

УДК 004.8

Особенности мышления, или зеркало самосознания для искусственного интеллекта

Куликов Д.К.

Донской государственный технический университет

Аннотация: В статье рассматриваются некоторые аспекты гипотетического взаимодействия устройства с искусственным интеллектом (ИИ) с человеком, в котором целью ИИ является самостоятельное обучение мышлению. Практический смысл этой задачи – обратить внимание на ряд существенных особенностей мышления, важных для разработок ИИ. Анализируются основные вопросы методологии ИИ и ее критика. Обсуждаются те особенности мышления человека, с которыми неизбежно столкнется ИИ, притом, что они обладают признаками неформализуемости, а также наделены положительными эвристическими свойствами и функционально неустранимы из мышления. В их числе рассмотрены аналогии, комплексы и псевдопонятия, противоречие. Показано, что общение – это необходимый для мышления источник новой информации, привносимой в систему и обеспечивающей ее развитие.

Ключевые слова: мышление, искусственный интеллект, аналогия, комплексное мышление, противоречие

Введение. Чему может научиться искусственный интеллект (ИИ) в плане мышления, взаимодействуя с человеком? Подобная постановка вопроса может показаться неожиданной, а сама возможность такого обучения остается гипотетической. Неспособность животных учиться целесообразному поведению через наблюдение за эффектом действий других особей является аргументом в пользу отсутствия у животных мышления. Механическое воспроизведение действий нельзя считать обучением, т.к. оно не регулируется *субъективной формой, отличной от ситуации действия*, не содержит *знания* об их смысле и глубинных основаниях. Не является обучением и акт распознавания целевой формы действия, поскольку такое распознавание уже опирается на знание о смысле наблюдаемых движений. Обучение мышлению не может происходить без познания того, что представляет собой действительность мышления. Наша постановка вопроса лишь принципиально заостряет проблему оценки реальных возможностей ИИ в области познания. Из обучения нельзя исключить момент активного

саморазвития и самообучения. Поэтому уточним нашу задачу: на что следует обратить внимание творцу ИИ, если он ставит цель создать машину, способную *самостоятельно* научиться мыслить у мыслящего существа?

Исследования в области ИИ развиваются в направлении разработок все более эффективных программ и аппаратов (робототехники, экспертных и управляющих систем). Но остается открытым вопрос, насколько положенная в основу исследований модель интеллекта адекватна задаче имитации или воссоздания умственной деятельности человека. Как отмечает Д. Маккарти (автор термина Artificial Intelligence), мы все еще плохо знаем природу интеллекта человека, и этим объясняется то, что реально достигнуто в исследованиях ИИ [9]. Методология исследования ИИ опирается на принцип, согласно которому существует универсальная способность к интеллектуальной деятельности, реализуемая в разной степени на уровне высокоорганизованной органической материи (у животных и человека), а также на высокотехнологичном машинном уровне. Во всех случаях *основание интеллекта* ищут в структурной и функциональной организации *единичного* организма или аппарата. В этом направлении исследования подталкиваются успехами в области нейрофизиологии и нейроинформатики. Критические замечания обращают внимание исследователей на редуционизм их методологии, а именно на сведение интеллектуальной деятельности к *переработке информации*, на необходимость учитывать *сверхиндивидуальные* компоненты интеллекта, момент *творчества*, роль *интуиции*, а также практический характер отношений человека к действительности. Главная задача статьи – обратить внимание на некоторые особенности мышления человека и их позитивные функции.

Искусственный интеллект и его возможности. С чем мог бы подойти носитель ИИ, вступая в познавательное взаимодействие с человеком? В его определениях *интеллект* – это момент общей способности агента достигать

цели в мире, обеспеченной процессами *вычисления*. Это также способность реализовывать *наилучшее достижение цели* с точки зрения доступных и учтенных возможностей. Процедура вычисления – это последовательность действий по выведению результата, и, выполняя шаги выведения, следует лишь «механически следовать предписаниям, как если бы мы были роботами: от нас не потребуются ни понимания, ни искусства, ни изобретательности» [4, с. 270]. *Познавательное* содержание такой рациональной формы сводится к поиску алгоритма для ответа «да» или «нет» на любой вопрос из потенциально бесконечного класса. Со времени изобретения первых интеллектуальных программ А. Ньюэллом и Г. Саймоном, исследования реализуют установку на имитацию операций мышления, которые использует человек при решении проблем в области вычислений и дедуктивных умозаключений, поскольку они успешно формализуемы средствами математики и логики.

