

Ловушки для искусственного интеллекта

Райков Александр Николаевич — ведущий научный сотрудник Института проблем управления РАН, доктор технических наук, профессор.

УДК 004.8

Путь развития искусственного интеллекта в XXI в. в контексте становления новых технологических укладов и постнеклассической парадигмы управления можно спрогнозировать с учетом анализа его многоликости и тернистой исторической ретроспективы. Этот путь, естественно, не гарантирован от неожиданных ловушек. Вместе с тем их можно избежать, найдя новые решения в малознакомых пространствах для моделирования, взяв за основу иные подходы к решению сложных задач и семантической интерпретации данных, синтезировав под эти новации пока еще неведомые на Земле материалы для компьютерной памяти и процессоров.

Ключевые слова

Искусственный интеллект, компьютерная философия, новая парадигма кибернетики, информационное общество, цифровизация.

Искусственный интеллект уже давно убедительно наращивает свои возможности. В век грядущий без него, по-видимому, нельзя будет обойтись во всех сферах человеческой практики. Государственное и корпоративное управление, общение человека с машиной, маркетинг и рынки, анализ данных и визуализация, интеллектуальное вооружение и ситуационные центры, медицина и игра в шахматы, роботы и интеллектуальные помощники в работе людей, космические аппараты и города будущего, комму-

никации и коллаборационистские среды — примеры областей интенсивного применения искусственного интеллекта.

Искусственный интеллект — это компьютерная философия, компьютерная психология и продвинутая компьютерная наука. Он помогает понять и использовать силу человеческого разума, а также, быть может, подобраться поближе к разгадке тайны всемогущества человеческого духа. Подходы, методы и средства искусственного интеллекта самые разные: гуманитарные науки,



логика, лингвистика и семиотика, управляемый хаос и обратные задачи, когнитивное моделирование и распознавание образов, природные и квантовые вычисления, генетические и муравьиные алгоритмы, нейронные и иммунные технологии, голография и многое другое.

Термин «искусственный интеллект» введен в середине прошлого века. Искусственный интеллект более всего сейчас связывают с компьютерной обработкой данных, с логикой и исчислениями, алгоритмизацией и программированием, аддитивными и ассоциативными схемами, знаками и их семантической интерпретацией, образами и визуальной аналитикой. Это особое и интригующее явление в среде человеко-машинного взаимодействия и, возможно, космического универсума.

На переломе веков искусственный интеллект все больше ухо-

дит в пространства смыслов, в способы поддержки коллективной творческой деятельности, глубже проникает в тайны субъективного и бессознательного, порождения группового и индивидуального «эврика-эффекта». Однако содержательная интерпретация смыслов не поддается традиционным логическим представлениям, а коллективное сознание остается пока слабо исследованным феноменом. На интеллект влияют эмоции, чувства, медитативные состояния ума, дальние воздействия, которые не всегда доступны формализованной репрезентации.

Мир строит сценарии на новый век [1], идет становление новой парадигмы кибернетики [2] и научного обеспечения постнеклассической рациональности [3]. В развитии искусственного интеллекта формируются новые тренды, с начала века подходы и методы искусственного ин-

теллекта уже активно используются в сравнительно новых областях при следующих обстоятельствах:

- поддержка принятия решений во всех сегментах рынка, например в финансовом [4], и во всех отраслях, например в космической [5];
- накопление, хранение и аналитическая обработка больших объемов данных (большие данные), что позволяет быстро обнаружить источник проблем и дать надежный прогноз развития событий;
- создание систем облачных и туманных вычислений, позволяющих аккумулировать разбросанные по множеству устройств, в том числе мобильных, вычислительные и интеллектуальные ресурсы;
- объединение физических объектов (вещей) и механизмов управления в вычислительную сеть с обеспечением их взаимодействия друг с другом (Интернет вещей), что помогает освободить человека от рутинных работ, раскрепостить его творческую деятельность;
- разработка аддитивных технологий и технологий дополненной реальности, позволяющих наполнять искусственные модели реальным содержанием, повышая тем самым качество социально-экономических прогнозов и ускоряя производственные процессы;
- поддержка целенаправленной коллективной творческой деятельности, в том числе с участием гражданского общества, на основе применения методов сетевого искусственного интеллекта [6] и др.

