

Компьютерное моделирование интеллектуальных функций¹

Настоящее выступление будет посвящено самым общим методологическим основаниям исследований по созданию искусственного интеллекта, которые занимают в данный момент в сфере междисциплинарных научно-практических областей одно из центральных мест. Это направление представляет собою интерес для философского исследования, поскольку воплощает в компьютерных системах важнейшие идеи и подходы, выработанные в рамках целого ряда гуманитарных дисциплин и в рамках современного естествознания. Компьютерное моделирование и искусственный интеллект — это воспроизведение на искусственном носителе не естественного интеллекта, но наших *представлений о естественном интеллекте и его функциях*. Как пишет отечественный исследователь проф. А.П. Огурцов, искусственный интеллект — это не «органопроекции нашего интеллекта, а органопроекции наших представлений об интеллекте и его актах»².

Объектом философского исследования в данном случае оказываются теоретическая и практическая деятельность, связанная с осознанием структуры интеллекта и попытками реализации этих структур в компьютерных науках. Как правило, представления о функциях мышления заимствуются из теоретизированных концепций, принадлежащих области философии и таких конкретных наук, как: психология, логика, кибернетика, нейробиология, физиология высшей нервной деятельности и т.п. Заимствование представлений о характере и структуре интеллектуальной функции носит фрагментарный характер, и полного изоморфизма между структурами теоретических концептов и принципами организации компьютерных моделей не наблюдается. В работах по искусственному интеллекту представление о моделируемых функциях остается почти не эксплицированным.

В сегодняшнем выступлении я попытаюсь привести примеры, с одной стороны, теоретических концепций, репрезентирующих некоторое представление о функциях и устройстве естественного интеллекта, с другой стороны, компьютерных моделей интеллектуальных функций. В ходе исторического развития человечество накопило огромное количество определений естественного интеллекта. Однако не все идеи оказали влияние на реальное развитие исследований в области искусственного интеллекта.

Кратко остановимся на ключевых этапах становления исследований в области искусственного интеллекта. Официальной датой рождения искусственного интеллекта считается 1956 год. По инициативе одного из самых авторитетных специалистов по данной тематике Дж. Маккарти в Дартмутском колледже состоялся специализированный двухмесячный семинар. В роли соорганизаторов выступали М. Минский, А. Ньюэлл, Г. Саймон. В работе семинара приняли участие десять крупнейших американских ученых, включая Т. Мура из Принстонского университета, Р. Соломонова и О. Селфриджа из Массачусеттского технологического института. На семинаре были сформулированы основные задачи, которые предстояло решить но-

¹ Доклад основан на материалах исследований, проведенных в рамках гранта Российского Фонда Фундаментальных Исследований №07-06-00066а.

² Огурцов А.П. Достижения и трудности в моделировании интеллектуальных актов // Материалы Всероссийской междисциплинарной конференции «Философия искусственного интеллекта», г. Москва. М.: ИФ РАН, 2005. С. 56.

вому научному направлению. Были рассмотрены такие аспекты искусственного интеллекта, как: компьютерное программирование, способность машины к пониманию естественного языка, нейронные сети, проверяемость, креативность и многие другие. Благодаря Дартмутскому семинару в научном мире закрепилось название «Artificial Intelligence».

На самом раннем этапе развития исследований по искусственному интеллекту ключевую роль в формировании области сыграли работы У. Маккалока и У. Питтса¹, но в особенности идеи А. Тьюринга², изложенные им в статье «Computing machinery and intelligence». В результате разработок ранних идей У. Маккалока, У. Питтса³ по созданию нейронных сетей, а также благодаря работам Ф. Розенблатта⁴ сформировались так называемые *коннекционистские*, или структурные, модели интеллектуальных систем. В 70-е годы своим развитием это направление обязано, в первую очередь, работам Дж. Хопфилда⁵.