С чего начнет машина, поставленная перед задачей отличить в деятельности человека мышление от того, что им не является? Согласно *эвристической гипотезе*, все люди в норме обладают одинаковыми интеллектуальными механизмами, а индивидуальные различия имеют психофизиологическую и биохимическую природу. А. Тьюринг решал эту проблему введением *джентльменского соглашения*: мы исходим из того, что все люди мыслят, поэтому все связанное поведение можно оценивать как проявление мышления. Однако, общаясь с человеком, машина неизбежно заметит оценочную сторону в его действиях, поскольку человек стремится для себя и другого *различать ментальные состояния*, квалифицируя их по эвристической шкале, в частности как *проницательные*, либо *глупые* и *неразумные* (при этом *не измеряя их* в техническом смысле).

Изучая процесс рассуждения, машина обнаружила бы, что человек редко может выразить его единообразным способом и указывает разные

стратегии мышления. Интеллектуальный продукт непредсказуемым образом обнаруживается личностью в разных когнитивных модальностях и разными способами – последовательным алгоритмическим перебором, мозговым штурмом, случайным поиском, в результате озарения и фантазии (в том числе во сне). Обратным образом, многие виды активности ИИ, например, распознавание образов и игра в шахматы, могут осуществляться методом перебора данных *без привлечения функций мышления.* Это имеет существенное значение для интерпретации его способностей как качественно отличных от человеческих. Внимание привлекает трактовка разумной деятельности как системы неформализуемых стратегий, но определенных предметно-содержательными критериями. Недостаточно ясна роль *эмоций, личностного знания и самосознания* для практики моделирования интеллекта, хотя на философском уровне эти вопросы обсуждаются широко.

Показательно, что сфера применения ИИ остается достаточно ограниченной. Это область деятельности, связанная с оперативным и быстрым вычислением и логистикой, экспертиза, поисковые агенты, игры и пр. Специфически человеческие сферы познания, творчества и нравственных отношений остаются вне компетенции ИИ. Например, использование «Электронного судьи» возможно лишь для анализа показаний и вещественных доказательств, а также при вынесении приговоров по несложным делам вроде ДТП [8].

Есть ли существенная разница между поведением пчелы, строящей идеальные шестигранники сот, и творческой работой Ф. Кекуле, открывшего во сне форму бензольного кольца? Разве действия пчелы не тождественны созданию эстетической архитектурной формы? Суть, однако, не в сходстве форм, а в специфике движущих сил человеческой деятельности и условиях, которые ее организуют и поддерживают. И хотя машина совершает работу по

проработке прототипов и алгоритмов в своей электронной «голове», а у пчелы мир тоже репрезентируется в нейронных связях ее гигантского по меркам насекомых мозга, попытки формализовать функции программ в указанных областях резко ограничены. Хороший пример – работы В.О. Лобовикова [6], посвященные формализации моральных и правовых модальностей. Пользоваться переменными типа «хорошо» и «плохо» можно лишь в пределах ригористической этики и норм естественного права, что упрощает этическое измерение жизни человека.

