ЯЗЫК В СИСТЕМЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: СИНТАКСИС И СЕМАНТИКА

В.А. Ладов

Каково соотношение человеческого сознания и системы искусственного интеллекта? Способен ли компьютер на осуществление мыслительных операций? Можем ли мы всерьез называть такие операции проявлением разумной деятельности? Как показывает полувековая традиция AI-философии (философии искусственного интеллекта), такие вопросы не являются праздными и вовсе не предполагают, как могло бы показаться с первого взгляда, однозначные отрицательные ответы. Дело в том, что для определения степени разумности системы ИИ необходимо сначала установить критерии осуществления разумных действий по отношению к самому человеческому сознанию. Необходимо определить, что мы будем понимать под существенными признаками разумности вообще, чтобы потом попытаться обнаружить эти признаки в системах ИИ. С того момента, как философия обратилась к обсуждению проблем, связанных с появлением искусственного интеллекта, было предложено немало теорий, каждая из которых по-своему определяла существенные признаки разумности. Отсюда и ответы об интеллектуальных способностях искусственных систем также оказывались различными – одни приписывали им признаки мыследеятельности и сознательного поведения, другие отрицали их.

Статья посвящена анализу одного из подходов к определению разумности. Исследователи, разрабатывающие этот подход, утверждают, что существенные признаки сознательного поведения можно выявить, исходя из рассмотрения того, каким образом человек использует свой естественный язык при осуществлении коммуникативных действий. В таком случае вышеуказанная общая проблема начинает обсуждаться на лингвистическом уровне. Искусственный интеллект является также коммуникативной системой. В данном случае используется особый язык-интерфейс, который обеспечивает интерактивную форму взаимодействия человека и компьютера, т. е. их коммуникацию. Возникает вопрос: Использует ли компьютер язык точно так же, как и человек? Если мы сможем ответить на этот вопрос положительно, то системе ИИ будет приписано свойство разумного поведения, если же ответ будет

отрицательным, значит, нам удастся выяснить то, чего не достает компьютеру, чтобы подняться до уровня человеческого сознания.

Одним из самых известных в истории AI-философии критических аргументов относительно разумности искусственных систем в рамках лингвистического подхода стал так называемый «аргумент китайской комнаты», выдвинутый американским философом Д. Серлом в начале 80-х годов [1]. Суть этого аргумента сводится к следующему. Допустим, человека, владеющего только английским, помещают в изолированную от внешнего мира комнату и предоставляют ему для чтения текст на китайском. Естественно, ввиду того, что он не имеет ни малейшего представления о значении китайских иероглифов, текст оказывается для него набором чернильных закорючек на листе бумаги – человек ничего не понимает. Затем ему дают еще один лист бумаги, исписанный по-китайски, и в придачу к этому определенную инструкцию на родном ему английском о том, как можно было бы сравнить два китайских текста. Эта инструкция научает выявлению тождественных символов и определению закономерности их вхождения в более общий контекст. Когда приносят третий китайский текст, к нему прилагают вторую английскую инструкцию о сравнении последнего с двумя предыдущими и т. д. В итоге после продолжительных упражнений испытуемому приносят чистый лист бумаги и просят что-нибудь написать по-китайски. К этому времени человек из китайской комнаты настолько хорошо освоил формальные символические закономерности, что, на удивление, действительно оказался способным написать вполне связный и понятный любому грамотному китайцу текст. Ну и, наконец, чтобы произвести должный эффект, человека выводят из комнаты на обозрение широкой публике и представляют как англичанина, изучившего китайский, что сам виновник презентации не замедлит подтвердить своим безукоризненным знанием иероглифического письма.

Так понимает ли наш испытуемый китайский? Серл дает категорически отрицательный ответ на этот вопрос. Понимание должно сопровождаться актами так называемой первичной интенциональности, в которых сознание, еще до всякого обращения к каким-либо материальным носителям, т. е. к речи или письму, способно концентрироваться на внутренних интенциональных (смысловых) содержаниях, как не редуцируемых ни к чему другому фактах автономной психической жизни. Интенциональность языка производна, она возникает при намеренном наделении изначально пустых знаков значением, посредством замещения внутреннего смыслового содержания пропозициональным содержанием синтаксически организованных структур.