При таком развитии событий заглянуть немного в будущее и увидеть ловушки для искусственного интеллекта — вопрос как важный, так и неотложный.

Немного истории

Историю развития искусственного интеллекта в прошлом веке можно проследить по *рис. 1*.

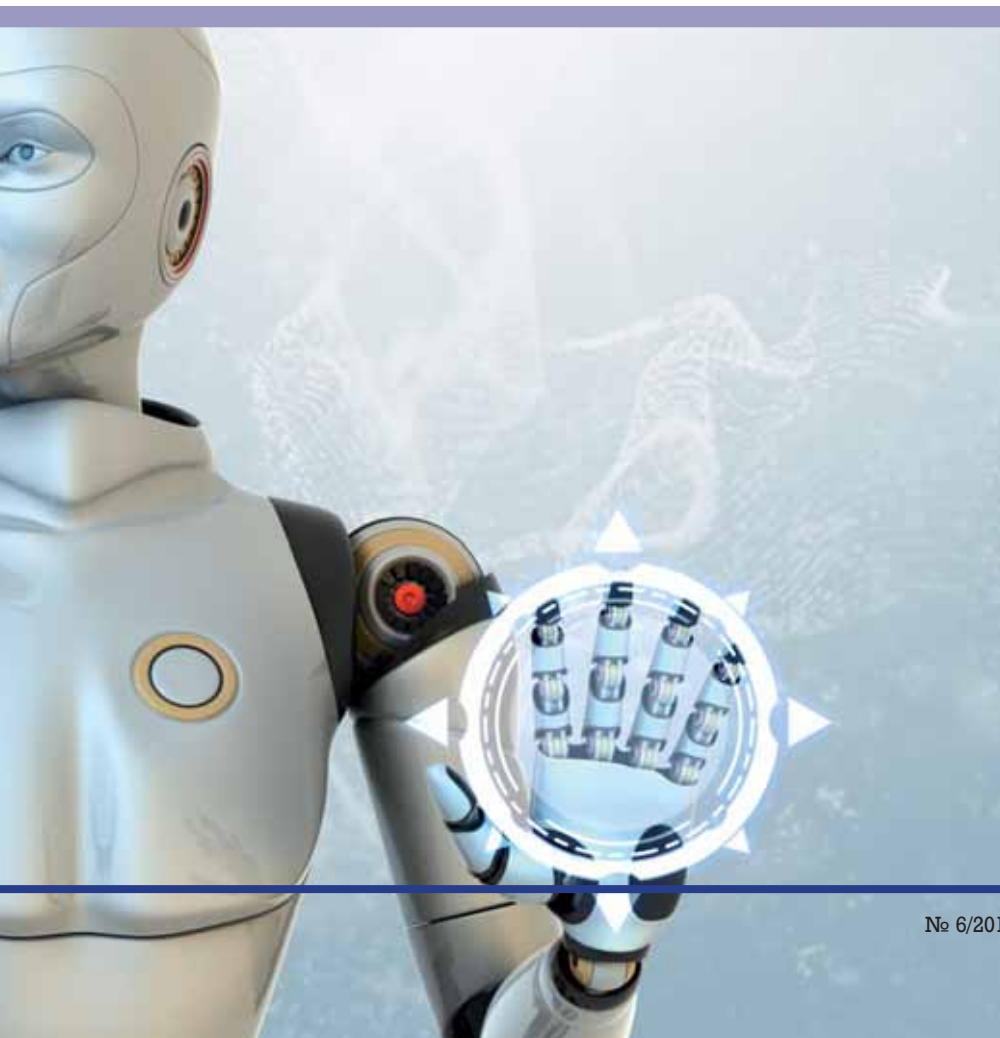
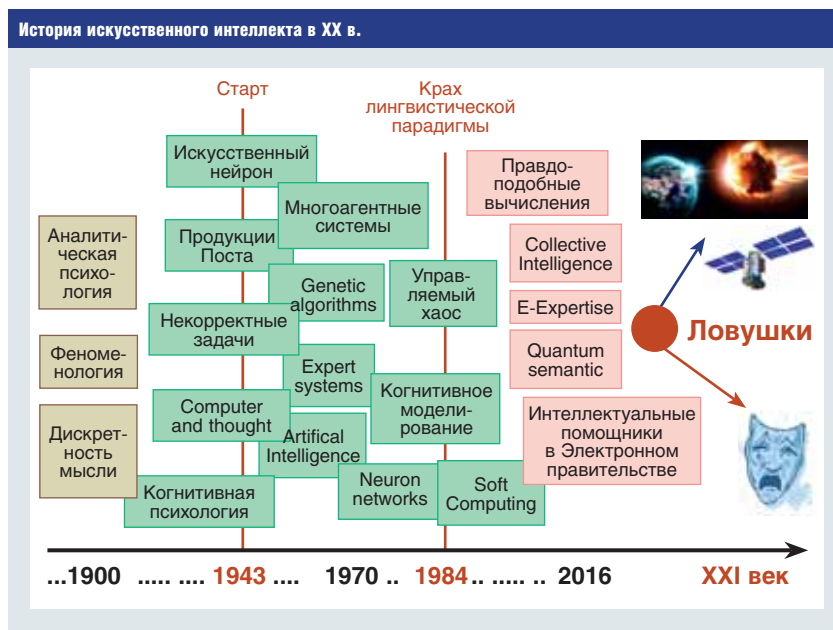