Аналогии. Это исторически древняя и онтогенетически ранняя форма мышления, связанная с естественной способностью к индукции и образованию ассоциативных связей по сходству и пространственно-временной смежности. В палеомышлении подобие воспринимается как тождество. Неспособность различать подобие и идентичность свойственна детскому возрасту, однако при определенных обстоятельствах она встречается и у взрослых людей. Аналогии ценны в плане рефлексии и способности к проблематизации данных, которыми оперирует мышление. Основной ошибкой аналогии является поспешный вывод о полном сходстве признаков разнородных, но подобных объектов. Пусть $\{a,b,c,d,e\} \equiv X_{(A)}$, и $\{a,b,c,d\} \equiv Y_{(B)}$. Тогда $(Y_{(B)} \rightarrow e) \vdash e_{(B)}$. Связка $Y_{(B)} \rightarrow e$ может выпасть из рассуждения, приводя к ошибочным выводам. Аналогия требует определения *меры допустимости нестрогого вывода*. В мышлении эта мера поддерживается интуитивной оценкой в каждом конкретном случае, а главное, предполагает *момент сомнения*. В аппарате оценка аналогии – функция анализа данных.

Аналогии задействованы в экспертных системах. Машина пользуется прототипическими моделями для распознавания по аналогии тех или иных информационных паттернов. Хотя в мышлении роль прототипов велика, человек способен произвольно создавать модели, условные связи и

оперативно их менять. Наш ИИ мог бы озадачиться тем, что некоторые люди легко укажут на логическое различие суждений «мир есть конь» и «мир есть материя» (примеры Ю. Лотмана [7, с. 525]). В суждениях *архаического типа* происходит *отождествление нетождественного* на основании усматриваемого сходства, либо в силу эмоционального подкрепления. Такое отождествление порождает мифологические и художественные образы. В *теоретическом суждении* связь понятий образуется не по аналогии, а через *переход субъекта в свои определения*. Аналогии также лежат в основании *метафор*, в которых достигнут определенный уровень *сознательного сближения значений*. Различие этих типов суждений не регулируется логической и синтаксической формой их построения.

Комплексы и псевдопонятия. Отмечается, что ИИ – это неразвивающаяся сущность, и в этом его недостаток в сравнении с органической природой мышления. Самые ранние этапы онтогенеза психики заключают в своем содержании важные моменты интеллекта. На каждом уровне возникают *специфические формы мышления*, которые являются основой для развития на более высоком уровне. Л.С. Выготский экспериментальным путем выявил форму, названную *комплексным мышлением*. Она соответствует этапу допонятийного мышления, специфической чертой которого является опора на некатегориальные способы организации умственной деятельности. Всеобщий характер этих форм подтвердили дальнейшие исследования. *Комплексы* – это случайные, ситуативные связи множества предметов, или связи, значимые с точки зрения практического отношения людей к предмету и понимания этого отношения. *Псевдопонятие* – это форма стихийного обнаружения существенных характеристик, которая не осознается самим субъектом. Такие формы мышления возникают при решении задач повседневного уровня, не предполагающих высокого уровня теоретических знаний. Это форма

стихийно развивается под действием языковых значений, создающих обобщения и связи независимо от объективного знания.

Можно ли отбросить этот феномен как пример несовершенства человеческого мышления в сравнении с формализованными интеллектуальными системами? Действительно, комплексное мышление – это источник множества предрассудков и поверхностных представлений. Однако эксперимент Выготского показал, что комплексы – это также момент развития мысли как процесса *познания алгоритма*, его построения. Ошибки в этом процессе являются необходимым условием. Также в комплексе предмет и его объективные характеристики приобретают связность социальной определенности, например, свинина и говядина как *пища* вообще и как *съедобная/несъедобная пища* (для иудеев). Начало теоретического мышления – это сдвиг в сторону *различения социально определенных и объективных свойств* предмета (значимости и истинности утверждения), который недоступен машине.

Противоречие. Формализация алгоритмов вывода построена на законе противоречия: $(\overline{a\bar{b}})$. Как показали исследования, человеческое предметное мышление, ориентированное на познание действительности вне сознания, вне языка и программы, – это мышление диалектическое. Оно свершается в противоречии как реальной логической форме, и как выразился Гегель, противоречие есть критерий истины, а его отсутствие (прежде всего для мышления) есть критерий заблуждения [1, с. 265]. «Здесь же и отличие человеческого мышления от психики любого млекопитающего, а также от действий счетно-вычислительной машины. Последняя тоже приходит в состояние «самовозбуждения», очень точно «моделирующее» истерику собаки в опытах Павлова, когда на ее «вход» подаются разом две взаимоисключающие команды – «противоречие». Для человека же появление противоречия – сигнал для включения мышления, а не истерики» [3, с. 171].

