

РОССИЙСКАЯ АКАДЕМИЯ НАУК

Научный совет по методологии искусственного интеллекта  
и когнитивных исследований при президиуме Российской академии наук

# ЧЕЛОВЕК И СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*Под редакцией  
академика РАН В.А. Лекторского*



Санкт-Петербург  
Издательство «Юридический центр»  
2022

УДК 004.8  
ББК 16.6  
Ч39

**Рецензенты:**

**Жуков В.И.** — академик РАН, главный научный сотрудник  
Института государства и права РАН, Заслуженный деятель науки РФ  
**Тосунян Г.А.** — академик РАН, сопредседатель Научного совета РАН  
по проблемам защиты и развития конкуренции, Заслуженный деятель науки РФ

**Авторы:**

**Лекторский В.А., Васильев С.Н., Макаров В.Л., Хабриева Т.Я.,  
Кокошин А.А., Ушаков Д.В., Валueva Е.А., Дубровский Д.И.,  
Черниговская Т.В., Семёнов А.Л., Зискин К.Е., Любимов А.П.,  
Целищев В.В., Алексеев А.Ю.**

Ч39      **Человек и системы искусственного интеллекта** / Под ред.  
акад. РАН В.А. Лекторского. — СПб.: Издательство «Юридический  
центр», 2022. — 328 с.  
ISBN 978-5-94201-835-1

В книге обсуждаются новые возможности и проблемы, порождённые ускоряющимся использованием систем искусственного интеллекта во всех областях жизнедеятельности человека. Обосновывается идея о необходимости разработки таких систем искусственного интеллекта, которые не враждебны, а доброжелательны в отношении человека, которые будут не его хозяевами, а его помощниками.

Авторы книги — известные учёные, специалисты в области философии, математики, нейронауки, психологии, права, лингвистики, экономики.

Книга предназначена для широкого круга читателей: как исследователей, так и практиков. Она также может быть использована в учебном процессе.

УДК 004.8  
ББК 16.6

*Научная монография рекомендована к печати Научным Советом  
по методологии искусственного интеллекта и когнитивных исследований  
при президиуме Российской академии наук*

ISBN 978-5-94201-835-1

© Коллектив авторов, 2022  
© Лекторский В.А., отв. ред., 2022  
© Издательство «Юридический центр», 2022

---

---

# СОДЕРЖАНИЕ

---

---

<b>Предисловие</b> (акад. РАН, <i>Лекторский В.А.</i> ) . . . . .	7
<b>Искусственный интеллект в изучении человека, человек в мире, создаваемом искусственным интеллектом</b> (акад. РАН, <i>Лекторский В.А.</i> ) . . . . .	10
Искусственный интеллект и когнитивные исследования . . . . .	11
Глобальная цифровизация как антропологический вызов. . . . .	18
<b>Искусственный интеллект и общество</b> (акад. РАН, <i>Васильев С.Н.</i> ) . . . . .	30
Искусственный интеллект и интеллектуальные системы . . . . .	31
Об интеграции средств искусственного интеллекта . . . . .	34
Об интеграции ИС на базе подстановочных логик . . . . .	45
К устойчивому развитию . . . . .	51
<b>Как современные цифровые технологии меняют представление об окружающем мире (искусственное общество)</b> (акад. РАН, <i>Макаров В.Л.</i> ) . . . . .	59
Вычисления как основа для объяснений устройства мира . . . . .	—
Цифровые двойники, искусственное общество . . . . .	61
Моделирование сознания . . . . .	63
Моделирование духовного мира. . . . .	65
Цифровые двойники в повседневной жизни. . . . .	68
Как человечество движется от материального мира к духовному . . . . .	—
<b>Право, искусственный интеллект, цифровизация</b> (акад. РАН, <i>Хабриева Т.Я.</i> ) . . . . .	71
<b>Вопросы применения технологий и систем искусственного интеллекта в военной сфере</b> (акад. РАН, <i>Кокошин А.А.</i> ) . . . . .	98
Общие вопросы . . . . .	—
Направления (сферы) возможного применения технологий (элементов) и систем ИИ . . . . .	99

<b>Вызовы искусственного интеллекта для психологии</b> (акад. РАН, <i>Ушаков Д.В.</i> , канд. психол. наук, <i>Валуева Е.А.</i> ) . . . . .	107
Искусственный и естественный интеллект: анalogии и различия . . . . .	—
Взаимодействие человека с искусственным интеллектом . . . . .	117
Индивидуальный цифровой ангел . . . . .	123
Ассимиляция искусственного интеллекта как цивилизационный вызов. . . . .	126
<b>Сознание, мозг, общий искусственный интеллект: новые стратегические задачи и перспективы</b> (докт. филос. наук, проф., <i>Дубровский Д.И.</i> ) . . . . .	128
О соотношении понятий общего и узкого, сильного и слабого, естественного и искусственного интеллекта . . . .	129
Сознание и AGI. Нужна ли «Трудная проблема сознания»? . . . . .	139
Субъективная реальность: ее общие свойства, ценностно-смысловые и интенционально-волевые оперативные структуры. . . . .	147
<b>Естественный и искусственный интеллект: смыслы или структуры?</b> (член-корр. РАО, <i>Черниговская Т.В.</i> ) . . .	160
<b>Расширенная личность как основной субъект и предмет философского анализа. Следствия для образования</b> (акад. РАН, <i>Семёнов А.Л.</i> , канд. пед. наук, <i>Зискин К.Е.</i> ) . . . . .	172
Информационные революции в истории человека. . . . .	—
Результаты информационных революций. Контрреволюция . . . . .	174
Расширения сознания. Расширенная личность человека. . . . .	176
Устная речь в контексте Четвертой революции . . . . .	181
Цифровая компетентность (Computational Thinking) . . . . .	184
Цифровая трансформация образования. . . . .	185
Организация образовательного процесса . . . . .	187
Результаты образования. Универсальные навыки . . . . .	189
Школьные предметы . . . . .	192
Нельзя игнорировать контрреволюцию. . . . .	195
Исторические истоки, параллели и ассоциации . . . . .	197

<b>О стандартизации понятий и терминов, связанных с искусственным интеллектом</b> (докт. юрид. наук, проф., <i>Любимов А.П.</i> ) . . . . .	201
Общие вопросы и направления стандартизации понятийного аппарата в сфере применения искусственного интеллекта. . . . .	—
Правовые проблемы, возникающие в процессе проектирования, производства и эксплуатации технических средств с искусственным интеллектом . . . . .	203
Федеральный проект «Искусственный интеллект» . . . . .	211
Отраслевая и функциональная стандартизация понятийного аппарата . . . . .	212
Национальный стандарт в области больших данных. . . . .	219
Отраслевые и функциональные проблемы искусственного интеллекта. . . . .	221
Организационные и гуманитарные проблемы искусственного интеллекта. . . . .	223
<b>Аргумент Лукаса–Пенроуза–Гёделя и парадоксы искусственного интеллекта</b> (докт. филос. наук, проф., <i>Целищев В.В.</i> ) . . . . .	227
<b>70 лет тесту Тьюринга: может ли компьютер всё?</b> (докт. филос. наук, проф., <i>Алексеев А.Ю.</i> ) . . . . .	249
Комплексный тест Тьюринга в отношении «человек — система ИИ». . . . .	253
Может ли система ИИ делать всё то, что способен делать человек? . . . . .	255
Человек и системы особенного ИИ . . . . .	258
Тест Лавлейс 2.0 . . . . .	260
Тест Лавлейс 3.0 . . . . .	261
Собираательный подход к системе общего ИИ . . . . .	267
Определительный подход к системе общего ИИ . . . . .	270
Наблюдательный подход к системе общего ИИ . . . . .	273
Праксеология отношений «человек — система ИИ». . . . .	275
<b>Приложение. Национальная стратегия развития искусственного интеллекта на период до 2030 года. . . . .</b>	280
<b>Список использованной литературы . . . . .</b>	301

---

---

# CONTENT

---

---

<b>Preface</b> ( <i>Lektorsky V.A.</i> ) . . . . .	7
<b>Artificial Intelligence in Human Studies, the Human being in the World, created by Artificial Intelligence</b> ( <i>Lektorsky V.A.</i> ) . . . . .	10
<b>Artificial intelligence and society</b> ( <i>Vassilyev S.N.</i> ) . . . . .	30
<b>How modern digital technologies are changing perceptions of the world around us (artificial society)</b> ( <i>Makarov V.L.</i> ) . . . . .	59
<b>Law, artificial intelligence, digitalization</b> ( <i>Khabrieva T.Y.</i> ) . . . . .	71
<b>Issues of application of artificial intelligence technologies and systems in the military sphere</b> ( <i>Kokoshin A.A.</i> ) . . . . .	98
<b>Challenges of Artificial Intelligence for Psychology</b> ( <i>Ushakov D.V., Valueva E.A.</i> ) . . . . .	107
<b>Consciousness, brain, general artificial intelligence: new strategic tasks and prospects</b> ( <i>Dubrovsky D.I.</i> ) . . . . .	128
<b>Artificial and Natural Intelligence: Meaning or Structures?</b> ( <i>Chernigovskaya T.V.</i> ) . . . . .	160
<b>Extended personality as the main subject and subject of philosophical analysis. Implications for education</b> ( <i>Semenov A.L., Ziskin K.E.</i> ) . . . . .	172
<b>On standardization of concepts and terms related to artificial intelligence</b> ( <i>Lyubimov A.P.</i> ) . . . . .	201
<b>The Lucas-Penrose-Goedel Argument and the paradoxes of artificial intelligence</b> ( <i>Tselishchev V.V.</i> ) . . . . .	227
<b>70 years of the Turing test: Can a computer do everything?</b> ( <i>Alekseyev A.Yu.</i> ) . . . . .	249
<b>Application.</b> . . . . .	280
<b>List of used literature</b> . . . . .	301

---

---

## ПРЕДИСЛОВИЕ

---

---

Сегодня человек живёт в особом, никогда раньше не существовавшем мире, в котором созданные им самим технологии ИИ заставляют его по-новому жить, иначе понимать самого себя, в мире, в котором возникают новые возможности и новые угрозы, в котором вопросы, обсуждавшиеся философами веками, требуют нового осмысления и, не переставая быть теоретическими, становятся жгучими практическими проблемами.

Возникновение разработок в области ИИ побудило многих исследователей попытаться понять естественный интеллект на основании вычислительных моделей, разработанных в исследованиях по ИИ. Так возникла когнитивная наука, развивающаяся на протяжении 50 лет, в которую вошли когнитивная психология, когнитивная нейронаука, когнитивная лингвистика, философия познания и сознания. В рамках когнитивной науки было сделано немало, при этом за 50 лет в ней существенно менялись подходы и концепции. Вместе с тем эта наука продолжает оставаться ареной для больших дискуссий по принципиальным проблемам, прежде всего по вопросу о возможностях и границах вычислительного подхода в понимании рациональности, познания, психической жизни и работы мозга. Главный из дискуссионных вопросов — вечный философский вопрос о характере и природе сознания, человеческой субъективности, свободы выбора, понимания и творчества, вопрос о природе норм — логических, познавательных, моральных, правовых, о взаимоотношении сущего и должного, ибо из того, что нечто существует, не следует, что это должно быть. Сами по себе вычислительные модели этих вопросов не решают.

Ускоряющееся использование систем ИИ во всех областях жизнедеятельности — в информационных технологиях, в экономике, в оборонном комплексе, в культуре, в образовании, в человеческих отношениях — создало новые колоссальные возмож-

ности и вместе с тем породило ряд проблем, несущих угрозу тем ценностям, принятие которых делают человека человеком.

Это прежде всего ценность индивидуальной автономии, свободы выбора, понимание человеком смысла своих действий и вытекающая из этого ответственность за результаты. В том случае, когда искусственная интеллектуальная система рекомендует человеку определённые действия, и к тому же когда он не может понять, на основании чего даётся такого рода рекомендация (и система не понимает сама себя), он перестаёт воспринимать себя в качестве самостоятельной личности.

Возможность обработки больших данных с помощью ИИ и выявление таким путём эмпирических корреляций ставит вопрос о будущем научных исследований: высказываются мнения о том, что в этих условиях больше нет необходимости ни в гипотезах, ни в теориях, ни в экспериментах. Обнаруженные корреляции дают возможность предсказания без выявления причинных зависимостей, без объяснения и понимания.

Использование систем ИИ в собирании и обработке персональных данных взламывает границу между приватным и публичным пространством, создаёт возможность для манипуляции и управления поведением.

Всемирная информационная сеть и «Интернет вещей» ставят проблему границ индивидуальности и сохранения этих границ. Юристы обсуждают вопросы о возможности наделения правосубъектностью систем ИИ и о т. н. «машиночитаемом» праве, которое становится непонимаемым не только обычными гражданами, но и специалистами.

Словом, сегодня мир столкнулся с проблемой «дегуманизации». Исторически считалось, что именно разум, рациональность отделяет человека от животного. Сегодня в лице систем ИИ человек имеет дело с такого рода рациональностью, которая нередко ему чужда, в ряде случаев она непонятна и грозит ему лишением фундаментальных человеческих качеств.

Где же выход? Он может быть только одним: разработка таких систем ИИ, которые не враждебны, а доброжелательны в отношении человека, которые будут не хозяевами человека,



а его помощниками. Это значит включение в программы искусственных интеллектуальных системы моральных норм (т. н. этический и «доверенный» ИИ). Это предполагает разработку моральными философами и моральными психологами таких норм, учитывая при этом, что нормы, принятые в разных культурах, могут в чём-то различаться, что существуют, например, разногласия в понимании социальной справедливости, политических свобод и других ценностей.

Иными словами, разработка и использование систем ИИ ставит человека перед необходимостью нового понимания самого себя, своих возможностей и границ, перед необходимостью самосохранения и создания возможностей для саморазвития. Развитие систем ИИ в острейшей форме ставит главный мировоззренческий вопрос: «Что есть человек?».

Статьи этой книги написаны специалистами в разных областях, входящих в общее направление когнитивных исследований: специалистами в области искусственного интеллекта, философами, математиками, психологами, исследователями в области нейронаук, экономистами, юристами. Все авторы — члены Научного Совета по методологии искусственного интеллекта и когнитивных исследований при Президиуме Российской Академии Наук. Авторы развивают разные подходы, их позиции не во всём совпадают. В целом эта книга — попытка разобраться в одной из острейших проблем нашего времени.

Председатель Научного совета  
по методологии искусственного интеллекта  
и когнитивных исследований при президиуме  
Российской академии наук,  
академик РАН *В.А. Лекторский*

## **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ИЗУЧЕНИИ ЧЕЛОВЕКА, ЧЕЛОВЕК В МИРЕ, СОЗДАВАЕМОМ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ**

---

---

Возникновение области исследований и разработок, получившей название «искусственный интеллект», относится к середине XX века. Имелась в виду возможность искусственного воспроизведения интеллектуальных операций человека. С самого начала по этому вопросу существовали разные мнения. Одни считали, что лишь некоторые из естественных интеллектуальных действий могут быть воспроизведены искусственно. Но были и энтузиасты, которые утверждали, что не только мышление, но и все познавательные процессы и даже все психические состояния могут быть имитированы искусственным образом. С тех пор прошло более полувека. В разработках моделей искусственного интеллекта получены большие результаты. Хотя споры о возможностях искусственного воспроизведения того, что всегда считалось отличительной особенностью человека — его разумности и сознания, не утихают. Вместе с тем использование систем искусственного интеллекта для решения экономических, соци-

---

*\* Лекторский Владислав Александрович* — академик РАН, главный научный сотрудник Института философии РАН. Председатель Научного совета по методологии искусственного интеллекта и когнитивных исследований при президиуме РАН.

*Lektorsky V.A.* — Full member of the Russian Academy of sciences, Principal researcher, Institute of Philosophy, Russian academy of sciences.

альных, политических, оборонных, межчеловеческих проблем, связанный с этим процесс глобальной цифровизации создают совершенно необычную среду человеческой жизнедеятельности и порождают необходимость нового понимания человека, его возможностей и его будущего. Это можно считать настоящей антропологической революцией.

### **Искусственный интеллект и когнитивные исследования**

Примерно 50 лет тому назад возникло междисциплинарное движение, получившее название когнитивной науки (Cognitive Science). Идея о том, что естественный интеллект не просто может искусственно имитироваться, но что он по своей сути является не чем иным, как сложно устроенным вычислением, была воспринята как рядом психологов (появилась когнитивная психология), так и некоторыми лингвистами, специалистами в области нейронаук и другими учёными. В рамках когнитивной науки стало возможным взаимодействие разных дисциплин: как эмпирических и экспериментальных, так и теоретических. Это было определено тем обстоятельством, что дисциплины, вошедшие в когнитивную науку, исходили из следующих двух установок:

1. Все познавательные процессы могут быть поняты как вычисления.

2. Последние осуществляются на носителях информации — ментальных репрезентациях.

Если речь идёт о разработке конкретной научной теории, о проведении экспериментов на той или иной теоретической базе, о технологическом использовании научных результатов, то к философскому пониманию теории можно и не обращаться. Но, как сегодня ясно всем, кто изучает науку, научная теория существует не сама по себе, а строится в рамках той или иной исследовательской программы. Последняя же включает картину мира, т. е. концепцию неких предельных оснований мироустройства.

И формулировка этой концепции означает вступление на почву философии, что предполагает учёт наработок в этой области, включение в делящиеся веками философские дебаты.

Упомянутые два основания возникшей когнитивной науки определённо предполагали философскую интерпретацию.

Прежде всего нужно было выяснить, что следует понимать под вычислением? Ведь это не просто математический, но также и философский вопрос. Какие вычисления вообще возможны? Можно ли непрерывные процессы понять как вычислительные? Можно ли взаимодействие между нейронами в головном мозгу при переработке когнитивной информации истолковать как вычисление? А интенциональность (специфическая особенность актов сознания), а феноменальный характер сознания (т. н. проблема «квалиа»), а интроспекцию, т. е. осознание состояний своего сознания?

Что же касается ментальных репрезентаций, то нужно было определить отношение носителей информации, над которыми производятся вычисления (так понимаются ментальные репрезентации в когнитивной науке), к тем репрезентациям, которые давно исследовались в философии и о характере которых философы вели и ведут споры: восприятия, представления, понятия, суждения (пропозиции), категории, когнитивные схемы и др.

Недаром философия одной из первых была включена в число дисциплин, образовавших когнитивную науку, и именно философы стали в ней главными теоретиками.

Это, например, известный американский философ Х. Патнем, который в начальный период развития когнитивной науки сформулировал исследовательскую программу функционализма, согласно которой все ментальные состояния (не только когнитивные в узком смысле слова, но и такие, как желания, намерения, эмоции и т. д.) могут быть поняты в качестве функций других состояний (ментальных и не-ментальных) агента познания и действия<sup>1</sup>. Функционализм стал философским обоснованием

---

<sup>1</sup> Putnam H. The Nature of Mental States // Readings in Philosophy of Psychology / Ed. by N. Block. 1980. Vol. 1. P. 223–231. Cambridge, Mass. Harvard University Press.

вычислительного понимания ментальности. Из такого подхода вытекает ряд важных следствий относительно характера исследований в рамках когнитивной науки и взаимоотношения входящих в неё дисциплин. Ведь согласно функционализму, природа ментального состояния определяется только его функцией, которая может пониматься как вычисление и не зависит от того, на каких носителях эта функция реализуется (тезис о «множественной реализуемости ментальных состояний»). Но в соответствии с подобным пониманием исследование ментальных состояний человека никак не зависит от изучения процессов переработки информации нейронными структурами головного мозга: когнитивная психология может использовать в качестве объясняющих моделей результаты разработок в области искусственного интеллекта и вовсе не обязана считаться с исследованиями нейронаук: ведь мозг человека с точки зрения функционализма — это лишь один из многих способов реализации ментальных состояний.

Другой американский философ Дж. Фодор, развивая идеи в рамках функционализма, выдвинул теорию врождённого языка мысли, который определяет правила переработки информации когнитивной системой человека и может быть уподоблен программе компьютерного устройства. Как компьютер сам по себе не имеет дела со смысловым характером перерабатываемой информации (этот смысл возникает при интерпретации результатов этой работы человеком), так и мозг человека согласно Фодору имеет дело исключительно с формальными, а не с содержательными характеристиками информации, т. е. не с семантикой, а лишь с синтаксисом.<sup>1</sup>

По мнению Фодора, вообще многие идеи, связанные с пониманием познания и сознания и сформулированные первоначально в чисто философской форме, получают вторую жизнь в когнитивной науке.

Между тем новые программы в когнитивной науке продолжали появляться. В начале 80-х годов прошлого столетия стал

---

<sup>1</sup> *Fodor J. Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology* // Fodor J. *Representations*. Cambridge. Mass. MIT Press, 1981.

популярен т. н. коннекционизм, опиравшийся на представления о том, что ментальные состояния возникают в результате не последовательного, а параллельного и распределённого процесса переработки информации<sup>1</sup>. Если первые программы в когнитивной науке использовали наработки в области символической логики (пережившей взлёт в XX веке) и философские идеи о ментальных репрезентациях, пропозициях, пропозициональных установках и способах рассуждений, обсуждавшихся в аналитической философии первой половины XX века, то коннекционистская программа опиралась на другие традиции: философский эмпиризм, сенсуализм, индуктивизм и психологический ассоцианизм. Правда, эти традиции получили существенно новый вид с помощью разработки моделей искусственных нейронных сетей и представления о характере передачи сигналов от одних групп нейронов другим. Появление коннекционистской программы породило дискуссии о характере вычислительного подхода в анализе когнитивных процессов и стало стимулом для разработки новых моделей искусственного интеллекта.

Результаты, полученные в исследовании работы мозга и в выявлении связи нейронных структур с субъективно переживаемыми ментальными состояниями к началу 90-х годов прошлого века, заставили большинство исследователей в когнитивной науке отказаться от функционалистской программы и от тезиса о множественной реализуемости когнитивных процессов. На одно из первых мест выдвинулась когнитивная нейронаука. Стала популярной идея о том, что именно результаты в рамках этой науки помогут создать искусственный сверх-мозг и на этой основе понять природу человеческой субъективности, расшифровать способы мозгового кодирования информации и в итоге создадут возможность читать чужие переживания и мысли с помощью доступа к мозгу другого человека. Понятно, что, если бы можно было осуществить нечто подобное, последствия для жизни человека были бы неописуемы. Большинство исследователей

---

<sup>1</sup> *Smolensky P. On the Proper Treatment of Connectionism // Behavioral and Brain Sciences. 1988. Vol. 11. P. 1–74.*

в области когнитивной нейронауки считают, что понять характер переработки информации головным мозгом можно с помощью моделей искусственного интеллекта, хотя это должны быть новые модели, учитывающие те факты о работе мозга, которые были получены в последнее время (вычислительная когнитивная нейронаука).

Между тем в последние 20 лет в когнитивных исследованиях становится всё более влиятельным другой подход, который исходит из того, что само по себе изучение работы нейронных сетей, сколь бы детальным оно ни было, не может дать ответ на вопрос о природе сознания, ибо последнее определяется не просто работой мозга (хотя без этой работы оно невозможно), а отношением познающего и действующего агента к внешнему миру, включающему как мир природного окружения, так и в случае человека мир, созданный самим человеком — мир культуры. Этот подход получил название 4 E Cognition, что расшифровывается как понимание познания и сознания в качестве Embodied (телесно-воплощённого), Embedded (вписанного в среду), Enacted (связанного с действием) и Extended (расширенного — это относится к человеку и к использованию им в познавательной деятельности разнообразных искусственных предметов). Подход 4 E Cognition становится сегодня всё более популярным в философии сознания, эпистемологии, в когнитивной психологии, когнитивной лингвистике<sup>1</sup>. Даже в когнитивной нейронауке появилось новое направление — культурная когнитивная нейронаука, исходящая из того, что человеческий мозг — не просто биологическое, но и культурное образование: биоартефакт. Нужно заметить, что многие идеи новейшего направления в когнитивных исследованиях — 4 E Cognition — не столь уж новы для нас, так как близки к тому, что у нас называлось культурно-историческим и деятельностным пониманием познания и сознания (Л.С. Выготский<sup>2</sup>,

---

<sup>1</sup> Chemero A. *Radical Embodied Cognitive Science*. Cambridge. Mass. MIT Press. 2009.

<sup>2</sup> *Выготский Л.С. Мышление и речь*. М. Изд. Лабиринт, 1999.

С.Л. Рубинштейн<sup>1</sup>, А.Н. Леонтьев<sup>2</sup>, А.Р. Лурия<sup>3</sup>, В.В. Давыдов<sup>4</sup> и др. психологи, Э.В. Ильенков<sup>5</sup>, Ф.Т. Михайлов<sup>6</sup>, Г.П. Щедровицкий<sup>7</sup>, В.С. Стёпин<sup>8</sup>, В.С. Швырёв<sup>9</sup> и др. философы, включая автора этих строк<sup>10</sup>). Но если учёт культуры столь важен для понимания сознания и субъективного мира человека, то это ставит вопрос о возможных пределах применения вычислительного подхода в изучении когнитивных процессов человека. Многие исследователи обращают внимание на то, что невозможно представить человеческие действия и функционирование социальных институтов как набор алгоритмизируемых правил. Как, например, показывают когнитивные исследования, уже восприятие не есть просто действие по правилам, а является извлечением такой информации из окружающего мира, характер которой часто нельзя предугадать<sup>11</sup>. Не подчиняется правилам и решение творческих задач — все попытки разработать «логику творчества» оказались безуспешными (эвристики, используемые при создании интеллектуальных систем, такими правилами не являются). Не может быть набором

---

<sup>1</sup> Рубинштейн С.Л. Бытие и Сознание. Человек и Мир. СПб.: Изд. дом «Питер», 2003.

<sup>2</sup> Леонтьев А.Н. Деятельность, сознание, личность. М.: Смысл, 2005.

<sup>3</sup> Лурия А.Р. Лекции по общей психологии. СПб.: Изд. дом «Питер», 2006.

<sup>4</sup> Давыдов В.В. Лекции по общей психологии. М.: Академия, 2008.

<sup>5</sup> Ильенков Э.В. Философия и культура. М.: Политиздат, 1992.

<sup>6</sup> Михайлов Ф.Т. Загадка человеческого Я. М.: Политиздат, 1976.

<sup>7</sup> Щедровицкий Г.П. Мышление-Понимание-Рефлексия. М.: Изд-во «Наследие ММК», 2005.

<sup>8</sup> Стёпин В.С. Теоретическое знание. М.: Прогресс-Традиция, 2000.

<sup>9</sup> Швырёв В.С. Научное познание как деятельность. М.: Политиздат, 1984.

<sup>10</sup> Лекторский В.А. Субъект. Объект. Познание. М.: Наука, 1980; *Lektorsky Vlad-sislav. 2021 / The Activity Approach in Late Soviet Philosophy // The Palgrave Book in Russian Thought / Ed. Bykova M.F., Forster M.N., Steiner L. Palgrave MacMillan. Cham, Switzerland., 1980. P. 407–422.*

<sup>11</sup> Гибсон Дж. Экологический подход к зрительному восприятию. М.: Прогресс, 1988; *Clark A. Being There: Putting Brain, Body and World Together. Mass. MIT Press. 1998.*



правил и плодотворный диалог с другим человеком, предполагающий эмпатию, умение понять позицию собеседника — ход такого диалога невозможно предугадать, но именно поэтому он и ценен. Вообще как раз то, что делает человека человеком, не есть действия по технологизируемым правилам: свободный выбор, понимание другого, признание, взаимная помощь, сострадание, самопожертвование. Моральные предписания (возлюби ближнего, относись к другому не как к средству) не указывают правила их выполнения. Но и в тех случаях, когда есть правила достижения той или иной цели (и в этой связи возможность использовать определённые технологии), сама постановка цели зависит от принимаемых человеком ценностей: от того, что он считает хорошим и плохим, достойным и недостойным, желательным и нежелательным и т. д. А ценности — не набор правил и не сумма технологий. Но ведь именно ценности лежат в основе той или иной культуры и определяют характер социальных институтов.

Исследование когнитивных процессов на протяжении последних 50 лет так и не привело к созданию единой когнитивной науки, о чём когда-то мечтали её основатели. Сегодня многие предпочитают говорить не о единой когнитивной науке, а о когнитивных науках. Тем не менее нужно признать, что в рамках междисциплинарного когнитивного движения возникло интенсивное и плодотворное взаимодействие между такими дисциплинами, например, между лингвистикой и нейронаукой, или же между психологией и разработками в области искусственного интеллекта, которое невозможно было представить себе в совсем недавнем прошлом. Разработки в области искусственного интеллекта породили т. н. «когнитивную революцию». Современные когнитивные исследования невозможны без обращения к моделям искусственного интеллекта, без обсуждения возможностей и границ применимости этих моделей при изучении человека.

## **Глобальная цифровизация как антропологический вызов**

Цифровизация — лозунг нашего времени. О ней говорят политические деятели разных стран, с нею связывают перспективы экономического развития, работу сферы финансов, административного управления, обороноспособность страны. Она интенсивно вторгается в область образования, как это воочию продемонстрировано во время эпидемии коронавируса. Идёт цифровизация биомедицины. Есть серьёзные основания считать, что именно те страны, которые сумеют осуществить глобальную цифровизацию жизни, станут в ближайшем будущем мировыми лидерами.

Но ведь то, что имеется в виду под цифровизацией, не что иное, как новейший этап информационного общества, которое успешно развивается во всём мире уже на протяжении последних 30 лет. Цифровизация — не что иное, как особый (цифровой) способ кодирования информации. И это такой способ, который позволяет колоссально увеличить скорость распространения информации, создаёт принципиально новые возможности для обработки огромных её массивов.

Цифровизация предполагает использование систем искусственного интеллекта. Без неё были бы невозможны персональные компьютеры, смартфоны и айпеды. Без неё не мог бы существовать Интернет.

Без неё не мог бы существовать сам современный человек, который настолько интегрирован в различные информационные потоки, настолько погружён в виртуальную цифровую реальность, что не представляет себе иной жизни.

Сегодня решение разных практических проблем ищется на пути цифровизации, и эти решения, как правило, находятся.

Однако, как это обычно бывает, решение одних проблем порождает новые. При этом многие новые проблемы, возникающие в связи с глобальной цифровизацией, настолько серьёзны, что заставляют думать о том, что мы стоим перед лицом экзистенциального вызова. Некоторые современные мыслители говорят

в этой связи об антропологической революции, другие — об антропологической катастрофе, третьи — о том, что это и первое, и второе одновременно.

Попробую осмыслить эту ситуацию.

Кажется, что на пути цифровизации человечество может осуществить то, о чём люди мечтали на протяжении столетий. Можно иметь свободу в получении информации из источников, разбросанных по всему миру, просчитывать последствия тех или иных экономических и политических действий, избавиться от недостатков существующих политических систем посредством т. н. «электронной демократии» (выражение политического волеизъявления через Интернет и социальные сети, т. е. нечто вроде прямой — в отличие от представительной демократии), получать полноценное образование независимо от места проживания (можно жить в глухой африканской деревушке и слушать лекции лучших профессоров Оксфорда), становится возможным многое другое. Человек как бы вырывается за пределы пространства — места своего проживания и за пределы времени — истории той или иной культурной общности, так как теперь люди сами могут создавать общности по интересам, члены которых обитают в разных регионах мира. Выгоды от цифровизации действительно неоспоримы.

Выясняется, однако, что новые возможности сопровождаются новыми и очень серьёзными проблемами. Так, например, возможность просчитывать результаты тех или иных действий в сфере экономики означает если не исчезновение рынка, то в любом случае радикальную трансформацию рыночных отношений. Цифровизация финансовых операций (блок-чейн) ставит под вопрос необходимость существования банков. Автоматизация всё большего количества видов деятельности влечёт за собою исчезновение огромной массы профессий и создаёт проблему занятости для множества людей. «Электронная демократия» не может отменить необходимость существования профессиональных политиков и политических экспертов, а это значит, что проблемы, связанные с их деятельностью (которая может руководствоваться узко групповыми интересами), остаются. Обучение

онлайн не заменяет личного общения преподавателя и ученика в реальном, а не виртуальном мире: чтение лишь электронных учебников («цифровое чтение») и коммуникация через Интернет не дают полноценного образования.

Это серьёзные проблемы, решение которых связано с важнейшими сторонами человеческой жизни. Их нужно обсуждать специально. Я в данном тексте остановлюсь лишь на некоторых из них, связанных с происходящей трансформацией коренных условий человеческого бытия. А эти условия претерпевают сегодня серьёзную ломку именно в связи с цифровизацией, с внедрением в нашу жизнь искусственных интеллектуальных систем.

Кант считал, что философия должна ответить на три главных вопроса: что я могу знать, что я должен делать, на что я могу надеяться? Ответ на эти вопросы, как он справедливо считал, определяет, что такое человек. Попробую показать, какую форму эти вопросы и возможные ответы на них приобретают в эпоху цифровизации.

Сначала о том, что можно знать в этих условиях.

Адепты и фанатики цифровизации считают, что сегодня «знать» может означать только одно: уметь вычислить. С этой точки зрения полноценно о чём-то знать могут только интеллектуальные технические устройства, умные машины. Именно на такую машину нужно теперь полагаться даже в простейших жизненных ситуациях. Искусственный интеллект будет знать вас лучше, чем вы знаете самого себя: он будет постоянно отслеживать состояние вашего здоровья, будет заказывать необходимые вам лекарства, следить за тем, чтобы вы эти лекарства принимали или даже будет сам вводить их в ваш организм, не считая необходимым ставить вас об этом в известность. «Умный дом» будет знать, что вам нужно иметь у себя дома, например, какие продукты должны быть в холодильнике, он будет сам заказывать эти продукты, рекомендовать вам меню и образ жизни.

Индивидуальная идентичность конструируется с помощью знания о своём прошлом. Как считает один современный философ, «Я — это центр нарративной гравитации», где нарратив — это рассказ себе о собственном прошлом, автобиографическая

память<sup>1</sup>. Но идущая сегодня цифровизация памяти создаёт принципиально новые возможности для вмешательства в человеческую жизнь. Ведь действия человека оставляют цифровые следы, которые доступны внешним инстанциям, способным его контролировать. В некоторых странах передвижения, человеческие контакты, разговоры уже сегодня регистрируются и оцениваются в баллах, в соответствии с которыми предоставляются те или иные жизненные блага или накладываются санкции.

Если ваша опредмеченная память попадает в общее публичное пространство, она становится достоянием других людей, о которых вы не знаете. Они могут обращаться с ней, не спрашивая вашего разрешения и не считаясь с вашими интересами. Сегодня нередко фотографии, помещённые в Инстаграм, используются другими людьми без ведома тех, кому они принадлежат, и даже во вред им. Но если ваша опредмеченная память в виде фотографий, видеороликов или аудиозаписей — это не что иное, как объективированная часть вашей личности, важная составляющая вашей идентичности, то это значит, что другие люди могут манипулировать вами, вашей идентичностью и вашей памятью и посягать на вашу автономию.

Сегодня с помощью цифровых манипуляций легко создавать фотографии таких событий, которых в действительности не было. Создаются особого рода «цифровые памятники» ушедших людей, в которые включаются написанные ими тексты, фотографии, видеоролики, аудиозаписи. Самое интересное в том, что можно, некоторым образом, взаимодействовать с этой оцифрованной памятью и даже вступать в особого рода «переписку» с покойным.

Большинство людей ведут записи о событиях собственной жизни с помощью смартфонов (цифровые фотографии с указанием дат их создания), блогов, участия в социальных сетях. Популярна программа Life Log, которая позволяет не только фиксировать время вашего сна и количество сделанных за день шагов, но также места вашего пребывания и сделанный путь.

---

<sup>1</sup> Dennett D. *Consciousness Explained*. Basil Bay Books. N-Y., Boston. L. 1991. P. 418.

Представим себе, что в недалёком будущем вы будете записывать цифровым способом все события вашей жизни (есть немало энтузиастов этой идеи). Кажется, что это позволит объективно сохранить память о себе и передать её потомкам. Представляется, что таким путём вы можете как бы обессмертить себя, продолжать вечно существовать на цифровом носителе в сознании тех, кто будет жить после вас. Но оказывается, что между памятью, сознательно переживаемой вами, и вашей памятью, оцифрованной и вынесенной во вне, существует серьёзная разница, и это создаёт ряд проблем. Дело в том, что каждое воспоминание о событии вашей жизни не похоже на предыдущие воспоминания о том же событии. Ибо ваша жизнь постоянно меняется, и то, что происходило, всякий раз осмысливается в новом контексте. Но если память о чём-то записана в цифровом виде, то это сделано жёстким способом, и само по себе обращение к памяти не может в ней ничего изменить. Это «неживая» память, она как бы отделяется от вас и выступает как нечто внешнее. Есть ещё одна проблема. Вы оживляете в виде воспоминаний только то, что для вас важно и нужно сейчас. Но оцифрованная память не обладает таким свойством. Она навязывает вам то, что в данный момент вам не нужно, о чём вы не хотели бы сейчас вспоминать, а, возможно, вообще хотели бы забыть. К тому же в этой памяти слишком много всего записано. И это не помогает вам жить, а сковывает вас. У этой памяти есть ещё одна неприятная особенность. Если ваши воспоминания вынесены в глобальное цифровое пространство (существуют в виде файлов в Интернете, записей в блогах, фотографий в Инстаграме), то возникает возможность взламывания вашего цифрового архива и подделки воспоминаний. А если вы привыкли к жизни в цифровом мире и стали больше полагаться не на те воспоминания, которые всплывают в сознании, а на те, которые оцифрованы, то начнёте принимать подделанные воспоминания за то, что происходило с вами на самом деле, т. е. начнёте жить не своей жизнью, а той, которая кем-то искусственно создана за вас<sup>1</sup>.

---

<sup>1</sup> Michaelian K. Is external memory memory? Biological memory and extended mind // *Consciousness and cognition*. 2012. 21, 1154–1165.

Вы перестаёте быть владельцем информации о своей жизни и её хозяином. Ваше личное пространство оказывается как бы взломанным, и вы становитесь предметом управления со стороны других людей. Современный немецкий исследователь М. Майер — Шёнбергер в книге с характерным названием «Удалить: ценность забвения в цифровую эпоху» считает, что впервые за свою историю человечество вступает в такую стадию, когда возникла возможность существования «вечной памяти» без забвения, и это серьёзная угроза индивидуальной автономии и частной жизни, всему человеческому в человеке<sup>1</sup>.

Намерение приобрести с помощью цифровизации невиданную свободу оборачивается тем, что человек перестаёт понимать окружающий мир и самого себя, теряет автономию, попадает в рабство к искусственному интеллекту.

Зато с помощью последнего возможен анализ т. н. «больших данных», что не в силах сделать естественный интеллект. На основе такого анализа делаются успешные предсказания в тех областях жизни, которые до недавних пор были полны неопределённости: сферы бизнеса, администрирования, здравоохранения. Обработка «больших данных» сегодня используется в ряде естественных, социальных и гуманитарных наук: биологии, социологии, лингвистике, литературоведении, истории искусства и др. Появился даже термин “Digital Humanities” — «цифровые гуманитарные науки»<sup>2</sup>. Некоторые философы стали писать о том, что наступил «конец теории», ибо вместо выдвижения теоретических гипотез, которые должны определённым путём сопоставляться с экспериментальными фактами (что непросто и в интеллектуальном, и в финансовом отношении) можно прибегнуть к машинной обработке огромного массива данных и выявить корреляции, существующие между разными данными<sup>3</sup>. В этой связи

---

<sup>1</sup> Meyer-Schonberger V. Delete: the virtue of forgetting in the digital age. Princeton: Princeton University Press, 2011.

<sup>2</sup> Цифровая культура // Международный журнал исследования культуры. 2012. СПб.

<sup>3</sup> Anderson S. The End of Theory // Wired. June 2008.

популярен тезис о том, что большие данные «говорят сами за себя», что в философской теории познания должен произойти новый возврат к эмпиризму на основе понимания познания как вычислительной обработки больших массивов информации.

Но что это значит с точки зрения познания и знания?

В этой связи я хочу обратить внимание на два обстоятельства.

Первое. Научную теорию невозможно непосредственно вывести из эмпирических данных, сколь не было бы велико их количество. Это связано с тем, что теория строится на основе некоторых идеализаций и идеального экспериментирования. Так, например, одно из основоположений классической механики, возникновение которой означало начало современной науки, гласит, что любое тело, на которое не оказывается никакое внешнее воздействие, либо покоится, либо движется бесконечно, прямолинейно и равномерно. Очевидно, что ничего подобного в опыте в принципе не может наблюдаться. Однако же именно на основе этой и некоторых других идеализаций было построено здание классической механики, с помощью которой стало возможно не только понять опытные данные, но и делать успешные предсказания и создавать новые технические устройства. Если бы в XVII и XVIII вв. существовали способы цифровой обработки массива «больших данных», и они использовались бы вместо построения теорий, ясно, что никакой науки в нашем смысле слова мы не имели бы.

Правда, как я отметил, некоторые философы считают, что возникновение способов машинной обработки «больших данных» как раз и означает «конец теории». Согласно этой точке зрения научная теория отныне не нужна: искусственный интеллект может хорошо делать предсказания, обходясь без всяких таких построений.

Но в этой связи я хочу обратить внимание на второе обстоятельство для понимания того, каким становится познание и знание в такой ситуации.

Дело в том, что даже если можно делать хорошие прогнозы на основе обработки «больших данных» с помощью искусственного интеллекта (а это действительно делается), то для человека такой прогноз часто является неожиданным. В отличие от того, что имело



место в отношении предшествующих технических устройств, поведение систем, работающих на базе искусственного интеллекта, неопределенно и непредсказуемо для человека. Цифровая технология выходит из-под его контроля. Прогнозы, производимые на основе анализа «больших данных» и основанные на обнаружении корреляций между разными группами событий, как правило, не выявляют причинных зависимостей, не содержат объяснений и поэтому не помогают понять того, что происходит и почему нечто ожидается<sup>1</sup>. До сих пор возможность справиться с неопределённостью предполагала получение контроля над ситуацией, а это, в свою очередь, означало понимание существующих причинных зависимостей. Именно на основе понимания таких зависимостей можно создавать технические устройства, которыми человек может управлять и с помощью которых он способен контролировать результаты своих действий. А в случае цифровой обработки «больших данных» с целью предсказаний человек оказывается в положении слуги, который должен выполнять идущие от хозяина предписания, смысла которых он не понимает (правда, их смысла не понимает и сам искусственный интеллект). Для человека жизнь не становится более определённой, но положение автономного деятеля он утрачивает, а жизнь его обесмысливается.

Теперь о том, что я должен делать в эпоху глобальной цифровой революции.

Очевидно, что в таких условиях, когда вы теряете индивидуальную автономию, не вы будете решать, что нужно делать: именно умные технические устройства будут предписывать вам эти решения. Хочу обратить внимание ещё на некоторые связанные с этим обстоятельства.

На основе анализа «больших данных», собранных о поведении некоторого человека, умная машина может предсказать большую вероятность совершения им противоправного поступка, а может быть, даже преступления. Возникает большой соблазн

---

<sup>1</sup> Мейер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живём, работаем и мыслим. М.: Изд. «Манн, Иванов и Фербер», 2014.

использовать эти предсказания для предотвращения нежелательных (а, может быть, даже весьма опасных) событий. В интересах общественной безопасности определённые инстанции могут решить, что не нужно дожидаться, когда этот человек совершит опасный поступок: нужно помешать ему сделать это, т. е. арестовать его. Но ведь подобные действия подрывают тот принцип, который лежит в основе всей нашей культуры и всей правовой системы: человек отвечает только за те действия, которые он совершил, а не за то, что он мог бы сделать, так как обладает свободой воли, т. е. в любой ситуации может поступить так или по-другому. Отказ от принципа свободы воли ломает все представления об индивидуальной самостоятельности и ответственности. Нужно, однако, иметь в виду, что рост рискованных ситуаций и вместе с тем возможность с помощью анализа «больших данных» прогнозировать поведение не только больших масс людей, но и отдельных индивидов, может соблазнить контролирующие инстанции: толкнуть их к тому, чтобы пожертвовать принципом свободы воли и тем самым отказаться от одного из коренных условий человеческого бытия.

Наконец, на что же можно надеяться в этой ситуации.

Некоторые адепты глобальной цифровизации предлагают такую надежду. Они исходят из того, что человеческая личность — это не что иное, как некоторый массив информации, записанный в мозгу с помощью цифровых кодов. В недалёком будущем, считают они, будет можно расшифровать эту информацию и, если ваше тело состарилось, то информация о вас будет перенесена на другое тело, более молодое, а когда состарится и это тело, подобную операцию можно будет повторить. И так далее до бесконечности. Есть и более радикальное предложение: переносить информацию о вас не на человеческое тело, а на цифровой носитель, который практически не стареет. Это значит, что вы будете жить вечно. Тем самым, говорят сторонники этой идеи, будет осуществлена всегда владевшая людьми мечта о бессмертии. Это будет «цифровое бессмертие». Сегодня в мире существует немало людей (физики, нейроучёные, психологи, философы, футурологи), которые не только верят в подобную возможность,

но и пытаются вести практическую работу в этом направлении. На сайтах футурологов даже указывается, когда примерно осуществится «цифровое бессмертие» — считается, что это произойдёт уже в этом столетии.

Между тем эта идея сомнительна и опасна.

Попробую показать, почему.

Начну с того, что даже если считать человеческую личность массивом информации, то нужно иметь в виду, что важнейшей частью этой информации является представление человека о самом себе — то, что называется «образом Я» (или «Я-концепцией»). А в этот образ входит представление о собственном теле. Конечно, тело каждого человека меняется — стареет, подвергается увечьям, возможно даже, что некоторые его части заменяются протезами. Но в любом случае представление о собственном теле и о непрерывном ряде его изменений — неотъемлемая часть «образа Я». Ведь именно с этим телом, которое находилось в те или иные моменты времени в том или ином месте и с помощью которого человек вступал в контакты с другими, связана его жизнь, его автобиографическая память, являющаяся основой индивидуальной идентичности. Если бы даже удалась пересадка индивидуальной психики на другой носитель (хотя это вряд ли осуществимо), то идентичность не удалось бы сохранить: возник бы другой человек (в известном рассказе Кафки «Превращение» описываются драматические переживания героя, который однажды обнаружил, что его «Я» пересажено в тело какого-то другого существа).

Но дело не только в этом. Если бы человек стал бессмертным, он перестал бы быть человеком. Ибо в таком обществе потеряли бы смысл важнейшие ценности, делающие человека человеком: забота о близких, самопожертвование, мужество, страдание. Утрата смысла смерти ведёт к утрате смысла жизни. Если бы «цифровое бессмертие» было возможно, нас ожидало бы не райское блаженство, а муки ада <sup>1</sup>.

---

<sup>1</sup> Лекторский В.А. Возможно ли постчеловеческое будущее? // Философия, познание, культура. М.: Канон-плюс, 2012.

И — самое главное — не следует забывать, что даже если бы удалось осуществить «цифровое бессмертие», как об этом мечтают некоторые футурологи, в ситуации глобальной цифровизации жизни человек в любом случае оказался бы управляемым и контролируемым сверхмощным искусственным интеллектом. Если бы в этом случае у него осталось что-то человеческое, вряд ли он захотел бы такого бессмертия.

Выходит, что в этих условиях надеяться не на что.

В этой связи я хочу обратить внимание на ещё одно важное обстоятельство.

Дело в том, что цифровизация и искусственный интеллект нередко понимаются как способ справиться с риском неопределённости — а последняя сегодня растёт, как об этом ещё раз свидетельствует вспышка эпидемии коронавируса. Кажется, что с помощью технических устройств можно в принципе контролировать протекание всех процессов в природе и обществе, влиять на них и направлять в желательное для человека русло. Но, как ясно сегодня, контролировать можно отнюдь не все процессы и не во всех случаях. Когда имеются сложно организованные и развивающиеся системы — а это большинство природных и социальных систем, то в тех ситуациях, когда последние близки к т. н. точке бифуркации, нельзя точно предсказать их поведение: можно только строить несколько сценариев их будущего, при этом степень вероятности того или иного сценария меняется по ходу дела. Но это значит, что сколь бы ни была развита вычислительная техника и сколь угодно большими ни были бы возможности искусственных интеллектуальных систем, нельзя в принципе преодолеть неопределённость в предсказании будущих событий. А это значит, что никакая технология не создаёт возможность управления процессами некоего рода.

Наш известный учёный и философ Н.Н. Моисеев специально выделял процессы, которыми можно управлять, и такие, которые можно только направлять, т. е. создавать условия, способные влиять на ход этих процессов в сторону, благоприятную

для человека — при этом никогда не будет полной уверенности в том, что процесс пойдет именно в желаемом направлении<sup>1</sup>.

Цифровизация не помогает уйти от неопределённости.

Это не может делать никакой сверхмощный искусственный интеллект. Неопределённость наряду с определённой заложена в самой структуре мироздания. И это неплохо, ибо если бы всё было жёстко определено, невозможно было бы то, что человек считает наиболее ценным: творчество, диалог, любовь, моральный поступок.

Сегодня человечество стоит перед лицом антропологических вызовов. Многие из них связаны с глобальной цифровизацией на основе использования систем искусственного интеллекта. Цифровизация приносит немало благ, решает множество проблем, но также может создавать угрозы самому бытию человека. А это значит, что предлагаемые в рамках цифровизации проекты нуждаются в серьёзной философской экспертизе, направленной на гуманизацию информационных технологий и на культивирование высших человеческих ценностей: свободы, личной автономии, достоинства, идентичности, творчества, понимания, взаимопонимания<sup>2</sup>. Философия бывает максимально востребована в периоды культурных и познавательных кризисов. Мы переживаем именно такое время.

---

<sup>1</sup> Моисеев Н.Н. Судьба цивилизации. Путь разума. М.: Языки русской культуры, 2000.

<sup>2</sup> Фролов И.Т. Прогресс науки и будущее человека. М.: Политиздат, 1975; Лекторский В.А. Человек и культура. СПб.: Изд-во Санкт-Петербургского университета профсоюзов, 2018.

---

---

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ОБЩЕСТВО

---

---

Рассматриваются проблематика искусственного интеллекта, эволюция разработок и их интеграция на базе логических и нейросетевых технологий. Обсуждаются влияние искусственного интеллекта и цифровых технологий на общество и их значение для экономики, экологии и противодействия природным, техногенным и другим угрозам. Акцентируется роль искусственного интеллекта и роботизации в реиндустриализации восточных и других регионов страны и значение цифровых технологий для перехода цивилизации к ноосферному состоянию с приоритетами науки и информации, как ресурсов развития, над энергией и веществом и с главенством духовно-нравственных ценностей.

**Введение.** Настоящее и будущее «цифровизации» среды жизнедеятельности человека сегодня связываются с IT-технологиями в области искусственного интеллекта (ИИ), с проектами телемедицины, умного дома-города-общества и т. п. Пандемия с коронавирусом и нарастание инфодемии обратили мир в новую реальность, обнажив слабые и сильные стороны государств, их конституционных строев, социальных, научно-технологических, экономических и экологических моделей развития. Пандемия Covid-19 — первая глобальная эпидемия в эпоху социальных

---

\* **Васильев Станислав Николаевич** — академик РАН, Институт проблем управления им. В.А. Трапезникова РАН, г.н.с., заместитель председателя Научного совета при Президиуме РАН по методологии искусственного интеллекта и когнитивных исследований, Москва.

**Vassilyev S.N.** — academician of the Russian Academy of Sciences, Trapeznikov Institute of Control Sciences of RAS, chief researcher.

сетей, породившая не только инфодемию, но и широкое использование цифровых средств, в т. ч. при вынужденной дистанционной работе человека.

Государства должны опереться на платформы цифрового мониторинга, моделирования и управления социумом, его средой обитания и экономикой на разных уровнях власти и государственного управления для ситуационного анализа обстановки и динамики ее развития, обеспечивая качество жизни и здоровье, сохраняя приватность частной жизни и не подменяя обработку данных манипулятивным социальным контролем. Государство должно обеспечить цифровую безопасность человека и содействовать обществу и его выборным органам в разработке инструментов контроля медиасредств по духовно-нравственным критериям оценки их влияния на человека.

Между тем, продолжают разноречивые оценки успешности борьбы Китая с Covid-19: «образ т. н. “развития без демократии” потускнел» (читай — не справились) или «цифровой диктатуре легче» (справились, но ценой нарушения приватности частной жизни). К тому же борьба за лидерство в мире обострилась тем, что в 2017 г. Китай объявил о программе достижения глобального лидерства в области ИИ.

При всеобщей перегрузке СМИ вчера темой пандемии, а сегодня объективной и фейковой информацией о событиях на Украине, тематика IT-средств и отношений человека с цифровыми платформами и средствами ИИ неизменно находятся в центре внимания

## **Искусственный интеллект и интеллектуальные системы**

Проблемы создания партнерской атмосферы в развивающемся роботизируемом и цифровом мире, проблемы единения человечества на максимально гуманизированной и экологизированной основе при обострении национальных проблем цифрового суверенитета и роботизации производительных и вооруженных

сил — новый вызов устойчивому развитию человечества, хотя об ИИ и его перспективности широко заговорили еще в 50-е годы прошлого века.

Тогда же стал использоваться и сам термин «ИИ», обозначающий именно *область исследований и IT-разработок*. Первые достижения по автоматизации решения математических, игровых и других задач логическими, эвристическими и другими методами окрылили авторов, но затем энтузиазм несколько ослаб, поскольку явно не доставало производительности компьютеров, да и цифровизация данных с бумажных носителей требовала времени.

А через два десятилетия — новый всплеск интереса к ИИ: создавались прикладные экспертные системы на логических правилах и системы нейросетевой обработки данных с усиленными методами машинного обучения, хотя сыграло свою роль и усиление производительности тогдашних ЭВМ.

В настоящее время рост интереса и бурное развитие ИИ можно объяснить следующими факторами. Прежде всего, возможностями<sup>1</sup> существенного сокращения ручного труда и повышения прибыльности бизнеса на основе внедрения цифровых технологий с интеллектуализацией; масштабом накопленных и используемых цифровых данных и развитием коммуникационных средств с уменьшением возрастной асимметричности пользователей, что особенно проявилось в условиях пандемии; предпочтением удаленной работы, высказываемым как сотрудниками, так и некоторыми их руководителями, отмечающими повышение производительности их сотрудников на «удаленке»; улучшением характеристик компьютеров — их производительности, объема памяти, мобильности, энергоэкономичности; расширением дистанционно предоставляемых сервисов. Это не только информационные и вычислительные, но и аппаратные сервисы, благодаря доступности автономной (беспилотной) и другой удаленной техники, опробуемой не только для информационных целей мони-

---

<sup>1</sup> Васильев С.С. Задача озадачиться // ЕГНС «Поиск». 2019. № 50 (1592). С. 10–11; [www.poisknews.ru](http://www.poisknews.ru)



торинга, но и в киберфизических задачах доставки материалов и предметов на расстояние и пр.

Функциональная нацеленность ИИ разнообразна. Это задачи автоматического и автоматизированного приема, передачи, накопления, обработки информации и управления. В числе этих задач — распознавание видеообразов и речи, анализ трехмерных сцен, автоматизация рассуждений, межкузыковий перевод естественно-языковых и других текстов и речи, машинное обучение, выявление закономерностей, визуализация и анализ больших данных, семантические модели и вычисления, управление в сетях и сетями, облачные вычисления, интернет вещей и многое другое.

Перечисленные IT-разработки хотя и претендуют называться интеллектуальными, правильнее называть их *интеллектными* (*intelligent*)<sup>1</sup>, т. е. использующими наработки ИИ как области исследований. Автономные интеллектные системы (АИС), с их реактивными и другими регуляторными механизмами поведения и способностями решать сложные неалгоритмизированные задачи, не обладают способностями человека понимать мир, рефлексировать, действовать адекватно обстановке, собственным интересам, знаниям, воспитанию и эмоциональному настрою (эмоциональным мотивам). *Интеллектуальными* (*intellectual*) будут называться будущие разработки со способностями не только автоматического достижения цели, но и самоозадачивания, т. е. обладающие, помимо механизмов целенаправленного поведения, также и механизмами мотивации потребностей и целеполагания.

АИС будущего, развив способности адекватного (рационального) поведения с целеполаганием, освободятся от их априорной сегодняшней нацеленности лишь на те или иные фиксированные классы задач, с малыми вариациями постановок задач пусть даже

---

<sup>1</sup> Задача озадачиться // ЕГНС «Поиск». 2019. № 50(1592). С. 10–11; [www.poisknews.ru](http://www.poisknews.ru); Åström K.J., McAvoy T.J. Intelligent Control: An Overview and Evaluation // Handbook of Intelligent Control. Neural, Fuzzy, and Adaptive Approaches / Ed. by D.A. White, D.A. Sofge. New York: Van Nostrand Reinhold, 1992. P. 3–34; Васильев С.Н., Жерлов А.К., Федосов Е.А., Федунев Б.Е. Интеллектуальное управление динамическими системами. М.: Физматлит, 2000. 352 с.

из разных предметных областей. Что касается промышленных, медицинских и других роботов, то это пока в основном роботы-манипуляторы или подвижные роботы с более или менее контролируемой последовательностью не только целей, но и действий.

Так что в ИИ — огромное поле интереснейших задач, в т. ч. по нейросимвольному моделированию адаптивного поведения агентов (живых организмов и АИС), их мотиваций, целеполагания и планирования индивидуальных и групповых действий агентов, эволюционного подавления внутривидовой и межвидовой конкуренции и усиления кооперативности.

Сегодня же к интеллектуальным можно отнести человеко-машинные интеллектуальные системы с целеполаганием человека и встроенными в их состав АИС. Применительно к таким человеко-машинным системам Президент В. Путин употребил термин «партнерские системы». На Международном форуме “*Artificial Intelligence Journey*” (Москва, 8–9.11.2019 г.) он акцентировал также внимание на необходимости создания «свода этических правил» взаимодействия между людьми и продуктами ИИ, что важно для роботизируемого и цифрового мира. При должном решении проблем этики и доверительности ИИ именно технологии ИИ могут и должны помочь гасить фейковость СМИ и инфодемию в социальных сетях, как явления в гуманитарно-правовой и социально-политической сферах, но возникшие на технологической основе.

### **Об интеграции средств искусственного интеллекта**

Проблемы интеграции разработок в области ИИ стоят давно<sup>1</sup> и работы по их решению ведутся; правда, как отмечается, в области ИИ<sup>2</sup> по-прежнему еще имеет место «фрагментарность

---

<sup>1</sup> Свириденко Д.И., Витяев Е.Е. Задачный подход к искусственному интеллекту и его теоретическая и технологическая база // Сб. трудов XVIII Национальной конференции по искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 36–44.

<sup>2</sup> См. там же.

исследований с претензией на самостоятельность отдельных направлений (машинное зрение, перевод и понимание естественных языков, робототехника и другие)».

Среди ранних интегративных работ в части приложений средств ИИ в области управления в реальном времени можно назвать гибридную систему управления марсоходом<sup>1</sup> (the Marsokhod planetary rover, NASA Research Center). В человеческом симбиозе образного и вербального мышления образы чувственного восприятия — необходимый фрагмент картины мира. По частичной аналогии с этим, в марсоходе интегрированы нейрораспознавание образов и продукционное управление на логико-лингвистических правилах “if-then”, а ниже на исполнительном уровне — робастное управление (по принципу обратной связи и при параметрических возмущениях). При этом нейросеть обеспечивает уточнение настройки на распознаваемые условия и обстоятельства планеты, априори неизвестные с требуемой точностью.

В человеко-машинных интеллектуальных системах (ЧМИС) человек — не просто пользователь, а партнер, формирующий постановку задачи, ее цели и критерии качества решения. Так, в ЧМИС управления он ответственен по крайней мере за верхний, интеллектуальный, уровень многоуровневого управления. Ниже — уровень АИС управления, в частности автоматического планирования действий по решению поставленной задачи, и еще ниже уровни исполнительного управления (может быть, предиктивного, адаптивного, робастного управления, управления с обратной связью, программного управления)<sup>2</sup>.

Выходом АИС, т. е. уровня интеллектуального управления (часто в логико-лингвистической форме), является план задействования разных внутренних (бортовых) и внешних средств достижения

---

<sup>1</sup> *Handelman D.A., Stengel R.F.* An Architecture for Real-time Rule-based Control // Proc. American Control Conference. 1987. P. 1636–1642.

<sup>2</sup> *Васильев С.Н., Кудинов Ю.И., Пащенко Ф.Ф. и др.* Intelligent Control Systems and Fuzzy Controllers, 1. Fuzzy Models, Logical-Linguistic and Analytical Regulators // J. Automation and Remote Control. 2020. Vol. 81. № 1. P. 171–191.

цели управления, образующий дерево «краевых задач» нижних уровней управления, например ПИД-регулятора. Помимо задач, не критичных к быстродействию, например в обучении, АИС и ЧМИС (далее — ИС) могут функционировать в режимах жесткого и мягкого реального времени, например в пилотируемых летательных аппаратах и обитаемых подводных объектах, в транспортных средствах, ситуационных центрах<sup>1</sup>.

Известные аварии и катастрофы на транспорте, в промышленности, энергетике и др. (Тримайлайлэнд, Бхопал, Чернобыль, Фукусима и др.) связываются с человеческим фактором (перегрузкой операторов, их халатностью) и/или с недостаточным качеством проектирования систем, возникновением нештатных ситуаций неуправляемости, что дополнительно повышает требования к повышению проектного потенциала ИС, в т. ч. интеграцией функциональных возможностей разных ИС.

Разработки гражданского и военного назначения, в т. ч. в стратегически значимой для страны сфере управления автономными (беспилотными) мобильными объектами космической, авиационной, морской, наземной и многосредовой природы, критически нуждаются в АИС для обработки сигналов, данных и знаний<sup>2</sup>. Таковы, в частности, задачи группового преследования и маршрутизации в конфликтных средах, которые в силу своей сложности, помимо классических средств, нуждаются в ИС. При этом традиционные вычислительные средства (моделирования, анализа, оптимизации и др.) требуется комплексировать с ИС.

---

<sup>1</sup> Федунов Б.Е. Интеллектуальные системы тактического уровня: состав, базы знаний // Сб. трудов XVIII Национальной конференции по искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 87–94.

<sup>2</sup> См., напр.: Aksaray D., Vasile C.-I., Belta C. Dynamic Routing of Energy-Aware Vehicles with Temporal Logic Constraints // Proc. International Conference ICRA on Robotics and Automation. Stockholm, 2016. P. 3141–3146; Galyaev A.A. On the Detection Functional in Motion of an Object in a Threat Environment // J. Automation and Remote Control. 2011. Vol. 71. № 4. P. 634–639; Leahy K., Zhou D., Vasile C.I., Oikonomopoulos K., Schwager M., Belta C. Provably Correct Persistent Surveillance for Unmanned Aerial Vehicles Subject to Charging Constraints // Experimental Robotics. 2016. P. 605–619; Васильев С.Н., Галеев А.А. Логико-оптимизационный подход в задачах преследования группы целей // ДАН. 2017. Т. 474. № 6. С. 1–7; Zabaranin M., Uryasev S., Murphey R. Aircraft Routing under the Risk of Detection // J. Naval Research Logistics. 2006. Vol. 5. № 8. P. 728–747.

Наиболее активно развиваются сейчас направления с широким применением в малом и крупном бизнесе: разрабатываемые АИС включают программы (софтботы в сетях и пр.), беспилотный транспорт (складской, пассажирский), роботы-манипуляторы в сборочном и других производствах. В проблематике т. н. интеллектуального анализа больших данных ЧМИС являются основным средством генерации нового знания в различных приложениях от медицины до наук о человеке и обществе.

В ЧМИС оперативно-советующего типа возможны случаи отказа человека (пользователя, оператора) использовать рекомендации, выработанные интеллектуальными средствами, входящими в ЧМИС. Случаи отказа, даже при способности ИС как-то объяснить и обосновать свою рекомендацию, целесообразно фиксировать и аккумулировать у разработчика ИС с объяснениями человека о причинах его отказа<sup>1</sup>. Отказы и объяснения при их обоснованности полезны для совершенствования ИС и свидетельствуют о квалификации человека, служа, может быть, основанием для его поощрения, а отсутствие отказа от неверной рекомендации может означать необходимость проверки квалификации и ответственности оператора. Возможны случаи и эмоционального отказа человека от верной или неверной рекомендации системы, который человеку не удастся или соответственно будет трудно объяснить логически.

В эпоху цифровизации сказанное о ЧМИС оперативно-советующего типа и других ИС важно во многих приложениях ИС, в т. ч. при разработке и использовании ИС в медицинской практике<sup>2</sup>.

---

<sup>1</sup> Михееenkova М.А., Финн В.К. Интеллектуальный анализ данных и его реализация в партнерских человеко-машинных системах // Материалы XII Мультиконференции по проблемам управления (МКПУ-2019). Ростов н/Д.; Таганрог: Изд.-во ЮФУ, 2019. Т. 1. С. 113–115; Федунев Б.Е. Интеллектуальные системы тактического уровня: состав, базы знаний // Сб. трудов XVIII Национальной конференции по искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 87–94.

<sup>2</sup> Kobrinskii B.A. Images in Logical-and-Linguistic Artificial Intelligence // J. Bio-medical Engineering and Medical Imaging. 2019. Vol. 6. № 1. P. 1–8.

В следующей таблице приводятся достоинства и недостатки некоторых средств ИИ.

*Таблица*

### Сравнение некоторых средств ИИ

Средства ИИ	Типовые достоинства	Типовые недостатки
<b>I. Нейросетевые</b>	<ol style="list-style-type: none"> <li>1. Применимы в многофакторных проблемах с плохой формализуемостью закономерностей</li> <li>2. Высокая степень распараллеливаемости, высокое быстродействие</li> <li>3. Способность к обучению</li> </ol>	<ol style="list-style-type: none"> <li>1. Необходимость обучающей информации — представительного набора примеров «вход-выход». («скорее — глаз, чем — мозг»)</li> <li>2. Обычно отсутствуют элементы символично-логического мышления</li> </ol>
<b>II. Эволюционные (генетические)</b>	<ol style="list-style-type: none"> <li>1. Высокая степень распараллеливаемости и быстродействия</li> <li>2. Аналогии с природной эволюцией</li> </ol>	<ol style="list-style-type: none"> <li>1. Априорная неизвестность эффективности в приложении</li> <li>2. «Скорее самоорганизация природной стихии, чем творческий процесс»</li> </ol>
<b>III. Продукционные (логико-лингвистические правила «if-then», в т. ч. с рефлексией и нечеткие)</b>	<ol style="list-style-type: none"> <li>1. Возможность представления дескриптивно-конструктивных знаний и рефлексии</li> <li>2. Естественность правил</li> </ol>	<ol style="list-style-type: none"> <li>1. Сложность исполнения больших баз продукций, недостаточная структурируемость</li> <li>2. Сложность обеспечения корректности в них</li> </ol>
<b>IV. Объектно-ориентированные (семантические сети, фреймы и т. д.)</b>	<ol style="list-style-type: none"> <li>1. Хорошая структурируемость</li> <li>2. Высокое быстродействие механизмов наследования свойств, умолчания и др.</li> </ol>	<ol style="list-style-type: none"> <li>1. Сложность программирования</li> <li>2. Ограниченность выразительных возможностей</li> </ol>
<b>V. Логические (символьные и семантические методы, АДТ, логическое программирование)</b>	<ol style="list-style-type: none"> <li>1. Высокая выразительность языка</li> <li>2. Корректность</li> <li>3. Высокая сложность решаемых офлайн-задач</li> </ol>	<ol style="list-style-type: none"> <li>1. Недостаточное быстродействие</li> <li>2. Неразрешимость богатых логик</li> <li>3. Недостаточность одной логики</li> </ol>

Несколько комментариев к таблице. Начнем с эволюционных (генетических) средств ИИ, поскольку они ближе других АИС к когнитивистике, а с другой стороны, интуитивно понятны как средства, подобные природному механизму развития популяций (с внутривидовой и межвидовой конкуренцией и кооперацией). Трудно назвать это творчеством (проявлением интеллекта), но разве этот эволюционный механизм не причастен к возникновению удивительной самоорганизации муравейника. Как пример, в работе Редько В.Г.<sup>1</sup> успешно моделируется эволюция мышления человека и познавательных способностей биоорганизмов, а в контексте движения к устойчивому развитию предлагаются некоторые механизмы устранения излишней конкуренции и противостояния. Есть примеры эволюционного синтеза оптимального управления динамическими системами<sup>2</sup>.

По своей типичной функциональности, как отмечено в таблице, нейросетевые средства ИИ напоминают скорее, например, глаз, чем мозг. Опытный врач поставит диагноз, даже если симптоматика заболевания искажена и ему не приходилось ранее ее наблюдать или абстрактно осмысливать, и механизм диагностики врача при этом напоминает скорее комбинацию распознавания на подсознательном уровне с рассуждениями на вербально-логическом уровне, не исключая и эмоциональные мотивы (факторы). Эффективное сочетание параллельных систем образного и вербально-логического мышления и эмоционального фактора — преимущество мозга перед ИС, несмотря на достигнутую и ожидаемую производительность суперкомпьютеров<sup>3</sup>.

В традиционных же нейросетях символично-логическая обработка знаний отсутствует, а рефлекторный механизм преобразо-

---

<sup>1</sup> Редько В.Г. Моделирование когнитивной эволюции: на пути к теории эволюционного происхождения мышления. 2-е изд. М.: ЛЕНАНД, 2019. 264 с.

<sup>2</sup> Diveev A.I., Balandina G.I., Konstantinov S.V. Binary Variational Genetic Programming for the Problem of Synthesis of Control System // Proc. 13<sup>th</sup> International Conference on Neural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSCD-2017). 2017. P. 165–170.

<sup>3</sup> Каляев И.А. Камо грядеши? // Экономические стратегии. М.: ИЭС, 2019. Т. 21. № 5 (163). С. 6–15.

вания входных сигналов в выходные, построенный на примерах прошлого опыта, обеспечивает нейросети в типичной проблематике распознавания и классификации целостность восприятия, повышенное быстроедействие и высокую реактивность по сравнению с логическими механизмами обработки знаний. Работы по обновлению нейросетевых моделей мозга с учетом новых данных нейробиологии весьма актуальны для выявления когнитивных механизмов<sup>1</sup>.

Что касается продукционных систем, то их выразительность по сравнению, например, с логическими средствами ограничена и это сужает класс задач, решаемых продукционной, обычно экспертной, системой. Когда количество правил большое, это может приводить к потере целенаправленности функционирования и ненадежности. В случае, когда набор правил рефлексирован, т. е. в правых частях некоторых из них фигурируют действия по изменению самого набора правил, большая ответственность возлагается на разработчиков по верификации и обеспечению корректности функционирования продукционной системы.

Нужны и развиваются более мощные системы «обдумывания ситуации», способные обрабатывать знания более общего вида вместо простого использования правил, заранее заготовленных или модифицируемых в процессе функционирования мало надежным способом. Использование объектно-ориентированных экспертных систем (на семантических сетях и фреймах) упрощает структурирование знаний, но усложняет программирование экспертной системы.

Системы автоматического доказательства теорем (АДТ), как и системы логического программирования (ЛП) в разных его парадигмах, работают обычно в подмножествах первопорядковой логики (ППЛ) или ее модификаций. Выразительность языка ППЛ повышает совместимость АИС с традиционными средствами многоуровневого управления (другие агенты, сенсоры, средства

---

<sup>1</sup> Кузнецов О.П., Базенков Н.И., Болдышев Б.А. и др. Асинхронная дискретная модель химических взаимодействий в простых нейронных системах // Искусственный интеллект и принятие решений. 2018. № 2. С. 3–20.



связи, исполнительные органы). Предикатные языки позволяют формализовать более широкие знания для последующей обработки машинно-ориентированными логическими правилами вывода. Каждый шаг вывода формализует не столько специальный (локальный) переход от условий к действию (как при использовании продукционных знаний), выражающий более или менее очевидным образом отдельный шаг приближения к цели управления, сколько представление человека о правильности (логичности) умозаключений, т. е. имеет более универсальный характер и более широкие возможности решения задач управления, сводимых к АДТ, хотя, возможно, и с увеличением требований к ресурсам выработки управления.

Компьютерные программы АДТ суть автономные или человеко-машинные усилители способностей человека. В их приложениях в математике создаются программы-ассистенты в помощь математикам<sup>1</sup> и АИС-пруверы, существенно превосходящие сегодня другие средства ИИ с точки зрения сложности теорем, доказываемых ими. Решены некоторые открытые задачи, поставленные известными математиками<sup>2</sup>, причем планка сложности неуклонно повышается. Небезосновательно высказываются мнения о том, что примеры математических теорем, доказательства которых будет трудно проверить даже профильным специалистам и которые будут верифицироваться средствами АДТ, будут множиться.

С другой стороны, эти достижения в области систем АДТ принадлежат сфере офлайн-задач без характерных ресурсных ограничений: это — сфера математических задач, программирования и вообще автоматизации строгих рассуждений, допускающих формализацию в некотором логическом языке, когда нет особой критичности в вопросе о ресурсах, отпущенных на

---

<sup>1</sup> Urban J., Vyskocil J. Theorem Proving in Large Formal Mathematics as an Emerging AI Field / arXiv:1209.3914v2(cs.AI) 16Dec2012

<sup>2</sup> Воз Л. Решение некоторых открытых проблем с помощью программы для автоматического доказательства теорем // Киберн. сб. Нов. сер. Вып. 21. М.: Мир, 1984. С. 235–263.

проведение автоматических выкладок. Проникновение полных первопорядковых логик и методов АДТ в сферу интересов и компетентности специалистов по теории управления — важный фактор развития ИС<sup>1</sup>.

За рамками компьютерной парадигмы развития ИИ — объединение области ИИ с когнитивной наукой и кибернетикой как наукой «управления в животном и машине» (по Н. Винеру). Пока же, например, коннекционизм (нейросетевизм), как и другие направления ИИ, ограничен феноменологическим подходом, а отчужденность ИИ от нейрофизиологии и других когнитивных наук, порой «горделивая», тормозит его развитие<sup>2</sup>, хотя в то же время известно много близких концепций комплексирования разработок в основном внутри ИИ, в т. ч. концепции создания «сильного ИИ рациональных агентов»<sup>3</sup>, «общего ИИ»<sup>4</sup>, «дружественного ИИ»<sup>5</sup> и других.

Внутри ИИ развивается «нейросимволизм» на пути интеграции коннекционизма (нейросетевизма) с символизмом. В частности, реализация логического (дедуктивного) вывода на нейросети<sup>6</sup> открывает

---

<sup>1</sup> Васильев С.Н., Жерлов А.К., Федосов Е.А., Федунев Б.Е. Интеллектуальное управление динамическими системами. М.: Физматлит, 2000. 352 с.

<sup>2</sup> Хокинс Дж. Об интеллекте. М.: Вильямс, 2007. 240 с.; Кузнецов О.П. Искусственный интеллект и когнитивные науки // Информационно-измерительные и информационные системы. 2013. Т. 11. № 5. С. 16–24.

<sup>3</sup> Рассел С., Норвиг П. Искусственный интеллект: современный подход. 2-е изд., перераб. М.: Вильямс, 2006. 1408 с.

<sup>4</sup> Goertzel B. Toward a Formal Characterization of Real-World General Intelligence // J. Advances in Intelligent Systems Research. 2010. Vol. 10. P. 1–6; Besold T.R., Schmid U. Why Generality Is Key to Human-Level Artificial Intelligence // J. Advances in Cognitive Systems. 2016. Vol. 4. P. 13–24.

<sup>5</sup> Udrowsky E. Complex Value Systems in Friendly Artificial Intelligence // Proc. 4<sup>th</sup> International Conference on Artificial General Intelligence (AGI-2011). Lecture Notes in Computer Science 6830. Springer, 2011. P. 388–393.

<sup>6</sup> См., напр.: Komendantskaya E. First-order Deduction in Neural Networks // Proc. 1st International Conference on Language and Automata Theory and Applications (LATA-2007). Tarragona. 2007. P. 307–319.

возможности обучения ИС стратегиям комбинирования правил вывода, что расширяет набор решаемых ею задач.

Работами по нейросимволизму являются работы, в которых анализ изображений ведется совместно с анализом связанных с ними естественно-языковых текстов в процессе обучения нейросети распознаванию образов и ЕЯ-текстов<sup>1</sup>.

Методология интеграции, именуемая семантическим моделированием<sup>2</sup>, базируется на машинном обучении и обеспечивает генерирование онтологий проектируемого из согласованной системы онтологических паттернов его внутренних спецификаций.

К нейросимвольной интеграции можно отнести и верификацию обученности нейросетей на основе проверки выполнимости логических спецификаций<sup>3</sup>, что особенно актуально для ответственных объектов управления<sup>4</sup>.

Приведем примеры гибридизации ИС, когда в качестве базовой составляющей технологии интеграции выступают не коннекционизм, а символично-логические средства (при том, что они могут входить и в состав интегрируемого).