Благодаря разработкам А. Ньюэлла, Г. Саймона⁶ оформился подход, основанный на эвристическом поиске. Затруднения на этом пути вызвали к жизни новое решение, получившее название моделей, основанных на знаниях. Инициаторами данного направления принято считать группу ученых Стенфордского университета (Э. Фейгенбаум, Б. Бьюкенен, Дж. Ледерберг), разработавших первую экспертную систему DENDRAL⁷. Сложившееся направление обычно классифицируют как *логическую парадигму* в исследованиях искусственного интеллекта.

Примерами реализации указанного логического подхода, или даже логической парадигмы, как называет его Д.А. Поспелов, являются многочисленные работы в области автоматического доказательства теорем, разработки языков представления знаний логического типа (язык PROLOG). Такой подход породил «экспертные системы, основанные на продукционных правилах, теорию реляционных баз данных, теорию решателей и планировщиков»⁸. Рассмотрим вопрос о том, какие теоретические представления об интеллектуальных функциях были использованы при разработке конкретных кибернетических моделей в рамках этого направления.

Предлагаю более подробно остановиться на описании исторически первой реализации логического подхода — программ автоматического доказательства теорем Логик-теоретик (Logic Theorist, далее LT) и Общий Решатель Задач (General Problem Solver, далее GPS).

¹ McCulloch W.S., Pitts W. A Logical Calculus of Ideas Immanent in Nervous Activity // Bull. Mathematical Biophysics, 1943. V. 5.

² Turing A. Computing machinery and intelligence. Mind. 59, 1950. С. 433—460. Русский перевод: Тьюринг А. Может ли машины мыслить? М.: ГИФМЛ, 1960.

³ McCulloch W.S., Pitts W. A Logical Calculus of Ideas Immanent in Nervous Activity // Bull. Mathematical Biophysics, 1943. V. 5.

⁴ Rosenblatt F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spatan, Chicago, 1962.

⁵ Hopfield J.J. Neurons with graded responses have collective computational properties like those of two-state neurons. Proceedings of National Academy of Science of the United States of America, 79. С. 2554—2558.

⁶ Newell A., Simon H. Empirical explorations with the Logic Theory Machine: a case study in heuristics // Feigenbaum E.A., Feldman J., ed. Computers and Thought. New York: McGraw-Hill, 1963. Русский перевод: Ньюэлл А., Саймон Г. Эмпирические исследования машины «Логик-теоретик»; пример изучения эвристики // Фейгенбаум Э., Фельдман Дж. Вычислительные машины и мышление. М.: Издательство «Мир», 1967.

⁷ Lindsay R.K., Buchanan B.G., Feigenbaum E.A., Lederberg J. Applications of artificial intelligence for organic chemistry: the DENDRAL project. NY: McGraw-Hill, 1980.

⁸ Поспелов Д.А. Десять «горячих точек» в исследованиях по искусственному интеллекту // Интеллектуальные системы (МГУ). 1996. Т.1, вып.1—4. С. 47.

Впервые работа программы LT¹ была продемонстрирована её создателями А. Ньюэллом и Г. Саймоном на знаковой для всей истории области искусственного интеллекта уже упоминавшейся Дартмутской конференции летом 1956 года. Программа LT продемонстрировала свои возможности, предложив доказательства для большинства теорем, содержащихся во второй главе *Principia Mathematica* Б. Рассела и А. Уайтхеда. Известно также, что программа предложила более короткое доказательство одной из теорем, нежели чем это было предложено Расселом и Уайтхедом². Сегодня предлагаю более подробно остановиться на идеях, лежащих в основе программы GPS. В программе GPS целью работы, как и в LT, является приведение некоторого данного исходного выражения к некоторому заданному конечному выражению.