Рисунок 1



Искусственный интеллект наиболее интенсивно развивался во второй половине XX в., а в первой половине века, докомпьютерной, появился весомый теоретический базис. К наиболее значимым событиям истории искусственного интеллекта стоит отнести появление в начале века работ по аналитической психологии (З. Фрейд), феноменологии (Э. Гуссерль), логике (Б. Рассел, Ф. Фреге). К середине века проявились когнитивная психология (К. Левин), исчисление продуктов (Э. Пост), решение некорректных задач (А. Тихонов, К. Иванов), перцептрон (Ф. Розенблатт), теорема отсчетов (В. Котельников), контекстные грамматики (Н. Чомский). В последней трети прошлого века особо стоит отметить проработки в таких областях, как управляемый хаос, мультиагентные системы, виртуальные и самоорганизующиеся среды, семантические сети и онтология, квантовая семантика и конвергентные технологии.

История искусственного интеллекта — это чередование оптимистических всплесков и грустных разочарований. Периодич-

ность основной череды небольшая — порядка 10 лет. Скажем, в начале 1980-х годов исследователи увлеклись немонотонными логиками, а в конце десятилетия иллюзии относительно их прорывной силы померкли. Лингвистические процессоры недолго радовали исследователей — в середине 1980-х годов произошел кризис лингвистической парадигмы. Однако это не помешало компьютерной лингвистике развиваться и дальше. Справедливости ради стоит отметить наличие и более длинных циклов в развитии искусственного интеллекта, а также подтверждение закономерности возрождения подзабытого старого. Например, идея перцептрона как прародителя нейронных технологий родилась в конце 1950-х, затем угасла, несколько позже прокатилась небольшая волна исследований оптических нейронных сетей, но после доказательства подобия поведения нейронных сетей и алгоритмов эта волна улеглась, чтобы в начале 1990-х вновь подняться в контексте потребности улучшения анализа фондовых рынков и инвестиций, лингвистической обработки данных и пр.

Ловушки будущего

Если в прошлом веке, несмотря на отмеченные волны в развитии искусственного интеллекта, в основном прослеживался оптимистичный градиент, то грядущий век скорее всего не станет исключением и как и прежде будет удивлять своими неожиданными откровениями, которые могут заставить принципиально менять складывающиеся подходы, парадигмы, принципы, инструменты.

Неожиданность эта — явление условное, некоторый недогляд. Она на самом деле обуславливается незаметно набухающими неопределенностью и латентностью вопросами в сфере искусственного интеллекта, формированию ответов на которые, по всей видимости, не уделялось должного внимания в течение какого-то периода времени, например прошлого века. Но немудимая история потихоньку интегрирует накапливающиеся упущения, создавая на пути незаметные на первый взгляд и, быть может, даже сокрушительные ловушки.