Интеграцию логических средств и автоматных моделей динамики автономных систем для планирования и реализации действий обеспечивает методология проверки выполнимости

---

<sup>1</sup> *Mao J., Gan C., Kohli P., Tenenbaum J.B., Wu J.* The Neuro-symbolic Concept Learner: Interpreting Scenes, Words, and Sentences from Natural Supervision // ICLR. 2019. P. 1–28.

<sup>2</sup> *Хорошевский В.Ф.* Проектирование систем программного обеспечения под управлением онтологий: модели методы, реализации // Онтологии проектирования. 2019. Т. 9. № 4 (34). С. 429–448.

<sup>3</sup> *Katz G., Barrett C.W., Dill D.L., Julian K., Kochenderfer M.J.* An Efficient SMT Solver for Verifying Deep Neural Networks // CAV. 2017. P. 97–117; *Pulina L., Tacchella A.* Challenging SMT Solvers to Verify Neural Networks // J. Artificial Intelligence Communications. 2012. Vol. 25. № 2. P. 117–13.5.

<sup>4</sup> *Калев И.А.* Камо грядеши? // Экономические стратегии. М.: ИЭС, 2019. Т. 21. № 5 (163). С. 6–15.

логических спецификаций на модели (model checking)<sup>1</sup>. При ее применении в ИС для построения плана действий ищется автоматная модель, на которой как на интерпретации некоторого логического языка были бы выполнимы спецификации требуемого поведения ИС.

Методология логической интеграции, именовавшаяся автомами<sup>2</sup> семантическим программированием, а ныне семантическим моделированием и задачным подходом<sup>3</sup>, использует т. н. формульную определимость<sup>4</sup>, а именно определимость логических спецификаций средств решаемой задачи в некотором специальном подмножестве языка ППЛ. Название методологии объясняется тем, что она основывается на выполнимости спецификаций на конструктивной модели задачи. Методология обеспечивает адекватность и корректность проектируемого. В частности, в семантическом моделировании большое значение придается возможности решения задач с помощью так называемых «оракулов» как внешних решателей. Они осуществляют поиск ответа на обращенный к ним запрос, что позволяет в виде оракулов использовать как строгие математические, так и эвристические модели.

---

<sup>1</sup> *Lomuscio A., Michaliszyn J. Verifying Multi-Agent Systems by Model Checking Three-valued Abstractions // Proc. 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2015). Istanbul, 2015. P. 189–198.*

<sup>2</sup> *Гончаров С.С., Свириденко Д.И. Проблемы цифровизации и семантическое моделирование // Материалы XII Мультиконференции по проблемам управления (МКПУ-2019). Ростов н/Д.; Таганрог: Изд-во ЮФУ, 2019. Т. 1. С. 16–19.*

<sup>3</sup> *Свириденко Д.И., Витяев Е.Е. Задачный подход к искусственному интеллекту и его теоретическая и технологическая база // Сб. трудов XVIII Национальной конференции по искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 36–44.*

<sup>4</sup> *Еришов Ю.Л. Определимость и вычислимость. 2-е изд., испр. и доп. Новосибирск: Науч. кн., 2000. 318 с.*

## Об интеграции ИС на базе подстановочных логик

Развиваемая методология дедуктивно-абдуктивной интеграции на базе подстановочных логик тоже относится к символьно-логическим подходам, объединяя возможности АДТ, традиционного логического программирования и логического планирования действий как самопрограммирования. Эта методология логической интеграции базируется не на семантическом (модельном) подходе, а на синтаксическом, т. е. на выводимости.

Здесь, в отличие от упомянутого ранее<sup>1</sup>, базовой составляющей технологии интеграции (как и в работе<sup>2</sup>) выступает логический формализм, а не коннекционизм, и как отмечено выше, в отличие от изданий<sup>3</sup> используется не выполнимость в той или иной интерпретации логически формализованных требований (спецификаций) к поведению ИС, а их выводимость. Используются логические исчисления разной семантики, в том числе конструктивной, на базе подмножеств первопорядкового языка, именуемых подстановочными.

Всякий подстановочный язык состоит из типово-кванторных формул, т. е. образован из типовых кванторов (в т. ч. «фиктив-

---

<sup>1</sup> *Komendantskaya E.* First-order Deduction in Neural Networks // Proc. 1st International Conference on Language and Automata Theory and Applications (LATA-2007). Tarragona, 2007. P. 307–319; *Mao J., Gan C., Kohli P., Tenenbaum J.B., Wu J.* The Neuro-symbolic Concept Learner: Interpreting Scenes, Words, and Sentences from Natural Supervision // ICLR. 2019. P. 1–28.

<sup>2</sup> *Свириденко Д.И., Витяев Е.Е.* Задачный подход к искусственному интеллекту и его теоретическая и технологическая база // Сб. трудов XVIII Национальной конференции по искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 36–44.

<sup>3</sup> *Lomuscio A., Michaliszyn J.* Verifying Multi-Agent Systems by Model Checking Three-valued Abstractions // Proc. 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2015). Istanbul. 2015. P. 189–198; *Гончаров С.С., Свириденко Д.И.* Проблемы цифровизации и семантическое моделирование // Материалы XII Мультиконференции по проблемам управления (МКПУ-2019). Ростов н/Д.; Таганрог: Изд-во ЮФУ, 2019. Т. 1. С. 16–19.

ных», например, в пропозициональном подмножестве) при разных ограничениях на их типовые условия. Подстановочность означает, что замена в формуле ее главных подформул на логически не менее сильные образует не менее сильные формулы. При этом типовые условия кванторов и их подформулы не считаются главными. Подстановочность используется при построении правил вывода и обосновании корректности и полноты исчислений.

Планирование решения задачи осуществляется с учетом возможности обращения как к внутренним (встроенным), так и внешним средствам (источникам информации и решателям). Кроме того, с помощью гипотезирования как правдоподобного вывода могут формироваться причины наблюдаемых явлений для обновления базы знаний.

Конструктивная логика была аксиоматизирована А. Гейтингом, а А.Н. Колмогоров в 1932 г. предложил интерпретацию интуиционистских исчислений как исчислений задач<sup>1</sup>. Развиваемые нами логические средства разработаны в духе исчислений задач в конструктивной, декларативной и немонотонной семантиках<sup>2</sup>.

Интеграция при этом логических средств достоверного и правдоподобного выводов между собой и с другими средствами ИС, например оптимального управления и нейрораспознавания образов, осуществляется путем дедуктивно-абдуктивного вывода в разработанном исчислении  $J$  с одноместными прави-

---

<sup>1</sup> Kolmogoroff A.N. Zur Deutung der Intuitionistischen Logik // Math. Z. 1932, № 35. P. 58–65.

<sup>2</sup> Васильев С.Н., Галяев А.А. Логико-оптимизационный подход в задачах преследования группы целей // ДАН. 2017. Т. 474. № 6. С. 1–7; Васильев С.Н., Жерлов А.К., Федосов Е.А., Федунев Б.Е. Интеллектуальное управление динамическими системами. М.: Физматлит, 2000. 352 с.; Buzikov M., Galyaev A., Guryev Yu., Titov K., Yakushenko E., Vassilyev S. Intelligent Control of Autonomous and Anthropocentric On-board Systems // Procedia Computer Science. 2019. Vol. 150. P. 10–18; Vassilyev S.N. Abductive Inference Method in Problems of Explaining the Observed // J. Computer and Systems Sciences International. 2021. Vol. 60. № 1. P. 153–161; Васильев С.Н. Формализация знаний и управление на основе позитивно-образованных языков // Информационные технологии и вычислительные системы. 2008. № 1. С. 3–19.

лами обработки знаний на разных подмножествах языка ППЛ<sup>1</sup>. Планирование целенаправленного решения задач управления осуществляется благодаря конструктивной семантике выводов в исчислении  $J$  целевой спецификации  $G$  из знаний ИС, включающих спецификации источников данных, знаний и средств их обработки, не только встроенных, но и агентов-партнеров (об их готовности к тем или иным действиям, приказы, разрешения и т. п.), что важно в задачах с неполной информацией и конфликтностью среды функционирования ИИ.

Устаревание знаний о мире, например из-за действий самого объекта управления, может преодолеваться немонотонной семантикой дедуктивного правила вывода. Кроме того, недостаток знаний или встроенных средств учитывается абдуктивным правилом, формирующим гипотезы  $\Delta$  как дополнительные предположения о мире. В числе этих гипотез — объяснения наблюдаемого эффекта  $G$  (явления, данных разведки).

Понятие абдуктивного вывода чаще всего определяется следующим образом<sup>2</sup>. Рассматриваются предметная область и её базовая теория  $T$ . Наблюдается некоторое явление  $G$ . Ищется нетривиальная гипотеза-объяснение  $\Delta$ , не противоречащая теории  $T$  и такая, что из  $T$  и  $\Delta$  выводимо наблюдаемое  $G$ .

В разработанном для этого исчислении  $J$  дедуктивно-абдуктивных выводов при формализации теории  $T$  используются подмножества первопорядкового языка, которые подстановочного типа и, в частности, могут состоять из структурно позитивных фор-

---

<sup>1</sup> Васильев С.Н., Галяев А.А. Логико-оптимизационный подход в задачах преследования группы целей // ДАН. 2017. Т. 474. № 6. С. 1–7; Buzikov M., Galyaev A., Guryev Yu., Titov K., Yakushenko E., Vassilyev S. Intelligent Control of Autonomous and Anthropocentric On-board Systems // Procedia Computer Science. 2019. Vol. 150. P. 10–18; Васильев С.Н. Формализация знаний и управление на основе позитивно-образованных языков // Информационные технологии и вычислительные системы. 2008. № 1. С. 3–19.

<sup>2</sup> Poole D. Representing Diagnosis Knowledge // J. Annals of Mathematics and Artificial Intelligence. 1994. Vol. 11. P. 33–50; Kowalski R. Logic Programming // In book: Computational Logic, Edition: In the History of Logic series / edited by D. Gabbay and J. Woods. Elsevier, 2014. P. 523–569; Финн В.К. Об эвристиках ДСМ-исследований // НТИ. 2019. Сер. 2. № 10. С. 1–34.

мул без дизъюнктивных ветвлений с ограничениями на входящие в них литералы: в каноническом представлении формулы каждая ее базовая подформула<sup>1</sup> не содержит противоположных литералов.

В исчислении  $J$  гипотезы могут формироваться построением логически минимальной гипотезы (миноранты) и затем переходом некоторыми упрощающими преобразованиями к ее достаточным условиям  $\Delta$  (например, в терминах т. н. абдуцентов). В научной литературе<sup>2</sup> приводятся примеры применения такого дедуктивно-абдуктивного вывода для объяснения наблюдаемых аномалий в задачах технической и медицинской диагностики.

В литературе в работах по абдуктивному гипотезированию на теорию  $T$  предметной области часто накладываются ограничения с целью максимального упрощения синтеза объяснения и его обоснования. Например, содержание утверждений теории должно иметь форму “if-and-only-if” либо может требоваться свойство хорновости утверждений теории  $T$  или его некоторые обобщения. Мы не используем подобных требований к представлению знаний. Свойство подстановочности языка  $L$  не сужает выразительной силы, придает ему крупноблочность, повышает его наглядность и совместимость с эвристиками, что важно для комбинирования стратегий вывода с эвристиками, например, ассоциируемыми с результатами нейрообработки.

Возможности актуализации базы знаний ИС полезны, например, в задачах поиска или обеспечения скрытности мобильных объектов в конфликтных средах. Такова, например, проблематика патрулирования роботами зоны возможного проникновения нарушителя, планирования действий по их безопасному переходу в оперативный район, выбору и преследованию ими мобильных объектов противника, уклонению от них и т. п.<sup>3</sup>

---

<sup>1</sup> Васильев С.Н. Формализация знаний и управление на основе позитивно-обработанных языков // Информационные технологии и вычислительные системы. 2008. № 1. С. 3–19.

<sup>2</sup> Vassilyev S.N. Abductive Inference Method in Problems of Explaining the Observed // J. Computer and Systems Sciences International. 2021. Vol. 60. № 1. P. 153–161.

<sup>3</sup> Galyaev A.A. On the Detection Functional in Motion of an Object in a Threat Environment // J. Automation and Remote Control. 2011. Vol. 71. № 4. P. 634–639;



Развиваемые средства не требуют трансляции знаний в сильно неестественную клаузалную форму (clausal form), обычную для резолюционных пруверов. Типичным недостатком нерезолюционных пруверов является многоместность и количество используемых ими правил вывода. Например, нерезолюционный прувер из работы J. Otten<sup>1</sup> использует 4 трехместных правила дедукции, что заметно повышает размерность пространства поиска выводов по сравнению с двумя одноместными правилами вывода, используемыми в исчислении  $J$ .

Формализация знаний в подстановочном языке с ослаблением требования позитивности дополнительно расширяет синтаксические возможности представления знаний и повышает его естественность, а возможная потеря на используемом подмножестве языка  $L$  полноты дедуктивного правила относительно выводимости может компенсироваться правилом абдукции.

В составе интегрируемых средств на технологической основе символизма дедуктивно-абдуктивных исчислений перспективно применение и других механизмов рассуждений и планирования действий, например решения подзадач на основе аналогий и прецедентов (Case-based Reasoning) и индуктивного вывода<sup>2</sup>.

---

Leahy K., Zhou D., Vasile C.I., Oikonomopoulos K., Schwager M., Belta C. Provably Correct Persistent Surveillance for Unmanned Aerial Vehicles Subject to Charging Constraints // *Experimental Robotics*. 2016. P. 605–619; Васильев С.Н., Галяев А.А. Логико-оптимизационный подход в задачах преследования группы целей // *ДАН*. 2017. Т. 474. № 6. С. 1–7; Buzikov M., Galyaev A., Guryev Yu., Titov K., Yakushenko E., Vassilyev S. Intelligent Control of Autonomous and Anthropocentric On-board Systems // *Procedia Computer Science*. 2019. Vol. 150. P. 10–18; Васильев С.Н., Галяев А.А., Рубинович Е.Я. Интеллектуальное управление мобильными средствами в конфликтных средах // *Материалы Международной конференции «Проблемы механики и управления»*. Махачкала, 2018. С. 117–119.

<sup>1</sup> Otten J. NanoCoP: a Non-clausal Connection Prover // *Proc. International Joint Conference on Automated Reasoning*. 2016. P. 302–312.

<sup>2</sup> Das R., Godbole A., Monath N., Zaheer M., McCallum A. Probabilistic Case-based Reasoning for Open-World Knowledge Graph Completion // *CaseBR arXiv:2010.03548v2 [cs.CL]*; Muggleton S.H., Lin D., Pahlavi N., Nezhad A.T. Meta-interpretive Learning Application to Grammatical Inference // *J. Machine Learning*. 2014. № 94. P. 25–49.

В числе интегрируемых программных разработок могут быть средства когнитивной эволюции<sup>1</sup>, эволюционной оптимизации<sup>2</sup> и др., а помимо нейросетей и алгоритмов их обучения могут использоваться конкурирующие с ними вероятностные, логические и алгебраические методы распознавания и эволюционные, вероятностные и другие алгоритмы обучения<sup>3</sup>, что альтернирует и сам формируемый план целенаправленных действий. Альтернативный отбор спецификаций средств для их учета в выводимом плане действий и выбор наилучшего из альтернативных планов требуют методов многокритериальной оптимизации, адекватных режимам реального времени<sup>4</sup>.

В проблематике ИС управления полезны средства картирования уровня рисков<sup>5</sup> по данным бортовых и внешних детекторов аномалий и анализаторов потенциальных угроз. Оптимизация маршрутов может учитывать данные карты рисков, скорость движения объекта управления по маршруту, продолжительность перехода в целевую область, режимы работы некоторых подсистем и другие характеристики, влияющие на выбор маршрута и скрытность<sup>6</sup>.

---

<sup>1</sup> Редько В.Г. Моделирование когнитивной эволюции: на пути к теории эволюционного происхождения мышления. 2-е изд. М.: ЛЕНАНД, 2019. 264 с.

<sup>2</sup> Diveev A.I., Balandina G.I., Konstantinov S.V. Binary Variational Genetic Programming for the Problem of Synthesis of Control System // Proc. 13<sup>th</sup> International Conference on Neural Computation, Fuzzy Systems and Knowledge Discovery (ICNCFSCD-2017). 2017. P. 165–170.

<sup>3</sup> Vinogradov D.V. VKF-method of Hypotheses Generation // J. Communications in Computer and Information Science. 2014. Vol. 436. P. 237–248; Skrynnik A., Panov A.I. Hierarchical Reinforcement Learning with Clustering Abstract Machines // J. Communications in Computer and Information Science. 2019. Vol. 441. P. 30–43.

<sup>4</sup> Федунев Б.Е. Интеллектуальные системы тактического уровня: состав, базы знаний // Сб. трудов XVIII Национальной конференции по искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 87–94.

<sup>5</sup> Buzikov M., Galyaev A., Guryev Yu., Titov K., Yakushenko E., Vassilyev S. Intelligent Control of Autonomous and Anthropocentric On-board Systems // Procedia Computer Science. 2019. Vol. 150. P. 10–18.

<sup>6</sup> Buzikov M., Galyaev A., Guryev Yu., Titov K., Yakushenko E., Vassilyev S. Intelligent Control of Autonomous and Anthropocentric On-board Systems // Procedia Com-

## К устойчивому развитию

Интересный подход к проблеме устойчивого развития общества по результатам моделирования когнитивной эволюции биоорганизмов предложен в работе В.Г. Редько<sup>1</sup>. В частности, некоторыми выводами по результатам моделирования устранимости конкуренции биоорганизмов было бы интересно дополнить концепцию т. н. «суперинтеллектуального» общества 5.0 из издания У. Норицугу<sup>2</sup> (по-другому — «суперумного» общества<sup>3</sup>), где, помимо оценок обоснованности суждений о влиянии на общество искусственного интеллекта, общество 5.0 характеризуется оптимальным сочетанием ресурсов общества посредством интеграции реального (физического) и виртуального (цифрового) пространства в условиях рыночной конкуренции. Проблема преодоления сложности такой конкуренции рассматривается также в работе И.А. Каляева<sup>4</sup>.

Covid-19 — не исключение, а одна из череды проблем прошлого и будущего. И мир должен быть подготовлен лучше к встрече с ними. Странам необходимо сотрудничать и учиться управлять в науке и искусстве противодействия угрозам, в т. ч. неокOLONIALИЗМУ государств и транснациональных компаний.

---

puter Science. 2019. Vol. 150. P. 10–18; *Galayev A.A.* On the Detection Functional in Motion of an Object in a Threat Environment // J. Automation and Remote Control. 2011. Vol. 71. № 4. P. 634–639; *Васильев С.Н., Галаяев А.А.* Логико-оптимизационный подход в задачах преследования группы целей // ДАН. 2017. Т. 474. № 6. С. 1–7; *Васильев С.Н., Галаяев А.А., Рубинович Е.Я.* Интеллектуальное управление мобильными средствами в конфликтных средах // Материалы Международной конференции «Проблемы механики и управления». Махачкала, 2018. С. 117–119.

<sup>1</sup> *Редько В.Г.* Моделирование когнитивной эволюции: на пути к теории эволюционного происхождения мышления. 2-е изд. М.: ЛЕНАНД, 2019. 264 с.

<sup>2</sup> *Норицугу У.* Общество 5.0: взгляд Mitsubishi Electric // Экономические стратегии. 2017. № 4. С. 2–11.

<sup>3</sup> *Гончаров С.С., Свириденко Д.И.* Проблемы цифровизации и семантическое моделирование // Материалы XII Мультиконференции по проблемам управления (МКПУ-2019). Ростов-на-Дону-Таганрог: Изд. ЮФУ, 2019. Т. 1. С. 16–19.

<sup>4</sup> *Каляев И.А.* Камо грядеши? // Экономические стратегии. М.: ИЭС, 2019. Т. 21. № 5 (163). С. 6–15.

Трагедия пандемии должна инициировать стремление к росту, развитию и добрососедству. Так, президент В. Путин молниеносно отреагировал по одному из важнейших направлений. Учитывая, что разрастание эпидемии связано со скученностью людей и циркуляцией людского потока через объекты высокой плотности (популярные мировые туристские объекты), потока — разносящего болезнь по миру, он дал установку экстренно развивать внутренний туризм. А это значит индустрию, практически новую, точнее забытую старую для России, с необходимым качеством и доступностью во многих регионах. Индустрию — с новыми требованиями к культуре, строительству, транспорту и многому еще.

Государства должны создать механизмы предсказания и борьбы с опасными явлениями. Covid-19 показал, что любое государство в первую очередь беспокоится за свою страну. Неприятие отдельными государствами, в т. ч. США, единых требований по экологии, — другой пример «некооперативности». Развитие международной обстановки в связи с событиями на Украине и успехами специальной военной операции России не впервые говорит о многополярности современного мира. Очевидна потребность в парадигме развития стран и человечества в целом, учитывающей разные виды опасностей, в т. ч. гегемонистических амбиций государств и корыстных интересов транснациональных компаний.

Проблемы противостояния опасностям актуализируют целеполагание и единение взглядов на развитие России в XXI веке<sup>1</sup>. Стране необходимо применить принцип стратегического управления — превратить угрозы в возможности, слабые стороны (проблемы) — в сильные. Соответственно: *«Проблемы России — это ее пока еще не использованные возможности»*. И в этом отношении у России объективные предпосылки для стремительного прорывного развития существуют: самая обширная терри-

---

<sup>1</sup> Васильев С.Н., Гончаренко С.С. От природно-техногенных и других угроз к устойчивому развитию // Научные труды Вольного экономического общества России. 2020. № 3 (223). С. 203–212.

тория в мире, в два раза превышающая территорию каждого из ближайших по ранжиру соседей (Китая и США) и огромный экологический резерв человечества; самый большой запас сырьевых ресурсов, в несколько раз превышающий ближайших по запасам соседей (Китая и США); обширная зона плодородия, проходящая широтной полосой через всю страну и имеющая 44% всех черноземов Земли.

Однако состояние и перспективы наращивания ВВП ведущих стран мира по оценкам МВФ в 2021 г. показывают необходимость *увеличения номинального ВВП России* на порядок для выхода на уровень лидеров. Другие проблемы России — а) *низкая производительность труда*, б) *недостаточная численность населения* для столь обширной территории и такого объема сырьевых ресурсов мирового уровня, в) *асимметрия заселенности и освоения* различных регионов, г) известные проблемы в сфере *образования и науки* (они же снова причины других проблем) и др.<sup>1</sup>

Стране нужна миссия, поддерживаемая всем народом, миссия, вызывающая общенациональный творческий, производственный порыв<sup>2</sup>. Порыв, которому подчинены все силы, средства и механизмы государства. Кто и что мы — Россия, для своего народа и для всего мира, для стран СНГ, для стран дальнего зарубежья, дружественных и недружественных. Миссии есть у многих стран. Миссия России определена, но пока нет единения интересов и интеграции усилий. Миссия — в реализации национальной стратегии устойчивого развития с преодолением

---

<sup>1</sup> Васильев С.Н., Гончаренко С.С. От природно-техногенных и других угроз к устойчивому развитию // Научные труды Вольного экономического общества России. 2020. № 3(223). С. 203–212.

<sup>2</sup> Новая парадигма развития России в XXI веке. Комплексные исследования проблем устойчивого развития: идеи, результаты / Под ред. В.А. Коптюга, В.М. Матросова, В.К. Левашова. М.: Изд-во “Academia”, 2000. 397 с.; Селезнев Г.Н., Христенко В.Б., Залиханов М.Ч., Львов Д.С., Матросов В.М., Гранберг А.Г., Левашов В.К., Урсул А.Д., Шелехов А.М. Научная основа стратегии устойчивого развития Российской Федерации / Под общ. ред. М.Ч. Залиханова, В.М. Матросова, А.М. Шелехова. М.: Изд-во Госдумы РФ, 2002. 392 с.

внешних и внутренних угроз, системных кризисов, построением общества высокого качества жизни и ответственным вкладом в развитие ноосферных процессов на планете. Миссия — показать пример успешного начала реализации этой стратегии, увлекающей страны и народы в ноосферу с глобальным устойчивым со-развитием природы и техносферы, с приоритетами нравственности и сознания над социальным бытием, интеллектуально-духовных процессов над материальными, приоритетами науки и информации, как ресурсов развития, над энергией и веществом.

Ранее Россия показывала миру, теперь уже классические, примеры победного выхода не только из войн, но и из системных кризисов экономики. «Новый курс Рузвельта» с выходом США из Великой депрессии не являлся пионерным. В его основе — опыт СССР, его первых пятилеток, реализовавший стратегический план ГОЭЛРО. При этом план реализовывался в еще более критической ситуации, когда не было перепроизводства и достаточной индустриальной и транспортной основы.

Образование, культура и СМИ должны соответствовать миссии России — быть опережающими, воспитывающими и ориентирующими на высоконравственные цели. *«Человеческое знание состоит не из одной математики и технологий... это еще только низшее знание, высшее объемлет собой мир нравственный»*, — в предвидении технократического кризиса человечества писал еще В.Г. Белинский.

Успехи фундаментальной и прикладной биомедицины и не только на поле боя с Covid-19, как и успехи всей науки, тоже относятся к числу важнейших факторов авторитета государства и весомости его голоса в мировой ноосферной политике. Российская наука нуждается в усилении поддержки государства.

Устойчивое развитие человечества и само его существование подвержены глобальным рискам в условиях автаркии стран, изоляционизма или блоковости. Однако эти риски, сужая, не исключают разрешимости общих проблем человечества объединением усилий сильных стран-партнеров.

Для ускоренного прорывного развития России крайне необходимо масштабное расширение индустриального пространства с выверенной и оптимальной моделью ускоренной реиндустриализации Сибири, Дальнего Востока и Российского Севера в виде Генеральной схемы развития и размещения производительных сил, аналогичной разработке 1968 года.

Обоснование этой необходимости, по настоящему не востребованное страной, содержится в исследованиях отечественных ученых (в т. ч.<sup>1</sup>, с целой системой фондо- и трудоемких стратегических проектов, и в числе первых — развитие транспортной инфраструктуры для новых транспортно-промышленных комплексов (ТПК). Главным транспортно-промышленным поясом должен быть Северный широтный экономический пояс на базе сквозной Северо-Российской евразийской широтной железнодорожной магистрали, включающей Сахалинскую дорогу, Байкало-Амурскую магистраль, Северо-Сибирскую магистраль и железнодорожные выходы на Баренцево, Белое и Балтийское моря. Этот пояс на перспективных месторождениях с мощными промышленными объектами кратно увеличит ВВП страны.

Проблема дефицита трудовых ресурсов вызовет сильнейшую потребность в автоматизации производственных и вспомогательных процессов и повышении качества жизни людей в новых ТПК и вообще в новых городах Сибири, Дальнего Востока и Российского Севера, актуальность строительства которых

---

<sup>1</sup> Гончаренко С.С., Персианов В.А. Роль социальной инфраструктуры в экономике России. М.: НП «Центр стратегического партнерства», 2013. 339 с.; Проблемные регионы ресурсного типа: азиатская часть России. Новосибирск: ИЭОПП СО РАН, ИИ СО РАН, ИСЭМ СО РАН, 2005. 377 с.; Транспортно-промышленное освоение Сибири и Дальнего Востока и приграничных территорий — системный фактор прорывного развития экономики России, ее интеграции в мировую систему («Горизонт-2030»), ч. II: Аналитический доклад Байкальского экономического форума / Под науч. ред. С.Н. Васильева, В.И. Суслова, В.А. Персианова, С.С. Гончаренко, Т.Н. Есиковой. М.; Новосибирск: ИПУ РАН, ИЭОПП СО РАН, ЕАТИЦ, 2006. 192 с.; Прокофьева Т.А., Гончаренко С.С., Семенов Н.Н., Элларян А.С. Стратегическая доктрина развития транспорта и формирования интегрированных транспортно-логистических систем в регионах Европейского Севера, Сибири и Дальнего Востока. М.: ОАО «ИТКОР», 2020. 226 с.

акцентировалась и в последних выступлениях высших должностных лиц страны. Интеллектуализация и применение роботов и включающих их распределенных робототехнических комплексов (РТК), в т. ч. с коллективным управлением<sup>1</sup>, позволит не только сократить дефицит работников, но и повысить производительность труда.

По данным Центра экономических показателей при Лондонской школе экономики, в развитых странах до 10% прироста ВВП за последние годы обусловлено внедрением промышленной робототехники, у которой в России пока ряд проблем<sup>2</sup>. От объемов ВВП США и Китая — эти 10% равны примерно объему сегодняшнего ВВП России и это будучи лишь дополнением к приросту за счет самой реиндустриализации страны.

Переход к роботизированным производственным системам в Стратегии научно-технологического развития Российской Федерации (указ Президента от 01.12.2016, № 642) представлен в числе первых приоритетов. Дополнительные формы преобразования страны (свободные экономические зоны, кластеры) потребуют учета факторов риска в условиях нового мирового порядка. Создание широкого спектра масштабируемых малолюдных производств на основе РТК и, как следствие, увеличение ВВП страны будет амбициозным компонентом стратегии устойчивого развития и источником психологического подъема энтузиазма российского народа, как это бывало не раз в нашей истории.

При упреждающем создании новых рабочих мест и перекалфикации работников роботизация не вызовет безработицы, а сегодняшняя нехватка инженеров и математиков-программистов в сфере РТК устранима мерами повышения привлекательности этих профессий, предпринимаемыми сейчас в стране.

---

<sup>1</sup> Каляев И.А., Гайдук А.Р., Капустян С.Г. Модели и алгоритмы коллективного управления в группах роботов. М.: Физматлит, 2009. 278 с.

<sup>2</sup> Перепелица А.А. Промышленная робототехника в России. Состояние и проблемы развития. Видение интератора. [http://alphajet.ru/knowledge/articles/stati\\_45.html](http://alphajet.ru/knowledge/articles/stati_45.html)



В создании РТК и роботов нуждаются не только производственные системы промышленности, но и медицинская, космическая и другие отрасли страны. Эволюция этих разработок в направлении от интеллектуальных систем к созданию автономных интеллектуальных систем — актуальное направление научно-технологического развития государства.

Требуют постоянного внимания проблемы развития электроники и интеллектуализации программно-аппаратных платформ цифровизации, а также проблемы цифрового суверенитета — гарантированной защиты цифровых платформ и инфраструктур государства с противодействием утечке данных и инноваций, а также злонамеренному их использованию, включая проблемы государственного и международного правового обеспечения атмосферы цифрового мира и проблемы противодействия использованию цифровых средств для разрушения духовно-нравственных устоев человека.

**Заключение.** Рассмотрены проблематика искусственного интеллекта, эволюция разработок и их интеграция на базе логических и нейросетевых технологий. Отмечены влияние искусственного интеллекта и цифровых технологий на общество и их значение для экономики, экологии и противодействия природным, техногенным, неоколониальным и другим угрозам. В контексте стратегического развития России акцентированы роль искусственного интеллекта и роботизации в реиндустриализации восточных и других регионов страны и значение цифровых технологий для сохранения мира в условиях трансформации мирового порядка и желаемого перехода цивилизации к ноосферному состоянию с приоритетами науки и информации, как ресурсов развития, над энергией и веществом и с главенством духовно-нравственных ценностей.

Исследования и разработки в области ИИ — стратегическое направление научно-технологического развития и безопасности государства, в том числе на основе интеллектуализации цифровых платформ государственного управления, социально-экономического развития страны и ее защиты от

природно-техногенных и других угроз. Актуальным является интенсификация междисциплинарных проектов<sup>1</sup> в области ИИ и когнитивных исследований с социально-эколого-экономическими приложениями в поддержку формирования и выполнения национальной стратегии устойчивого развития<sup>2</sup>. Дело за реализацией, а для этого необходимо дальнейшее единение в стране интересов и усилий.

Миссия России — успешная реализация этой стратегии как пример устойчивого развития, увлекающий страны в глобальное устойчивое со-развитие человечества, природы и техносферы в противовес трендам неокOLONиализма и возрождения нацизма.

---

<sup>1</sup> Лекторский В.А. Философия, искусственный интеллект и когнитивная наука // Искусственный интеллект. Междисциплинарный подход / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИнтелЛ, 2006. С. 13–26.

<sup>2</sup> Соколов И.А. Теория и практика искусственного интеллекта // Вестник Российской академии наук. 2019. Т. 89. № 4. С. 365–370; Васильев С.Н., Гончаренко С.С. От природно-техногенных и других угроз к устойчивому развитию // Научные труды Вольного экономического общества России. 2020. Т. 89. № 3 (223). С. 203–212; Редько В.Г. Моделирование когнитивной эволюции: на пути к теории эволюционного происхождения мышления. 2-е изд. М.: ЛЕНАНД, 2019. 264 с.

## КАК СОВРЕМЕННЫЕ ЦИФРОВЫЕ ТЕХНОЛОГИИ МЕНЯЮТ ПРЕДСТАВЛЕНИЕ ОБ ОКРУЖАЮЩЕМ МИРЕ (ИСКУССТВЕННОЕ ОБЩЕСТВО)

---

---

В современном цифровом мире многие сложившиеся представления о мире, в том числе мнения известных философов, существенно меняются. Нижеприведенный текст для удобства восприятия разделен на шесть частей:

1. Вычисления как основа для объяснений устройства мира.
2. Цифровые двойники, искусственное общество.
3. Моделирование сознания.
4. Моделирование духовного мира.
5. Цифровые двойники в повседневной жизни.
6. Как человечество движется от материального мира к духовному.

1. Начнем с понятия алгоритма, известного со времен Аристотеля.

Прорыв произошел в первой половине XX века, когда математики дали несколько точных с их точки зрения определений понятия алгоритма. Например, рекурсивные функции, алгоритмы Маркова, машины Тьюринга и др. Серьезный прорыв в этой

---

\* *Макаров Валерий Леонидович* — академик РАН, Научный руководитель ЦЭМИ РАН, заместитель председателя Научного совета при Президиуме РАН по методологии искусственного интеллекта и когнитивных исследований, Москва.

*Makarov V.L.* — Academician, Scientific Director of CEMI RAS.

области произошел, когда Алан Тьюринг сформулировал свою универсальную машину. А именно машину, которая с помощью введенной программы может осуществить любой алгоритм. Тогда появились электронные вычислительные машины, любые алгоритмы стало возможным записывать на программных языках (к примеру — АЛГОЛ, ФОРТРАН, JAVA, Питон и пр.), которых теперь имеется несколько сотен.

Появилось весьма распространенное мнение, что любое явление, наблюдаемое в мире, можно сформулировать в форме алгоритма. То есть все в мире имитируется компьютерным вычислением. И мир есть не что иное, как вычисления.

Один из современных крупнейших ученых Стивен Вольфрам обосновывает это утверждение в своей книге «Новый вид науки»<sup>1</sup>. Однако все-таки преобладает мнение, что мир устроен сложнее. Есть вещи и явления, которые невозможно объяснить в терминах алгоритмов<sup>2</sup>. Детально эта тема обсуждается в книге В.Ф. Петренко, А.П. Супрун<sup>3</sup>.

Понятие духовного мира, идущее еще от Платона, трудно объяснить алгоритмически. Но цифровой мир развивается сейчас невиданными темпами.

Независимо от разных мнений и споров бурно развивается деятельность по созданию т. н. цифровых двойников (digital twins). Цифровые двойники оказываются востребованными в разных областях, поскольку экономят время и деньги. Есть двойники у автомобилей, самолетов, атомных бомб и вообще всего, созданного человеческими руками. Двойник атомной бомбы, очевидно, понадобился, чтобы совершенствовать бомбу без испытаний в натуре, что, как известно, запрещено. Что касается автомобилей и им подобных, то эксперименты, испытания и пр. в компьютере намного проще, быстрее и дешевле, чем

---

<sup>1</sup> *Wolfram Stephen. A New Kind of Science. 2003.*

<sup>2</sup> *Пенроуз Р. Новый ум короля. О компьютерах, мышлении и законах физики. М., 2011.*

<sup>3</sup> *Петренко В.Ф., Супрун А.П. Методологические пересечения психосемантики сознания и квантовой физики. М.; СПб.: Нестор-История, 2017.*

на реальном образце. Вопрос лишь в том, насколько точно цифровой двойник совпадает с его реальным образцом.

В результате в настоящее время активно развивается ИТ<sup>1</sup> — отрасль создания разнообразных цифровых двойников. Особенно эта отрасль стала востребованной после появления нового явления — интернета вещей. Дело в том, что людям надо знать, в каком состоянии находятся те или иные вещи. Например, получающий популярность каршеринг обязывает хозяев автомашин постоянно знать где находится машина, сколько бензина в баке, какие поломки случились и т. д. Становится привычным иметь информацию о ситуации в доме, что есть в холодильнике и пр. Принципиальные изменения ждут пожарную охрану. Любые сигналы о пожарной опасности моментально поступают в соответствующие службы.

2. Принципиальный скачок в осознании смысла цифровых двойников произошел после появления понятия под названием искусственное общество.

Термин «искусственное общество» стал популярен после выхода статьи об агенто-ориентированных моделях.<sup>2</sup> При этом под искусственным обществом стали понимать агенто-ориентированные модели (сокращенно АОМ), которые стали основным инструментом для представления общества в форме компьютерной программы. Подробно об этом изложено в книге В.Л. Макарова, А.Р. Бахтизина.<sup>3</sup> Популярный международный журнал JASSS (Journal of Artificial Societies and Social Simulation) практически полностью посвящен рассмотрению общества как АОМ. Кстати сказать, понятие и конструкция АОМ в математическом плане совпадает с понятием игры многих лиц, которую взяли на

---

<sup>1</sup> Макаров В.Л. Получение нового знания методом компьютерного моделирования // Искусственный интеллект. Междисциплинарный подход. М.: ООО «ИИИ-телеЛП», 2006.

<sup>2</sup> Epstein Joshua M. and Axtell Robert Growing Artificial Societies: Social Science from the Bottom Up. Brooking Institution Press and MIT Press, Washington DC.

<sup>3</sup> Макаров В.Л., Бахтизин А.Р. Социальное моделирование — новый компьютерный прорыв (Агент-ориентированные модели). М.: Изд-во Экономика, 2013.

вооружение многие нобелевские лауреаты по экономике, в частности Нэш и Эрроу.

Чтобы сконструировать искусственное общество, которое будет имитировать реальное общество и, тем самым, использоваться при принятии решений в реальном обществе, надо иметь детальное представление, что такое человеческое общество. Существует обширная литература в этой области, множество различных представлений о человеческом обществе, часто друг другу противоречащих.

Стоит упомянуть Никласа Лумана, который написал более 70 книг на эту тему, посвятив всю свою жизнь описанию, что такое человеческое общество. Ключевое слово у Лумана — это *коммуникация*.

Коммуникацию Луман понимает очень широко, гораздо шире, чем, скажем, это выражается в современных социальных сетях. Его понимание общества ассоциируется с пониманием известного философа и математика Альфреда Уайтхеда. О представлении Уайтхеда стоит поговорить подробнее.

Он специально подчеркивает явление разнообразия как свойство общества. Разнообразные объединения людей: семьи, команды, клубные и прочие группы, население домов, деревень, районов, городов, наконец, стран. По другим показателям — национальности, языку, религии, идеологии — также получают объединения, иногда очень мощные. Эти объединения то есть «организмы», по Уайтхеду, должны обладать особыми свойствами, не известными более низким уровням. Социальные науки, изучающие общество во всех его проявлениях, образовали набор понятий, относящихся к этим особым свойствам. Это мода, общественное мнение, уровень коллективизма, патриотизма, народного гнева и пр. Отдельно речь пойдет о таких объединениях, как религии, политические партии, объединения на базе идеологии и т. п.

Как правило, поведение людских объединений рассматривалось, понималось, изучалось несколько однобоко. Группы людей не выглядели и не понимались как единый организм, у которого может быть свое мнение, свое мышление, свое сознание наконец.

Не исключено, что в таких организмах могут появиться новые свойства, которых нет у отдельного человека.

Альфред Уайтхед<sup>1</sup> замечает, что «организм» низшего уровня действует в соответствии с правилами «организма» более высокого уровня, в котором он находится. Он приводит примеры с электроном, клеткой живого организма, железом в магнитном поле. Применительно к человеку в группе это очевидно. Один и тот же человек в разных группах действует по-разному. Похоже, что именно это имел в виду Никлас Луман, когда говорил о своем понимании коммуникации.

3. При построении АОМ принципиальную роль играет поведение агентов, в частности, как они принимают решения и действуют. И тут естественным образом появляется необходимость формализовать явление сознания у агентов.

Возникает естественный вопрос: как в искусственном обществе агенты — люди могут действовать как обладающие сознанием? В противном случае они не имеют главного свойства, отличающего человека от другого существа. Ведь люди в таком обществе должны описываться компьютерной программой, что представляется утопией.

В настоящее время не сложилось твердого мнения, что искусственный интеллект, и соответственно, искусственное общество могут, хотя бы теоретически, полностью быть адекватными (быть полной копией) человеку и обществу людей. Сознание пока или в принципе не может быть объяснено вычислительным способом. Мир или «вычислительный мир», согласно С. Вольфраму и его сторонникам, или бесконечно сложен, как думают великие философы прошлого и сторонники Р. Пенроуза, считающие сознание необъяснимым вычислительной логикой.

На практике мы все больше встречаемся с этой ситуацией, поскольку цифровой мир наступает. Человек управляет машиной без водителя совсем не так, как сидя за рулем. Поэтому самый совершенный вариант искусственного общества, как инструмент принятия решений, может приводить человека к ситуациям,

---

<sup>1</sup> Уайтхед Альфред Н. Избранные работы по философии. М.: Прогресс, 1990. С. 210.

которые обычно называются тупиковыми. А из тупика выведет только человеческое сознание.

Желание понять суть человека и общества всерьез начало обсуждаться древнегреческими философами, начиная с Платона и Аристотеля. И это обсуждение продолжается до сих пор, привлекая огромные научные достижения, например, в области психологии и квантовой механики. Недавно, например, вышла книга Габриэля Маркуса,<sup>1</sup> молодого немецкого философа, где подробно рассматриваются самые различные точки зрения на данную проблему. То есть современных философов проблема сознания чрезвычайно интересует.

Вернемся опять к Уайтхеду и его пониманию мира, идущему от физического понимания, начинающегося с атомов и кончая сложными предметами и существами.

По-видимому, есть «приматы», то есть самые маленькие частицы, существование которых еще не доказано экспериментально.<sup>2</sup> Следующий уровень принадлежит элементарным частицам — электронам, протонам, позитронам, нейтрину и пр. Далее из них формируются атомы. Следующий уровень — молекулы. За ними клетки. И уж из молекул и клеток формируется мир вещей, которые даны в ощущениях. Миру вещей принадлежат и живые организмы, в том числе люди. Люди обладают мыслительными способностями, сознанием и, стало быть, свободой воли. То есть они могут осознанно принимать решения относительно своего поведения.

И вот теперь самое интересное. Следующий уровень — это разнообразные объединения людей: семьи, команды, клубные и прочие группы, население домов, деревень, районов, городов, наконец, стран. По другим показателям — национальности, языку, религии, идеологии — также получают объединения, иногда очень мощные. Эти объединения — то есть «организмы»,

---

<sup>1</sup> Маркус Габриэль. Я не есть мозг. Философия духа для XXI века. М.: URSS, 2020

<sup>2</sup> Уайтхед Альфред Н. Избранные работы по философии. М.: Прогресс, 1990. С. 195.



по Уайтхеду, должны обладать особыми свойствами, не известными более низким уровням. Социальные науки, изучающие общество во всех его проявлениях, образовали набор понятий, относящихся к этим особым свойствам. Это мода, общественное мнение, уровень коллективизма, патриотизма, народного гнева и пр.

В представлении о мире, навеянном описанной схемой Уайтхеда, имеется очень любопытное противоречие. На каждом следующем слое разнообразие увеличивается, и значительно. При этом существование и поведение субъектов каждого уровня становятся все более независимыми. Но на предпоследнем и последнем слое информационная зависимость становится все больше. Мир чрезвычайно разнообразен, но при этом информационное пространство опутывает всех агентов и объединяет их в нечто целое. Это целое сейчас модно называть цифровым миром. Стивен Вольфрам в своей книге<sup>1</sup> доказывал, что вселенная — это своеобразный компьютер, где все происходит, так сказать, алгоритмически. Это утверждение весьма сомнительно, о чём дальше будет сказано подробнее.

Подводя итоги вышесказанному, логично распространить такие термины, как мышление, мнение, сознание, точка зрения, которые обычно применяются только к отдельному человеку, на самые разнообразные группы людей, формируемые и действующие в обществе. Коллективное мышление и сознание оказываются не выдумкой, а реальностью.

4. Возникает естественный вопрос: можно ли моделировать (с помощью алгоритмов, с помощью компьютеров) духовный мир человека? Понятие духовного мира появилось еще у Платона, но не изучалось научными методами сколько-нибудь глубоко.

При моделировании мира агентов главными сущностями являются продукты, понимаемые в самом широком смысле, и действия над ними. Как это ни удивительно, представляется, что в духовном мире также имеются свои «продукты» и свои действия над этими «продуктами», которые осуществляют агенты — люди и группы людей.

---

<sup>1</sup> *Wolfram Stephen. A New Kind of Science. 2002.*

Ученые, особенно философы, с давних времен разделяли материальный и духовный миры. Лейбниц даже написал эссе под названием «Монадология», где сравнивает атомы материального мира с монадами духовного. Как известно, атомы понимаются очень широко, начиная с Демокрита. То же можно сказать и о монадах. Термин «монада» идет от древних греков, но стал основным понятием духовного мира именно благодаря Лейбницу. Однако видов атомов не так много, это конечное множество. И поэтому познание материального мира происходит сравнительно быстро. Монады же все абсолютно индивидуальны, и их бесконечное множество. И поэтому познание духовного мира неисчерпаемо. Загадки духовного мира будут существовать всегда.

Связь материального и духовного миров многообразна. Особенно важно описать окружение агента в материальном мире и его ощущение уровня удовлетворения от данного окружения. Например, человек имеет постоянную работу, квартиру, любимую жену, детей, нормальное здоровье. Однако возможен вопрос: удовлетворен ли он своей жизнью? Ответ на него неоднозначен, он зависит от многих факторов, в том числе духовного свойства.

Есть индивидуализм, есть коллективизм, между ними существует соборность. Духовная составляющая есть комбинация сознания и совокупности чувств.

Можно также сделать следующий шаг. В мировых религиях и общественном сознании присутствует понятие «душа». Обычно душу связывают с человеком как отдельным материальным существом. Но коль скоро действующими агентами в обществе являются разные группы (клубы, предприятия, страны, партии...), то логично распространить понятие души на этих агентов. Кстати сказать, в иудаизме есть направление «каббала», которое не отрицает наличие душ у других существ, кроме человека.

Одно из возможных направлений для понимания того, как устроен духовный мир, исходит из аналогии его миру материальному. Материальный мир, как многократно отмечалось, состоит из атомов, которые формируют более сложные структуры и т. д. Можно гипотетически предположить, что атомами в духовном мире являются первичные ощущения: слух, обоняние, зрение и

пр. Из них формируется следующий слой: страх, удовольствие, радость и многое другое. А дальше возникает страдание, любовь разных типов, привязанность, верность. Недаром литераторов называют инженерами человеческих душ, ибо в их произведениях все эти структуры проявляются в разных ипостасях. Как такое понимание духовного мира поможет в построении искусственного общества, увидим дальше.

Кстати сказать, в логических построениях, например в математике, бесспорные утверждения называются аксиомами, другие утверждения выводятся из аксиом с помощью логических правил. В математике это называется доказательствами.

Как уже отмечалось, агентами можно объявить кого угодно. Главное, что агент обладает определенной свободой выбора своих действий, в том числе и в духовном мире. Для этого агента надо наделить разными чувствами, упомянутыми выше. Тогда вместо производства и потребления материальных продуктов будут производиться и потребляться «чувства». Я ставлю термин «чувства» в кавычки, поскольку они понимаются в самом широком смысле слова. Например, к таковым относятся неудовлетворенность, лень, зависть, желание, стремление, мечта и так далее, что не всегда идентифицируется как чувства. Напоминаю, что «чувства» относятся не только к индивидуумам, но и к группам — клубам, предприятиям, партиям, странам и пр. Главный вопрос при компьютерном моделировании, как всегда: как алгоритмически определить и как измерять.

Траектория и эволюция материального мира состоит в производстве и потреблении видов продуктов, понимаемых в самом широком смысле. А что есть эволюция духовного мира? Что происходит с агентами в духовном мире? В частности что они производят и потребляют, если пользоваться этими терминами? Производят агенты духовные продукты, которые довольно часто совпадают с продуктами материального мира, в частности знаниями и другой информацией. Кстати, это обеспечивает связь между указанными мирами. Но есть и чисто духовные продукты, например, молитвы или впечатления от посещения музея, театра и тому подобное. Среди агентов, представляющих группы людей,

естественно, выделяются храмы различных религий, театры, школы, и все объединения, связанные с культурой. Особую роль при этом играют школы и университеты, в которых кроме обычного образования присутствует духовная компонента.

5. Возвращаясь к понятию цифровых двойников, следует остановиться на их использовании в обычной жизни людей. Понятно, что двойник самолета или атомной бомбы существенно упрощает дальнейшее использование и совершенствование изделия. Следующий шаг — это объединения цифровых двойников, например, двойники — индивидуальные агенты и двойники — группы. Таким образом, возникают «матрешки» или, по-научному, фракталы, в которых процесс моделирования усложняется. Хорошим примером здесь является процесс планирования, который возрождается практически во всех странах, где цифровизация идет полным ходом. Планирование начинается с изделий-двойников, потом процесса их производства, то есть двойников предприятий, далее взаимодействия предприятий и т. д. В конечном счете, получается искусственное общество с экономикой — цифровым двойником. Это планирование хорошо тем, что можно составлять самые разнообразные варианты планов, из которых общество выбирает наилучший. В частности, прежде, чем объявлять противнику войну, надо проиграть все возможные варианты, при этом учитывать, что противник делает то же самое и, в конечном счете, война происходит лишь в виртуальной реальности.

Применительно к отдельному человеку можно рассуждать об индивидуальном устройстве типа смартфона, который содержит мир искусственного общества данного индивидуума. Человек, прежде чем принять окончательное решение, проигрывает в своем смартфоне различные варианты. То есть цифровое планирование переходит на индивидуальный уровень. А поскольку это касается всех людей, то общество таких людей будет выглядеть иначе, чем современное общество. Будут ли люди счастливее, это вопрос.

6. В цифровом обществе устройство группы агентов существенно усложняется, когда уже с меньшей натяжкой мож-

но говорить о самосознании и мышлении группы как целого. Примером может служить проект китайского правительства под названием «Система социального кредита» или доверия.<sup>1</sup> Суть ее состоит в том, что некий центральный орган в автоматическом и непрерывном режиме собирает информацию о действиях всех граждан. Чем совершеннее становится система слежения за гражданами, тем больше информации о каждом человеке накапливается. Например, уже сейчас во многих странах камеры слежения практически полностью отслеживают движение автомобилей и соответственно фиксируют любые нарушения.

На основании собираемой информации формируется текущий рейтинг граждан, который устанавливает ограничения на их поведение в обществе. В Японии также имеется план создания общества 5.0., где роль информационных технологий огромна. К сожалению, многоаспектного глубокого анализа будущего общества подобного типа пока нет. Хочется верить, что нас ждет большое разнообразие типов общества на планете, в рамках которого сохранятся культурные, исторические, языковые и прочие особенности, которые человечество не должно потерять.

То, что реализуется быстрая замена монотонных действий и осуществляющих их профессий на роботов и другие автоматические устройства, видно повсеместно. Имеется уже несколько исследований, касающихся тенденции перехода от простого труда к творческому.<sup>2</sup>

Появился новый термин «сетевой интеллект».<sup>3</sup> У группы видение мира отличается от видения отдельного человека.

В настоящее время назрело создание цифровой платформы для хранения, анализа, рейтингования и прочих научных результатов, которые публикуются в форме журнальных статей. В силу специфики современной науки узкие направления, которые полностью отслеживаются сравнительно небольшой группой

---

<sup>1</sup> Кириллов Андрей. ТАСС в КНР.

<sup>2</sup> Доклад Яковлевой Н.Г. на Научном Совете МГУ 15 сент. 2021 г.

<sup>3</sup> *Танскотт, Дон.* «Электронно-цифровое общество. Плюсы и минусы эпохи сетевого интеллекта» / Пер. с англ. 1996 г.

ученых, перестали быть сколько-нибудь ведущими и значимыми. Мировая тенденция ведет к комплексным исследованиям, которые невозможно отслеживать прочитыванием журнальных статей. Ситуация дополнительно осложняется устоявшимся порядком оценки научных результатов по количеству публикаций и ссылок. Необходим анализ текстов методом искусственного интеллекта, создание своеобразного компьютерного языка для фиксации научных результатов. Компьютерная программа должна определять степень новизны, количество новых идей, прорывной характер идей.

Определенные движения в указанном направлении есть. Например, ведущие цифровые гиганты типа Гугла имеют свои базы знаний, внутри которых оценки новых знаний несколько иные. Следует также отдельно выделять создание нового знания, доказательство и обоснование его правомерности с помощью в частности, дополнительных аргументов, а также распространение знания в обществе, в частности реклам.

Прогнозировать суть нового типа общества в эпоху цифровизации особенно трудно из-за многообразия новых направлений. Но ясно, что большинство профессий будут более творческими, более креативными. Причем это касается всех видов человеческой деятельности, включая не только культуру, искусство и другие креативные по определению направления, но и все виды развлечений, занятий для преодоления скуки и пр. Требование и лозунг простых римлян: «Хлеба и зрелищ» остается актуальным и в современную эпоху, разве что зрелища должны стать более креативными.

## ПРАВО, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ЦИФРОВИЗАЦИЯ

---

---

Появление и развитие технологий, именуемых искусственным интеллектом, а также их интенсивное применение в различных сферах человеческой деятельности и жизни общества стало своеобразным вызовом для права и юридической доктрины. В общем русле цифровизации и становления антропотехносферы процессы создания этих технологий, их усовершенствования и применения очень скоро стали демонстрировать социально значимые эффекты, указывающие на необходимость упорядочения соответствующих процессов и правового регулирования

---

\* *Хабриева Талия Ярулловна* — директор Института законодательства и сравнительного правоведения при Правительстве РФ, заместитель президента Российской академии наук, академик РАН, заслуженный деятель науки Российской Федерации, действительный член Международной академии сравнительного права (IACL, штаб-квартира — Париж, Франция), доктор юридических наук, профессор, заслуженный юрист Российской Федерации, заслуженный юрист Республики Татарстан.

Адрес: 117218, г. Москва, ул. Большая Черемушkinsкая, д. 34.

office@izak.ru Страница в сети Интернет: <https://izak.ru/institute/sotrudniki/khabrieva-taliya-yarullovna/>

*Khabrieva Taliya Y.* — Director of the Institute of Legislation and Comparative Law under the Government of the Russian Federation, Deputy President of the Russian Academy of Sciences, Member of the Russian Academy of Sciences, Doctor of Law, Professor, Titular Member of the International Academy of Comparative Law, Honored Lawyer of the Russian Federation, Honoured Lawyer of the Republic of Tatarstan.

office@izak.ru

<https://izak.ru/institute/sotrudniki/khabrieva-taliya-yarullovna/>

опосредующих их общественных отношений. Искусственный интеллект, как и ряд других продуктов технологической революции, послужил предпосылкой к формированию специфического предмета правового регулирования, что поставило перед юридической практикой задачу поиска адекватных правовых форм и методов регулирования.

К настоящему времени накопились факты для научного анализа эмпирики, позволяющей проследить реакцию российского и зарубежных государств, международных институтов на развитие нового сегмента в правовой сфере, а также оценить их подходы к регулированию общественных отношений, связанных с созданием и эксплуатацией технологий искусственного интеллекта.

Искусственный интеллект осваивается правоведами в качестве как прикладной (практико-ориентированной), так и фундаментальной проблемы. Есть основания констатировать появление самостоятельного научного направления в юридической науке, объединяющего теоретическую и практическую проблематику правового регулирования общественных отношений, связанных с созданием и эксплуатацией технологий искусственного интеллекта. Внутри него уже даже проведены классификации состоявшихся научных исследований, систематизирован широкий спектр вопросов<sup>1</sup>. В их числе, например, понятийно-категориальный аппарат права, необходимый для создания юридических форм в области применения искусственного интеллекта, правовой статус соответствующих технологий и их носителей, юридическая ответственность за вред, причиненный искусственным интеллектом, и т. д. Доктриной предложены вариативные правовые модели, позволяющие на текущем этапе придать наиболее важным общественным отношениям в области применения искусственного интеллекта правовую форму и купировать реальные и потенциальные риски для человека и общества.

---

<sup>1</sup> См.: Морхат П.М. Искусственный интеллект: правовой взгляд. М.: Буки Веди, 2017; Сеницын С.А. Российское и зарубежное гражданское право в условиях роботизации и цифровизации. Опыт междисциплинарного и отраслевого исследования: научно-практическое пособие. М.: Инфотропик Медиа, 2020.



Правовые исследования в этой области можно сгруппировать следующим образом:

1) общие вопросы правового обеспечения создания и использования искусственного интеллекта<sup>1</sup>;

2) исследования, направленные на поиск путей создания оптимального технологического, экономического и правового ландшафта для использования технологии искусственного интеллекта в социальной и юридической практике<sup>2</sup>. Разработки ведутся в области регулирования автономных и полуавтономных транспортных систем<sup>3</sup>, автоматизированных биржевых консультантов (*robo-advising*)<sup>4</sup>, банковской деятельности<sup>5</sup>, применения искусственного интеллекта в корпоративных процедурах<sup>6</sup>;

3) отраслевые правовые исследования, оценивающие возможности, последствия и риски опосредования технологий искусственного интеллекта и общественных отношений, связанных с его эксплуатацией различными отраслями права и законодательства (антимонопольного<sup>7</sup>, законодательства о защите

---

<sup>1</sup> См.: *Oskamp A. & Lodder A.R.* Introduction: Law, Information Technology, and Artificial intelligence // *Lodder A.R. & Oskamp A. (eds.) Information Technology and Lawyers.* Dordrecht, Springer. 2006. P. 1–22.

<sup>2</sup> См., напр.: *Palmerini E. et al.* RoboLaw: Towards a European framework for robotics regulation // *Robotics and Autonomous Systems.* 2016. Vol. 86. P. 78–85.

<sup>3</sup> См.: *Autonomous Vehicles — Self-Driving Vehicles Enacted Legislation* // National Conference of State Legislatures. 2018. URL: <http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx>

<sup>4</sup> См.: *Miller R. M.* Don't Let Your Robots Grow Up to Be Traders: Artificial Intelligence, Human Intelligence, and Asset-Market Bubbles // *Journal of Economic Behavior and Organization*, Forthcoming. URL: <http://dx.doi.org/10.2139/ssrn.415220>

<sup>5</sup> См.: *Wall L.D.* Some financial regulatory implications of artificial intelligence // *Journal of Economics and Business*, 2018. 100. P. 55–63.

<sup>6</sup> См.: *Möslein F.* Robots in the Boardroom: Artificial Intelligence and Corporate Law (September 15, 2017) // *Woodrow Barfield and Ugo Pagallo (eds), Research Handbook on the Law of Artificial Intelligence*, Edward Elgar, (2017/18, Forthcoming). URL: <https://www.law.ox.ac.uk/business-law-blog/blog/2017/11/robots-boardroom-artificial-intelligence-and-corporate-law>

<sup>7</sup> См., напр.: *Ezrachi Ariel and Stucke Maurice E.* Artificial Intelligence & Collusion: When Computers Inhibit Competition (April 8, 2015). University of Illinois Law

персональных данных<sup>1</sup>, об интеллектуальных правах<sup>2</sup>), с адаптацией к этому отраслевых и комплексных правовых институтов (юридическая ответственность, персональные данные, кибербезопасность и т. д.)<sup>3</sup>;

4) использование искусственного интеллекта в юриспруденции, в том числе для выполнения определенных задач юриста<sup>4</sup>;

5) комплексные исследования, нацеленные на разработку специализированного «права роботов», коррелирующего с кибернетической этикой и существующими программными правилами, имплементированными в машинные алгоритмы<sup>5</sup>.

---

Review. Vol. 2017. Oxford Legal Studies Research Paper. No. 18/2015, University of Tennessee Legal Studies Research Paper. No. 267, Available at SSRN: <https://ssrn.com/abstract=2591874>

<sup>1</sup> См., напр.: Artificial Intelligence, Robotics, Privacy and Data Protection. Room document for the 38th International Conference of Data Protection and Privacy Commissioners. 2016. URL: [https://edps.europa.eu/data-protection/our-work/publications/other-documents/artificial-intelligence-robotics-privacy-and\\_en](https://edps.europa.eu/data-protection/our-work/publications/other-documents/artificial-intelligence-robotics-privacy-and_en)

<sup>2</sup> См., напр.: *Yu R.* The Machine Author: What Level of Copyright Protection is Appropriate for Fully Independent Computer Generated Works? University of Pennsylvania Law Review. 2017. 165 (5). P. 1241–1270.

<sup>3</sup> См.: *Ashley K.D.* Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge, Cambridge University Press, 2017. Available from: doi: 10.1017/9781316761380

<sup>4</sup> См.: *Barfield W.* Issues of Law for Software Agents within Virtual Environments. Presence: Teleoperators and Virtual Environments. 2005. 14 (6). P. 741–748. Available from: doi: 10.1162/105474605775196607; *Sartor G.* Cognitive Automata and the Law: Electronic Contracting and the Intentionality of Software Agents. Artificial Intelligence and Law. 2009. 17 (4). P. 253–290; *Allen R.J.* Legal evidence scholarship meets artificial intelligence. Applied Artificial Intelligence. 2004. 18 (3–4). P. 367–389. Available from: doi: 10.1080/08839510490280003

<sup>5</sup> См.: *Balkin J.M.* The Three Laws of Robotics in the Age of Big Data // Ohio State Law Journal. 2017. Vol. 78. Forthcoming. Yale Law School, Public Law Research Paper N 592. URL: <https://ssrn.com/abstract=2890965>; *Leenes R.E., Lucivero F.* Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design // Law, Innovation and Technology. 2014. Vol. 6. № 2. P. 194–222. URL: <https://ssrn.com/abstract=2546759>

Изучается вопрос о субъектности искусственного интеллекта<sup>1</sup>, взаимодействии человека и робототехники в процессе принятия решений<sup>2</sup>. В частности, искусственный интеллект осваивается с позиций когнитивных наук, с точки зрения способности больших данных и искусственного интеллекта влиять на решения человека в культурном, экономическом и политическом контексте<sup>3</sup>. Отдельный блок поиска на стыке права и нейронауки<sup>4</sup> направлен на выяснение возможностей искусственного интеллекта в юридической аргументации<sup>5</sup>;

б) теоретико-методологические исследования бытия искусственного интеллекта в правовой сфере, создания правовой инфраструктуры взаимодействия человека и технологий искусственного интеллекта<sup>6</sup>.

---

<sup>1</sup> См.: *Saripan H.* Are Robots Human? A Review of the Legal Personality Model // *World Applied Sciences Journal*. 2016. 34 (6). P. 824–831. Available from: doi:10.5829/idosi.wasj.2016.34.6.15672; *Лантев В.А.* Понятие искусственного интеллекта и юридическая ответственность за его работу // *Право. Журнал Высшей школы экономики*. 2019. № 2. С. 87–90; *Понкин И.В., Редькина А.И.* Искусственный интеллект с точки зрения права // *Вестник РУДН. Серия: Юридические науки*. 2018. Т. 22. № 1. С. 91–109; *Eidenmueller H.* The Rise of Robots and the Law of Humans (March 26, 2017) // *Oxford Legal Studies Research Paper*. 2017. № 27. URL: <http://dx.doi.org/10.2139/ssrn.2941001>

<sup>2</sup> См.: *Solum L.B.* Legal Personhood for Artificial Intelligences // *North Carolina Law Review*. Vol. 70. P. 1231, 1992. Illinois Public Law Research Paper № 09–13. URL: <https://ssrn.com/abstract=1108671>

<sup>3</sup> См.: *Sunstein C.* Of Artificial Intelligence and Legal Reasoning (November 2001). University of Chicago Law School Roundtable. Vol. 8. 2001. URL: <http://dx.doi.org/10.2139/ssrn.289789>

<sup>4</sup> См.: *Kolber A.* Will There Be a Neurolaw Revolution? // *Indiana Law Journal*. 2014. Vol. 89. P. 807. URL: <https://ssrn.com/abstract=2398071>

<sup>5</sup> См.: *Feteris E., Kloosterhuis H.* Law and Argumentation Theory: Theoretical Approaches to Legal Justification (June 21, 2013). URL: <http://dx.doi.org/10.2139/ssrn.2283092>

<sup>6</sup> См., например: *Хабриева Т.Я., Черногор Н.Н.* Будущее права. Наследие академика В.С. Степина и юридическая наука. М.: Российская академия наук; Институт законодательства и сравнительного правоведения при Правительстве Российской Федерации; ИНФРА-М, 2020; *Юридическая концепция роботизации / отв. ред.*

Основные трудности, с которыми сталкиваются как ученые, так и практики, это, во-первых, неоднородность рассматриваемых технологий, одновременное развитие «сильного» и «слабого», «общего» и «специального», «высокого» и «низкого» искусственного интеллекта<sup>1</sup>, множественность форм искусственного интеллекта (как собственно технологии, так и технические системы, имеющие в своем составе или использующие искусственный интеллект — беспилотные автомобили и пр.<sup>2</sup>). Его разновидности потенциально и реально порождают разные социальные эффекты, несут различные риски для человека и общества, формируют запрос на учет их особенностей при разработке правовых механизмов, коррелирующих с возникающим социальным взаимодействием по поводу применения соответствующих технологий и систем. Отсюда развилка, на которой оказались как теоретики, так и практики и которая еще не пройдена. Необходимо сделать выбор либо в пользу унифицированного правового регулирования общественных отношений, связанных с созданием, совершенствованием (в том числе посредством машинного обучения и самообучения), распространением и эксплуатацией в пределах той или иной юрисдикции вне зависимости от специфики технологий, которые по тем или иным признакам могут быть идентифицированы как искусственный интеллект, либо в сторону дифференциации этого регулирования с учетом особенностей «интеллектуальных» технологий и их классификации.

Во-вторых, это высокая динамика развития технологий искусственного интеллекта и расширения сферы их применения. В настоящее время они используются в сферах оказания государственных услуг и публичного управления; правоохранительной,

---

Ю.А. Тихомиров, С.Б. Нанба. М.: Проспект, 2019; *Залоило М.В.* Искусственный интеллект в праве. М.: Инфортропик Медиа, 2021.

<sup>1</sup> См. об этом: Future of Artificial Intelligence Act of 2017 H.R.4625 US; European Commission. White paper on artificial intelligence—a European approach to excellence and trust. 2020 etc.

<sup>2</sup> NIST (2019) U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools: [https://www.nist.gov/system/files/documents/2019/08/10/ai\\_standards\\_fedengagement\\_plan\\_9aug2019.pdf](https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf).

судебной и таможенной деятельности, здравоохранении; автоматизированном транспорте, экологии, банковской деятельности, подборе персонала, на рынке недвижимости и др.

И наконец, в-третьих, это специфическая природа самого феномена этих технологий, которая обуславливает потенциальные возможности по-разному категорировать в праве и позиционировать в правовой сфере искусственный интеллект. Исследования показали, что искусственный интеллект обладает свойствами (или может их приобрести в перспективе), которые создают основания для того, чтобы с юридической точки зрения идентифицировать его как минимум в трех ипостасях:

1) в качестве объекта субъективного права или правового режима (например, объекта исключительных прав, определенного режима эксплуатации);

2) как инструмент правового регулирования — вспомогательного или основного. Впрочем, отнесение искусственного интеллекта к тому или другому виду не абсолютно. Данные технологии, применяемые в настоящее время как вспомогательный инструмент, имеют перспективы стать одним из основных средств регулирования. Изучение влияния так называемой цифровизации на социальное, в том числе правовое регулирование показало наличие у технологий искусственного интеллекта необходимого для этого потенциала. Строго говоря, инструментальный подход к позиционированию искусственного интеллекта в правовой сфере указывает на возможность идентификации этой технологии как объекта субъективного права или правового режима.

Так, технология, принятая к использованию государством для целей правоприменения (например, «электронный судья»), находится в чьей-то собственности, кто-то обладает исключительными правами на нее, она не может использоваться в правоприменении вне юридической формы. Однако речь идет об особенном объекте, так как он относительно автономно полностью или частично осуществляет функцию государства. Есть основания для предположения, что технологии искусственного интеллекта в перспективе будут использоваться (или примут участие) в правообразовании, выявлении, интерпретации и даже генерации политической

воли (в качестве общей воли, отражающей борьбу и согласование свободных волей в их взаимодействии и взаимообусловленности, государственной, классовой или иной). Предпосылки для этого уже складываются. Технологии искусственного интеллекта используются в правотворчестве, для прогнозирования последствий принимаемых нормативных правовых актов, выявления противоречий в законодательстве. И это только начало технологического переоснащения правотворчества. Поэтому с доктринальной точки зрения продуктивно рассматривать искусственный интеллект еще и в инструментально-правовом ракурсе, особенно в контексте перспективы машинизации права и его персонализации, на характеристике которых мы остановимся ниже;

3) в качестве субъекта права, как бы это ни шокировало юристов, отстаивающих постулаты классической юриспруденции и традиционные, проверенные столетиями юридические конструкции правосубъектности. В доктрине ведется острая полемика по этому вопросу, предложены модели правосубъектности искусственного интеллекта (особенно популярна в этом плане концепция электронного лица), а в практике известны случаи наделения юнитов искусственного интеллекта статусом гражданина.

Такая дисперсия правовой сущности искусственного интеллекта делает его если не уникальным, то, как минимум, неординарным явлением в праве. Выбор одного из этих акцентов или иного варианта категорирования искусственного интеллекта в праве (возможно, некоего синтетического, комплексного) определит модель правового регулирования соответствующих отношений и вектор его развития. В настоящее время этот выбор не сделан. Отсюда неспешность в создании норм позитивного права и международно-правовых регуляторов, интенсивное применение экспериментальных правовых режимов<sup>1</sup>, преимущественное

---

<sup>1</sup> См., например: Федеральный закон от 24.04.2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных».

по сравнению с правом использования этических норм<sup>1</sup> для упорядочения процессов создания и эксплуатации технологий искусственного интеллекта. Соответственно, преждевременно говорить о состоявшемся стабильном правовом регулировании, существовании его конвенционной концепции, четком понимании пределов и т. п., что в силу указанной причины на данном этапе развития технологий искусственного интеллекта и правового знания является нормальным. В формирующихся практикой прототипах моделей государственного регулирования, опосредующих развитие и использование технологий искусственного интеллекта общественных отношений преобладают акты государственного стратегического планирования<sup>2</sup>.

---

<sup>1</sup> См., например, разработанные в Европейском союзе Этические принципы надежного искусственного интеллекта, Европейскую этическую хартию об использовании искусственного интеллекта в судебных системах и окружающих их реалиях, принятые в Японии Этические стандарты для разработчиков искусственного интеллекта.

<sup>2</sup> Например, в России действуют Национальная стратегия развития искусственного интеллекта до 2030 года в Российской Федерации (утв. Указом Президента РФ от 10 октября 2019 г. № 490), Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года (утв. распоряжением Правительства РФ от 19 августа 2020 г. № 2129-р.) и др. В Японии — Стратегия развития робототехники 2015 года. В Южной Корее принята Хартия этических норм для роботов. В Объединенных Арабских Эмиратах реализуется Стратегия 2017 г. в сфере искусственного интеллекта. В Европейском Союзе принята Дорожная карта развития робототехники в Европе “Robotics 2020”, в 2018 году подписана Декларация о сотрудничестве в области искусственного интеллекта. В Китайской Народной Республике действует ряд документов — Глобальная государственная программа развития «Сделано в Китае — 2025», Тринадцатый пятилетний план развития на 2016–2020 годы, Руководство о содействии производству промышленных роботов 2014 г., План развития робототехнической отрасли 2016 г., План развития технологий искусственного интеллекта нового поколения 2017, на местном уровне — соответствующие программы развития. В США реализуется Дорожная Карта развития робототехники и Национальная робототехническая инициатива, в Канаде — Монреальская декларация об ответственном развитии искусственного интеллекта. Национальные стратегии развития искусственного интеллекта приняты в большинстве государств — членов Европейского Союза (Бельгии, Германии, Дании, Италии, Кипре, Латвии, Литве, Люксембурге, Мальте, Нидерландах, Португалии, Словакии Финляндии, Франции, Чехии, Швеции, Эстонии).

Важно обратить внимание на то обстоятельство, что технологии искусственного интеллекта являются не только «раздражителем» для государств и права, вызывающим определенную рефлексию, но и важным фактором эволюции права. К такому выводу приводит, прежде всего, следование в осмыслении процессов и перспектив взаимодействия права и искусственного интеллекта теоретическому посылу социально-экономической обусловленности права, а также его взаимосвязанности с технологическим укладом.

По оценкам ученых, мировое сообщество находится на пороге вступления в шестой технологический уклад и постиндустриальный период своего развития. Несмотря на то, что даже развитые государства только вступили в шестой технологический уклад и его протяженность пока неизвестна, можно прогнозировать, что его временные контуры будут ограничены непродолжительным периодом<sup>1</sup>. Предполагается, что для завершения формирования шестого уклада и создания заделов седьмого потребуется 15–20 лет<sup>2</sup>. Ядро шестого технологического уклада, по прогнозам ученых, составят «конвергирующие друг с другом, создающие сложные самоорганизующиеся комплексы нано-, био-, инфотехнологии, направленные на дальнейшее радикальное сращивание антропо- и техносред»<sup>3</sup>.

Принципиальным отличием седьмого технологического уклада от всех предыдущих, по оценкам специалистов, будет, в частности, то, что технологический способ производства составят интеллектуальные системы, принимающие решения без участия человека, «безлюдные» технологии, а рабочей силой будет

---

<sup>1</sup> См.: Глазьев С.Ю. Современная теория длинных волн в развитии экономики // Экономическая наука современной России. 2012. № 2. С. 30; Его же. О неравномерности современного экономического роста как процесса развития и смены технологических укладов // Социология. 2013. № 4. С. 42–53.

<sup>2</sup> См.: Рязанцев А.П., Рязанцева Д.А. Интеграционные процессы и технологические уклады // Национальные интересы: приоритеты и безопасность. 2014. № 16. С. 54.

<sup>3</sup> Асеева И.А. Аксиологические приоритеты VI технологического уклада // Эпистемология и философия науки. 2017. Т. 51. № 1. С. 129.



искусственный интеллект (раньше это был человек). В качестве ключевого фактора и средства производства прогнозируются когнитивные технологии и предиктивная аналитика. Основу функционирования хозяйственного механизма составят интеграционные структуры субъектов рынка (государства, частных корпораций, науки и образования), технопарки, а социальными последствиями такой модернизации станут машинное обучение, производственные процессы, налаженные без участия человека<sup>1</sup>. Такого рода очертания будущего технологического и экономического уклада жизни общества позволяют предположить существенные трансформации права и правового порядка, принимая во внимание известную закономерность о соотношении экономического базиса и надстройки, неизбежной трансформации последней вслед за экономикой. В свете таких прогнозов в несколько ином свете предстают идеи и полемика относительно правового статуса искусственного интеллекта в настоящее время и в некоторой перспективе, в том числе «проекты» его правосубъектности, которые многими учеными подвергаются резкой критике.

Они не добавляют оптимизма и в части потенциальных рисков для человека. Искусственный интеллект, претендуя на статус одного из главных «фигурантов» нового технологического и экономического укладов, имеет высокие шансы на конкуренцию с человеком. С экономической точки зрения человек эту конкуренцию проигрывает. Ее можно выдержать во многом при помощи правовой идеологии, этики и права, которые для защиты человека, вероятно, будут вступать в противоречие с экономическим базисом и технологическим укладом. Поэтому одной из важнейших задач юридической доктрины видится поиск путей преодоления этого противоречия в пользу человека. Возможными направлениями этого поиска могут стать аксиология и антропология права.

---

<sup>1</sup> См.: Вдовина А.А. Понятие «технологический уклад» в системе экономических категорий и новые технологические уклады общественного развития // Креативная экономика. 2019. Т. 13. № 4.

К умозаключениям о влиянии искусственного интеллекта на право приводят эмпирические исследования трансформации последнего и динамики правовой сферы. Если ранее технологическое развитие было внешним фактором генезиса права, то в настоящее время технологии начинают определять его существенные признаки и свойства. Так, исследования зафиксировали «точки роста» новых направлений в эволюции права, возникших на ее современном этапе.<sup>1</sup>

В их числе наметившиеся *дегуманизация* права, его *машинизация*, *персонализация* и *натурализация*. Эти направления не являются взаимоисключающими, скорее наоборот — они с разных позиций в определенной части характеризуют общий вектор развития права под влиянием технологий искусственного интеллекта и посредством его использования в социальной, в том числе в правовой регуляции.