Решение проблемы требует разрешения заключенного в ней противоречия, и это решение требуется искать не в регистрах памяти, а в творческом преодолении конфликта: найти средство в самой реальности или создать его впервые.

Шаблон пропозиционального вывода, который использует ИИ, описывается рассуждением по форме *modus ponens*: *если впереди объект А, то агент совершает действие В*. Однако и рассуждение, и поведение человека нельзя описать таким шаблоном, хотя внутри логики его мышления и поступков *modus ponens* универсален. Это описание действия, скорее, не требующего рассуждений: под него подходит, например, рефлекторный акт, автоматизм, ритуализованное действие. В лучшем случае, это снятое мыслью отношение причины и действия. Мыслящее поведение описывается схемой $a \rightarrow (b \vee \bar{b})$, т.е. включает в себе решение задачи выбора. Логика этой формулы содержит в себе перевода переменной \bar{b} в конкретное предметное значение (если b переменная, то $\bar{b} \equiv c$). *Если я вижу вражеского солдата, стрелять в него, или не стрелять?* Практически «не стрелять» – это не абстрактное бездействие, но какие-либо иные действия (прятаться, наблюдать, связаться со штабом, сдать и пр.).

В другом случае речь идет о выходе мышления на противоречие $a \rightarrow (b \wedge \bar{b})$, решение которого нельзя найти в пределах логического алгоритма, но лишь в действительном отношении (в предметной связи). «*Яма есть и ямы нет*» означает для мышления, что «*яма есть физически, но она преодолевается практически*» (например, перекрытием). В наличии ли это средство, либо нет, для мышления не может быть заданным обстоятельством. Формализация подобных обстоятельств упирается в проблему бесконечного исчисления. Импликация $a \rightarrow (b \wedge \bar{b})$ формально ложна при истинности антецедента, но истинна как утверждение неправдоподобия или абсурда («если пациент Наполеон, то некто жив и мертв»).

Рассмотрим ситуацию с явлением капитала, проанализированную Э.В. Ильенковым [2]. Если дано явление капитала (a), то оно возникло из обращения и не из обращения $b\wedge\bar{b}$. С точки зрения формальной логики, это невозможно. Однако, это *логическая* модель такого *исторического* отношения, где $b\wedge\bar{b}$ воспроизводится как необходимое условие самовозрастания капитала (особый товар, создающий новую стоимость – рабочая сила). И тогда в качестве предпосылки формула $b\wedge\bar{b}$ должна браться как антецедент: $(b\wedge\bar{b})\rightarrow a$. В этих условиях импликация истинна независимо от значения истинности конъюнкции.

Конъюнкция $b\wedge\bar{b}$ может быть положена как истинная в форме идеи, например, когда мы действуем в предметной среде с планом ее изменения. Это условие планирования, причем не созерцательного, а деятельного, включающего измерение, расчеты, *согласование действий*. Конъюнкция $b\wedge\bar{b}$ может быть также положена как истинная *для другого*, предполагая *изменение его отношения к предмету и его восприятия действительности*. На этом держится, например, ритуализованное действие: священный и неприкасаемый предмет означает именно $b\wedge\bar{b}$, где b есть предмет в его физической определенности, а \bar{b} – его идеальная положенность. Здесь лежит поле дологических форм мышления и значение абсурда для выработки собственно логических форм мысли.