Принципиально в программе различаются: объекты, то есть некоторые ситуации, например, структуры задач математической логики; цели, то есть ситуации, которые следует достичь; а также операторы. К программе прилагается список таких операторов, то есть способов преобразования (12 правил), которые могут быть использованы в отношении используемых программой объектов. В качестве объектов могут выступать формулы из букв (P, Q, R, ...) и логических связок ($\sim$ — «нет», $\supset$ — «влечет за собой», $\bullet$ — «и», $\vee$ — «или»). В программе используется три типа целей. Первый тип цели состоит в том, чтобы преобразовать объект a в объект b. Для этого производится их сравнение и находится различие d. Если различия нет, то цель считается достигнутой, если есть, то подцелью считается уменьшить это различие d путем преобразования a посредством данных операторов в c. Далее требуется снова преобразовать уже c в b. Второй тип цели состоит в выяснении возможности применения некоторого оператора q к a. Если такое применение возможно, то его производят, если нет, то необходимо выяснить различие между a и некоторым объектом, к которому оператор q применим. Далее путем применения цели первого типа a преобразуется в a1, к которому далее применяется оператор q. Третий тип цели — это уменьшение отличия между объектами при помощи поиска подходящего оператора. В целом используемая в GPS модель решения задач основана на двух компонентах: процедура сравнения двух объектов и уменьшения различий между ними (1); таблица из 12 операторов, описывающая взаимосвязи различий (2).

Создатели GPS считают, что не следует проводить четкую границу между попыткой выполнить при помощи машины задачи, решаемые человеком, и попыткой моделировать процессы, которые человек действительно использует при решении этих же задач. GPS позиционируется как программа, которая «совмещает оба эти подхода с пользой как для одного, так и для другого»³. В своей статье «GPS — программа, моделирующая процесс человеческого мышления» А. Ньюэлл, Г. Саймон сравнивают методику решения одной и той же формальной задачи по преобразованию исходного символического выражения в целевое символическое выражение человеком и программой GPS. При этом испытуемый, в рассматриваемом случае студент, может использовать тот же набор операторов, что и программа GPS. Целью эксперимента является доказать, что испытуемый пользуется в ходе решения задачи теми же эвристически-

¹ Описание программы опубликовано в следующей статье: Newell A., Simon H. Empirical explorations with the Logic Theory Machine: a case study in heuristics // Feigenbaum E.A., Feldman J., ed. *Computers and Thought*. New York: McGraw-Hill, 1963, или Ньюэлл А., Саймон Г. Эмпирические исследования машины «Логик-теоретик»; пример изучения эвристики // Фейгенбаум Э., Фельдман Дж. Вычислительные машины и мышление. М.: Издательство «Мир», 1967. С.113–144

² Рассел С., Норвич П. Искусственный интеллект: современный подход. М.: Издательский дом «Вильямс», 2006. С. 56.

³ Ньюэлл А., Саймон Г. GPS — программа, моделирующая процесс человеческого мышления // Фейгенбаум Э., Фельдман Дж. Вычислительные машины и мышление. М.: Издательство «Мир», 1967. С. 283.

ми методами, которые использует и машина, а именно, описанным выше методом уменьшения различия. При решении задачи испытуемый не только записывает на доске свои промежуточные результаты, но также по просьбе экспериментаторов рассуждает вслух. Устные и письменные рассуждения испытуемого вносятся в письменный протокол.

Проанализировав протокол испытуемого и результаты работы программы GPS, авторы выделили несколько различий. Приведем некоторые из них. Во-первых, некоторые правила используются испытуемым, но он не сообщает об этом ни письменно, ни устно, из чего можно сделать вывод, что он производит некоторые операции либо подсознательно, либо просто «в уме». В любом случае некоторые мысленные операции оказываются в случае с человеком скрытыми от наблюдения. Авторы указывают на неспособность GPS различать между внутренним и внешним миром субъекта. Во-вторых, испытуемый, судя по его устным высказываниям, рассматривает несколько вариантов решения параллельно, на что не способна предлагаемая программа. В-третьих, фраза испытуемого «мне следовало бы...» выражает способность человека возвращаться по ходу решения на несколько шагов назад и просматривать решение заново, что, как считают авторы, необходимо включить в возможности программы.