Дегуманизация права проявляется в становлении идейной основы и практических попытках включения в состав субъектов права квазисубъектов «нового поколения» (нового вида), порожденных технологическим укладом «цифрового» общества. Истории права и государства (в том числе новейшей) известно немало примеров отнесения к субъектам права природных объектов (рек, растений, животных), неодушевленных предметов (например, бытовой техники). Другие социальные регуляторы (мораль, религия) никогда не демонстрировали ничего подобного. Но в условиях развития конвергентных технологий, в частности технологий искусственного интеллекта, доктрина и практика пошли по пути теоретического обоснования и практических экспериментов по наделению правосубъектностью так называемых цифровых сущностей, технических устройств, функционирующих на основе программных продуктов, технологий искусственного интеллекта. Так, в 2017 г. Европейский парламент в своей резолюции вместе с рекомендациями Комиссии «Нормы гражданского права о робототехнике» обозначил допустимость опре-

---

<sup>1</sup> См.: Хабриева Т.Я. Идентификация права в современной социальной регуляции // Вопросы философии. 2021. № 12. С. 5-17.

деления особого правового статуса для роботов в долгосрочной перспективе, чтобы хотя бы наиболее сложные автономные роботы могли быть наделены статусом электронных лиц, которые несут ответственность за свои действия и могут принимать независимые решения или иным образом независимо взаимодействовать с третьими лицами<sup>1</sup>. В этом же году робот по имени София стал(а) подданным(ой) Саудовской Аравии. Следует обратить внимание на то, что прав у нее в некотором смысле больше, чем у женщин в Саудовской Аравии. София не носит абайю, может общаться с мужчинами, и ее вряд ли арестуют за нарушение законов шариата.

Доктрина уже готова предложить правовые модели, обеспечивающие максимальную автономизацию юнитов искусственного интеллекта в социальных и правовых процессах. Так, их предлагается наделить правом на функционирование, энергосбережение, самообучение. Обсуждаются вопросы о распространении на новых «субъектов» прав человека (в их конституционном смысле), юридического оформления их правового статуса в качестве «электронного лица» (модели которого уже разработаны)<sup>2</sup>.

Наделение этих творений правосубъектностью объясняется прагматичными соображениями, в числе которых необходимость решения ставших уже очевидными проблем их правовой идентификации в социальном и правовом пространстве, упрощения и оптимизации применения к ним действующих правовых режимов, содействия интеграции и сотрудничеству между человеком и машиной<sup>3</sup>, а также тем, что спецификой программ, которые в процессе самообучения становятся независимыми от

---

<sup>1</sup> См.: European Parliament Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) // European Parliament. URL: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN>

<sup>2</sup> См.: Правосубъектность: общетеоретический, отраслевой и международно-правовой анализ: сборник материалов к XII Ежегодным научным чтениям памяти С.Н. Братуся. М.: ИЗиСП, Статут, 2017; Морхат П.М. Право и искусственный интеллект / под. ред. И.В. Понкина. М.: ЮНИТИ-ДАНА, 2017. С. 277, 303–358.

<sup>3</sup> См.: Морхат П.М. Указ. соч. С. 277–358.

своего создателя. Такие программы могут генерировать результаты, которые не охватывались замыслом разработчика, но стали возможны благодаря обобщающей способности лежащего в их основе алгоритма<sup>1</sup>.

Как в теоретической, так и в практической плоскостях вопросы правосубъектности искусственного интеллекта обсуждаются преимущественно в контексте трех сюжетов: право интеллектуальной собственности, ответственность за гражданско-правовые деликты, картельные сговоры. Так, в области права интеллектуальной собственности искусственный интеллект рассматривается как новый источник творчества, способный продуцировать новые результаты. С развитием машинного обучения искусственный интеллект научился продуцировать самые разнообразные результаты: изображения, видео, музыку, дизайн<sup>2</sup>. В доктрине обсуждается вопрос о признании результата, созданного искусственным интеллектом без вмешательства человека, потенциально охраноспособным, а в качестве правообладателя — саму технологию<sup>3</sup>. Сторонники признания исключительных прав за искусственным интеллектом предлагают модернизировать ключевую дефиницию права интеллектуальной собственности, распространив понятие «автор» не только на людей, но и на машины. Признание искусственного интеллекта автором созданного им результата позволит наделять искусственный интеллект исключительными правами на созданный результат<sup>4</sup>. Вопросы правосубъектности искусственного интеллекта дискутируются и в контексте проблематики

---

<sup>1</sup> См.: Войниканис Е.А., Семенова Е.В., Тюляев Г.С. Искусственный интеллект и право: вызовы и возможности самообучающихся алгоритмов // Вестник ВГУ. Серия: Право. 2018. № 4. С. 137.

<sup>2</sup> См.: Guadamuz A. Artificial intelligence and copyright // WIPO Magazine. 2017. № 5.

<sup>3</sup> См., например: Kalin H. Artificial Intelligence and the Copyright Dilemma (September 1, 2016) // IDEA: The IP Law Review. Vol. 57. 2017. № 3.

<sup>4</sup> Оппоненты этой идеи считают, что включение машин в число авторов будет иметь негативный эффект, так как внесет еще больше неопределенности в процесс правоприменения и породит больше вопросов, чем ответов (см.: Kalin H. Op. cit.).

патентного права<sup>1</sup>. Диапазон точек зрения достаточно широк — от предложений о точечной коррекции<sup>2</sup> патентного права до идеи его кардинального реформирования или отказа от включения в сферу действия патентного права изобретений, созданных искусственным интеллектом<sup>3</sup>.

Наиболее важной причиной актуальности обоснования соответствующих правотворческих решений является стремление перенести на сами юниты искусственного интеллекта бремя юридической ответственности за причинение вреда человеку, имуществу, охраняемым законом общественным отношениям. Причинение вреда жизни или здоровью человека в процессе эксплуатации систем с искусственным интеллектом из разряда теоретических предположений и потенциальных рисков перешло в разряд реальных фактов. Так, в 2018 г. беспилотная машина Uber сбила пешехода, переходившего улицу по пешеходному переходу<sup>4</sup>. Этот факт еще более актуализировал дискуссию об установлении иммунитетов для разработчиков программы, производителей системы с искусственным интеллектом и пользователей системы<sup>5</sup> и деликтоспособности технологий и систем, их использующих.

---

В качестве аргумента используется довод о отсутствии определенности в том, каким образом машины будут распоряжаться этими правами.

<sup>1</sup> См.: *Abbott R.* I Think, Therefore I Invent: Creative Computers and the Future of Patent Law // *Op. cit.* URL: <https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3522&context=bclr>

<sup>2</sup> См.: *Abramowicz M.* Orphan Business Models: Toward a New Form of Intellectual Property, 124 // *Harvard Law Review*. 2011.

<sup>3</sup> См.: *Yanisky-Ravid, Shlomit and Liu, Xiaoqiong (Jackie).* When Artificial Intelligence Systems Produce Inventions: The 3A Era and an Alternative Model for Patent Law (March 1, 2017). 39 *Cardozo Law Review*, 2215–2263 (2018), Available at SSRN: <https://ssrn.com/abstract=> or <http://dx.doi.org/10.2139/ssrn.2931828>.

<sup>4</sup> Uber settles with family of woman killed by self-driving car. URL: <https://www.theguardian.com/technology/2018/mar/29/uber-settles-with-family-of-woman-killed-by-self-driving-car>

<sup>5</sup> См.: European Commission. Workshop on liability in the area of autonomous systems and advanced robots and Internet of Things systems. 2017. URL: <https://ec.europa>.

*Практика использования технологий искусственного интеллекта на торгах породила такое явление, как «цифровые картели».* Важной особенностью искусственного интеллекта является то, что он действует самостоятельно по независимому от разработчика и пользователя сценарию. Так как результат программы искусственного интеллекта независим от разработчиков и пользователей программы, заранее сложно предсказать, приведет ли использование такой программы к сговору или же будет способствовать эффективному ведению бизнеса в рамках закона.

Ключевой проблемой для антимонопольного права применительно к цифровым картелям является то, что цифровые картели открывают возможность сговора без участия человека, в автоматическом режиме. Например, два хозяйствующих субъекта начинают использовать основанную на искусственном интеллекте программу с целью оптимизации ценообразования (например, с целью продавать машины по наиболее оптимальной цене). Однако программа в качестве самой оптимальной стратегии выбирает стратегию поддержания цены на уровне цены конкурентов. Итогом является сговор в автоматическом режиме, произошедший без участия самих хозяйствующих субъектов. Ключевая проблема, связанная с цифровыми картелями, состоит в поиске ответа на вопрос, кто является субъектом ответственности за цифровые картели, то есть за сговор в автоматическом режиме<sup>1</sup>. Или же квалифицировать данный факт как объективно противоправное деяние, а не правонарушение.

В доктрине отмечается, что ответственность искусственного интеллекта должна сводиться не к карательно-воспитательным мерам, а к установлению действенного механизма управления рисками<sup>2</sup>. Вместе с тем оценивается перспектива применения к

---

eu/digital-single-market/en/news/workshop-liability-area-autonomous-systems-and-advanced-robots-and-internet-things-systems; *Murphy R., Woods D.D.* Beyond Asimov: the three laws of responsible robotics. IEEE Intelligent Systems. 2009. Jul. 24 (4).

<sup>1</sup> См.: *Войниканис Е.А., Семенова Е.В., Тюляев Г.С.* Указ. соч. С. 145.

<sup>2</sup> См.: Правовые аспекты использования искусственного интеллекта: актуальные проблемы и возможные решения: доклад НИУ ВШЭ / рук. авт. кол: В.Б. Нау-

искусственному интеллекту мер уголовно-правового характера и появление в уголовных кодексах раздела, посвященного мерам уголовно-правового характера для электронных субъектов<sup>1</sup>.

Это далеко не полная характеристика практических и доктринальных предпосылок к развитию идей о правосубъектности искусственного интеллекта. Однако и их осмысление в сопряжении с формирующимся теоретическим обоснованием приводит к выводу, во-первых, о том, что юридическое оформление правосубъектности искусственного интеллекта — это вполне реальная перспектива.

Во-вторых, что реализация подобного сценария может привести к стиранию в праве фундаментальных различий между привычными субъектами и объектами правовых отношений — людьми и предметами. Такая тенденция уже обозначилась. Есть основания полагать, что на данном этапе человеческой истории право людей будет противостоять праву имущества на фоне продолжающейся дегуманизации права и его инфантилизации (расширения субъектного состава).

Рассматриваемый процесс еще нельзя признать состоявшимся трендом в эволюции права. Его проявления пока слишком незначительны для того, чтобы произошла революция в праве. Право по-прежнему антропоцентрично, нацелено на то, чтобы защитить человека от разного рода угроз, в том числе генерируемых стремительно развивающейся антропотехносферой. Однако его динамика сигнализирует о наличии предпосылок для изменения одной из фундаментальных основ права. Этому способствует развитие антропотехносферы вообще и технологий искусственного интеллекта в частности.

*Машинизация права (в терминологии зарубежной доктрины — алгоритмизация).* В правовом регулировании (как и в со-

---

мов, С.А. Чеховская, А.Ю. Брагинец, А.В. Майоров. М.: Издательский дом Высшей школы экономики, 2021 // 480106412.pdf (hse.ru). С. 12.

<sup>1</sup> См.: Radutniy O.E. Criminal liability of the artificial intelligence.doi: 10.21564/2414-990x.138.105661 UDC 343.22+343.412:004.056//cyberleninka.ru/article/n/criminal-liability-of-the-artificial-intelligence/viewer

циальном регулировании в целом) все больше используются алгоритмы, лежащие в основе программных продуктов. На данном этапе развития права и технологий это можно продемонстрировать на примере создания машиночитаемых норм (машиночитаемого права) — некоей машиночитаемой кодовой «реплики» текстуально выраженных юридических норм, то есть их модификации в программный код. Уже заявлены стратегические намерения разных государств о создании машиночитаемого права (в указанном смысле слова)<sup>1</sup>. Существуют документы стратегического планирования, предусматривающие создание такого рода массивов<sup>2</sup>, известны состоявшиеся эксперименты по переводу в машиночитаемый вид некоторых законов<sup>3</sup>.

Доктринальная характеристика этого направления развития права уже дана правоведами, причем как в статике (на основе изучения современного положения дел), так и в динамике (посредством ретроспективного анализа и прогнозирования следующих этапов алгоритмизации права). Учеными обозначены риски, практические и теоретические проблемы машинизации права, предложены решения некоторых из них<sup>4</sup>. Но это лишь «верхушка айсберга», начальный этап глубокой трансформации права и правового регулирования. Поступательное внедрение современных технологий обработки данных и искусственного интеллекта в социальное регулирование дает основания для прогнозирования превращения машинизации права из вспомогательного приема, позволяющего праву достигать своей цели, в магистральное на-

---

<sup>1</sup> См., например: Паспорт федерального проекта «Нормативное регулирование цифровой среды». URL: [https://www.economy.gov.ru/material/directions/gosudarstvennoe\\_upravlenie/normativnoe\\_regulirovanie\\_cifrovoy\\_sredy/](https://www.economy.gov.ru/material/directions/gosudarstvennoe_upravlenie/normativnoe_regulirovanie_cifrovoy_sredy/)

<sup>2</sup> 27 сентября 2021 г. Правительственная комиссия по цифровому развитию утвердила разработанную Минэкономразвития России Концепцию развития технологий машиночитаемого права. URL: [https://www.economy.gov.ru/material/news/v\\_pravitelstve\\_utverdili\\_koncepciyu\\_razvitiya\\_tehnologiy\\_mashinochitaemogo\\_prava.html](https://www.economy.gov.ru/material/news/v_pravitelstve_utverdili_koncepciyu_razvitiya_tehnologiy_mashinochitaemogo_prava.html)

<sup>3</sup> URL: [https://apolitical.co/solution\\_article/new-zealand-explores-machine-readable-laws-to-transform-government/](https://apolitical.co/solution_article/new-zealand-explores-machine-readable-laws-to-transform-government/); <https://www.digital.govt.nz/dmsdocument/95-better-rules-for-government-discovery-report/html>

<sup>4</sup> См.: Хабриева Т.Я., Черногор Н.Н. Указ. соч. С. 134–152.



правление его развития, сопряженное с приобретением им новых свойств и качественных состояний.

С машинизацией права тесно связана его *персонализация*, которая предполагает разработку и реализацию индивидуализированных «правовых норм» (под конкретных лиц), на основе обработки данных. При помощи новейших технологий может быть обработан значительный массив данных о конкретном индивиде и выявлены его поведенческие особенности, предпочтения, которые затем будут учтены в процессе разработки персонализированных правовых норм. Современные девайсы делают возможным оперативное «вменение» предписаний индивидам и доведение информации о них правоприменителям. В этой логике правовые предписания и право в целом могут быть весьма существенно (или даже полностью) персонализированы<sup>1</sup>.

Например, вместо установления общего для всех скоростного режима движения транспортных средств соответствующие режимы могут быть «спроектированы» под конкретных водителей на основании их опыта, водительской истории и дорожных условий в режиме реального времени<sup>2</sup>. Правореализация в отношении субъектов может происходить мгновенно посредством смартфонов, умных автомобилей и других устройств. Программа способна учесть все имеющие значение факты, моментально проанализировать, какое поведение допустимо «в каждом возможном сценарии», и, как следствие, максимизировать полезный эффект правового регулирования<sup>3</sup>.

---

<sup>1</sup> См.: *Grigoleit Hans Christoph and Bender, Philip Maximilian*. The Law between Generality and Particularity — Potentials and Limits of Personalized Law (December 13, 2019). Algorithmic Regulation and Personalized Law: A Handbook (Christoph Busch & Alberto De Franceschi eds., 2021). P. 115–136, Available at SSRN: <https://ssrn.com/abstract=3503376>

<sup>2</sup> *Elkin-Koren, Niva and Gal, Michal S.* «The Chilling Effect of Governance-by-Data on Data Markets» // *University of Chicago Law Review*: Vol. 86: Iss. 2, Article 6. 2019. P. 403–431. Available at: <https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/6>

<sup>3</sup> *Jordan M. Barry, John William Hatfield, and Scott Duke Kominers*. To Thine Own Self Be True? Incentive Problems in Personalized Law, 62 *Wm. & Mary L. Rev.* 723 2021. URL: <https://scholarship.law.wm.edu/wmlr/vol62/iss3/2>

Расширение использования технологий искусственного интеллекта в социальном регулировании, их возрастающая конкуренция с традиционными социальными регуляторами с высокой долей вероятности могут поставить под вопрос такое свойство права, как нормативность, или, как минимум, существенно изменить баланс между нормативным и индивидуальным правовым регулированием в сторону последнего.

Одним из фундаментальных признаков правовой нормы является то, что она рассчитана на абстрактного субъекта, на персонально неопределенный круг лиц. Однако в случае с персонализированным правом эта принципиальная черта правовой нормы целенаправленно преодолевается. Точнее будет говорить об индивидуально установленных правилах поведения лица в правовой сфере. Вероятно, произойдет утрата нормативности права. Сформулированные в общем виде правила поведения (нормы) уступят место индивидуальным предписаниям. Скорее всего, возникнет новый паттерн, первичный элемент системы права, который в современном «не-алгоритмизированном праве» представлен юридической нормой. Развитие баз данных и технологий искусственного интеллекта создает предпосылки для такого рода изменений.

По некоторым оценкам, персонализированное право может существенно повысить эффективность правового регулирования. В то же время оно несет в себе многочисленные риски, в числе которых: дискриминация субъектов права (в ее различных формах); препятствия для разрешения споров; ограничение свободы выбора для лица и ущемление свободы воли индивида; вытеснение договорного регулирования государственным; детерминированность содержания законодательных актов не дискуссией народных представителей, а решениями экспертов и ученых в области информационно-коммуникационных технологий; изменение концепции гражданина, обладающего определенным набором прав, подмена ее концепцией потребителя, характеризующегося лишь определенным набором предпочтений<sup>1</sup>.

---

<sup>1</sup> См.: Philip M. Bender. (2020). «Limits of Personalization of Default Rules Towards a Normative Theory», Working Papers tax-mpg-rps-2020-02, Max Planck In-

Перспектива персонализации права, ее выход на новый уровень во многом сопряжены с так называемой *натурализацией права*, под которой понимается использование в праве природоподобных технологий. В юридической науке это явление связывается главным образом с разработками нейронауки, охватывающей междисциплинарные исследования нервной системы человека и объединяющей достижения во многих областях научного знания.

Оптимистичный и вместе с тем идеализированный прогноз состоит в том, что в перспективе коллаборация права и нейронауки может привести к построению и практической реализации модели права, адаптированной к природе человека и способу функционирования его мозга, основанной на влиянии на работу мозга на нейронном уровне, воздействии на волю человека (свободу воли), эмоции и мотивационные процессы<sup>1</sup>.

С технологической точки зрения такой сценарий видится вполне возможным. Известны данные о разработках компании Илона Маска и его конкурентов (американской компании Synchron) по созданию интерфейса «мозг-компьютер». В США уже одобрены процедуры имплантации электродов в мозг для лечения болезни Паркинсона, эпилепсии, а некоторые компании продают переносные устройства для мониторинга мозговой активности, способные отслеживать настроение и эмоции. Компания Synchron получила разрешение Управления по санитарному надзору за качеством пищевых продуктов и лекарств

---

stitute for Tax Law and Public Finance. URL: <https://ideas.repec.org/p/mpi/wpaper/tax-mpg-rps-2020-02.html>; *Verstein Andrew*. (2019). Privatizing Personalized Law. University of Chicago Law Review: Vol. 86: Iss. 2, Article 1. Available at: <https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/1>; *Busch Christoph and De Franceschi Alberto*. Granular Legal Norms: Big Data and the Personalization of Private Law (March 06, 2018, 17 Pages). Forthcoming in Vanessa Mak, Eric Tjong Tjin Tai and Anna Berlee (eds), Research Handbook on Data Science and Law, Edward Elgar 2018, Available at SSRN: <https://ssrn.com/abstract=3181914>

<sup>1</sup> Опустим оценку и полемику по поводу данного термина. По нашему мнению, он не совсем удачный, однако достаточно широко, хотя и неоднозначно, используется в зарубежной доктрине (см., например: *Станек Ю.* Право и нейронаука — точки пересечения. URL: <https://cyberleninka.ru/article/n/pravo-i-neyronauka-tochki-peresecheniya/viewer>).

США (FDA) на проведение клинических испытаний вживляемого нейрочипа (устройства под названием Stentrode) на людях<sup>1</sup>.

Ученые отмечают, что проверенная на животных возможность изменения состояния нейронов означает принципиальную возможность вмешательства в мысли человека (например, не исключено, что нейротехнологии позволят менять предпочтения людей без их согласия для маркетинговых или других целей). Управление изображением, которое получает мозг через подключенное устройство, может показаться фантастическим сюжетом, но такие технологии уже реальность. По оценкам экспертов, самые фаталистичные технологии станут доступными в течение 10 лет. И хотя массовое распространение таких технологических новинок еще только в перспективе, бесспорно одно — исследования в области науки нейроправа и практическая реализация их результатов применительно к правовому регулированию способны коренным образом изменить подход к праву, его пониманию, характеристике свойств, правотворчеству и правореализации.

Так, право традиционно представляло собой нормативную систему (наряду с другими социальными регуляторами, такими, как мораль, религия). В этом качестве оно рассматривалось учеными и практиками (в том числе в соотношении с содержанием нормативов и практикой их применения). Развитие нейрокогнитивных технологий ведет к преобразованию права в *нормативно-когнитивную систему*, а соответствующая трансформация права — к формированию иной структурной организации его системы, внутренних и внешних связей, принципов и механизмов ее построения, развития, функционирования и т. д. Кроме того, натурализация права предполагает киборгизацию субъектов правоотношений, а возможно, и субъектов права в целом. Практическая реализация такого масштабного проекта требует решения ряда проблем не только фундаментального, теоретического и даже мировоззренческого порядка, но и прикладного характера. Например, формирование сети для подключения субъектов права или адаптация уже существующей, обеспечение

---

<sup>1</sup> URL: <https://trends.rbc.ru/trends/innovation/6107560d9a79474e216a2fc9>

ее локализации в рамках государства или глобализации (регионализации). При этом неизбежно потребуются идентифицировать сеть в правовом поле. Возможные варианты такой идентификации состоят в том, чтобы позиционировать ее в качестве институционально-технологической основы правопорядка (в этом случае сеть и технологии могут быть использованы для установления связи индивида с конкретным правопорядком, государством, для создания новой формы гражданства), объекта или субъекта правового регулирования. Нельзя исключать и комбинации решений, результатом которых будет некий «гибридный» статус подобной сети. Очевидно, что функционирование такой сети будет осуществляться если не на основе технологий искусственного интеллекта, то, как минимум, при их широкомасштабном использовании.

Теоретической проработки потребуют вопросы о правовом положении человека, не подключенного к сети; о степени обязательности подключения, его форме (например, должно ли оно быть постоянным на основе кибернейроподключения или переменным — путем использования коммуникаторов (смартфонов)). Предстоит поиск корреляций между диктуемыми «технологическим императивом» предпочтениями, технологическими решениями и фундаментальными правами человека, базовыми принципами права и конституционным регулированием. Осознание этой проблемы и возможной угрозы манипуляции с разумом при помощи технологий глубокой стимуляции мозга уже послужило мотивом для научных инициатив о внесении во Всеобщую декларацию прав человека «нейроправ», призванных защитить от злоупотреблений новыми технологиями. В их числе права на собственную личность, свободу воли, ментальную приватность, равный доступ к средствам расширения возможностей мозга и защиту от алгоритмической предвзятости<sup>1</sup>. В Чили разработан законопроект, нацеленный на защиту данных человеческого

---

<sup>1</sup> Речь идет об идее Р. Юсте (руководителя Morningside Group и профессора Колумбийского университета в г. Нью-Йорк) // <https://www.reuters.com/article/us-global-tech-rights-idUSKBN28D3HK>; <http://www.humanrightscolumbia.org/events/neurorights-human-rights-guidelines-neurotechnology-and-ai>

разума (нейроданных), установление ограничений на нейротехнологию чтения и письма в мозге, закрепление справедливого распределения и доступа к этим технологиям, наложение ограничений на нейроалгоритмы<sup>1</sup>.

С натурализацией права сопряжена трансформация когнитивной платформы правопорядка: появление специальных правовых интерфейсов «человек — машина» и «машина — человек»); сужение сферы действия принципа свободы воли. «Искусственный интеллект» лишен возможности коммуницировать с человеком непосредственно. Для такого взаимодействия необходим специальный интерфейс, функционирование которого должно обеспечиваться не только программно-аппаратными средствами, но также специально-правовыми и организационными. Специфика правового регулирования такова, что фактически потребуются создание не одного интерфейса, а двух, с тем чтобы обеспечить корректное включение «искусственного интеллекта» в аппарат правового принуждения.

В свою очередь, включение «искусственного интеллекта» в аппарат правового принуждения обуславливает изменение сферы действия принципа свободы воли, который является системообразующим для современного правопорядка. Появление в аппарате правового принуждения элемента, лишённого самоидентификации, подчиненного алгоритмам, повлечет сужение сферы действия названного принципа. Человек, наделенный волей, будет вынужден подчиняться не другому человеку, наделенному волей, а «машине», подчиненной алгоритму. Человеческая воля в данном случае будет реализовываться только опосредованно, через соответствующий алгоритм. И ее значение будет снижаться пропорционально уровню самообучения «машины».

---

<sup>1</sup> Законопроект от 7 октября 2020 г. (Boletín № 13.828–19 “Proyecto de ley, iniciado en moción de los Honorables Senadores señor Girardi, señora Goic, y señores Chahuán, Coloma y De Urresti, sobre protección de los neuroderechos y la integridad mental, y el desarrollo de la investigación y las neurotecnologías”). URL: <https://www.camara.cl/verDoc.aspx?prmID=14152&prmTIPO=INICIATIVA>; <https://www.senado.cl/defensa-de-los-neuroderechos-una-tarea-para-los-parlamentos-a-nivel-global/senado/2020-10-07/132033.html>

В практическом плане обозначенные векторы развития права (машинизация, персонализация, натурализация), вероятно, будут трудноотделимы друг от друга, так как в конечном счете они соотносятся друг с другом как процесс (машинизация) и результат (персонализация, натурализация, машиночитаемое право), условие (машинизация, натурализация) и свойство (персонализация), особенность будущего состояния права, его базовая характеристика, отражающая принципиально новый модус правотворчества, и реализации права, в том числе его применения. Однако все они непосредственно связаны с технологиями искусственного интеллекта, без которых невозможна реализация указанного сценария.

Одним из совокупных эффектов, сопровождающих развитие права, станет его *сакрализация* — доступность и понятность только «посвященным», оборотной стороной которой будет недоступность его постижения для многих (если не большинства) граждан и даже законодателей в силу закрытости алгоритмов, отсутствия специальных знаний в области языков программирования, правовых ограничений доступа к содержанию алгоритмов и технологиям, выполняющим регуляторные функции. В качестве одного из следствий сакрализации права можно ожидать кризис доверия граждан к праву, институтами механизмам правореализации, в том числе правоприменения.

Данный сценарий развития права непременно повлечет изменения в свойствах права, таких как нормативность и формальная определенность. Еще более важно, что «под давлением» окажутся принципы права — гуманизм, равенство и справедливость, а также фундаментальные права человека — на свободу (личную, воли, выбора), неприкосновенность частной жизни, достоинство личности.

Утрата правом нормативности по указанным выше причинам или существенное изменение баланса между нормативным и индивидуальным правовым регулированием в пользу последнего обострит проблему равенства в праве. Такого рода метаморфозы права, в свою очередь, повлекут изменение еще одного важного основания права — принципа справедливости, концепт которого (вырабатанный доктриной и судебной практикой) соединяет в себе идеи

равенства и соразмерности<sup>1</sup>. Возможная девальвация равенства в праве повлечет изменения содержания принципа справедливости.

Формальная определенность права, во всяком случае в привычном для юристов понимании, тоже может приобрести алгоритмическую форму или уступить место какому-либо из свойств алгоритмов.

«Точки роста» указанных направлений маркируют контур новой «парадигмы» развития права, его системы координат, приблизительное очертание которой уже можно обозначить. Дальнейшая эволюция права обусловлена становлением гибридного правового ландшафта — реальности социально значимого взаимодействия человека, адаптированного к новой среде обитания, в том числе посредством собственного технологического апгрейда, гибридов природы и культуры (квазиобъектов или «субъект-объектов»), «преодолевающих границы между культурой и природой, деятелем и материалом»<sup>2</sup>.

Одной из общих закономерностей развития права станет его гибридизация в указанном смысле, в результате конвергенции с технологическими инновациями и практиками. При этом конвергенция будет происходить на уровне первичных элементов, «элементарных частиц» (в терминологии философов — конвергенция на «наноуровне»<sup>3</sup>), когда нормы права и единицы информации, юридические нормы и алгоритмы станут взаимозаменяемыми. Такой прогноз дает повод для размышлений о конвергенции содержания и формы права, нормативного и индивидуального правового регулирования, а также основание для утверждения о том, что черты и свойства гибрида приобретут не только право и его структуры, но и правовая реальность. Совокупным эффек-

---

<sup>1</sup> См.: Зорькин В.Д. Справедливость — императив цивилизации права // Вопросы философии. 2019. № 1. С. 6.

<sup>2</sup> Дэвис Э. Техногнозис: мир, магия и мистицизм в информационную эпоху. Екатеринбург: Ультра; Культура, 2008.

<sup>3</sup> См.: Аршинов В.И., Буданов В.Г. Парадигма сложности и социогуманитарные проекции конвергентных технологий // Вопросы философии. 2016. № 1. С. 59–70.



том от действия факторов механизации, персонализации и натурализации права будет превращение права из нормативной в автоматизированную нормативно-когнитивную систему.

Данный сценарий развития права непременно повлечет изменения в свойствах права, таких как нормативность и формальная определенность. Еще более важно, что «под давлением» окажутся принципы права — гуманизм, свобода воли, равенство и справедливость, а также фундаментальные права человека — на свободу (личную, воли, выбора), неприкосновенность частной жизни, достоинство личности. Может измениться и функционал права — на смену регулированию общественных отношений, согласованию свободных волей, интересов и ценностей, разрешению конфликтов придет регулирование поведения индивидов, исключающее столкновения и конфликты.

Перспективы изменения «парадигмы» эволюции права, его развития в новой логике, преобразования свойств и значения в жизни общества весьма реальны. Будущее состояние права и идентифицирующие его маркеры во многом будут зависеть от того, что станет с идеей права, когда цивилизационный переход будет завершен, какое место в системе социальных ценностей займут идеи гуманизма, свободы воли, равенства и справедливости, являющиеся базовыми для права, в соотношении с упомянутыми технологиями и сохранятся ли они в качестве исходного начала в регулировании общественных отношений и поведения людей. Именно данное обстоятельство определит вектор дальнейшей эволюции права. В связи с этим важной задачей юридической науки видится проблематизация ценностных оснований права, выявление и фиксация их смыслов и корреляций с ценностями формирующегося общества.

## ВОПРОСЫ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ И СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ВОЕННОЙ СФЕРЕ

---

---

### Общие вопросы

Многие ученые и эксперты исходят из того, что искусственный интеллект — свойство автоматических систем брать на себя отдельные функции интеллекта человека, например выбирать и принимать оптимальные решения на основе обработанного должным образом полученного опыта и рационального анализа внешних воздействий. Один из отечественных классиков исследований в области ИИ В.К. Финн справедливо пишет о том, что ИИ — это направление, в котором разрабатываются и средства имитации, и средства усиления умственной деятельности человека<sup>1</sup>.

На обозримую перспективу, по-видимому, речь идет о т. н. «слабом искусственном интеллекте»; вопросы применения «сильного искусственного интеллекта» практически единодушно относятся к более отдаленному периоду времени. «Сильный искусственный интеллект» (в Национальной стратегии развития ИИ

---

\* *Кокошин Андрей Афанасьевич*, академик РАН, директор Института перспективных стратегических исследований НИУ ВШЭ.

*Kokoshin Andrey Afanasievich* — Academician of the Russian Academy of Sciences, Director of the Institute for Advanced Strategic Studies of the HSE, Moscow, Russia.

<sup>1</sup> *Финн В.К.* Искусственный интеллект. Методология. Применения. Философия. М.: КРАСАНД, 2011. С. 7.

в РФ назван как «универсальный (сильный) искусственный интеллект»)<sup>1</sup> — максимально возможно имитирующий поведение человека, способный решить широкий круг задач. Под «слабым искусственным интеллектом» подразумеваются технологии и системы ИИ, призванные решать различные сравнительно узко специализированные задачи.

В результате анализа и обобщения планов и практических действий по использованию технологий искусственного интеллекта для нужд вооруженных сил различных зарубежных государств можно выделить ряд формирующихся сфер применения таких технологий. Для некоторых направлений технологии и даже системы ИИ в интересах вооруженных сил, военных ведомств могут быть с небольшими доработками заимствованы из гражданской сферы с обеспечением необходимых мер специфики и безопасности (секретности, защищенности).

Считается, что оружие, оснащенное перспективными системами искусственного интеллекта, может существенно повысить «операционную гибкость» в применении новейших средств поражения, позволяя командирам адаптировать огневые, радиоэлектронные, кибернетические средства поражения под специфические и быстро меняющиеся условия войн и вооруженных конфликтов. Но применение ИИ, о чем речь пойдет далее, не сводится только к оружию.

### **Направления (сферы) возможного применения технологий (элементов) и систем ИИ**

В настоящих условиях на основе исследования планов, программ и конкретных мероприятий по развитию технологий ИИ в КНР, США, Индии, Франции и в ряде других стран представляется возможным идентифицировать следующие направле-

---

<sup>1</sup> Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (а также «Национальная стратегия развития искусственного интеллекта на период до 2030 года»).

ния (сферы) возможного применения технологий (элементов) и систем ИИ.

- **Обеспечение более эффективного проведения «операций в киберпространстве».**

Многими зарубежными специалистами отмечается, что технологии ИИ могли бы брать на себя все более значительную часть работы высококвалифицированных специалистов по «кибервойне», так как средства ИИ потенциально способны значительно быстрее находить уязвимости в киберсетях и разрабатывать коды и машинные алгоритмы.

В ряде зарубежных исследований отмечается, что «кибероружие», управляемое ИИ, может поражать киберсети противника строго дозировано, в заранее определенных масштабах, действуя при этом в автоматическом режиме. (Это, в том числе, относится к воздействию на сети, управляющие критически важной инфраструктурой, включая энергетическую.) Одновременно говорится о том, что средства ИИ призваны оказывать противодействие «кибератакам» на объекты военной инфраструктуры, на информационно-коммуникационные сети, на различные ударные средства, включая пилотируемые самолеты и БПЛА.

Считается, что технологии ИИ в сфере кибербезопасности позволяют осуществлять не просто мониторинг, но и анализ сетевой обстановки в реальном масштабе времени, в том числе обеспечивая обнаружение попыток кибератак и загрузки вредоносных программ. Одна из задач для систем ИИ состоит в обнаружении признаков аномального поведения в соответствующих сегментах киберпространства и в устранении в режиме реального времени возникающих уязвимостей.

- **Развитие интеллектуальных средств и методов радиоэлектронной борьбы (РЭБ),** которые, в том числе, как обнаруживали бы более оперативно и надежно источники излучения, так и маскировали бы их путем разбрасывания ложных источников облучения.

- **Повышение боевых возможностей для беспилотных летательных аппаратов** в плане ускоренной и более качествен-

ной обработки разведданных, повышения ситуационной осведомленности операторов БПЛА.

*В США одним из наиболее продвинутых в данной области является масштабный проект «Мейвен», стартовавший несколько лет назад. В рамках данного проекта с использованием технологий искусственного интеллекта осуществляется анализ больших объемов видеоматериалов, поступающих с разведывательных БПЛА, его обобщение и представление оператору-аналитику обработанной информации для последующего формирования «базы целей», предназначенных для уничтожения. Конечная информация должна поступать в виде целеуказаний на ударные БПЛА.*

• **Повышение возможностей пилотируемых самолетов ударной авиации, в том числе путем создания с использованием ИИ принципиально новой интегрированной человеко-машинной системы управления воздушным боем.** Здесь одним из наиболее продвинутых считается программа ММТ Министерства вооруженных сил Франции. Она включает в себя такие компоненты, как «виртуальный помощник» — «умная кабина», «человеко-машинное взаимодействие», «управление миссией», «интеллектуальные датчики», «роботизированная поддержка и обслуживание» и др.

*В этой системе должно быть задействовано большое число датчиков, предоставляющих полномасштабные данные о техническом состоянии узлов и агрегатов летательного аппарата, о воздушной обстановке, о состоянии членов экипажа. Предполагается, что такая система будет иметь потенциал для развития и «глубокого обучения». У экипажа должен быть «виртуальный помощник», который будет способен восполнять «низкоуровневые операции», разгружая летчиков от ряда второстепенных задач. Это должно предоставить возможность экипажу сосредоточиться на тактических решениях по выполнению боевой задачи, имея максимальную осведомленность о ситуации. Большая роль отводится «интеллектуальным датчикам», которые должны приспосабливаться к изменениям потребностей в управлении, связанным с динамичными изменениями обстановки и эксплуатационными потребностями.*

• **Повышение эффективности работы автономных систем** (воздушных, морских, наземных) в сложных условиях (пересеченный ландшафт, высокая влажность, разная скорость ветра и состояние моря, подавление электромагнитного излучения и др.), в том числе посредством оптимизации человеко-машинного взаимодействия.

• **Совершенствование технологий управления «роем» беспилотных аппаратов**, с тем чтобы уменьшить число задействованных операторов, время, необходимое на их подготовку, а также нагрузку на каждого оператора во время осуществления операций.

*Многие специалисты отмечают, что на протяжении сравнительно длительного времени дешевых, надежных и высокоскоростных средств обмена информацией между компонентами «роя» не было, однако исключительно быстрое развитие телекоммуникационных средств за последние годы стало радикально менять ситуацию в человеко-машинной системе управления воздушным боем.*

В то же время, как считают многие эксперты, на пути создания «роевых группировок» встают в том числе вопросы о сохранении полного контроля над их действиями со стороны человека. (Это относится и к автономным необитаемым подводным аппаратам, способным оказать существенное влияние на вооруженную борьбу на море.)

• **Повышение уровня ситуационной осведомленности командиров и командующих, штабов различных уровней.**

Обработка с повышенной скоростью больших объемов данных: видео, аудио, текстовых сообщений, данных с РЛС, разведывательных спутников и др. Анализ изображений со спутников и тепловизоров, обработка аудио и видеоизображений в реальном времени; распознавание объектов, отслеживание движений; распознавание лиц и голоса. Считается возможным за счет технологий ИИ обеспечение полной автоматизации некоторых видов разведки, в частности радиоразведки.

Считается, что на более продвинутом уровне за счет технологий ИИ может осуществляться синтез и сопоставление

данных радиоэлектронной и оптико-электронной разведки, сообщения агентурной разведки, от разведки, работающей в киберсетях, от обработки открытых источников информации с применением средств анализа «больших данных». В целом речь идет о совмещении в системах ИИ информационных сетей, использующих различную семантику и различные стандарты обмена данными.

Более эффективное решение за счет использования ИИ, в т. ч. задачи по распознаванию подготовки внезапного нападения; вскрытие с помощью ИИ дезинформации, ее отфильтровывание.

Повышение уровня информируемости о положении дел с собственными силами и средствами с активным использованием средств и технологий, аналогичных тем, что используются применительно к задачам разведки, целеуказания.

• **Создание систем поддержки принятия решений (СППР) для боевого управления** (на тактическом, оперативном и стратегическом уровнях). СППР при этом должны опираться на обширную базу соответствующих прецедентов. С этим связывают создание, в том числе, перспективных компьютерных систем с ИИ для проведения военных игр тактического и оперативного уровня, а также игр политико-военного уровня с полномасштабным учетом, в том числе, информационного противоборства и борьбы в киберпространстве.

Многие специалисты обоснованно отмечают, что использование неверных прецедентов может вести к опасным ошибкам в работе СППР.

*В Минобороны США, в том числе, рассматривают вопросы создания СППР для проведения гуманитарных операций и действий контингентов Вооруженных сил США, когда они привлекаются к борьбе со стихийными бедствиями.*

• **Разработка «цифрового помощника для солдат»** («виртуального собеседника»), который позволит военнослужащим вплоть до уровня подразделений получать нужную им информацию о тактике ведения боя или инструкции по ремонту техники в режиме реального времени с использованием рации и электронных гаджетов.

- **Создание персональных автоматизированных переводчиков**, в том числе устной речи, для общения с местным населением на соответствующих языках.

- **Прогнозирование состояния здоровья военнослужащих и гражданского персонала**, включая анализ медицинских показаний, записей врачей, классификацию медицинских снимков, данных разнообразных анализов, оценку последствий посттравматических стрессовых расстройств, анализ, нацеленный на предотвращение самоубийств и т. п.

- **Использование искусственного интеллекта для развития телемедицины**, для повышения качества проводимых хирургических операций военными медиками.

- **Создание тренажеров-имитаторов с применением технологий виртуальной реальности**, максимально приближенной к реальным условиям, — для подготовки пилотов ВВС, операторов ЗРК, танкистов и пр., а также для офицеров — операторов по управлению разнородными силами и средствами.

*Значительным достижением Израиля считается создание на основе технологий искусственного интеллекта тренажеров тактического и оперативного уровня (Command and Staff Training — CST) для проведения «объединенных операций», в том числе в очень сложных городских условиях, вплоть до операций без прямого применения средств поражения («операции вне войны»).*

- **Повышение эффективности систем тылового, в том числе материально-технического обеспечения (логистических систем)**, включая оптимизацию цикла пополнения складов с учетом стоимости поддержания на определенном уровне товарно-материальных запасов. Решение задачи сокращения масштабов потерь в деятельности органов тылового, материально-технического обеспечения. Оптимизация логистических и транспортных планов.

- **Предиктивное ремонтное обслуживание.** Использование данных с сенсоров, установленных на танках, грузовиках и дру-



гих видах техники, об их состоянии. Выявление неработающих и работающих неоптимально систем.

Содействие поддержанию оптимального уровня функциональности данных систем. Упреждающее оповещение о необходимости ремонта конкретных единиц техники или их отдельных агрегатов и деталей.

- **Совершенствование бюджетно-финансового планирования и бухгалтерского учета с эффективным мониторингом всех статей военных расходов** в целях оптимизации распределения средств между различными структурами военного ведомства, компонентами вооруженных сил в соответствии с определенными приоритетами военного строительства, в целях повышения эффективности расходов на военные НИОКР и закупки вооружений, военной и специальной техники (ВВСТ), на подготовку кадров и др.

- **Оптимизация за счет использования ИИ кадровой работы** на всех уровнях вооруженных сил, процессов обеспечения военными кадрами и гражданским персоналом вооруженных сил.

- **Персонализация подготовки военнослужащих** и оценки ее результатов.

\* \* \*

Многие специалисты обращают внимание на то, что сохраняется необходимость интерпретации человеком (высококвалифицированным оператором) тех результатов, которые получаются в результате функционирования систем ИИ. Это еще раз подчеркивает важность развития такого направления, как «человеко-машинное взаимодействие», необходимость активной подготовки соответствующих кадров, так как системы ИИ для использования в вооруженных силах должны соответствовать уровню подготовленности соответствующих операторов, и наоборот (то есть системы ИИ не должны становиться для оператора «черным ящиком»).

**Вопросы контроля человеком (оператором) систем ИИ, обеспечение надежности и безопасности применения оружия с использованием ИИ** можно считать среди наиболее сложных при рассмотрении проблем широкого внедрения технологии и систем ИИ. Особенно это относится к средствам поражения личного состава, техники, других материальных объектов противника, что тесно связано с проблемами гуманитарного знания, в том числе с политико-правовыми вопросами.

Подавляющее большинство экспертов и должностных лиц сходятся в том, что любые решения о применении ударных (летальных) средств должны быть в руках исключительно человека.

## ВЫЗОВЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПСИХОЛОГИИ\*\*

---

---

Из всей совокупности общих для психологии и искусственного интеллекта тем ниже будут рассмотрены две, наиболее существенные. Первая относится к сходствам и различиям устройства естественного и искусственного интеллекта, как их видят психология и информатика. Вторая — к психологическим проблемам, сопутствующим проникновению в жизнь технологий искусственного интеллекта.

### Искусственный и естественный интеллект: анalogии и различия

Технические устройства нечасто воспроизводят естественные биологические структуры. Подъемные силы самолета создаются не так, как у птицы, колесо автомобиля не похоже на конечности животного, а плавники рыбы — на винт подводной лод-

---

\* *Ушаков Дмитрий Викторович* — академик РАН, директор Института психологии РАН.

*Ushakov Dmitry Viktorovich* — Academician of the Russian Academy of Sciences, Director of the Institute of Psychology of the Russian Academy of Sciences.

*Валужева Екатерина Александровна* — научный сотрудник Института психологии РАН, кандидат психологических наук.

*Valueva Ekaterina Aleksandrovna* — is a researcher at the Institute of Psychology of the Russian Academy of Sciences, Candidate of Psychological Sciences.

\*\* Работа выполнена при поддержке РФФИ, грант № 18-29-22095мк.

ки. Искусственный интеллект в реализованных на сегодняшний день проектах добивается эффективности решения задач в большинстве случаев не теми средствами, что человек, и опирается на другие вычислительные ресурсы и «элементную базу». Во многих областях искусственному интеллекту еще очень далеко до рационального поведения человека, в то время как в других, например, в игре в шахматы или го, он уже намного ушел вперед. Однако трудно указать такую отрасль, где интеллектуальная машина делала бы в точности то и так, что и как человек. Если ей удастся догнать человека, она не останавливается и, опираясь на вычислительное быстроедействие, вырывается вперед.

Тем не менее, сколь ни парадоксально это звучит на фоне только что сказанного, между исследованиями в двух областях прочерчиваются глубокие параллели, которые сопровождаются двухсторонними заимствованиями и перетеканием идей и принципов. Возьмем нейронные сети глубокого обучения, с которыми связан последний бум в области искусственного интеллекта. Корни идеи нейронных сетей уходят вглубь времен, далеко предшествующих формированию наук об искусственном интеллекте. Согласно философско-психологическому ассоцианистскому подходу<sup>1</sup>, ассоциативные связи составляют сеть, пронизывающую всю совокупность идей, которыми располагает человек. Эта идеология составляет фактически начало пути к современным сетям глубокого обучения, хотя и содержит довольно сильные отличия<sup>2</sup>.

Дальнейшее движение по этому пути шло силами как психологов, так и психофизиологов: взаимодействие психологии с искусственным интеллектом нельзя представить вне биологии, в первую очередь таких ее разделов, как психо- и нейрофизиология. В конце XIX — первой половине XX века трудами

---

<sup>1</sup> Идеи ассоцианизма при более внимательном изучении можно найти еще в Античности — у Аристотеля и даже Платона, но подробное рассмотрение этого вопроса выходит за рамки данной статьи.

<sup>2</sup> Так, узлы сетей ассоцианистов связаны с определенными понятиями, в то время как в промежуточных слоях сетей глубокого обучения узлы работают на субсимволическом уровне, а понятия кодируются распределенно, связаны с несколькими узлами.

И.П. Павлова, В.М. Бехтерева, Э. Торндайка, Б.Ф. Скиннера и др. представления об ассоциировании были распространены (через понятия условного и оперантного научения) с области идей на область поведенческих актов, обучения и обогащены впечатляющим корпусом экспериментальных данных.

Следующий шаг был сделан в 1949 г. нейропсихологом Д. Хеббом, который наложил ассоцианистские и рефлекторные идеи на сформировавшиеся к тому времени представления о нейронах и их связях, введя понятия о гипотетическом синапсе, который формирует связь между нейронами в случае совпадения моментов их разрядов. Тем самым фактически предлагался первый алгоритм обучения нейронных сетей, делался последний шаг к идеологии распределенного хранения знания — поведенческий акт или понятие связаны в памяти с работой целой группы нейронов, возможно расположенных в различных отделах мозга.

Сам Хебб, однако, не пытался исследовать, какими способностями будет обладать сеть, собранная из описанных им нейронов. Собственно и сегодня нет каких-либо оснований считать, что такая сеть способна обучиться распознавать конфигурации или совершать другие важные когнитивные действия. Первой моделью сети, показавшей способность к распознаванию, стал в 1957 г. Персептрон, изобретатель которого Ф. Розенблатт при всей разносторонности его интересов может быть охарактеризован скорее как нейрофизиолог или психолог, чем как математик или программист.

Персептрон фактически открыл эру нейроморфных устройств, способных к обучению. Наложением физиологической концепции на компьютерное устройство было показано, что построенная по определенным принципам нейронная сеть может научиться распознаванию конфигураций. При этом алгоритм обучения Персептрона — метод коррекции ошибки — не совпадал с работой нейрона Хебба. Нейрон Хебба образует новую связь под воздействием положительного подкрепления, в то время как алгоритм Персептрона меняет связи только в случае ошибки, отрицательного подкрепления.

Дальнейшее развитие Персептрона — многоуровневые сети глубокого обучения — позволили добиться выдающихся результатов в плане функционирования, но еще дальше ушли от подобия работе мозга. Обратное распространение ошибки — алгоритм обучения многоуровневых сетей, предложенный исходно А.И. Галушкиным и П. Вербосом, подхваченный и развитый такими специалистами, как Д. Румельхарт и Д. Хинтон и позволивший достичь поразительных результатов в области машинного обучения — вряд ли аналогичен нейрофизиологическим процессам. Если бы этот алгоритм соответствовал биологическим реалиям, то в человеческом мозге должны были бы присутствовать мощные пути обратного распространения ошибки, с чем физиологи, однако, не встречаются.

Синапс Хебба, напротив, до сих пор продолжает казаться физиологически правдоподобным. Так, данные, полученные с помощью оптогенетических технологий, свидетельствуют, что активация нейрона световым воздействием в то время, как организм совершает какую-либо поведенческую реакцию (например, оборонительную), приводит к тому, что последующая активация этого нейрона вызывает ту же реакцию. В то же время модели самообучающихся сетей, построенных по принципу синапса Хебба, создать не удалось, и неясно, как это можно было бы сделать.

Нейросети, обучающиеся при помощи механизмов обратного распространения ошибки, порой показывают удивительно «человекоподобные» результаты даже там, где речь идет о неправильном или ошибочном поведении. Так, например, в классических работах Д. Румельхарта и Д. Макклелланда было показано, что при выучивании прошедших времен английских глаголов нейросеть демонстрирует те же стадии развития, что и англоязычный ребенок.

Таким образом, пример нейросетей показывает сложную и неоднозначную картину: идеи психологов и разработчиков искусственного интеллекта взаимодействуют, взаимно используются, технологии искусственного интеллекта в некоторых моментах оказываются очень похожими на естественные когнитивные системы, но при этом всегда сохраняется несовпадение, обуслов-

ленное тем, что психологи обязаны следовать фактам об объекте их исследований, а специалисты по искусственному интеллекту движимы стремлением к созданию наиболее эффективно работающих систем.

Еще один пример подобного частичного совпадения нейрофизиологических процессов и машинного обучения представляет зрительная кора. Д. Хьюбел и Т. Визель получили в 1981 г. Нобелевскую премию за открытие нейронов-детекторов, отвечающих на опознание определенных зрительных стимулов, например, линии с заданным углом наклона. Эти нейроны имеют разные рецептивные поля: некоторые из них отвечают на релевантные стимулы лишь в некоторых частях зрительного поля, другие — по всему полю. На этой основе сложилась четкая картина работы нейросетей в сфере зрительного восприятия, на которой были построены т. н. сверточные (convolutional) сети: первичные слои сети отвечают на точечные стимулы, далее распознаются отдельные признаки типа прямых линий в отдельных частях зрительного поля, еще более высокие слои обнаруживают эти признаки по всему полю. Сверточные сети показали себя весьма успешными в решении задач искусственного интеллекта, особенно в задачах распознавания видеоизображений. Однако недавний анализ баз данных по функционированию клеток зрительной коры показал, что лишь примерно 10% из них работают по тем принципам, которые установили Д. Хьюбел и Т. Визель.

Таким образом, вероятно, что нейронные сети в мозге человека способны к самообучению по механизмам, в чем-то сходным, но не совпадающим полностью с современными алгоритмами глубокого обучения. Прогресс нашего знания предполагает, что постепенно понимание нейрофизиологических механизмов будет углубляться, приводя к прояснению механизмов обучения нейронные сети в мозгу. Но даже углубление понимания этих механизмов не будет означать, что технические решения в сфере искусственного интеллекта непременно последуют за образцами, данными природой.

Представляется, что наиболее горячей зоной обмена идеями психологов со специалистами по искусственному интеллекту могут стать когнитивные архитектуры (Kotseruba, Tsotsos, 2018)<sup>1</sup>.

Сегодня, когда революция глубокого обучения уже принесла блестящие плоды, становится все очевиднее односторонность этой парадигмы в ее чистом виде, необходимость для дальнейшего движения дополнения нейросетей другими технологиями.

Так, при всех удивительных достижениях нейросети все равно проигрывают когнитивной системе человека по ряду направлений, например — по скорости обучения и широте использования результатов научения. Например, программа АльфаГо может сегодня победить любого человека в игре го. Однако подсчеты показывают, что АльфаГо для достижения такого уровня сыграла и изучила порядка миллиона партий, в то время как профессиональные игроки в го достигают хорошего уровня, сыграв, вероятно, в тысячи раз меньше. Победа достигнута, таким образом, за счет не столько алгоритма обучения, сколько быстродействия и неумоимости компьютера, т. е. экстенсивным путем. Кроме того, перенос мастерства на другие варианты игры (например,

---

<sup>1</sup> Формально работы по когнитивным архитектурам относятся к области искусственного интеллекта, поскольку предполагают создание программ, осуществляющих определенные когнитивные функции, однако имеют две существенные особенности. Во-первых, когнитивные архитектуры представляют собой устройства не специализированные, а направленные на решение большого количества задач. Они имеют перцептивные функции, позволяющие кодировать и анализировать поступающую на сенсоры информацию, структуры, проводящие логический вывод и рассуждение, а также системы принятия решений и планирование действий. Возможность решения разнородных задач предполагает и разнообразие режимов работы, а значит, различные процессы, связанные со сменой режимов, в том числе — внимание. Во-вторых, эти устройства тестируются на соответствие экспериментальным данным, получаемым на человеке. Таким образом, когнитивные архитектуры представляют собой фактически сконцентрированное видение устройства когнитивной системы человека сквозь призму понятий искусственного интеллекта. Очевидно, что прикладное значение когнитивных архитектур по сравнению с обычными моделями искусственного интеллекта ниже: специализированное устройство способно решать конкретную задачу лучше неспециализированного. Однако такие архитектуры за счет стремления к сходству с когнитивной системой человека играют роль моста между двумя областями исследований.



для поля другой формы и размеров, измененных правил и т.д.) для нейросети проблематичен.

При этом добавление к нейросети механизмов обучения за счет конструирования моделей объектов и событий в ряде случаев позволяет приблизить достижения нейросетей к человеческим. Основанный на построении моделей подход не только выглядит более похожим по механизмам на естественный интеллект, но и позволяет приблизиться к человеку как в плане узнавания с одного примера, так и возможности генерировать новые образцы (например, Lake, Salakhutdinov, Tenenbaum, 2015).

Подобные примеры улучшения функций искусственного интеллекта и приближении его к человеку при дополнении нейросети глубокого обучения механизмами создания моделей объектов можно продолжить. Американский фонд DARPA, анонсируя в 2018 г. новую программу работ по искусственному интеллекту AI Next, особое внимание уделил дополнению нейросетей глубокого обучения элементами контекстного рассуждения и объяснения решений.

Одним из перспективным направлений представляется комбинация нейросетей глубокого обучения с символьным искусственным интеллектом, что удивительно резонирует с историей и современными исследованиями психологии.

Идеология нейросетей глубокого обучения в информатике родственна идеологии «диффузных ассоциаций» в науках о человеческом интеллекте, и выше мы видели, как первая произошла из таких течений второй, как ассоцианистская психологии и исследования условного и оперантного обучения. В психологии, как и в информатике, делались попытки свести всю организацию когнитивной системы к диффузным ассоциациям, что, однако, также наталкивалось на критику. Критика связана с тем, что механизм диффузных ассоциативных связей плохо подходит для формирования структурированного знания и выполнения умственных действий, даже таких элементарных, как арифметические. Эта критика впервые прозвучала еще в возражениях немецких ученых против ассоцианизма, ведущего родословную из Англии и США. Так, в 1910–20-х гг. представители Вюрцбургской

школы, такие как Н. Ах, показывали, что в интроспективном опыте человека (самонаблюдении) присутствуют феномены, не сводимые к ассоциациям, и параллельно утверждали, что путем одних только ассоциаций объяснить рассуждения человека нельзя. В дополнение к ассоциативным принципам они предполагали функционирование структурирующих процессов (детерминирующая тенденция и т. д.) и даже пытались средствами своего времени посчитать, насколько эти структурирующие процессы могут противостоять ассоциативным. Аналогичным образом ассоциативный бихевиоризм начиная с 1930-х гг. стал дополняться идеями о таких несводимых к ассоциациям структурированных образованиях, как когнитивные карты (Э. Толмен).

Современные исследования также показывают частичный характер объяснения когнитивных процессов через диффузные ассоциации. Чисто ассоцианистские механизмы выявления регулярностей в окружающем мире выступают у человека в форме т. н. имплицитного обучения, при котором индивид выучивает некоторые правила, не сознавая самого факта, что он их выучил (А. Ребер). Прототипическим примером имплицитного обучения может служить усвоение родного языка, на котором люди говорят по правилам, которые не сознают. Например, читатель может спросить себя, в каком случае он говорит «два», а в каком — «двое». Скорее всего, если читатель — носитель русского языка, то употребляет эти слова без ошибок, но определит правило, по которому это употребление происходит, лишь приложив специальные усилия, а именно, воспроизведя контексты, в которых употребляются оба слова, и проанализировав различие этих контекстов. Это сильно напоминает глубокое обучение нейросетей, в отношении которых нужно принимать специальные меры, чтобы установить, на каком основании они выносят то или иное решение (т. н. объяснимый ИИ — explainable AI).

При этом в психологии несомненно, что более распространенным и, вероятно, важным случаем, чем имплицитное обучение, является обучение эксплицитное. Подобно тому, как в искусственном интеллекте в качестве дополнения к глубокому обучению часто рассматриваются символьные системы, в психо-

логии когнитивную систему можно понять, только вводя наряду с ассоциативными процессами также и действия с хорошо структурированным знанием.

Наиболее сбалансированные психологические модели когнитивных процессов сегодня носят комбинированный характер, включая как ассоциативную сеть, так и процессы создания и интерпретации структурированных репрезентаций. Примером может служить теория Я.А. Пономарева, согласно которому когнитивная система человека имеет двухполюсный характер, где один полюс связан с хорошо структурированным знанием, а другой — с диффузным, но постоянно пополняемым за счет взаимодействия человека с окружающим миром (Пономарев, 1976). Когнитивная деятельность человека, согласно теории Пономарева, связана с «перемещением» между этими полюсами, при котором происходят постоянные переходы между диффузным и хорошо структурированным знанием. Можно показать, что такая двухполюсная организация является эффективным способом обеспечить возможность для когнитивной системы к открытию нового при одновременном удержании ранее выработанных схем структурированного поведения (Ушаков, 2006).

Параллели между психологическими исследованиями когнитивной системы и моделями искусственного интеллекта вряд ли случайны. В конечном счете психология, нейрофизиология, когнитивная лингвистика и искусственный интеллект вторгаются с разных сторон в одну и ту же область. Разгадывание загадки человеческого мышления было бы равнозначно колоссальному прогрессу в области создания устройств искусственного интеллекта.

Тем не менее, встречи туннелей, прорываемых с разных сторон, пока не происходит. В этом месте с новой силой возникает известный с середины прошлого века вопрос о том, может ли в принципе техническое устройство воспроизвести в полной мере человеческое мышление, а далее превзойти его. Этот вопрос по-прежнему составляют предмет дискуссий, которые основываются на разведении сущностей, с одной стороны, относящихся к вычислимости и доступных техническим устройствам типа машины Тьюринга, а с другой — связанных с сознанием, таких как

субъективные переживания (в терминах философов — квалиа), истинность, логическая необходимость, понимание или моральная ответственность. Вопрос заключается в том, параллельны ли состояния сознания работе материального вычислительного устройства или для их появления требуется нечто большее?

Наиболее серьезное обсуждение в этой связи вызывает т. н. Геделевский аргумент, который приобрел широкую известность после работ британского философа Дж. Лукаса. Аргумент основывается на теореме Геделя, согласно которой в любой формальной системе, включающей в себя арифметику, присутствуют истинные, но не выводимые из аксиом утверждения. На этом основании Лукас доказывал, что человек способен понимать истинность утверждений, которые не могут быть вычислены. Таким образом, производится разотождествление сознания истинности и вычислений, на которые способна машина.

Более изощренную аргументацию на основе Геделевского аргумента развил Нобелевский лауреат по физике Р. Пенроуз, доказавший специальную «теорему» на этот счет (Пенроуз, 2003). На этом основании Пенроуз предполагает, что собственно мышление не может осуществляться машиной Тьюринга, и связывает его с квантовыми процессами, протекающими в микротрубочках нейронов<sup>1</sup>. Более того, как стало выясняться, сам К. Гедель был склонен интерпретировать свою теорему о неполноте в духе принципиального барьера между вычислениями, которые в принципе могут быть осуществлены материальным устройством, и истиной, лежащей в Платоновом мире идей.

---

<sup>1</sup> Фактически Р. Пенроуз подводит новую аргументацию под представления Р. Декарта, согласно которым мышление привносится в мир души, которая влияет на протекание физических процессов в местах разрыва детерминационных цепочек. Понятно, что в соответствии с таким подходом между науками о человеке и искусственным интеллектом будет всегда стоять непреодолимый барьер, а технические устройства не приобретут собственно человеческие возможности. Этому подходу противостоит другой, восходящий к Г.В. Лейбницу, согласно которому сознательные явления строго параллельны физическим процессам. В этом плане символично и, возможно, неслучайно, что Лейбниц был создателем арифмометра, который рассматривается как один из прообразов современного компьютера.

Тем не менее, фундаментальное математическое исследование в этой области не находит, что геделевский аргумент ведет к твердому заключению о превосходстве человека над машиной (Ершов, Целищев, 2012). Авторы процитированного исследования рассматривают итог спора вокруг Геделевского аргумента пока как ничейный.

Существуют и аргументы другого типа, согласно которым разумное поведение технического устройства еще не означает наличие у него ментальных состояний<sup>1</sup>. Однако такого рода аргументы не могут не ставить под сомнение и существование ментальных состояний у других людей — ведь о ментальных состояниях каждый знает непосредственно только по себе, а о других судит лишь на основании признаков осознанности в поведении.

В целом в качестве итога можно утверждать, что сегодня не существует твердых аргументов о невозможности создания мыслящих и сознательных машин (например, Чалмерс, 2015), как, конечно, нет и гарантии, что они когда-либо будут созданы человеком.

### **Взаимодействие человека с искусственным интеллектом**

Предсказывать отдаленное будущее технологий — неблагодарное занятие. Технологии — результат движения мысли, а мысль, по определению, непредсказуема. Тем не менее, мы создаем более или менее глубокие и реалистичные программы дальнейшего раз-

---

<sup>1</sup> Наиболее часто в этом контексте обсуждается мысленный эксперимент «Китайская комната» американского философа Дж. Серля. Представим себе человека, который не знает китайского языка, но обучен правилам отвечать на определенную комбинацию китайских иероглифов другой их комбинацией так, что внешне он неотличим от носителя китайского языка. Эксперимент призван показать отсутствие равенства между внешним разумным поведением системы и наличием у нее ментального опыта. Однако сам Серль не использует свою аргументацию для отрицания возможности машинного мышления, направляет полемику на более локальные цели — опровержение представлений, что «инстанциация» адекватной компьютерной программы сама по себе является достаточным условием понимания.

вития технологий искусственного интеллекта и должны представлять себе их потенциальное включение в жизнь общества. Велики или не очень потенциальные возможности искусственного интеллекта, однако они заставляют задуматься о том, как выстроить отношения человека с ними. Сегодня идет внедрение в практику технологий искусственного интеллекта, пока еще далеких по своим возможностям от того, что просматривается вдалеке. Однако сопоставление с этими далекими и не очень ясными перспективами позволяет увидеть глубину происходящего.

Идея опасности для человека созданных им же самим разумных существ идет еще со времен Античности. Чудовище Франкенштейна будоражило воображение в XIX в., а с началом разработки технологий искусственного интеллекта был найден наиболее подходящий кандидат на роль существа, восстающего против человека.

С. Хокинг, например, выражал опасение, что супермощный искусственный интеллект может расправиться с человечеством, как с семьей муравьев. Мало кто из людей, - аргументировал он, - разрушит муравейник со зла. Но если вам нужно поднять уровень воды для обеспечения работы ГЭС, то вряд ли вы будете принимать в расчет затопление муравейника. Сверхумный ИИ будет очень хорошо справляться с задачами, и если его и наши цели не будут совпадать, у нас будут огромные проблемы.

Образ, нарисованный Хокингом, предполагает, что решение задач — высшая ценность для ИИ. Если люди утратят интеллектуальное лидерство, их ценности перестанут главенствовать и не будут выступать чем-то, с чем нужно считаться, торя магистральную дорогу истории.

По мнению И. Гуда, способность человека создать интеллектуальную машину, превосходящую его возможности, приведет ко все ускоряющемуся созданию машин с нарастающими интеллектуальными способностями: первая ультраинтеллектуальная машина является последним изобретением, которое человек должен когда-либо делать.

Во главу угла ставятся всеобщие ценности, квинтэссенцией которых является «вселенский прогресс интеллекта». Пафос

движения за и через человека и «слишком человеческое» иногда проскальзывает в постгуманизме. В этом чувствуется романтика готовности к гибели, чтобы вырастить великое будущее.

Противоположная позиция — абсолютная ценность индивидуального человеческого существования. Она находит отражение, например, у Н.Ф. Федорова в идее использования технологического прогресса для достижения ценностных идеалов. В этом плане цивилизационное назначение техники, в т. ч. искусственного интеллекта — реализовать исконное стремление человека к справедливости, счастью, бессмертию.

Большинство людей, безусловно, примыкают ко второй ценностной позиции. Проблема, однако, заключается в возможности, во-первых, людям, странам и социальным институтам договориться о целях создания искусственного интеллекта и не превратить его в средство или арену борьбы, а во-вторых, не допустить случайного формирования могучих искусственных агентов с бесчеловечными ценностями. История часто состоит в том, что совокупные действия множества людей приводят к результату, которого не хотел никто из них.

Представляется важным держать эту более удаленную оптику, рассматривая сегодняшние проблемы искусственного интеллекта.

Теперь перейдем к описанию тех процессов в человеческом обществе, которые возникают уже в связи с нынешним уровнем искусственного интеллекта и его ближайшим прогнозируемым развитием. Внедрение любых технологий перестраивает жизнь людей и требует психологической адаптации, но искусственный интеллект в этом плане занимает особое место. Даже те устройства, которые просто решают, лучше или хуже, задачи, решаемые также человеком, могут влиять на различные процессы в обществе.

Одно из ближайших по времени событий в этом плане — внедрение систем искусственного интеллекта в управление транспортом. На наших глазах формируются устройства, обеспечивающие автономное управление автомобилем — от навигаторов и круиз-контроля до системы удержания автомобиля в полосе, автопарковщика и т. д. Это, казалось бы, вполне пози-

тивное и прогрессивное явление оказывается, однако, на деле не столь простым, например, в плане обеспечения безопасности. Тонким получается момент обратного принятия человеком функции управления на себя, как показал опыт Гугломобиля. Возможность положиться на работу искусственного водителя в какой-то момент снимает у человека чувство ответственности за происходящее на дороге, что может не позволить ему в полной мере включиться в процесс управления, когда это потребуется.

В то же время необходимость полагаться на надежность электронного водителя может приносить душевный дискомфорт. Внедрение технологий искусственного интеллекта должно повышать качество жизни, увеличивать безопасность, психологический комфорт. Однако в действительности эффект может быть иным, и побочные продукты внедрения искусственного интеллекта без должного социо-гуманитарного сопровождения могут вести к повышению тревожности, напряженности у людей.

Маркетологи автокорпораций предсказывают значительный период сосуществования обычного автомобиля с автономным, причем последний, будучи существенно дороже, должен быть способен предложить потребителю серьезные преимущества. В каком-то смысле преимущество очевидно, поскольку происходит избавление от труда. Однако при этом теряется и удовольствие. На протяжении десятилетий автопроизводители рекламировали потребителям удовольствие от вождения их автомобилей, «драйверские автомобили» и т. д. Теперь, однако, предлагается от этого удовольствия отказаться, заплатив серьезные деньги. Более того, у водителя, передавшего управление автомобилем автомату, возникает тревога, связанная с потерей того, что психологи называют «иллюзией контроля» — человек испытывает большую уверенность в благоприятном исходе, если сам может каким-либо образом влиять на ситуацию.

В более общем плане возникает проблема передачи машинам контроля над сферами жизни, которые до сих пор находились в ведении людей. Искусственный интеллект, заменяя человека в его когнитивной деятельности, конечно, способен избавить от многочисленных малоприятных и утомительных дел и драматически



повысить производительность труда, точнее — предоставить на благо человечества почти не ограниченные объемы производимого автоматическими устройствами труда. Однако он и лишает удовольствий, более того — грозит затронуть функции главного действующего лица, субъекта собственной жизни, т. е. смыслообразующий стержень человеческой психики. Наконец, это означает передачу ответственности устройству, которое не обладает сознанием и, как следствие, не является вменяемым в юридическом и психиатрическом смыслах. Все это ставит перед социо-гуманитарными науками задачу выработки такой концепции передачи некоторых когнитивных функций искусственным устройствам, которая бы способствовала реализации человеком своих смыслов в жизни.

Далее, передача когнитивных функций человека искусственным устройствам неизбежно связана с уничтожением рабочих мест для людей и возможной безработицей по причине автоматизации. Так, в связи с тем же автономным автомобилем работу могут потерять профессиональные водители — такси, общественного и грузового транспорта. Для России, например, это — миллионы человек.

А если искусственный интеллект окажется способен работать в области бухгалтерского учета и логистика, торговли и архитектурного проектирования, перевода и регистрации документов, то с неизбежностью произойдет перераспределение сфер занятости, причем не совсем понятно, где остановится экспансия искусственного интеллекта.

Этот сюжет фактически повторяет на новом ветке историю начала XIX века — развитие информационных технологий грозит привести к тем же последствиям, к каким привело широкое внедрение машин, а именно к протестам теряющих трудовые места работников, вылившимся в движение луддитов. На своей родине в Англии луддизм составил для правительства проблему, соизмеримую по масштабам с проходившими примерно в то же время наполеоновскими войнами.

В случае искусственного интеллекта машины, заменяя людей, создают вместе с этим и средства существования. Отсюда

возникает перспектива всеобщего минимального гарантированного дохода, при которой приобретение профессии и труд перестают быть необходимым условием выживания человека. Жизнь вроде бы станет проще, но для психологов понятно, что свободное время и свобода выбора жизненных проектов с большой остротой ставят вопрос о смысле существования.

Еще один пласт опасений связан с тем, что искусственный интеллект может анализировать людей и социальные процессы, более того — активно на них влиять. Так, технологии искусственного интеллекта уже сегодня способны оценивать свойства человека по различным его проявлениям, прогнозировать поведение в различных ситуациях. Но также эти технологии способны создавать средства влияния: например, генерировать новые объекты — тексты, картинки, видео — с заранее заданным значением, например, создавать новые изображения известных предметов и людей или их видеоизображения. Сформировалось понятие «глубинной фальсификации» *deep fakes*, которое обозначает технологию, использующую глубокое обучение (*deep learning*) в целях создания достаточно правдоподобной ложной информации (*fakes*). В дальнейшем можно ожидать применения такого рода объектов как в психотерапии, так и для индоктринации.

Сказанное означает, что стараниями искусственного интеллекта для реципиента новостей мир чьих-то фантазий становится трудноотличимым от действительно происходящих событий.

Искусственный интеллект вызывает подозрения в контексте использования «Большим Братом», т.е. для выявления неблагонадежных и чрезмерно свободомыслящих граждан. Впрочем, пока эти технологии скорее используются коммерческими предприятиями в целях выявления потенциального интереса людей к совершению тех или иных покупок и последующего таргетирования рекламы. отслеживать передвижения пользователя, его предпочтения в социальных сетях, а также запросы в поисковых системах Пусть это использование и относительно безобидно, но оно может вести к отработке технологий выявления индивидуальных уязвимостей людей с дальнейшим воздействием на них в нужном кому-либо направлении.

Искусственный интеллект в целом создает невиданные возможности вычисления психологических тайн людей и их слабостей, а также средств воздействия. В истории человечества происходит уменьшение физического принуждения, но повышение более эффективного принуждения путем манипуляций.

### **Индивидуальный цифровой ангел**

Быстрое и точное распознавание паттернов человеческого поведения искусственным интеллектом позволяет решать в реальном времени задачи, которые лежали до последнего времени в сфере человеческой компетенции. Создается возможность индивидуального взаимодействия искусственного интеллекта с человеком, при котором с учетом особенностей человека для него создаются оптимальные условия в сферах, например, образования или личностной поддержки. В связи с этим проф. С.Ю. Степановым и автором этих строк предложена концепция индивидуального цифрового ангела, интегрирующего различные технологии искусственного интеллекта в интересах человека.

Цифровой ангел должен способствовать человеку в самых разнообразных сферах его жизни. Так, он может помогать в образовании и саморазвитии. Сегодня уже разработаны модели создания индивидуальных образовательных траекторий для учеников, наиболее продвинутой из которых на сегодняшний день, пожалуй, является модель Фальманя. Индивидуальная образовательная траектория предполагает, что ученики получают задания наиболее подходящей для них сложности и направленности, объяснения нужного материала и в нужной форме.

Можно выделить три последовательных уровня, которые должен пройти искусственный интеллект в этом направлении. На первом уровне в систему закладываются экспертные мнения профессионалов-преподавателей. На втором уровне искусственный интеллект обучается на больших данных, собранных в процессе выполнения учениками заданий. На третьем уровне

электронный учитель сам придумывает наиболее подходящие задания для учеников, основываясь на цифровой модели его знаний, умений и способностей.

Образование — это лишь одна из сторон саморазвития, и технологии искусственного интеллекта могут помогать людям узнавать свои сильные и слабые стороны, а также развиваться в разных областях — самообладании, общительности, настойчивости и т. д. Они могут помогать и с самоорганизацией, устройством распорядка дня, совладанием со стрессом, отслеживанием возникновения и преодолением зависимостей.

Саморазвитие связано с самореализацией, в том числе — в профессиональной сфере. Электронные ассистенты могут осуществлять профориентацию и своего рода индивидуальный коучинг — основываясь на оценке индивидуальных особенностей человека, учете его сильных сторон, они могут консультировать по вопросам выбора профессиональной траектории.

Одной из интенсивно развивающихся областей является автоматическая психотерапия. Сегодня уже эффективность такого рода воздействий считается доказанной результатами ряда исследований. Электронные ассистенты смогут оказывать определенную помощь в трудных жизненных ситуациях, консультировать и поддерживать людей. Искусственный психотерапевт способен вести беседу, например, с пожилыми людьми и оказывать психологическую помощь. Искусственный интеллект уже в настоящее время довольно успешно применяется для помощи аутистам.

Эта деятельность хорошо дополняется растущими возможностями искусственного интеллекта по распознаванию состояний людей. В частности, развиваются методы распознавания состояний по видеоизображениям человеческого лица и фигуры, по контенту и акустике речи, а также на основе психофизиологических индикаторов — кожно-гальванической реакции, сердечного ритма, электроэнцефалографии, излучения в ближнем инфракрасном спектре и др. Кроме того, развиваются методы распознавания психологического состояния индивидов и групп на основе больших данных, в первую очередь — анализа поведения в социальных сетях.

На основе оценки состояния человека на протяжении определенных временных промежутков электронный ассистент может давать ему советы, оказывать помощь в регуляции состояний, например, с помощью биологической обратной связи.

Но искусственный интеллект способен давать своему хозяину и информацию о состоянии и свойствах других людей, что может улучшать социальные взаимодействия. Взаимодействие — деловое и межличностное — вообще составляет еще одну важную область, где интеллектуальные ассистенты могут помогать людям. Такие устройства в принципе могут оценивать не только состояния, но и свойства людей, давать обратную связь, способствовать рефлексии. В них могут быть заложены алгоритмы, применяемые психологами для улучшения диадического и группового взаимодействия.

Индивидуальные ассистенты могут обмениваться информацией с интеллектуальными устройствами, обеспечивающими психологический климат коллективов — школ, университетов, компаний.

Наконец, умные ассистенты должны обеспечивать и защиту от злонамеренных психологических воздействий, которые, как отмечалось выше, с течением времени могут становиться все интенсивнее. Феномен дип-фейка свидетельствует о том, что маскировка ложной информации под истинную становится искуснее и интеллектуальные системы могут стать необходимыми для различения одного и другого. Необходима также защита от захвата внимания, который формирует игроманов и Интернет-зависимых.