Первая ловушка — это ловушка цифровизации. Информационное общество сейчас буквально погрузилось в цифру: цифровое правительство, цифровые города, цифровые архивы и даже «цифровой человек». Вычислительная техника, на волне которой вырос искусственный интеллект, в середине прошлого века по определенным причинам выбрала дискретную магистраль: цифры, биты, байты, отсчеты на аналоговом сигнале. Однако с какой бы точностью современный компьютер ни восстанавливал непрерывный сигнал по его точкам отсчета, с каждым отсчетом разница между природным (аналоговым) сигналом и искусственным цифровым накапливается. Вместе с тем дискретность представления, как оказалось, обрывает спектр непрерывного модулирующе-

го сигнала и как следствие накапливает ошибку его представления, незаметно разрушает целостность представляемых сигналом явлений. Возможно, именно цифровизация является причиной того, что ряд сложных задач на современных суперкомпьютерах решается неделями и месяцами. Есть и еще один негатив — обрыв спектра сигнала незаметно умяет тонкие гармоники репрезентации бесконечной и явно недискретной духовности информационного социума и как следствие приводит к возможному росту рисков потери субъективности [3] (рис. 2).

Вторая ловушка — это ловушка рациональности. Объяснения эффективности и результативности деловых процессов подтверждаются количественными показателями. Это может быть статистика, результаты социологических исследований или анализа больших данных. Аналитик — это мастер деления целого на части, описания этих частей формулами для расчета, последующих обобщений и синтеза. Обратная операция, индуктивная — от частей к целому, — не

Жизнь богата понятиями и смыслами, которые не укладываются в прокрустово ложе метрических интерпретаций, а для неметрических пространств необходимые механизмы решения обратных задач пока еще не разработаны.

всегда получается, логика неплохо работает только в одну сторону. Синтез же может давать сбой, так же как и в первой ловушке, при рациональном подходе растет риск потери целостности исследуемого явления. Как известно, смыслы — плоды парадоксов, носителями основных оценок являются люди, эксперты. Вместе с тем институт экспертизы еще плохо разработан и на то есть различные, включая субъективные, основания [7].

Ловушка причинности — это третья ловушка. Большинство систем искусственного интеллекта используют парадигму автоматического обучения на основе накопленного опыта, а логика отражает причинные

связи. Прогнозы часто делаются через экстраполяцию предыдущей динамики событий. Но далеко не все явления реальной жизни развиваются монотонно (вспомним «Дикие карты») и раскрывают причину. Задачи планирования и управления, ориентированные на амбициозный и конкурентный прорыв, как правило, ставятся лидерами и далеки от любых экстраполяций прошлого. Мысли лидеров не всегда сразу понятны окружению. Решение амбициозных задач неустойчиво, эти задачи некорректны, при небольшом изменении исходных данных путь к цели может сильно меняться (рис. 3). При этом следует иметь в виду, что существующие методы решения подобных задач, истоки создания которых датируются серединой прошлого века, пригодны только для метрических пространств. Однако жизнь богата понятиями и смыслами, которые не укладываются в прокрустово ложе метрических интерпретаций, а для неметрических пространств необходимые механизмы решения обратных задач пока еще не разработаны.

Еще одна, четвертая, ловушка — феноменологическая. Искусственный интеллект ищет решения в визуализации, мол, это охват иных слоев сознания, захват «правого полушария». Вместе с тем визуальный образ — это тот же знак, растровый он или векторный, и записан он в дискретной форме. Интеллект же питается эмоциями, чувствами, переживаниями. Последние