Ученые ставили своей целью не только создание машины, способной решать задачи, требующие интеллектуальных усилий, аналогичным человеку способом, они также ставили себе целью создание теории процесса решения задач человеком. Так, например, «можно мыслить себе «разумную» программу, которая будет манипулировать символами так же, как и наш испытуемый; при подаче на ее вход выражений символической логики она будет на выходе выдавать последовательность примененных правил, совпадающих с правилами, примененными испытуемым. Если мы будем наблюдать извне за действиями такой программы, мы обнаружим, что она рассматривает различные правила и оценивает различные выражения, то есть делает то же самое, что мы обнаруживаем в протоколе опытов с человеком»¹. Эта мысль интересна для данного исследования не столько с точки зрения стремления ученых-кибернетиков повлиять на психологическую теорию, сколько для указания на одно принципиальное свойство создаваемой разумной машины, а именно на способность порождать объяснения.

В рассмотренной компьютерной модели проявляется ряд существенных характеристик интеллекта, источниками которых являются мировоззренческие установки, понимание интеллекта, сформировавшееся в философии Античности, философии Нового времени, классической немецкой философии, в рамках классической логики и философии математики. Остановимся на некоторых ключевых характеристиках. Для теории познания философии Древней Греции характерны такие черты, как: *различение чувственного познания и умозрения* как различных источников и способов познания, направленность процесса познания на постижение истины. В философии Аристотеля мы встречаем попытку систематизации тех способов, которыми ум достигает достоверного знания. Рассуждение, исходящее из истинных посылок, проводимое по правилам силлогизмов, то есть по правилам формальной логики, является доказательством и ведет к истинному знанию. Тем самым, мышление представляется как процесс доказательства, вывода корректных заключений из истинных посылок по некоторым формальным правилам, правилам, которые сформулированы логикой.

При рассмотрении попыток автоматизированного доказательства, таких, как программы LT и GPS, все более необходимым становится обращение к неявным структурам человеческого мышления, выявлению таких характеристик интеллектуальной

¹ Там же. С. 289.

деятельности, которые не были рассмотрены логикой. Структура, о которой здесь идет речь, касается процедур выбора последующего шага, ограничения перебора возможных вариантов.

Интересно проследить теоретические предпосылки представления о таком структурном компоненте интеллекта, как *интуиция*. Обратимся к философии математики XX века и более ранним идеям, сформировавшимся в рамках рационализма Нового времени и немецкой классической философии.

Сходные мысли по поводу получения нового математического результата высказывал А. Пуанкаре. «Творчество состоит как раз в том, — пишет он, — чтобы не создавать бесполезных комбинаций, а строить такие, которые оказываются полезными. Творчество — это отличать, выбирать»¹. Пуанкаре обращает внимание на важность отбора значимых комбинаций, который оказывается ключевым моментом и в разработке современных автоматических систем доказательства теорем.

Пуанкаре пишет об этом процессе следующее: «Математическое доказательство представляет собой не просто какое-то нагромождение силлогизмов: это силлогизмы, расположенные в известном порядке, причем этот порядок расположения элементов оказывается гораздо более важным, чем сами элементы»². Процесс доказательства представляется ему не просто механическим выводом из посылок всех возможных заключений по некоторым правилам, поскольку такой подход порождает множество бесполезных вариантов, но выбор среди полезных комбинаций. Человек обладает уникальной способностью «освобождать себя от труда создавать эти бесполезные комбинации»³.