Последнее логически подводит к вопросу об уязвимостях цифровых ангелов. Потенциальная возможность захвата цифрового ангела человека путем вирусной атаки или иного взлома означал бы не просто похищение персональной информации, но и чрезвычайные возможности управления и манипуляции человеком. Падение ангела легко может превратить его в самого опасного искуителя и врага человека.

## **Ассимиляция искусственного интеллекта как цивилизационный вызов**

Вышесказанное подводит к мысли, что технологии искусственного интеллекта могут крайне благотворно повлиять на жизнь человечества, но и наоборот стать для него большим несчастьем. Причем дело здесь не в искусственном интеллекте, а в человеке. Вопрос в том, возобладают ли в обществе силы конкуренции или кооперации. ИИ увеличивает общие доступные человечеству ресурсы, делая тем самым менее необходимой конкуренцию за них, но в то же время создает и невероятные ранее по своей изощренности методы этой конкуренции.

Силы конкуренции очень сильны — они пронизывают всю жизнь от межгосударственных отношений до бизнеса и политики внутри стран. Конец истории Ф. Фукуямы, похоже, отменился и баталии на всех уровнях продолжаются. Еще одна проблема связана с возможным резким увеличением скорости социальных и жизненных изменений при широком внедрении искусственного интеллекта. О постоянном возрастании темпов эволюции писал в XIX в. Э. Геккель, а за ним — Ф. Энгельс, а позже подобный феномен исследовал В.И. Вернадский. Перемены с внедрением технологий искусственного интеллекта грозят идти со все возрастающей скоростью, оставляя нам минимум времени на раздумья.

Создание технологий искусственного интеллекта выглядит полем для нового Армагедона и, возможно, сформирует основную линию напряжения грядущего времени.

## **ЛИТЕРАТУРА**

1. *Ершов Ю.Л., Целищев В.В.* Алгоритмы и вычислимость в человеческом познании. Новосибирск: Изд-во Сибирского отделения Российской акад. наук, 2012.

2. *Пенроуз Р.* Новый ум короля: О компьютерах, мышлении и законах физики. М., 2003.

3. Степанов С.Ю., Оржековский П.А., Ушаков Д.В. Проблема цифровизации и стратегии развития непрерывного образования // Непрерывное образование: XXI век. 2020. № 2 (30). С. 2–15. — DOI 10.15393/j5.art.2020.5684.

4. Чалмерс Д. Сознательный ум: В поисках фундаментальной теории. М.: УРСС: Книжный дом «ЛИБРОКОМ», 2015.

5. Kotseruba I., Tsotsos J.K. (2018) 40 years of cognitive architectures: core cognitive abilities and practical applications // Artificial Intelligence Review <https://doi.org/10.1007/s10462-018-9646-y>

6. Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B. (2015) Human-level concept learning through probabilistic program induction. Science 350(6266):1332–38. [arBML, MB, ED, NK].

7. Stepanov S.Yu., Ushakov D.V. Artificial intelligence and digital angel technology in education // Сборник трудов «Восемнадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2020. Москва, 10-16 октября 2020 г.» / Под ред. В.В. Борисова, О.П. Кузнецова. М.: МФТИ, 2020. С. 105–112.

8. Russell S, Norvig P. Artificial intelligence: a modern approach. Prentice Hall, Upper Saddle River.1995.

## СОЗНАНИЕ, МОЗГ, ОБЩИЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НОВЫЕ СТРАТЕГИЧЕСКИЕ ЗАДАЧИ И ПЕРСПЕКТИВЫ

---

---

Основные цели и задачи развития в нашей стране искусственного интеллекта (далее — ИИ) четко сформулированы в «Национальной стратегии развития искусственного интеллекта на период до 2030 года», утвержденной Указом Президента Российской Федерации от 10 октября 2019 года. В настоящее время развитие ИИ подошло к новому стратегическому рубежу, т. е. к задаче создания **Общего ИИ** (в принятой международной номенклатуре — **AGI: Artificial General Intelligence**). Эта задача соответствует мировому тренду развития ИИ в условиях нарастающей конкуренции между крупнейшими научными центрами и специализированными корпорациями, а в более широком масштабе — между Россией и такими лидирующими в области ИИ государствами, как США, Китай, Япония, страны Западной Европы. Факт наличия острой конкуренции очевиден, Россия пока существенно отстает от своих главных конкурентов.

---

\* *Дубровский Давид Израилевич* — доктор философских наук, профессор, главный научный сотрудник Института философии РАН, заместитель председателя Научного совета при Президиуме РАН по методологии искусственного интеллекта и когнитивных исследований, Москва, Россия.

*Dubrovsky David Israilevich* — Doctor of Philosophy, Professor, Chief Researcher at the Institute of Philosophy of the Russian Academy of Sciences, Deputy Chairman of the Scientific Council under the Presidium of the Russian Academy of Sciences on the Methodology of Artificial Intelligence and Cognitive Research, Moscow, Russia.



Догонять их по основным сложившимся приоритетным направлениям развития ИИ — бессмысленно.

Нам нужен прорыв! Прорыв на принципиально новый уровень развития ИИ, о чем говорил Президент России В.В. Путин. Для всех является понятным, в чем заключается стратегическое значение прорыва, от которого зависит наше экономическое развитие, решение проблем безопасности, многих важных социальных вопросов. Но возможно ли это? Теоретически прорыв (несмотря на наше положение в настоящее время) возможен с учетом того, что проблемное поле ИИ в высшей степени многомерно. В нем назревает сегодня множество мыслимых и немыслимых пока новаций разного рода и разного порядка. Вопрос состоит в том, как своевременно диагностировать и использовать или создать такие из них, которые позволили бы добиться выхода на принципиально новый уровень разработки Общего интеллекта (AGI). Наша страна обладает высокоразвитой научно-технологической базой, сильной математической школой для подобных исследований, располагает рядом крупных достижений и преимуществ в решении теоретических, организационных и практических проблем развития ИИ, которые могут быть успешно использованы в нынешней конкурентной ситуации.

Фундаментальный прорыв означает формирование новой теоретической и методологической основы ИИ, предусматривает разработку новых подходов и методов, которые способствовали бы созданию AGI, т. е. системы, приближающейся по функциональным возможностям к естественному интеллекту (далее — ЕИ).

### **О соотношении понятий общего и узкого, сильного и слабого, естественного и искусственного интеллекта**

В актуализации, организации, стимулировании исследований в области AGI ведущую роль сыграла активная позиция Сбербанка России, которому удалось создать за счет собственных средств мощную научно-исследовательскую команду и вычислительную инфраструктуру в области ИИ и робото-

техники. В рамках этой структуры, объединяющей ведущих специалистов, созданы экспертные советы, на которых подробно обсуждают ключевые проблемы ИИ и AGI, проводятся крупнейшие мировые конференции по искусственному интеллекту “Artificial Intelligence Journey” (8–9 ноября 2019 года и 4–5 декабря 2020 года), где выступал Президент России В.В. Путин.

В 2021 году вышла первая в нашей стране книга, посвященная основательному обзору проблематики AGI, анализу состояния ее разработки, основных трудностей и возможных способов их преодоления<sup>1</sup>. В предисловии к ней Президент, председатель правления Сбербанка России Герман Греф так оценивает современное состояние разработок ИИ: «В большинстве своем существующие решения являются примерами реализации технологий узко специализированного искусственного интеллекта, требующего настройки и перепроверки со стороны человека. Чтобы так же хорошо решать разнообразные комплексные задачи, как это делают люди, машины должны научиться строить причинно-следственные модели окружающей среды и ориентироваться в разных контекстах, а не просто максимизировать успех при решении какой-то узкой задачи. Они должны понимать физические, психологические и другие законы нашего мира и уметь связывать новую информацию в общую картину с тем, что уже знают. Чтобы добиться этого, нам необходимо преодолеть очередной технологический рубеж — создание Общего искусственного интеллекта или AGI»<sup>2</sup>. Греф подчеркивает принципиальное значение разработки в этих целях междисциплинарных проблем AGI и формулирует стратегическую установку для достижения успеха: «Крайне важным фактором создания и развития технологий Общего искусственного интеллекта является обеспечение сквозного целеполагания между прикладными (или индустриальными) задачами, фундаментальными исследованиями и системой образования — так называемая триада Practice —

---

<sup>1</sup> Сильный искусственный интеллект. На подступах к сверхразуму / Александр Ведяхин и др. М.: Интеллектуальная Литература, 2021. 232 с.

<sup>2</sup> Там же. С. 5–6.

Education — Research»<sup>1</sup>. Указанные три составляющие ввиду их неотъемлемой взаимообусловленности должны быть в центре внимания государственных органов, призванных содействовать реализации программы AGI.

Рассмотрим теоретико-методологические вопросы философского плана. Для создания AGI в первую очередь следует преодолеть ограниченности традиционной методологии разработки ИИ, основы которой заложены А. Тьюрингом. Согласно этой методологии, понятие «интеллект» трактуется в сугубо функционально-операциональном смысле, исключающем роль сознания, вся сложная проблематика которого выносится за скобки. Исследование сознания в данном случае полагается излишним: существует когнитивная или практическая задача, которая формулируется и решается посредством сугубо операциональных методов с помощью компьютерных программ.

Именно методология Тьюринга открыла компьютерную эру, обеспечила выдающиеся успехи ИИ, развитие информационных технологий, качественное изменение земной цивилизации. Однако сегодня мы переходим к новому этапу, требующему существенных теоретико-методологических обновлений. Парадигма функционализма, в формирование которой А. Тьюрингом внесен первостепенный вклад, естественно, остается действующей, но она требует более широкой интерпретации с учетом ее роли в объяснении сознания и способов использования новейших результатов исследований сознания для развития ИИ.

Тенденция «реабилитации» сознания и потребность коррективов методологии Тьюринга отчетливо проявлялась в разных формах за последние два десятилетия<sup>2</sup>. В этом направлении нам еще предстоит глубокая теоретико-методологическая работа. К тому же немало специалистов в области ИИ опираются на традиционные методы, позволяющие и в настоящее время успешно

---

<sup>1</sup> Там же. С. 7.

<sup>2</sup> См., например: Efimov A. Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence // Artificial General Intelligence (AGI). 2020. Vol. 12177. P. 83–94.

решать многие практические задачи, и они убеждены, что для реализации намечаемых новых задач достаточно совершенствования, улучшения существующих методов. Они скептически относятся к привлечению проблемы сознания, считая, что она влечет за собой множество неопределенностей, «сплошной туман», что нет смысла в разработке т. н. трудной проблемы сознания, только еще более усложняющей ситуацию. Подобные убеждения проявились у некоторых видных специалистов, выступавших на экспертных обсуждениях проекта разработки AGI, проводимых Сбербанком России. Участвуя в указанных обсуждениях, я полемизировал по этим вопросам.

Подобные настроения присущи, конечно, лишь некоторым специалистам. Существуют разные взгляды, разные направления разработки ИИ. Среди них значительное место занимают концептуальные подходы, в которых внимание обращено на изучение ЕИ, моделирование его специфических свойств, создание новых когнитивных архитектур с целью сближения функциональных возможностей ИИ и ЕИ. Одним из тех, кто считает ЕИ главным источником и ресурсом нового этапа развития ИИ (что отвечает задачам AGI), является крупный отечественный специалист, активный член нашего Научного совета при Президиуме РАН по методологии искусственного и когнитивных исследований (далее сокращенно — НСМИИ) В.К. Финн. Он основательно анализирует функциональные когнитивные структуры ЕИ и, главное, придает первостепенное значение философско-методологическому подходу к решению актуальных проблем современного этапа развития ИИ, использует его в своих исследованиях<sup>1</sup>. В.К. Финном создан широко известный ДСМ-метод, на основе которого автором получены значительные результаты<sup>2</sup>. Развитие этого направления может быть перспективным и для

---

<sup>1</sup> Финн В.К. Искусственный интеллект: методология, применения, философия. М.: URSS, Красанд, 2011.

<sup>2</sup> Финн В.К. О классе ДСМ-рассуждений, использующих изоморфизм правил индуктивного вывода // Искусственный интеллект и принятие решений. 2016. № 3. С. 48–61.

разработки AGI. Существуют и многие другие подходы, применяющие в этом аспекте результаты исследований ЕИ, которые неоднократно обсуждались на заседаниях НСММИ<sup>1</sup>. Особенно важны в этом плане результаты разработки положений постнеклассической эпистемологии применительно к оценке и стимулированию когнитивных исследований и проблематики ИИ в целом (В.А. Лекторский<sup>2</sup>), а также для решения актуальных теоретических вопросов развития информационных технологий, повышения их социальной значимости (В.И. Аршинов, В.Е. Лепский, В.Г. Буданов, В.А. Глазунов и др.).

Наряду с представителями компьютерных и философских дисциплин НСММИ объединяет специалистов в области нейронауки, психологии, лингвистики, математики, экономики, различных отраслей социогуманитарного знания, что позволяет обсуждать и продуктивно разрабатывать широкий диапазон междисциплинарных проблем ИИ — необходимое условие для формирования искомых прорывных направлений в развитии ИИ. Большой вклад в развитие ИИ внесен академиками РАН В.Л. Макаровым, С.Н. Васильевым, рядом других крупных специалистов в области ИИ. Свидетельством этого может служить настоящее издание.

Несомненно, традиционные методы ИИ с учетом их совершенствования еще долго будут успешно использоваться для решения типичных прикладных задач. Но для прорыва нужны большие творческие и волевые усилия, связанные с преодолением привычных ментальных клише, выбором (и созданием) наиболее перспективных направлений исследований, концентрацией сил и средств для разработки принципиально новых теоретических и методологических концепций развития ИИ.

---

<sup>1</sup> Кузнецов О.П. Избранные труды. Автоматы, языки и искусственный интеллект. М.: Наука, 2016. 464 с.

<sup>2</sup> Лекторский В.А. Эпистемология классическая и неоклассическая. М., 2001; *Он же*. Человек и культура. СПб., 2018; *Он же*. Философия, искусственный интеллект и когнитивная наука // Искусственный интеллект: Междисциплинарный подход / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИнтелЛ, 2006. С. 13–21.

В связи с этим важное место занимают вопросы уточнения понятий ИИ и ЕИ: сильного и слабого, общего (AGI) и узкого ИИ, особенно *соотношение AGI с ЕИ*. К этому добавляются вопросы о прото-AGI и узкого AGI, которые рассматриваются как стадии продвижения к полноценному AGI. Все это получило широкое освещение в указанной книге, посвященной проблемам AGI, которая, подчеркнем еще раз, вносит большой вклад в его разработку.

Под сильным интеллектом чаще всего имеют в виду человеческий интеллект, наделенный сознанием, т. е. ЕИ. Все существующие формы ИИ чаще относят к слабому ИИ, хотя в решении некоторых узких задач он может быть чрезвычайно сильным. Но то, что касается Общего интеллекта (AGI) и его отношения к ЕИ и к обычному ИИ, как правило, сопровождается вопросами, ответы на которые истолковываются неоднозначно. Термин «общий интеллект» появился в 1997 году, когда компьютер *Deep Blue* обыграл чемпиона мира в шахматы. На базе рассмотрения проблемы AGI было предложено делить ИИ не на сильный и слабый, а на *общий* (способный адаптироваться к решению разных проблем) и *узкий* (способный решать определенную задачу). В 2007 году состоялась первая конференция, посвященная проблемам развития Общего ИИ, которые впоследствии стали предметом возрастающего внимания специалистов в области ИИ. Тем более что подавляющее большинство систем ИИ характеризуются как узкоспециализированные, способные качественно решать определенную узкую задачу. Поэтому критики не раз говорили о том, что компьютер, запрограммированный на решение тысячи задач, не способен самостоятельно научиться решать тысячу первую.

В указанной книге об AGI подробно изложена история развития ИИ, отмечено, что оно шло в направлении снижения зависимости технологий ИИ от человека и к повышению способности решения более широкого круга задач. Но эти достижения по-прежнему далеки от решения проблемы AGI. Авторами дана общая оценка нынешнего положения дел: во-первых, с точки зрения адаптивности все существующие системы ИИ на основе любых известных подходов предполагают функционирование

в ограниченном наборе задач и условий и не способны самообучаться функционированию в условиях существенно новых. То есть они являются программируемыми, хотя сложность их программирования заметно снизилась, а способность к обучению возросла. Во-вторых, с точки зрения автономности систем ИИ все существующие системы не являются автономными и не могут полноценно функционировать без живого оператора, отвечающего за запуск и остановку, техобслуживание, целеполагание и определение режимов работы в зависимости от тех или иных условий или задач. То есть они остаются управляемыми»<sup>1</sup>. «В третьих, с точки зрения интегративности современные системы ИИ являются не системами, обладающими интеллектом как таковым (даже ограниченным), а системами компьютерного зрения, обработки естественного языка, анализа данных (машинного обучения), обработки символьной информации (рассуждений на основе знаний) и т. д., то есть интегративными не являются»<sup>2</sup>. Тем самым четко обозначены главные свойства, требуемые для построения действительного AGI.

Интерес представляет то, как решается вопрос об определении AGI, его степени общности и отношения к ЕИ: «Общий интеллект — это способность достигать сложных целей в сложных средах» (Бен Герцель). «Интеллект — это способность системы адаптироваться к своей среде, работая при недостаточных знаниях и ресурсах» (Пей Ванг)<sup>3</sup>. От себя же авторы книги добавляют следующее: «Делая акцент на широком диапазоне сред, область AGI позволяет нам избавиться от антропоцентричных предпочтений и предлагает сфокусироваться на общих решениях, пригодных для разных агентов (человека, животных, роботов, ботов и т. д.), действующих в разных условиях...», в том числе и «на неизведанных планетах»<sup>4</sup>.

---

<sup>1</sup> Сильный искусственный интеллект. С. 52.

<sup>2</sup> Там же.

<sup>3</sup> Там же. С. 28.

<sup>4</sup> Там же. С. 29.

Столь широкая трактовка интеллекта теоретически правомерна, соответствует принципу изофункционализма систем. Не исключено, что она может быть полезной и за пределами не только антропологического, но и биологического контекста, оставляя простор для творчества технологических суперноваций, для осмысления гибридных и иных новообразований в ходе нарастающей антропотехнологической эволюции.

Однако возникает вопрос о том, почему сразу отодвигается на задний план «сильный» интеллект, наш ЕИ как высшая форма из всех известных форм интеллекта, наделенная сознанием, благодаря которому мы создаем все искусственные интеллекты. Ведь именно в ЕИ присутствуют драгоценные и желательные для AGI свойства.

Авторы разъясняют свою позицию следующим образом: «Определение интеллекта в области AGI может казаться слишком абстрактным и далеким от наших представлений о естественном интеллекте. А еще широко распространено мнение, что человеческий интеллект — единственный пример интеллекта. Почему бы не опираться на него? Даже если мы предположим, что для характеристики AGI необходимы или достаточны критерии человеческого интеллекта, описание последнего тоже основано на довольно зыбких понятиях»<sup>1</sup>. В подтверждение о «зыбкости» дано определение интеллекта из Википедии, далеко не надежного в научном отношении источника. В нем, наряду с указанием на то, что он является «качеством психики», перечислены в несистематизированном виде его многочисленные «способности», от приспособления к новым ситуациям до ощущений, мышления, воображения, внимания, воли и рефлексии. Приведенное определение трудно принять в полной мере, поскольку непонятно, каким образом упорядочить, соотнести между собой эти способности и использовать их для целей AGI. В итоге предложено не вполне ясное заключение: «Определение интеллекта в области AGI не включает эти способности, но и не отвергает их»<sup>2</sup>.

---

<sup>1</sup> Там же. С. 30.

<sup>2</sup> Там же. С. 31.



Заслуживает внимания и такой аргумент: «Как повышение уровня решения узких задач вплоть до сверхчеловеческого не потребовало “сильных” качеств, так и расширение общности методов решения задач вовсе не обязательно подразумевает преднамеренное движение в сторону сильного ИИ»<sup>1</sup>. Под «сверхчеловеческим» имеются в виду действительно достижения «узкого» ИИ: например, победа компьютера над чемпионом мира по игре в Го, многие другие результаты, связанные с задачами анализа больших данных, быстроедействие и т. д.

Но далее читаем: «Можно предположить, что *некоторые аналоги некоторых “сильных” качеств у действительно общего ИИ должны быть* (здесь и далее курсив мой. — Д.Д.). Например, наверняка Общий ИИ должен иметь *способность к интроспекции* — анализу *собственных мыслительных процессов* или даже оптимизации лежащих в их основе алгоритмов. При этом у такого ИИ будет некий *образ себя как часть картины мира*. Но это не обязательно означает, что у него будет самосознание в философском смысле и уж тем более личность сродни человеческой, хотя по глубине рефлексии он вполне может и превосходить человека. Наверняка он проявит “понимание” тех областей, в которых действует успешнее человека. Но это не значит, что такое понимание будет сопровождаться у него субъективными переживаниями, схожими с человеческими. Наверняка у него будет многомерная система мотивации, включающая аналоги, например, любопытства и удивления. Но его вовсе не обязательно пытаться наделить всеми человеческими эмоциями. Хотя, скажем, для социальных роботов это может быть полезно, но даже они способны лишь симулировать чувства и эмоции, а не испытывать их»<sup>2</sup>.

Приведенные суждения весьма показательны. Они свидетельствуют о том, что специфические свойства и функции сознания, присущие нашему мыслительному процессу, все-таки **должны** привлекаться для построения AGI. Но возможно ли это

---

<sup>1</sup> Там же. С. 32.

<sup>2</sup> Там же. С. 32–33.

за вычетом состояния *субъективной реальности* — неотъемлемого качества сознания?

Теоретически возможно, но при условии способности *адекватного вычленения и описания оперативных функциональных аналогов* определенных состояний субъективной реальности (далее — СР) и затем их *воспроизведения на небиологическом субстрате* (до чего нам пока очень далеко). Но что имеется в виду под «интроспекцией» и «собственными мыслительными процессами» или «образом себя как части картины мира»? Эти понятия выражают специфические особенности реальной сознательной деятельности (исследуемого нами ЕИ). Они являются весьма сложными, многомерными, взаимодополняющими друг друга по содержанию. В когнитивном плане каждое из них включает в себя возможность множества аналитических и синтетических оперативных действий. **Поэтому не вполне ясно, о каком аналоге идет речь.** Приведенные понятия представлены слишком абстрактно. Следовательно, указания на подобные аналогии вряд ли могут иметь предполагаемое значение для моделирования когнитивных структур AGI.

Несмотря на это, установка на поиск аналогий заслуживает поддержки, т. к. обращена к задаче использования специфических свойств сознательной деятельности для разработки AGI. Нужны результаты профессионального философского и специального феноменологического анализа операциональных структур СР, допускающих корректное определение их *аналогов*, пригодных для построения когнитивных архитектур AGI.

Такая установка отличается от жестких высказываний о том, что общий искусственный интеллект «не лежит между слабым и сильным ИИ, а просто находится в стороне от них и определяется без отсылки к человеческому интеллекту как ИИ, способный решать широкий круг задач»<sup>1</sup>. Но она, очевидно, в то же время противостоит и упомянутым выше случаям прямого отрицания роли сознания в разработке проблем ИИ и AGI. Специалист в области ИИ, занимающий эту позицию, **демонстрирует**

---

<sup>1</sup> Там же. С. 27.

**парадоксальную ситуацию: он как бы забывает о том, что является носителем ЕИ и обладает искомыми свойствами AGI, постоянно применяет их в своей профессиональной работе, несмотря на то, что они не выделяются им, остаются скрытыми для него в процессах его сознательной деятельности, направленной на разработку AGI.**

Большинство специалистов в области ИИ погружены в решение прикладных задач. Они вряд ли могут быть достаточно компетентны в философско-методологических вопросах проблемы сознания, анализа его ценностно-смысловых и деятельно-волевых структур; им «хватает» своих дел, трудностей и проблем. Но сегодня важно понимать, что профессионалы, занимающиеся *теорией познания, методологией науки, феноменологией субъективной реальности*, способны быть активными помощниками в общем деле теоретического осмысления роли сознания на современном этапе развития ИИ. Разумеется, многие специалисты в области ИИ отдают себе в этом отчет и готовы к сотрудничеству.

## **Сознание и AGI.**

### **Нужна ли «Трудная проблема сознания»?**

Сознание обладает специфическим и неотъемлемым качеством СР. Его изучение и понимание зачастую выносятся за скобки при решении задач сугубо функционалистского, бихевиорального типа, в частности, теми, кто занимается ИИ, не выходя за рамки классической методологии Тьюринга. В большинстве случаев у них нет необходимости выделять и анализировать качество СР, поскольку их внимание сосредоточено на описании поставленной задачи и ее программировании.

Часто мы как бы не замечаем это качество, даже когда напругаем, корректируем собственное мышление, ибо оно подобно воздуху, которого мы тоже «не замечаем» во многих интервалах жизни. Качество СР становится слишком «заметным», когда оно нарушается в психопатологических случаях

(феномен деперсонализации) или когда предметом специального исследования оказывается наше мышление и сознание. Тогда как раз и возникают принципиальные вопросы «трудной проблемы сознания». Ведь ее суть связана именно с трудностью научного объяснения качества СР, которое всегда служило камнем преткновения в естественнонаучных дисциплинах. Особенно остро это проявляется в нейронаучных исследованиях сознания (классическая проблема «Сознание и мозг», Mind-Brain Problem),

Дело в том, что *явлениям СР нельзя приписывать физические свойства* (массу, энергию, пространственные характеристики). Как же в таком случае можно теоретически корректно объяснить связь явлений СР с мозговыми процессами — и шире — с телесными и физическими процессами вообще (несмотря на то, что эта связь на каждом шагу очевидна: моя мысль легко управляет моей рукой). Вопросы, касающиеся такого объяснения, как раз и принято объединять под названием «Трудная проблема сознания». Она более 60 лет находится в центре внимания западной аналитической философии, служит предметом постоянных дискуссий. Ей посвящено за это время поистине огромное число публикаций. Однако общепринятое решение пока отсутствует. В большей части концепций предлагается тот или иной редукционистский подход, чаще всего в форме отождествления явлений СР с физическими процессами (см. подробный критический анализ этих концепций<sup>1</sup>).

«Трудная проблема сознания» рассматривается в двух главных категориальных планах:

1. В *онтологическом плане* ставятся вопросы объяснения связи явлений СР с деятельностью мозга, их способности служить причиной изменения телесных процессов, способности управлять ими (т. е. объяснения *психической причинности*), а также более широкие вопросы происхождения СР в ходе

---

<sup>1</sup> Дубровский Д.И. Субъективная реальность и мозг: к вопросу о полувековом опыте разработки проблемы сознания в аналитической философии // Дубровский Д.И. Проблема сознания: Теория и критика альтернативных концепций. М.: ЛЕНАНД, 2019. С. 140–183.

эволюции как особой, *виртуальной*, реальности, форм ее существования, роли в физическом мире (при условии, что ей нельзя приписывать физические свойства);

2. В *эпистемологическом плане* возникают вопросы: а) об отсутствии прямых логических связей между понятиями, описывающими явления сознания, СР (интенциональность, смысл, ценность, целеполагание, воля, вера, творческое воображение и др.) и понятиями, описывающими физические процессы (масса, энергия пространственные характеристики, молекулярные, электронные процессы и др.); этот концептуальный разрыв называют в аналитической философии «провалом в объяснении»; б) вопросы о переходе от индивидуально-субъективного опыта, присущего только данному индивиду (выражаемого в отчетах от первого лица), к *интерсубъективным*, общезначимым утверждениям (от третьего лица) и обоснованию истинного знания. Здесь также возникают значительные теоретические трудности, которые являются предметом эпистемологических исследований<sup>1</sup>.

Предлагаемое мною теоретическое решение основных вопросов «Трудной проблемы сознания» основано на информационном подходе и противостоит редукционистским концепциям. Оно было в основных чертах представлено в моих работах еще в 60-х и 70-х годах прошлого столетия<sup>2</sup>, и затем длительное время развивалось с опорой на результаты нейронаучных исследований сознания и достижения в области ИИ. Оно получило

---

<sup>1</sup> См., напр.: Дубровский Д.И. Актуальные аспекты проблемы интерсубъективности // Естественный и искусственный интеллект: Методологические и социальные проблемы / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: Канон+, 2011. С. 129–148.

<sup>2</sup> Дубровский Д.И. Физиологическое и логическое // Вопросы философии. 1966. № 8; Он же. Мозг и психика: О необоснованности философского отрицания психофизиологической проблемы // Вопросы философии. 1968. № 8. С. 119–135; Он же. Психические явления и мозг. Философский анализ проблемы в связи актуальными задачами нейрофизиологии, психологии и кибернетики. М.: Наука, 1971. 386 с. (Второе, доп. издание: М.: ЛЕНАНД, 2021. 400 с.); Он же. Информация, сознание. Мозг. М.: Высшая школа, 1980. 286 с. (Второе, доп. издание: Информация, сознание мозг: Расшифровка мозговых кодов психических явлений. М.: ЛЕНАНД, 2021. 304 с.).

наиболее четкое и систематическое изложение в посвященной ей книге<sup>1</sup> и в специальной, сравнительно небольшой по объему статье, опубликованной в международном нейронаучном журнале<sup>2</sup>. Напомню кратко основные ее положения, но хочу сразу подчеркнуть, что главный вопрос онтологического плана состоит именно в объяснении связи явлений СР с мозговыми процессами. Для этого принимаются три исходные посылки:

1. *Информация необходимо воплощена в своем физическом носителе, не существует вне и помимо его.*

2. *Информация инвариантна по отношению к физическим свойствам своего носителя, т. е. может кодироваться по-разному, храниться и передаваться носителями, имеющими разные физические свойства (например, с помощью звуковых сигналов, письменных знаков, электронных сообщений, разных языков и т. д.*<sup>3</sup>

3. *Явление СР (например, переживаемый мной в данном интервале образ данного дерева, есть информация о данном объекте (обозначим его А). Как свидетельствует современная нейронаука, мозговым носителем явления СР, в данном случае А, выступает определенная нейродинамическая система (обозначим ее Х).*

Если эти три исходные посылки принимаются, то из них логически выводятся искомые объяснительные следствия.

**Явление СР необходимо связано с соответствующим мозговым процессом как информация со своим носителем.** Хотя нейродинамическая система *Х* необходимо состоит из физических компонентов, ее функциональная специфика не может быть

---

<sup>1</sup> Дубровский Д.И. Проблема «Сознание и мозг». Теоретическое решение. М.: Канон+, 2015. 208 с.

<sup>2</sup> Dubrovsky David I. "The Hard Problem of Consciousness". Theoretical solution of its main questions //AIMS Neuroscience. 2019. Vol. 6. N 2. P. 85–103.

<sup>3</sup> Мои оппоненты часто заявляют, что информационный подход не может использоваться в качестве основы для теоретического объяснения, поскольку нет общепринятой теории информации. Существуют две альтернативные концепции — атрибутивная и функциональная. Тем не менее есть ряд общепринятых в современной науке положений об информации. Именно такие общепринятые положения используются мной в качестве исходных посылок; они хорошо согласуются как с атрибутивной, так и с функциональной концепцией.

объяснена на основе физических свойств и закономерностей (поскольку, как известно, описание функциональных отношений логически независимо от описания физических отношений). Это показывает и анализ характера необходимой связи *A* и *X*.

Связь между *A* и *X* не является причинной, это особый вид *функциональной* связи: *A* и *X* — явления **одновременные и однопричинные**; они находятся в отношении **взаимооднозначного соответствия**; *X* — есть кодовое воплощение *A* или, короче, код *A*. Такого рода связь можно назвать **кодовой зависимостью**, она образуется в филогенезе и онтогенезе самоорганизующейся системы и служит элементом ее организации. Отсюда выводится ряд следствий, имеющих прямое отношение к разработке проблем AGI. Попытаемся их кратко рассмотреть:

1) Это, прежде всего, результаты анализа кодовой зависимости и основательное осмысление теоретических и методологических вопросов кодирования информации и расшифровки кодов, что служит созданию новых интерфейсов «мозг-компьютер-машина», решению задач эффективной коммуникации между человеком и роботом, между роботом и внешними объектами, между самими роботами. Операции кодирования информации и расшифровки кода — ключевые факторы передачи и приема («понимания» агентом) информации, организации системы управления, а постольку играют первостепенную роль в области создания компьютерных программ на пути к AGI. Этот круг вопросов, на мой взгляд, пока еще слабо изучается под углом проблематики AGI.

Между тем направление нейронауки, именуемое «Чтением мозга» (“Brain-Reading”), достигло в последние годы значительных успехов в расшифровке мозговых кодов психических явлений. Приведу лишь два примера. Нейроученые из Киотского университета совместно со специалистами по ИИ предложили новый, весьма эффективный способ непосредственно визуализировать по данным фМРТ изображения, которые видит человек в момент сканирования мозга<sup>1</sup>. Впечатляющий результат получен

---

<sup>1</sup> Guohua Shen, Tomoyasu Horikawa, Kei Majima, Yukiyasu Kamitani. Deep image reconstruction from human brain activity” | BioRxiv | doi: 10.1101/240317 Posted December 28, 2017.

в области прямого перевода из мозга (т. е. декодирования) текста из 30–50 предложений на английском языке с помощью отведения сигналов примерно от 250 пунктов перисильвиевой коры мозга испытуемого. При этом средняя частота ошибок в словах не превышала 3%<sup>1</sup>. Эти достижения указывают на возможности новых подходов и перспектив в разработке AGI.

2) Предлагаемый информационный подход, включающий анализ *кодовой зависимости*, позволяет обосновать специфику **информационной причинности** по сравнению с физической причинностью. Это различие определяется *принципом инвариантности информации по отношению к физическим свойствам ее носителя*. Возьмем простой пример: я говорю вам: «Подайте мне, пожалуйста, эту книгу», и вы даете ее мне. Очевидно, что это действие зависит не от самих по себе физических свойств звукового сигнала, а именно от информации на основе сложившейся у вас кодовой зависимости. То же самое действие я могу вызвать у вас множеством других сигналов с другими физическими свойствами. В этом и состоит специфика информационной причинности, которая не противоречит физической причинности, но представляет по сравнению с ней *новый тип* вызываемых агентом *следствий*, характерный для причинно-следственных отношений в функционировании биологических, социальных и ряда технических систем. **Психическая, ментальная причинность**, стоявшая для естествознания под большим вопросом, **является видом информационной причинности**. Тем самым она получает научное объяснение, позволяет дать обоснованный ответ на классический вопрос о воздействии ментального на физическое. Это открывает новые возможности моделирования и реализации человеко-машинных взаимодействий, особенно в интеллектуальной робототехнике, в создании эффективных систем гибридного интеллекта, что имеет прямое отношение к проблематике AGI. Ведь для нынешних систем ИИ характерно то, что они

---

<sup>1</sup> Joseph G. Makin, David A. Moses and Edward F. Chang (2018). Machine translation of cortical activity to text with an encoder–decoder framework // Nature-Neuroscience Technical Report <https://doi.org/10.1038/s415938>



плохо работают с причинно-следственными связями. Для агента же, способного действовать в широком диапазоне сред, необходимо «понимание» причинно-следственных связей, объективно присущих той среде, в которой он должен учиться ориентироваться и действовать. Если речь идет об агенте уровня AGI, то он должен научиться различать физические и информационные причины, а среди последних — ментальные причины, развивать способность «понимать их значение» в процессах коммуникации. В этой связи допустима возможность моделирования основных параметров изменяющейся среды на языке информационных причинно-следственных отношений с целью расширения способности самообучения в новой ситуации систем ИИ, развивающихся в направлении к AGI.

3) Поскольку  $A$  и  $X$ , т. е. явление CP и его мозговой носитель (нейродинамическая кодовая структура) выступают как *однопричинные, одновременные* и находятся в отношении *взаимодозначного соответствия*, то отсюда следует, что описание ряда функциональных свойств данного явления CP может служить в качестве первичной модели учета аналогичных свойств ее нейродинамической кодовой структуры. Речь идет о таких параметрах явления CP, как выделенный его временной интервал, формальные признаки (категоризованность и др.), специфический контент, операциональные характеристики и т. д. Это служит основанием для нейронаучных исследований по установлению корреляций между определенными психическими явлениями и их нейродинамическими показателями, которые нередко довольно четко воспроизводятся при повторных экспериментах. Наряду с вопросом, в какой мере эти корреляты отображают реальную кодовую структуру данного явления CP (они остаются весьма дискуссионными), остаются большие трудности корректного выделения и описания того явления CP, которое является объектом нейронаучного исследования. Материалы таких исследований представляют все же определенный интерес для проблемы расшифровки кодов, взятой в широком смысле, в том числе и в плане развития ИИ.

4) Обозначенный в начале предыдущего пункта характер взаимосвязи  $A$  и  $X$  позволяет вывести еще одно важное следствие.

Наша психическая деятельность включает не только непосредственную данность мне, моему Я определенного явления СР, но и *способность оперировать, управлять им в довольно широком диапазоне*. Скажем, я вижу сейчас **A**, но в следующий момент хочу видеть предмет **B**, переключая внимание и переживая его образ. Или я просто лежу с закрытыми глазами и «листаю» из памяти (по своему желанию) различные образы, обдумывая с определенной целью каждый из них. Виды произвольного управления своими явлениями СР могут быть исключительно разнообразными. Здесь важно подчеркнуть именно фактор произвольности. Но что означает, что, переживая образ **A**, я в следующий момент переживаю по своей воле образ **B**? Ведь образ **A** воплощен в нейродинамической системе **X**, а образ **B** воплощен в другой нейродинамической системе **Y**. Это означает, что, управляя по своей воле переходом от **A** к **B**, я управляю переходом от **X** к **Y**, т. е. соответствующими своими/мозговыми нейродинамическими системами. Это можно выразить в общем виде: способность управлять своими явлениями СР, своими образами, мыслями, есть способность управлять определенным классом своих мозговых нейродинамических систем. Каждый из нас постоянно делает это (часто не лучшим для себя образом), хотя не подозревает о такой способности своего Я. Эта способность выражает фундаментальное свойство Эго-системы нашего мозга (осуществляющей функции нашего Я) — свойство самоорганизации и саморегуляции. Произвольное действие в этом отношении является актом самодетерминации. Вопросы самоорганизации и саморегуляции стоят на первом плане, когда речь идет о создании AGI. Анализ опыта нейронаучных исследований **функций самодетерминации и саморегуляции**, взятых в единстве их информационного содержания (представленного в форме СР) и, если так можно сказать, «аппаратного исполнения» головным мозгом, может быть весьма полезен для осмысления подходов к моделированию указанных функций в процессе разработки AGI.

Все изложенное выше в четырех пунктах получило более развернутое рассмотрение и обоснование в упоминавшейся книге «Проблема “Сознание и мозг”: Теоретическое решение».

Но и то, что сказано здесь, является ответом на вопрос: зачем нужна «Трудная проблема сознания?» для тех, кто действительно настроен на решение новых и трудных теоретико-методологических и практических задач, стоящих перед нами на пути к AGI.

### **Субъективная реальность: ее общие свойства, ценностно-смысловые и интенционально-волевые оперативные структуры**

Исследования сознания с его специфическим и неотъемлемым качеством СР позволяют выяснить и описать основные системные, структурные, функциональные, операциональные характеристики естественной мыслительной деятельности человека. Этот круг вопросов подробно обсуждался в ряде работ<sup>1</sup>. Указанные характеристики СР и анализ ее общих свойств может служить важным ресурсом для моделирования когнитивных структур, отвечающих задачам построения AGI.

Рассмотрим подробнее в этом плане само качество СР и основные динамические структурные и функциональные свойства СР в ее ценностно-смысловых и интенционально-волевых проявлениях.

1. Понятие СР означает всякое осознаваемое состояние, в отличие от бессознательных процессов, которые существуют одновременно с ним и служат его неперменной основой. Оно охватывает как отдельные осознаваемые явления СР и их виды (ощущения, восприятия, чувства, мысли, желания, целеполагания, волевые усилия и т. д.), так и их целостное образование, объединяемое нашим Я.

СР *персональна*, переживается индивидом как «текущее настоящее», т. е. *сейчас*, хотя может отображать прошлое

---

<sup>1</sup> Лекторский В.А. Субъект, объект, познание. М., 1980; Дубровский Д.И. Проблема идеального. Субъективная реальность. М.: Канон+, 2002. 368 с.; *Он же*. Субъективная реальность // Философская электронная энциклопедия. М.: Институт философии РАН, 2020. URL: [https://iphlib.ru/library/collection/elphilenc/document/dubrovskii\\_subektivnaya\\_realnost](https://iphlib.ru/library/collection/elphilenc/document/dubrovskii_subektivnaya_realnost)

и будущее. Состояние СР удостоверяет для меня реальность моего существования. Речь идет о непрерывном процессе, который временно прерывается в условиях глубокого сна без сновидений, общего наркоза, обморока, в иных случаях временной потери сознания (в этих интервалах я для себя не существую). Состояние СР выражает актуально протекающий процесс мышления, когнитивную и деятельную активность личности. Не говоря уже о повседневной, обыденной сознательной деятельности, все новые фундаментальные идеи в науке и культуре первично возникают, оформляются и развиваются в форме явлений СР отдельных личностей.

2. СР в своем специфическом качестве присуща не только сознанию человека, но и психике животных, о чем свидетельствуют опыт общения с ними и данные зоопсихологии об их ощущениях, восприятиях и свойственных высшим животным формах мышления (наличие качества СР отчетливо подтверждается, например, известными опытами с воздействием галлюциногенов на собак).

СР — находка биологической эволюции. У первых одноклеточных организмов носителем информации и регулятором поведения служили химические процессы. Речь идет о *допсихическом уровне информационной реальности*. Следующим этапом ее развития стало зарождение психики, обладающей качеством СР, возникшим у многоклеточных животных, способных активно передвигаться в постоянно изменяющейся внешней среде.

Это ознаменовало возникновение **нового типа информационного управления**, регуляции поведения, позволило решить проблему эффективной самоорганизации многоклеточного организма. Ведь его элементами являются отдельные клетки, которые представляют собой самоорганизующиеся системы со своими жесткими программами, «отработанными» эволюцией в течение многих сотен миллионов лет. Но теперь последние должны были согласовываться с общеорганизменной программой и наоборот. Это — весьма сложная задача, решение которой предполагало нахождение оптимальной *меры централизации и автономизации* контуров управления, способной обеспечить

сохранение и укрепление целостности сложной живой системы, т. е. меры централизации управления, которая не нарушает фундаментальные программы отдельных клеток, и меры автономности их функционирования, которая, наряду с кооперативными и конкурентными способами их взаимодействия между собой, не препятствует их содружественному участию в реализации программ управления целостным организмом.

Аналогичная проблема эффективной самоорганизации личности, социальных и технических систем актуальна и для ИИ, когда появляются подобные задачи взаимоотношения глобальных и локальных уровней управления в сложных системах, взаимодействия множества компьютеров (и роботов), человека и роботов, автономной системы роботов и т. п. Данная проблема решается в сложной биологической системе *эффективнее*, чем в случаях взаимоотношений человека и общества, человека и технических устройств, в организации технических систем. Это требует внимания и изучения, когда речь идет о новых разработках в области ИИ, особенно же AGI и интеллектуальной робототехники, связанных с аналогичными задачами эффективного управления.

Возникновение у животных СР стало исторически первой формой *виртуальной реальности*, открывающей по мере развития все более широкий диапазон способностей к абстрагированию, обобщению, прогнозированию, планированию, *пробным действиям в виртуальном плане*, иным схожим операциям (в форме «мысленных экспериментов»), повышающим приспособляемость к среде. Это качество *виртуальности* во многом определяет способность формирования новых навыков, приспособления к изменившейся среде, т. е., по существу, главное свойство, необходимое для AGI, — способность определять и самостоятельно решать задачи «в широком диапазоне сред».

3. В ходе антропогенеза произошло качественное развитие СР. Возникает сознание (как высшая форма СР), а с ним язык. Особенностью сознания по сравнению с животной психикой является то, что психическое отображение и управление становятся объектом отображения и управления. Появляется способность, по сути, неограниченного производства информации об инфор-

мации, возможность высокой степени абстрагирования, резкого расширения диапазона мысленных действий, моделирования ситуаций, проектирования, целеполагания и волеизъявления, фантазирования, творческого решения задач.

Качество СР человека и язык открыли принципиально новый этап развития информационных процессов. СР представляют собой два основных свойства психической деятельности: 1) **способность индивида иметь информацию в «чистом» виде** — в том смысле, что ее мозговой носитель для нас элиминирован: когда я переживаю образ дерева, мне дана информация об этом предмете и отображение мной данной информации, т. е. знание о том, что именно я вижу это дерево; но я ничего не знаю, не чувствую, что при этом происходит в моем головном мозгу; а в нем происходят очень сложные, многоступенчатые сетевые процессы переработки информации, предшествующие, например, переживанию мной ощущения красного или образа бегущего ко мне человека, т. е. мне дан **итоговый результат** сложных информационных процессов в той **интегральной** форме («простой» и понятной), которая необходима для адекватного поведения и решения задач в данной меняющейся ситуации; 2) вместе с тем в явлениях СР нам дана и **способность оперировать этой информацией в «чистом» виде по своей воле** в широком диапазоне (контролировать собственное Я, переключать внимание, направлять движение своей мысли и концентрировать ее на определенном объекте, давать волю своему воображению, ставить определенные цели и решать мысленно разные задачи). Это означает, что наше Я является **самоорганизующейся** структурой, способной к эффективной **саморегуляции**. Возникает острый вопрос о свободе воли, который вызывает немало трудностей в нейронаучных исследованиях сознания, но может получить объяснение с позиций информационного подхода<sup>1</sup>.

Именно данность информации в «чистом» виде и способность управлять ею выражают специфические черты СР, присущие всякому когнитивному, мыслительному процессу.

---

<sup>1</sup> Dubrovsky D.I. The Problem of Free Will and Modern Neuroscience // Neuroscience and Behavioral Physiology. 2019. Vol. 49. N 5. P. 629–639.

4. Динамическая структура СР с ее многообразными способами продления и смены переживаемого контента (наличного «содержания») включает в себя такие функциональные регистры, как единство *рефлексивного* и *арефлексивного*, *актуального* и *диспозиционного*. Рефлексивное — это длящееся *сейчас осознаваемое* переживание данного контента, которое в следующий момент «уходит» в память, становится арефлексивным (заменяется другим рефлекслируемым контентом, но затем может снова «возвращаться», становится рефлекслируемым). Актуальное — это рефлексивное, но взятое в его отношениях и зависимости от *диспозиционного*, которое представляет собой ценностно-смысловые структуры, хранящиеся в памяти или на более глубоком психическом, а затем и на еще более глубоком допсихическом (генетическом) уровне неосознаваемых информационных процессов. Диспозиционное в большей степени выражает *устойчивые, ценностно-смысловые образования нашей психики* (усвоенные нормы культуры, убеждения, проверенные опытом знания, привычки, умения, интересы, предпочтения и т. п.), которые задают направленность актуализуемых в данном интервале желаний, мыслей и действий; в ряде случаев такая направленность становится для личности непреодолимой или труднопреодолимой (как, например, при наркотической зависимости).

В аспекте рассмотрения регистров рефлексивного и арефлексивного, актуального и диспозиционного для разработки ИИ представляет интерес функционирование нашей памяти в связи с тем, что компьютерам присуща архитектура фон Неймана, основанная на постоянном разделении памяти и вычисления. СР демонстрирует принципиально иное устройство памяти, способы ассоциативной, контекстуально значимой выборки определенного контента из памяти, замены его другим, управления этими информационными процессами.

5. Охарактеризуем ключевые функциональные структуры СР и присущие ей *операциональные способности*, которые могут иметь непосредственное отношение к моделированию когнитивных архитектур, соответствующих уровню AGI.

Элементарный анализ раскрывает фундаментальное структурно-операциональное свойство всякого явления СР — его **дву-мерность**, т. е. способность одновременного отображения в нем некоторого объекта и вместе с тем самого себя, т. е. **единство иноотображения и самоотображения**. СР в любых ее интервалах и проявлениях обладает в определенной степени способностью самоотображения на уровне нашего Я. Эта **фундаментальная** способность именуется в психиатрии **чувством принадлежности**; его нарушение вызывает психическое расстройство — различные формы **деперсонализации**<sup>1</sup>. При этом деперсонализация влечет состояние **дереализации** — различные формы нарушений отображения внешних объектов и структуры собственного тела, что подтверждает органическую связь, взаимозависимость самоотображения и иноотображения. Данные психопатологии и психоневрологии предоставляют для этого исключительно ценный материал<sup>2</sup>.

Указанная взаимосвязь отчетливо проявляется на уровне зрительного восприятия в исследованиях зеркальных нейронов, которые одновременно осуществляют «кодирование сенсорной информации в моторных терминах», т. е. «создают единство зрительного восприятия и его вероятной моторной репрезентации»<sup>3</sup>. Благодаря зеркальным нейронам, мы способны незамедлительно распознавать определенные типы действий, производимых другими, достигать **понимания** другого.

Важно учитывать, что на моторных репрезентациях наблюдаемых действий, создаваемых зеркальными нейронами, **«основана способность обучения при помощи подражания»**<sup>4</sup>.

---

<sup>1</sup> Меграбян А.А. Деперсонализация. Ереван: Армгосиздат, 1962.

<sup>2</sup> См.: Назлоян Г.М. Концептуальная психотерапия. Портретный метод. М.: ПЕР СЭ, 2002; Рамачандран В.С. Мозг рассказывает. Что делает нас людьми. М.: Карьера Пресс, 2014; Литвак Л.М. «Жизнь после смерти»: предсмертные переживания и природа психоза. Опыт самонаблюдения и психоневрологического исследования / Под ред. и со вступительной статьей Д.И. Дубровского. М.: Канон+, 2007

<sup>3</sup> Рицолатти Дж., Синигалья К. Зеркала в мозге: О механизмах совместного действия и сопереживания. М.: Языки славянских культур, 2012. С. 97.

<sup>4</sup> Там же.



Это обстоятельство может служить стимулом для поиска путей построения моделей саморасширения диапазона «навыков» интеллектуального робота, необходимого для AGI.

6. Взаимозависимость иноотображения и самоотображения полнее всего выражена на уровне многомерной динамической структуры нашего Я, т. е. в том, что можно назвать *базисной динамической структурой* СР. Она представляет собой *единство и переменное соотноснение противоположных модальностей «Я» и «не-Я»*. Это единство представлено в каждом наличном интервале СР, оно формирует его ценностно-смысловой каркас и деятельно-волевые векторы. В динамическом биполярном контуре «Я» — «не-Я» совершается движение «содержания» (контента) субъективной реальности. Это «содержание» способно переходить из модальности «Я» в модальность «не-Я» и наоборот (например, когда «содержание», относящееся к модальности «Я», мои личностные свойства, мои оценки себя становятся для меня объектом внимания, анализа и оценки, а значит, выступают в данном интервале уже в модальности «не-Я» и т. п.). Такого рода взаимопреобразования, перемена модальности переживаемого «содержания» — это *механизм эффективного отображения и освоения действительности*, в том числе самой СР, ее саморегуляции, а вместе с тем механизм освоения социального опыта и осуществления творческой деятельности.

Взаимопереходы модальностей «Я» и «не-Я» постоянно сохраняют биполярную структуру СР в любом ее интервале, не нарушая идентичности персонального Я (исключая психопатологию). *Каждая из модальностей определяется лишь через противопоставление другой и соотносение с ней*. Поэтому в самом общем виде «Я» есть то, что противопоставляется «не-Я» и соотносится с ним; и наоборот, «не-Я» есть то, что противопоставляется «Я» и соотносится с ним. Первостепенный интерес для понимания операциональных структур системы «Я», «Самости» представляют исследования выдающегося нейропсихолога Антонио Дамасио, прежде всего его анализ «автобиографической самости»<sup>1</sup>.

<sup>1</sup> Дамасио А. Так начинается «я». Мозг и возникновение сознания. М.: Карьера-Пресс, 2018. 384 с.

Базисная динамическая структура субъективной реальности раскрывается конкретнее, когда в ходе феноменологического анализа выявляются основные виды противопоставления и соотношения «Я» с «не-Я». Если взять за систему отсчета модальность «Я», то в первом приближении «Я» выступает по отношению к «не-Я» как отношение «Я»: 1) к внешним объектам, процессам (внешней предметности); 2) к собственному телу; 3) к самому себе; 4) к другому «Я» (другому человеку); 5) к «Мы» (той социальной общности, группе, с которой «Я» себя идентифицирует, к которой оно себя в том или ином отношении причисляет, ценности которой разделяет); 6) к «Они» (той общности, социальной группе, которой «Я» себя противопоставляет или, по крайней мере, от которой оно себя отделяет, ценности которой резко отрицает или равнодушно не принимает, не понимает); 7) к «Абсолютному» («Мир», «Бог», «Космос», «Природа» и т. п.).

Таков один из мыслимых способов выделения основных видов «содержания» (контента) «не-Я», а следовательно, самого «Я», поскольку оно полагает и раскрывает себя посредством своего «не-Я». Иными словами, таковы типичные смысловые (когнитивные и ценностные) измерения нашего «Я» (**категории контентов Я**). В этой биполярной динамической контентной многомерности «Я» и полагает себя как свое «не-Я», выступающее в форме «знания», «оценки» и «действия». В ней непрерывно совершаются процессы самоотображения и самоорганизации структуры СР, формируется и реализуется ее активность, ее **дейтельно-волевые функции**. Именно эти свойства и функциональные регистры СР определяют необходимые для АГІ способность к самообучению и саморегуляции в изменяющейся среде, высокую степень автономности агента.

7. Приведенные результаты феноменологического анализа динамической структуры СР для тех специалистов кто нацелен на разработку АГІ, могут быть полезны в ряде отношений. Прежде всего, речь идет о ясном понимании той огромной дистанции, которая отделяет нынешний уровень ИИ от ЕИ, от специфических функциональных свойств естественного процесса мышления, к моделированию и реализации которых у нас пока

нет технологического доступа. Но, занимая реалистическую позицию, мы сохраняем веру в наши *творческие* способности для того, чтобы добиваться шаг за шагом сближения с ЕИ, т. е. **добиваться все более глубокого самопознания и самопреобразования, т. к. эти функции исходно присущи нашему ЕИ.**

На каждом этапе мы используем методологические установки, позволяющие корректно сокращать чрезвычайную многомерность СР (в ее динамических операциональных и творческих проявлениях), выбирать и осваивать *некоторые* ее специфические функциональные свойства, опираться на приемлемые аналогии в операциональных действиях мышления человека сравнительно с интеллектуальным роботом (поскольку он совершает физически реальные действия во внешней среде, подлежащие строгой проверке на их эффективность).

Если рассматривать *единство иноотображения и самоотображения как необходимое условие эффективности мыслительных и практических действий человека*, то интеллектуальный робот обладает в известной мере аналогичными функциями. У него есть четко заданное иноотображение (и соответствующий набор внешних действий). Существует и нечто, подобное самоотображению, которое может быть заложено в надежности его программного обеспечения и технического устройства, в механизмах слежения за их исправностью. Но сразу при этом мы видим и качественное различие. У человека функция самоотображения обладает способностью *саморегуляции*, которая постоянно стимулируется, корректируется, совершенствуется процессами иноотображения, как и наоборот: иноотображение стимулируется, корректируется, повышает результативность под воздействием процессов самоотображения, т. е. обладает способностью саморегуляции. В этом и заключается свойственное сознанию, структуре СР единство функций иноотображения и самоотображения, их органическая взаимообусловленность, которая особенно ярко проявляется на уровне базовой динамической бимодальной структуры субъективной реальности — в единстве и переменном соотношении модальностей «Я» и «не-Я». Что касается робота, то присущая ему функция иноотображе-

ния и то, что условно можно было бы отнести у него к функции самоотображения, фактически разделены.

Это указывает на проблемы, требующие разработки при создании интеллектуальных роботов нового поколения, т. е. достигающих уровня AGI. В моделировании интеллектуальных роботов могут использоваться результаты исследования сознания, что ведет к расширению «мира» робота, его «техно-умвельта», тем самым и диапазона его успешных действий (А.Р. Ефимов). В этом направлении, видимо, будет возможность использовать ресурсы, связанные с развитием методов глубоких нервных сетей для создания программ указанной выше **двумерной функциональности** в поведении интеллектуального робота.

Разумеется, пунктирно намеченные возможности новых подходов нуждаются в дальнейшей конкретизации и критическом обсуждении. Но они свидетельствуют о широкой творческой перспективе применения феноменологических исследований операциональной специфики сознательной деятельности.

8. Изложенные выше результаты феноменологических исследований СР заставляют обратить внимание на чрезвычайную сложность «содержания» (множества различных *контентов*) нашей сознательной деятельности, взаимосвязь, взаимовлияние различных по своему виду (классу) контентов. Они приблизительно определены и перечислены нами выше. Эти различные категориальные структуры *диспозиционально* укоренены в нашем сознании и на бессознательном уровне психики. Они обычно вовлечены в той или иной степени во все мыслительные процессы, задают широкий по их «содержанию» *спектр операций абстрагирования, анализа и синтеза*. Такие операции различаются по когнитивным свойствам в зависимости от того или иного категориального контента. Одно дело операции абстрагирования, анализа и синтеза при исследовании предметной области внешнего природного мира, другое — при исследовании нашего Я или социокультурных реалий.

Различие «онтологий» обуславливает соответственные особенности эпистемологических подходов. Тем не менее при решении сложных теоретических и особенно практических задач

мы, как правило, используем широкий спектр этих операций. Многие из них включают *интуитивный уровень*, связанный порой с новыми, находящимися в стадии формирования творческими результатами. Из многообразного спектра в системах ИИ используется лишь небольшая часть когнитивных операций, но в достаточной степени определенных. В этом состоит одно из главных отличий ЕИ от ИИ. Все это должно служить поводом для размышлений при разработке AGI, поскольку в отношении него не требуется способность решения всего разнообразия человеческих задач. Но важно уточнять те их классы, которые соответствуют его целям и реальным возможностям на нынешнем этапе. Среди них особую актуальность в последнее время приобретают этические проблемы развития систем ИИ, которые требуют специального основательного исследования.

9. В многомерных планах проблемы сознания целесообразно выделить еще один вопрос. Он касается не просто когнитивных операций (абстрагирования, обобщения и др.) в различных категориальных контентах, указанных выше, а учета множества экзистенциально значимых состояний сознательной деятельности, которые, хотя и включены в процесс познания, играют в нем существенную роль, но не могут быть достаточно четко выражены в принятых когнитивных терминах. Все они, однако, требуют учета и охвата для наличия целостного образа сознания, полноты анализа познавательного процесса и оценки его целей, т. к. в настоящее время, в условиях нарастания глобального кризиса нашей потребительской цивилизации, особое значение приобретают вопросы о подлинных смыслах личной и социальной деятельности, о противостоянии нарастающему абсурду в нынешних процессах социального развития.

Известный феноменолог М. Мерло-Понти писал: «Жизнь сознания — познающая жизнь, жизнь желания, или жизнь перцептивная — скрепляется “интенциональной дугой”, которая проецирует вокруг нас наше прошлое, будущее, наше житейское окружение, нашу физическую, идеологическую и моральную ситуации или, точнее, делает так, что мы оказываемся вовлечены во все эти отношения. Эта интенциональная дуга и создает единство

чувств, единство чувств и мышления, единство чувствительности и двигательной функции»<sup>1</sup>. Понятие *интенциональной дуги* выполняет интегративную роль, т. е. средства охвата и концептуального учета всего диапазона «участников» познавательного процесса, от перцептивного уровня до идеологических, моральных и иных, экзистенциально значимых составляющих познавательного процесса, часть которых нередко остается в тени.

М. Мерло-Понти не назвал в приведенном перечне эстетическую составляющую, эстетическую интенцию, которая в той или иной степени всегда присутствует в познавательных процессах (часто в неявной форме, но, тем не менее, дает о себе знать в виде равнодушия познающего субъекта к красоте и безобразию, к тому, что называют возвышенным и низменным, трагическим и комическим, *гармоничным*, изящным и т. п.). Все это приобретает высокую степень выражения и ценности в художественном познании, в творчестве выдающихся представителей литературы и искусства. Справедливо признать, что наиболее глубокие и разносторонние сведения о сознании даны в произведениях Пушкина, Толстого, Чехова, других великих писателей и поэтов.

Несомненно важное значение поэзии для эпистемологического анализа обыденного и научного мышления. Поэтические образы и метафоры — мощный творческий стимул развития интеллекта, т. к. разрушают устоявшиеся комфортные клише, решительно и убедительно пересекают границы привычных противоположных смыслов, обнаруживают ранее не известные степени свободы в нашем мышлении, создают в нем новые ценностно-смысловые образования и векторы активности, просветляют наш духовный мир.

Эстетическая интенция в познавательных процессах не менее значима, чем этическая, что находит отражение в контексте ЕИ, и должна учитываться при обсуждении проблем разработки AGI. Следует иметь в виду не только когнитивный аспект, способ решения определенной задачи, но и оценку ее допустимости с позиций этических норм и принятых социальных критериев.

---

<sup>1</sup> Мерло-Понти М. Феноменология восприятия. СПб.: Ювента, 1999. С. 182.

Особенно важная и трудная тема в проблематике развития ИИ — выяснение и использование для целей AGI операций бессознательно-психического уровня и *допсихического*, генетически заданного уровня, на котором осуществляется управление сложнейшими процессами в организме. Биологическая эволюция выработала многочисленные информационные компетенции эффективного управления сложными системами, которые унаследованы человеком, но остаются малоисследованными. Подтверждением может служить следующий известный пример. В нашем теле присутствует 100 000 километров кровеносных русел, и эта колоссальная динамическая система эффективно управляется при постоянном изменении внешних и внутренних физических условий. Несомненно, что допсихический уровень информационных процессов оказывает, пусть опосредствованно, существенное влияние на нашу когнитивную деятельность, что требуется учитывать.

В заключение важно еще раз обратить внимание на ключевую роль совершенствования системы образования для успешной разработки AGI, о чем говорил в обращении к читателям книги об AGI Герман Греф: «Современные образовательные программы готовят инженеров, умеющих работать с существующими технологиями, а не творчески мыслящих исследователей, способных совершать новые прорывы»<sup>1</sup>. В этом проявляется слабость, формализм, в ряде отношений интеллектуальное убожество современного образования. Чтобы воспитывать творческую устремленность, оно должно делать акцент на фундаментальных теоретических вопросах развития ИИ, формировать у молодого инженера широкий духовный кругозор, интерес к философским и социальным проблемам познания и самопознания, к поэзии, гениальной художественной литературе как источникам и стимулам творческого мышления.

---

<sup>1</sup> Сильный искусственный интеллект. С. 55.

## ЕСТЕСТВЕННЫЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: СМЫСЛЫ ИЛИ СТРУКТУРЫ?

---

---

Само понимание искусственного интеллекта, как ни странно, недостаточно разработано и понимается очень по-разному не только представителями разных областей знаний, но даже внутри самого этого профессионального пространства. Между тем без договоренности об определениях дискуссия о том, сможет ли ИИ обрести сознание и оставаться управляемым людьми, бессмысленна. Одним из важнейших исследований и разработок точной эпистемологии и эвристик на этом поле является книга В.К. Финна «Интеллект, информационное общество, гуманитарное знание и образование»<sup>1</sup> и подробный разбор её В.Б. Тарасова<sup>2</sup> (2021 г.). Эта книга анализирует интеллектуальный процесс как

---

\* *Черниговская Татьяна Владимировна* — доктор биологических наук, доктор филологических наук, профессор Санкт-Петербургского государственного университета, директор Института когнитивных исследований СПбГУ, член-корр. РАО.

*Tatiana V. Chernigovskaya* — doctor of Sciences in Biology, doctor of Sciences in Philology, professor at St Petersburg University, Head of the Institute for Cognitive Studies, member of the Russian Academy of Education.

<sup>1</sup> *Финн В.К.* Интеллект, информационное общество, гуманитарное знание и образование. М.: ЛЕНАНД, 2021.

<sup>2</sup> *Тарасов В.Б.* Философия искусственного интеллекта, система когнитивных наук, компьютерная поддержка научных исследований и образования // Искусственный интеллект и принятие решений. 2021. № 2. С. 93–109.



таковой — его стадии и компоненты, и важнейшие вопросы — не только технического, но даже экзистенциального характера. Но даже эта замечательная работа не решает вопросы семиозиса, и важнейший из них — о том, возможны ли смыслы в потенциальных мирах искусственных интеллектуальных систем.

Наше время особенно остро ставит перед наукой вопросы, никогда не уходившие из философского дискурса, но странным образом занимавшие периферическое место в нейронауке. При обсуждении т. н. *трудных* проблем, в частности психофизической, имеющей долгую историю, многократно подчеркивалось, что их понимание также отличается внутри разных областей знания (работы В.А. Лекторского<sup>1</sup>, Д.И. Дубровского<sup>2</sup>, Y.I. Alexandrov, M.E. Sams<sup>3</sup>; Ю.И. Александрова<sup>4</sup>, Д. Чалмерса<sup>5</sup>, А.Я. Каплана<sup>6</sup>, Т.В. Черниговской<sup>7</sup>).

---

<sup>1</sup> *Лекторский В.А.* Исследование интеллектуальных процессов в современной когнитивной науке: философские проблемы // Естественный и искусственный интеллект / Под ред. Д.И. Дубровского, В.А. Лекторского. М.: Канон+, 2011. С. 3–16.

<sup>2</sup> *Дубровский Д.И.* Актуальные проблемы интересубъективности // Естественный и искусственный интеллект / Под ред. Д.И. Дубровского, В.А. Лекторского. М.: Канон+, 2011. С. 129–148.

<sup>3</sup> *Alexandrov Y.I., Sams M.E.* Emotion and consciousness: Ends of a continuum // Cognitive brain research. 2005. Vol. 25(2). P. 387–405.

<sup>4</sup> *Александров Ю.И.* От теории функциональных систем к системной психофизиологии // Психология сегодня: теория, образование, практика / Под ред. А.Л. Журавлева, Е.А. Сергиенко, А.В. Карпова. М.: Институт психологии РАН, 2009. С. 13–56.

<sup>5</sup> *Чалмерс Д.* Сознательный ум. В поисках фундаментальной теории. М.: URSS, Либроком, 2013.

<sup>6</sup> *Каплан А.Я.* Мы — это больше, чем наш мозг. В поисках субъективного начала // Труды кафедры богословия Санкт-Петербургской духовной академии. Теология и современные исследования сознания, 2019. №2 (4). С. 25–34.

<sup>7</sup> *Черниговская Т.В.* Языки сознания: кто читает тексты нейронной сети? // Человек в мире знания: в честь 80-летия акад. В.А. Лекторского. М.: РОССПЭН, 2012. С. 403–415; *Черниговская Т.В.* Чеширская улыбка кота Шрёдингера: мозг, язык и сознание. М.: АСТ, 2021; *Черниговская Т.В.* Нейронаука в поисках смыслов: мозг как барокко? // Вопросы философии. 2021. № 1. С. 17–26; *Chernigovskaya T.V.* (2020). Biology, Environment, and Culture: From Animal Communication to Human Language and Cognition. Vestnik Sankt Peterburgskogo universiteta. Filosofii i konfliktologii. Vol. 36 (1). P. 157–170.

Экспериментальная нейронаука по умолчанию считала, что, если мы узнаем свойства нейронов и их взаимодействий, то поймем, что такое сознание и дух (F. Crick, C. Koch, 2007)<sup>1</sup>. Так, конечно, считают отнюдь не все, но большинство представителей естественных наук. Есть и другой полюс: например, Дж. Сёрль говорит, что сознание нередуцируемо, для него иллюзия — такая же реальность, что естественная наука принять не может<sup>2</sup>. В программной статье К.В. Анохин выделяет *главную* проблему «разум-мозг» и *трудную* проблему «сознание-мозг» и описывает задачи, которые для продвижения к успеху должны быть решены<sup>3</sup>. Это не отменяет парадоксов, что мозг находится в мире, а мир находится в мозге (В.А. Лекторский, 2011)<sup>3</sup>, что внешний мир строится изнутри (В.П. Зинченко, 2010)<sup>4</sup>. Нельзя не согласиться, что в рамках существующих представлений о работе мозга и с помощью принятых подходов главные проблемы решены быть не могут. В этой статье, как и в опубликованной ранее статье «Нейронаука в поисках смыслов: мозг как барокко?»<sup>11</sup>, я предлагаю подходить к этой сверх-сложной задаче, не только разрабатывая традиционные для нейронауки методы и парадигмы, но и используя «археологию» ментальных принципов, проявившихся в семиотическом поведении человека, в частности, в искусстве.

Вечные вопросы по-прежнему требуют ответов от естественных наук, в частности от физиологии и генетики, но и от антропологии, а теперь и сферы искусственного интеллекта.

Предлагается свести вопросы к четырем блокам:

- Достаточно ли средств естественных наук, чтобы найти принципы и механизмы, по которым действует мозг в его

---

<sup>1</sup> Crick F., Koch C. A neurobiological framework for consciousness. M. Velmans, S. Schneider (eds.). The Blackwell Companion to Consciousness. Blackwell. 2007. P. 567–579.

<sup>2</sup> Сёрл Дж. Открывая сознание заново. М.: Идея-Пресс, 2002.

<sup>3</sup> Анохин К.В. Когнитом: в поисках фундаментальной нейронаучной теории сознания // Журнал высшей нервной деятельности. 2021. Т. 71. №1. С. 39–71.

<sup>4</sup> Зинченко В.П. Сознание и творческий акт. М.: Языки славянских культур, 2010.

высших, а не только простых умениях, сопоставимых с возможностями других живых существ или искусственными интеллектуальными объектами?

- Язык, сознание, мышление, *qualia* — это уникальные характеристики человека? Как мы можем (если можем) это доказать?

- Как это устроено в мозгу и в каком мозгу? Модули, сети и их свойства, гиперсети, коннектомы и когнитомы...

- Можем ли мы это воспроизвести? не только наш интеллект, но наш внутренний мир?

- Релевантен ли вопрос о смыслах, или речь идет только об имитации семиотического поведения небиологическими системами?

С очевидностью вопросу о свойствах и перспективах развития искусственных интеллектуальных систем сопутствует вопрос — *зачем мы их создаем?* Варианты ответов немногочисленны, что не делает ситуацию проще: (а) Чтобы понять, как устроен человеческий мозг, т. е. для дальнейшей разработки теории. (б) Чтобы понять, что вообще в ментальном мире возможно, в том числе и в том, которого мы не знаем или которого пока нет? (в) Чтобы понять, как нас можно биологически или технически «улучшить», создать новых людей, которые будут быстрее, умнее, с огромной памятью, таких, которые все делают лучше, чем современные люди? (г) Чтобы сделать «франкенштейна», или ещё амбициознее — оцифровать наш ментальный и эмоциональный мир и достичь тем самым бессмертия?

Как пишет философ концептуального искусства Б. Гройс<sup>1</sup>, «на протяжении долгого времени человеку онтологически отводилась средняя позиция между Богом и животным. При этом казалось более престижным стоять ближе к Богу и дальше от животного. Но в Новое время мы обычно располагаем человека между животным и машиной. В этом новом контексте кажется, что лучше быть животным, чем машиной. Интеллектуальными самообучающимися программами типа AlphaZero взяты практи-

---

<sup>1</sup> Гройс Б.Е. Под взглядом теории. М.: Ad Marginem, 2013 // <https://theoryandpractice.ru/posts/7146-groys>.

чески все рубежи: шахматы, го, сёги, даже покер... Было множество восторженных разборов шахматных партий AlphaZero с прежним чемпионом — программой StockFish. AlphaZero побеждала за счет своей “глубокой интуиции”, перебирая “всего” 80 тысяч позиций в секунду (тогда как StockFish — 70 млн), и тем не менее выигрывала. Считается, что она играла более холистично, подчиняя все ходы единой цели, делая, казалось бы, нелепые и даже неверные ходы, если смотреть не вдолгую, и вынуждала оппонента к тому, что называется Zugzwang. Программа использовала “искусственную интуицию” — в противоположность жесткой переборной логике (С.Е. Perez, 2017)<sup>1</sup>. Эту манеру игры описывают как “инопланетную” — так не играют люди, так не играли и программы, созданные людьми. Семантический провал между интуицией и логикой преодолен, и это похоже на когнитивную атаку или даже цивилизационный вызов нашим представлениям об интеллектуальных возможностях человека» (ср.: Д.В. Ушаков, 2011)<sup>2</sup>. Анализ разгромных для лучших игроков в го матчей поразил экспертов: победа над Lee Sodol в 2016 г. и над Ke Jie в 2017 г. показала примерно то же: таких ходов люди не делают и такие стратегии им в голову никогда не приходили (программа может нести потери в тех или иных моментах игры, но отыгрывает эти потери на следующих ходах). Этот результат на этапе проигрышных ходов неочевиден игроку-человеку и наблюдателю, поэтому решения программы порой кажутся странными; программа может жертвовать одним или несколькими камнями, чтобы добиться обладания тактической инициативой — игроки-люди обычно так не поступают<sup>3</sup> (W. Knight, 2017)<sup>4</sup>.

---

<sup>1</sup> Perez C.E. (2017) AlphaZero: How Intuition Demolished Logic // <https://medium.com/intuitionmachine/alphazero-how-intuition-demolished-logic-66a4841e6810>.

<sup>2</sup> Ушаков Д.В. Психология интеллекта и одаренности. М.: Изд-во «Институт психологии РАН», 2011.

<sup>3</sup> Моя благодарность А. Ефимову и С. Шумскому за комментарии.

<sup>4</sup> Kight W. (2017) A Stronger AlphaGo Defeats the World's Number One Player // <https://www.technologyreview.com/2017/05/23/151572/a-stronger-alphago-defeats-the-worlds-number-one-player/>.

Итак, мы столкнулись с когнитивно новым пространством. Конечно, можно сказать, что и человеческий мозг, поиграв в го ещё пару тысяч лет, возможно, додумался бы и до таких стратегий. А если нет? Не следует ли из этого, что нейронаука ищет в мозгу лишь то, о чем знает (что естественно, но недальновидно). Я привожу здесь эти примеры возможностей искусственного интеллекта, так как вижу параллель с обращением к «археологии» мышления/сознания/интуиции через искусство.

Мы замахнулись на повторение повторений — создать тех, кто будет творить миры параллельно или даже вместо нас. На поверхности это выглядит как ученическая задача: пиши как Моцарт, Дюрер, Пушкин; играй с невысказанными скоростями и технически безупречно (этот вектор мы видим и в человеческом искусстве — ещё недавно нельзя было представить себе такие запредельные скорости и техники в исполнении — инструментальном и пластическом). Конечно, при некотором прогрессе, программы будут это делать «лучше» людей. Если, конечно, свести искусство к технике и убрать личность, душу, ум, интерпретации, состояние, т. е. человеческое, слишком человеческое...

Основная работа мозга — семиозис, и он имеет долгую биологическую историю (работы U. Eco<sup>1</sup>, J. Hoffmeyer, K. Kull<sup>2</sup>, K. Kull<sup>3</sup>, Yu. Natchin, T.V. Chernigovskaya<sup>4</sup>). Дискуссии о соотношении социального и биологического в человеке ведутся давно, но с небольшим прогрессом, потому что фактически нет взвешенной позиции (см. исследования А.Г. Козинцева<sup>5</sup>

---

<sup>1</sup> Eco U. Kant and the Platypus. Essays on Language and Cognition. NY: Harcourt Brace & Company, 2000.

<sup>2</sup> Hoffmeyer J., Kull K. Baldwin and biosemiotics: What intelligence is for // B.H. Weber, D. Depew, (eds.). Evolution and Learning: The Baldwin Effect Reconsidered. Cambridge: MIT Press, 2003. P. 253–272.

<sup>3</sup> Kull K. Towards a Theory of Evolution of Semiotic Systems // Chinese Semiotic Studies. 2014. Vol. 10 (3). P. 485–495.

<sup>4</sup> Natchin Yu., Chernigovskaya T. From Archebiosis to Evolution of Organisms and Informational Systems // Biological Communications. 2020. Vol. 65 (3). P. 215–227.

<sup>5</sup> Козинцев А.Г. Происхождение и ранняя история вида *Homo sapiens*: новые биологические данные // Фундаментальные проблемы археологии, антропологии и

и Е.Н. Панова<sup>1</sup>). Умберто Эко в книге «Кант и утконос»<sup>22</sup> обсуждает истоки семиозиса и задает острейшие вопросы: Почему мы пользуемся знаками? Насколько надежны и стабильны связи между ними и тем, что они обозначают? Что вообще заставляет нас высказываться (фило- и онтогенетически)? Нельзя не вспомнить глубокую мысль Пирса о том, что в основе лежит внимание к объекту, причем не ко всем его чертам, а только к релевантным для данной ситуации или конвенции<sup>2</sup>. Это, конечно, ведет к еще более общей проблеме происхождения языка и даже необходимости дефиниций — пониманию того, что мы имеем в виду, когда говорим «язык» — со всем веером вариантов от структурных до функциональных (см. работы Т.В. Черниговской<sup>3</sup>, З.А. Зориной и А.А. Смирновой<sup>4</sup>; Д. Бикертон<sup>5</sup>, Р. Данбара<sup>6</sup>, У.Т. Фитча<sup>7</sup>; Е.А. Cartmil с соавторами<sup>8</sup>). Поразительным образом сходятся взгляды биолога и поэта: Т. Дикон говорит о ко-эволюции языка и мозга, при этом настаивает на том, что «язык оккупировал мозг»<sup>9</sup>. Иосиф Бродский в своей Нобелевской лекции формулирует это

---

этнографии Евразии. К 70-летию академика А.П. Деревянко. Новосибирск: Изд-во ИАЭТ СО РАН, 2013. С. 519–535.