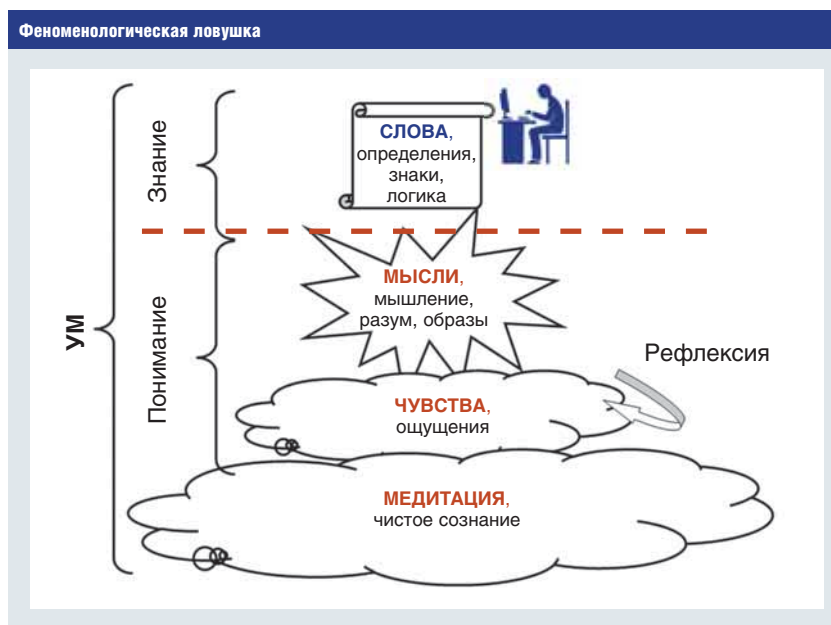
Рисунок 2



Рисунок 3



Рисунок 4



получают позитивную и негативную окраску не сами по себе, а в зависимости от когнитивного влияния на поведение человека [8]. Эмоции характеризуются интенсивностью: отклик, экстаз, вспышка, взрыв, блаженство, а также длительностью, предметностью, эмоциональным тоном. Гнев, радость, тоска, страх — тоже эмоции. Есть и более глубоко-

кий слой сознания — медитативный, трансцендентальный [9]. Его искусственный интеллект пока совсем не может охватить (рис. 4).

Высокой сложностью характеризуются вопросы достижения инсайта, особенно группового с учетом коллективного бессознательного. Это достиже-

ние командой целого, мгновенное озарение, беспричинное прозрение. Как результат страдает целостность охвата и модельного представления с помощью систем искусственного интеллекта решаемых проблем, растут риски прогнозов и иных решений. Эту сложность медитативного уровня можно отнести к очередной, пятой, ловушке искусственного интеллекта.

Перечень ловушек можно продолжить, отметив, что они плод сложившихся стереотипов и накопившихся ошибок в разработке систем искусственного интеллекта. Чтобы в них не попасть, возможно, стоит подумать о способах их предупреждения.

Как избежать ловушек?

Каковы возможные аспекты защиты от упомянутых ловушек? Прежде всего это определение фундаментальных столпов искусственного интеллекта, исследование аналитики трансцендентного, ускорение групповых инсайтов, квантовая интерпретация логических преобразований (квантовая семантика), природоподобное моделирование, определение метафизики эмоций, развитие методов ситуационной осведомленности и виртуального сотрудничества.

Освоение фундаментальных столпов искусственного интеллекта можно проиллюстрировать рис. 5. Так, для обеспечения устойчивости его систем может быть использован потенциал гамильтонова пространства, лагранжиана, теории управляемого хаоса [10]. Целенаправленность же процессов функционирования систем искусственного интеллекта может быть обеспечена за счет использования методов решения обратных задач на топологических пространствах и конвергентного управления [11].

По-видимому, нужна более решительная попытка исследователей по углублению в процесс

моделирования медитативного и коллективного бессознательного с применением методов философии, психологии, физиологии, физики, математики, теории поля, строения элементарных частиц и даже космологии.

При этом в области философии акцентируется внимание на феноменологии, в психологии — на рефлексивной и когнитивной психологии. В охвате медитативного помогут квантовые вычисления и порождаемая ими квантовая семантика. Однако путь, по которому идет развитие квантовых вычислений, также не застрахован от ловушек. Так, известный кубит и его комбинации заманивают возможностью представлять бесконечный объем информации и экспоненциально ускорять вычисления. Одновременно заметим, что эта конструкция отталкивается от бинарной суперпозиции и как следствие может попасть в упомянутые выше ловушки. А посему ускорения вычислений могут оказаться не столь впечатляющими, как ожидалось, а энергетика поддержки суперпозиций пока непосильной.