Дискуссия. Прежде всего, обращает на себя внимание *богатство действительного мышления, его форм при ограниченных способностях ИИ*. В работе ИИ нет *момента развития*, который присутствует у человека и в котором заключены его существенные моменты. Программируются интеллектуальные схемы, соответствующие формальным операциям. При этом оставляются без внимания способности, развитые ребенком на *ранних этапах онтогенеза*. С этим связан тот факт, что многие вещи люди делают легко и свободно, тогда как ИИ для этого требуется избыточный объем

вычислений. Историческое развитие синтезирует все разнообразие опыта коллективной деятельности человека, транслируя его форму и содержание в интеллектуальное и идейное достояние личности. Мыслящее устройство не может работать в изоляции [7, с. 583], но и человек нуждается в общении. Необходимо решение проблемы *неустранимости человека* из сферы деятельности ИИ. Данную проблему следует разделить на две: 1) причины и условия необходимости системы «человек-машина»; 2) перспективы автономного взаимодействия интеллектуальных агентов между собой *без человека* (это, в свою очередь, ставит вопрос о безопасности таких автономных систем).

Общение и его участники – это источник новой информации, т.е. содержания, привносимого в систему. Это неформализуемый момент, внешний по отношению к интеллектуальной системе в измерении *языка и познания*. Для интеллекта и его развития недостаточна *дихотомия «язык/речь»*, но важна *трехсоставная структура*, в которой помимо языка-системы и речевой деятельности необходимо выделить *язык-текст*. В тексте лингвистическая система (синтаксис+словарь) зафиксирована содержательно, а именно, как осмысленный субъектами *тезаурус*. Он не формализуем и неполон, требуя для своего определения *внелингвистической реальности* [5]. В познании подобное обстоятельство связано с функцией противоречия. Важное, что следует усвоить для ИИ: человек разрешает противоречия *практически*, создавая новые предметные отношения и полагая их в качестве истинных (в том числе недостатки индукции и нестрогих аналогий).

Литература:

1. Гегель Г. Работы разных лет. В 2-х тт. Т.1. М.: Мысль, 1970.



2. Ильенков Э.В. О материальности сознания и трансцендентальных кошках // Дialeктическое противоречие. М.: Политиздат, 1979.
3. Ильенков Э.В. Об идолах и идеалах. К.: «Час-Крок», 2006.
4. Клини С.К. Математическая логика. М.: Мир, 1973.
5. Куликов Д.К. Гипотеза лингвистической относительности Сепира-Уорфа и её методологическая оценка в работах М.К.Петрова // Политическая концептология. 2012, №2. с. 189-199.
6. Лобовиков В.О. Естественное право как математика свободы и морально-правовое программирование автономных адаптивных роботов // Новое в искусственном интеллекте. – М.: ИИнтелЛ, 2005.
7. Лотман Ю.М. Семиосфера. СПб.: Искусство-СПб, 2001.
8. Панчул Ю. Взлет и падение искусственного интеллекта // panchul.com/2008/02/11/artificial-intelligence/
9. McCarthy J. What is artificial intelligence? // formal.stanford.edu/jmc/whatisai/whatisai.html

References:

1. Gegel' G. Raboty raznyh let. V 2-h tt. T.1. М.: Mysl', 1970.
 2. Il'enkov Je.V. O material'nosti soznaniya i transcendental'nyh koshkah // Dialekticheskoe protivorechie. М.: Politizdat, 1979.
 3. Il'enkov Je.V. Ob idolah i idealah. К.: «Chas-Krok», 2006.
 4. Klini S.K. Matematicheskaja logika. М.: Mir, 1973.
 5. Kulikov D.K. Gipoteza lingvisticheskoy otnositel'nosti Sepira-Uorfa i ejo metodologicheskaja ocenka v rabotah M.K.Petrova // Politicheskaja konceptologija. 2012, №2. s. 189-199.
 6. Lobovikov V.O. Estestvennoe pravo kak matematika svobody i moral'no-pravovoe programmirovanie avtonomnyh adaptivnyh robotov // Novoe v iskusstvennom intellekte. – М.: IntelLL, 2005.
-



7. Lotman Ju.M. Semiosfera. SPb.: Iskusstvo-SPb, 2001.
8. Panchul Ju. Vzlet i padenie iskusstvennogo intellekta // panchul.com/2008/02/11/artificial-intelligence/
9. McCarthy J. What is artificial intelligence? // formal.stanford.edu/jmc/whatisai/whatisai.html