<sup>1</sup> *Панов Е.Н.* Человек — созидатель и разрушитель. Эволюция поведения и социальной организации. М.: Издательский дом ЯСК, 2017.

<sup>2</sup> *Pierce C.S.* Semiotica. Torino: Einaudi, 1980.

<sup>3</sup> *Черниговская Т.В.* Homo Loquens: эволюция церебральных функций и языка // Журнал эволюционной биохимии и физиологии. 2004. Т. 40. № 5. С. 400–406.

<sup>4</sup> *Зорина З.А., Смирнова А.А.* О чем рассказали «говорящие» обезьяны: способны ли высшие животные оперировать символами? М.: Языки славянских культур, 2006.

<sup>5</sup> *Бикертон Д.* Язык Адама: Как люди создали язык. Как язык создал людей. М.: Языки славянских культур, 2012.

<sup>6</sup> *Данбар Р.* Лабиринт случайных связей: Рассказ о том, как мы общаемся, а главное — зачем. М.: ЛомоносовЪ, 2012.

<sup>7</sup> *Фитч У.Т.* Эволюция языка. М.: Языки славянских культур, 2013.

<sup>8</sup> *Cartmill E.A., Roberts S., Lyn H., Cornish H.* (2014) “The Evolution of Language”, Proceedings of the 10th International Conference (EVLANG10) Vienna, 14–17 April 2014, World Scientific Publ., Vienna. P. 479–481.

<sup>9</sup> *Deacon T.W.* (2013) Incomplete nature: How mind emerged from matter, W.W. Norton & Company, NY.

не менее жестко: «Поэт — есть средство существования языка... тот, кем язык жив... Это и есть тот момент, когда будущее языка вмещивается в его настоящее»; и далее: «Пишущий стихотворение пишет его потому, что язык ему подсказывает или просто диктует следующую строчку... стихосложение — колоссальный ускоритель сознания, мышления, мироощущения» (с. 764–765) и наконец: «...при помощи языка анатомизируешь свой опыт» (с. 724)<sup>1</sup>.

Идея повторения миров не нова. Человек — это его мозг, и он вовсе не только обрабатывает информацию, поступающую через органы чувств: он и творит миры, никогда прежде не бывшие, мозг (и не только человеческий в ограниченных объемах) — семиотическое устройство (об этом см. работы J. von Uexküll<sup>2</sup>, С.Т. Золяна с соавторами<sup>3</sup>), и это серьезное препятствие к тому, чтобы переносить сведения о поведении и мозге животных на человека. Смыслы важнее алгоритмов, и из-за сложности — их продукции и расшифровки требуют огромных энергетических затрат: при овладении языком маленький ребенок усваивает гигантские объёмы информации в день (в основном это лексическая семантика), и ясно, что для развития языка смыслы важнее синтаксических структур, позволяющих язык строить и, конечно, являющихся специфичными для человека (F. Mollica, S.T. Piantadosi, 2019)<sup>4</sup>. Разыскания в области происхождения языка и его эволюции прямо связаны с базисными взглядами исследователей, сводимыми к оппозиции структуры и функций и поиском особых зон в мозгу человека, отличных от близких биологических видов (исследования авторов Т. Деакон<sup>5</sup>,

<sup>1</sup> Бродский И.А. Сочинения. Екатеринбург: У-Фактория, 2002.

<sup>2</sup> von Uexküll J. Wiesehe wir die Natur und wiesiehtsiesichselber? // Die Naturwissenschaften. 1922. Vol. 10 (12). P. 265–281.

<sup>3</sup> Золян С.Т., Ильин М.В., Сладкевич Ж.Р., Тульчинский Г.А. Смысл в жизни и смысл в тексте // Слово.ру: балтийский акцент. 2020. Т. 11. №1. С. 7–33.

<sup>4</sup> Mollica F., Piantadosi S.T. Humans store about 1.5 megabytes of information during language acquisition // Royal Society Open Science. 2019. Vol. 6 (3). 181393. <https://doi.org/10.1098/rsos.181393>

<sup>5</sup> Deacon T. Monkey homologues of language areas: computing the ambiguities // Trends in Cognitive Sciences. 2004. Vol. 8 (7). P. 288–290.

L. Pylkkänen<sup>1</sup>, S. Neubauer с соавторами<sup>2</sup>), именно человеческих генетических механизмов и их предпосылок (об этом писали W.R. Clark и M. Grundstein<sup>3</sup>, M. Heide с соавторами<sup>4</sup>). Мы знаем, что в нашем мозгу есть связи, соединяющие передние и латеральные височные отделы с фронтальными отделами лобных долей, образующие так называемый *uncinate fasciculus* — крючковидный пучок, характерный только для мозга человека, хотя и имеющий эволюционные предпосылки (F. Balezeau с соавторами, 2020)<sup>5</sup>. Речь идет о вентральном языковом потоке, обеспечивающем обработку семантических аспектов языка. Данные нейронауки показывают, какие характеристики коннектома обеспечивают функционирование сложнейшей структуры ментального лексикона, позволяют удерживать разные слои памяти, отличать реальность от галлюцинаций, и т. д. (работы авторов К. Hugdahl<sup>6</sup>, М. Kireev с соавторами<sup>7</sup>).

Да, человек своим мозгом порождает смыслы, но результат не может быть достигнут одним индивидом: смысл всегда на пересечении творца и реципиента, это всегда диалог, и успех зависит

---

<sup>1</sup> Pylkkänen L. The neural basis of combinatory syntax and semantics // Science. 2019. Vol. 366. P. 62–66.

<sup>2</sup> Neubauer S., Gunz Ph., Scott N., Hublin J.-J., Mitteroecker Ph. Evolution of brain lateralization: A shared hominid pattern of endocranial asymmetry is much more variable than in great apes // Science advances. 2020. Vol. 6 (7). eaax9935. P. 1–11.

<sup>3</sup> Clark W.R., Grundstein M. (2000) Are we Hardwired? The Role of Genes in Human Behavior, Oxford University Press, Oxford.

<sup>4</sup> Heide M., Haffner Chr., Murayama A., Korotaki Y., Shinohara H., Okano H., Sasaki E., Hattner W. Human-specific ARHGAP11B increases size and folding of primate neocortex in fetal marmoset // Science. 2020. Vol. 369 (6503). P. 546–550.

<sup>5</sup> Balezeau F., Wilson B., Gallardo G., Dick F., Hopkins W., Anwander A., Frederici A., Griffiths T., Petkov Ch. (2020) “Primate auditory prototype in the evolution of the arcuate fasciculus”, Nature Neuroscience, Vol. 23 (5). P. 611–614.

<sup>6</sup> Hugdahl K. Experimental Methods in Neuropsychology. NY: Kluwer Academic Publishers, 2002.

<sup>7</sup> Kireev M., Slioussar N., Korotkov A.D., Chernigovskaya T.V., Medvedev S.V. Changes in functional connectivity within the fronto-temporal brain network induced by regular and irregular Russian verb production // Frontiers in Human Neuroscience. 2015. Vol. 9 (36). P. 193–220.



от такого взаимодействия, способности понять собеседника, что возможно только при сопоставимой культурной и интеллектуальной базе (работы авторов S. Kitayama<sup>1</sup>, Y.I. Alexandrov, M.E. Sams). Сознание не может развиваться «в темноте» не только в смысле концепций Чалмерса и Нагеля, но и в контексте Мамардашвили и Пятигорского, говоривших о *жутком труде мысли, происходящем на пределе человечески возможного, напряжения всех сил, в контексте культуры, социума*<sup>2</sup>. Лотман неоднократно писал о полисемантичности текста, о самовозрастающем логосе, ибо текст знает больше автора<sup>3</sup>. Смыслы рождаются в диалоге, от диалога культур до диалога разных типов мышления внутри самого мозга (об этом писали Ю.М. Лотман<sup>4</sup>, Т.В. Черниговская и В.Л. Деглин<sup>5</sup>, S. Kitayama, К.Р. Арутюнова и Ю.И. Александров<sup>6</sup>). Нестабильность смыслов в зависимости от меняющегося контекста, от отношения означаемого и означающего, имеющая социальное и личностное значение — очень острая тема в нашем обсуждении, особенно с оглядкой на искусственные интеллектуальные программы. Для понимания сложности смысла нужно иметь подготовку. Эту динамику смены

---

<sup>1</sup> Kitayama S. Culture and basic psychological processes- toward a system of culture: Comment on Oyserman et al. // Psychological Bulletin. 2002. Vol. 128 (1). P. 89–96.

<sup>2</sup> Мамардашвили М.К. Картезианские размышления // Философские чтения. СПб.: Азбука-классика, 2002; Пятигорский А.М. Непрерываемый разговор. СПб.: Азбука классика, 2004.

<sup>3</sup> Лотман Ю.М. Мозг — текст — культура — искусственный интеллект / Избр. статьи: В 3 т. Т. I: Статьи по семиотике и типологии культуры. Таллинн: Александра, 1992. С. 25–33.

<sup>4</sup> Лотман Ю.М. О семиосфере // Структура диалога как принцип работы семиотического механизма / Под ред. Ю.М. Лотмана и др. Тарту, 1984. (Труды по знаковым системам. Т. 17; Учен. зап. Тарт. гос. ун-та. Вып. 641). С. 5–23.

<sup>5</sup> Черниговская Т.В., Деглин В.Л. Проблема внутреннего диалогизма (нейрофизиологическое исследование языковой компетенции) // Структура диалога как принцип работы семиотического механизма / Под ред. Ю. М. Лотмана. Тарту, 1984. (Труды по знаковым системам. Т. 17; Учен. зап. Тарт. гос. ун-та. Вып. 641). С. 48–67.

<sup>6</sup> Арутюнова К.Р., Александров Ю.И. Мораль и субъективный опыт. М.: Институт психологии РАН, 2019.

приоритетов, а значит и интерпретации текста можно зафиксировать в научном эксперименте, регистрируя процесс «считывания текста» с разными когнитивными акцентами с помощью современных методик.

Естественные науки ищут ответы на главные вопросы о человеке «микроскопами и телескопами», но так нельзя найти смыслы, а именно они нам и нужны. Смыслы важнее структур. Археология цивилизации не менее важна, чем идентификация и измерение объектов, в том числе и в самом мозгу. Тем более, что стабильных объектов там нет вообще, а есть многомерные облака, одновременно вложенные друг в друга и сами в себя, согласно неведомой нам пока мозговой математике. Человеческий мозг — вершина эволюции — достиг невероятных высот, разработав языки математики, музыки, пластики, поэзии... и бесполезно их анатомизировать. Искусство показывает нам, каков наш мир. Мир людей. Но это не копии, а «ощупывание» мира, ментальные и эмоциональные усилия по его постижению.

Поразительно, что вне контекста собственно культурного дискурса явно недооценено опережение когнитивных, говоря современным языком, открытий, сделанных совсем ненаучными методами (см., например, работы авторов J. Lehrer<sup>1</sup>, Ю.Е. Шелепина<sup>2</sup>). Мозг показывает в искусстве то, что он как когнитивный инструмент умеет. Высвечивает лучом, как в живописи барокко, центры интереса, играет в «несуществующие» комбинации, как в сюрреализме, пробует описать мир геометрическими формами, отказывается от форм вообще или использует только формы — «врожденные идеи» и абстракции, сжимает все формы и цвета в черном квадрате... И это только в изобразительной части искусств. А игра с временем в музыке и поэзии и гравитацией и пространством в балете и архитектуре... Мозг, создавая произведения искусства, как бы пробует на прочность законы природы, в том числе и ещё не открытые. Стили в искусстве открывают внимательному наблюдателю реестр работы со смыслами.

---

<sup>1</sup> Lehrer J. Proust Was a Neuroscientist, NY: HMH, 2008.

<sup>2</sup> Шелепин Ю.Е. Введение в нейроэкономику. СПб.: Троицкий мост, 2017.

*Правильно ли мы идём, пытаясь понять, что такое человек, насколько биология определяет культуру и vice versa, насколько современные всё усложняющиеся технологии определяют успех на этом пути? Всё яснее просвечивает значимость рисунка, звука, пластики, а не только формулы... Всё яснее видно и то, что гениальный Гия Канчели называл «сложная простота». Дороги искусства и науки — внезапно стали сходиться. Без понимания картины мира интересующего нас в некий момент биологического существа, не говоря о потенциально возможных искусственных интеллектуальных субъектах — коммуникация и любое взаимодействие является фикцией. Привычный комментарий, что у ИИ будет только то, чему мы его обучим — несостоятелен: у эволюции свои законы, и сложные системы любого генеза могут развиваться сами по себе, с мало предсказуемым результатом.*

*(Исследование настоящей главы поддержано грантом РНФ № 21-18-00429 «Когнитивные механизмы обработки мультимодальной информации: тип текста и тип реципиента».*

*The study was funded by RFBR № 21-18-00429).*

**РАСШИРЕННАЯ ЛИЧНОСТЬ  
КАК ОСНОВНОЙ СУБЪЕКТ И ПРЕДМЕТ  
ФИЛОСОФСКОГО АНАЛИЗА.  
СЛЕДСТВИЯ ДЛЯ ОБРАЗОВАНИЯ**

---

---

**Информационные революции  
в истории человека**

Скорость цивилизационных изменений растет: на протяжении даже не жизни нынешнего поколения, как обычно говорят в таких случаях, а нашей собственной жизни. В течение последнего десятилетия в окружающем нас мире произошли разнообразные фундаментальные изменения. И это — не просто появление тех или иных информационных технологий, с которыми

---

\* *Семенов Алексей Львович* — Московский государственный университет им. М.В. Ломоносова, заведующий кафедрой математической логики и теории алгоритмов, Россия, Москва, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, директор Института кибернетики и образовательной информатики им. А.И. Берга, Россия, Москва, НИУ «ВШЭ», профессор, Россия, Москва, академик РАН, академик РАО.

*Alexei Semenov* — Lomonosov Moscow State University, Head of the department of mathematical logic and theory of algorithms, Moscow, Russia FRC CSC RAS, Director of Axel Berg Institute of Cybernetics and Educational Computing, Moscow, Russia. HSE, professor, Moscow, Russia.

*Зискин Константин Евгеньевич* — Московский государственный университет им. М.В. Ломоносова, заведующий отделом Центра стратегии развития образования, Россия, Москва. Автономная некоммерческая организация дополнительного профессионального образования «Высшая школа компетенций», генеральный директор, Россия, Москва. Кандидат педагогических наук.

*Konstantin Ziskin* — Lomonosov Moscow State University, Center for Education Development Strategy, Head of the department, Moscow, Russia. Nongovernmental organization for further education “Higher school of competences”, CEO, Moscow, Russia.

мы сталкиваемся буквально каждый день. Вот очевидная констатация, вероятно, кто-то ее высказывал раньше, но мы, к сожалению, не нашли первоисточника:

*Никогда еще изменения не были такими быстрыми и уже не будут такими медленными, как сегодня — в XXI веке.*

Но не скорость, а радикальность изменений заставляет нас рассмотреть весьма длительную историческую перспективу, начиная с самого появления *Homo sapiens*. Первая после этого появления информационная революция в истории человечества — это появление речи. По-видимому, человек, при некотором определении того, что такое *Homo sapiens*, существовал ещё до речи. По-видимому, до возникновения устной речи также существовали способы общения людей (жесты, прикосновения, запахи, нечленораздельные звуки) и шла мыслительная деятельность человека, и с этого, по-видимому, и стоит отсчитывать события. Технология речевого общения, в сочетании с видимым действием, составила первую информационную революцию, о которой говорилось выше. Изобразительные технологии, включая рисунок и письмо, связанные с фиксацией движения человека, сопровождающегося передачей информации, — это следующая технология. И наконец, технология искусственного интеллекта, к которому мы относим и автоматизацию вычислений, и написание текста, и проектирование самих технологий, представляется следующей существенной революцией.

Но вот 50–100 тысяч лет назад произошли радикальные изменения, связанные с тем, что возникла эффективная технология передачи информации, а именно — речь. Несколько тысяч лет назад, может быть 4–6 тысяч лет, возникло письмо. А в последние 100 лет возникло то, что можно назвать искусственным интеллектом (ИИ). Практически экспоненциальный рост! Сущность понятия «искусственный интеллект» — важная философская проблема, но мы здесь ее рассматривать не будем. Отметим лишь, что в том понимании, в котором этот термин будет использоваться в статье, искусственный интеллект будет пониматься широко. Это все способы автоматизации, механизации интеллектуальной деятельности человека. Есть ручной труд, есть

машинное, индустриальное производство, а есть интеллектуальная деятельность. В какой-то момент человек начал использовать те или иные инструменты для этой деятельности, и вот об этих цифровых, электронных инструментах здесь идет речь. В связи с этим можно упомянуть и об алгоритмическом искусственном интеллекте, попытке прямого моделирования формализованной интеллектуально-логической деятельности человека, и о том, что называется «машинное обучение» — попытке машинного моделирования человеческой интуиции.

Следует особо отметить, что ИИ сегодня охватывает и коммуникацию. Предшествующие революции начинались именно с коммуникации, а потом захватывали все сферы жизни человека. Можно считать, что и предтечами идущей сейчас революции были телеграф и телефон. Сегодня разные формы коммуникации во многом слились благодаря достижениям в сфере ИИ. Возможно, в силу тех же экспоненциальных тенденций, мы присутствуем сегодня при начале следующей информационной революции, основанной на прямом интерфейсе между человеком и машиной, между мыслительными процессами, сопровождающимися электронными, электрическими импульсами в мозге человека и внешним компьютерным миром.

## **Результаты информационных революций. Контрреволюция**

В начале 30-х годов XX века Лев Семенович Выготский, на своей лекции в Комакадемии<sup>1</sup>, выдвинул фундаментальное положение, что информационные технологии меняют то, как человек думает, общается и действует. (Конечно, Выготский использовал

---

<sup>1</sup> *Выготский Л.С.* Инструментальный метод в психологии // Собр. соч.: В 6 т. Т. 1. 1982. [http://elibrary.gnpbu.ru/text/vygotsky\\_ss-v-6tt\\_t1\\_1982/go,108;fs,1/](http://elibrary.gnpbu.ru/text/vygotsky_ss-v-6tt_t1_1982/go,108;fs,1/); *Vygotsky L.S.* The instrumental method in psychology (1981), <https://www.marxists.org/archive/vygotsky/works/1930/instrumental.htm>; *Vygotsky L.S.* Mind in society: The development of higher psychological processes. Harvard University Press (1980).

терминологию своего времени, но его примеры и пояснения оправдывают наш «перевод».) Эти изменения прослеживаются во всех упомянутых информационных революциях. При этом информационные (как и другие) технологии не только что-то добавляют, но и что-то «отнимают» у человека. Как говорил Выготский, очередная информационная технология «отменяет и делает ненужным целый ряд естественных процессов, работу которых выполняет то или иное интеллектуальное орудие». С этим положением связано, в частности, и понятие, которое можно назвать «контрреволюцией». Например, приблизительно в VII–IV веках до нашей эры в античной Греции завершилась революция перехода от устной цивилизации, от устного общения к письменному. В связи с этим Платон цитирует<sup>1</sup> возражение своего учителя — Сократа против письменной речи, как того способа коммуникации, который убивает мудрость, общение и учение. Не исключено, что Сократ не пользовался письменной речью, по крайней мере, свидетельств этого не сохранилось, разумеется, это более очевидно для Гомера. Но ведь и официальные данные о грамотности населения нашей страны в XIX веке показывают неграмотность большинства, следует ли отказать этому большинству в мышлении, общении и учении, или напротив — считать, что только оно и было способно к этим видам деятельности?

Л.С. Выготский указывал, что примерами «психологических орудий», то есть, в нашем смысле, технологий обработки информации, являются «язык, различные формы нумерации, исчисления, мнемотехнические приспособления, алгебраическая символика, произведения искусства, письмо, схемы, диаграммы, карты, чертежи, всевозможные условные знаки» и так далее. Выготский утверждал, что «включение этих орудий в процесс поведения, во-первых, вызывает деятельность целого ряда новых функций, связанных с использованием данного орудия, с управлением им. Во-вторых, отменяет и делает ненужным целый ряд естественных процессов, работы которых выполняет орудие. В-третьих, видо-

---

<sup>1</sup> Платон. Сочинения в 4 томах. М.: Мысль 3 (1971): 455–542. <http://psylib.org.ua/books/plato01/27timei.htm>

изменяет протекание отдельных моментов, входящих в состав инструментального акта психических процессов, замещают одни функции другими, то есть пересоздает, перестраивает всю систему, структуру поведения совершенно так же, как техническое оружие пересоздает весь строй трудовых операций». Именно с этим связаны и те трудности, и проблемы, которые возникают, с которыми сталкивается все человечество и отдельный человек, переходя к использованию новых интеллектуальных орудий. И с этим связаны и те явления, которые в данной работе мы характеризуем как «контрреволюцию» — попытку сохранить то, что новый инструмент «делает ненужным» или вообще «отменяет».

### **Расширения сознания. Расширенная личность человека**

Интересно сравнить положения Выготского 1930 года с высказываниями 2003 года британского философа Энди Кларка<sup>1</sup>: «Расширяющая сознание технология приходит в удивительном разнообразии форм. Они включают в себя лучшие из наших старых технологий — ручку, бумагу, карманные часы, блокнот художника, логарифмическую линейку, а также смартфоны, связывающие пользователя со все более отзывчивой всемирной паутиной, а скоро и с повседневными объектами, которые населяют наши дома и офисы. И дальше: «Некоторых деятелей от образования это пугает, а для меня же это естественное торжество способности, делающее нас созданиями столь мощными в решении проблем. Наш мозг более чем у любого другого существа на планете способен искать и использовать интимные отношения с не биологическими ресурсами, которые, в конечном счете, обеспечивают наши способности к яркому абстрактному мышлению. Мы прирожденные киборги, всегда готовы слиться в постижении мира нашу умственную деятельность операциям пера, бумаги, электроники».

---

<sup>1</sup> Clark A. Being there: Putting brain, body, and world together again. MIT press, 1998. <http://www2.econ.iastate.edu/tesfatsi/BeingThere.AClark1998.EntireBook.pdf>



Это — представление о «расширенном сознании», представление о том, что современное сознание — это не сознание, находящееся только под черепной коробкой. Наглядный пример того, что можно называть расширенным сознанием, приводит Мишель Серр, в частности, в своей книжке 2012 г., переведенной на русский язык, “Petite Poucette” — «Девочка с пальчик»<sup>1</sup> (как некоторая антитеза «Мальчику-с-пальчик»). Там он говорит о судьбе святого Дионисия Ареопагита, когда обезглавленный Дионисий встал с колен, подобрал голову и, держа ее в руках, продолжил карабкаться вверх по склону Монмартра и проповедовать. Эта легенда породила даже целую плеяду святых кефалофоров (носителей собственных голов). На иконе<sup>2</sup> изображен Иисус, который общается с несколькими кефалофорами, у которой сознание располагалось вне тела. И Серр продолжает:

«Девочка с пальчик открывает свой компьютер... Там в отрыве от себя она держит свою доселе бывшую внутренней познавательную способность подобно тому, как святой Дионисий нес в руках свою голову... С недавних пор мы все превратились в святых Дионисиев — мозг отделился от головы с ее костями и нервными клетками. Компьютер, ящик у нас в руках, вмещает в себя и приводит в действие то, что мы называем «нашими способностями», память, причем тысячекратно более мощную, чем наше воображение, наштапкованное миллионами образов, и даже разум, ведь компьютерные программы могут решать сотни задач, с которыми мы в одиночку не справились. Наша голова лежит перед нами в овеществленном ящике мысли».

Все это не абстрактные философские построения, они, как это будет показано ниже, непосредственно связаны с проблематикой образования и буквально с сиюминутной работой современной

---

<sup>1</sup> Серр Мишель. Девочка с пальчик. М.: Ад Маргинем Прессе, 2016. (Оригинал: Serres M. Petite Poucette. Éditions Le Pommier, Paris, 2012.)

<sup>2</sup> Святые защитники Цюриха Феликс, Регула и Экзюперантий. Их обезглавили в III в. в ходе гонений на христиан. После казни все трое чудесным образом поднялись, взяли свои головы и помолились Господу на близлежащей горе. [https://farm1.staticflickr.com/272/20180532756\\_ec91196301\\_o.jpg](https://farm1.staticflickr.com/272/20180532756_ec91196301_o.jpg)

школы Российской Федерации. Вернемся к эпистемологической перспективе. Мишель Серр говорит: «Что по отсечении головы остается у нас на плечах? Интуиция, изобретательная и неуёмная: знания загружены в ящик, но охота к изобретению по-прежнему с нами. И любопытство: неужели мы обречены быть умными? Когда появилось книгопечатание, Монтень... предпочел ясную голову накопленным знаниям, поскольку их запас, уже объективированный, стоял у него на полках в виде книг. До Гутенберга историкам требовалось знать наизусть сочинения Фукидида и Тацита. Те, кто интересовался физикой, заучивал труды Аристотеля и греческих механиков, а те, кто совершенствовался в ораторском искусстве, — сочинения Демосфена и Квинтилиана. Иначе говоря, все они пичкали голову. Способ сэкономить — на запоминание места книги на полке уходит меньше ресурсов памяти, чем на хранение всего ее содержания. Новый способ сэкономить более радикальный — нет надобности помнить даже место книги, этим занят поисковик. Отрубленная голова Девочки-с-палец уже не та, что прежние головы Монтеня, для которых быть ясными лучше, чем напичканными наукой». Серр здесь ссылается на Монтеня<sup>1</sup>. На самом деле, исчезает и сама дихотомия между ясностью и многознанием.

Это действительно совершенно актуализованная эпистемология. И замечательно, что отсылки наших предшественников ведут не к сегодняшнему дню. Речь идет о Гутенберге, где-то упоминаются карманные часы или вообще часы, сопровождающие нашу жизнь. Насколько точно мы умеем определять время сегодня без часов? Возникает вопрос: а почему, собственно, без часов? Часы-то есть у нас. Но точно так же возникает вопрос: насколько хорошо мы помним без компьютера? И сегодня можно спросить: а почему, собственно, без компьютера? Зададим другой странный вопрос: насколько хорошо мы что-то помним без сотового телефона? Кто знает наизусть 10 телефонных номеров? А актуальных, действующих сегодня, кто знает? Сегодня трудно найти человека, который помнит 10 телефонных номеров, действующих

---

<sup>1</sup> Монтень М. Опыты. М.-Л., Изд-во АН СССР, 1960, кн. I. Гл. XXVI «О воспитании детей». С. 189–212. (Оригинал: Michel de Montaigne, Les Essais.)

сегодня. Но еще интереснее, кто 20 или 30 лет назад каждую неделю обращался к Большой Советской Энциклопедии или толковому словарю? — Немногие. Кто сегодня, по крайней мере, раз в неделю смотрит какое-нибудь слово, понятие, статью в интернете, Википедии и так далее? — Существенно большее количество людей.

Эти изменения произошли, сознание уже расширилось. Но повторим: наличие карманных часов — это уже расширенное сознание, помимо субъективного чувства времени, как и календарь на стене, который объясняет, какой, собственно, сейчас сезон — вообще, урожай уже пора собирать или нет (в дополнение к наблюдению и интуиции). На это всё обращает наше внимание, в том числе, и Мишель Серр.

Сейчас, как уже было сказано, мы вступаем в завершающий этап не четвертой индустриальной, а четвертой именно информационной (считая от возникновения сознания) революции, относящийся к расширенному сознанию. И в этом смысле смартфон, только что упоминавшийся Серром, именно не ноутбук, а тот самый смартфон, за которым мы постоянно лезем в карман, когда хотим или с кем-то пообщаться, или выяснить какой-нибудь непонятный для нас термин — вот важный элемент этого современного расширения сознания. Голосовой ввод текста, который становится реальностью на наших глазах, квалифицированный перевод с одного языка на другой... 10 лет назад странно было видеть профессионального переводчика, который использует машинный перевод, хотя сама по себе идея машинного перевода принадлежит, конечно, 50-м годам прошлого века. Но сегодня большинство профессионалов используют не ручку с бумагой, а клавиатуру, как и автоматизированный перевод с одного языка на другой, а потом уже его правят, редактируют, уточняют, в том числе интерактивно и так далее.

Кохлеарный имплантат — устройство, воздействующее непосредственно на слуховой нерв. Сегодня более миллиона жителей Земли получают речевую информацию без того, чтобы воспринимать звук непосредственно по обычному каналу. Под кожу вживляется приемник электромагнитных сигналов, выво-

дящий их на слуховой нерв. Получает приемник сигналы от передатчика, похожего на обычный слуховой аппарат. Передатчик использует нетривиальные методы обработки речевого сигнала. Это — добавление внешних объектов, расширение внутренних механизмов и вытеснение старых, о котором говорит Выготский. Процесс расширения использования имплантатов сопровождается индивидуальным и коллективным протестом. И это — тоже четвертая информационная революция.

Очень вероятно, что глухой ребенок, в раннем детстве получивший кохлеарный имплантат (и именно в этом возрасте наиболее эффективно начать его использовать), не освоит языка жестов (языка глухих), или, по крайней мере, не будет бегло говорить на нем. Аналогично этому слепые теряют свою традиционную грамотность за счет современных технологий. Большинству слепых среднего возраста уже не нужно знать Брайля, большинству глухих с кохлеарным имплантатом не нужно владеть языком жестов. И мы слышим протесты против технологии от видных деятелей этих человеческих сообществ.

Рассмотрим контрреволюционные ответы на революционные изменения. Например, «мысль изреченная есть ложь» есть реакция на то, что без речи все было честнее — ты мог посмотреть человеку в глаза или посмотреть, как он двигает рукой, или пинает тебя ногой — и это чистая правда. Любой разговор, вообще говоря, затемняет существо дела, там и наврать можно и так далее. Далее, следуя Сократу в изложении Платона, мы можем спросить: а вот автор романа — он его знает или нет? Если не знает, то это какое-то жульничество. Вот Гомер свои поэмы знал — спроси его и он всю Илиаду, как считается, от первой до последней строчки произнесет близко к тексту. А уже Тургенев или Достоевский своих романов наизусть не знали. Тогда это жульничество в понимании Сократа: что-то написал на бумаге, а сам за это не отвечает. И математик часто не держит в уме целиком вычисления или полное доказательство теоремы. Они записаны и уже находятся в его расширенном сознании. Это видно в ситуациях с материальными технологиями на примере движений разрушителей машин — луддитов.

Видный представитель современной интеллектуальной и культурной, но прежде всего религиозной, жизни владыка Тихон — Тихон Шевкунов, несколько лет назад поддержал создание и обсуждение небольшого телевизионного фильма<sup>1</sup>, в котором молодым людям задавались вопросы, относящиеся к российской культуре и истории. Например: «Кто написал роман «Бесы»?» Можно представить себе, что происходило. Происходило примерно то же, что и в 70-е годы. Когда ты задаешь молодому человеку вопрос про роман «Бесы», это вне его контекста, он знает, так сказать, что происходит вокруг него, и всё. Ответ был такой: «Некрасов, Пушкин, можно мы погуглим...» У девушки спросили, чем известен Михаил Ломоносов. «Михаил Ломоносов — ну как же, Ломоносов, конечно, все знают, но сейчас я погуглю...» Интернет, как говорится в таких случаях, взорвался миллионами откликов — насколько же малограмотна наша молодежь!

Это наглядная демонстрация современного расширенного сознания, то есть для владыки Тихона и большинства населения нашей страны это доказательство тех ужасов, которые принес нашу жизнь Интернет. Хотя в 70-е годы молодые люди были бы примерно такими же. Но посмотрим на ситуацию объективно: в каждый момент, когда девушку тянуло загуглить, оператор не дожидаясь, чтобы она ответила. Это очень интересная история, фильм можно легко найти в Интернете. Это демонстрация расширенного сознания, как оно выглядит сегодня — со всеми плюсами, минусами и соответствующей контрреволюцией.

### **Устная речь в контексте Четвертой революции**

Следующий вопрос — это устная речь в той самой перспективе информационной революции, революции мышления, революции коммуникации. Выше говорилось о том, что в какой-то

---

<sup>1</sup> <https://rg.ru/2017/03/21/nuzhna-li-klassika-pokoleniiu-ru.html> <https://vimeo.com/208770296>

момент произошел сдвиг — устная культура начала вытесняться письменной. Это то самое революционное изменение, которое происходило во времена Платона и Сократа. Вальтер Онг<sup>1</sup> рассмотрел эффект, связанный с современными технологиями и с возвращением устности. Он замечает, что без письма грамотный ум не будет и не может мыслить так, как он мыслит не только когда он занят письмом, но когда он формулирует свои мысли в устной форме. Больше, чем любое другое изобретение, письменность преобразила человеческое сознание. Но мы уже слышали это от Выготского. (Онг на него здесь не ссылается.) Одновременно Онг пишет и о том, что сегодня возникает «вторая устность». Заметим, попутно, что устность — официальный термин. Это один из принципов судопроизводства Российской Федерации, где устная коммуникация является первичной: устное участие сторон, присутствие на процессе и устное изложение аргументов. Судья принимает решение на основании дискуссии, происходящей устно. Новая устность, которая возникает сегодня, становится все более и более важным форматом общения. Величайшие речи XX века — это речи, воспринимающиеся устно. “I have a dream” — это устное заявление, это не текст речи Мартина Лютера Кинга, как та или иная речь Ленина, Сталина, Трумэна и так далее. Рассмотрение подобных эффектов было, в частности, предметом деятельности Маршалла Маклюэна, но и в более общем контексте, когда все это синтезируется и объединяется, становится новым форматом взаимодействия, включая даже и до-речевые, до-устные форматы взаимодействия. В частности, замечательно название книги Маклюэна и Фиоре<sup>2</sup> “The Medium is the Massage” (трансформация раннего термина Маклюэна), то есть не “message”, не сообщение, а именно «массаж» — как возвращение к тактильности.

Обратим внимание на то, что Мераб Мамардашвили считал, что его “message” — это то, как нечто происходит здесь и сейчас

---

<sup>1</sup> Walter J. Ong. *Orality and Literacy. The Technologizing of the World*. Methuen & Co., 1982.

<sup>2</sup> McLuhan, Marshall and Fiore, Quentin. *The Medium is the Massage*. New York: Bantam Books, 1967.

в устном формате. И он носил всегда собой диктофон Grundig<sup>1</sup>, записывая то, что происходит. Устное сообщение — это дело, которое он делает, а не текст, который существует в предыдущей письменной цивилизации. То есть Мамардашвили — это пример возвращения к устной цивилизации, отказ от письменности, от письменного.

Жак Деррида<sup>2</sup> так это выразил: «Эпоха господства линейного письма и соответствующей модели мысли [sic! — АС и КЗ] заканчивается повсюду — в литературе, философии, науке, а это, в свою очередь, предполагает новую организацию пространства и времени».

Итак, письмо сегодня оказалось иным. Оно оказалось слитым с устным сообщением, оно оказалось многоязычным. Грамматика выносится за пределы человеческого мозга, устная речь немедленно трансформируется в письменную или в речь на другом языке. Итак, с точки зрения речи мы сейчас находимся в состоянии некоторого нового формата, связанного тоже с расширенным сознанием. Происходит неминуемое массовое падение орфографической и даже письменно-функциональной грамотности, как это было с устной грамотностью в эпоху Сократа — Платона. Есть устройство, которое мгновенно читает вслух любые письменные тексты — просто подносят к страничке, к любому месту страницы. Другое устройство может выделять голос одного говорящего в шуме, одновременно узнавать лица и голоса и вспоминать подробности о человеке и деле. Часть твоего расширенного сознания — маленькая коробочка, могут и в телефоне те же функции находиться, которая сопровождает тебя: «С кем это я разговариваю сейчас?», ага, ну понятно. Вот опять-таки, сколько лиц мы помним сегодня, сколько лиц мы будем помнить через 10 лет, обладая всеми теми же устройствами.

---

<sup>1</sup> <https://story.ru/istorii-znamenitostej/legenda/govoryashchiy-filosof/>

<sup>2</sup> *J. Derrida. Le problème de la genèse dans la philosophie de Husserl. Paris: Presses Universitaires de France, 1990.* The parallelism is made particularly relevant by Foucault's criticism of Husserl.

Впервые в истории становится слышен всеобщий голос — это еще одна из идей Серра о том, что, на самом деле, теперь все слышны, видимы и так далее. «Не возвещают ли эти детские лепеты, эти немолчные гомоны мира наступление новой эры — второй эры устной речи и вместе с тем небывалой эры виртуального письма? И не поглотят ли ее волны эру страницы, сформировавшей нас? Мне поступь новой вокальной эры, растущей из виртуального, слышится уже давно — как всеобщее требование слова, подобное тому, более частному, что исходит от «девочек-с-пальчик» в школах и университетах, от пациентов в больницах, от трудящихся на предприятиях. Все хотят говорить, все общаются друг с другом в бесчисленных сетях». Вместо одностороннего общения от начальства к подчиненному, от врача к больному, от учителя к ученику Серр видит многоголосый шум и гам. Теперь мы переходим к тому, что в данном контексте означают перспективы образовательного пространства.

### **Цифровая компетентность (Computational Thinking)**

Имеется широкий спектр представлений о том, что такое Computational Thinking. Нам представляется разумным развитие линий Андрея Петровича Ершова<sup>1</sup> и Симора Паперта<sup>2</sup>. Это означает:

- рефлексия — осознание себя как расширенной личности;
- умение и желание использовать возможности и средства цифрового мира;

---

<sup>1</sup> *Ершов А.П.* Программирование — вторая грамотность. Русская версия доклада на 3-й Всемирной конференции ИФИП и ЮНЕСКО по применению ЭВМ в обучении в Лозанне, Швейцария, 1981. [http://ershov.iis.nsk.su/ru/second\\_literacy/article](http://ershov.iis.nsk.su/ru/second_literacy/article) // Проблемы информатики. 2015. № 4. С. 71–85.

<sup>2</sup> *Papert S.* Mindstorms: Children, Computers and Powerful Ideas. New York, USA: Basic Books Inc. Publishers, 1980. 252 p.



- способность к цифровому моделированию всего мира и цифровому взаимодействию с ним.

Вот что пишет Серр: «Объективность, коллектив, технология, организация и т. д. регулируются в наши дни не столько декларативными абстракциями, которые два с лишним тысячелетия вынашивались науками и словесностью и утверждались философией, сколько алгоритмической или процедурной когнитивностью». Это слишком сжато сформулировано, но смысл в том, что это вычислительное мышление занимает место многих других видов знания и отношения к миру.

## **Цифровая трансформация образования**

Представление о роли цифровой трансформации в развитии школы отражены в «Хартии цифрового пути школы»<sup>1</sup>.

Укажем основные факторы, входящие в наше представление о цифровой трансформации школы:

- передача технических умений и навыков от биологической к цифровой части расширенной личности; высвобождение времени и энергии — мотивации нецифровой части личности учащегося для творческой и исследовательской деятельности;
- фиксация в цифровой образовательной среде (на цифровой платформе) всей учебной работы учащегося и его взаимодействия с другими, возможность эффективной обратной связи, оценивания и прогнозирования работы ученика; возможность аттестации учителя исходя из обеспеченной им реализации образовательного потенциала учащихся;
- персонализированный выбор целей, проектирование путей к ним и обязательное их достижение.

В частности, важнейшим требованием реальной цифровой трансформации должно быть то, чтобы школа и в своей деятельности, и в оценке результатов этой деятельности апеллировала к

---

<sup>1</sup> <https://rffi.1sept.ru/document/charter>

представлению об учащемся человеке с расширенным сознанием. То есть спрашивать у ребенка, который сейчас час, запрещая посмотреть на часы, или просить его написать сложный текст или перемножить два числа именно ручкой на бумаге, выглядит неоправданным, неэффективным, а возможно, даже преступным, ограничением. И надо спросить себя, имеем ли мы право на это ограничение? Чем мы оправдываем то, что, становясь людьми с расширенным сознанием, ребенку в этом праве быть таким человеком отказываем?

При этом именно цифровая трансформация образования дает возможность ребенку вернуться к естественно присущему человеку способу учения. Человек, который рождается, обладает способностью к постижению окружающего мира, к обучению, он обладает любопытством, склонностью к коммуникации и творчеству. И современная школа, как отмечают многие, занимается тем, что последовательно вытесняет, если не сказать, убивает эту склонность к удивлению, к творчеству и общению. Когда ученик пытается спросить что-то, что ему действительно интересно, на уроке, как правило, учитель говорит, что это не относится к теме занятия, или одергивает ребенка: «Сядь!» или «Выйди из класса!» Если мы считаем, что очень важным является общение с другими, то на уроке мы слышим «Не подсказывай!», «Не шумите!» и т. д.

Ребенок приходит в школу из детского сада уже с мобильником в кармане, который дала ему мама. Он выкладывает себе в Инстаграм то, что ему интересно, и то, что он хочет об этом рассказать, показать другим, и спрашивает нас: «Зачем писать, если можно сказать или показать? Зачем учить правила, если машина их знает и учитывает?» То есть мы оказались в более тяжелой ситуации, чем были двадцать, тридцать, пятьдесят, двести лет назад, когда было понятно, зачем все это нужно, сегодня это стало непонятно, и это часть общего вопроса: «А зачем вообще я пришел в школу?» Вот сегодня этот вопрос становится все более и более актуальным по самым разным причинам, не только тем, которые обозначены в данной работе, но и другим.

Цифровая трансформация образования одновременно:

- адекватна потребностям общества и личности XXI века;
- позволяет сделать образование органичным, впервые

в истории человечества возвращаясь к исконному состоянию человеческого бытия, прежде всего бытия ребенка.

## **Организация образовательного процесса**

Обсуждение того, как, используя современные цифровые инструменты, сделать работу школы в целом более эффективной, выходит за пределы настоящей работы, но все же приведем простой пример из школьной жизни. В современной школе довольно важную роль играют домашние задания. Что происходит с домашним заданием сейчас? Учительница на очередном уроке задает на дом написать какое-нибудь сочинение, эссе или выполнить задачу и так далее. На следующий урок дети приносят тетради, эти тетради они сдают учительнице, учительница забирает их домой в сумке — это пара килограммов, дома проверяет и приносит на следующий урок. Картинка естественная и очевидная, но обратим внимание — она все-таки довольно сильно парадоксальная — дети что-то принесли на урок, они выполнили это дома, потом принесли, никто с ними не пытается это обсуждать. Все про это забывают, и им раздают тетради еще на следующем уроке, когда им уже совершенно это неинтересно, помимо других вещей, которые не интересны детям и учителю. А наличие цифровой среды меняет этот картинку таким образом: дети накануне что-то сделали, послали это учителю посредством электронной цифровой коммуникации, учительница дома это проверила, пришла на следующий урок и дальше обсуждает именно это с детьми, говоря, что вот тут, когда я давала вам задание, я забыла сказать некоторую важную вещь. Из-за этого получилось что-то не то у некоторых из вас, давайте это сейчас обсудим. А у одного из вас, у Васи, возникла какая-то идея, вообще говоря, она неправильная, но она замеча-

тельная, давайте про эту идею тоже поговорим и так далее. Это возвращает ситуацию с выполнением задания от учителя, данного детям, к некоторой естественности — задание дано, оно выполнено, оно обсуждено, переходя от существующей сегодня в нынешней ситуации, когда все происходит некоторым искаженным образом.

Если нам удастся разрешить использование цифровых ресурсов детям, которыми пользуются в жизни взрослые, то дальше возникает действительно использование цифровой платформы, на которой все это можно интегрировать. Можно заниматься персонализированным планированием и сбором данных о каждом из учащихся, и перейти к образованию, ориентированному на результаты, а не на затраты ресурсов ученика и учителя. И дальше можно, используя эти данные, уже прогнозировать дальнейшую деятельность и построить систему адаптивного образования, но подчеркнем, что первым этапом является все-таки то, чтобы сама деятельность ученика учитывала бы цифровой мир, то есть использовала бы все инструменты, которые у нас уже есть сегодня.

Возможно, что-то на описанном пути уже происходит в массовом масштабе, скажем, в некоторых экспериментах, идущих в Китайской Народной Республике. Например, Squirrel AI Learning<sup>1</sup>, компания-единорог, собравшая более миллиарда долларов при своём старте, где создаются индивидуализированные пути обучения для каждого ученика. Но и в нашей стране тоже уже приняты решения по тому, чтобы платформы персонализированного обучения реализовывались в различных регионах страны. Есть поручение Президента страны В.В. Путина Правительству Российской Федерации и Сбербанку заниматься этими изменениями. Создание платформы, в которой отражается все то, что происходит с каждым учащимся.

---

<sup>1</sup> Squirrel AI Learning. 1st AI-powered adaptive education provider in China. <http://squirrelai.com/>

## Результаты образования. Универсальные навыки

Современная цивилизация, общество, экономика, политика повышает значимость универсальных навыков (soft skills), как в абсолютном смысле, так и в сравнении с рутинными, техническими, специализированными навыками (hard skills). Последние все чаще передаются роботу, цифровой технологии. Универсальные навыки, например способность к коллективному действию, к лидерству, управление собственной энергией и временем, социально-эмоциональный интеллект и т. д. (в данной работе нет возможности и особого смысла вдаваться в детали), существовали всегда, позитивно оценивались в XX веке. При этом, как уже было сказано, массовая традиционная школа, в том числе — советская, часто их подавляла, подменяла другими, имитировала, искажая содержание.

Цифровая трансформация позволяет эти качества личности формировать в рамках школы.

*Преадаптивность.* Итак: никогда еще изменения не были такими быстрыми и не будут такими медленными, как сегодня. В связи с этим видный деятель российского образования Александр Григорьевич Асмолов<sup>1</sup> говорит о таком явлении, как пре-адаптивность — это готовность человека сегодня к НЕожиданным, НЕпредсказуемым изменениям, которое становится важнейшим качеством личности в современном мире. При этом заметим, что некоторая предсказуемость, во многом иллюзорная, уютного, «лампового» XX века — похоже результат научно-технической революции. В прошлые века проблема предсказуемости решалась, в основном, при помощи религиозных или оккультных практик, как мы теперь знаем, не особенно эффективно. Таким образом, современная актуальность готовности

---

<sup>1</sup> *Asmolov A., Guseltseva M.* Education as a space of opportunities: from human capital to human potential // European Proceedings of Social and Behavioural Sciences, издательство Future Academy (online). Т. 64. № 6. P. 40–45; *Asmolov A.G.* Race for the Future: «...Now Here Comes What's Next» // Russian Education and Society, издательство M. E. Sharpe Inc. (United States). Т. 60. № 5. P. 381–391.

к НЕожиданному, напоминает «диктофон Мамардашвили», как возврат к Сократовской «изустности».

И если говорить об образовании, то надо сказать, что сегодня современная школа как общественное явление, отношение к школе, к образованию всего общества является, пожалуй, наиболее явным примером описанной выше контрреволюции. Разумеется, мы не пытаемся использовать этот термин в оценочном смысле и уж, тем паче, не политизируем ситуацию. Современная школа не нейтральна к цифровым технологиям, она им очевидно враждебна. Но хотелось бы все-таки готовить детей к тому, что на самом деле нужно, и перестать готовить их к несуществующим уже сегодня видам деятельности в будущем. Хотелось бы сократить цифровую пропасть между образованием и миром. В данном случае «хотелось бы» — это, к сожалению, не мнение большинства. Но стоит, по крайней мере, обсудить это положение. Сегодня цивилизация стала цифровой. Школа все еще находится во вчерашней — не цифровой цивилизации, а дети во многом оказываются «в завтра» — в набегавшем сингулярном будущем. Школа тем не менее пытается во многом противостоять окружающим изменениям. При этом цифровая трансформация, безусловно, возможна, и в том числе с использованием всех тех инструментов искусственного интеллекта, о которых говорилось выше.

*Рефлексия.* Частью нашего представления о расширенной личности является ее способность к рефлексии именно как расширенной личности, понимание своих возможностей и ограничений в цифровом мире. Инструментами рефлексии и развития могут быть: внутренняя и внешняя цифровая фиксация того, что с человеком происходит; цифровое проектирование и планирование будущего, напоминание о планах, компенсация прокрастинации, отчет о выполнении, автоматизированная коррекция.

*Цифровая грамотность.* Развивая положение о цифровой компетентности расширенной личности в современном мире, мы приходим к необходимости:

- осваивать базовые цифровые технологии, которые на самом деле повсеместно используются, то есть являются элементом

общей грамотности человека, скажем, калькулятор, или редактор текста;

- осваивать цифровое моделирование окружающего мира;
- формировать представление о том, как цифровой мир «работает», наряду с представлениями о том, как работает человеческий организм, сознание, природа и технология: почему движется автомобиль, или почему встает солнце, или почему мне надо время от времени пить воду, или как работает искусственный интеллект.

*Ориентация в мире и большие идеи.* Традиционно образовательные цели сводятся к тому, чтобы школьники знали на память различные тексты, алгоритмы действия и имели бы в своем распоряжении некоторые технические навыки. В частности, умели определять, что это за явление, что это за объект и так далее. В сегодняшней ситуации цифрового мира важнее становятся совсем другие элементы образования — это ориентация в мире. Не знание конкретно на память того или другого, а то, как это можно найти, как узнать, что это, как это называется, как это работает. Не знание готового алгоритма на память или какого-то объекта, а возможность это узнать, возможность это найти. Для того чтобы этой возможностью воспользоваться, нужно иметь общую ориентацию в мире, понимание следующего уровня. Нужно иметь общую методологию деятельности, то есть то, как нужно подойти к некой проблеме, разобраться, понять, изобрести, действовать и так далее. Нужно обладать пре-адаптивностью, которая строится на том, что ребенку дают все время новые задачи — непонятные, неизвестные. Нужно уметь разобраться, как решать квадратное уравнение, а не отрабатывать умение сто раз решать одно и то же уравнение с другими числами.

Все это связано с концепцией самостоятельного и постепенного, при помощи учителя, формирования учащимся больших идей о мире.

## Школьные предметы

Официально провозглашается (Концепция преподавания русского языка<sup>1</sup>, ФГОС<sup>2</sup>), что при освоении в школе предмета «Русский язык» самым важным является развитие способности к коммуникации. Если изучить документы традиционной программы по русскому языку, учебники, мы увидим, что реально они не имеют отношения к формированию у ребенка этой способности. Курс русского языка в нашей школе воспитывает способность к коммуникации в лучшем случае в 5% времени учащегося. Уж точно никакой устной коммуникации школа не формирует, да и письменной не особенно, если еще считать, что в литературе написание сочинений как-то относительно-ограничено этому помогает, то в русском языке совсем нет. Основное время — 75% в среднем — мы тратим на то, чтобы добиться орфографической грамотности ребенка. И дальше — лингвистическое знание, школьное, как некоторый элемент (не очень) научной картины мира.

В цифровом мире можно все это перестроить, можно действительно предложить реальную коммуникацию на реальные темы. Например, тот же самый ребенок в начальной школе может записать воспоминания своей бабушки о первом событии в ее жизни и так далее. Дальше можно показать на том же самом экране, можно сопровождать своё устное выступление (обратим внимание — устное) какими-то тезисами, и это поможет школе действительно дать реальный толчок к развитию детской коммуникации. Результат грамотности может быть достигнут, конечно, если говорить опять о неграмотности под твоей черепной коробкой, а грамотности использования тех самых инструментов,

---

<sup>1</sup> Концепция преподавания русского языка и литературы в РФ. Утверждена Распоряжением Правительства РФ от 9 апреля 2016 г. № 637-п. <http://static.government.ru/media/files/GG2TF4pq6RkGAtAIJKHYKTXDmFlMAAOd.pdf>

<sup>2</sup> Приказ Минобрнауки России от 6.10.2009 г. № 373 «Об утверждении и введении в действие федерального государственного образовательного стандарта начального общего образования». <http://base.garant.ru/197127/> (актуально на 3.11.2019).



которые используют взрослые люди. Все мы пишем тексты на компьютере, компьютер отмечает наши ошибки, мы понимаем, когда это на самом деле ошибка, когда это на самом деле некоторое новое слово, новая конструкция, ситуация, которую мы хотим использовать, и это часть нашей сегодняшней грамотности в расширенном сознании.

Об освоении теории языка — лингвистики. Большая идея, общая парадигма современного образования, несуществующая в реальности, но потенциально рассматриваемая многими образовательными философами и идеологами, — это образование как исследование, образование как самостоятельное открытие чего-то нового, образование как решение неизвестных заранее задач. В школьном языке эту идею тоже можно реализовать, у нас имеется прекрасный Национальный корпус русского языка<sup>1</sup>, можно его исследовать и отыскивать в нем те или иные закономерности: а можно ли писать букву «ю» после буквы «ш»? Давайте посмотрим, есть ли такие слова в русском языке, и создадим теорию про это написание.

О литературном образовании. В 2019 году во всероссийском конкурсе «Учитель года» победила волгоградская учительница русского языка и литературы Лариса Арачашвили. Вот что она говорит о своем подходе к преподаванию литературы: «Я почти не использую учебники в классах, потому что там тяжелые и почти неподъемные для школьников статьи. У современных детей совсем другое мышление»<sup>2</sup>. Обратим внимание, что другое мышление здесь используется не в качестве ругательства, «у них клиповое мышление», а как факт — другое у них мышление. «Поэтому я иногда прибегаю к интерактивным урокам или циклам лекций, которые читает Быков. Конечно, не так, что включила и ушла... Естественно, мы с учениками используем гаджеты. Например, когда изучаем поэзию, включаем приложение Mentimeter.

---

<sup>1</sup> Национальный корпус русского языка. <https://ruscorpora.ru/new/>

<sup>2</sup> Лариса Арачашвили. «Я знала, что зарплата у меня будет невысокая. И я не жалуюсь, а работаю». Платформа «Мел», 2019 г. [https://mel.fm/ucheba/uchitelya/416782-teacher\\_of\\_russia\\_2019](https://mel.fm/ucheba/uchitelya/416782-teacher_of_russia_2019)

...Еще мы делаем буктрейлеры...». Впечатляет, как устроен не просто какой-то учитель, а Учительница года, которую выдвинули в Учителя года уважаемые люди. Эти столпы российского образования сочли, что это и есть современный Учитель года, это очень показательно. На сайте «Учитель года России» об Учителе России 2016 года Александра Шагалова<sup>1</sup> говорится: «Методическая фишка» Александра Шагалова — использование аудиозаписей, на которых писатели и поэты читают свои произведения». И это действительно главное, что позволило ему победить. Шагалов брал оцифрованные пластинки — Ахматова, Пастернак читают свои произведения. Вот это то, как происходит, конечно, не с каждым учителем, но с теми учителями, которые, с одной стороны, что-то делают нестандартное, с другой стороны, ощущаются как лидеры образовательного сообщества в стране сегодня.

Примерно понятно, как можно говорить о других школьных предметах. Переход от традиционных целей, например естествознания, к современным целям, к самостоятельному открытию, формированию методологии больших идей. Сегодня ты можешь поднести свой мобильник к любому цветочку или листочку, и он тебе скажет, что это такое. В чем идея систематики? Это становится важным и интересным вопросом, в отличие от знания конкретных фактов.

Конечно, очень важна математика. Обратим внимание только на один аспект. Мы традиционно провозглашаем те или иные цели, которые на самом деле в очень малой степени реализуем. В математике мы заботимся о вычислительных навыках. Очень хорошо воспитывать у человека аккуратность и безошибочность действий, особенно это было хорошо 100 или 200 лет назад. Но сегодня, конечно, важнее, когда человек умеет обращать внимание на реакцию среды, других людей, начальства, в том числе, или потребителя. Значит, наличие обратной связи, той самой кибернетической обратной связи, является куда более важным

---

<sup>1</sup> <https://teacher-of-russia.ru/index2018.php?page=winners18>

<https://rg.ru/2016/10/17/reg-ufo/uchitel-goda-shagalov-klassikov-nado-chitat-oshchushchaia-shelest-stranic.html>

для современного человека. Его рефлексивность, чем заведомая безошибочность, которую мы пытаемся воспитать в математике. Мы говорим, что геометрия учит человека мыслить — могла бы, конечно, учить, если бы он там самостоятельно что-нибудь доказывал. Фактически он должен выучить чужие доказательства, и это далеко не лучший способ научиться что-то делать, если ты учишь чужие доказательства. В цифровом мире принципиальным становится функция, на которую математика должна была бы претендовать — это моделировать реальность, но, на самом деле, сегодня это не происходит. В цифровом виде это становится совершенно реальным и возможным.

Выше уже было сказано об обратной связи, что вместо того чтобы снижать балл за ошибки, нужно посмотреть, насколько ты можешь свою ошибку понять и объяснить, выявить и исправить, экспериментировать с математическими объектами. И важная математика как решение именно новых задач, именно как самостоятельное открытие тех или иных подходов, методов и алгоритмов, фактов из математической реальности, потом с передачей их компьютеру.

Это трансформация математики.

## **Нельзя игнорировать контрреволюцию**

Несмотря на очевидное движение системы образования в сторону цифровизации, очень сильно и противодействие. В Интернете можно найти огромное количество контрреволюционных текстов. Вот, например, перечень, приведенный в статье Андрея Афанасьева<sup>1</sup> «Цифровизация образования, все минусы электронной школы. Что будет с детьми?»:

1. Непроверенные технологии.
2. Утрата навыков письма, как следствие — утрата способностей к творчеству.

---

<sup>1</sup> <https://vc.ru/flood/43800-cifrovizaciya-obrazovaniya-vse-minusy-elektronnoy-shkoly-chto-budet-s-detmi>

3. Утрата способностей воспринимать большие тексты.
4. Экранная зависимость.
5. Снижение социальных навыков.
6. Цифровое слабоумие. Утрата умственных способностей.
7. Использование вай-фай в школах. Электромагнитное излучение.
8. Проблемы с речевым развитием у детей.
9. Проблемы со зрением.
10. Компьютерная, игровая зависимость.
11. Отказ от бумажных учебников.
12. Разница между чтением с экрана и с бумаги.
13. Электронное досье на каждого ребенка, контроль за семьей.
14. Зарубежный опыт цифрового образования.
15. Чего ждать учителям?
16. Чипизация.

Луддизм — неистребим. Страх перед непредсказуемым будущим — неотъемлемое свойство человеческой природы. Так что предстоит большая просветительская работа среди старшего поколения.

И все это опять возвращает нас к вопросу о том, правы ли мы в сегодняшней школе, что мы отрицаем способность ребенка ко всему этому? Отрицаем его право на то, чтобы обращаться к электронному цифровому миру? Открыв Интернет, можно найти такой текст: «Они разрушат этот мир: пугающее поколение Z»<sup>1</sup>. Вот очередной пример такого контрреволюционного текста. И фактически что можно было бы сделать, если все-таки считать, что нужно двигаться в этом направлении? Прежде всего, конечно, не нужно приказывать всем перейти на какие-нибудь интерактивные доски или мобильные телефоны, или что-то еще. Речь идет о том, чтобы разрешить школе, разрешать учителям, разрешать детям использовать, в том числе при сдаче всяких экзаменов, цифровые ресурсы, которые самими взрослыми

---

<sup>1</sup> Андрей Афанасьев. Они разрушат этот мир: пугающее поколение Z // РИА новости, 7 января 2020 г. <https://ria.ru/20200107/1562255259.html>

используются в жизни и в работе. С другой стороны, нужно было бы уважать позицию родителей, которые желают оградить своих детей от цифрового мира, и позицию родителей, желающих обеспечить освоение своими детьми доцифровых навыков — знание текстов наизусть (надо помочь, в том числе с помощью цифровых технологий, выучить «Евгения Онегина», включая и последние главы), писать нажимом тонким пером, заниматься ментальной арифметикой, если родители этого хотят, а ребенку это оказывается по тем или иным причинам интересным.

### **Исторические истоки, параллели и ассоциации**

Это не первая попытка модернизации отечественной школы. В заключение приведем пример того, как это происходило в конце 50-х годов, в момент отказа от перьевого письма в школе.

Письмо Минпроса РСФСР от 25 апреля 1958 года «О пользовании авторучками учащимися общеобразовательных школ»<sup>1</sup>.

«В настоящее время в Министерство просвещения поступают письма с просьбой разрешить учащимся V–X классов семилетней средней школы пользоваться авторучками при выполнении письменных работ. Некоторые методисты предлагают начать постепенный переход на письмо авторучкой даже в I–IV классах, в отдельных школах (210-я школа г. Ленинграда, 129-я школа г. Москвы) с разрешения местных органов народного образования уже проводится опыт применения авторучки в этих классах. В результате годовой опытной работы в этих школах достигнуты положительные результаты.

Рассмотрев поступившие предложения и обсудив опыт вышеуказанных школ, Министерство просвещения РСФСР считает возможным:

---

<sup>1</sup> Письмо Минпроса РСФСР от 25.04.1958 № 81-М «О пользовании авторучками учащимися общеобразовательных школ». [http://gospismo.ru/library/pisma-ussr/sssrpismo\\_32/](http://gospismo.ru/library/pisma-ussr/sssrpismo_32/)

а) разрешить учащимся V–X классов пользоваться авторучкой при выполнении письменных работ (то есть просто повсюду *разрешить* — А.Л.);

б) с целью расширения опытной работы по применению авторучки в начальных классах рекомендовать Министерством просвещения АССР, край-(обл-гор-) ОНО организовать в 1958/59 учебном году под руководством институтов усовершенствования учителей и педагогических институтов такую же работу в 1–2 школах республики, края (области), города.

О результатах опытной работы сообщите в Министерство просвещения по окончании 1958/59 учебного года.»

Тем самым уже есть прецедент создания дырки в заборе, отгораживающем школу от жизни.

А вот о калькуляторах — цифровой трансформации изучения математики в 1982 году: «Эксперимент показал, чтобы существенную помощь оказывают микрокалькуляторы на уроках математики, физики и химии в старших классах, где они используются не только для упрощения вычислительной работы, но и взамен математических таблиц и для подведения учащихся к некоторым гипотезам, которые доказываются в дальнейшем», «Использование микрокалькулятора в учебном процессе позволяет высвободить время для более глубокого изучения учебного материала и более прочного овладения основными знаниями и умениями.» «Однако в использовании микрокалькулятора на занятиях с младшими школьниками следует проявлять большую осторожность, поскольку учащихся необходимо научить свободно пользоваться письменными алгоритмами арифметических действий с действительными числами, устно выполнять вычисления, встречающиеся в алгебраических преобразованиях, а также при контроле вычислительной работы, выполняемой прикидкой (приближенной оценкой величины результата). Преувеличенное же внимание к счету с помощью микрокалькулятора может отрицательно сказаться на уровне вычислительных действий учащихся.» «Основываясь на результатах эксперимента, можно утверждать, что применение микрокалькуляторов в условиях сформированных вычислительных навыков не приводит

к их утрате.»<sup>1</sup> И в заключение этого раздела — историческое Постановление ЦК КПСС и Совета Министров СССР от 28 марта 1985 года № 271 «О мерах по обеспечению компьютерной грамотности учащихся средних учебных заведений и широкого внедрения электронно-вычислительной техники в учебный процесс»<sup>2</sup>. В нем были уже заложены основные положения цифровой трансформации школы.

*Заключение.* Приведем основные выводы, которые можно сделать исходя из нашего представления о человеке как расширенной личности в применении к современному обществу:

1. Человек в результате революции искусственного интеллекта за последние 100 лет приобрел принципиально новые способности по сравнению с человеком предыдущих поколений, среди них способность, умение и желание применять цифровые технологии. При этом, и не прибегая в какой-то ситуации к этим технологиям, он иначе мыслит, общается и действует. Происходящие изменения приводят к утрате важности в жизни и развитии человека многих традиционных качеств и к «контрреволюции» — установке, общественной позиции и действиям, направленным на дискредитацию и торможение изменений.

2. Центром и бастионом контрреволюции, сопротивления изменениям, определяемых расширением личности, является система образования, прежде всего — общеобразовательная школа. Система образования — в лице многих своих работников, руководителей, активных представителей родительской общности исходит из представлений о результатах образования для не расширенной личности, изолированной от современного

---

<sup>1</sup> Об использовании микрокалькуляторов в учебном процессе (Инструктивно-методическое письмо), НИИ содержания и методов обучения АПН СССР и Главное управление школ Министерства просвещения СССР, Математика в школе, № 3, 1982, С. 6–8. <https://uch-lit.ru/periodika/dlya-uchiteley-i-uchashhihsya/matematika-v-shkole/matematika-v-shkole-metodicheskii-zhurn-115>

<sup>2</sup> Постановление ЦК КПСС и Совета Министров СССР от 28 марта 1985 года № 271 «О мерах по обеспечению компьютерной грамотности учащихся средних учебных заведений и широкого внедрения электронно-вычислительной техники в учебный процесс». Вопросы образования. 2005. № 3. С. 341–346.

мира. Такой подход к результатам, выраженный, в частности, в государственной итоговой аттестации, может стать главным препятствием для гармоничного развития личности XXI века.

(Благодарности)<sup>1</sup>

---

<sup>1</sup> Первый автор благодарен очень многим, среди них (неполный список): Симору Паперту, который вместе со своими коллегами из МедиаЛаб помог осознать возникшее у автора благодаря Николаю Николаевичу Константинову представление об образовании и роли технологии в нем; Владиславу Александровичу Лекторскому за приглашение выступить на семинаре «Философско-методологические и научно-теоретические проблемы ИИ» Научного совета по методологии искусственного интеллекта при Отделении общественных наук РАН, а особенно за обсуждение, организованное Владиславом Александровичем после доклада, и его собственное участие в этом обсуждении. Он благодарит также Елену Анатольевну Труфанову за ее исходное выступление на пусковом семинаре программы РФФИ и всех участников Программы РФФИ «Фундаментальные основы цифровой трансформации общего образования», Артема Симоновича Соловейчика за обсуждение «Хартии» и хода Программы; Александра Изотовича Адамского и Александра Григорьевича Асмолова за настойчивые попытки понять автора и возникшее в результате понимание автором себя.

Список благодарностей второго автора должен быть еще больше, но дабы не утомлять читателя, ограничусь наиболее значимыми для этого текста персонами: прежде всего, это мой соавтор Алексей Львович Семенов, который заставил меня взглянуть на проблему цифровизации образования в широком антропологическом контексте. Это мои Учителя в педагогике Юлия Иосифовна Турчанинова и Алексей Иванович Пискунов — люди, благодаря которым, как мне кажется, у меня появилось представление о педагогической традиции и о трансформации образования. Это выдающиеся психологи Вадим Артурович Петровский и Станислав Гроф, в беседах с которыми я осознал, насколько сложное и неоднозначное явление — личность человека.

Авторы также благодарят Татьяну Александровну Рудченко и В. Кондратьева за помощь в подготовке данного и других текстов.



## О СТАНДАРТИЗАЦИИ ПОНЯТИЙ И ТЕРМИНОВ, СВЯЗАННЫХ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

---

---

### Общие вопросы и направления стандартизации понятийного аппарата в сфере применения искусственного интеллекта

Вопросы создания цифровых технологий и искусственного интеллекта приобретают новый, актуальный смысл, и это характерно для всех развитых стран мира, в том числе и для России<sup>1</sup>.

---

\* *Любимов Алексей Павлович* — доктор юридических наук, профессор, заместитель Главного ученого секретаря Президиума РАН, заместитель председателя Научного совета при Президиуме РАН по методологии искусственного интеллекта и когнитивных исследований, Москва.

*Alexey P. Lyubimov* — Doctor of law, Professor, Deputy Chief Scientific Secretary of the RAS Presidium.

<sup>1</sup> *Васильев С.Н.* Задача — озадачиваться. Искусственный интеллект научат ставить цели // Поиск. № 50. 13 декабря 2019 г. С. 10; Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (а также «Национальная стратегия развития искусственного интеллекта на период до 2030 года»); Указ Президента РФ от 01.12.2016 № 642 «О Стратегии научно-технологического развития Российской Федерации» // Собрание законодательства РФ, 05.12.2016, № 49, ст. 6887; Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы» // Собрание законодательства РФ, 15.05.2017, № 20, ст. 2901; Любимов А.П., Пономарева Д.В., Барабашев А.Г. К вопросу о понятии искусственного интеллекта в российском праве // Актуальные вопросы экономики, управления и права: сборник научных трудов (ежегодник). 2019. № 2–3. С. 16–34;

Искусственный интеллект стал проникать во все сферы жизни человека. Важнейшим направлением использования ИИ, являющегося одним из ветвей современной информатики, является разработка и использование аппаратно-программных средств, позволяющих ставить и решать сложные интеллектуальные задачи пользователям, не имеющим специальной профессиональной подготовки и использующим ограниченный словарь естественного языка.

Появление новых возможностей и выгод при расширении сферы использования искусственного интеллекта связано со многими сферами деятельности человека и его потребностями. К ним, в частности, можно отнести:

- технологии искусственного интеллекта в экономике: на крупных и малых производствах;
- IT-технологии в органах управления;
- IT-технологии в законотворческой деятельности;
- IT-технологии в судопроизводстве;
- IT-технологии в экспертной деятельности;
- отраслевые IT-технологии в космосе, авиации, оборонном секторе, на транспорте, в связи;
- IT-технологии в науке, высшей школе, образовании;
- коллективный общественный интеллект, искусственный интеллект в глобальных информационных и социальных сетях;
- домовая и квартирная техника: от технологического искусственного охранника до умного дома;
- бытовая техника на основе информационных технологий и применения искусственного интеллекта;
- офисная техника. Компьютер, планшет и телефон как современный мобильный офис;
- IT-технологии в области мега-данных, ситуационной видеоаналитики;

---

*Любимов А.П., Пономарева Д.В., Барабашев А.Г.* Искусственный интеллект в европейской правовой доктрине // Вестник Дипломатической академии МИД России. Международное право. 2019. №1 (10). С. 32–47; *Любимов А.П., Пономарева Д.В., Барабашев А.Г.* Основные понятия искусственного интеллекта. М.: ООО «Сам Полиграфист», 2019. 116 с.

— **киборгизация человека и животных** — замена частей или всего тела кибернетическими аналогами;  
— и многое другое.

По мнению В.В. Рубанова, применение искусственного интеллекта в экспертной области сосредоточено на распознавании объектов и реагировании на изменения их состояний по некоторым шаблонам. Подобные технологии могут оказывать информационную поддержку экспертам и управленцам, но не заменять их<sup>1</sup>. Нам представляется, что невозможное сегодня завтра будет обыденным явлением искусственного интеллекта. Это подтверждает деятельность нашего Нобелевского лауреата, академика РАН Ж.И. Алферова, который сначала работал с простейшими проводниками, а затем довел эти технологии до искусственного интеллекта<sup>2</sup>.

### **Правовые проблемы, возникающие в процессе проектирования, производства и эксплуатации технических средств с искусственным интеллектом**

Если с IT-технологиями и искусственным интеллектом в бытовой сфере, а равно и в промышленности, многое понятно, то в более рискованных и опасных сферах не всё обстоит так просто. Об этом свидетельствуют примеры использования ИИ в области разработки вооружения, что приводит к опасному соперничеству и, соответственно, правовым ограничениям.

Проблемы разного уровня при использовании ИИ могут возникнуть во многих сферах деятельности человека: от экономической, военной, транспортной, энергетической сфер, при

---

<sup>1</sup> Рубанов В.В. Не подражать живым системам // Независимая газета. 23.03.2020. URL: [http://www.ng.ru/stsenarii/2020-03-23/9\\_7824\\_consciousness.html](http://www.ng.ru/stsenarii/2020-03-23/9_7824_consciousness.html) (дата обращения: 23.08.2021).

<sup>2</sup> Любимов А.П. Жорес Алферов — легенда мировой науки: от полупроводников до искусственного интеллекта // Представительная власть — XXI век. 2019. № 1–2. С. 38.

имплантации киберфизических систем в тело человека, создании дополненной реальности — до социальной сферы (например, при составлении социальных рейтингов). В связи с этим необходимо определить само понятие «искусственный интеллект» как основу для создания и регулирования цифрового рынка.

В учебном смысле впервые термин, приближенно обозначающий понятие «искусственный интеллект», был использован в 1956 году на семинаре в Дартсмутском колледже (США) и был определён как «умение рассуждать разумно» (AI — artificial intelligence). Позднее, в 1969 году, в Вашингтоне, на I Международной конференции по искусственному интеллекту название термина ИИ было официально использовано в научном обороте.

В настоящее время идёт разработка Европейской стратегии развития цифровой экономики, в основе которой лежат новейшие информационно-коммуникационные технологии с использованием искусственного интеллекта<sup>1</sup>.

Понятия, определяющие виды, сущность и опасности, сопровождающие всеобъемлющее распространение искусственного интеллекта, зависят от сферы его применения. Так, применение искусственного интеллекта в сфере деятельности вооружённых сил опасно возможностью нарушения самого основного и неотъемлемого права человека — права на жизнь. Известно, что во всем мире новейшие достижения науки всегда брались в разработку и внедрение в первую очередь военными ведомствами. Однако предоставление возможности какой-либо военной структуре с искусственным интеллектом принимать решение о начале военных действий, об уничтожении людей или объектов создаёт проблему определения субъекта ответственности в случае ошибки при принятии таких решений, что формирует обширное поле для преступной безнаказанности. Из этого следует одно из требований, которое, на наш взгляд, должно быть применено в области действия искусственного интеллекта и закреплено в определении

---

<sup>1</sup> Парфенова Л.Б. Европейская стратегия развития цифровой экономики: региональная дифференциация // Вестник Тверского государственного университета. Серия: Экономика и управление. 2018. № 3. С. 30–38.

его понятия — действие структур с искусственным интеллектом должно происходить исключительно в рамках опознавательных факторов, к которым можно отнести: регистрацию, информацию о владельце и управлении, о фактах перехода контроля в том числе путём удаленного доступа и т. п.<sup>1</sup>

По мнению С.Ю. Кашкина и А.В. Покровского, другая перспективная сфера применения искусственного интеллекта — это развитие киберфизических систем, в том числе их имплантация в тело человека<sup>2</sup>.

Киберфизические системы, состоящие из различных природных объектов, искусственных подсистем и управляющих контроллеров, призваны обеспечивать тесную связь между физическими процессами и компьютероинтегрированными встроенными устройствами или системами. Киберфизические системы возникают на стыке Интернета людей, вещей и сервисов. Они распространяются практически на все виды человеческой деятельности, включая промышленные, транспортные, энергетические, военные и экономические системы, а также на все механизмы жизнеобеспечения: от медицины до создания умных домов и городов. Уже сейчас применение искусственного интеллекта в киберфизических системах усиливает аналитические способности человека посредством создания интерактивных систем нового уровня при сохранении человека в контуре управления.

Искусственный интеллект постепенно завоевывает все сферы нашей жизни. Переход человеческого разума в полностью виртуальное пространство уже не кажется мечтами фантастов. На наших глазах происходит слияние человеческого разума и компьютера. На базе развития идеологии философии трансгума-

---

<sup>1</sup> *Любимов А.П.* Основные подходы к определению понятия «искусственный интеллект» // Научно-техническая информация. Серия 2: Методика и организация информационной работы. 2020. №9. С. 1–6.

<sup>2</sup> *Кашкин С.Ю., Покровский А.В.* Искусственный интеллект, робототехника и защита прав человека в Европейском союзе // Вестник Университета им. Кутафина (МГЮА). 2019. № 4 (56). С. 78.

низма идёт формирование нового, улучшенного постчеловека с искусственным интеллектом.

Однако переход к новой ступени эволюции путём модификации тела с использованием технологий искусственного интеллекта, предполагаемый трансгуманистами, оставляет без внимания вопросы духовности, и возникает опасность утраты людьми тех качеств и свойств, которые формируют человечество как биологический вид. Чипы, встроенные в тело человека, уже сейчас позволяют идентифицировать людей при совершении ими покупок, при управлении автомобилем или системой «умный дом». Таким путём запущен процесс расчеловечивания личности, что неминуемо ведёт к расслоению общества, так как объективное отсутствие, по разным причинам, равного доступа к новейшим технологиям модификации тела создаёт социальное неравенство.