Для общественности, которая оценивала результаты обучения человека из китайской комнаты, возникла иллюзия того, что экзаменуемый действительно овладел китайским. Причина этой иллюзии кроется в той привычке, в соответствии с которой люди предположили за пропозициональными содержаниями продуцированных человеком синтаксических форм его внутренние интенциональные содержания, явившиеся основой первых. Но на деле обучение в китайской комнате принесло прямо противоположные результаты. Человек научился формальным операциям со знаковой системой без какого-либо собственного «интенционального участия» в этом предприятии. Пропозициональные содержания представленного на обозрение китайского письма имели смысл только для тех, кто действительно мог подкрепить их более фундаментальными интенциональными содержаниями своей психики. Человек из китайской комнаты сам не понял ничего из того, что написал.

По мысли Серла, действия испытуемого англичанина полностью аналогичны работе AI. Искусственный интеллект, несмотря ни на какие интенсификации в сфере технологий, никогда не сможет достичь уровня человеческого сознания именно из-за невозможности преодолеть фундаментальный разрыв между первичной и производной интенциональностями. С помощью специальных программ, настраивающих на формальное оперирование символическими образованиями, AI может создавать иллюзии мощнейшей мыслительной активности, многократно превышающей способности человеческого сознания. Результаты такой деятельности AI оказываются, в самом деле, чрезвычайно полезными для человека. И тем не менее у нас нет никаких оснований тешить себя иллюзией существования «братьев по разуму». AI не мыслит. Всю работу по содержательному наполнению пустых символических структур берет на себя человек, «прикрепляя» последние к внутренним интенциональным содержаниям – подлинным элементам разумной жизни.

С позиции лингвистического подхода в AI-философии серлевский аргумент утверждает то, что язык искусственных систем не имеет семантики. Вся работа в системе интерфейса человек – компьютер со стороны машины происходит исключительно на синтаксическом уровне. Компьютер «обучен» определенным программам-алгоритмам связи символических элементов знаковой системы так, что возникает впечатление относительно их семантической нагруженности.

Возьмем в качестве примера работу географической электронной энциклопедии. Система ИИ запрограммирована так, чтобы, получив от человека запрос: «Как называется столица Непала?», выдавать ответ: «Катманду». При этом очевидно, что компьютер не понимает, что собственно стоит за теми знаками языка, которые использованы в данном

запросе, семантически они для него пусты. Просто в соответствующей программе дана директива: «при запросе, представляющем собой один синтаксический комплекс, выдавать в качестве ответа другой».

Машина может действовать как формальный логик. Отвлекаясь от какого-либо содержательного наполнения, она способна к оценке истинности сложных высказываний на основании анализа истинностных функций составляющих их простых атрибутивных суждений. Она способна оценить истинность вывода как в случае содержательно наполненного высказывания «Если на улице идет дождь, то асфальт мокрый», так и в случае высказывания, продуцированного на «тарабарском» языке: «Если жунсы губеют, то брунсы тернеют». Путем объединения логических электронных микросхем, принцип работы которых будет соответствовать истинностным функциям для логических союзов, в общую схему мы можем построить электрическую цепь, моделирующую весьма сложные дедуктивные рассуждения. Мы знаем, что в логике, посредством метода истинностных таблиц, можно оценить достоверность вывода из некоторого числа посылок. Возьмем для примера рассуждение: «В преступлении подозреваются трое: Иванов, Петров, Сидоров. Они дали следующие показания. Иванов утверждал, что если преступление совершил Сидоров, то Петров его не совершал. Петров настаивал на том, что если Иванов совершил преступление, то Сидоров тоже в этом участвовал; но если Иванов ни при чем, то это сделали либо он сам, либо Сидоров. Сидоров отрицал свою вину, но настаивал на виновности либо Петрова, либо Иванова. При условии, что все говорили правду, кто виновен?» Формализуем его так, чтобы оценить достоверность вывода о виновности каждого из подозреваемых:

$$((С \supset \neg П) \,\&\, ((И \supset С) \,\&\, ((\neg И \supset (П \vee С))) \,\&\, (\neg С \,\&\, (П \vee И))) \supset И$$
$$((С \supset \neg П) \,\&\, ((И \supset С) \,\&\, ((\neg И \supset (П \vee С))) \,\&\, (\neg С \,\&\, (П \vee И))) \supset П$$
$$((С \supset \neg П) \,\&\, ((И \supset С) \,\&\, ((\neg И \supset (П \vee С))) \,\&\, (\neg С \,\&\, (П \vee И))) \supset С$$

