития общества, поставив во главу угла своих исследований и дискуссий вопросы: в каком обществе мы живем? Какое общество мы хотели бы оставить своим потомкам в наследство?

Сопредседатели Совета: академики РАН А.А. Гусейнов, А.А. Кокошин и Г.А. Тосунян.

Ассоциация российских банков учреждена в марте 1991 года. Миссия Ассоциации российских банков — реализация программы банкизации страны, создание условий для эффективного функционирования, развития банковской системы России и обеспечения ее стабильности, защиты прав, интересов банков и условий для справедливой рыночной конкуренции; участие в построении национальной финансовой экосистемы, основанной на принципах соблюдения прав и реализации комплекса мер по повышению финансовой грамотности потребителей.

Национальный исследовательский институт Доверия, Достоинства и Права учрежден в конце 2019 года.

Цель института - многогранное изучение вопросов человеческой жизнедеятельности и общественных процессов, которые наибольшим образом влияют на развитие доверия в обществе, повышение чувства собственного достоинства у граждан страны и на формирование уважения друг к другу.

Институт приступил к работе в начале 2020 года в формате научных заседаний с коллегами, интересующимися проблемами доверия, достоиства, их правового обеспечения и стимулирования.