В 1997 году Ник Бостом (шведский физик и философ) вместе с британским предпринимателем и философом Дэвидом Пирсом основали в Великобритании Всемирную Трансгуманистическую Ассоциацию, объединившую более 3 тыс. членов (в 2008 году ассоциация поменяла своё название на Humanity+). Эти трансгуманисты позиционируют себя преемниками античного гуманизма и утилитаризма, построенного на продвижении идеологии оправдания применения высоких технологий для продления жизни и достижения бессмертия финансовой олигархии. Акцент делается на расширении применения искусственного интеллекта, оцифровки человеческого сознания, крионике (заморозка умерших людей для последующего оживления), внедрении нанороботехники в медицине и т. п. В 2000 году и в нашей стране (в Москве) было образовано Российское трансгуманистическое движение, представители которого позиционируют себя как «продолжатели дела Н.Ф. Фёдорова, К.Э. Циолковского, В.И. Вернадского и других представителей русского космизма» и пропагандируют в России «идеи научного иммортализма и трансгуманизма, а также новые исследования и технологии, которые позволяют каждому превзойти «естественные» пределы возможностей

своего тела, своего разума, своей жизни»<sup>1</sup>. В США сформирована и действует Трансгуманистическая партия, и её представитель, Золтон Иштван, в 2016 году даже баллотировался на пост президента США. Эта партия является частью международного движения, в основе которого лежит мировоззрение, что биологический человек настоящего — это не окончательная ступень эволюции, а перспектива развития сферы применения технологий искусственного интеллекта ведёт к неизбежности и целесообразности полной замены органов человека и отмены физического тела<sup>2</sup>.

Ускоряющийся рост числа устройств со встроенными процессорами и средствами хранения большого объёма данных приводит к лавинообразному увеличению объема информации, необходимой для принятия решений. Это вступает в противоречие с ограниченными когнитивными способностями человека, эволюционирующими медленнее, чем кибернетические возможности, в результате этого нарастает опасность вытеснения человека из процесса управления.

Необходимость обработки большого объёма информации и возникновение глобального оборота данных обусловили разработку и внедрение новейших технологий искусственного интеллекта (ИИ). Поэтому при формулировании понятийного аппарата ИИ ключевой позицией является описание функции сбора и анализа огромного массива фактов для принятия на их основе необходимых оптимальных решений.

Для осуществления своей стратегии по искусственному интеллекту, принятой в апреле 2018 года, Европейская комиссия 7 декабря того же года представила согласованный план, подготовленный вместе с государствами-членами для содей-

---

<sup>1</sup> Всемирная Трансгуманистическая Ассоциация / Российское трансгуманистическое движение. URL: <http://transhumanism-russia.ru> (дата обращения: 20.05.2020); Communication from the European Commission of 07.12.2018 // Coordinated Plan on Artificial Intelligence. 2018. COM 795. P. 2; European Parliament. Reports A8–0005/2017 of 27.01.2017 with recommendation to the Commission on Civil Law Rules on Robotics (2015/ 2103(INL)). P. 15.

<sup>2</sup> *Соснов А.Я.* Роли для робота // Поиск. 2019. № 46. 15 ноября. С. 13.

ствия развитию и использованию ИИ в Европе. В этом плане декларируется, что «осмысливая огромные объёмы данных для предложения эффективных решений, искусственный интеллект совершенствует продукты, процессы и бизнес-модели во всех секторах экономики»<sup>1</sup>.

Европейский парламент ещё в 2017 году обратился с рекомендацией разработать для Европейской комиссии по гражданско-правовому регулированию робототехники меры для «ограничения рисков, связанных с возможностью взлома или отключения оперативных систем программирования, встроенных в организм человека, либо стирания их памяти, что может поставить под угрозу здоровье человека, а в особых случаях даже его жизнь»<sup>2</sup>.

Понятийный аппарат искусственного интеллекта должен лечь в основу модернизированных и отвечающих запросам современности правовых норм защиты интеллектуальной собственности и прав человека, так как неизбежное расширение сферы применения ИИ изменяет направления развития цивилизации и влияет на трансформацию понятий главных ценностей человечества. На материальный объект или услугу, создаваемые с использованием искусственного интеллекта, должны распространяться определённые права на интеллектуальную собственность, которые в перспективе станут основным источником современного синергетического и мультипликационного эффекта интеллектуального и материального производства.

Сложности в определении понятия искусственного интеллекта заключаются в том, что мы сталкиваемся с совершенно новым для человечества явлением, объединяющим человеческие характеристики (что предопределяет необходимость поднять вопрос о защите прав человека) и механические свойства искусственного интеллекта. Правосубъектность единицы искусственного интеллекта может быть определена только при

---

<sup>1</sup> Communication from the European Commission of 07.12.2018 // Coordinated Plan on Artificial Intelligence. 2018. COM 795. P. 2.

<sup>2</sup> European Parliament. Reports A8-0005/2017 of 27.01.2017 with recommendation to the Commission on Civil Law Rules on Robotics (2015/ 2103(INL)). P. 15.



чётком разграничении ответственности между юридическим и физическим лицами. Поэтому возникает необходимость введения новой формы правосубъектности — электронное лицо<sup>1</sup>.

Таким образом, становится очевидным, что для введения юридически значимого определения понятия «искусственный интеллект» (как для применения в процессе проектирования, производства и эксплуатации технических средств с искусственным интеллектом, так и для использования в процессе судебной деятельности) необходимо проработать не только сущностные характеристики ИИ, обозначить возможные сферы применения, легальные объекты воздействия и механизмы защиты интеллектуальных прав на создаваемый продукт, но и ввести юридически значимые обязательства по соблюдению определённых стандартов безопасности с обеспечением должного уровня контроля действий устройств с участием искусственного интеллекта<sup>2</sup>.

При определении понятия искусственного интеллекта следует учитывать, что существующий класс IT-продуктов (к которым относятся программно-аппаратные системы и технологии автоматического и автоматизированного решения многообразных задач приёма, передачи, накопления, обработки и управления данными и знаниями, в том числе с имитацией некоторых интеллектуальных функций человека, таких как распознавание видеообразов, анализ трёхмерных изображений, обработка и перевод на разные языки текстов и речи, машинное обучение, выявление закономерностей, визуализация и анализ больших объемов данных, облачные вычисления и т. п.), по большому счёту, не является в полной мере искусственным интеллектом, так как пока не обладает способностью человека «самоозада-

---

<sup>1</sup> *Кашкин С.Ю., Покровский А.В.* Искусственный интеллект, робототехника и защита прав человека в Европейском союзе // Вестник Университета им. Кутафина (МГЮА), 2019. № 4 (56). С. 78.

<sup>2</sup> *Любимов А.П.* Основные подходы к определению понятия «искусственный интеллект» // Научно-техническая информация. Серия 2: Методика и организация информационной работы. 2020. № 9. С. 1–6.

чиваться», т. е. не умеет сам себе ставить задачи, осознавая ситуацию и себя в ней. Для них применяются такие определения, как «слабый искусственный интеллект» или «интеллектуальные системы и технологии», поскольку они используют нарабатываемое и уже созданное в области искусственного интеллекта<sup>1</sup>.

Специалисты в области теории управления, математической кибернетики и системного анализа, а также такие учёные, как С.Н. Васильев, Е.Н. Ведута, В.В. Рубанов, при определении ИИ выделяют в основном два понятия: «сильный» и «слабый» искусственный интеллект. К «сильному» ИИ они относят будущие системы и технологии со способностями не только автоматического достижения цели, но и самоозадачивания, то есть обладающие помимо реактивных и других регуляторных механизмов поведения также механизмами целеполагания<sup>2</sup>. К «слабому» ИИ относят интеллектуальные человеко-машинные системы и технологии, в которых функция самоозадачивания реализуется человеком. По мнению В.В. Рубанова, те, кто непосредственно связан с предметными практиками, не спешат называть подобные программы и оснащенные ими устройства искусственным интеллектом, предпочитая более скромный термин — «высокоавтоматизированные системы». Термином «искусственный интеллект» чаще пользуются «институты развития» и бенефициары программ «цифровизации» — для обоснования затрат на проекты по соответствующей графе бюджета<sup>3</sup>.

---

<sup>1</sup> См.: там же.

<sup>2</sup> Васильев С.Н. Задача — озадачиваться. Искусственный интеллект научат ставить цели // Поиск. 2019. № 50. 13 декабря. С. 10; Ведута Е.Н., Любимов А.П., Джакубова Т. Н., Ряскова Е.С. Концепция национальной программы создания проактивного искусственного интеллекта // Представительная власть — XXI век. 2019. № 4. С. 22–29; Рубанов В.В. Не подражать живым системам // Независимая газета. 23.03.2020. URL: [http://www.ng.ru/stsenarii/2020-03-23/9\\_7824\\_consciousness.html](http://www.ng.ru/stsenarii/2020-03-23/9_7824_consciousness.html) (дата обращения: 23.08.2021).

<sup>3</sup> Рубанов В.В. Указ. соч.

## **Федеральный проект «Искусственный интеллект»**

Минэкономразвития России в целях реализации Национальной стратегии развития искусственного интеллекта на период до 2030 года, утвержденной Указом Президента Российской Федерации от 10 октября 2019 года № 490, разработало паспорт федерального проекта «Искусственный интеллект» при участии Минцифры России, Центра компетенций по Федеральному проекту (ПАО Сбербанк), рабочей группы при АНО «Цифровая экономика» по Федеральному проекту (представители бизнеса и экспертного сообщества), заинтересованных федеральных органов исполнительной власти.

Паспорт федерального проекта «Искусственный интеллект» утверждён на заседании Президиума Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности 27 августа 2020 года. При этом датой вступления в силу Федерального проекта является дата утверждения измененной национальной программы «Цифровая экономика Российской Федерации», включающей Федеральный проект.

Финансирование Федерального проекта «Искусственный интеллект» в 2021–2024 годах составит 31,5 млрд рублей, из которых 24,6 млрд рублей будет выделено из федерального бюджета, также будет привлечено из внебюджетных источников 6,9 млрд рублей.

Также предполагается обеспечить дополнительное финансирование на внедрение ИИ в госсекторе за счёт государственных средств, предусмотренных на цифровую трансформацию ведомств и соответствующие стандарты.

## **Отраслевая и функциональная стандартизация понятийного аппарата**

В 2020 и 2021 годах утверждён ряд новых национальных стандартов для специалистов в области цифровых технологий<sup>1</sup>. К ним относятся:

### **ГОСТ Р 59276-2020 «Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения».**

Стандарт распространяется на системы искусственного интеллекта, обеспечивающие решение конкретных практически значимых задач:

- определено понятие доверия к системам искусственного интеллекта;
- приведена классификация факторов, влияющих на качество и способность систем искусственного интеллекта вызывать доверие на стадиях жизненного цикла;
- формализована взаимосвязь качества и способности систем искусственного интеллекта вызывать доверие;
- приведена классификация основных способов обеспечения доверия к системам искусственного интеллекта.

Стандарт направлен на обеспечение доверия к системам искусственного интеллекта со стороны потребителей результатов работы этих систем и, при необходимости, со стороны организаций, ответственных за регулирование вопросов создания и применения систем искусственного интеллекта на основе подтверждения их качества. Применяется на этапах создания и эксплуатации систем ИИ.

### **ГОСТ Р 59277-2020 «Системы искусственного интеллекта. Классификация систем искусственного интеллекта».**

В настоящем стандарте установлена схема классификации, отражающая основные особенности систем искусственного

---

<sup>1</sup> Консорциум Кодекс: <https://kodeks.ru/news/read/utverjdeny-novye-nacionalnye-standarty-dlya-specialistov-v-oblasti-cifrovyyh-tehnologiy-2021-06-01> (дата обращения 08.09.2021).

интеллекта (СИИ) для решения прикладных задач, помогающая определить направления их стандартизации:

- определены принципы классификации систем искусственного интеллекта;

- установлена схема классификации систем искусственного интеллекта.

Установление классификации систем искусственного интеллекта позволит сравнивать различные решения по таким параметрам, как: вид деятельности, структура знаний, функции контура управления, безопасность, конфиденциальность, степень автоматизации, методы обработки информации, интеграция/интероперабельность, комплексность системы, архитектура, специализация. Это позволит заинтересованным сторонам выбирать надлежащие решения для их приложений и сравнивать качество доступных решений.

**ГОСТ Р 59278-2020 «Информационная поддержка жизненного цикла изделий. Интерактивные электронные технические руководства с применением технологий искусственного интеллекта и дополненной реальности. Общие сведения».**

Стандарт устанавливает:

- общие требования к интерактивному электронному техническому руководству (ИЭТР) с применением технологий искусственного интеллекта (ИИ) и дополненной реальности (ДР);

- требования к составу ИЭТР с применением технологий ИИ и ДР;

- требования к функциональности ИЭТР с применением технологий ИИ и ДР;

- требования к разработке ИЭТР с применением технологий ИИ и ДР.

В состав современных ИЭТР, помимо электронных документов об изделии, входят комплексные программно-технические средства, в том числе использующие технологии виртуальной и дополненной реальности, что позволяет предоставлять пользователям ИЭТР качественно новые функциональные возможности при работе с изделиями, в частности, при выполнении типовых

операций в ходе эксплуатации, технического обслуживания и ремонта изделий.

Использование технологий искусственного интеллекта и дополненной реальности в составе ИЭТР направлено на расширение функциональных возможностей ИЭТР для повышения качества решения задач ИЭТР.

**ГОСТ Р 56045-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Рекомендации по оценке мер обеспечения информационной безопасности».** Утверждён приказом Росстандарта от 20 мая 2021 года № 421 — *ст. вводится в действие на территории РФ с 30 ноября 2021 года.*

Стандарт предоставляет рекомендации по оценке реализации и функционирования мер обеспечения ИБ, включая оценку технического соответствия мер обеспечения ИБ информационных систем, согласно установленным в организации стандартам по ИБ. В стандарте предлагаются инструкции по анализу и оценке мер обеспечения ИБ, используемых в рамках системы менеджмента ИБ, описание которой приводится в ИСО/МЭК 27001. Стандарт применим для организаций всех видов и любых размеров, включая акционерные общества открытого и закрытого типов, государственные учреждения и некоммерческие организации, проводящие оценки ИБ и технического соответствия.

**ГОСТ Р 59381-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Основы управления идентичностью. Часть 1. Терминология и концепции».** Утверждён приказом Росстандарта от 20 мая 2021 года № 413 — *ст. вводится в действие на территории РФ с 30 ноября 2021 года.*

Стандарт определяет термины, относящиеся к управлению идентичностью, основные концепции идентичности и управления идентичностью, а также их взаимосвязь. Положения стандарта применимы для любой информационной системы, обрабатывающей идентификационную информацию, и используются совместно с документами по стандартизации, регламентирую-

щими вопросы идентификации. Стандарт предназначен для применения путем включения нормативных ссылок на него в соответствии с действующим законодательством и (или) прямого использования устанавливаемых в нем положений.

**ГОСТ Р 59382-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Основы управления идентичностью. Часть 3. Практические приемы».** Утвержден приказом Росстандарта от 20 мая 2021 года № 412 — *ст. вводится в действие на территории РФ с 30 ноября 2021 года.*

Стандарт представляет собой руководство по управлению идентичностью и обеспечению уверенности в том, что система управления идентификационными данными соответствует его требованиям. Стандарт применим для систем управления идентификационными данными, в которых осуществляется получение, обработка, хранение, передача или использование связанных с сущностями идентификаторов и/или идентификационной информации (в том числе персональных данных) с целью идентификации или аутентификации сущностей и/или с целью принятия решений с применением атрибутов сущностей. Практические приемы управления идентификационными данными могут также рассматриваться в других стандартах. Положения стандарта не исключают применение криптографических методов (алгоритмов) при управлении идентичностью, но не устанавливают требования по их реализации.

**ГОСТ Р 59383-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Основы управления доступом».** Утверждён приказом Росстандарта от 20 мая 2021 года № 414 — *ст. вводится в действие на территории РФ с 30 ноября 2021 года.*

Стандарт определяет основы управления доступом и безопасного управления процессом доступа к информации и ресурсам средств информационно-коммуникационных технологий. В стандарте приведены концепции, термины и определения, применимые для методов управления распределенным доступом

в сетевой среде, представлены разъяснения, касающиеся взаимосвязанной архитектуры, компонентов и функций управления. Положения стандарта могут быть использованы при управлении доступом к информационным ресурсам, вычислительным ресурсам средств вычислительной техники, ресурсам автоматизированных (информационных) систем, средствам вычислительной техники и автоматизированным (информационным) системам в целом. Описание характеристик и качества средств управления физическим доступом, задействованных в системах управления доступом, выходят за рамки применения настоящего стандарта.

**ГОСТ Р 59385-2021 «Информационные технологии. Искусственный интеллект. Ситуационная видеоаналитика. Термины и определения».**

Вводимые в настоящем стандарте понятия отражают внутренние логические связи, соответствующие пути от оцифровки изображений с помощью устройств формирования видеоизображений, в том числе с применением нанотехнологических решений, к результатам когнитивного анализа посредством технических систем искусственного интеллекта.

Стандарт устанавливает единое терминологическое обеспечение для осуществления взаимопонимания между всеми участниками процесса: разработчиками, поставщиками, пользователями, прочими заинтересованными сторонами. В стандарте описаны:

- общие термины и определения;
- термины и определения, относящиеся к видам анализа;
- термины и определения, относящиеся к сцене видеонаблюдения;
- термины и определения, относящиеся к ситуациям и сценариям.

Настоящий стандарт является первым в группе стандартов, устанавливающих нормативные требования в области ситуационной видеоаналитики, которые будут регламентировать эксплуатационные характеристики, методики испытаний и требования к размещению оборудования технических систем ситуационной видеоаналитики. Термины и определения, установленные



настоящим стандартом, рекомендуются для применения во всех видах документации и литературы, входящих в сферу действия работ по стандартизации интеллектуальных систем ситуационной видеоаналитики и (или) использующих результаты этих работ. Понятия, введенные в настоящем стандарте, рекомендуются к использованию при разработке технических заданий, нормативных документов, производстве продукции предметной области данного стандарта.

**ГОСТ Р 59391-2021 «Средства мониторинга поведения и прогнозирования намерений людей. Аппаратно-программные средства с применением технологий искусственного интеллекта для колесных транспортных средств. Классификация, назначение, состав и характеристики средств фото- и видеофиксации».**

Стандарт является первым в комплексе стандартов по установлению требований применения технологий искусственного интеллекта (ИИ) на транспорте и последовательно развивает серию национальных стандартов РФ «Средства мониторинга поведения и прогнозирования поведения людей» с целью повышения доверия к технологиям ИИ, развития интеллектуальных систем без средств физического воздействия на водителя, а также повышения уровня безопасности на транспорте. Требования стандарта направлены на обеспечение безопасности дорожного движения, жизни и здоровья людей, сохранности их имущества и охраны окружающей среды. В стандарте приводятся:

- минимально необходимый перечень функций средств мониторинга водителя;

- общие требования к интерфейсу средств мониторинга водителя;

- требования к составу (функциональному и аппаратно-программному) и установке средств мониторинга водителя.

Стандарт предназначен для применения в технологических и производственных процессах на предприятиях и в организациях, осуществляющих деятельность по оказанию транспортных услуг с использованием колесных транспортных средств, оснащенных аппаратно-программными средствами с применением

технологий искусственного интеллекта для мониторинга поведения и прогнозирования намерений водителя.

**ГОСТ Р ИСО/МЭК 27000-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Общий обзор и терминология».** Утверждён приказом Росстандарта от 19 мая 2021 года № 392 — *ст. вводится в действие на территории РФ с 30 ноября 2021 года.*

Стандарт содержит общий обзор систем менеджмента информационной безопасности (СМИБ). В нем приведены термины и определения, используемые в семействе стандартов СМИБ. Стандарт применим к организациям любого типа и размера (например, коммерческим предприятиям, правительственным учреждениям, некоммерческим организациям). Термины и определения, представленные в настоящем стандарте: охватывают общие термины и определения, используемые в семействе стандартов СМИБ; не охватывают все термины и определения, применяемые в семействе стандартов СМИБ; не ограничивают применение новых терминов в семействе стандартов СМИБ.

**ГОСТ Р ИСО/МЭК 27033-4-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Безопасность сетей. Часть 4. Обеспечение безопасности межсетевого взаимодействия с использованием шлюзов безопасности».** Утверждён приказом Росстандарта от 19 мая 2021 года № 391 — *ст. вводится в действие на территории РФ с 30 ноября 2021 года.*

В стандарте содержится руководство по обеспечению безопасности межсетевого взаимодействия с использованием шлюзов безопасности (межсетевых экранов, межсетевых экранов уровня приложений, систем предотвращения вторжений и т. д.), в соответствии с документированной политикой информационной безопасности (ИБ) для шлюзов безопасности, включающее в себя: выявление и анализ угроз безопасности сети, связанных со шлюзами безопасности; определение требований по обеспечению безопасности сетей для шлюзов безопасности на основе

анализа угроз; использование методов проектирования и реализации для устранения угроз и решения вопросов управления для типовых схем построения сетей; решение проблем, связанных с внедрением, эксплуатацией, мониторингом и проверкой шлюзов безопасности сети.

### **Национальный стандарт в области больших данных**

В 2021 году в России приняли первый Национальный стандарт в области больших данных.

В стандарте содержатся терминологические нормативы для участников рынка больших данных. Это первый подобный стандарт в России.

**ГОСТ Р ИСО/МЭК 20546-2021 «Информационные технологии (ИТ). Большие данные. Обзор и словарь»** от 13 июля 2021 года соответствует международному стандарту Information technology — Big data — Overview and vocabulary. Его разработали в Национальном центре цифровой экономики МГУ и Институте развития информационного общества в рамках бюджетной программы государственной поддержки центра компетенций «Национальной технологической инициативы» (НТИ) по технологиям хранения и анализа больших данных МГУ.

Как говорится в описании стандарта, термины, которые установлены в нём, обязательны для применения во всех видах документации и литературы по данной научно-технической отрасли.

В ГОСТе приводятся переводы англоязычных терминов, относящихся к большим данным, и расшифровка их значений. Так, «аналитика данных» (data analytics) — это «составное понятие, охватывающее получение, сбор, проверку и обработку данных, включая их количественную оценку, визуализацию и интерпретацию». В свою очередь, «вариативность данных» (data variability) — это «изменения в скорости передачи, формате или структуре, семантике или качестве массива данных».

Кроме того, в стандарте содержатся ключевые характеристики больших данных, в том числе их объем, скорость обработки, разнообразие и вариативность, с объяснением значения каждой.

«Принятие первого национального стандарта задает вектор совместных действий власти, бизнеса, научно-образовательного сообщества по развитию экономики данных в России. Мы ставим перед собой задачу сократить отставание в стандартизации больших данных»<sup>1</sup>, — отмечает председатель совета директоров Института развития информационного общества Юрий Хохлов. По его словам, после введения терминологического стандарта к нему добавятся и остальные.

**Первый международный стандарт** в сфере big data «Большие данные — Требования на основе облачных вычислений и их возможности» был утверждён Международным союзом электросвязи в 2015 году. Спустя три года Международная организация по стандартизации (ISO) утвердила стандарты эталонной архитектуры больших данных, сценариев использования big data, а также направления дальнейшей стандартизации больших данных. В 2020 году были опубликованы стандарты безопасности при использовании персональных данных в big data. В настоящее время ведется работа над серией стандартов о качестве больших данных для аналитики и машинного обучения.

По словам Ю. Хохлова, предстоит принять ещё восемь национальных стандартов в сфере больших данных, касающихся эталонной архитектуры, безопасности, методов анализа, сценариев использования и прочего. Часть из них уже разработана и прошла общественное обсуждение, а другая часть будет разрабатываться одновременно с международными стандартами.

В конце 2019 года Boston Consulting Group оценивала объём российского рынка больших данных в 45 млрд рублей с темпом прироста 12% в течение последних пяти лет. По информации Ассоциации больших данных, объём рынка больших данных в России составляет 10–30 млрд рублей. К 2024 году ожидается его рост до 300 млрд рублей.

---

<sup>1</sup> Хохлов Ю. [https://habr.com/ru/news/t/567940/?bx\\_sender\\_conversion\\_id=402488](https://habr.com/ru/news/t/567940/?bx_sender_conversion_id=402488)

Как отмечает председатель технического комитета «Киберфизические системы» при Росстандарте Никита Уткин, терминологический стандарт позволит участникам рынка больших данных общаться на одном языке при составлении закупочной документации, техзаданий и техдокументации. Кроме того, этот стандарт нужен для разработки регуляторных документов<sup>1</sup>.

Директор Института исследований Интернета Карен Казарян считает действия Росстандарта логичными. Однако, по его словам, для бизнеса этот стандарт ничего не меняет, скорее, он нужен как словарь для чиновников, которые не совсем представляют, что такое big data<sup>2</sup>.

В 2021 году в рамках мер по поддержке IT-отрасли начали готовить нормативную базу для создания репозитория государственной big data и правила доступа к нему разработчиков в сфере искусственного интеллекта. Ассоциация больших данных, в которую входят крупные банки, IT-компании и операторы связи до этого назвала передачу государственных данных предпринимателям как одну из задач в развитии рынка больших данных к 2024 году.

## **Отраслевые и функциональные проблемы искусственного интеллекта**

Возникает много вопросов по стандартизации понятийного аппарата искусственного интеллекта в различных отраслях экономики. В связи с этим, а также в целях реализации Федерального закона «О стандартизации в Российской Федерации»<sup>3</sup> и повышения эффективности работ по стандартизации на национальном, межгосударственном и международном уровнях и по согласованию с заинтересованными организациями, приказом

---

<sup>1</sup> Уткин Н. Там же.

<sup>2</sup> Казарян К. Там же.

<sup>3</sup> Федеральный закон от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации».

Росстандарта создан технический комитет по стандартизации «Искусственный интеллект»<sup>1</sup>. В соответствии с этим приказом за техническим комитетом закреплены объекты стандартизации с кодами ОКПД2, имеющие смежную область деятельности:

- оборудование коммуникационное (в области искусственного интеллекта);
- средства транспортные и оборудование (в области искусственного интеллекта);
- продукты программные и услуги по разработке программного обеспечения; консультационные и аналогичные услуги в области информационных технологий (в области искусственного интеллекта);
- а также многочисленные услуги:
- по изданию прочего программного обеспечения (в области искусственного интеллекта);
- в информационных технологиях (в области искусственного интеллекта);
- юридические (в области искусственного интеллекта);
- при проведении геологических, геофизических и взаимосвязанных изыскательных работ и прочих, не включенные в другие группировки (в области искусственного интеллекта);
- при техническом регулировании, стандартизации, метрологии, аккредитации, каталогизации продукции (в области искусственного интеллекта);
- в картографии (в области искусственного интеллекта);
- работы, связанные с научными исследованиями и экспериментальными разработками (в области искусственного интеллекта);
- при обеспечении безопасности и проведении расследований (в области искусственного интеллекта);

---

<sup>1</sup> Приказ Росстандарта от 1 ноября 2019 г. № 2612 «Об утверждении Программы национальной стандартизации на 2020 год»; Приказ Росстандарта от 25 июля 2019 года № 1732 «О создании технического комитета по стандартизации «Искусственный интеллект».

- в высшем и среднем образовании (в области искусственного интеллекта);
- в медицине (в области искусственного интеллекта).

Закреплённые объекты стандартизации с кодами ОКПД2 по смежной области деятельности имеют свою отраслевую операциональную специфику в определениях. Термин «операциональные определения» ввёл известный учёный по теории управления Эдвардс Деминг.

Приведенные выше примеры свидетельствуют о том, что предстоит большая и серьёзная работа по системному выстраиванию сущностных характеристик и систематизации понятийного аппарата в области искусственного интеллекта для различных сфер и областей общественной и экономической деятельности. В этой работе возможны ошибки и неудачные решения. В.Я. Цветков писал об этом: «При получении информации человек может вносить погрешности в измерения за счёт несовершенства приборов и методик измерения. Всё это приводит к тому, что в результате познания человек создаёт своё собственное информационное поле, которое следует считать искусственным»<sup>1</sup>.

## **Организационные и гуманитарные проблемы искусственного интеллекта**

Возникает много вопросов и по другим проблемам искусственного интеллекта. Примерами могут служить:

- обозначение пределов внедрения технологий искусственного интеллекта в деятельность государственных органов власти;
- проблемы доверия к системам искусственного интеллекта;
- влияние искусственного интеллекта на стратегическую стабильность, меры повышения прозрачности в киберпростран-

---

<sup>1</sup> *Цветков В.Я.* Естественное и искусственное информационное поле // Международный журнал прикладных и фундаментальных исследований. 2014. № 5. (часть 2). С. 178.

стве, современные и будущие технические разработки в сфере нераспространения и связанные с этим ядерные угрозы.

С учётом этих позиций члены РАН оценивали в своих выступлениях на осенней сессии Общего собрания РАН в 2018 году перспективы и проблемы внедрения технологий искусственного интеллекта<sup>1</sup>.

Создание в России научных центров мирового уровня в рамках национального проекта «Наука» предусматривает организацию до конца 2023 года не менее девяти научных центров, выполняющих исследования и разработки по приоритетным научно-техническим областям.

Совет по государственной поддержке создания и развития научных центров утвердил перечень из семи приоритетных направлений развития науки<sup>2</sup>. В первую очередь — это цифровые технологии и искусственный интеллект с использованием роботизированных системы на базе создания и применения материалов нового поколения.

По мнению специалистов, есть опасность, что по мере того, как искусственный интеллект будет возвышаться, естественный интеллект (в своей массе) будет опускаться. У учёных возникает вопрос: зачем тогда прогнозная аналитика про будущее, когда массовый спрос будет формироваться рефлексамии про настоящее?

Успехи в движении к искусственному интеллекту напрямую связаны с успехами в формировании языка человеко-машинных коммуникаций. Однако современные информационные системы пользуются инструментами времён «доисторического материализма»: классификаторами, словарями, тезаурусами и т. п. Их стали даже величаво называть онтологиями. Но на такой основе невозможно создание «умных» агентов и «размышляющего» искусственного интеллекта. Причина в том, что понятия словарей

---

<sup>1</sup> Любимов А.П., Щитов А.Н. Современные научно-технологические приоритеты РАН // Представительная власть — XXI век. 2018. № 7–8. С. 26–33.

<sup>2</sup> Постановление Правительства РФ от 30.04.2019 № 538 «О мерах государственной поддержки создания научных центров мирового уровня».



(семантика), не объединяемые причинно-следственными связями (синтаксис), не позволяют строить из них комбинации, несущие смысл (прагматика).

Возможно ли создание нового языка для цифровой реальности? В.В. Рубанов полагает, что эта актуальная задача хотя и суперсложна, но разрешима. Следует только подчеркнуть, что для её решения нужна не традиционная лингвистика, а относительно новая наука семиотика<sup>1</sup>. Считается, что подобный язык понятен человеку, а также легко переводим в алгоритмы для компьютерных вычислений.

*Заключение.* За IT-технологиями и искусственным интеллектом большое будущее. Ещё более могущественными они станут, если человечество и его лучшие представители объединят усилия и предложат технологии совершенствования биологического человека с большими геномными возможностями. Такое научное направление предполагает сотрудничество учёных и специалистов из различных областей: от гуманитариев до технарей; от медиков до селекционеров и т. д.<sup>2</sup>

Технологии самой значительной сложности должны быть воплощены в очень простых решениях, которые будут управляться человеческим голосом и мозгом: инновационные электронные очки и линзы с широким спектром подключенных к ним функций, умные голосовые компьютеры/планшеты, мобильные телефоны с огромными базами данных и неограниченными объёмами памяти и т. д. В недалеком будущем наши технологии станут технически живыми, то есть природоподобными! Цивилизация приближается к массовой киборгизации людей и животных, предусматривающей замену частей или всего тела кибернетическими аналогами. В этой связи на первый план выходят во-

---

<sup>1</sup> Рубанов В.В. Не подражать живым системам // Независимая газета 23.03.2020. URL: [http://www.ng.ru/stsenarii/2020-03-23/9\\_7824\\_consciousness.html](http://www.ng.ru/stsenarii/2020-03-23/9_7824_consciousness.html) (дата обращения: 23.04.2020).

<sup>2</sup> Бабкин В.В., Промоненков В.К., Овчаренко М.М., Любимов А.П. Инновационная концепция средств защиты растений в Российской Федерации // Химическая промышленность сегодня. 2017. № 8. С. 50–54.

просы культуры общества<sup>1</sup>, создания этических основ и правил правового регулирования новых общественных отношений, формирующихся в связи с применением систем искусственного интеллекта и робототехники<sup>2</sup>, введения правовых ограничений для разработчиков и пользователей технологий ИИ<sup>3</sup>, стандартизации и принятия законодательных актов как охранительных инструментов по отношению к технологиям и системам ИИ<sup>4</sup>: от национальных стратегий ИИ до принципов применения ИИ в конкретных сферах.

При этом не простым вопросом остается доверие к системам, технологиям и устройствам искусственного интеллекта, который включает в себя множество аспектов из различных областей общественных отношений.

---

<sup>1</sup> Любимов А.П., Щитова Н.Г. Нужен ли нам новый Закон о культуре? (По материалам парламентских слушаний в Государственной Думе 19 апреля 2010 г.) // Представительная власть — XXI век. М., 2010. С. 12–16.

<sup>2</sup> Гусейнов А.А. Великие моралисты. М., 1995; Любимов А.П. Философия права. М.: Юрайт, 2019. 257 с.

<sup>3</sup> Любимов А.П. Качество подготовки законопроекта субъектом права законодательной инициативы // Российская юстиция. 2006. № 2. С. 47–52.

<sup>4</sup> Любимов А.П. Конституция Российской Федерации: вопросы и проблемы толкования // Представительная власть — XXI век. 2003. № 6. С. 12–14; Любимов А.П. 20 лет Конституции РФ // Представительная власть — XXI век. 2013. № 7–8. С. 14–17.

## **АРГУМЕНТ ЛУКАСА–ПЕНРОУЗА–ГЁДЕЛЯ И ПАРАДОКСЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

---

---

Термин «искусственный интеллект» употребляется в самых разнообразных контекстах и целях. Изначально он употреблялся в связи с разговорами о возможности создания «механического» устройства, которое бы осуществляло операции, сравнимые с человеческим мозгом. Сопоставление возможностей интеллекта человека и «интеллекта» машины имеет множество оттенков, но при попытках ввести этот разговор в более строгие рамки два результата являются особо значимыми. Так называемый тест Тьюринга является операциональным способом обнаружения равных возможностей человека и машины. Более важным и теоретически сложным является исследование следствий теорем Гёделя о неполноте арифметики. Как оказалось, метаматематические результаты говорят нечто существенное о возможностях и ограничениях человеческого интеллекта.

Интерпретация теорем Гёделя сужает вопрос, ограничивая термин «мышление» «математическим мышлением», то есть заключениями, которые обладают математической достоверностью. (Именно на этом настаивал сам Гёдель). Далее, на основании ма-

---

\* *Целищев Виталий Валентинович* — доктор философских наук, профессор. Научный руководитель Института философии и права СО РАН.

*Tselishchev Vitaly Valentinovich* — Doctor of Philosophy, Professor, Scientific Director of the Institute of Philosophy and Law of the Siberian Branch of the Russian Academy of Sciences, Moscow, Russia.

тематических аргументов осуществляется выбор между двумя вариантами: либо машина эквивалентна человеку в плане интеллекта, либо человек превосходит машину. Аргументация в пользу первой точки зрения разделяется «механицистами», а второй — «менталистами». Развернувшаяся с 1960-х и продолжающаяся по настоящее время полемика представителей двух направлений была поначалу неравноправной. Механисты были представлены оксфордским философом Дж. Лукасом, опубликовавшим в 1961 г. получившую большую известность статью<sup>1</sup>, а также известным математическим физиком Р. Пенроузом, с двумя его книгами «Новый ум короля»<sup>2</sup> и «Тени разума»<sup>3</sup>. На другой стороне было практически все сообщество специалистов по математической логике, уличавших этих двух в непрофессионализме в технических деталях при интерпретации теорем Гёделя. Ситуация драматически изменилась, когда из посмертно расшифрованных записок Гёделя стало ясно, что сам он был менталистом, пополнившим эту небольшую компанию. Многие критики менталистов, узнав об этом обстоятельстве, как язвительно выразился один из участников полемики, «молчаливо отошли в сторону», что вполне понятно в силу огромного веса Гёделя. Тут же выяснилось, как сухо заметил весьма авторитетный логик С. Феферман<sup>4</sup>, что несмотря на все допущенные Пенроузом математические ошибки они не влияют на общий ход его рассуждения и не обесценивают его аргументов.

Таким образом, в последние три десятка лет в полемике менталистов с механицистами было обнаружено множество тонких следствий интерпретации гёделевских теорем, показавших реальную сложность построения искусственного интеллекта как эквивалента человеческого мышления. В данной статье представлены лишь несколько из наиболее значимых *pro et contra* аргументов, обнаруженных в ходе этой полемики, которые

---

<sup>1</sup> Lukas J. Minds, Machines and Gödel // Philosophy. Vol. 36. 1961.

<sup>2</sup> Пенроуз Р. Новый ум короля. М.: URSS, 2002.

<sup>3</sup> Пенроуз Р. Тени разума. М.: Институт компьютерных исследований, 2003.

<sup>4</sup> Feferman S. Penrose's Gödelian Argument // Psyche 2 (7), May 1995.

не позволяют поставить точку в сложном вопросе, можно ли категорически утверждать о превосходстве одного из двух направлений и, что более важно, о превосходстве человеческого интеллекта над машинным.

Существуют многочисленные интерпретации теорем Гёделя о неполноте, но наиболее интересной из них оказалась интерпретация самого Гёделя, которая получила название «дизъюнкции Гёделя». Ее формулировка крайне парадоксальная уже в том отношении, что она выводит вопрос о «состязании» человека и машины в область гораздо более тонких вещей. Она формулируется так:

Либо математика неполна (незавершаема) в том смысле, что очевидные аксиомы никогда не могут быть вмещены в конечные правила, то есть, что человеческий ум (даже в сфере чистой математики) бесконечно превосходит возможности любой конечной машины, либо существуют абсолютно неразрешимые диофантовые проблемы определенного типа.<sup>1</sup>

В менее парадоксальной формулировке, приближающей к теме данной статьи, «дизъюнкция» выглядит так: Во-первых, человеческий ум не способен к формулировке (или механизации) всех математических интуиций; если они сформулированы, появляются новые, например, о непротиворечивости. Во-вторых, либо человеческий ум превосходит все машины (более точно, может решить больше теоретико-числовых проблем), либо существуют такие теоретико-числовые проблемы, которые неразрешимы для человеческого ума.

Вопрос о том, в какой степени эти идеи Гёделя в философском отношении совпадают с идеями Лукаса или Пенроуза, является сложным уже по причине существенного расхождения их общих философских позиций. В частности, по самой неопределенной и общей классификации, Дж. Лукас является «идеалистом», по-

---

<sup>1</sup> Gödel K. Some Basic Theorems on the Foundations of Mathematics and their Implications // Gödel K. Collected Works. Vol. 3 / eds. Feferman et al. Oxford: Oxford University Press, 1995. P. 310. (Русский перевод В. Целищев: *Гёдель К. Некоторые основные теоремы в основаниях математики и их следствия* // Хинтиikka Я. О Гёделе. К. Гёдель. Статьи. М.: Канон+, 2014. С. 177.)

скольку цель его аргументации, по его собственному признанию, состоит в опровержении материализма, Р. Пенроуз — материалистом, поскольку считает, что физика может, будучи перестроена, объяснить ментальные явления, а К. Гёдель — дуалистом. Как видно, здесь выдержан полный спектр онтологических философских позиций соотношения ума и материи.

Вернемся к началу полемики, поднятой Лукасом. Популярное понимание сути теорем Гёделя о неполноте состоит в том, что Гёдель сконструировал предложение, которое истинно, но не доказуемо в достаточно богатой формальной системе. В основе рассуждения Лукаса лежит следующее соображение:

Первая теорема Гёделя должна прилагаться к кибернетическим машинам, потому что сущность их состоит в том, что они должны быть конкретизацией формальной системы. Отсюда следует, что если дана некоторая машина, непротиворечивая и способная выполнять простые арифметические операции, существует формула, которую невозможно представить для машины как истинную, то есть, формула, недоказуемая-в-системе, но такая, которую мы считаем истинной. Отсюда следует, что никакая машина не может быть полной или адекватной моделью ума, и что ум существенно отличен от машин. Другими словами, человеческий ум способен увидеть истинность гёделева неразрешимого предложения, а машина — нет.<sup>1</sup>

Несмотря на внешнюю простоту (и видимость ошибки) такой аргументации, в ней есть очень много подводных камней, которые и стали поводом для значительной полемики. Прежде всего, очень интересным вопросом является выяснение причин, по которым гёделево неразрешимое предложение признается истинным. Одно объяснение идет от самого Гёделя: предложения «Я не доказуемо» (далее — G) — и на самом деле недоказуемо и поэтому истинно, сейчас признано не объяснением, а просто эвристикой. Другое объяснение, на котором настаивал М. Даммит, объяснял истинность G тем, что речь идет о намеренной интер-

---

<sup>1</sup> *Lukas J. Minds, Machines and Gödel // Philosophy. Vol. 36. 1961. P. 44.*

претации формальной системы.<sup>1</sup> Третье объяснение взывало к расширению исходной формальной системы, в которой *G* оказывалось истинным. Есть еще «семантический аргумент» Н. Теннанта<sup>2</sup>, и другие интересные соображения о природе самореферентности. Но в данном контексте мы просто принимаем истинность *G* как установленный факт.

Лукас говорил от лица философов, и по этой причине он адресовал эту аргументацию к людям, которые «могут понять теорему Геделя», а не к профессиональным математикам. Но апелляция к популярным формулировкам может приводить к концептуальным ошибкам. Так, у Лукаса вторая теорема Геделя излагается в таком виде: «Если система *S* непротиворечива, это не может быть доказано в *S*». Х. Патнэм отмечает, что ошибка Лукаса состоит в том, что тот путает два различных утверждения, которые могли бы быть названы «утверждение, что *S* непротиворечива». Лукас путает обычное утверждение, что используемые математиками методы непротиворечивы, с очень сложным математическим утверждением, которое возникает при применении теоремы Геделя к гипотетической формализации этих методов. В этом отношении Пенроуз более аккуратен, используя не саму теорему Геделя, а теорему Тьюринга, которую он применяет ко всей мировой активности всех математиков, и говорит, что их творческая активность не может быть полностью схвачена никаким алгоритмом, никаким множеством жестких правил, которыми запрограммирована машина Тьюринга. Другими словами, с точки зрения Пенроуза, никакая машина Тьюринга не может симулировать человеческий мозг.

Общие дискуссии о возможности моделирования ума машиной слишком расплывчаты, чтобы их можно было серьезно обсуждать без ряда технических деталей. Вполне может оказаться, что вопрос не является правильно поставленным

---

<sup>1</sup> *Dummett M.* The Philosophical Significance of Gödel's Theorem // *Truth and Other Enigmas*. Harvard: Harvard University Press, 1995.

<sup>2</sup> *Tennant N.* Deflationism and Gödel Phenomena // *Mind*. Vol. 111. № 443, 2002. P. 551–582.

и не имеет вполне определенного ответа. Подобный скепсис имеет под собой основание, поскольку даже в отношении собственно «гёделевской версии» гёделевского аргумента есть вполне обоснованные сомнения. Так, Дж. Булос утверждает следующее: «Имеется пропасть между утверждением, что никакая конечная машина, с определенными слабыми условиями на нее, не может напечатать определенное формальное предложение (которое зависит от машины), и утверждением, что если ум есть конечная машина, тогда существуют истины, которые не могут быть установлены никаким доказательством, которое может воспринять человеческий ум. Весьма неясно, что означает, что человеческий ум есть конечная машина, то есть машина Тьюринга. И сказать, что ум (по крайней мере, в отношении доказательства теорем), может быть представлен машиной Тьюринга, значит оставить полностью открытым вопрос, как именно он так представлен».<sup>1</sup>

Но как бы то ни было, начиная со статьи Лукаса, полемика началась, и значительная часть ее заключалась в том, в какой степени тезис Лукаса о преимуществе человеческого ума над машиной можно сделать предметом точной аргументации. Собственно стратегия Лукаса состояла в том, чтобы использовать для опровержения механистической гипотезы метод доказательства от противного: предположение о том, что человеческий ум представлен машиной Тьюринга, приводит к противоречию. Действительно, ум при таком предположении был бы сравним с формальной системой, а его выход — с теоремами. Формальная система должна включать формальную систему Арифметики Пеано, где формулируется гёделёво неразрешимое предложение, которое, тем не менее, может рассматриваться как истинное компетентным математиком, понимающим доказательство Гёделя. Отсюда, никакое представление человеческого ума машиной Тьюринга не может быть правильным, так как для любого такого представления будет гёделёво предложение, которое машина

---

<sup>1</sup> Boolos G. Introductory Note // Gödel K. Collected Works. Vol. 3 / eds. Feferman et al. Oxford: Oxford University Press, 1995. P. 293.



не может доказать, и таким образом, представить как истинное, но которое математик видит истинным.

Одна из интересных реконструкций аргументация Лукаса может быть представлена следующим образом<sup>1</sup>:

1. «Механизм» представляет собой тезис, что все человеческие существа, в частности, все хорошие логики, являются машинами.

2. Любая машина, по словам Лукаса, есть «конкретное проявление формальной системы».

3. Поэтому если некоторый хороший логик Альф есть машина, некоторая формальная система  $S$  представлена Альфом.

4. Но теорема Гёделя показывает, что для любой непротиворечивой формальной системы, достаточно сильной для того, чтобы содержать элементарную арифметику, имеется правильно построенное предложение системы, которое при вполне защищаемых предположениях любой хороший логик признает истинным, но которое не доказуемо в системе. Хороший логик может обнаружить истинность этого предложения, если мы предположим, что хорошие логики знают, что элементарная арифметика непротиворечива, и может утверждать, что данная система непротиворечива, если непротиворечива арифметика.

5. Поэтому для любой формальной системы  $X$ , которая одновременно непротиворечива и достаточно сильна для того, чтобы вместить элементарную арифметику, имеется гёделево предложение, которое не доказуемо в  $X$ , но которое Альф, как хороший логик, может рассматривать как истинное.

6. Но если бы Альф был машиной, представляющей формальную систему  $S$ , которая одновременно непротиворечива и достаточно сильна для того, чтобы вместить элементарную арифметику, он не был бы способен обнаружить истинность гёделева предложения системы  $S$ , потому что оно не было бы доказуемо в  $S$ .

7. Следовательно, Альф не машина, и «Механизм» ложен.

---

<sup>1</sup> Kirk R. Mental Machinery and Gödel // Synthese. Vol. 66. 1986. P. 437–452.

Ясно, что предыдущие строки полны неясностей, — что имеется в виду под «выводом», под «доказательством», и т. д. Сам Лукас полагает, что понятие доказательства отнюдь не исчерпывается тем, что называется формальным доказательством. Для этого у него есть резоны такого рода. Доказательство есть способ избегания противоречия, и если человек утверждает нечто, он претендует на собственную непротиворечивость. Однако он не может доказать своей собственной непротиворечивости, и этот факт Лукас полагает следствием второй теоремы Геделя о неполноте. Однако он отмечает, что теорема говорит о невозможности доказательства непротиворечивости системы в самой системе. Таким образом, компьютер, если он непротиворечив, не может произвести в качестве истинного утверждение о своей собственной непротиворечивости. Но если человеческий ум есть машина, тогда и человек не может дать формального доказательства непротиворечивости той формальной системы, которая «схватывает» его ум. Однако человек может выйти за пределы формальной системы, и тогда не может быть препятствий для «менее формальных и менее систематизированных» доказательств непротиворечивости. Такие неформальные аргументы не могут быть полностью формализованы, но нельзя их по этой самой причине полагать менее ценными. Больше того, человек вправе полагать, что он знает о своей собственной непротиворечивости, и далее, он просто обязан предполагать эту самую непротиворечивость. Другими словами, правильность, или непротиворечивость, есть неперенная посылка нашего мышления. Мы, в некотором смысле, решаем быть непротиворечивыми. Таким образом, непротиворечивость является нормой рационального мышления, но нам надо для реализации такой нормы допустить ценность неформального вывода, или неформального доказательства. Вот с таким видом «доказательства» имеет дело аргумент Лукаса.

В ответ на критику своего аргумента Лукас неоднократно жаловался, что именно излишняя формализация аргумента выявляет предполагаемые огрехи. Больше того, именно стараниями математических логиков за аргументом закрепилась репутация «ложного, но интересного». Между тем, уже было сказано ранее, изначально

мотив Лукаса в интерпретации теорем Гёделя о неполноте состоял в неприятии Лукасом крайнего материализма, в частности, в защите им философской аргументации о свободе воли, а такая аргументация сторонится строгого размышления. Как известно, свободой воли обладает человек, а лишены ее механизм, машина, компьютер, которые жестко детерминированы в своем поведении. Так что попытки формальной реконструкции своей аргументации Лукас не признает убедительными. В этой связи Х. Патнэм предложил такую реконструкцию аргумента Лукаса<sup>1</sup>.

Основа аргументации состоит в том, что субъект «видит» истинность гёделева предложения. Посылка Лукаса состоит в том, что этим субъектом является человек, но никак не машина. Пусть «Видение Я» означает «Я (человек) могу видеть в качестве истинного предложение...», и пусть «Видение Ф» означает «Ф (машина Тьюринга) может произвести в качестве истинного предложение...». Тогда аргумент Лукаса может представлять следующий логический вывод:

1. Если Ф непротиворечиво, тогда неверно, что машина эквивалентна человеку ( $\Phi = Я$ ).
2. Имеется «Видение Я» (Если Ф непротиворечиво, тогда G).
3. Если Ф непротиворечиво, тогда имеется «Видение» Я (G).
4. Если Ф непротиворечиво, тогда неверно, что имеется «Видение» Ф (G).
5. Если Ф непротиворечиво, тогда неверно, что  $\Phi = Я$ .

Патнэм полагает, что данный логический вывод не является корректным по той причине, что переход от 2 к 3 незаконен. Действительно, можно ли из посылки о том, что если Ф непротиворечиво, тогда G истинно, вывести заключение, что если Ф непротиворечиво, тогда человек может видеть истинность G? Конечно, говорить о строгой логической формализации аргумента не приходится, потому что самый важный ингредиент «Видение» не регламентировано строгими логическими правилами употребления. Как видно, позиция Лукаса, который смешивает

---

<sup>1</sup> *Puntam H. Minds and Machines // Minds and Machines / ed. Anderson A. N.Y.: Prentice Hall, 1964.*

вает формальный и неформальный дискурсы в некую «диалогическую» форму, полна тонких моментов, которые скорее запутывают ситуацию, нежели ее проясняют.

Вступление в спор Р. Пенроуза делает вопрос гораздо более четким. Характерной особенностью подхода Пенроуза является упор не на полноте или неполноте, а на обоснованности. Исходная постановка проблемы состоит в том, что человек «видит» истинность геделева предложения в системе  $\Phi$ , и в то же время, согласно теореме Геделя, система  $\Phi$  не может быть использована для постижения истинности  $G$ , если человек (математик) знает, что формальная система  $\Phi$  обоснована. Обоснованность системы состоит в том, что все доказуемые утверждения являются истинными. С логической точки зрения обоснованность системы равносильна ее непротиворечивости. Но для реализации менталистской позиции требуется нечто большее, а именно привязывание понятия обоснованности к платонистской философии математики.

Суть аргументации Пенроуза такова: обоснованность предполагает доступ математика к истинности математических утверждений. Это требует понимания того, о чем собственно говорит формальная система  $\Phi$ , то есть требует знания намеренной интерпретации формальной системы. Именно благодаря такому доступу мы знаем, что аксиомы системы истинны и правила ее ведут от истинного утверждения к истинному утверждению. Непосредственный контакт с платонистской реальностью у Пенроуза вполне созвучен со знаменитой концепцией Гёделя о том, что математическая интуиция в познании математических объектов аналогична чувственным ощущениям в познании материальных объектов. Именно такой контакт с внечувственными объектами позволяет видеть истинность предложения  $G$ .

Обоснованность формальной системы эквивалентна ее непротиворечивости. Но согласно второй теореме Геделя о неполноте, в рамках формальной системы ее собственная непротиворечивость не может быть доказана. Психологически это не такой уж удивительный факт. В конце концов, формальная

система может быть настолько сложной, что установление ее непротиворечивости может представлять просто невыполнимую задачу. Если же предположить, что человеческий ум при всей его сложности есть машина, тогда нет особых причин надеяться на доказательство непротиворечивости такой машины.

Но тогда трудно говорить о том, что человек может видеть истинность гёделева предложения. Единственный способ избежать такого заключения состоит в том, чтобы считать, что человеческий ум является обоснованным, но мы не можем знать об этом. Вот только попытка представить человеческий ум как обоснованную систему является неверным шагом. Каковы аргументы Пенроуза в пользу того, что человеческий ум обоснован, и что ум не может быть схвачен обоснованной формальной системой, чьей обоснованности в этой системе нельзя доказать? Д. Чалмерс полагает, что Пенроуз выдвигает аргумент, справедливость которого зависит от того, знаем ли мы, что  $\Phi$  есть формальная система, которая схватывает наше мышление.<sup>1</sup> Если мы могли бы знать, что формальная система  $\Phi$  схватывает наше мышление, тогда первый аргумент выглядел бы так:

- 1) Мы знаем, что мы обоснованы;
- 2) Мы знаем, что  $\Phi$  схватывает наше мышление; следовательно, мы знаем, что  $\Phi$  обосновано.

Обе посылки не выглядят особо правдоподобными. Но если выбирать из них более слабую, то, с точки зрения Пенроуза, посылка (2) весьма сомнительна. Даже если наше мышление схватывается  $\Phi$ , тогда вряд ли мы можем знать этот факт. Очень важным обстоятельством в данном рассуждении является то, что в попытках разобраться, можем мы ли установить факт схватывания некоторой формальной системой  $\Phi$  нашего мышления, мы не можем прибегать к помощи посторонних средств типа эмпирических или психологических исследований. Аргументация должна быть чисто теоретической, не выходящей за пределы формальной системы. Именно в такой стратегии и заключается «гёделевский аргумент». В конечном счете, и сам Гёдель

---

<sup>1</sup> *Chalmers D. Minds, Machines, and Mathematics // Psyche, 2 (9), June 1995.*

полагал, что установление обоснованности человеческого ума как машины может быть проведено и эмпирическими средствами, но тогда это будет уже другой, «не-геделевский аргумент».

Итак, говорит Пенроуз, «мы пытаемся серьезно рассмотреть возможность того, что математическое понимание на деле эквивалентно некоторому сознательно познаваемому алгоритму, однако эквивалентность эта принципиально непознаваема».<sup>1</sup> Другими словами, мы не можем убедиться в том, что некоторая конкретная, известная нам, формальная система лежит в основе математического понимания. Тот факт, что математическое понимание основывается именно на конкретной алгоритмической процедуре, остается неосознаваемым, так и непознаваемым. Если бы такое познание было возможным, то пришлось бы признать и обоснованность формальной системы  $\Phi$ . Действительно, если не допустить обоснованность  $\Phi$ , тогда придется не допустить и следствия  $\Phi$ , в число которых входят уже признанные истинными предложения. Таким образом, для выполнения своей задачи Пенроуз должен аргументировать, что если мы обоснованная формальная система  $\Phi$ , тогда мы должны были бы определить, что  $\Phi$  обоснована, независимо от всякого знания о том, что мы  $\Phi$ . То есть, он должен сделать так, чтобы если  $\Phi$  представлено нам, мы могли бы определить ее обоснованность через анализ самой лишь  $\Phi$ . Это бремя Пенроуз берет на себя в разделе 3.3 («Тени разума»). Именно этот раздел несет на себе основной вес аргументации. Если Пенроуз в нем не преуспевает, тогда рушится вся аргументация.<sup>2</sup>

Каков в этом случае ответ Пенроуза? Предположим, что существует познаваемая процедура  $\Phi$ , непознаваемым образом эквивалентная человеческому математическому пониманию. Мы можем рассматривать  $\Phi$  как систему аксиом и правил. Ясно, что мы можем постичь с помощью математического понимания и интуиции, что каждая из аксиом неопровержимо истинна.

---

<sup>1</sup> Пенроуз Р. Тени разума. М.: Институт компьютерных исследований, 2003. С. 212.

<sup>2</sup> Chalmers D. Minds, Machines, and Mathematics // Psyche, 2 (9), June 1995.

То же относится и к правилам вывода, потому что в высшей степени сомнительно, чтобы мы опирались на такие правила вывода, которые были бы «фундаментально неясны». В результате мы можем заключить, что система  $\Phi$  обоснована. Отсюда следует, что система непротиворечива. В силу этой непротиворечивости утверждение  $G$  в  $\Phi$  должно быть не просто истинным, а неопровержимо истинным. А теперь следует важное соображение: поскольку система  $\Phi$  эквивалентна мышлению человека, она включает в себя всю совокупность того, что доступно нашему пониманию. Но это, в свою очередь, означает, опять-таки в силу того, что предполагается эквивалентность ума и формальной системы, что утверждение  $G$  в  $\Phi$  представляет собой теорему системы  $\Phi$ . Но такое возможно только в случае противоречивости системы, следствием чего будет принятие человеком в качестве истины таких утверждений как « $1 = 2$ ». Полученное противоречие должно быть интерпретировано как опровержение посылки аргумента, что существует познаваемая процедура  $\Phi$ , непознаваемым образом эквивалентная человеческому математическому пониманию.

Этот интересный аргумент основан на ряде посылок, некоторые из которых представляются весьма спорными. Анализ Пенроуза ограничивается  $P_1$ -предложениями, хотя формальная система, симулирующая мозг, может выглядеть гораздо более сложным образом. Больше того, она может быть такой сложности, что трудно будет разложить ее на аксиомы и правила вывода, а это разложение весьма существенно для того, чтобы установить обоснованность формальной системы.

В частности, «правила вывода» могут быть столь же сложными, как это имеет место в человеческом мозге. В этом отношении наиболее «продвинутой» является гипотеза коннекционизма, согласно которой мыслительные операции идут параллельным образом, так что выделение правил вывода оказывается под вопросом. В этом случае будет невозможно убедиться в корректности правил. Дело не в том, что эти правила могут привести к неправильным результатам в мышлении, что, правда, и случается часто, а в том, что процедуры, управляющие динамикой нашего мозга, слишком сложны для нас, чтобы мы могли полагать их обоснованными.

В целом, аргументация Пенроуза сводится к тому, что человеческое мышление находится за пределами вычислений. Поэтому никакая компьютерная программа не может быть столь же разумной, как человеческое существо. Пенроуз предполагает, что существует компьютерная программа, столь же разумная, как и человек, и показывает, что это ведет к противоречию.

Вот как это выглядит в сводном виде:

1. Противоречие, о котором говорит Пенроуз, состоит в следующем: предположим, что мыслительные способности некоторого математика  $X$  полностью описываются некоторой формальной системой  $\Phi$ . Это означает, что для каждого математического утверждения  $S$  в языке  $\Phi$ , которое  $X$  полагает «неопровержимо истинным»,  $S$  есть теорема  $\Phi$ , и наоборот. Далее мы предполагаем, что  $X$  знает, что  $\Phi$  описывает его мышление.

2. Согласно  $X$ , вера, что  $\Phi$  описывает его мышление, влечет веру в обоснованность  $\Phi$ .  $X$  оправдывает это предположение так: «С математической точки зрения было бы неразумно позволять себе не верить в то, что представляет основания нашей неопровержимой системы вер».

3. По теореме Гёделя, так как  $\Phi$  обоснована,  $G$  в  $\Phi$ , гёделево утверждения для  $\Phi$ , должно быть истинно, но не должно быть теоремой  $\Phi$ . Следовательно, так как  $X$  верит, что  $\Phi$  обоснована, он должен заключить, что  $G$  в  $\Phi$  «неопровержимо истинно». Поэтому имеется нечто (а именно,  $G$  в  $\Phi$ ), что  $X$  полагает неопровержимо истинным, но что не является теоремой  $\Phi$ . Это противоречит предположению, что  $\Phi$  полностью описывает мыслительные силы  $X$  (включая его знание, что  $\Phi$  имеет это свойство).

Аргументация Пенроуза представляет собой сложную смесь различных дискурсов. Естественно, это сопровождается неоднозначностью трактовки многих ключевых терминов, в частности, неоднозначным употреблением термина «формальная система  $\Phi$ ». Пенроуз подразумевает по крайней мере три интерпретации термина «формальная система  $\Phi$ ». Одна интерпретация  $\Phi$  состоит в том, что  $\Phi$  представляет врожденную (внутренне присущую) способность к размышлению самого математика. Это чисто иннативистская характеристика математи-



ческого мышления, и именно к ней относится уверенность в истинности некоторых математических истин, которые «истинны неопровержимо». Эта интерпретация позволяет объявить внутреннюю непротиворечивость человеческого ума. Другая интерпретация включает в себя, помимо врожденной способности к математическим заключениям с математической определенностью, еще и эмпирический опыт математика, со всеми эвристическими и индуктивными приемами получения знания. Третья возможность состоит в том, что  $\Phi$  представляет пределы того, что могло бы быть известно математику, независимо от того, как приобретается это знание, через размышление или эмпирически. Различия между этими интерпретациями приобретают значение, когда решается вопрос о том, знает ли математик, что его мышление описывается посредством  $\Phi$ . Формальная система  $\Phi$  является дедуктивной структурой, и полученные математиком знания о роли  $\Phi$  не отражены в самой  $\Phi$ . Пенроуз отдает себе отчет в слабости своей аргументации в этом отношении, и поэтому вводит в рассмотрение так называемый «новый аргумент». В нем рассматривается новая система  $\Phi^*$ , которая включает в себя  $\Phi$  плюс все, что следует из информации, что способности мышления были описаны в  $\Phi$ . Этот новый аргумент более тонок, и его не так просто опровергнуть. Он будет рассмотрен позднее.

Следствием неоднозначности термина «формальная система» является безапелляционное утверждение Пенроуза о том, что неопровержимо истинным считается утверждение об обоснованности математического мышления. Потому что «неопровержимая истинность» есть истинность иннативистского толка. Другими словами, следует полагать, что математическое мышление человека обосновано. Таким образом, мы впадаем в порочный круг, когда обоснованность объясняется в терминах внутренней непротиворечивости человеческого ума, а непротиворечивость, будучи эквивалентной обоснованности, объясняется в терминах обоснованности математического мышления. Именно это обстоятельство, судя по всему, явилось причиной появления у Гёделя фразы «доказательство с математической определенностью», ибо чем еще может быть математическая определенность,

как неуверенностью во внутренней непротиворечивости человеческого ума.

При обсуждении вопроса об обоснованности математического мышления поднимается труднейший эпистемологический вопрос об ошибочности мышления и его причинах. Человек может по тем или иным причинам полагать некоторое утверждение «неопровержимым», когда оно по ряду критериев таковым не является. МакКаллох полагает, что это вопрос даже не эпистемологический, а психологический. Поэтому аргумент Геделя не доказывает, что человеческое мышление должно быть невычислимым — она лишь доказывает, что если человеческое мышление вычислимо, тогда оно должно быть либо необоснованным, или же для нас принципиально (*inherently*) невозможно знать одновременно, каковы наши мыслительные способности, и являются ли они обоснованными. Пенроуз отмечает возможность того, что мы знаем наши мыслительные силы, но не знаем, обоснованы ли они (раздел 3.2. «Тени разума»). Пенроуз говорит, что если мы знаем, что некоторая конкретная компьютерная программа  $\Phi$  эквивалентна человеческому мышлению, тогда мы были бы вынуждены заключить, что  $\Phi$  обосновано... Для меня этот вопрос является скорее делом психологии, чем математики. Пенроуз полагает некоторые свои веры относительно математики «неопровержимо истинными», и даже не рассматривает возможность того, что некоторые веры могут быть ошибочными. Если он придерживается этой веры, то вовсе не следует, что мышление Пенроуза не вычислимо. Отсюда следует лишь то, что Пенроуз никогда не был убежден в том, что оно вычислимо. Для людей вроде меня, кто имеет более гибкую позицию в отношении непогрешимости своей веры, то есть в отношении того, что их мышление может быть неосновательным, аргумент Пенроуза не очень весом.<sup>1</sup>

В конечном счете мы действительно имеем полупсихологическую уверенность математика в истинности некоторого утверждения в качестве единственного критерия того, что это утверждение «неопровержимо истинно». Дело в том, что

---

<sup>1</sup> *McCullough D.* Can Humans Escape Gödel? // *Psyche*, 2 (4), April 1995.

«неопровержимость» является весьма произвольным критерием. Например, могут существовать такие сложные истины, которые доказаны с математической определенностью, то есть, всеми доступными математику надежными средствами, но которые трудно считать интуитивно истинными. Будут ли такие истины неопровержимыми? Далее, вполне должен быть случай, что ничего ложного не может быть «неопровержимой истиной». Однако может быть показано, что даже если допустить, что ничего ложного не может быть признано «неопровержимой истиной», этот факт не может быть «неопровержимой истиной». Можно показать, что понятие «неопровержимой истины» не может само быть неопровержимым.

Дискуссия вокруг механицистского тезиса исходит из предположения, что имеется множество всех аналогов арифметических теорем и только их, которые известны неопровержимо, или, с математической определенностью. Это предложения, сформулированные на языке первого порядка. Предполагается, что идеализированный математик никогда не утверждает ложные математические предложения. Коль скоро речь идет об идеализированном математике, такие истинные предложения являются познаваемыми или доказуемыми арифметическими предложениями. В данном случае мы имеем две интерпретации одного и того же оператора формальной системы: доказуемость и познаваемость.

Проблема демаркации познаваемых математических истин и всех математических истин поднимает множество вопросов о природе математических утверждений. Особенно в этом отношении характерен крайний платонизм Гёделя, с точки зрения которого существует сфера платонистских истин, которые превосходят все, что в принципе может быть известно человеку. Сам Гёдель в этом контексте говорит об «объективной математике» и «субъективной математике». Каково же соотношение этих двух математик? Ясно, что познаваемость математических истин связана с некоторыми эпистемологическими процедурами «доступа» к ним, и в первую очередь, это математическое доказательство. Для точного представления такого рода доступа изобретается формальное доказательство, которое, в свою очередь, является

частью формальной аксиоматической системы. Ясно, что такая формальная система должна «схватывать» интуитивное содержание математических утверждений. Интуитивно мы полагаем, что содержательные математические утверждения не противоречат друг другу, и стало быть, адекватность формальной системы для представления содержательного математического знания заключается в непротиворечивости формальной системы.

Адекватность формальной системы означает, что интуитивно истинные предложения математики должны быть доказуемы, и доказуемые утверждения — истинными. Как известно, для достаточно богатых формальных систем арифметики такого рода адекватности добиться невозможно — согласно первой теореме Геделя о неполноте существуют интуитивно истинные, но не доказуемые в формальной системе утверждения. К таким утверждениям, согласно второй теореме Геделя о неполноте, относится и утверждение о непротиворечивости формальной системы. Таким образом, не все математические истины представимы доказуемыми утверждениями в формальной системе, и эта принципиальная неполнота формальных систем является важным обстоятельством.

Обозначим множество истин «субъективной» математики через  $K$  (далее мы будем называть их истинами «человеческой», в противоположность «платонистской» математики). Далее, пусть множество всех математических (платонистских) истин, выраженных в языке первого порядка, будет  $T$ . Интуитивно, исходя уже из этимологии терминов, ясно, что  $K \subseteq T$ . Рассмотрим гипотетическую возможность  $K = T$ , которая означает, что все математические истины рано или поздно будут известны человеку. Согласно теореме Тарского,  $T$  не определимо в языке арифметики. Это значит, что  $T$  не является рекурсивно перечислимым. Если множество  $K$  совпадает с множеством  $T$ , тогда  $K$  также не является рекурсивно перечислимым. Но в отношении познаваемых истин это была бы странная позиция, которую можно было бы понять только следующим образом. Человек обладает такими когнитивными способностями, которые позволяют ему «превосходить» в своем познании множество рекурсивно

перечислимых истин. Это означает, что человек превосходит по своим возможностям чисто механическое накопление истин, которое и делает возможным рекурсивно перечислимый характер множества познаваемых утверждений. Таким образом, для сторонников «механизма», доктрины, согласно которой мышление человека успешно моделируется компьютером, или машиной Тьюринга, неверно, что  $K = T$ . Тогда в множестве  $T$  остаются такие истины, которые не познаваемы в принципе.

Вернемся к упомянутому ранее «новому» аргументу Пенроуза, который учитывает, что при сопоставлении когнитивных способностей человека и компьютера нужно исходить из некоторых математических утверждений, которые являются базисными для обеих категорий когнитивных созданий. Можно вести речь о базисе человеческой математики, реализованном в аксиоматике. Компьютер может рассматривать эти аксиомы как «неопровержимые» математические утверждения (они могут быть «сообщены компьютеру в качестве таковых человеком»). Больше того, программа компьютера может быть устроена так, что компьютер может полагать эти утверждения неопровержимо истинными согласно его собственным критериям.

Противопоставление человека и компьютера опирается на спор вокруг тезиса, опирается ли математическое понимание на набор механизмов, скажем, некоторых механизмов  $\Phi$ . В качестве таких механизмов можно указать алгоритмы. Предполагается, что если в случае человека такой набор механизмов представляется крайне проблематичным, то уж в случае компьютера (или робота, как у Пенроуза) «мышление» его управляется таким набором механизмов. Так называемый «новый» аргумент Пенроуза состоит в том, что даже для компьютера это неверно. Точнее, даже компьютер будет вынужден отвергнуть возможность того, что его математическое понимание опирается на набор механизмов  $M$ , независимо от того, как обстоит дело в действительности.

Мы имеем здесь весьма сильный аргумент против «механистов». В определенном отношении мы имеем дело с тактикой гамбита: сначала мы даем преимущество «механистам», гово-

ря о том, что компьютер обладает достаточным «сознанием» для того, чтобы иметь веру в неопровержимые математические утверждения. Но это преимущество «противника» быстро рассеивается, потому что Пенроуз показывает (или пытается показать), что компьютер может поверить в то, что управляется набором механизмов только ценой противоречия, что недопустимо для компьютера. Таким образом, «механизм» терпит двойное крушение, потому что даже при «ослаблении» ему мы имеем противоречие.

Ряд исследователей считают этот аргумент весьма изобретательным и тонким, требующим тщательного анализа. Суть этого аргумента состоит в том, что человек (идеализированный, никогда не ошибающийся математик) или компьютер не может непротиворечиво верить, что некоторый данный алгоритм перечисляет доказуемые арифметические предложения (или даже  $\Pi_1$  — утверждения). Если же предположить, что в основе постижения математических истин лежит некоторый алгоритм, то оказывается, что мы не можем описать его. Поскольку мы не можем писать его, мы не можем утверждать, что этот алгоритм перечисляет  $K$ , множество арифметических познаваемых истин. Если мы предположим, что  $K$  является перечислимым, мы немедленно должны отвергнуть гипотезу, что в основе постижения математических утверждений лежит некоторый алгоритм.

С. Шапиро предлагает такую экспликацию «нового» аргумента Р. Пенроуза.<sup>1</sup> Предположим, что множество познаваемых или доказуемых математических утверждений  $K$  рекурсивно перечислимо. Это означает, что имеется некоторая машина Тьюринга, которая перенумеровывает  $K$ . Обозначим геделево число этой машины Тьюринга через  $e$ , и множество перечисляемых при этом утверждений через  $W_e$ . Тогда  $K = W_e$ . Далее, пусть  $K'$  будет множеством арифметических предложений, относительно которых идеализированный математик или компьютер неопровержимо знает, что они следуют из гипотезы, что  $K = W_e$ .

---

<sup>1</sup> Shapiro S. Mechanism, Truth, and Penrose's New Argument // Journal of Philosophical Logic. Vol. 32. 2003. P. 19–42.

Структура аргумента Пенроуза такова (Пенроуз воспроизводит возможную аргументацию компьютера или идеализированного математика):

(1) Предположим, что  $K = W_e$ .

(2) Тогда каждый член  $K'$  истинен, и таким образом,  $K'$  непротиворечиво (это следует из утверждения об обоснованности «понимания» компьютера).

(3) Если справедливо (1), тогда  $K'$  рекурсивно перечислимо и идеализированный субъект (компьютер или идеализированный математик) может написать геделево предложение  $G'$  для  $K'$  (используя  $e$ ).

(4) Поэтому, по (1), если  $K'$  непротиворечиво, тогда  $G'$  истинно, но не в  $K'$ .

(5) Но из (1) и (2) следует, что  $K'$  непротиворечиво. Поэтому  $G'$  истинно, но не в  $K'$ .

(6) Поэтому мы знаем, что если  $K = W_e$ , тогда  $G'$ . Из утверждений (1) и (5) следует  $G'$ , и поэтому можно избавиться от предположения-гипотезы (1). Таким образом,  $G'$  содержится в  $K'$  (по определению  $K'$ ).

(7) Также, если  $K = W_e$ , тогда  $G'$  не есть в  $K'$ . Это следует из предложений (1) и (5), что позволяет избавиться от (1).

(8) Таким образом, из предложений (6) и (7) следует  $K \neq W_e$ .

Обсуждаемый «новый» аргумент, в свою очередь, имеет еще одну версию (первая изложена в разделе 3.16, а вторая — в фантастическом диалоге робота и человека в разделе 3.23.). Как свидетельствует Д. Чалмерс, «насколько я могу определить, этот аргумент свободен от явных недостатков, которые поражают геделевы аргументы, такой как аргумент Лукаса и ранний аргумент Пенроуза. Если он и имеет недостатки, то они лежат более глубоко. Создается впечатление, что заключение аргумента получается, как по волшебству. И хотя имеются различные направления, по которым можно критиковать аргумент, нет никакого сногсшибательного ответа. По этой причине это настоящий вызов сторонникам искусственного интеллекта».<sup>1</sup>

---

<sup>1</sup> *Chalmers D. Minds, Machines, and Mathematics // Psyche, 2 (9), June 1995.*

Размер данной статьи не позволяет изложить детали этой аргументации Пенроуза. Отметим только, что помимо самого Д. Чалмерса, интересную формализацию его представил П. Линдстрем.<sup>1</sup>

Таким образом, в данной статье изложены лишь краткие сведения о полемике механицистов и менталистов. «За кадром» остались многие интересные вопросы, например, проблема «видения» истинности гёделева предложения и человеком и компьютером, вопрос о том, что имеется в виду под «знанием» компьютера, исследование т. н. принципов рефлексии, что происходит при трансфинитном расширении формальной системы и т. п.<sup>2</sup> Отличие Гёделем «человеческой» или «субъективной» математики от «объективной» является фактически философским обобщением проблемы, в каком смысле человеческое мышление имеет «некомпьютерный» характер. Но это вовсе не ставит точку в споре механицистов и менталистов, и можно считать, что в «игре» человека с компьютером с точки зрения строгого логического анализа пока фиксируется ничья.

---

<sup>1</sup> *Lindström P.* Penrose's New Argument // *Journal of Philosophical Logic*. Vol. 30/ 2001. P. 241–250.

<sup>2</sup> Систематическое рассмотрение этих вопросов представлено в работе: *Целищев В.В.* «Алгоритмизация мышления: гёделевский аргумент» (1-е издание. Новосибирск, Параллель, 2005; 2-е издание. М.: URSS, 2021.



## 70 ЛЕТ ТЕСТУ ТЬЮРИНГА: МОЖЕТ ЛИ КОМПЬЮТЕР ВСЁ?

---

---

*Введение.* Недавно, в 2020 году, исследователи искусственного интеллекта отметили два важных события. Первый юбилей посвящался 70-летию теста Тьюринга: в 1950 году в философском журнале “Mind” была опубликована небольшая статья Алана Тьюринга<sup>1</sup>, переведенная на русский язык под названием «Вычислительные машины и интеллект»<sup>2</sup>. В самом начале этой статьи прозвучал, пожалуй, основной вопрос философии искусственного интеллекта «Могут ли машины мыслить?». Был критически аргументирован положительный ответ на этот вопрос. Так же был адресован прямо к сегодняшнему дню прогноз о том, что человек и машина (система искусственного интеллекта) станут равноправными партнерами. Этот прогноз по сути реализовался, с погрешностью в одно-два десятилетия. Но не подтвердился наивный тьюринговский оптимизм о равноправности

---

\* *Алексеев Андрей Юрьевич* — доктор философских наук, профессор, ведущий научный сотрудник кафедры философии и методологии науки философского факультета МГУ имени М.В. Ломоносова.

*Alekseyev Andrey Yurievich* — Doctor of Philosophy, Professor, Leading researcher of the Department of Philosophy and Methodology of Science of the Faculty of Philosophy of Lomonosov Moscow State University, Moscow, Russia.

<sup>2</sup> *Turing, A.* (1950), Computing Machinery and Intelligence. *Mind* 59 (236). P. 433–460.

<sup>3</sup> *Алан М. Тьюринг.* Может ли машина мыслить? / Под ред. Б.В. Бирюкова. М., 1960. С. 19–158.

человеко-машинного партнерства: сегодня системы искусственного интеллекта достаточно успешно стали вытеснять человека из исконно человеческих, творческих сфер деятельности: из медицины, журналистики, образования, науки.

Принято считать, что в юбилейной статье А. Тьюринга обозначен тест на компьютерную имитацию человеческого интеллекта. Это не совсем так. Буквальный смысл статьи состоит в доказательстве возможностей имитации компьютером человека, который пытается подтвердить или опровергнуть свою половую идентичность. При чем здесь имитация интеллекта? Этот вопрос резко поставила С.А. Яновская в предисловии к русскому переводу статьи<sup>1</sup>. Так же и другой исследователь, Дж. Лассаж, прямо утверждал, что А. Тьюринг стал известен за то, что никогда не предлагал: тест Тьюринга на искусственный интеллект — это вовсе не тест, и все то, что утверждается в статье, никакого отношения к интеллекту не имеет<sup>2</sup>.