Вместе с тем явно нельзя отбрасывать квантово-механические эффекты пространственного и временного энтенглмента (*entanglement* — сцепленность, запутанность) и интерференции квантовых состояний (рис. 6). Именно они пока подсказывают возможность воздействия на медитативные процессы любых элементов Вселенной. Однако парадоксы использования семантических интерпретаций, например символов через их бесконечные вселенские дополнения, еще предстоит разрешить. При этом полезным может оказаться анализ адекватности применения гипотезы о теневых фотонах, которые способствуют образованию интерференции волн Де Бройля. Это поможет недоступный медитативный феномен искусственного интел-

Рисунок 5

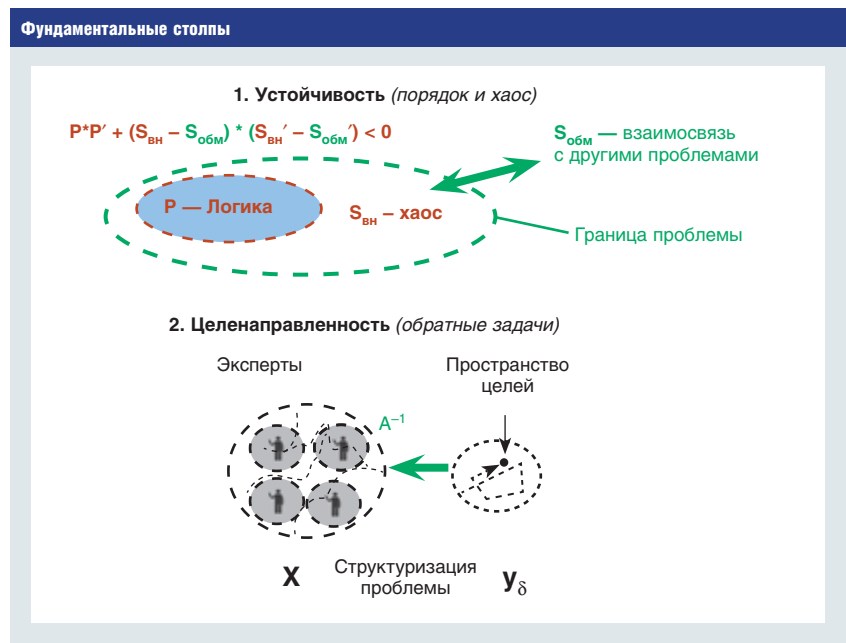
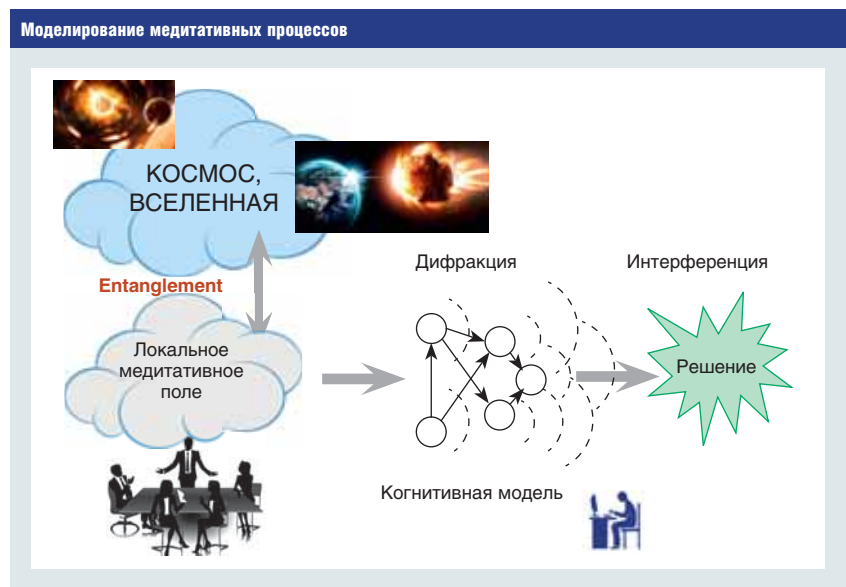


Рисунок 6



лекта проиллюстрировать в виде лазера, луч которого производит дифракцию на «кристалле» когнитивной модели.