Когда речь заходит о тех возможных правилах, которыми руководствуется человек, делая подобную выборку, они оказываются «крайне тонкого и деликатного характера, они явственно чувствуются, но плохо поддаются формулировке», в итоге оказывается, что это результат интуиции. Работа интуиции, по Пуанкаре, происходит на подсознательном уровне. Результат этой работы проявляется на сознательном уровне уже в качестве «исходных точек», набора «полезных комбинаций». Процесс математического доказательства Пуанкаре подразделяет на два этапа: первый — бессознательный, второй — сознательный. На бессознательном уровне происходит предварительная работа по выбору наиболее вероятных полезных комбинаций. Затем эти комбинации попадают на сознательный уровень: «в поле сознания появляются действительно полезные комбинации, да еще некоторые другие, которые он <исследователь>, правда, отбросит в сторону, но которые не лишены характера полезных комбинаций. Все происходит подобно тому, как если бы изобретатель был экзаменатором второй ступени, имеющим дело лишь с кандидатами, успешно прошедшими через первое испытание»⁴.

Второй уровень представляет собой проверку бессознательных догадок. Такие вычисления отличаются строгостью и сложностью, поэтому, как пишет Пуанкаре, «требуют дисциплины, внимания, воли и, следовательно, сознания»⁵. Задача, которая стоит перед Пуанкаре, — доказать, что на бессознательном уровне действительно работают именно интуитивные механизмы.

Тезис о том, что бессознательная работа представляет собой некоторую механическую работу по получению максимального количества комбинаций, представляется ему неправомерным. С одной стороны, если признать механический характер бессознательной работы, то «таинственной» представляется причина

¹ Пуанкаре А. Наука и метод // О науке. М.: Наука, 1990. С. 403.

² Там же. С. 402.

³ Там же. С. 408.

⁴ Там же. С. 404.

⁵ Там же. С. 413.

того, почему «среди тысяч продуктов нашей бессознательной деятельности одним удастся переступить порог сознания, тогда как другие остаются за его порогом»¹. С другой стороны, Пуанкаре указывает на определенный характер результата бессознательной работы. Этот результат не соответствует по форме результату механической работы. «Никогда не случается, чтобы бессознательная работа доставила вполне готовым результат сколько-нибудь продолжительного вычисления, состоящего в одном только применении определенных правил. Казалось бы, что абсолютное «я» подсознания в особенности должно быть способно к такого рода работе, являющейся в некотором роде исключительно механической. Казалось бы, что думая вечером о множителях какого-нибудь произведения, можно надеяться найти при пробуждении готовым самое произведение или, еще иначе, что алгебраическое вычисление, например, проверка, может быть выполнено помимо сознания. Но в действительности ничего подобного не происходит, как то доказывают наблюдения»².

В противоположность сознательному этапу вычислений на бессознательном этапе имеет место отсутствие дисциплины и порядка. Таким образом, математическое доказательство задействует как бессознательные, так и сознательные механизмы, причем, часто подсознательному этапу предшествует этап предварительной сознательной работы. На бессознательном уровне господствует интуиция, на сознательном — строгие, четкие правила логики. Пуанкаре однозначно ставит проблему перехода к осознанию полезных вариантов. Он пишет, «лишь некоторые среди них оказываются гармоничными, а, следовательно, полезными и прекрасными в то же время; именно они сумеют разбудить ту специальную восприимчивость математика, последняя же, однажды возбужденная, со своей стороны, привлечет наше внимание к этим комбинациям и этим даст возможность переступить через порог сознания»³. Пуанкаре четко фиксирует мысль о том, что творческая выборка полезных вариантов представляет собой бессознательную работу человеческой интуиции. В.Ф. Асмус указывает на то, что у Пуанкаре «интуиция» выступает и как синоним математической «догадки», математического вдохновения, и как принцип математического рассуждения, как условие математической дедукции. В данном рассмотрении ценность представляет именно понимание интуиции как процесса, предполагающего *наличие в доказательстве, в качестве необходимого условия, элементов, не входящих в строгую дедуктивную структуру*. Пуанкаре показывает, что из системы математических рассуждений оказывается невозможным удалить полностью те элементы, которые в своей основе имеют не логику, но интуицию, что ограничивает возможности логического подхода к моделированию интеллекта.