На наш взгляд, А. Тьюринг просто-напросто ошибся, отождествив имитацию гендера с имитацией мышления. Подобного рода «ошибки» уже имели место в истории вычислительной техники. Например, первая программа была «разработана» Адой Лавлейс путем исправления типографской ошибки, допущенной при переводе на английский язык статьи об аналитической машине Ч. Бэббиджа, представленной на французском языке итальянским инженером Л. Монабреа<sup>3</sup>. Значимость таких «ошибок» для развития культуры иногда бывает колоссальной. Они требуют герменевтической обработки, например, понимания того, что автор не мог оценить значимость своей работы, а подлинный смысл её распознается интерпретаторами, причем в неизвестном будущем.

---

<sup>1</sup> Яновская С.А. Предисловие к русскому переводу // Там же. С. 5.

<sup>2</sup> Lassegue, J. 1988, What Kind of Turing Test did Turing Have in Mind? *Tekhnema* 3. P. 37–58. URL: <http://tekhnama.free.fr/3Lasseguearticle.htm>

<sup>3</sup> Гутнер Р.С., Полунов Ю.Л. Августа Ада Лавлейс и возникновение программирования // Кибернетика и логика. Математико-логические аспекты становления идей кибернетики и развития вычислительной техники. М.: Наука, 1978. 334 с. С. 57–101.

Поэтому мы предлагаем рассматривать тест Тьюринга как собирательное понятие для различных версий и версий этих и других версий этой оригинальной тьюринговой игры в имитацию пола игрока. Сегодня тест Тьюринга охватывает чрезвычайно широкие горизонты когнитивной феноменологии. Может ли компьютер мыслить, жить, осознавать, самосознавать, страдать, понимать, любить, творить, быть личностью, обществом? Эти и многие другие мировоззренческие вопросы поднимаются исследователями теста Тьюринга — инженерами, математиками, программистами, психологами, филологами, социологами, культурологами, биологами, врачами, учителями и др. Они изучают принципиальные вопросы бытия, сознания, личности, общества, творчества, свободы наряду с принципиальными вопросами вычислимости, алгоритмичности, программируемости, моделируемости, реализуемости. То есть они начинают выполнять роли философов. Это, как правило, весьма специфичные, не «школьные» философы. Однако несмотря на отсутствие философского профессионализма, из-за неожиданного всплеска государственной и общественной значимости ИИ, происшедшего несколько лет назад, философия таких философов потихоньку начинает доминировать над всеми сферами философского знания, а концептуальная предпосылка философствования — тест Тьюринга — становится основой отношений «человек — мир». Тест Тьюринга раскалывает «мир» на естественный и искусственный. Но он пытается и соединить этот мир, расколотый им же.

Этому соединению способствует второе важное событие, шестидесятилетний юбилей которого в 2020 году отмечался в честь публикации статьи «Разумы и машины»<sup>1</sup>. Хилари Патнэм впервые предложил в едином формате совместить идеи разума, мозга и компьютера. Традиционная психофизическая проблематика революционно дополнилась психотехнической проблемой. Теперь не только биологический, естественный мозг связывается с сознанием. С психическим стал связываться технический, искусственный мозг. В годы написания упомянутой статьи ее

---

<sup>1</sup> “Minds and machines”, Патнэм.

основная идея — машинный функционализм, пожалуй, казался шизофренической аномалией: как можно в здравом уме утверждать о том, что *мои* психические состояния — это логические состояния некоторой абстрактной машины Тьюринга? Однако сегодня эта форма функционализма обрела рациональные основания, стала необычайно модной, плодотворной, перестала встречать веские возражения. Функционализм машинных состояний, как и полстолетия назад, продолжает координировать частные теории сознания: он обогащает бихевиоризм и физикализм идеями программируемых состояний, функционально инвариантных относительно субстрата реализации. Но сегодня функционализм преодолевает частное мелкотемье когнитивистики. Как парадигма объяснения естественного интеллекта и построения искусственного интеллекта он стал претендовать на построение единой картины мира. Ведь если системы ИИ неотчуждаемо прочно вошли в нашу жизнь, стали имманентны реальности, то и картина мира должна базироваться на концепции вычислимости этой реальности.

Потому попытаемся охарактеризовать философские, фундаментальные параметры отношения «человек — система ИИ». Для этого дадим четкое представление о функциях теста Тьюринга как истинного, настоящего, *комплексного теста*, изучающего всевозможные фундаментальные мировоззренческие вопросы связи «Я — компьютерный мир». При этом важно изучить систему ИИ в философских рамках, то в контексте противоречивых диспозиций. Для этого изучим возможности системы ИИ обеспечивать как общие, так и особенные стороны интеллектуальных решений. Одновременно покажем на примере изучения системы особенного ИИ то, что комплексный тест достаточно просто конструировать. Наконец, при рассмотрении системы общего ИИ, мы покажем возможность постнеклассического позиционирования человека в исследованиях ИИ: вместо того, чтобы вгонять *Меня*, человека, в машину Тьюринга, как это принято в патнэмовском машинном функционализме, следует *Меня* сделать тьюринговым наблюдателем. В новом, тестовом функционализме, *Я*, сохраняя субъективное богатство естественной личности, наблюдаю

и совершенствую естественно-искусственный мир посредством эпистемологического инструментария, посредством комплексного теста Тьюринга.

### **Комплексный тест Тьюринга в отношении «человек — система ИИ»**

*Функции комплексного теста Тьюринга.* О комплексном тесте Тьюринга я пишу на протяжении двадцати лет. Имеется ряд статей, монографий, докторская диссертация на эту тему, доклады в авторитетной аудитории на 55-, 60- и 65-летних рубежах истории развития идеи теста Тьюринга<sup>1</sup>. Основная идея этих работ состоит в следующем: достойно оценить вклад главной идеи искусственного интеллекта — теста Тьюринга — мы сможем, когда систематически изучим и интегрируем эти «версии версий», выявим, обобщим и доопределим основные функции, которые выполняет такой комплексный тест в современном информационном обществе, наметим стратегии реализации когнитивной феноменологии развивающегося компьютерного мира. Этот тест как истинный, настоящий тест Тьюринга представляется

---

<sup>1</sup> Алексеев А.Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИИнтелл, 2013. 304 с.; Алексеев А.Ю. Возможности искусственного интеллекта: можно ли пройти тесты Тьюринга? // Искусственный интеллект: междисциплинарный подход / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИнтелл, 2006. 448 с. С. 223–242; Алексеев А.Ю. Идея Комплексного Теста Тьюринга // «Тест Тьюринга: философские интерпретации и практические реализации»: Материалы научно-практической конференции, посвящённой 60-летию публикации статьи Алана Тьюринга “Computing Machinery and Intelligence” / Под общ. ред. А.А. Костиковой, редакция и составление Н.Ю. Ключевой. М.: Алькор Паблишерс, 2011. 136 с. (Серия кафедры философии языка и коммуникации). С. 8–19; Алексеев А.Ю. Роль комплексного теста Тьюринга в современной философии искусственного интеллекта // 10 лет — Научному совету РАН по методологии искусственного интеллекта. Материалы симпозиума, 26 марта 2015 г., Институт философии РАН, г. Москва / Ред.: Алексеев А.Ю., Дубровский Д.И., Лекторский В.А. М.: ИИнтелл, 2016. 142 с. (Сер.: Философия искусственного интеллекта). С. 118–127.

для меня концептуальным базисом новой философии — философии искусственного интеллекта. В связи с большим количеством работ и ввиду наличия явных ссылок на них<sup>1</sup> в данной книге я не буду раскрывать подробности построения и использования комплексного теста. Идея «комплексности», «умной составленности», «искусности» (“Comprehensive Turing Test”) такого теста представляется очень простой: надо собрать воедино всевозможные способы изучения мировоззренческих проблем человека в компьютерном мире, формализовать, систематизировать, унифицировать, операционализировать и далее в интегральном формате начать решать конкретные значимые задачи искусственного интеллекта.

В литературе выделяется определенное количество крупных версий теста Тьюринга, т. н. «совершенных частных тестов Тьюринга». В этих версиях четко просматриваются базовые функции, предложенные в оригинальной игре в имитацию:

1) интеррогативная функция, характеризующая содержание и форму вопросов, которые задает тьюринговый судья (interrogator, наблюдатель) при изучении *x*-системы на предмет ее *y*-способностей;

2) дефинитная функция, обеспечивающая определение когнитивного феномена с учетом ее вычислительных коннотаций;

3) конструкторская функция, раскрывающая принципы работы компьютера, способного реализовать когнитивный феномен, исследуемый в тесте;

4) критическая функция, отражающая суть полемики по поводу возможности или невозможности компьютерной реализации когнитивного феномена;

5) конститутивная функция, позиционирующая отношение судьи (наблюдателя) к когнитивным аспектам компьютерной реальности, или, конкретнее, человека к системе ИИ.

---

<sup>1</sup> Алексеев А.Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга. Диссертация на соискание ученой степени доктора философских наук по специальности 09.00.08 «Философия науки и техники». М.: МГУ им. М.В. Ломоносова, 2015. 482 с.

Помимо совершенных версий, имеется большое количество несовершенных тестов. В них акцентируется внимание на, возможно, интересной, но незначительной детали тьюрингового тестирования. Таких несовершенных версий много, более сотни зарегистрированных.

### **Может ли система ИИ делать всё то, что способен делать человек?**

Внешне первичной функцией комплексного теста Тьюринга является интеррогативная функция. Она задает обобщенные вопросы к системе ИИ: «Может ли компьютер мыслить, понимать, жить, творить, любить, осознавать, быть личностью, обществом, обладать свободой и пр.? Короче, может ли компьютер делать как бы “всё”»? Здесь универсальный квантификатор пробегает на множестве частных тестов Тьюринга, надежных, апробированных, критически проверенных.

В попытке поставить эти вопросы и ответить на них комплексный тест выполняет функцию дефиниции «мышления», «понимания», «творчества», «сознания», «другого», «личности», «Я» и пр. При этом исключается концептуальный дискомфорт употребления компьютерно-ориентированных понятий: «машинное мышление», «компьютерное сознание», «человек-компьютер», «любовь робота». Эти дефиниции имеют функциональный статус: в системе ИИ определяются *x*-феномены *u*-систем, *z*-субстраты которых не существенны или вовсе не известны. Главное — роль (функция), которую играют люди и системы ИИ в отношениях реализации когнитивных феноменов и их трансляция в компьютерный мир.

В-третьих, комплексный тест Тьюринга исполняет функцию конструирования новых принципов компьютерной технологии. Зачем перегонять «из пустого в порожнее» метафизические конструкции «философии сознания и ИИ»? Надо решать другие, более прагматичные задачи, например, междисциплинарного

моделирования «творчества». Такая переориентация, возможно, и есть направление «главного удара» в доктрине развития искусственного интеллекта. Помимо подобных общих вопросов, комплексный тест Тьюринга способствует макетному демонстративному изучению новых принципиальных проектов систем ИИ.

В-четвертых, критическая функция. Комплексный тест — это концептуальный инструментарий «чистых» дискуссий в познании когнитивных феноменов компьютерного мира. К аргументации не примешиваются технологические, материальные, идеологические, политические, экономические, религиозные, морально-этические, эстетические и иные соображения. Такая форма фундаментального познания современного мира в финансовом плане стоит ровно столько, сколько стоит «мышление», а сегодня такой абстрактный труд стоит много меньше труда уборщицы или, скажем, охранника. Однако в когнитивистском отношении данная форма познания предполагает крупные мысленные баталии, чрезвычайно важные для судеб цивилизации. Надо «сесть в кресло и подумать». В ходе оперирования мысленными экспериментами легко доказывается, например, то, что большинство современных компьютерных систем не следует относить к системам ИИ: они реализуют не эвристики, но всего лишь «большие» алгоритмы. Достаточно просто критически оценить современные социальные проекты, в которые вовлекается идея вычисления когнитивных феноменов. Например, ЕГЭ — а это суть обычная автоматизированная система — не способствует познанию и пониманию. Напротив, она отдаляет эти самые важные параметры образования: затрудняет познание, отдаляет понимание как предмета понимания, так и понимание понимания этого предмета. Перспективы общества, в котором калькуляции когнитивной сферы возведены в социальную норму, раскрываются тьюринговым тестом философских зомби.

В-пятых, конститутивная функция комплексного теста обеспечивает вписывание концепта Я в картину высокотехнологичного мира. В конвергентное развитие и функционирование систем ИИ надо включить мое самосознание, самость, личность, Я. Если мы согласимся с машинным функционализмом, то получим



что-то из сферы научной фантастики. В тьюринговой же методологии вполне корректно поднимается вопрос о реализации этого концепта посредством системы ИИ. Реализация организуется по схеме четырех «ре-». Возможна *репликация*, то есть имитация *Я* в формате полюбившегося многим «цифрового двойника»; *репрезентация* — моделирование *Я* для получения новых знаний; *репродукция* — воспроизводство функций *Я* в искусственной среде; *рекреация* — продуцирование оригинального «квази*Я*», которого нет в естественной реальности, но, возможно, проявляется в дополненной или виртуальной реальности.

Акцентирую внимание на следующем. Если мы попытаемся реализовать *Я* посредством патнэмовского функционализма, то последовательно попадем в тупик «когнитивного физикализма»: слишком сложным оказывается изучение и общества и мозга. Не до функций с ролями. Когнитивный физикализм, в свою очередь, невозможен в той же мере, в какой невозможен социальный физикализм О. Конта, считал сам Х. Патнэм<sup>1</sup>. Если же мы попытаемся вписать *Я* в систему ИИ методами постнеклассического, тестового функционализма, приписав человеку роль тьюрингового судьи, т. е. наблюдателя этой системы ИИ, то не возникает никакого концептуального дискомфорта. Человек остается человеком, конечно, усложненным человеком, но не фикцией «ре-человека».

Систему ИИ, как систему компьютерной реализации комплексного теста Тьюринга, можно описать подробно, максимально полно, употребив при этом сотни зарегистрированных частных тестов. Но можно поступить по-философски, столкнув противоречия в интерпретациях системы ИИ. А для такой концептуальной contradикции предлагается рассмотреть полярные проекты: систему особенного ИИ и систему общего ИИ. Первый проект определяет уникальный, оригинальный способ функционирования системы ИИ. Второй проект, напротив, задает универсальный, всеохватный формат компьютерной реализации когнитивных функций.

---

<sup>1</sup> Putnam H. Reason, Truth, and History. Cambridge: Cambridge university Press, 1981.

## Человек и системы особенного ИИ

Системы особенного ИИ обеспечивают формирование уникальных, оригинальных, творческих знаний. В комплексном тесте Тьюринга такие системы принято изучать в рамках проблемы компьютерного творчества, в контексте исследований аргумента Лавлейс и теста Лавлейс.

*Тест Лавлейс.* Осмысление проекта системы особенного ИИ, который прекрасно демонстрирует образ креативной робототехники<sup>1</sup>, стоит начать со статьи С. Брингсйорда, П. Беллоу, Д. Феруччи: «Творчество, тест Тьюринга и (улучшенный) тест Лавлейс»<sup>2</sup>. В ней много примеров компьютерной генерации произведений изобразительного искусства, прозы, поэзии. Аргумент Ады Лавлейс — наиболее сильное из всех возражений тесту Тьюринга. Авторы его немного перефразировали: компьютер не может самостоятельно творить, так как творчество требует, как минимум, изобретения чего-либо нового. Но компьютеры не изобретают ничего нового; они всего лишь делают то, что программист «заставляет» их делать посредством программ.

Как имитировать креативность некреативным? Ведь если использовать УЦВМ, высшее достижение которой — решение арифметических задач, то это примерно то же самое, когда глупый пытается имитировать умного. Обратное — имитация умным глупого — конечно, не вызывает сомнения<sup>3</sup>. Могут различаться степени глупости. Тогда глупому может показаться умным ответ, который фактически является глупым, но, скажем, немного умнее общего уровня глупости. Примерно такая же релятивная ситуация обнаруживается и в примере с креативностью: отношение творчества асимметрично и градуально.

---

<sup>1</sup> Алексеев А.Ю. Принципы креативной робототехники // Технологос. 2020. № 4. С. 5–17. DOI: 10.15593/perm.kipf/2020.4.01.

<sup>2</sup> Селмер Брингсйорд, Пол Беллоу, Дэвид Феруччи. Творчество, тест Тьюринга и улучшенный тест Лавлейс / Пер. с англ. А. Ласточкина // Тест Тьюринга. Роботы. Зомби / Под ред. А.Ю. Алексеева. М.: МИЭМ, 2006. С. 62–85.

<sup>3</sup> Block N. Psychologism and Behaviorism // Philosophical Review. 1981. N 90. P. 5–43.

Человек-разработчик  $P$  создает компьютерную систему ИИ  $K$  (бота, робота, искусственного агента), который продуцирует артефакт  $A$  как уникальную, неповторимую, полезную вещь.  $P$  не способен определить, каким образом  $K$  продуцирует  $A$ , несмотря на то, что  $P$  полностью осведомлен об устройстве, функциях и поведении  $K$ .

Система  $K$  проходит тест Лавлейс тогда и только тогда, когда:

1)  $K$  продуцирует артефакт  $A$ ; 2) артефакт  $A$  произведен не путем обмана, случайного стечения обстоятельств или сбоя машин,  $K$  всегда может  $A$  воспроизвести (репродуцировать, повторить); 3) судья  $I$  не способен объяснить, как  $K$  создал  $A$ , хотя имеет полное представление о  $K$ : структуре баз данных, алгоритмах функционирования, техническом обеспечении и пр.

Предложенная аксиоматика, по мнению авторов теста Лавлейс, приводит к следующим соображениям:

1) Креативный тест невозможно пройти путем обмана. Подтасовка ответов характерна для прохождения оригинального теста Тьюринга, здесь же по условию 2 это невозможно.

2) Тест Лавлейс всем современным компьютерным системам присваивает статус некреативных. Все эти модели запрограммированы разработчиками, которые знают особенности их функционирования. Ничего «от себя»  $K$  добавить не могут.

3) Необходимым условием прохождения теста Лавлейс является способность  $K$  к программированию собственного программного кода. Но это невозможно из-за аргумента Геделя-Лукаса-Пенроуза: механическая система, которой является  $K$ , не способна к самоформализации.  $K$  может порождать программы методами «автогенеративного программирования». Но он не способен к автогенерации принципов автогенерации, т. е. не способен самостоятельно порождать автогенеративные аксиомы и правила вывода. Креативная система ИИ должна быть самопрограммируемой на уровне принципов программирования.

4) Но и самопрограммируемости недостаточно для прохождения теста Лавлейс. Дело в том, что все изменяемые правила носят «переходящий» характер. Результат обучения — это, к примеру, новая структура базы данных или изменённый программный код.

Но такие модификации неоригинальны, они происходят в контексте известного, по крайней мере, со стороны учителя.

5) Моделирование «свободной воли», имитация мотивационно-волевых механизмов человека, стремящегося к созданию чего-то нового, также некреативно: всё это осуществляется в известном пространстве решений.

Таким образом, тест Лавлейс показывает невозможность компьютерной реализации феномена творчества. При этом, однако, возможно компьютерное продуцирование квазикреативных («как бы» творческих) продуктов. То есть креативность, оригинальность, особенность системы ИИ напрямую зависит не от её объективных характеристик, а лишь от субъективной компетентности человека, оценивающего артефакт.

## Тест Лавлейс 2.0

Не так давно, в 2014 году, тест Лавлейс был дополнен небольшой статьёй Марка Ридла «Тест Лавлейс 2.0 на искусственные креативность и интеллект»<sup>1</sup>. Была предложена новая версия исходных предпосылок:

1)  $K$  продуцирует артефакт  $A$  типа  $T$ ; 2)  $A$  должен соответствовать набору ограничений  $C$ , где  $c_i \in C$  любой критерий, выражаемый на естественном языке; 3) человек-оценщик  $H$ , выбравший  $T$  и  $C$ , удовлетворен тем, что  $A$  является допустимым экземпляром  $T$  и соответствует  $C$ ; 4) человек-судья  $I$  определяет комбинацию  $T$  и  $C$  как нереальную для обычного человека.

В предложенной аксиоматике количество людей увеличилось. Помимо человека, выполняющего роль тьюрингового судьи, появляется человек ( $H$ ), компетентно оценивающий гениальность, талантливость, креативность артефакта  $A$  или, напротив, бесталанность, обыденность этого произведения. Происходит четкая

---

<sup>1</sup> Mark O. Riedl. The Lovelace 2.0 Test of Artificial Intelligence and Creativity. Proceedings of the AAAI Workshop: Beyond the Turing Test, Austin, Texas, 2015; <https://arxiv.org/abs/1410.6142>

спецификация критериев оцениваемого продукта (п. 2). Человек-оценщик и человек-судья должны быть экспертами в достаточно ограниченной области, т. е. нужен конкретный специалист, а не «человек вообще», утверждающий то, что машина может или не может творить.

Тест «Лавлейс 2.0» был предложен в пик интереса к моделям глубокого машинного обучения. Эти модели, помимо многофакторного взвешивания нейронных сетей, позволяют комбинировать фон и рисунок, накладывая, например, на выявленный стиль одного художника «рисунок» другого художника.

«Творчество не уникально для человеческого интеллекта. Но это одна из его отличительных черт. Многие формы творчества требуют интеллекта», — полагает М. Ридл<sup>1</sup>. На наш взгляд, это суждение дополняет тьюринговский подход: интеллектуальность определяется креативностью, среди всего прочего. Креативность и интеллектуальность — это взаимополагающие компетенции системы ИИ. Например, интеллектуальная робототехника основана на технологиях креативной робототехники. И, взаимно, креативная робототехника должна быть интеллектуальной.

### Тест Лавлейс 3.0

Продолжая нумерацию версий теста, предложим тест Лавлейс 3.0. По сути, идея нового теста на компьютерное творчество прозвучала в нашей ранней статье<sup>2</sup>. Оценивать компьютерную креативность можно с двух позиций: физикалистской и менталистской. Физикалистский (натуралистический) подход обеспечивает имитацию творческой деятельности за счет сложных и необычных комбинаций известных объектов. Объекты могут быть представлены явно.

---

<sup>1</sup> Ibid.

<sup>2</sup> *Пожарев Т., Алексеев А.* Креативные мультимедиа: физикалистский и менталистский подходы // *Философия творчества: материалы Всероссийской научной конференции*, 8–9 апреля 2015 г., Институт философии РАН, г. Москва / Под ред. Н.М. Смирновой, А.Ю. Алексеева. М.: ИИнтелл, 2015. С. 418–424; <https://aintell.info/21/book.pdf>

Но могут быть завуалированными, неявными, скрытыми, то есть выглядеть своеобразными ассамбляжами субобъектов. Нейронные сети глубокого обучения — а именно эти технологии вызвали появление теста Лавлейс 2.0 — интересны тем, что позволяют манипулировать с субобъектами, например со «стилями» художников. Второй, когнитивный (менталистский), подход ориентирован на работу с феноменами сознания, с идеальными объектами, со смыслами. Это абстракции, интенции, концепции. Эти когнитивные феномены семиотически закрепляются, выражаются и используются в объективном мире. Они значат, но их в нем нет как вещей.

Аксиоматика теста остается прежней. Немного уточняется п. 3:

3\*) оценщик *H* выявляет факт осмысленного продуцирования артефакта *A*.

Например, какая операция использовалась при создании *A*: абстрагирование или обобщение? Если оценщик *H* обнаружил абстрагирование, то возникает повод утверждать креативность. Если *H* обнаружил обобщение, то это повод отрицать креативность. Нейронные сети глубокого и неглубокого обучения не формируют идеальные объекты, не оперируют ими. Они способны обобщать и ограничивать. Формально-логически это означает то, что машина работает исключительно с объемами понятий, экстенционально. С содержанием понятий, со смыслами, интенционально, работает исключительно человек.

*Человеку — интенциональное, компьютеру — экстенциональное!* Именно таково фундаментальное отличие между человеком и системой ИИ.

Необходимо расширить концепцию вычислимости, чтобы экстенциональная составляющая системы ИИ представлялась достаточно широко. Принципы алгоритмической вычислимости, предложенные в 1937 году в формате машины Тьюринга в ряде работ<sup>1</sup>, дополнились исторической находкой конца XX века — машиной Корсакова, точнее семейством гомеоскопов. Если телескоп исследует физический макромир, микроскоп — физический

---

<sup>1</sup> Алексеев А.Ю. Протонейрокомпьютер Корсакова // Нейрокомпьютеры: разработка и применение. 2013. № 7. С. 6–17.

микромир, то гомеоскоп, по замыслу его создателя, русского инженера С.Н. Корсакова, изучает подобие в идеях, то есть исследует идеальный мир. Машина Корсакова обеспечивает визуализацию сложных идей и операциональную обработку рутинных ее составляющих. Предлагается коннекционистский способ «вычислений»: параллельный, ассоциативный, гомеоморфный. Он прекрасно дополняет классическую парадигму символьных вычислений. Принципиальные отличия машин Корсакова и Тьюринга соответственно таковы: коннекционизм/символизм; усиление/имитация человеческого интеллекта; активность/пассивность тьюрингового судьи; интерактивность/априорность теста; открытость/замкнутость компьютерной системы; произвольный/ведомый алгоритм; креативность/реактивность действий; формализация работы художника/математика и пр. В этих оппозициях подчеркивается то, что машина Корсакова двойственно дополняет машину Бэббиджа-Лавлейс-Тьюринга. Поэтому ее применение подводит к более универсальным теоретико-алгоритмическим принципам построения компьютерных систем, чем те, которые по сути не менялись почти два века, с момента описания Адой Лавлейс машины Ч. Бэббиджа.

Эти машины по отдельности не способны решить реформаторские задачи построения принципов креативной системотехники. В машине Корсакова отсутствует идея автоматных переходов. Машина Тьюринга, существующая благодаря этим переходам, тем не менее, примитивно целочисленно задает лингвистические конструкции. Это не поддерживает представления об индетерминированности, континуальности, многозначности, неформальности, незавершенности вычислительного процесса в системах особенного ИИ. Совместно же машины способны на многое. Самый простой способ комплексного использования машин состоит в следующем: вместо ленты Тьюринга использовать стопку перфокарт Корсакова. Тогда программируемые переходы наполнятся богатым репрезентативным содержанием. Важно то, что собственная инструкция автоматного перехода формируется путем обобщения либо ограничения наборов субсимвольных дескрипций этой инструкции.

Такая комплексная машина Корсакова-Тьюринга может выступить принципиальным проектом системы особенного ИИ.

Она способна программировать богатое репрезентативное содержание и охватывать широкую экстенциональную сферу компьютерной техники. Как было отмечено выше, вряд ли тест Лавлейс 3.0 можно пройти, этот тест задает абсолютную точку измерения «творчества», оригинальности, особенности системы ИИ. Однако, на наш взгляд, вполне возможна реализация квазикреативных артефактов посредством машины Корсакова-Тьюринга. Продукты ее комплексных «вычислений», пожалуй, введут в заблуждение любого взыскательного эксперта.

Концептуальные принципы креативных систем ИИ (например, креативной робототехники) основаны на комплексном применении тьюринговых тестов. Тест Лавлейс — один из подобных тестов. В решении проблем компьютерного творчества он играет определяющую роль. Мы не утверждаем, что концептуальные основы особенных, креативных систем ИИ следует базировать именно на таком тесте Лавлейс в формате его различных версий. Но этот тест или тест, ему подобный, необходим в методологическом инструментарии любой креативной технологии. Он предельно обостряет дистинкцию творческого и нетворческого в целостной человеческой деятельности, которая уже сегодня начинает осуществляться совместно с креативными системами ИИ, например, с эмоциональными креативными роботами. По крайней мере, сценарии будущей совместной жизни человека с эмоциональными эмпатичными гиперреалистичными роботами сегодня массово формируются, обсуждаются, закрепляются миллионами пользователей игры “Detroit: Become Human”<sup>1</sup>.

Очевидно то, что конструирование частных тестов Тьюринга и создание новых, оригинальных систем ИИ представляется методологически приемлемой задачей. Но как быть с тем, что система ИИ, подобно человеку, должна работать с большим количеством предметных областей? Для обобщения этих «доменов» целесообразно рассмотреть т. н. «Общий искусственный интеллект» как направление научно-теоретических и инженерно-технологических исследований, ориентированное на

---

<sup>1</sup> [https://ru.wikipedia.org/wiki/Detroit:\\_Become\\_Human](https://ru.wikipedia.org/wiki/Detroit:_Become_Human)



построение и применение компьютерных имитаций, моделей и репродукций когнитивных феноменов самого широкого спектра жизненных, психических, личностных и социальных проявлений. Методологической основой системы общего ИИ выступает общий компьютерный функционализм, отмеченный в начале данной работы как современная интерпретация психофизиологической теории машинного функционализма Х. Патнэма 60-х гг. прошлого века.

В формировании общего функционализма выделяются собирательный, определительный и наблюдательный подходы. *Собирательный подход* — это сбор, идентификация, координация, формализация, систематизация, унификация, кодификация всевозможных функционалистских теорий. *Определительный подход* — это анализ и выявление главных функционалистских характеристик, отношений, закономерностей, причинностей, инвариантных относительно содержания когнитивных феноменов. *Наблюдательный подход* позволяет оценивать с позиции человека или социальной общности, погруженных в коммуникативные «волны» виртуального и реального мира, различные статусы технологических реализаций общего функционализма: онтологические, эпистемологические, логические, лингвистические, аксиологические, эстетические, этические и праксеологические особенности проектов искусственной жизни, искусственного сознания (мозга), искусственной личности, искусственного общества.

Система общего искусственного интеллекта базируется на таких принципах построения интеллектуальной системы, которые пригодны для компьютерной реализации витальных, ментальных, персональных и социальных феноменов. Это принципы создания и развития логических, математических, программных, информационных, языковых, технических средств интеллектуальных компьютерных технологий. Они инвариантны как для традиционных программируемых устройств, так и для аватаров, роботов, киборгов. Принципы действенны, т. к. включены в технологию общего ИИ, играя роль ее *концептуальной организации*. Концептуальная организация технологии систем общего ИИ —

это система идей, категорий, понятий, мысленных экспериментов, теоретических дискуссий, логических моделей и других составляющих индивидуальной и социальной деятельности, определяющей смыслы, значения, ценности и цели этой технологии.

Предваряя размышления по поводу общего функционализма, целесообразно выбрать некий общий, очевидный и демонстративный когнитивный феномен, относительно которого можно строить теории и модели систем общего ИИ. В исследованиях сознания таким феноменом является *боль*. На наш взгляд, другой когнитивный феномен — *потребность* — имеет преимущество из-за широты масштаба феноменологической очевидности, эпистемологической адекватности, онтологической фундированности.

Не вдаваясь в теоретические обоснования тезиса о первичности *потребности* перед всеми другими приватными и общественными явлениями когнитивно-компьютерной феноменологии систем общего ИИ, отметим лишь то, что все мы являемся субъектами поведения, инициированного нашими потребностями. Феномен *потребности* многообразен. Он междисциплинарно изучается естественными, социальными, гуманитарными и техническими науками. В перспективе исследования «искусственной потребности» понятие «искусственного интеллекта» представляется узким. Ведь если даже под «интеллектом» понимать, в кантианском стиле, общую человеческую способность упорядочивать чувственные данные, рационально рассуждать и разумно регулировать поведение, тем не менее тьюринговский тест на такой, весьма широкий, «интеллект» не включает изучения возможностей компьютерной имитации «жизни», «сознания», «самосознания», «творчества», «воли», «эмоций», «личностного» и многого другого из того, что требует феноменология общего ИИ. Поэтому предпочтительно в роли методологической основы систем общего ИИ избрать обозначенный выше комплексный тест Тьюринга как трудно обозримую совокупность многочисленных версий теста Тьюринга и версий этих версий. Совместно эти версии способствуют обоснованному атрибутированию (приписыванию) компьютерным системам феноменов

*потребления* витального, психического, личностного и социального содержания. То есть интеграция комплексного теста осуществляется не по причине системотехнического расположения, а в силу априорной «потребности», некоторой квазиинтенциональности. Теперь возникает возможность связать «потребности» в единую интегральную систему с учетом квазиалгоритмического статуса их реализации и цифрового представления. Таким образом, исследования *потребностей*, скажем, в выживании, образовании, свободе, справедливости получают единую когнитивно-компьютерную базу, интегральный инструментарий, включенный в глобальную инфраструктуру экономических, политических, социальных, материально-производственных, научных, военных и др. технологий. Подобная методологическая основа интегрирует различные версии функционализма и подчеркивает инвариантность различных когнитивных феноменов относительно субстрата их реализации.

### Собирательный подход к системе общего ИИ

Охарактеризуем данный подход, отталкиваясь от классической функционалистской формулировки «боли» — того ментального состояния, в котором пребывает большинство зарубежных функционалистов: *боль* (*c*) — это функциональное состояние (*f*), возникающее из-за телесного повреждения (*p*); выражается в стогах (*b*), которые могут слышать (*m*) другие люди (*s*), проявляется в телесных вздрагиваниях (*b*); информирует меня (*m*) о том, что с телом что-то не в порядке (*p*), продуцирует беспокойство (*m*) и желание (*m*) устранить (*p*) это состояние (*f*) при отсутствии конфликтующих намерений (*m*). В этой нестрогой формуле, производной от работы Дж. Левин<sup>1</sup>, термины суждения проиндексированы обозначениями: *c* — когнитивное, *m* — мен-

---

<sup>1</sup> Levin J. Functionalism // The Stanford Encyclopedia of Philosophy (Fall 2013 Edition) (<http://plato.stanford.edu/archives/fall2013/entries/functionalism/>).

тальное,  $p$  — физическое,  $f$  — функциональное,  $b$  — бихевиоральное,  $s$  — социальное.

Придумаем аналог для наших целей: *Потребность* ( $c$ ) — это функциональное состояние ( $f$ ), возникающее из-за какого-то недостатка в индивидуальном или социальном организме ( $p$ ); проявляется в реактивном ( $b$ ), целенаправленном ( $m$ ) или ценностно-ориентированном поведении ( $s$ ); формирует информацию ( $m$ ) о том, что в организме ( $p$ ) чего-то не хватает ( $f$ ), продуцирует интенцию ( $m$ ) преодолеть ( $p$ ) это состояние ( $f$ ) при отсутствии конфликтующих ( $s$ ) намерений ( $m$ ).

Если преобразовать эту формулу в контексте *функционализма машинных состояний*, то получится, что возникновение нужды в чем-то, испытывание нужды, намерение ее удовлетворить — все это описывается программой машины Тьюринга. На мой взгляд, показательно такое высказывание: ощущать *потребность* ( $m$ ) — значит быть в некотором функциональном состоянии  $F_i$  которое возникает как следствие входных данных  $i_j$  ( $b$ ), других функциональных состояний ( $f$ ) и служит причиной поведенческой реакции  $o_k$  ( $b$ ), множественно реализуемой ( $f$ ) на субстратах различной природы ( $p$ ), по программе ( $f$ ), представимой автоматной таблицей ( $f$ ) машины Тьюринга ( $p$ ).

Сравним с классической для функционализма работой Х. Патнэма «Психика и машины»: «1) боль  $P$  тождественна стимуляции  $C$ -волокон»<sup>1</sup>; 2) «Состояние  $S$  тождественно включению триггера № 36»<sup>2</sup>. Например, *потребность* может быть реализована путем введения ферментов в микротубулы нейронов коры головного мозга человека, собаки, других биологических существ, воображаемых аналогов этих компонентов у марсиан, социумов, роботов. Но такая же потребность может быть реализована не за счет физического воздействия, а в силу модификации информационной причинности функциональных компонент.

<sup>1</sup> Putnam H. Minds and machines // Hook S. (ed.) Dimensions of Mind. New York university Press, 1960.

<sup>2</sup> Ibid.

Психофункционализм не дает подробного описания состояний и переходов машины Тьюринга. То, что мы можем описать, представимо машиной Тьюринга, что не можем описать, но феноменально переживаем, описывается в виде черных ящиков (блоков) как «сущностей», играющих роли в осуществлении поведения. Н. Блок предлагает формулу. Если  $T$  — психологическая теория с  $n$  ментальными состояниями, из которых, например, 18-е является «потребностью» (у Н. Блока 17-е является «болью»), тогда психофункционалистская «теория потребности» задается следующим образом: ощущать потребность ( $c$ ) значит быть  $F_{18}$ -м при условии:  $F_1 \dots F_n (T(F_1 \dots F_n, i_1, \dots, o_1, \dots))$ , где  $F_p, \dots, F_n$  — функциональные термы, обозначающие самые разные когнитивные феномены (убеждения, мысли, ощущения, намерения и пр.);  $i_1, \dots, o_1, \dots$  и т. д. — это входы/выходы системы ( $p$ )<sup>1</sup>.

Аналитический функционализм формирует, обрабатывает, транслирует тематически-нейтральные функциональные термины, осуществляет аналитические операции обобщения и ограничения понятий, например, операции подведения отдельных психических феноменов, влияющих на *потребность*, под достаточно четко фиксированный когнитивный тип. Также этот нейтральный язык способен выделять элементарные свойства и фиксировать сингулярные события и, самое главное, коррелировать когнитивные феномены с их физическими реализаторами, в частности с компьютерными устройствами.

Функционализм тождества функциональных состояний — это развитие программы аналитического функционализма на базе крупных наработок физикалистской теории тождества. Так, для высказывания «память — это физико-химический процесс в мозге» проблема тождества заключена в прояснении того, что такое «это». Для анализа когнитивных феноменов применяется двумерный лингво-логический аппарат. Один подязык участвует в образовании психологических дефиниций, другой — естественнонаучных, в частности нейрофизиологических.

<sup>1</sup> Block N. Psychologism and Behaviorism // Philosophical Review. 1981. N 90. P. 5–43.

*Компьютерный функционализм (компьютеризм).* Состояния системы ИИ соотносятся с содержанием этих состояний<sup>1</sup>: Возникновение потребности означает включение триггера № 37 и выявление содержательных условий и контекстов его срабатывания.

В целом собирательному подходу к построению общего компьютерного функционализма присуща трудная проблема, которая заключена в необозримом разнообразии «функционализмов», собираемых в единую систему. И это разнообразие нелинейно растет.<sup>2</sup>

### Определительный подход к системе общего ИИ

Томас Полджер в своей работе “Natural minds”<sup>3</sup>, пожалуй, впервые вводит в научный оборот понятие «*Общий функционализм*» (Generic functionalism) для интеграции различных функционалистских теорий вокруг главного функционалистского отношения. По мнению Т. Полджера, таковым является отношение множественной реализации. Это функциональное отношение устанавливается между реализатором и супервентором. *Реализатор* фактически осуществляет (реализует) биологическое, нейрофизиологическое, психофизиологическое, персональное, социальное состояние и задает содержание этого состояния. *Супервентор* сопровождает процессы, осуществляемые реализатором. Причем в разных вариантах трактовки понятия «супер-

---

<sup>1</sup> Horst S. The Computational Theory of Mind // The Stanford Encyclopedia of Philosophy (Spring 2011 Edition) (<http://plato.stanford.edu/archives/spr2011/entries/computational-mind/>).

<sup>2</sup> Алексеев А.Ю. Функционализм машины Сеченова vs функционализм машины Тьюринга // Алексеев А.Ю. (ред.) 150 лет «Рефлексам головного мозга». Сборник научных трудов, посвященных изданию статьи И.М. Сеченова (23 ноября 1863 г.). М.: ИИнтелл, 2014. С. 138–185.

<sup>3</sup> Polger T.W. Natural minds. Massachusetts Institute of Technology: A Bradford book, 2004.

вентор» это сопровождение осуществляется перед реализатором, за ним, параллельно ему или в единстве с ним.

*Народно-психологический функционализм* в отношении супервентор (sup) / реализатор (re) формулируется следующим образом. Если агент  $a$  (sup) испытывает потребность ( $c$ ) в  $x$  ( $m$ ) и уверен ( $m$ ), что  $x$  можно достичь ( $m$ ), выполнив  $y$  ( $p$ ), и приписывает  $y$  ( $m$ ) лучший способ достижения  $x$  среди других возможных способов  $\{y', y'', y''', \dots\}$  ( $b$ ), тогда  $a$  стремится исполнить  $y$  ( $f$ )<sup>1</sup>.

К главному функционалистскому отношению можно относиться по-разному: определять его бытийственный статус, изучать каузальные контексты, соотносить механизмы функционирования с механизмами развития, рассматривать исследовательские предпочтения. По всей видимости, исходя из этих соображений, Т. Полджер предлагает набор функционалистских формулировок.

*Метафизический функционализм* обосновывает онтологические формулы следующего вида: «быть в состоянии потребности ( $c$ ) — значит быть в функциональном состоянии относительно входа/выхода системы, других ментальных состояний и системы в целом»<sup>2</sup>.

*Интенциональный функционализм* объясняет способность функционального состояния обладать интенциональным содержанием: как, за счет чего и почему некоторые состояния или функциональные элементы способны быть носителями значения или выступать непосредственно смыслами: «Для когнитивного состояния потребности ( $c$ ) обладать интенциональным или репрезентативным содержанием (или быть способным к этому) значит выполнять для  $m$  функциональные роли  $\{f_1, \dots, f_n\}$ »<sup>3</sup>.

*Семантический функционализм*: «содержание потребности ( $c$ ) определяется функциональными ролями  $\{f_1, \dots, f_n\}$ , выполняемыми в системе ИИ  $S$  относительно к внутренней

---

<sup>1</sup> В данной формулировке предложен пример, измененный из [Robb, Heil 2013]).

<sup>2</sup> Polger T.W. Natural minds. Massachusetts Institute of Technology: A Bradford book, 2004. P. 80.

<sup>3</sup> Ibid. P. 87.

когнитивной организации  $S$  либо к взаимодействию  $S$  с внешним окружением»<sup>1</sup>.

*Теоретический функционализм*: содержание, задаваемое термом или предикатом «потребность»  $P$  теории  $T$  определяется их функциональными ролями  $\{R_1, \dots, R_n\}$ , которые  $P$  выполняет в  $T$ <sup>2</sup>.

*Объяснительный функционализм* объясняет когнитивный феномен потребности посредством референции когнитивного термина к функции, которую выполняет некоторый реализатор в системе в целом. Имеются классические объяснительные схемы: телеологическая, объясняющая каузальные отношения стремлением к цели; адаптационная, строящая объяснение, исходя из понятия механизма приспособления и эволюции. Для последней схемы объяснение потребности ( $c$ ) как наличного, сохраненного или измененного когнитивного состояния в системе  $S$  ссылается на функциональные роли  $\{f_1, \dots, f_n\}$  которые  $t$  занимает в системе  $S$ , или на то, как некоторые свойства  $\{p_1, \dots, p_n\}$  способствуют конкретизации (демонстрации) посредством  $t$  функциональных состояний  $\{f_1, \dots, f_n\}$  в  $S$ <sup>3</sup>.

*Методологический функционализм* изучает феномен в терминах функций и функциональных ролей системы, ее подсистем, компонентов и элементов. Метод понимания потребности ( $c$ ) системы  $S$  заключается в определении функциональных ролей  $\{f_1, \dots, f_n\}$ , которые  $M$  играет в системе  $S$  или в некоторой надсистеме  $S^*$  (если  $S$  играет в  $S^*$  функциональную роль) или в определении функциональных блоков  $\{(f, i, o)_1, \dots, (f, i, o)_n\}$ , композиционно задающих ( $c$ )<sup>4</sup>.

Отметим, что определительный подход к построению общего компьютерного функционализма весьма существенно разнообразит функционализмы, предваряя тем самым собирательный подход. Так, конкретизация реализаторов (супервенторов) позволяет выделить физические, химические, биологические, нейрофи-

<sup>1</sup> Ibid. P. 90.

<sup>2</sup> Ibid. P. 93.

<sup>3</sup> Ibid. P. 93.

<sup>4</sup> Ibid. P. 107.



зиологические, психические, персоналистические, социальные и др. «функционализмы». Выявленные функционализмы, если их собрать и обработать, позволят сформировать иные определители функционалистских отношений. Модификация определителей приведет к новым разновидностям функционализма, которые вновь следует собрать и т. д. Получается, что совместно эти два подхода позволяют проект системы общего ИИ представить в синергетическом формате. Возникает вопрос о смысле этого: кому и зачем нужна такая самоорганизация? Решить вопрос помогает еще один, наблюдательный, подход.

### Наблюдательный подход к системе общего ИИ

Еще раз вернемся к идее машинного функционализма. Это чрезвычайно универсальная и общая форма функционализма. В самом деле, имеется концептуальная возможность описать *мои* когнитивные состояния как машинные состояния «мышления», где предметом «мышления» выступают различные теории мышления, среди которых есть функционалистские теории, а в них мы обнаружим теорию машинного функционализма. Столь универсальная форма функционализма явно претендует на доминирующую роль в методологии общего ИИ. Поэтому обозначим лозунг: *«назад, к машинному функционализму!»*. Однако при этом изменим роль человека. Не будем вгонять его в компьютерные программы и большие данные систем ИИ. Напротив, вынесем человека во вне, сделаем его наблюдателем, тьюринговым судьей, который оценивает собственные когнитивно-компьютерные компетенции и когнитивные компетенции нерегистрируемого множества других людей, животных, сообществ. Организуем *тестовый функционализм*, основанный на применении комплексного теста Тьюринга к цифровому обществу и электронной культуре. Снабдим человека концептуальным инструментарием комплексного тестирования компьютеризованной реальности и предоставим возможность решать мировоззренческие, собствен-

но человеческие проблемы личности, свободы, смысла жизни, бессмертия, социального идеала и т. п.<sup>1</sup>

Сформулируем идею *тестового компьютерного функционализма*: Я (с) — главный потребитель и конструктор цифрового мира, конституируемого в комплексе компьютерных репликаций, репрезентаций и репродукций моих витальных (b), ментальных (m), персональных (h), социальных (s) потребностей.

В процессе наблюдения за своими компьютерно реализованными компетенциями Я осуществляю операцию редукции и по-гуссерлиански отбрасываю некоторый сопутствующий процессу оцифровки неоцифровываемый и некомпьютируемый «осадок». В «осадок» выпадают качества субъективной реальности: смысл, интенциональность, содержание опыта, творчество, квалиа, самость, «Я». В ходе редукции когнитивного к функциональному, к аппроксимативно полному описанию всех возможных входов/выходов/состояний моей когнитивной системы из проблематики моих потребностей изгоняется собственно моя личная, приватно значимая «потребность»: она переходит в область функциональной дескрипции. Вместе с этим перестают обладать качеством субъективной реальности мои боли, надежды, мысли, моя свобода, творчество и собственно «Я», *мое сознание, мои потребности*. Они законно и незаконно отчуждены и искаженно трансформированы цифровым обществом. Но для «Я» они не утрачены. Тестовый функционализм возвращает меня себе, намечая перспективы компьютерной реализации того, что ранее считалось невозможным запрограммировать, и выделения моего «собственно человеческого» на очередном этапе построения системы общего ИИ.

Однако каково быть при этом системой ИИ, например роботом, который подробно и полно меня имитирует? Ответ на этот вопрос более прозрачен, нежели ответы на вопросы о том, каково

---

<sup>1</sup> Алексеев А.Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга. Диссертация на соискание ученой степени доктора философских наук по специальности 09.00.08 «Философия науки и техники». М.: МГУ им. М.В. Ломоносова, 2015. С. 435–443.

быть камнем, деревом, мышью, женщиной, социумом?<sup>1</sup> Отвечать на этот вопрос — значит повторить всю историю развития ИИ, а также представить возможные тренды его развития. Очевидным является то, что первичным для этой системы ИИ становится не цель или ценность, но потребность «выжить» и найти, исходя из собственных «интенций», источник питания для компьютерных реализаций моих когнитивных компетенций. Очевидно и то, что надо решить вопросы о возникновении искусственной потребности, ее переходе в цели, цели в задачи, задачи в решение и, далее, в новые потребности<sup>2</sup>.

### **Праксеология отношений «человек — система ИИ»**

*Заключение.* Система общего/особенного ИИ, предложенная в формате общего компьютерного функционализма собирательно-определятельно-наблюдательного типа, проходящая тест Лавлейс 3.0, позволяет совместить мои естественные потребности с искусственными потребностями, глубже понять и те, и другие, отделив истинные от ложных, понять, что надо для полной и детальной имитации моих потребностей в жизни, познании, общении. Такая система ИИ позволяет разобраться в том, что мне в действительности надо от себя, от естественного мира, в котором я родился, и искусственного мира, в который меня «забросили» и продолжают «забрасывать», используя, например, сегодня идеи государственной фантастики типа «Мультивёрса» (multiverse)<sup>3</sup>.

---

<sup>1</sup> Сломэн А. Что значит быть камнем? // Тест Тьюринга. Роботы. Зомби / Перевод с англ. В. Крючкова; под ред. А.Ю. Алексеева. М.: МИЭМ, 2006. С. 86–102.

<sup>2</sup> Vassilyev S.N. Temporal Reasoning on the Basis of New Logic for Intelligent Control // Proc. of the IEE Inter. Symposium on Intelligent Control, 2003. P. 843–847.

<sup>3</sup> «Людям необходимо использовать возможности метавселенных, чтобы общаться и учиться, несмотря на расстояния», — заявил Президент России Владимир Путин в ходе выступления на международной конференции по искусственному интеллекту и анализу данных Artificial Intelligence (AI) Journey, 12 ноября

Система ИИ создается для *меня*, а не для какой-то там “meta-universe-индустрии 5+”. Я реализую мои компетенции, создаю уникальные оригинальные артефакты, обогащаю компьютерный мир электронной культуры. Я полностью принимает реалии искусственного компьютерного мира. Однако Я дистанцируюсь от этих принципов в пользу естественной свободы, самости, смысла и многого другого внутри меня, но Я открыт «мультиверсу» и совмещаю три ипостаси наблюдения.

Наблюдатель-0. Это наивный взгляд на ИИ, повседневно «знание» о нем, базирующееся на обыденных навыках и знаниях, а также на убежденности в обоснованности сенсаций, технической дезинформации, фантастических фильмов.

Наблюдатель-1. Критическое выявление когнитивных феноменов и принципов их компьютерной реализации. Применение комплексного теста Тьюринга.

Наблюдатель-2. Система ИИ выступает как общественная форма сознания и форма социальной жизни. Именно на этом уровне и возможна реальная проработка культурных, этических, эстетических, политических и др. оснований систем общего/особенного ИИ. Здесь прорабатывается *сложностная* модель участия наблюдателя в системе «человек — система ИИ». Для исследования моей приватной жизни требуется изучить мой внешний мир, а углубленное его изучение — это изменение внутреннего мира. В этой модели все обозначенные уровни систематически замыкаются и компьютерно реализуются. Методологический базис такого «наблюдения» лишь частично затрагивает традиционный компьютерный функционализм. Компьютеризм совмещает машинный функционализм Х. Патнэма, для которого характерен эгологический тезис «Я — машина Тьюринга» с контентным функционализмом его ученика — Дж. Фодора. Но изменения в функционализмах несущественны: в «ментализме»,

---

2021 г. По итогам конференции прошлого года, по всей видимости, уже был создан сильный ИИ. В 2021 г. провозглашается новая фантастическая идея — создание Метавёрса как общегосударственной компьютерной стимуляции. <https://lenta.ru/news/2021/11/12/meta/>

во внутреннем языке мысли, мы просто замечаем иные формы описания машинных состояний индетерминированной машины Тьюринга, не более того.

Комплексный тест Тьюринга позволяет сформулировать новую форму функционализма, которая способствует смысловому конституированию компьютерного мира. Новый, постнеклассический функционализм (тестовый функционализм, функционализм теста Тьюринга в разнообразии современных функционализмов в наибольшей степени идентичен машинному функционализму. Но он восполняет то, что было упущено Х. Патнэмом — тест Тьюринга. Теперь *Я* снабжаюсь концептуальным инструментарием изучения компьютерного мира — комплексным тестом Тьюринга и вполне законно выдвигаю вопросы праксеологического содержания. Например, возможна ли свобода в условиях запрограммированной реальности? Исследуются альтер- и интересубъективные отношения *Я* и *другого* посредством критического анализа атрибуций (приписываний) мышления, переживания, творчества, сознания, самосознания, любви и других *x*-когнитивных феноменов у-системам. Осуществляется постнеклассическое позиционирование: *Я* — главный субъект электронной культуры, активно конституирующий её альтер- и интересубъективные смысловые пространства. *Я* освобождаюсь от алгоритмической зависимости, осуществляя своеобразное компьютерное «эпохэ»: при компьютерной репрезентации когнитивного феномена в формате функционалистской дескрипции *Я* включаю эту программу в «базу знаний» комплексного тестирования «мультиверса». Данная программа принадлежит *Я*, она полно и точно, как утверждают разные функционализмы, характеризует мое сознание и сознание другого. Но это — не *Я* и не другой. Поэтому такую программную реализацию (репликацию, репрезентацию, репродукцию и рекреацию) моих когнитивных способностей надо по-гуссерлиански отбросить за «скобки». Если комплексно отбрасывать атрибуции, исключается очень многое. Однако самое важное и приватное сохраняется «в остатке». Остается все то, с чем не справляется компьютерный

функционализм: самость, квалиа, интенция, проекция «каково быть собою или кем-то», содержание опыта и многое другое, сознательное. Не справляются системы ИИ без человека с этими вещами! Ни в каком мультёрсе никто никогда не обнаружит моей экзистенции.

В машинном и тестовом функционализмах роли *Я* и компьютера радикально различаются. Если в первом утверждается: «*Я — компьютер*», («*Я — машина Тьюринга*»), то во втором, напротив, «*Я — не компьютер*». Тестовый компьютеризм способствует программированию когнитивных феноменов, развивая компьютерный базис электронной культуры этой мультивселенной. Однако, в большей мере, он предназначен для решения мировоззренческих, собственно человеческих, хотя и компьютерно модифицированных проблем личности, свободы, смысла жизни, бессмертия, социального идеала.

Позвольте предложить онтологическую формулу тестового компьютеризма, весьма важную для изучения праксеологических особенностей отношения «человек — система ИИ»: когнитивный феномен компьютерного мира конституируется атрибуциями компьютерно непредставимого *Я*, исполняющего функции комплексного теста Тьюринга. Тезис таков: *Non computato, ergo sum: не компьютеризуем, следовательно, существуем*. Данный тезис, позиционирующий фундаментальные отношения между человеком и компьютерным миром, инспирирован идеей комплексного теста Тьюринга. Поэтому такой тест является основной концептуальной концепцией философских исследований искусственного интеллекта, а эта философия выполняет функции классической философии, но в современном контексте, в контексте электронных технологий компьютерного мира, в которые, пожалуй, уже навсегда «заброшен» сегодняшний человек.

Имеются и положительные моменты такой философии. Возможно усиление тьюрингового оптимизма о равноправности партнерства «человек — компьютер». Система ИИ все-таки создается не для захвата мира, как это предлагал сделать Президент РФ В.В. Путин в выступлении 31 августа 2019 года

перед студентами г. Ярославля <sup>1</sup>. Целью системы ИИ является самопознание человека. Причем таким самопознающим *Я* может быть самый конкретный и самый низовой человек. Подобная «достоевщина» по отношению к человеку, вступающему во взаимодействие с ИИ, лишь подчеркивает величие задачи построения системы общего/особенного ИИ. Именно такой человек значим для компьютерной реализации *Я* во вселенных мультивёрса, ничуть не меньше!

---

<sup>1</sup> В.В. Путин произнес следующее: «Искусственный интеллект — будущее не только России, это будущее всего человечества. Здесь колоссальные возможности и трудно прогнозируемые сегодня угрозы. Тот, кто станет лидером в этой сфере, будет властелином мира»: <http://www.kremlin.ru/events/president/news/55493>

---

---

# ПРИЛОЖЕНИЕ

---

---

**Указ Президента РФ от 10.10.2019 № 490**  
**«О РАЗВИТИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**  
**В РОССИЙСКОЙ ФЕДЕРАЦИИ»**  
**(вместе с «Национальной стратегией развития**  
**искусственного интеллекта на период до 2030 года»)**

В целях обеспечения ускоренного развития искусственного интеллекта в Российской Федерации, проведения научных исследований в области искусственного интеллекта, повышения доступности информации и вычислительных ресурсов для пользователей, совершенствования системы подготовки кадров в этой области постановляю:

1. Утвердить прилагаемую Национальную стратегию развития искусственного интеллекта на период до 2030 года.
2. Правительству Российской Федерации:
  - а) до 15 декабря 2019 г. обеспечить внесение изменений в национальную программу «Цифровая экономика Российской Федерации», в том числе разработать и утвердить федеральный проект «Искусственный интеллект»;
  - б) представлять Президенту Российской Федерации ежегодно доклад о ходе реализации Национальной стратегии развития искусственного интеллекта на период до 2030 года;
  - в) предусматривать при формировании в 2020–2030 годах проектов федеральных бюджетов на очередной финансовый год и на плановый период бюджетные ассигнования на реализацию настоящего Указа.
3. Настоящий Указ вступает в силу со дня его подписания.

Президент  
Российской Федерации  
В. ПУТИН

Москва, Кремль  
10 октября 2019 года. № 490



Утверждена  
Указом Президента  
Российской Федерации  
от 10 октября 2019 г. № 490

## **НАЦИОНАЛЬНАЯ СТРАТЕГИЯ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ПЕРИОД ДО 2030 года**

### **1. ОБЩИЕ ПОЛОЖЕНИЯ**

1. Настоящей Стратегией определяются цели и основные задачи развития искусственного интеллекта в Российской Федерации, а также меры, направленные на его использование в целях обеспечения национальных интересов и реализации стратегических национальных приоритетов, в том числе в области научно-технологического развития.

2. Правовую основу настоящей Стратегии составляют Конституция Российской Федерации, Федеральный закон от 28 июня 2014 г. № 172-ФЗ «О стратегическом планировании в Российской Федерации», указы Президента Российской Федерации от 7 мая 2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года», от 9 мая 2017 г. № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы», от 1 декабря 2016 г. № 642 «О Стратегии научно-технологического развития Российской Федерации» и иные нормативные правовые акты Российской Федерации, определяющие направления применения информационных технологий в Российской Федерации.

3. Настоящая Стратегия является основой для разработки (корректировки) государственных программ Российской Федерации, государственных программ субъектов Российской Федерации, федеральных и региональных проектов, плановых и программно-целевых документов государственных корпораций, государственных компаний, акционерных обществ с государственным участием, стратегических документов иных организаций в части, касающейся развития искусственного интеллекта.

4. Положения настоящей Стратегии должны учитываться при реализации следующих документов:

а) Стратегия развития информационного общества в Российской Федерации на 2017–2030 годы;

б) национальная программа «Цифровая экономика Российской Федерации» и иные национальные проекты (программы), федеральные и региональные проекты, в рамках реализации которых возможно использование технологий искусственного интеллекта;

в) планы мероприятий («дорожные карты») Национальной технологической инициативы;

г) государственные программы, программно-целевые документы, эффективность реализации которых может быть повышена за счет использования технологий искусственного интеллекта;

д) проекты, обеспечивающие достижение целей и показателей деятельности федеральных органов исполнительной власти (ведомственные проекты).

5. Для целей настоящей Стратегии используются следующие основные понятия:

а) искусственный интеллект — комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений;

б) технологии искусственного интеллекта — технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта;

в) перспективные методы искусственного интеллекта — методы, направленные на создание принципиально новой научно-технической продукции, в том числе в целях разработки универсального (сильного) искусственного интеллекта (автономное решение различных задач, автоматический дизайн физических объектов, автоматическое машинное обучение, алгоритмы решения задач на основе данных с частичной разметкой и (или) незначительных объемов данных, обработка информации на основе новых типов вычислительных систем, интерпретируемая обработка данных и другие методы);

г) смежные области использования искусственного интеллекта — технологии и технологические решения, в которых искусственный интеллект используется в качестве обязательного элемента, включая робототехнику и управление беспилотным транспортом;

д) набор данных — совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства Российской Федерации об информации, информационных технологиях и о защите информации и необходимых для разработки программного обеспечения на основе искусственного интеллекта;

е) разметка данных — этап обработки структурированных и неструктурированных данных, в процессе которого данным (в том числе текстовым документам, фото- и видеоизображениям) присваиваются идентификаторы, отражающие тип данных (классификация данных), и (или) осуществляется интерпретация данных для решения конкретной задачи, в том числе с использованием методов машинного обучения;

ж) аппаратное обеспечение — система взаимосвязанных технических устройств, предназначенных для ввода (вывода), обработки и хранения данных;

з) вычислительная система — предназначенные для решения задач и обработки данных (в том числе вычислений) программно-аппаратный комплекс или несколько взаимосвязанных комплексов, образующих единую инфраструктуру;

и) архитектура вычислительной системы — конфигурация, состав и принципы взаимодействия (включая обмен данными) элементов вычислительной системы;

к) общедоступная платформа — информационная система для сбора, обработки, хранения и опубликования наборов данных, доступная в информационно-телекоммуникационной сети «Интернет» (далее — сеть «Интернет»);

л) открытая библиотека искусственного интеллекта — набор алгоритмов, предназначенных для разработки технологических решений на основе искусственного интеллекта, описанных с использованием языков программирования и размещенных в сети «Интернет»;

м) технологическое решение — технология, программа для электронно-вычислительных машин (программа для ЭВМ), база данных или их совокупность, а также сведения о наиболее эффективных способах их использования.

## II. РАЗВИТИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РОССИИ И В МИРЕ

6. Развитие информационных систем, помогающих человеку принимать решения, началось с появления в 1950-х годах экспертных систем, описывающих алгоритм действий по выбору решения в зависимости от конкретных условий. На смену экспертным системам пришло машинное обучение, благодаря которому информационные системы самостоятельно формируют правила и находят решение на основе анализа зависимостей, используя исходные наборы данных (без предварительного составления человеком перечня возможных решений), что позволяет говорить о появлении искусственного интеллекта.

7. В связи с увеличением вычислительных возможностей программно-аппаратных комплексов, в том числе в результате использования графических процессоров и распределенных архитектур вычислительных систем, стало доступным широкое применение машинного обучения на базе множества вычислительных систем, организованных по принципу нейронных сетей (по аналогии с человеческим мозгом), что привело к значительному повышению качества разрабатываемых технологических решений.

8. Машинное обучение характеризуется рядом особенностей. Во-первых, для поиска вычислительной системой непредвзятого решения требуется ввести репрезентативный, релевантный и корректно размеченный набор данных. Во-вторых, алгоритмы работы нейронных сетей крайне сложны для интерпретации и, следовательно, результаты их работы могут быть подвергнуты сомнению и отменены человеком. Отсутствие понимания того, как искусственный интеллект достигает результатов, является одной из причин низкого уровня доверия к современным технологиям искусственного интеллекта и может стать препятствием для их развития.

9. Технологические решения, разработанные с использованием методов машинного обучения в соответствии с пунктом 8 настоящей Стратегии, являются примером искусственного интеллекта, способного решать только узкоспециализированные задачи (слабый искусственный интеллект). Создание универсального (сильного) искусственного интеллекта, способного, подобно человеку, решать различные задачи, мыслить, взаимодействовать и адаптироваться к изменяющимся условиям, является сложной научно-технической проблемой, решение которой находится на пересечении различных сфер научного знания — естественно-научной, технической и социально-гуманитарной. Решение этой проблемы может привести не только к позитивным изменениям в ключевых сферах жизнедеятель-

ности, но и к негативным последствиям, вызванным социальными и технологическими изменениями, которые сопутствуют развитию технологий искусственного интеллекта.

10. Стремительное развитие технологий искусственного интеллекта сопровождается существенным ростом как государственных, так и частных инвестиций в их развитие, а также в разработку прикладных технологических решений на основе искусственного интеллекта. По оценкам международных экспертов, инвестиции в технологии искусственного интеллекта выросли с 2014 по 2017 год в три раза и составили около 40 млрд долларов США. В 2018 году мировой рынок технологических решений, разработанных на основе искусственного интеллекта, составил 21,5 млрд долларов США и, по прогнозам экспертов, к 2024 году достигнет почти 140 млрд долларов США.

11. В настоящее время в мире происходит ускоренное внедрение технологических решений, разработанных на основе искусственного интеллекта, в различные отрасли экономики и сферы общественных отношений. По оценкам экспертов, ожидается, что благодаря внедрению таких решений рост мировой экономики в 2024 году составит не менее 1 трлн. долларов США. Указанные тенденции обусловлены следующими факторами:

а) общий («сквозной») характер применения прикладных технологических решений, разработанных на основе искусственного интеллекта;

б) высокая степень влияния технологических решений, разработанных на основе искусственного интеллекта, на результативность деятельности организаций и человека, в том числе связанной с принятием управленческих решений;

в) высокая доступность инструментов (в том числе программ для ЭВМ с открытым кодом) для разработки на основе искусственного интеллекта технологических решений;

г) потребность в обработке больших объемов данных, создаваемых как человеком, так и техническими устройствами, для повышения эффективности экономической и иной деятельности.

12. Благодаря реализации настоящей Стратегии должны быть созданы условия для эффективного взаимодействия государства, организаций, в том числе научных, и граждан в сфере развития искусственного интеллекта, что позволит российским технологиям искусственного интеллекта занять значительную долю мирового рынка.

13. Российская Федерация обладает существенным потенциалом для того, чтобы стать одним из международных лидеров в развитии и использовании технологий искусственного интеллекта. Этому способствуют

высокий уровень базового физико-математического образования, сильная естественно-научная школа, наличие компетенций в области моделирования и программирования. Российские команды регулярно занимают первые места на школьных и студенческих международных олимпиадах по математике, информатике, программированию. Россия входит в десятку стран — лидеров по количеству научных публикаций по физике, математике, химии. Кроме того, в Российской Федерации сформировано активное и постоянно растущее сообщество специалистов по обработке данных с использованием искусственного интеллекта.

14. Дополнительными благоприятными факторами развития технологий искусственного интеллекта в России являются современная базовая информационно-коммуникационная инфраструктура (высокий уровень доступа к сети «Интернет», развитие сети радиотелефонной связи третьего и четвертого поколений) и доступность мобильной передачи данных.

15. Продукты (услуги) отечественных организаций в сфере информационных технологий (например, поисковые и иные сервисы, социальные сети) пользуются высоким спросом и занимают лидирующие позиции на российском рынке и рынке Евразийского экономического союза. Технологические решения, разработанные в Российской Федерации на основе искусственного интеллекта (например, компьютерное зрение и обработка естественного языка), уже сейчас обладают значительной коммерческой привлекательностью и высоким экспортным потенциалом на мировом рынке.

16. Вместе с тем немногочисленные ведущие участники глобального рынка искусственного интеллекта предпринимают активные действия для обеспечения своего доминирования на этом рынке и получения долгосрочных конкурентных преимуществ, создавая существенные барьеры для достижения другими участниками рынка конкурентоспособных позиций.

17. Реализация настоящей Стратегии с учетом сложившейся обстановки на глобальном рынке искусственного интеллекта и среднесрочных прогнозов его развития является необходимым условием вхождения Российской Федерации в группу мировых лидеров в области развития и внедрения технологий искусственного интеллекта и, как следствие, технологической независимости и конкурентоспособности страны.

18. Согласно прогнозам долгосрочного социально-экономического развития Российской Федерации в случае недостаточного развития и использования конкурентоспособных технологий искусственного интеллекта реализация приоритетных направлений научно-технологического развития страны замедлится, что впоследствии повлечет за собой ее экономическое и технологическое отставание.

### III. ОСНОВНЫЕ ПРИНЦИПЫ РАЗВИТИЯ И ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

19. Основными принципами развития и использования технологий искусственного интеллекта, соблюдение которых обязательно при реализации настоящей Стратегии, являются:

а) защита прав и свобод человека: обеспечение защиты гарантированных российским и международным законодательством прав и свобод человека, в том числе права на труд, и предоставление гражданам возможности получать знания и приобретать навыки для успешной адаптации к условиям цифровой экономики;

б) безопасность: недопустимость использования искусственного интеллекта в целях умышленного причинения вреда гражданам и юридическим лицам, а также предупреждение и минимизация рисков возникновения негативных последствий использования технологий искусственного интеллекта;

в) прозрачность: объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей продуктов, которые созданы с использованием технологий искусственного интеллекта, к информации о применяемых в этих продуктах алгоритмах работы искусственного интеллекта;

г) технологический суверенитет: обеспечение необходимого уровня самостоятельности Российской Федерации в области искусственного интеллекта, в том числе посредством преимущественного использования отечественных технологий искусственного интеллекта и технологических решений, разработанных на основе искусственного интеллекта;

д) целостность инновационного цикла: обеспечение тесного взаимодействия научных исследований и разработок в области искусственного интеллекта с реальным сектором экономики;

е) разумная бережливость: осуществление и адаптация в приоритетном порядке существующих мер, направленных на реализацию государственной политики в научно-технической и других областях;

ж) поддержка конкуренции: развитие рыночных отношений и недопустимость действий, направленных на ограничение конкуренции между российскими организациями, осуществляющими деятельность в области искусственного интеллекта.

#### **IV. ПРИОРИТЕТНЫЕ НАПРАВЛЕНИЯ РАЗВИТИЯ И ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

20. Приоритетные направления развития и использования технологий искусственного интеллекта определяются с учетом национальных целей и стратегических задач, определенных Указом Президента Российской Федерации от 7 мая 2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года».

21. Использование технологий искусственного интеллекта в отраслях экономики носит общий («сквозной») характер и способствует созданию условий для улучшения эффективности и формирования принципиально новых направлений деятельности хозяйствующих субъектов, в том числе за счет:

а) повышения эффективности процессов планирования, прогнозирования и принятия управленческих решений (включая прогнозирование отказов оборудования и его превентивное техническое обслуживание, оптимизацию планирования поставок, производственных процессов и принятия финансовых решений);

б) автоматизации рутинных (повторяющихся) производственных операций;

в) использования автономного интеллектуального оборудования и робототехнических комплексов, интеллектуальных систем управления логистикой;

г) повышения безопасности сотрудников при выполнении бизнес-процессов (включая прогнозирование рисков и неблагоприятных событий, снижение уровня непосредственного участия человека в процессах, связанных с повышенным риском для его жизни и здоровья);

д) повышения лояльности и удовлетворенности потребителей (в том числе направление им персонализированных предложений и рекомендаций, содержащих существенную информацию);

е) оптимизации процессов подбора и обучения кадров, составления оптимального графика работы сотрудников с учетом различных факторов.

22. Использование технологий искусственного интеллекта в социальной сфере способствует созданию условий для улучшения уровня жизни населения, в том числе за счет:

а) повышения качества услуг в сфере здравоохранения (включая профилактические обследования, диагностику, основанную на анализе изображений, прогнозирование возникновения и развития заболеваний, подбор оптимальных дозировок лекарственных препаратов, сокращение угроз пандемий, автоматизацию и точность хирургических вмешательств);



б) повышения качества услуг в сфере образования (включая адаптацию образовательного процесса к потребностям обучающихся и потребностям рынка труда, системный анализ показателей эффективности обучения для оптимизации профессиональной ориентации и раннего выявления детей с выдающимися способностями, автоматизацию оценки качества знаний и анализа информации о результатах обучения);

в) повышения качества предоставления государственных и муниципальных услуг, а также снижения затрат на их предоставление.

## **V. ЦЕЛИ И ОСНОВНЫЕ ЗАДАЧИ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

23. Целями развития искусственного интеллекта в Российской Федерации являются обеспечение роста благосостояния и качества жизни ее населения, обеспечение национальной безопасности и правопорядка, достижение устойчивой конкурентоспособности российской экономики, в том числе лидирующих позиций в мире в области искусственного интеллекта.

24. Основными задачами развития искусственного интеллекта являются:

а) поддержка научных исследований в целях обеспечения опережающего развития искусственного интеллекта;

б) разработка и развитие программного обеспечения, в котором используются технологии искусственного интеллекта;

в) повышение доступности и качества данных, необходимых для развития технологий искусственного интеллекта;

г) повышение доступности аппаратного обеспечения, необходимого для решения задач в области искусственного интеллекта;

д) повышение уровня обеспечения российского рынка технологий искусственного интеллекта квалифицированными кадрами и уровня информированности населения о возможных сферах использования таких технологий;

е) создание комплексной системы регулирования общественных отношений, возникающих в связи с развитием и использованием технологий искусственного интеллекта.

25. Для выполнения указанных в пункте 24 настоящей Стратегии задач необходимы:

а) создание новых высокопроизводительных рабочих мест и повышение уровня занятости населения;

б) обеспечение конкурентоспособного уровня материального вознаграждения для специалистов в области искусственного интеллекта, создание благоприятных условий для их работы, в том числе дистанционной;

в) обеспечение необходимых условий для привлечения, в том числе из иностранных государств, лучших специалистов в области искусственного интеллекта;

г) поддержка экспорта российских продуктов (услуг), созданных (оказываемых) с использованием искусственного интеллекта, и их продвижения на мировой рынок;

д) создание стимулов для привлечения частных инвестиций в развитие корпоративной науки, научных исследований и разработок в области искусственного интеллекта;

е) формирование комплексной системы безопасности при создании, развитии, внедрении и использовании технологий искусственного интеллекта.

26. Основными механизмами развития искусственного интеллекта являются:

а) обеспечение роста предложения конкурентоспособных в мире российских продуктов (услуг), созданных (оказываемых) с использованием искусственного интеллекта;

б) обеспечение роста спроса со стороны российских граждан, организаций и государственных органов на продукты (услуги), созданные (оказываемые) с использованием искусственного интеллекта.

27. Основными показателями, характеризующими рост предложения продуктов (услуг), созданных (оказываемых) с использованием искусственного интеллекта, являются:

а) увеличение количества организаций, разрабатывающих технологические решения на основе искусственного интеллекта и занимающих лидирующие позиции в мире;

б) увеличение количества результатов интеллектуальной деятельности в области искусственного интеллекта, которые прошли государственную регистрацию либо учтены иным способом в соответствии с общепринятой мировой практикой и применяются в промышленности.

28. Показателем, характеризующим рост спроса на технологии искусственного интеллекта, является увеличение количества организаций, в том числе организаций социальной сферы, и государственных органов, использующих искусственный интеллект для повышения эффективности своей деятельности.

**Поддержка научных исследований в целях обеспечения  
опережающего развития искусственного интеллекта**

29. В целях создания и развития российских технологий искусственного интеллекта необходимо обеспечить приоритетную поддержку соответствующих фундаментальных и прикладных научных исследований.

30. Фундаментальные научные исследования должны быть направлены на создание принципиально новых научных результатов, в том числе на создание универсального (сильного) искусственного интеллекта, и решение иных задач, предусмотренных настоящей Стратегией, включая реализацию следующих приоритетов:

а) алгоритмическая имитация биологических систем принятия решений, в том числе распределенных коллективных систем, таких как пчелиный рой или муравейник;

б) автономное самообучение и развитие адаптивности алгоритмов к новым задачам;

в) автономная декомпозиция сложных задач, поиск и синтез решений.

31. Осуществление непрерывной государственной поддержки фундаментальных научных исследований в области искусственного интеллекта, прежде всего с использованием существующих механизмов оказания такой поддержки, должно быть направлено на обеспечение лидерства Российской Федерации в создании и использовании перспективных методов искусственного интеллекта.

32. Для развития фундаментальных и прикладных научных исследований в области искусственного интеллекта необходима реализация следующих мер:

а) приоритетная долгосрочная поддержка научных исследований в области искусственного интеллекта, в том числе проводимых в субъектах Российской Федерации (включая обоснованное увеличение штатной численности научного и инженерного персонала);

б) стимулирование привлечения инвестиций юридических и физических лиц в разработку технологий искусственного интеллекта;

в) реализация междисциплинарных исследовательских проектов в области искусственного интеллекта в различных отраслях экономики;

г) проведение патентных исследований и их регулярная актуализация с участием российских организаций — лидеров в реализации приоритетных направлений научных исследований в области искусственного интеллекта;

д) развитие исследовательской инфраструктуры и обеспечение доступа научных работников (исследователей) к вычислительным ресурсам, базам и наборам данных;

е) развитие международного сотрудничества Российской Федерации, включая обмен специалистами и участие отечественных специалистов в российских и международных конференциях в области искусственного интеллекта;

ж) повышение эффективности оценки научных работников (исследователей), в том числе посредством применения новых критериев результативности их деятельности (помимо научных публикаций).

33. К 2024 году основным показателем, характеризующим успешную реализацию мер по поддержке научных исследований в области искусственного интеллекта, должен стать существенный рост:

а) количества и индекса цитируемости в ведущих мировых научных изданиях научных статей российских ученых на тему, посвященную искусственному интеллекту;

б) количества зарегистрированных (учтенных) результатов интеллектуальной деятельности в области искусственного интеллекта;

в) количества разработанных на основе результатов интеллектуальной деятельности в области искусственного интеллекта прикладных технологических решений, используемых в практической деятельности.