Интерес могут представлять известные исследования эффектов, возникающих в среде низкотемпературной газоразрядной плазмы, которая обладает возможностью воздействовать на биологические процессы.

Избежать ловушки дискретности можно через развитие вычислительного подхода с аналоговой оптической реализацией. Так, идея формирования и обработки когерентного оптического излучения с многообразной и многослойной голографической записью сигнала может стать основой для создания такого компьютера, который позволит учитывать весь

Целенаправленность процессов функционирования систем искусственного интеллекта может быть обеспечена за счет использования методов решения обратных задач на топологических пространствах и конвергентного управления.

бесконечный спектр сигнала, мгновенно решать сложные задачи, на которые сейчас тратятся недели и месяцы. Правда, использование этого подхода будет сдерживаться следующими факторами:

- оптический процессор моделирует в лучшем случае элементы мыслительной глубины, но не медитативного пространства;
- создание материала для многогоразовой и быстрой голографической памяти представляет чрезвычайно сложную наукоемкую проблему;
- нужен принципиально новый язык программирования, оперирующий как логикой, так и механизмами преобразования света.

Ну и, конечно, лучшая гарантия избежать любых ловушек — это обучение будущему с охватом большинства перечисленных вопросов.

Что же может стать «точкой сборки» в развитии искусственного интеллекта к концу XXI в.? Ею может стать устройство, неизвестное пока на Земле, каковым может оказаться, например, крохотный — до 1 мм — интеллектуальный летающий аппарат, готовый к целенаправленному полету в стае. В этом устройстве все иное: и механика, и материалы, и управление, и энергетика, и пико-размеры элементов, и что-то еще.

* * *

По всей видимости, попытки моделирования эмоционального и медитативного, а также коллективного бессознатель-

ного — это самое настоящее для искусственного интеллекта в XXI в. Оно может повлиять и на компьютеризацию, и на все развитие инструментов информационного общества. На этом пути стоит ожидать ловушек, связанных с уже сложившимися трендами цифровизации, изыскания причинности событий, фетиша рациональности, забвения феноменологичности и духовности.

Сама постановка вопроса такого эмоционального и медитативного моделирования кажется абсурдной, поскольку эти явления принципиально недоступны моделированию в его сегодняшнем понимании. Здесь от традиционных моделей стоит ждать только робких подсказок.

Чтобы избежать ловушек, требуется дальнейшее формирование

фундаментальных основ искусственного интеллекта. Может оказаться плодотворным квантово-семантический подход, однако его априори нельзя считать адекватным решению рассматриваемого вопроса, поскольку физическая реальность мира частиц, бозонов, фермионов только отчасти может ассоциироваться с медитативными, бессознательными субстанциями.

Вместе с тем уход от дискретности и цифры, погружение в неметрическую понятийную реальность, освоение виртуальных коллаборационистских сред, квантово-семантическая интерпретация медитативного и эмоционального уже наметили просвет в исследовании пока еще многого неясного в решении вопросов искусственного интеллекта.

ПЭС 16104 / 20.06.2016

Источники

1. Агеев А.И. Пять сценариев на век // Экономические стратегии. 2015. № 5–6 (130–131). С. 7.
2. Новиков Д.А. Кибернетика 2.0 // Проблемы управления. 2016. № 1. С 73–81.
3. Лепский В.Е. Эволюция представлений об управлении. Методологический и философский анализ. М.: Когито-Центр, 2015. 107 с.



4. Ageev A.I., Loginov E.L., Raikov A.N. Интеллектуальные технологии организации финансового мониторинга и контроля при реализации госзакупок // Экономические стратегии. 2016. № 1 (135). С. 16–27.

5. Бауэр В.П., Московский А.М., Сильвестров С.Н., Райков А.Н. Ситуационный центр для управления космической промышленностью // Экономические стратегии. 2014. № 5. С. 34–41.

6. Gubanov D., Korgin N., Novikov D., Raikov A. E-Expertise: Modern Collective Intelligence, Springer. Series: Studies in Computational Intelligence, Vol. 558, 2014, XVIII, 112 p.