Идеи Пуанкаре не только сыграли огромную роль в истории математики и в целом в науке XX века, но и повлияли на ученых, работающих в области кибернетики. Д.А. Поспелов считает, что вопросы, поднятые великим математиком и философом А. Пуанкаре, а также идеи философов-рационалистов XVII века имеют прямое отношение к трудностям, вставшим на пути современных специалистов, работающих над моделированием интеллекта. Мысль Пуанкаре, несомненно, базировалась на представлениях, сформировавшихся задолго до него в истории философии. Эти идеи о соотношении логики и интуитивного в познании волновали многих философов XVII века. Декарт, Лейбниц, Спиноза, философы и одновременно ученые, которые внесли существенный вклад в развитие математики и логики, неоднократно обращались к вопросу о возникновении и обосновании математических истин. В первую очередь интерес для философов XVII века представляли именно истины, лежащие в

¹ Там же. С. 409.

² Там же. С. 413.

³ Там же. С. 411.

основе математики как науки, роль в процессе их постижения играл процесс интеллектуальной интуиции.

При рассмотрении попыток автоматизированного доказательства, таких, как программы LT и GPS, все более необходимым становится обращение к неявным структурам человеческого мышления, выявлению таких характеристик интеллектуальной деятельности, которые не были рассмотрены логикой. Структура, о которой здесь идет речь, касается процедур выбора последующего шага, ограничения перебора возможных вариантов. С одной стороны, мы фиксируем понимание интеллекта как процесса доказательства, вывода корректных заключений из истинных посылок по некоторым формальным правилам, правилам, которые сформулированы логикой. С другой стороны, заметно расширение такого понимания за счет добавления к решению логической задачи установки на решение вычислительной задачи. При реализации первого типа задачи машина пользуется жестким путем решения, который заложен в неё программистом. В таком случае продолжение процесса всегда однозначно определено, и даже если имеется разветвление, то машина имеет информацию, которая позволяет ей сделать дальнейший шаг. При решении задач второго типа выбор в узле не является однозначно определенным. Задачи такого типа могут быть решены машиной путем полного или частичного перебора вариантов. Как отмечает Поспелов Д.А., главная особенность естественного интеллекта в том, что «в отличие от машины человек проводит целесообразный перебор, резко сокращает число рассматриваемых вариантов за счет априорной оценки многих из них как неперспективных»¹. Именно эта идея и лежит в основе программ LT и GPS, создатели которых А. Ньюэлл и Г. Саймон являются признанными основателями такого направления, как эвристическое программирование. Постановка такой задачи и пути её решения в явном виде изоморфны сформулированным ранее теоретическим концептам относительно *наличия в доказательстве, в качестве необходимого условия, элементов, не входящих в строгую дедуктивную структуру; бессознательного характера мыслительных процедур*².

В настоящем кратком выступлении нам удалось остановиться фактически лишь на одном варианте модели интеллектуальной функции, репрезентантом которой оказались программы GPS и LP. Заметим, что, на наш взгляд, все на настоящий момент представленные в компьютерных науках модели искусственного интеллекта предполагают заимствование ряда характеристик из теоретических представлений о функциях естественного интеллекта и включают моделирование двух различных типов: моделирование механизма организации интеллектуальной функции и моделирование результата действия интеллектуальной функции. Представляется, что в настоящем докладе был проиллюстрирован вариант моделирования результата действия интеллектуальной функции, что в целом характерно для логической парадигмы в области искусственного интеллекта³.

¹ Поспелов Д.А., Пушкин В.Н. Мышление и автоматы. М.: Изд-во «Советское радио», 1972. С. 21.

² Отметим, однако, что авторы GPS не разделяли уверенности рассмотренных в первой части исследования мыслителей относительно неформализуемого характера этих неявных процедур.

³ Более подробно вопрос о методологических основаниях исследований в области искусственного интеллекта рассматривается автором в диссертационном исследовании на соискание ученой степени кандидата философских наук на тему: «Отношение теоретических концепций и компьютерных моделей в исследованиях искусственного интеллекта». М., 2008 г. (работы автора до 2009 года выходили под фамилией Трушкина).