#### **Разработка и развитие программного обеспечения, в котором используются технологии искусственного интеллекта**

34. Основными направлениями разработки и развития программного обеспечения, в котором используются технологии искусственного интеллекта, являются:

а) создание благоприятных условий для специалистов в области искусственного интеллекта, работающих с программным обеспечением, в котором используются технологии искусственного интеллекта, включая:

организацию эффективного взаимодействия специалистов в области искусственного интеллекта с организациями, по заказу которых создается программное обеспечение;

оказание финансовой поддержки молодым специалистам в области искусственного интеллекта;

обеспечение участия специалистов в области искусственного интеллекта в российских и международных конференциях и соревнованиях в этой области;

введение упрощенного режима реализации пилотных проектов, необходимых для развития технологий искусственного интеллекта;

б) обеспечение условий для создания открытых библиотек искусственного интеллекта, в том числе стимулирование (включая материаль-

ное) специалистов к участию в российских и международных проектах по их созданию;

в) признание успешного участия специалистов в создании открытых библиотек искусственного интеллекта в качестве научного достижения;

г) разработка единых стандартов в области безопасности (в том числе отказоустойчивости) и совместимости программного обеспечения, эталонных архитектур вычислительных систем и программного обеспечения, а также определение критериев сопоставления программного обеспечения и критериев эталонных открытых тестовых сред (условий) в целях определения качества и эффективности программного обеспечения.

35. К 2024 году должен существенно повыситься уровень участия российских специалистов в международном обмене знаниями, их вклад в создание открытых библиотек искусственного интеллекта. Российскими специалистами должны быть созданы широко применяемые в мире открытые библиотеки (координируемые российскими исследователями и научным сообществом) и программное обеспечение, в которых используются технологии искусственного интеллекта.

36. К 2030 году должно быть разработано программное обеспечение, в котором используются технологии искусственного интеллекта, для решения задач в различных сферах деятельности. Российские организации, разрабатывающие такое программное обеспечение, должны войти в группу лидеров на мировом рынке.

### **Повышение доступности и качества данных, необходимых для развития технологий искусственного интеллекта**

37. Основными факторами развития технологий искусственного интеллекта являются увеличение объема доступных данных, в том числе данных, прошедших разметку и структурирование, и развитие информационно-коммуникационной инфраструктуры для обеспечения доступа к наборам таких данных.

38. Основными направлениями повышения доступности и качества данных, необходимых для развития технологий искусственного интеллекта, в соответствии с требованиями законодательства Российской Федерации являются:

а) разработка унифицированных и обновляемых методологий описания, сбора и разметки данных, а также механизма контроля за соблюдением указанных методологий;

б) создание и развитие информационно-коммуникационной инфраструктуры для обеспечения доступа к наборам данных посредством:

создания (модернизации) общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных;

хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций — разработчиков в области искусственного интеллекта;

установления приоритетного доступа российских государственных органов и организаций к общедоступным платформам.

39. Для реализации настоящей Стратегии требуется создание нормативно-правовой базы, предусматривающей обеспечение защиты данных, полученных при осуществлении экономической и научной деятельности, в том числе их хранение преимущественно на территории Российской Федерации, а также установление приоритетного доступа российских государственных органов и организаций к таким данным. Публикация данных должна осуществляться в соответствии с требованиями законодательства Российской Федерации, регулирующего доступ к данным и их оборот.

40. К 2024 году российским организациям должны быть доступны наборы данных, которые соответствуют методологиям их сбора и разметки и хранятся на общедоступных платформах. При этом персональные и иные данные, доступ к которым ограничен федеральными законами, должны быть надежно защищены, их обработка должна соответствовать международным обязательствам Российской Федерации в этой области.

41. К 2030 году объем опубликованных на общедоступных платформах наборов данных должен быть достаточным для решения всех актуальных задач в области искусственного интеллекта, в том числе за счет публикации звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения.

#### **Повышение доступности аппаратного обеспечения, необходимого для решения задач в области искусственного интеллекта**

42. Основными направлениями повышения доступности аппаратного обеспечения, необходимого для решения задач в области искусственного интеллекта, являются:

а) проведение фундаментальных научных исследований, направленных на разработку перспективных архитектур вычислительных систем (в том числе нейроморфных вычислительных систем, построенных на принципе подобию биологическим нейронным системам);

б) реализация мер государственной поддержки в части, касающейся разработки отечественных высокоскоростных и энергоэффективных процессоров и других элементов вычислительных систем, в том числе на этапах проектирования и производства опытных образцов, приобретения необходимых интеллектуальных прав на элементы вычислительных систем и системное программное обеспечение;

в) реализация мер государственной поддержки в части, касающейся создания и производства программно-аппаратных комплексов с использованием преимущественно отечественной электронной компонентной базы и оптических элементов;

г) поддержка создания и развития специальных центров коллективного пользования в целях разработки прототипов перспективных элементов электронной компонентной базы, их тестирования и сборки готовых изделий;

д) поддержка создания высокопроизводительных центров обработки данных и их развития посредством обеспечения разработчикам и научным работникам (исследователям) льготного доступа к вычислительным мощностям.

43. К 2024 году должна быть создана инфраструктура поддержки отечественных организаций, осуществляющих деятельность в области искусственного интеллекта, включая создание высокопроизводительных центров обработки данных. Также должны быть разработаны российские микропроцессоры, не уступающие мировым аналогам по скорости и энергоэффективности.

44. К 2030 году на российском и международном рынках должны быть широко представлены функционирующие образцы микропроцессоров с комплектом соответствующего программного обеспечения. Должны быть открыты специализированные центры обработки данных на основе российских микропроцессоров. Интеллектуальные устройства, в которых используются такие микропроцессоры, должны быть введены в обращение на соответствующем товарном рынке. Кроме того, должны быть разработаны принципиально новые типы архитектур вычислительных систем и зарегистрированы интеллектуальные права на них.

**Повышение уровня обеспечения российского рынка технологий  
искусственного интеллекта квалифицированными кадрами  
и уровня информированности населения о возможных сферах  
использования таких технологий**

45. Основными направлениями повышения уровня обеспечения российского рынка технологий искусственного интеллекта квалифицированными кадрами и уровня информированности населения о возможных сферах использования таких технологий являются:

а) разработка и внедрение образовательных модулей в рамках образовательных программ всех уровней образования, программ повышения квалификации и профессиональной переподготовки для получения гражданами знаний, приобретения ими компетенций и навыков в области математики, программирования, анализа данных, машинного обучения, способствующих развитию искусственного интеллекта. При этом в целях развития перспективных методов искусственного интеллекта приоритетное значение приобретает конвергентное знание, обеспечиваемое в том числе за счет интеграции математического, естественно-научного и социально-гуманитарного образования;

б) привлечение организаций, осуществляющих деятельность в области искусственного интеллекта, к участию в мероприятиях, направленных на развитие общего и профессионального образования;

в) повышение качества математического и естественно-научного образования обучающихся (в рамках как основных, так и дополнительных образовательных программ), его интеграция с социально-гуманитарным образованием, создание условий для привлечения обучающихся к углубленной подготовке по этим направлениям;

г) увеличение количества и повышение привлекательности конкурсов и олимпиад, направленных на развитие интеллектуальных и творческих способностей обучающихся;

д) стимулирование (в том числе материальное) работодателей к принятию мер, направленных на приобретение сотрудниками компетенций в области искусственного интеллекта и в смежных областях его использования;

е) создание благоприятных условий для привлечения ведущих российских специалистов, проживающих за рубежом, и иностранных специалистов мирового уровня к работе в Российской Федерации, включая:

обеспечение конкурентоспособного уровня заработной платы и условий для самореализации таких специалистов на всей территории Российской Федерации;

обеспечение простоты и удобства соблюдения иностранными специалистами требований российского трудового и миграционного законода-



тельства, в том числе при получении гражданства Российской Федерации и разрешений на работу;

ж) информирование населения и организаций о преимуществах и безопасности применения технологических решений, разработанных на основе искусственного интеллекта, а также о доступности программ обучения и переобучения.

46. К 2024 году должно существенно увеличиться число граждан, имеющих компетенции в области искусственного интеллекта и в смежных областях его использования, в том числе аспирантов и специалистов в области искусственного интеллекта, имеющих ученую степень. Российская Федерация должна стать привлекательной для трудоустройства квалифицированных специалистов в области искусственного интеллекта, в том числе в связи с высоким уровнем заработной платы и созданием благоприятных условий для работы.

47. К 2030 году в России должны реализовываться образовательные программы мирового уровня для подготовки высококвалифицированных специалистов и руководителей в области искусственного интеллекта. Российские образовательные организации высшего образования должны занимать лидирующие позиции в мире по направлениям в области искусственного интеллекта. Дефицит специалистов в этой области должен быть устранен, в том числе за счет привлечения ведущих иностранных специалистов, имеющих ученую степень.

**Создание комплексной системы регулирования общественных отношений, возникающих в связи с развитием и использованием технологий искусственного интеллекта**

48. Для стимулирования развития и использования технологий искусственного интеллекта необходимы адаптация нормативного регулирования в части, касающейся взаимодействия человека с искусственным интеллектом, и выработка соответствующих этических норм. При этом избыточное регулирование в этой сфере может существенно замедлить темп развития и внедрения технологических решений.

49. Основными направлениями создания комплексной системы регулирования общественных отношений, возникающих в связи с развитием и внедрением технологий искусственного интеллекта, являются:

а) обеспечение благоприятных правовых условий (в том числе посредством создания экспериментального правового режима) для доступа к данным, преимущественно обезличенным, включая данные, собираемые государственными органами и медицинскими организациями;

б) обеспечение особых условий (режимов) для доступа к данным, включая персональные, в целях проведения научных исследований, создания технологий искусственного интеллекта и разработки технологических решений на их основе;

в) создание правовых условий и установление процедур упрощенного тестирования и внедрения технологических решений, разработанных на основе искусственного интеллекта, а также делегирования информационным системам, функционирующим на основе искусственного интеллекта, возможности принятия отдельных решений (за исключением решений, которые могут ущемлять права и законные интересы граждан), в том числе при исполнении государственными органами государственных функций (за исключением функций, направленных на обеспечение безопасности населения и государства);

г) устранение административных барьеров при экспорте продукции (работ, услуг) гражданского назначения, созданной на основе искусственного интеллекта;

д) создание единых систем стандартизации и оценки соответствия технологических решений, разработанных на основе искусственного интеллекта, развитие международного сотрудничества Российской Федерации по вопросам стандартизации и обеспечение возможности сертификации продукции (работ, услуг), созданной на основе искусственного интеллекта;

е) стимулирование привлечения инвестиций посредством совершенствования механизмов совместного участия инвесторов и государства в проектах, связанных с разработкой технологий искусственного интеллекта, а также предоставления целевой финансовой поддержки организациям, осуществляющим деятельность по развитию и внедрению технологий искусственного интеллекта (при условии, что внедрение таких технологий повлечет за собой существенные позитивные эффекты для отраслей экономики Российской Федерации);

ж) разработка этических правил взаимодействия человека с искусственным интеллектом.

50. К 2024 году должны быть созданы необходимые правовые условия для достижения целей, решения задач и реализации мер, предусмотренных настоящей Стратегией.

51. К 2030 году в Российской Федерации должна функционировать гибкая система нормативно-правового регулирования в области искусственного интеллекта, в том числе гарантирующая безопасность населения и направленная на стимулирование развития технологий искусственного интеллекта.

## VI. МЕХАНИЗМЫ РЕАЛИЗАЦИИ НАСТОЯЩЕЙ СТРАТЕГИИ

52. Реализация настоящей Стратегии обеспечивается согласованными действиями федеральных органов государственной власти, органов государственной власти субъектов Российской Федерации, государственных органов, органов местного самоуправления, государственных академий наук, научных и образовательных организаций, фондов поддержки научной, научно-технической и инновационной деятельности, общественных организаций, предпринимательского сообщества, государственных корпораций, государственных компаний и акционерных обществ с государственным участием.

53. Координацию деятельности участников реализации настоящей Стратегии осуществляет Правительственная комиссия по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности.

54. Для координации деятельности бизнес-сообщества и научных организаций по реализации настоящей Стратегии создается объединение, в которое входят представители организаций, осуществляющих деятельность по развитию и внедрению технологий искусственного интеллекта.

55. Финансовое обеспечение реализации настоящей Стратегии осуществляется за счет средств бюджетов бюджетной системы Российской Федерации, средств государственных внебюджетных фондов и внебюджетных источников, включая средства институтов развития, государственных корпораций, государственных компаний, акционерных обществ с государственным участием и частные инвестиции.

56. Правительство Российской Федерации при участии объединения, указанного в пункте 54 настоящей Стратегии, в том числе в рамках федерального проекта «Искусственный интеллект» национальной программы «Цифровая экономика Российской Федерации», осуществляет:

а) разработку и утверждение плана мероприятий по реализации настоящей Стратегии, предусматривающего в том числе:

перечень показателей эффективности реализации настоящей Стратегии, включая показатели, установленные пунктами 27, 28, 33, 35, 36, 40, 41, 43, 44, 46, 47, 50 и 51 настоящей Стратегии, и методологию их расчета;

целевые значения показателей эффективности реализации настоящей Стратегии на 2024 и 2030 годы;

мероприятия по достижению целей и выполнению задач, предусмотренных настоящей Стратегией;

описание рисков реализации настоящей Стратегии и способов их минимизации;

- б) координацию деятельности по реализации настоящей Стратегии;
- в) мониторинг реализации настоящей Стратегии.

57. В целях аналитической поддержки реализации настоящей Стратегии проводятся научные исследования, направленные на прогнозирование развития технологий искусственного интеллекта, а также на прогнозирование социальных и этических аспектов их использования. Результаты этих исследований должны учитываться при принятии управленческих решений.

58. Одним из механизмов реализации настоящей Стратегии является выполнение комплексных научно-технических программ полного инновационного цикла, предусмотренных Стратегией научно-технологического развития Российской Федерации.

59. Корректировка настоящей Стратегии осуществляется по решению Президента Российской Федерации каждые три года на основании предложений, подготовленных Правительством Российской Федерации при участии объединения, указанного в пункте 54 настоящей Стратегии, с учетом результатов мониторинга ее реализации и динамики развития искусственного интеллекта. Результаты мониторинга реализации настоящей Стратегии и предложения по ее корректировке отражаются в совместном экспертно-аналитическом докладе Правительства Российской Федерации и указанного объединения.

---

---

# СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

---

---

## ДОКУМЕНТЫ

(законы, указы Президента РФ, постановления и распоряжения  
Правительства РФ, приказы, ГОСТы, письма и др.)

1. Федеральный закон от 24.04.2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных».

2. Федеральный закон от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации».

3. Указ Президента РФ от 10 октября 2019 г. № 490. «Национальная стратегия развития искусственного интеллекта до 2030 года в Российской Федерации».

4. Указ Президента РФ от 01.12.2016 № 642 «О Стратегии научно-технологического развития Российской Федерации» // Собрание законодательства РФ, 05.12.2016, № 49, ст. 6887.

5. Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы» // Собрание законодательства РФ, 15.05.2017, № 20, ст. 2901.

6. Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (а также «Национальная стратегия развития искусственного интеллекта на период до 2030 года»).

7. Постановление Правительства РФ от 30.04.2019 № 538 «О мерах Государственной поддержки создания научных центров мирового уровня».

8. Распоряжение Правительства РФ от 9 апреля 2016 г. № 637-р. Концепция преподавания русского языка и литературы в РФ.

9. Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р. Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года.

10. Приказ Минобрнауки России от 6.10.2009 г. № 373 «Об утверждении и введении в действие федерального государственного образовательного стандарта начального общего образования». <http://base.garant.ru/197127/> (актуально на 3.11.2019).

11. Приказ Росстандарта от 1 ноября 2019 г. № 2612 «Об утверждении Программы национальной стандартизации на 2020 год».

12. Приказ Росстандарта от 25 июля 2019 г. № 1732 «О создании технического комитета по стандартизации «Искусственный интеллект».

13. Инструктивно-методическое письмо. Об использовании микрокалькуляторов в учебном процессе. НИИ содержания и методов обучения АПН СССР и Главное управление школ Министерства просвещения СССР // Математика в школе, 1982. № 3. С. 6–8.

14. Паспорт федерального проекта «Нормативное регулирование цифровой среды». URL: [https://www.economy.gov.ru/material/directions/gosudarstvennoe\\_upravlenie/normativnoe\\_regulirovanie\\_cifrovoy\\_sredy/](https://www.economy.gov.ru/material/directions/gosudarstvennoe_upravlenie/normativnoe_regulirovanie_cifrovoy_sredy/)

15. Письмо Минпроса РСФСР от 25.04.1958 № 81-М О пользовании авторучками учащимися общеобразовательных школ. [http://gospismo.ru/library/pisma-ussr/sssrpismo\\_32/](http://gospismo.ru/library/pisma-ussr/sssrpismo_32/)

16. Постановление ЦК КПСС и Совета Министров СССР от 28 марта 1985 года № 271 «О мерах по обеспечению компьютерной грамотности учащихся средних учебных заведений и широкого внедрения электронно-вычислительной техники в учебный процесс» // Вопросы образования. 2005. № 3. С. 341–346.

17. Концепция развития технологий машиночитаемого права. URL: [https://www.economy.gov.ru/material/news/v\\_pravitelstve\\_utverdili\\_koncepciyu\\_razvitiya\\_tehnologiy\\_mashinochitaemogo\\_prava.html](https://www.economy.gov.ru/material/news/v_pravitelstve_utverdili_koncepciyu_razvitiya_tehnologiy_mashinochitaemogo_prava.html)

18. Artificial Intelligence, Robotics, Privacy and Data Protection.

19. Autonomous Vehicles — Self-Driving Vehicles Enacted Legislation // National Conference of State Legislatures. 2018. URL: <http://www.ncsl.org/research/>

20. Commissioners. 2016. URL: [https://edps.europa.eu/data-protection/our-work/publications/other-documents/artificial-intelligence-robotics-privacy-and\\_en](https://edps.europa.eu/data-protection/our-work/publications/other-documents/artificial-intelligence-robotics-privacy-and_en)

21. Communication from the European Commission of 07.12.2018 // Coordinated Plan on Artificial Intelligence. 2018. COM 795. P. 2.

22. European Commission. Workshop on liability in the area of autonomous systems and advanced robots and Internet of Things systems. 2017. URL: <https://ec.europa.eu/digital-single-market/en/news/workshop-liability-area-autonomous-systems-and-advanced-robots-and-internet-things-systems>.

23. European Parliament Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) // European Parliament. URL: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN>

24. European Parliament. Reports A8-0005/2017 of 27.01.2017 with recommendation to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). — P. 15.

25. NIST (2019) U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools: [https://www.nist.gov/system/files/documents/2019/08/10/ai\\_standards\\_fedengagement\\_plan\\_9aug2019.pdf](https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf).

26. Законопроект от 7 октября 2020 г. (Boletín № 13.828–19 «Proyecto de ley, iniciado en moción de los Honorables Senadores señor Girardi, señora Goic, y señores Chahuán, Coloma y De Urresti, sobre protección de los neuroderechos y la integridad mental, y el desarrollo de la investigación y las neurotecnologías»). URL: <https://www.camara.cl/verDoc.aspx?prmID=14152&prmTIPO=INICIATIV; https://www.senado.cl/defensa-de-los-neuroderechos-una-tarea-para-los-parlamentos-a-nivel-global/senado/2020-10-07/132033.html>

### КНИГИ, СБОРНИКИ, УЧЕБНИКИ

27. 10 лет — Научному совету РАН по методологии искусственного интеллекта [Текст] Материалы симпозиума, 26 марта 2015 г., Институт философии РАН. М.: ИИнтелл, 2016.

28. 150 лет «Рефлексам головного мозга» [Текст] Сборник научных трудов, посвященных изданию статьи И.М. Сеченова (23 ноября 1863 г.) — М.: ИИнтелл, 2014. 432 с.

29. Алан М. Тьюринг. Может ли машина мыслить? [Текст] (Под ред. Б.В. Бирюкова). М., 1960.

30. Алексеев А.Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты [Текст]. М.: ИИнтелл, 2013. 304 с.

31. Арутюнова К.Р., Александров Ю.И. Мораль и субъективный опыт [Текст]. М.: Институт психологии РАН, 2019.

32. Бикертон Д. Язык Адама: Как люди создали язык. Как язык создал людей [Текст]. М.: Языки славянских культур, 2012.
33. Васильев С.Н., Жерлов А.К., Федосов Е.А., Федунов Б.Е. Интеллектуальное управление динамическими системами [Текст]. М.: Физматлит, 2000. 352 с.
34. Выготский Л.С. Мышление и речь [Текст]. М.: Лабиринт, 1999.
35. Гибсон Дж. Экологический подход к зрительному восприятию. М.: Прогресс, 1988.
36. Гройс Б.Е. Под взглядом теории [Текст]. М.: Ad Marginem, 2013.
37. Гусейнов А.А. Великие моралисты [Текст]. М., 1995.
38. Гутнер Р.С., Полунов Ю.Л. Августа Ада Лавлейс и возникновение программирования. Кибернетика и логика. Математико-логические аспекты становления идей кибернетики и развития вычислительной техники [Текст]. М.: Наука, 1978. 334 с.
39. Давыдов В.В. Лекции по общей психологии [Текст]. М.: Академия, 2008.
40. Данбар Р. Лабиринт случайных связей: Рассказ о том, как мы общаемся, а главное — зачем [Текст]. М.: ЛомоносовЪ, 2012.
41. Дубровский Д.И. Информация, сознание, мозг [Текст]. М.: Высшая школа, 1980. 286 с.
42. Дубровский Д.И. Проблема «Сознание и мозг». Теоретическое решение [Текст]. М.: Канон+, 2015. 208 с.
43. Дубровский Д.И. Проблема идеального. Субъективная реальность [Текст]. М.: Канон+, 2002. 368 с.
44. Дубровский Д.И. Проблема сознания: Теория и критика альтернативных концепций [Текст]. М.: ЛЕНАНД, 2019. С. 140–183.
45. Дубровский Д.И. Психические явления и мозг. Философский анализ проблемы в связи актуальными задачами нейрофизиологии, психологии и кибернетики [Текст]. М.: Наука, 1971. 386 с. (Второе, доп. издание. М.: ЛЕНАНД, 2021. 400 с.).
46. Ершов Ю.Л. Определимость и вычислимость [Текст]. 2-е изд., испр. и доп. Новосибирск: Науч. кн., 2000. 318 с.
47. Ершов Ю.Л., Целищев В. В. Алгоритмы и вычислимость в человеческом познании [Текст]. Новосибирск: Изд-во Сибирского отд. Российской акад. наук, 2012.
48. Залоило М.В. Искусственный интеллект в праве [Текст]. М.: Инфортропик Медиа, 2021.



49. *Зинченко В.П.* Сознание и творческий акт [Текст]. М.: Языки славянских культур, 2010.
50. *Зорина З.А., Смирнова А.А.* О чем рассказали «говорящие» обезьяны: способны ли высшие животные оперировать символами? [Текст]. М.: Языки славянских культур, 2006.
51. *Ильенков Э.В.* Философия и культура [Текст]. М.: Политиздат, 1992.
52. Искусственный интеллект: междисциплинарный подход [Текст] / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИнтелл, 2006. 448 с.
53. *Калинин А.Ю., Любимов А.П.* Социология процессов правообразования в современной России [Текст]. Коломна: Государственный социально-гуманитарный университет, 2016. 355 с.
54. *Каляев И.А., Гайдук А.Р., Капустян С.Г.* Модели и алгоритмы коллективного управления в группах роботов. М.: Физматлит, 2009. 278 с.
55. *Кузнецов О.П.* Избранные труды. Автоматы, языки и искусственный интеллект [Текст]. М.: Наука, 2016. 464 с.
56. *Лекторский В.А.* Субъект, объект, познание [Текст]. М., 1980.
57. *Леонтьев А.Н.* Деятельность, сознание, личность [Текст]. М.: Смысл, 2005.
58. *Литвак Л.М.* «Жизнь после смерти»: предсмертные переживания и природа психоза. Опыт самонаблюдения и психоневрологического исследования [Текст] / Под ред. и со вступительной статьей Д.И. Дубровского. М.: Канон+, 2007.
59. *Лотман Ю.М.* Мозг — текст — культура — искусственный интеллект / Избр. статьи: В 3 т. Т. I: Статьи по семиотике и типологии культуры [Текст]. Таллинн: Александра, 1992. С. 25–33.
60. *Лурия А.Р.* Лекции по общей психологии [Текст]. СПб.: Изд. дом «Питер», 2006.
61. *Любимов А.П.* Конституция Российской Федерации. В таблицах и схемах. Учебное пособие. 3-е изд. [Текст]. М.: Экзамен, 2006. 176 с.
62. *Любимов А.П.* Философия права [Текст]. М.: Юрайт, 2019. 257 с.
63. *Любимов А.П., Пономарева Д.В., Барабашев А.Г.* Основные понятия искусственного интеллекта [Текст]. М.: ООО «Сам Полиграфист», 2019. 116 с.
64. *Мамардашвили М.К.* Картезианские размышления // Философские чтения [Текст]. СПб.: Азбука-классика, 2002.
65. *Маркус Габриэль.* Я не есть мозг. Философия духа для XXI века [Текст]. М.: URSS, 2020.

66. Меграбян А.А. Деперсонализация [Текст]. Ереван: Армгосиздат, 1962.
67. Мейер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живём, работаем и мыслим [Текст]. М.: Изд-во «Манн, Иванов и Фербер», 2014.
68. Мерло-Понти М. Феноменология восприятия [Текст]. СПб.: Ювента, 1999.
69. Михайлов Ф.Т. Загадка человеческого Я. [Текст]. М.: Политиздат, 1976.
70. Муссеев Н.Н. Судьба цивилизации. Путь разума [Текст]. М.: Языки русской культуры, 2000.
71. Монтень М. Опыты [Текст]. М.-Л.: Изд-во АН СССР, 1960. Кн. I. Гл. XXVI «О воспитании детей». С. 189–212. Оригинал: Michel de Montaigne, Les Essais.
72. Морхат П.М. Искусственный интеллект: правовой взгляд [Текст]. М.: Буки Веди, 2017.
73. Морхат П.М. Право и искусственный интеллект: монография [Текст] / Под ред. И.В. Понкина. М.: ЮНИТИ-ДАНА, 2017.
74. Назлоян Г.М. Концептуальная психотерапия. Портретный метод: монография [Текст]. М.: ПЕР СЭ, 2002.
75. Новая парадигма развития России в XXI веке. Комплексные исследования проблем устойчивого развития: идеи, результаты [Текст] / Под ред. В.А. Коптюга, В.М. Матросова, В.К. Левашова. М.: Изд-во «Academia», 2000. 397 с.
76. Панов Е.Н. Человек — созидатель и разрушитель. Эволюция поведения и социальной организации [Текст]. М.: Издательский дом ЯСК, 2017.
77. Пенроуз Р. Тени разума [Текст]. М.: Институт компьютерных исследований, 2003. С. 212.
78. Пенроуз Р. Новый ум короля [Текст]. М.: URSS, 2002.
79. Петренко В.Ф., Супрун А.П. Методологические пересечения психосемантики сознания и квантовой физики. Нестор-История [Текст]. М.-СПб., 2017.
80. Платон. Сочинения в 4-х томах [Текст]. М.: Мысль, 1971.
81. Правосубъектность: общетеоретический, отраслевой и международно-правовой анализ: сборник материалов к XII Ежегодным научным чтениям памяти С.Н. Братуся [Текст]. М.: ИЗиСП, Статут, 2017.
82. Прокофьева Т.А., Гончаренко С.С., Семенов Н.Н., Элларян А.С. Стратегическая доктрина развития транспорта и формирования интегри-

рованных транспортно-логистических систем в регионах Европейского Севера, Сибири и Дальнего Востока [Текст]. М.: ОАО «ИТКОР», 2020. 226 с.

83. Проблемные регионы ресурсного типа: азиатская часть России [Текст]. Новосибирск: ИЭОПП СО РАН, ИИ СО РАН, ИСЭМ СО РАН, 2005. 377 с.

84. *Пятигорский А.М.* Непрерываемый разговор [Текст]. СПб.: Азбука классика, 2004.

85. *Рамачандран В.С.* Мозг рассказывает. Что делает нас людьми [Текст]. М.: Карьера Пресс, 2014.

86. *Рассел С., Норвиг П.* Искусственный интеллект: современный подход. 2-е изд., перераб. [Текст]. М.: Вильямс, 2006. 1408 с.

87. *Селезнев Г.Н., Христенко В.Б., Залиханов М.Ч., Львов Д.С., Матросов В.М., Гранберг А.Г., Левашов В.К., Урсул А.Д., Шелехов А.М.* Научная основа стратегии устойчивого развития Российской Федерации / Под общ. ред. М.Ч. Залиханова, В.М. Матросова, А.М. Шелехова [Текст]. М.: Изд-во ГосДумы РФ, 2002. 392 с.

88. *Сёрл Дж.* Открывая сознание заново [Текст]. М.: Идея-Пресс, 2002.

89. Транспортно-промышленное освоение Сибири и Дальнего Востока и приграничных территорий — системный фактор прорывного развития экономики России, ее интеграции в мировую систему («Горизонт-2030»). Ч. II: Аналитический доклад Байкальского экономического форума / Под науч. ред. С.Н. Васильева, В.И. Суслова, В.А. Персианова, С.С. Гончаренко, Т. Н. Есиковой [Текст]. М.-Новосибирск: ИПУ РАН, ИЭОПП СО РАН, ЕАТИЦ. 2006. 192 с.

90. *Редько В.Г.* Моделирование когнитивной эволюции: на пути к теории эволюционного происхождения мышления. 2-е изд. [Текст]. М.: ЛЕНАНД, 2019. 264 с.

91. *Рубинштейн С.Л.* Бытие и Сознание. Человек и Мир [Текст]. СПб.: Изд. дом «Питер», 2003.

92. *Серр Мишель.* Девочка с пальчик [Текст]. М.: Ад Маргинем Пресс, 2016. (Оригинал: *Serres M. Petite Poucette. Éditions Le Pommier, Paris, 2012.*)

93. Сильный искусственный интеллект. На подступах к сверхразуму [Текст] / Александр Ведяхин и др. М.: Интеллектуальная Литература, 2021. 232 с.

94. *Стёпин В.С.* Теоретическое знание [Текст]. М.: Прогресс-Традиция, 2000.

95. *Тапскотт Дон.* Электронно-цифровое общество. Плюсы и минусы эпохи сетевого интеллекта [Текст]. М., 1999.

96. Тест Тьюринга. Роботы. Зомби [Текст] / Под ред. А.Ю. Алексеева. М.: МИЭМ, 2006. 120 с.

97. Уайтхед Альфред Н. Избранные работы по философии [Текст]. М.: Прогресс, 1990.
98. Ушаков Д.В. Психология интеллекта и одаренности [Текст]. М.: Изд-во «Институт психологии РАН», 2011.
99. Финн В.К. Интеллект, информационное общество, гуманитарное знание и образование [Текст]. М.: ЛЕНАНД, 2021.
100. Финн В.К. Искусственный интеллект: методология, применения, философия [Текст]. М.: URSS, Красанд, 2011.
101. Фитч У.Т. Эволюция языка [Текст]. М.: Языки славянских культур, 2013.
102. Фролов И.Т. Прогресс науки и будущее человека [Текст]. М.: Политиздат, 1975.
103. Хокинс Дж. Об интеллекте [Текст]. М.: Вильямс, 2007. 240 с.
104. Целищев В.В. Алгоритмизация мышления: геделевский аргумент [Текст] (1-е издание. Новосибирск: Параллель, 2005; 2-е издание. М.: URSS, 2021).
105. Чалмерс Д. Сознательный ум. В поисках фундаментальной теории [Текст]. URSS, Либроком, 2013.
106. Черниговская Т.В. Чеширская улыбка кота Шрёдингера: мозг, язык и сознание [Текст]. М.: АСТ, 2021.
107. Швырёв В.С. Научное познание как деятельность [Текст]. М.: Политиздат, 1984.
108. Шелепин Ю.Е. Введение в нейроиконику [Текст]. СПб.: Троицкий мост, 2017.
109. Щедровицкий Г.П. Мышление-Понимание-Рефлексия [Текст]. М.: Изд-во «Наследие ММК», 2005.
110. Юридическая концепция роботизации: монография [Текст] / Отв. ред. Ю.А. Тихомиров, С.Б. Нанба. М.: Проспект, 2019.

## СТАТЬИ

111. Александров Ю.И. От теории функциональных систем к системной психофизиологии // Психология сегодня: теория, образование, практика / Под ред. А.Л. Журавлева, Е.А. Сергиенко, А.В. Карпова. М.: Институт психологии РАН, 2009. С. 13–56.
112. Алексеев А.Ю. Возможности искусственного интеллекта: можно ли пройти тесты Тьюринга? // Искусственный интеллект: междисциплинарный подход / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИнтелл, 2006. С. 223–242.

113. *Алексеев А.Ю.* Принципы креативной работотехники // *Технологос*. 2020. № 4. С. 5–17.
114. *Алексеев А.Ю.* Протонейрокомпьютер Корсакова // *Нейрокомпьютеры: разработка и применение*. 2013. № 7. С. 6–17.
115. *Анохин К.В.* Когнитом: в поисках фундаментальной нейронаучной теории сознания // *Журнал высшей нервной деятельности*. 2021. Т. 71. № 1. С. 39–71.
116. *Аришинов В.И., Буданов В.Г.* Парадигма сложности и социогуманитарные проекции конвергентных технологий // *Вопросы философии*. 2016. № 1. С. 59–70.
117. *Бабкин В.В., Промоненков В.К., Овчаренко М.М.* и др. Инновационная концепция средств защиты растений в Российской Федерации // *Химическая промышленность сегодня*. 2017. № 8. С. 50–54.
118. *Васильев С.Н.* Задача — озадачиваться. Искусственный интеллект научат ставить цели // *Поиск*. 13 декабря. № 50. 2019. С. 10.
119. *Васильев С.Н.* Формализация знаний и управление на основе позитивно-образованных языков // *Информационные технологии и вычислительные системы*. 2008. № 1. С. 3–19.
120. *Васильев С.Н., Галяев А.А.* Логико-оптимизационный подход в задачах преследования группы целей // *ДАН*. 2017. Т. 474. № 6. С. 1–7.
121. *Васильев С.Н., Галяев А.А., Рубинович Е.Я.* Интеллектуальное управление мобильными средствами в конфликтных средах // *Материалы Международной конференции «Проблемы механики и управления»*. Махачкала, 2018. С. 117–119.
122. *Васильев С.Н., Гончаренко С.С.* От природно-техногенных и других угроз к устойчивому развитию // *Научные труды Вольного экономического общества России*. 2020. № 3 (223). С. 203–212.
123. *Васильев С.Н., Кудинов Ю.И., Пащенко Ф.Ф.* и др. Intelligent Control Systems and Fuzzy Controllers, 1. Fuzzy Models, Logical-Linguistic and Analytical Regulators // *J. Automation and Remote Control*. 2020. Vol. 81. N 1. С. 171–191.
124. *Вдовина А.А.* Понятие «технологический уклад» в системе экономических категорий и новые технологические уклады общественного развития // *Креативная экономика*. 2019. Т. 13. № 4.
125. *Ведута Е.Н., Любимов А.П., Джакубова Т. Н., Ряскова Е.С.* Концепция национальной программы создания проактивного искусственного интеллекта // *Представительная власть — XXI век*. 2019. № 4. С. 22–29.

126. Воз Л. Решение некоторых открытых проблем с помощью программы для автоматического доказательства теорем // Киберн. сб. Нов. сер. Вып. 21. М.: Мир, 1984. С. 235–263.
127. Войниканис Е.А., Семенова Е.В., Тюляев Г.С. Искусственный интеллект и право: вызовы и возможности самообучающихся алгоритмов // Вестник ВГУ. Серия: Право. 2018. № 4. С. 137.
128. Выготский Л.С. Инструментальный метод в психологии. Собр. соч.: В 6 т. Т. 1. М.: Педагогика, 1982. [http://elibr.gnpbu.ru/text/vygotsky\\_ss-v-6tt\\_t1\\_1982/go,108;fs,1/](http://elibr.gnpbu.ru/text/vygotsky_ss-v-6tt_t1_1982/go,108;fs,1/)
129. Гёдель К. Некоторые основные теоремы в основаниях математики и их следствия // Я. Хинтиikka, О Гёделе, К. Гёдель. Статьи. М.: Канон+, 2014. С. 177.
130. Глазьев С.Ю. Современная теория длинных волн в развитии экономики // Экономическая наука современной России. 2012. № 2. С. 30.
131. Глазьев С.Ю. О неравномерности современного экономического роста как процесса развития и смены технологических укладов // Социология. 2013. № 4. С. 42–53.
132. Гончаренко С.С., Персианов В.А. Роль социальной инфраструктуры в экономике России. М.: НП «Центр стратегического партнерства», 2013. 339 с.
133. Гончаров С.С., Свириденко Д.И. Проблемы цифровизации и семантическое моделирование // Материалы XII Мультиконференции по проблемам управления (МКПУ-2019). Ростов н/Д; Таганрог: Изд-во ЮФУ, 2019. Т. 1. С. 16–19.
134. Дамасио А. Так начинается «я». Мозг и возникновение сознания. М.: Карьера-Пресс, 2018. 384 с.
135. Дубровский Д.И. Актуальные аспекты проблемы интерсубъективности // Естественный и искусственный интеллект: Методологические и социальные проблемы / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: Канон+, 2011. С. 129–148.
136. Дубровский Д.И. Мозг и психика: О необоснованности философского отрицания психофизиологической проблемы // Вопросы философии, 1968. № 8. С. 119–135.
137. Дубровский Д.И. Субъективная реальность // Философская электронная энциклопедия. М.: Институт философии РАН, 2020.
138. Дубровский Д.И. Физиологическое и логическое // Вопросы философии. 1966. № 8.
139. Дэвис Э. Техногнозис: мир, магия и мистицизм в информационную эпоху. Екатеринбург: Ультра; Культура, 2008.

140. *Ершов А.П.* Программирование — вторая грамотность. Русская версия доклада на 3-й Всемирной конференции ИФИП и ЮНЕСКО по применению ЭВМ в обучении в Лозанне, Швейцария, 1981. [http://ershov.iis.nsk.su/ru/second\\_literacy/article](http://ershov.iis.nsk.su/ru/second_literacy/article) // Проблемы информатики. 2015. № 4. С. 71–85.

141. *Золян С.Т., Ильин М.В., Сладкевич Ж.Р., Тульчинский Г.А.* Смысл в жизни и смысл в тексте // Слово.ру: балтийский акцент. 2020. Т. 11. № 1. С. 7–33.

142. *Зорькин В.Д.* Справедливость — императив цивилизации права // Вопросы философии. 2019. № 1. С. 6.

143. *Каляев И.А.* Камо грядеши? // Экономические стратегии. ИЭС. 2019. Т. 21. № 5 (163). С. 6–15.

144. *Каплан А.Я.* Мы — это больше, чем наш мозг. В поисках субъективного начала // Труды кафедры богословия Санкт-Петербургской духовной академии. Теология и современные исследования сознания. 2019. № 2 (4). С. 25–34.

145. *Кашкин С.Ю., Покровский А.В.* Искусственный интеллект, робототехника и защита прав человека в Европейском союзе // Вестник Университета им. Кутафина (МГЮА). 2019. № 4 (56). С. 78.

146. *Козинцев А.Г.* Происхождение и ранняя история вида *Homo sapiens*: новые биологические данные // Фундаментальные проблемы археологии, антропологии и этнографии Евразии. К 70-летию академика А.П. Деревянко. Новосибирск: Изд-во ИАЭТ СО РАН, 2013. С. 519–535.

147. *Кузнецов О.П.* Искусственный интеллект и когнитивные науки // Информационно-измерительные и информационные системы. 2013. Т. 11, № 5. С. 16–24.

148. *Кузнецов О.П., Базенков Н.И., Болдышев Б.А.* и др. Асинхронная дискретная модель химических взаимодействий в простых нейронных системах // Искусственный интеллект и принятие решений. 2018. № 2. С. 3–20.

149. *Лекторский В.А.* Возможно ли пост-человеческое будущее? // Философия, познание, культура. М.: Канон-плюс, 2012.

150. *Лекторский В.А.* Исследование интеллектуальных процессов в современной когнитивной науке: философские проблемы // Естественный и искусственный интеллект / Под ред. Д.И. Дубровского, В.А. Лекторского. М.: Канон+, 2011. С. 3–16.

151. *Лекторский В.А.* Философия, искусственный интеллект и когнитивная наука // Искусственный интеллект: Междисциплинарный подход / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИнтелЛ, 2006. С. 13–21.

152. Литвак Л.М. «Жизнь после смерти»: предсмертные переживания и природа психоза. Опыт самонаблюдения и психоневрологического исследования / Под ред. и со вступительной статьей Д.И. Дубровского. М.: Канон+, 2007.

153. Лотман Ю.М. О семиосфере // Структура диалога как принцип работы семиотического механизма / Под ред. Ю.М. Лотмана и др. Тарту, 1984. (Труды по знаковым системам. Т. 17; Учен. зап. Тарт. гос. ун-та. Вып. 641). С. 5–23.

154. Луман Н. Теория общества // Теория общества: сб. М.: КАНОН-пресс Ц, 1999. С. 196–235.

155. Любимов А.П. Жорес Алферов — легенда мировой науки: от полупроводников до искусственного интеллекта // Представительная власть — XXI век. 2019. № 1-2. С. 38.

156. Любимов А.П. 20 лет Конституции РФ // Представительная власть — XXI век. 2013. № 7-8. С. 14–17.

157. Любимов А.П. Качество подготовки законопроекта субъектом права законодательной инициативы // Российская юстиция. 2006. № 2. С. 47–52.

158. Любимов А.П. Конституция Российской Федерации: вопросы и проблемы толкования // Представительная власть — XXI век. 2003. № 6. С. 12–14.

159. Любимов А.П. Основные подходы к определению понятия «искусственный интеллект» // Научно-техническая информация. Серия 2: Методика и организация информационной работы. 2020. № 9. С. 1–6.

160. Любимов А.П. Политико-правовая основа электронной демократии и культуры в России // Философские науки. 2017. № 2. С. 79–88.

161. Любимов А.П. Правовые основы использования информационных технологий в России // Информационные и телекоммуникационные технологии. 2016. № 31. С. 22–28.

162. Любимов А.П., Пономарева Д.В., Барабашев А.Г. Искусственный интеллект в европейской правовой доктрине // Вестник Дипломатической академии МИД России. Международное право. 2019. № 1 (10). С. 32–47.

163. Любимов А.П., Пономарева Д.В., Барабашев А.Г. К вопросу о понятии искусственного интеллекта в российском праве // Актуальные вопросы экономики, управления и права: сборник научных трудов (ежегодник). 2019. № 2-3. С. 16–34.

164. Любимов А.П., Пономарева Д.В., Барабашев А.Г. О национальной стратегии развития искусственного интеллекта // Представительная власть — XXI век. 2019. № 5-6. С. 1–7.



165. Любимов А.П., Черный В.В. Эволюция глобализма: от компьютеризации — к электронной демократии и цифровой экономике знаний // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2021. № 3. С. 1–7.

166. Любимов А.П., Щитов А.Н. Современные научно-технологические приоритеты РАН // Представительная власть — XXI век. 2018. №7-8. С. 26–33.

167. Любимов А.П., Щитова Н.Г. Нужен ли нам новый Закон о культуре? (По материалам парламентских слушаний в Государственной Думе 19 апреля 2010 г.) // Представительная власть — XXI век. М., 2010. С. 12–16.

168. Макаров В.Л. Бахтизин А.Р. Социальное моделирование — новый компьютерный прорыв (Агент-ориентированные модели). М.: Экономика, 2013.

169. Макаров В.Л. Получение нового знания методом компьютерного моделирования // Искусственный интеллект. Междисциплинарный подход. М.: ООО «ИИнтелл», 2006.

170. Михеенкова М.А., Финн В.К. Интеллектуальный анализ данных и его реализация в партнерских человеко-машинных системах // Материалы XII Мультиконференции по проблемам управления (МКПУ-2019). Р н/Д; Таганрог: Изд-во ЮФУ, 2019. Т. 1. С. 113–115.

171. Норичугу У. Общество 5.0: взгляд Mitsubishi Electric // Экономические стратегии. 2017. № 4. С. 2–11.

172. Парфенова Л.Б. Европейская стратегия развития цифровой экономики: региональная дифференциация // Вестник Тверского государственного университета. Серия: Экономика и управление. 2018. № 3. С. 30–38.

173. Перепелица А.А. Промышленная робототехника в России. Состояние и проблемы развития. Видение интегратора. [http://alphajet.ru/knowledge/articles/stati\\_45.html](http://alphajet.ru/knowledge/articles/stati_45.html)

174. Пожарев Т., Алексеев А. Креативные мультимедиа: физикалистский и менталистский подходы // Философия творчества: материалы Всероссийской научной конференции, 8–9 апреля 2015 г., Институт философии РАН, г. Москва / Под ред. Н.М. Смирновой, А.Ю. Алексеева. М.: ИИнтелл, 2015. С. 418–424.

175. Правовые аспекты использования искусственного интеллекта: актуальные проблемы и возможные решения: доклад НИУ ВШЭ / рук. авт. кол.: В.Б. Наумов, С.А. Чеховская, А.Ю. Брагинец, А.В. Майоров. М.: Издательский дом Высшей школы экономики, 2021. С. 12.

176. *Рицолатти Дж., Синигалья К.* Зеркала в мозге: О механизмах совместного действия и сопереживания. М.: Языки славянских культур, 2012. С. 97.

177. *Рубанов В.А.* Не подражать живым системам // Независимая газета 23.03.2020.

178. *Рязанцев А.П., Рязанцева Д.А.* Интеграционные процессы и технологические уклады // Национальные интересы: приоритеты и безопасность. 2014. № 16. С. 54.

179. *Свириденко Д.И., Витяев Е.Е.* Задачный подход к искусственному интеллекту и его теоретическая и технологическая база // Сб. трудов XVIII Национальной конференции по искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 36–44.

180. *Селмер Брингсйорд, Пол Беллоу, Дэвид Феруччи.* Творчество, тест Тьюринга и улучшенный тест Лайвлейс // Тест Тьюринга. Роботы. Зомби / Перевод с англ. А.Ласточкина; под ред. А.Ю.Алексеева. М.: МИЭМ, 2006. С. 62–85.

181. *Синицын С.А.* Российское и зарубежное гражданское право в условиях роботизации и цифровизации. Опыт междисциплинарного и отраслевого исследования: научно-практическое пособие. М.: Инфотропик Медиа, 2020.

182. *Сломэн А.* Что значит быть камнем? // Тест Тьюринга. Роботы. Зомби / Перевод с англ. В. Крючкова; под ред. А.Ю. Алексеева. М.: МИЭМ, 2006. С. 86–102.

183. *Соколов И.А.* Теория и практика искусственного интеллекта // Вестник Российской академии наук. 2019. Т. 89. № 4. С. 365–370.

184. *Соснов А.Я.* Роли для робота // Поиск. 2019. № 46. 15 ноября. С. 13.

185. *Станек Ю.* Право и нейронаука — точки пересечения. URL: <https://cyberleninka.ru/article/n/pravo-i-neyronauka-tochki-peresecheniya/viewer>).

186. *Степанов С.Ю., Оржековский П.А., Ушаков Д.В.* Проблема цифровизации и стратегии развития непрерывного образования // Непрерывное образование: XXI век. 2020. № 2 (30). С. 2–15.

187. *Тарасов В.Б.* Философия искусственного интеллекта, система когнитивных наук, компьютерная поддержка научных исследований и образования // Искусственный интеллект и принятие решений. 2021. № 2. С. 93–109.

188. *Федунов Б.Е.* Интеллектуальные системы тактического уровня: состав, базы знаний // Сб. трудов XVIII Национальной конференции по

искусственному интеллекту с международным участием (КИИ-2020). М.: МФТИ, 2020. С. 87–94.

189. Финн В.К. О классе ДСМ-рассуждений, использующих изоморфизм правил индуктивного вывода // Искусственный интеллект и принятие решений. 2016. № 3. С. 48–61.

190. Финн В.К. Об эвристиках ДСМ-исследований // НТИ. 2019. Сер. 2. № 10. С. 1–34.

191. Хабриева Т.Я., Черногор Н.Н. Будущее права. Наследие академика В.С. Степина и юридическая наука. М.: Российская академия наук; Институт законодательства и сравнительного правоведения при Правительстве Российской Федерации; ИНФРА-М, 2020.

192. Хорошевский В.Ф. Проектирование систем программного обеспечения под управлением онтологий: модели методы, реализации // Онтологии проектирования. 2019. Т. 9. № 4 (34). С. 429–448.

193. Цветков В.Я. Естественное и искусственное информационное поле // Международный журнал прикладных и фундаментальных исследований. 2014. № 5 (часть 2). С. 178.

194. Черниговская Т.В. Homo Loquens: эволюция церебральных функций и языка // Журнал эволюционной биохимии и физиологии. 2004. Т. 40. № 5. С. 400–406.

195. Черниговская Т.В. Нейронаука в поисках смыслов: мозг как барокко? // Вопросы философии. 2021. № 1. С. 17–26.

196. Черниговская Т.В. Языки сознания: кто читает тексты нейронной сети? // Человек в мире знания: в честь 80-летия акад. В.А. Лекторского. М.: РОССПЭН, 2012. С. 403–415.

197. Черниговская Т.В., Деглин В.Л. Проблема внутреннего диалогизма (нейрофизиологическое исследование языковой компетенции) // Структура диалога как принцип работы семиотического механизма / Под ред. Ю.М. Лотмана. Тарту, 1984 (Труды по знаковым системам. Т. 17; Учен. зап. Тарт. гос. ун-та. Вып. 641). С. 48–67.

## ДИССЕРТАЦИИ

198. Алексеев А.Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга. Диссертация на соискание ученой степени доктора философских наук по специальности 09.00.08 «Философия науки и техники». М.: МГУ им. М.В. Ломоносова, 2015. 482 с.

ЛИТЕРАТУРА НА ИНОСТРАННЫХ ЯЗЫКАХ

199. *Abbott R.* I Think. Therefore I Invent: Creative Computers and the Future of Patent Law // Op. cit. URL: <https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3522&context=bclr>
200. *Abramowicz M.* Orphan Business Models: Toward a New Form of Intellectual Property, 124 // *Harvard Law Review*, 2011.
201. *Aksaray D., Vasile C.-I., Belta C.* Dynamic Routing of Energy-Aware Vehicles with Temporal Logic Constraints // *Proc. International Conference ICRA on Robotics and Automation*. Stockholm, 2016. P. 3141–3146.
202. *Alexandrov Y.I., Sams M.E.* Emotion and consciousness: Ends of a continuum // *Cognitive brain research*. 2005. Vol. 25 (2). P. 387–405.
203. *Algorithmic Regulation and Personalized Law: A Handbook* (Christoph Busch & Alberto De Franceschi eds., 2021). P. 115–136, Available at SSRN: <https://ssrn.com/abstract=3503376>
204. *Allen R.J.* (2004) Legal evidence scholarship meets artificial intelligence. *Applied Artificial Intelligence*. 18 (3–4). P. 367–389. Available from: doi: 10.1080/08839510490280003
205. *Anderson S.* The End of Theory // *Wired*. June 2008.
206. *Ashley, K. D.* (2017) *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge: Cambridge University Press. Available from: doi: 10.1017/9781316761380
207. *Asmolov A. G.* Race for the Future: «...Now Here Comes What's Next» // *Russian Education and Society*, издательство M.E. Sharpe Inc. (United States). Том 60. № 5. С. 381–391.
208. *Asmolov A., Guseltseva M.* Education as a space of opportunities: from human capital to human potential // *European Proceedings of Social and Behavioural Sciences*, издательство Future Academy (online). Том 64. № 6. С. 40–45.
209. *Åström K.J., McAvoy T.J.* Intelligent Control: An Overview and Evaluation // *Handbook of Intelligent Control*. Neural, Fuzzy, and Adaptive Approaches / Ed. by D.A.White, D.A.Sofge. New York: Van Nostrand Reinhold, 1992. P. 3–34.
210. *Balezeau F., Wilson B., Gallardo G., Dick F., Hopkins W., Anwander A., Frederici A., Griffiths T., Petkov Ch.* Primate auditory prototype in the evolution of the arcuate fasciculus. *Nature Neuroscience* Vol. 23 (5). 2020. P. 611–614.

211. *Balkin J.M.* The Three Laws of Robotics in the Age of Big Data // Ohio State.
212. *Barfield W.* (2005) Issues of Law for Software Agents within Virtual Environments. Presence: Teleoperators and Virtual Environments. 14 (6). P. 741–748. Available from: doi: 10.1162/105474605775196607
213. *Besold T.R., Schmid U.* Why Generality Is Key to Human-Level Artificial Intelligence // J. Advances in Cognitive Systems. 2016, Vol. 4. P. 13–24.
214. *Block N.* Psychologism and Behaviorism // Philosophical Review. 1981. N 90. P. 5–43.
215. *Boolos G.* Introductory Note // Gödel K. Collected Works. Vol.3 / eds. Feferman et al. Oxford: Oxford University Press, 1995. P. 293.
216. *Busch Christoph and De Franceschi Alberto.* Granular Legal Norms: Big Data and the Personalization of Private Law (March 06, 2018, 17 Pages). Forthcoming in Vanessa Mak, Eric Tjong Tjin Tai and Anna Berlee (eds), Research Handbook on Data Science and Law, Edward Elgar 2018, Available at SSRN: <https://ssrn.com/abstract=3181914>
217. *Buzikov M., Galyaev A., Guryev Yu., Titov K., Yakushenko E., Vassilyev S.* Intelligent Control of Autonomous and Anthropocentric On-board Systems // Procedia Computer Science. 2019. Vol. 150. P. 10–18.
218. *Cartmill E.A., Roberts S., Lyn H., Cornish H.* “The Evolution of Language”, Proceedings of the 10th International Conference (EVOLANG10) Vienna, 14–17 April 2014, World Scientific Publ., 2014. P. 479–481.
219. *Chalmers D.* Minds, Machines and Mathematics // Psyche, 2 (9), June 1995.
220. *Chemero A.* Radical Embodied Cognitive Science. Cambridge: Mass. MIT Press, 2009.
221. *Chernigovskaya T.V.* Biology, Environment, and Culture: From Animal Communication to Human Language and Cognition. Vestnik Sankt Peterburgskogo universiteta. Filosofiya i konfliktologiya 2020. Vol. 36 (1), P. 157–170.
222. *Clark A.* Being there: Putting brain, body, and world together again. — MIT press, 1998. <http://www2.econ.iastate.edu/tesfatsi/BeingThere.AClark1998.EntireBook.pdf>
223. *Clark W. R., Grundstein M.* Are we Hardwired? The Role of Genes in Human Behavior. Oxford: Oxford University Press, 2000.

224. *Crick F., Koch C.* (2007). A neurobiological framework for consciousness. M. Velmans, S. Schneider (eds.). *The Blackwell Companion to Consciousness*. Blackwell, P. 567–579.

225. *Das R., Godbole A., Monath N., Zaheer M., McCallum A.* Probabilistic Case-based Reasoning for Open-World Knowledge Graph Completion // *CaseBR arXiv:2010.03548v2 [cs.CL]*.

226. *Deacon T.* Monkey homologues of language areas: computing the ambiguities // *Trends in Cognitive Sciences*. 2004. Vol. 8(7). P. 288–290.

227. *Deacon T. W.* (2013) *Incomplete nature: How mind emerged from matter*, W. W. Norton & Company, NY.

228. *Dennett D.* *Consciousness Explained*. Basil Bay Books. NY.; Boston, L. 1991. P. 418.

229. *Dilemma* (September 1, 2016) // *IDEA: The IP Law Review*. Vol. 57. 2017. № 3.

230. *Diveev A.I., Balandina G.I., Konstantinov S.V.* Binary Variational Genetic Programming for the Problem of Synthesis of Control System // *Proc. 13th International Conference on Neural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSCD-2017)*. 2017. P. 165–170.

231. *Dubrovsky D.I.* The Problem of Free Will and Modern Neuroscience // *Neuroscience and Behavioral Physiology*. 2019. Vol. 49. No. 5. P. 629–639.

232. *Dubrovsky David I.* The Hard Problem of Consciousness. Theoretical solution of its main questions // *AIMS Neuroscience*. 2019. Vol. 6. No. 2. P. 85–103.

233. *Dummett M.* *The Philosophical Significance of Gödel's Theorem // Truth and Other Enigmas*. Harvard: Harvard University Press, 1995.

234. *Eco U.* *Kant and the Platypus. Essays on Language and Cognition*. NY: Harcourt Brace & Company, 2000.

235. *Efimov A.* Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence // *Artificial General Intelligence (AGI)*. 2020. Vol. 12177. P. 83–94.

236. *Eidenmueller H.* The Rise of Robots and the Law of Humans (March 26, 2017) // *Oxford Legal Studies Research Paper*. 2017. № 27. URL: <http://dx.doi.org/10.2139/ssrn.2941001>

237. *Elkin-Koren Niva and Gal Michal S.* The Chilling Effect of Governance-by-Data on Data Markets, *University of Chicago Law Review*: 2019. Vol. 86: Iss. 2. Article 6. P. 403–431. Available at: <https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/6>

238. *Epstein Joshua M. and Axtell Robert*. Growing Artificial Societies: Social Science from the Bottom Up. Brooking Institution Press and MIT Press, Washington DC. 1996.

239. *Ezrachi Ariel and Stucke Maurice E*. Artificial Intelligence & Collusion: When Computers Inhibit Competition (April 8, 2015). University of Illinois Law Review, Vol. 2017. Oxford Legal Studies Research Paper. No 18/2015, University of Tennessee Legal Studies Research Paper. No 267, Available at SSRN: <https://ssrn.com/abstract=2591874>

240. *Feferman S*. Penrose's Gödelian Argument // *Psyche* 2 (7), May 1995.

241. *Feteris E., Kloosterhuis H*. Law and Argumentation Theory: Theoretical Approaches to Legal Justification (June 21, 2013). URL: <http://dx.doi.org/10.2139/ssrn.2283092>

242. *Fodor J*. Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology // Fodor J. Representations. Cambridge: Mass. MIT Press, 1981.

243. Future of Artificial Intelligence Act of 2017 H.R.4625 US; European Commission. White paper on artificial intelligence—a European approach to excellence and trust. 2020 etc.

244. *Galyaev A.A.* On the Detection Functional in Motion of an Object in a Threat Environment // *J. Automation and Remote Control*. 2011. Vol. 71. N 4. P. 634–639.

245. *Gödel K*. Some Basic Theorems on the Foundations of Mathematics and their Implications // Gödel K. Collected Works. Vol. 3 / eds. Feferman et al. Oxford: Oxford University Press, 1995. P. 310. Русский перевод В. Целищева.

246. *Goertzel B*. Toward a Formal Characterization of Real-World General Intelligence // *J. Advances in Intelligent Systems Research*. 2010. Vol. 10. P. 1–6.

247. Grigoleit Hans Christoph and Bender, Philip Maximilian. The Law between Generality and Particularity — Potentials and Limits of Personalized Law (December 13, 2019).

248. *Guadamuz A*. Artificial intelligence and copyright // *WIPO Magazine*. 2017. N 5.

249. *Guohua Shen, Tomoyasu Horikawa, Kei Majima, Yukiyasu Kamitani*. Deep image reconstruction from human brain activity” | *BioRxiv* | doi: 10.1101/240317 Posted December 28, 2017.

250. Handbook on the Law of Artificial Intelligence, Edward Elgar, (2017/18, Forthcoming). URL: <https://www.law.ox.ac.uk/business-law-blog/blog/2017/11/robots-boardroom-artificial-intelligence-and-corporate-law>

251. *Handelman D.A., Stengel R.F.* An Architecture for Real-time Rule-based Control // Proc. American Control Conference. 1987. P. 1636–1642.
252. *Heide M., Haffner Chr., Murayama A., Korotaki Y., Shinohara H., Okano H., Sasaki E., Hattner W.* Human-specific ARHGAP11B increases size and folding of primate neocortex in fetal marmoset//Science. 2020. Vol. 369 (6503). P. 546–550.
253. *Hoffmeyer J., Kull K.* Baldwin and biosemiotics: What intelligence is for // B.H. Weber, D. Depew, (eds.). Evolution and Learning: The Baldwin Effect Reconsidered. Cambridge: MIT Press, 2003. P. 253–272.
254. *Horst S.* The Computational Theory of Mind // The Stanford Encyclopedia of Philosophy (Spring 2011 Edition) (<http://plato.stanford.edu/archives/spr2011/entries/computational-mind/>).
255. *Hugdahl K.* Experimental Methods in Neuropsychology. NY: Kluwer Academic Publishers, 2002.
256. *J. Derrida.* Le problème de la genèse dans la philosophie de Husserl. Paris: Presses Universitaires de France, 1990. The parallelism is made particularly relevant by Foucault's criticism of Husserl.
257. *Jordan M.* Barry, John William Hatfield, and Scott Duke Kominers, To Thine Own Self Be True? Incentive Problems in Personalized Law, 62 Wm. & Mary L. Rev. 723 (2021). URL: <https://scholarship.law.wm.edu/wmlr/vol62/iss3/2>
258. *Joseph G. Makin, David A. Moses and Edward F. Chang* (2018). Machine translation of cortical activity to text with an encoder–decoder framework // Nature-Neuroscience Technical Report <https://doi.org/10.1038/s41593-018-0000-0>
259. *Kalin H.* Artificial Intelligence and the Copyright... 1, 2006
260. *Katz G., Barrett C.W., Dill D.L., Julian K., Kochenderfer M.J.* An Efficient SMT Solver for Verifying Deep Neural Networks // CAV. 2017. P. 97–117.
261. *Kight W.* (2017) A Stronger AlphaGo Defeats the World's Number One Player // <https://www.technologyreview.com/2017/05/23/151572/a-stronger-alpha-go-defeats-the-worlds-number-one-player/>.
262. *Kireev M., Slioussar N., Korotkov A.D., Chernigovskaya T.V., Medvedev S.V.* Changes in functional connectivity within the fronto-temporal brain network induced by regular and irregular Russian verb production // Frontiers in Human Neuroscience. 2015. Vol. 9 (36). P. 193–220.
263. *Kirk R.* Mental Machinery and Gödel // Synthese. Vol. 66. 1986. P. 437–452.



264. *Kitayama S.* Culture and basic psychological processes- toward a system of culture: Comment on Oyserman et al. // *Psychological Bulletin*. 2002. Vol. 128 (1). P. 89–96.
265. *Kobriniskii B.A.* Images in Logical-and-Linguistic Artificial Intelligence // *J. Biomedical Engineering and Medical Imaging*. 2019. Vol. 6. N 1. P. 1–8.
266. *Kolber A.* Will There Be a Neurolaw Revolution? // *Indiana Law Journal*. 2014. Vol. 89. P. 807. URL: <https://ssrn.com/abstract=2398071>.
267. *Kolmogoroff A.N.* Zur Deutung der Intuitionistischen Logik // *Math. Z.* 1932. N 35. P. 58–65.
268. *Komendantskaya E.* First-order Deduction in Neural Networks // *Proc. 1st International Conference on Language and Automata Theory and Applications (LATA-2007)*. Tarragona, 2007. P. 307–319.
269. *Kotseruba I., Tsotsos J. K.* (2018) 40 years of cognitive architectures: core cognitive abilities and practical applications // *Artificial Intelligence Review* <https://doi.org/10.1007/s10462-018-9646-y>
270. *Kowalski R.* Logic Programming // In book: *Computational Logic*, Edition: In the History of Logic series, edited by D. Gabbay and J. Woods. Elsevier, 2014. P. 523–569.
271. *Kull K.* Towards a Theory of Evolution of Semiotic Systems// *Chinese Semiotic Studies*. 2014. Vol. 10 (3). P. 485–495.
272. *Lake B.M., Salakhutdinov R., Tenenbaum J.B.* (2015) Human-level concept learning through probabilistic program induction. *Science* 350(6266):1332–38. [arBML, MB, ED, NK]
273. *Lassegue J.* 1988, What Kind of Turing Test did Turing Have in Mind? *Tekhnema* 3. P. 37–58.. URL : <http://tekhnama.free.fr/3Lasseguearticle.htm>
274. Law (September 15, 2017) // Woodrow Barfield and Ugo Pagallo (eds.), *Research Law Journal*. 2017. Vol. 78.
275. Law, Innovation and Technology. 2014. Vol. 6. N 2. P. 194–222. URL: <https://ssrn>.
276. *Leahy K., Zhou D., Vasile C.I., Oikonomopoulos K., Schwager M., Belta C.* Provably Correct Persistent Surveillance for Unmanned Aerial Vehicles Subject to Charging Constraints // *Experimental Robotics*. 2016. P. 605–619.
277. *Lehrer J.* Proust Was a Neuroscientist, NY: HMH, 2008.
278. *Lektorsky Vladislav.* 2021/ The Activity Approach in Late Soviet Philosophy // *The Palgrave Book in Russian Thought* // Ed. M.F. Bykova,

M.N. Forster, L. Steiner. Palgrave MacMillan. Cham, Switzerland. 1980. P. 407–422.

279. *Levin J.* Functionalism // The Stanford Encyclopedia of Philosophy (Fall 2013 Edition) (<http://plato.stanford.edu/archives/fall2013/entries/functionalism/>).

280. *Lindström P.* Penrose's New Argument // Journal of Philosophical Logic. Vol. 30. 2001. P. 241–250.

281. *Lomuscio A., Michaliszyn J.* Verifying Multi-Agent Systems by Model Checking Three-valued Abstractions // Proc. 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2015). Istanbul. 2015. P. 189–198.

282. *Lukas J.* Minds, Machines and Gödel // Philosophy. Vol. 36. 1961.

283. *Lyubimov A.P., Cherny V.V.* The Evolution of Globalism: From Computerization to e-Democracy and the Digital Economy of Knowledge // Automatic Documentation and Mathematical Linguistics. 2021. Vol. 55, No 2. P. 39–45.

284. *Mao J., Gan C., Kohli P., Tenenbaum J.B., Wu J.* The Neuro-symbolic Concept Learner: Interpreting Scenes, Words, and Sentences from Natural Supervision // ICLR. 2019. P. 1–28.

285. *Mark O. Riedl.* The Lovelace 2.0 Test of Artificial Intelligence and Creativity. Proceedings of the AAAI Workshop: Beyond the Turing Test, Austin, Texas, 2015; <https://arxiv.org/abs/1410.6142>

286. *McCullough D.* Can Humans Escape Gödel? // Psyche, 2 (4), April 1995.

287. McLuhan, Marshall and Fiore, Quentin. The Medium is the Massage. Bantan Books, New York, 1967.

288. *Meyer-Schonberger V.* Delete: the virtue of forgetting in the digital age. Princeton. Princeton University Press. 2011.

289. *Michaelian K.* Is external memory memory? Biological memory and extended mind. // Consciousness and cognition. 2012. 21, 1154–1165.

290. *Miller R. M.* Don't Let Your Robots Grow Up to Be Traders: Artificial Intelligence, Human Intelligence, and Asset-Market Bubbles // Journal of Economic Behavior and Organization, Forthcoming. URL: <http://dx.doi.org/10.2139/ssrn.415220>

291. *Mollica F., Piantadosi S.T.* Humans store about 1.5 megabytes of information during language acquisition//Royal Society Open Science. 2019. Vol. 6 (3). 181393. <https://doi.org/10.1098/rsos.181393>

292. *Möslein F.* Robots in the Boardroom: Artificial Intelligence and Corporate ...

293. *Muggleton S.H., Lin D., Pahlavi N., Nezhad A.T.* Meta-interpretive Learning Application to Grammatical Inference // J. Machine Learning. 2014. N 94. P. 25–49.

294. *Murphy R., Woods D.D.* Beyond Asimov: the three laws of responsible robotics. IEEE Intelligent Systems. 2009. Jul. 24 (4).

295. *Natochin Yu., Chernigovskaya T.* From Archebiosis to Evolution of Organisms and Informational Systems// Biological Communications. 2020. Vol. 65 (3). P. 215–227.

296. *Neubauer S., Gunz Ph., Scott N., Hublin J-J., Mitteroecker Ph.* Evolution of brain lateralization: A shared hominid pattern of endocranial asymmetry is much more variable than in great apes// Science advances. 2020. Vol. 6(7). P. 1–11.

297. *Oskamp A. & Lodder A.R.* (2006) Introduction: Law, Information Technology, and Artificial intelligence. In: Lodder, A.R. & Oskamp A. (eds.) Information Technology and Lawyers. Dordrecht, Springer. P. 1–22. Available from: [doi:10.1007/1-4020-4146-2\\_1](https://doi.org/10.1007/1-4020-4146-2_1).

298. *Otten J.* NanoCoP: a Non-clausal Connection Prover // Proc. International Joint Conference on Automated Reasoning. 2016. P. 302–312.

299. *Palmerini E.* et al. RoboLaw: Towards a European framework for robotics regulation // Robotics and Autonomous Systems. 2016. Vol. 86. P. 78–85.

300. *Papert, S.* Mindstorms: Children, Computers and Powerful Ideas. New York, NY, USA: Basic Books Inc. Publishers, 1980. 252 p.

301. *Perez C.E.* (2017) AlphaZero: How Intuition Demolished Logic // <https://medium.com/intuitionmachine/alphazero-how-intuition-demolished-logic-66a4841e6810>.

302. *Philip M. Bender*, 2020. «Limits of Personalization of Default Rules Towards a Normative Theory», Working Papers tax-mpg-rps-2020–02, Max Planck Institute for Tax Law and Public Finance. URL: <https://ideas.repec.org/p/mpi/wpaper/tax-mpg-rps-2020–02.html>;

303. *Verstein, Andrew* (2019). Privatizing Personalized Law. University of Chicago Law Review: Vol. 86: Iss. 2, Article 1. Available at: <https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/1>;

304. *Pierce C.S.* Semiotica. Torino: Einaudi, 1980.
305. *Polger T.W.* Natural minds. — Massachusetts Institute of Technology: A Bradford book, 2004.
306. *Poole D.* Representing Diagnosis Knowledge // J. Annals of Mathematics and Artificial Intelligence. 1994. Vol. 11. P. 33–50.
307. *Pronchev G.B., Mikhailov A.P., Lyubimov A.P., Solovyev A.A.* Particularities of the Internet-based virtual social environments within the context of information warfare // EurAsian Journal of BioSciences. 2020. Vol. 14. N 2. P. 3731–3739.
308. *Pulina L., Tacchella A.* Challenging SMT Solvers to Verify Neural Networks // J. Artificial Intelligence Communications. 2012. Vol. 25. N 2. P. 117–135.
309. *Putnam H.* Minds and machines // Hook S. (ed.) Dimensions of Mind. New York University Press, 1960.
310. *Putnam H.* Reason, Truth, and History. Cambridge: Cambridge University Press, 1981.
311. *Putnam H.* The Nature of Mental States // Readings in Philosophy of Psychology / Ed. by N. Block. 1980. Vol. 1. P. 223–231. Cambridge. Mass, Harvard University Press.
312. *Pylkkänen L.* The neural basis of combinatory syntax and semantics// Science. 2019. Vol. 366. P. 62–66.
313. *Radutniy O. E.* Criminal liability of the artificial intelligence. doi: 10.21564/2414–990x.138.105661 UDC 343.22+343.412:004.056//cyberleninka.ru/article/n/criminal-liability-of-the-artificial-intelligence/viewer
314. Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design //
315. *Russell S, Norvig P.* Artificial intelligence: a modern approach. Prentice Hall, Upper Saddle River. 1995.
316. *Saripan H.* (2016) Are Robots Human? A Review of the Legal Personality Model. World Applied Sciences Journal. 34 (6). P. 824–831. Available from: doi:10.5829/idosi.wasj.2016.34.6.15672
317. *Sartor G.* (2009) Cognitive Automata and the Law: Electronic Contracting and the Intentionality of Software Agents. Artificial Intelligence and Law. 17 (4). P. 253–290.
318. *Shapiro S.* Mechanism, Truth, and Penrose’s New Argument // Journal of Philosophical Logic. Vol. 32. 2003. P. 19–42.

319. *Skrynnik A., Panov A.I.* Hierarchical Reinforcement Learning with Clustering Abstract Machines // J. Communications in Computer and Information Science. 2019. Vol. 441. P. 30–43.
320. *Smolensky P.* On the Proper Treatment of Connectionism // Behavioral and Brain Sciences. 1988. Vol. 11. P. 1–74.
321. *Solum L.B.* Legal Personhood for Artificial Intelligences // North Carolina Law Review. 1992. Vol. 70. P. 1231. Illinois Public Law Research Paper № 09–13. URL: <https://ssrn.com/abstract=1108671>
322. Squirrel AI Learning. 1st AI-powered adaptive education provider in China. <http://squirrelai.com/>
323. *Stepanov S.Yu., Ushakov D.V.* Artificial intelligence and digital angel technology in education. М.: МФТИ, 2020. С. 105–112.
324. *Sunstein C.* Of Artificial Intelligence and Legal Reasoning (November 2001). University of Chicago Law School Roundtable. Vol. 8. 2001. URL: <http://dx.doi.org/10.2139/ssrn.289789>
325. *Tennant N.* Deflationism and Gödel Phenomena // Mind. Vol. 111. N 443. 2002. P. 551–582.
326. *Turing A.* Computing Machinery and Intelligence. Mind 59 (236). 1950. P. 433–460.
327. *Udkowsky E.* Complex Value Systems in Friendly Artificial Intelligence // Proc. 4th International Conference on Artificial General Intelligence (AGI-2011). Lecture Notes in Computer Science 6830. — Springer, 2011. P. 388–393.
328. *Urban J., Vyskocil J.* Theorem Proving in Large Formal Mathematics as an Emerging AI Field / arXiv:1209.3914v2(cs.AI) 16Dec2012
329. *Vassilyev S.N.* Abductive Inference Method in Problems of Explaining the Observed // J. Computer and Systems Sciences International. 2021. Vol. 60. N 1. P. 153–161.
330. *Vassilyev S.N.* Temporal Reasoning on the Basis of New Logic for Intelligent Control // Proc. of the IEE Inter. Symposium on Intelligent Control, 2003. P. 843–847.
331. *Vinogradov D.V.* VKF-method of Hypotheses Generation // J. Communications in Computer and Information Science. 2014. Vol. 436. P. 237–248.
332. *Vygotsky, L. S.* Mind in society: The development of higher psychological processes. Harvard University Press (1980).

333. *Vygotsky L. S.* The instrumental method in psychology (1981), <https://www.marxists.org/archive/vygotsky/works/1930/instrumental.htm>

334. *Wall L.D.* Some financial regulatory implications of artificial intelligence. *Journal of Economics and Business*, 2018. 100. P. 55–63.

335. *Walter J.* Ong. Orality an Literacy. The Technologizing of the World. Methuen & Co., 1982.

336. *Wolfram Stephen.* A New Kind of Science. 2003.

337. *Yanisky-Ravid, Shlomit and Liu, Xiaoqiong (Jackie).* When Artificial Intelligence Systems Produce Inventions: The 3A Era and an Alternative Model for Patent Law (March 1, 2017). 39 *Cardozo Law Review*, 2215–2263 (2018), Available at SSRN: <https://ssrn.com/abstract=> or <http://dx.doi.org/10.2139/ssrn.2931828>.

338. *Yon Uexküll J.* Wiesehenwir die Natur und wiesiehtsiesichselber? // *Die Naturwissenschaften*. 1922. Vol. 10 (12). P. 265–281.

339. *Yu R.* The Machine Author: What Level of Copyright Protection is Appropriate for Fully Independent Computer Generated Works? *University of Pennsylvania Law Review*. 2071. 65 (5). P. 1241–1270.

340. *Zabarankin M., Uryasev S., Murphey R.* Aircraft Routing under the Risk of Detection // *J. Naval Research Logistics*. 2006. Vol. 5. N 8. P. 728–747.

**Издательство «Юридический центр»**

**Мы ждем Ваши отзывы и предложения:**

**По вопросам сотрудничества**

8 (911) 977-67-25

**Электронная почта**

[urizdat@mail.ru](mailto:urizdat@mail.ru)

[damenia@rambler.ru](mailto:damenia@rambler.ru)

[www.juridicheskicentr.ru](http://www.juridicheskicentr.ru)

**ЧЕЛОВЕК И СИСТЕМЫ  
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

*Под ред. акад. РАН В.А. Лекторского*

**Издательство «Юридический центр»**

Корректор *С.С. Алмаметова*

Компьютерная верстка *С.Л. Кузьминой*

Подписано в печать 25.04.2022 г.  
Формат 60×90/16. П. л. 20,5. Уч.-изд.л. 22,5 л.  
Тираж 1000 экз. Заказ . Цена свободная.

Отпечатано в типографии ООО «Литография Принт»  
191119, Санкт-Петербург, ул. Днепропетровская, дом 8, офис 14  
[www.litobook.ru](http://www.litobook.ru) e-mail: [info@litobook.ru](mailto:info@litobook.ru)  
тел.: +7 (812) 712-02-08

ISBN 978-5-94201-835-1