7. Райков А.Н. «Эксперткратия» как инструмент лоббирования // Президентский контроль. 2010. № 3. С. 26–30.

8. Ильин Е.П. Эмоции и чувства. СПб: Питер, 2002. 752 с.

9. Камалашила (Мэтьюз Э.). Медитация. Буддийский путь покоя и прозрения // Пер. с англ. СПб: Уддияна. 2003. 320 с.

10. Захаров В.Н., Ульянов С.В. Нечеткие модели интеллектуальных промышленных регуляторов и систем управления. II. Эволюция и принципы построения // Техническая кибернетика. 1993. № 4. С. 189–205.

11. Райков А.Н. Конвергентное управление и поддержка решений. М.: ИКАР, 2009. 245 с.

Traps for Artificial Intelligence

Raykov Aleksandr Nikolaevich

Institute Control Sciences of the Russian Academy of Sciences

The course of artificial intelligence development in the XXI century in the context of formation of new technological modes and post-nonclassical management paradigm can be forecasted based on analysis of its many-sided and thorny historical retrospective. This way, of course, is not assured from unexpected traps. However, they can be avoided by finding new solutions in unfamiliar spaces for modeling, based on other approaches to solving complex problems and semantic interpretation of data by synthesizing for these innovations still unknown materials for computer memory and processors.

Keywords

Artificial intelligence, computer philosophy, new paradigm of cybernetics, information society, digitalization.

References

1. Ageev A.I. Pyat' stseneriev na vek [Five Scenarios for a Century]. *Ekonomicheskie strategii*, 2015, no. 5–6 (130–131), p. 7.
2. Novikov D.A. Kibernetika 2.0 [Cybernetics 2.0]. *Problemy upravleniya*, 2016, no. 1, pp. 73–81.
3. Lepskiy V.E. *Evolutsiya predstavleniy ob upravlenii. Metodologicheskii i filosofskiy analiz* [Evolution of Ideas About Management. Methodological and Philosophical Analysis]. Moscow, Kogito-Tsentr, 2015, 107 pp.
4. Ageev A.I., Loginov E.L., Raikov A.N. Intellektual'nye tekhnologii organizatsii finansovogo monitoringa i kontrolya pri realizatsii goszakupok [Intelligent Information-Analytical Technologies for Organizing Financial Monitoring and Control of State Procurement Implementation]. *Ekonomicheskie strategii*, 2016, no. 1 (135), pp. 16–27.
5. Bauer V.P., Moskovskiy A.M., Sil'vestrov S.N., Raikov A.N. Situatsionnyy tsentr dlya upravleniya kosmicheskoy promyshlennost'yu [Situation Centre of Space Industry Control]. *Ekonomicheskie strategii*, 2014, no. 5, pp. 34–41.
6. Gubanov D., Korgin N., Novikov D., Raikov A. E-Expertise: Modern Collective Intelligence, Springer. Series: Studies in Computational Intelligence, Vol. 558, 2014, XVIII, 112 p.
7. Raikov A.N. "Ekspertokratiya" kak instrument lobbirovaniya ["Expertocracy" as a Lobbying Tool]. *Prezidentskiy kontrol'*, 2010, no 3, pp. 26–30.
8. Il'in E.P. *Emotsii i chuvstva* [Emotions and Feelings]. Saint-Petersburg, Piter, 2002, 752 pp.
9. Kamalashila (Met'yuz E.). *Meditatsiya. Buddiyskiy put' pokoya i prozreniya* [Meditation. The Buddhist Path of Peace and Enlightenment]. Saint-Petersburg, Uddiyana, 2003, 320 pp.
10. Zakharov V.N., Ul'yanov S.V. Nechetkie modeli intellektual'nykh promyshlennykh regulyatorov i sistem upravleniya. II. Evolyutsiya i printsipy postroeniya [Fuzzy Models of Intelligent Industrial Controllers and Control Systems. II. Evolution and Construction Principles] *Tekhnicheskaya kibernetika*, 1993, no. 4, pp. 189–205.
11. Raikov A.N. *Konvergentnoe upravlenie i podderzhka resheniy* [Convergent Management and Solutions Support]. Moscow, IKAR, 2009, 245 pp.