При построении соответствующих таблиц истинности выяснится, что мы можем с достоверностью утверждать виновность Петрова. К точно такому же выводу придет и система ИИ, причем данные логические вычисления она проделает значительно быстрее человека. Однако эта «дедуктивная ловкость» машины будет только подтверждать тезис Серла – в работе компьютера нас завораживает именно эта невероятная синтаксическая мощь, скорость оперирования знаками. Тем не менее какими бы головокружительными ни были операции, связанные с синтаксисом формальной знаковой системы, компьютер никогда не сможет самостоятельно задать им какую-либо семантическую интерпретацию.

На основании информации о морфологии языка машина может даже имитировать продуцирование связанного текста. Известно, что те или иные части речи подчиняются специфическим законам словообразования и потому данный процесс также можно формализовать. Например, читая фразу: «Глокая куздра бодланула бокра и бодрячит бокренка» мы вполне можем предположить здесь наличие частей речи – существительных, глаголов и прилагательных, – структура которых формально будет напоминать обычные, т. е. семантически нагруженные термины нашего естественного языка. Данные морфологические элементы, опять же в соответствии с определенными синтаксическими правилами, будут занимать соответствующие места в предложениях – места подлежащего, сказуемого, определений и т. д. Если машину запрограммировать на соответствие всем этим правилам, то ее успехи также могут быть очень впечатляющими. Однако в данном случае система ИИ будет себя вести в точном подобии с упоминаемым выше персонажем из китайской комнаты, который научился лишь филигранному оперированию синтаксическими элементами знаковой системы по определенным правилам без какой-либо семантической интерпретации.

Итак, на основании серлевского аргумента китайской комнаты мы можем утверждать, что с точки зрения лингвистического подхода в AI-философии существенным признаком разумной деятельности будет считаться способность к семантической интерпретации знаковой системы. И именно этим признаком не обладают системы искусственного интеллекта.

Аргумент китайской комнаты вызвал бурные обсуждения в рамках традиции AI-философии. Здесь нашлись как его приверженцы (см., например, [2]), так и оппоненты (см., например, [3]). Одни утверждали, что системе ИИ действительно никак нельзя приписать способность к семантической интерпретации, другие настаивали на том, что в определенном смысле эту способность нельзя приписать даже и человеку. При этом все молчаливо согласились с тезисами относительно синтаксиса.

Интересный момент в исследовании данной проблемы, на который мы хотели бы обратить внимание в данной статье, заключается в том, что спустя десятилетие тот же Д. Серл [4] весьма оригинальным образом пересмотрел свой собственный аргумент. На этот раз вопрос был поставлен о синтаксисе. А можем ли мы, в самом деле, утверждать – так как мы это делали ранее, – что машина способна на выполнение синтаксических процедур в рамках заданной знаковой системы? Теперь американский философ дал отрицательный ответ и на этот вопрос.

Для того чтобы прояснить смысл серлевской аргументации, вновь вспомним, для начала, англичанина, изучающего китайский. Критиче-

ский аргумент относительно семантики начинался с того, что человек не понимает значений написанных на бумаге символов. Осваивая формальные правила операций с данными символами, он овладевает определенной синтаксической техникой, которая создает иллюзию семантической осведомленности. Однако не пропустили ли мы здесь, при описании данного лингвистического действия, один важный момент, на который нам следовало бы обратить внимание еще до начала формулировки критического аргумента относительно семантики? Что именно может увидеть человек на предоставленных в его распоряжение листах бумаги? Строго говоря, на физическом уровне на листе бумаги виден лишь хаотический набор чернильных пятен различной формы. Получается, что прежде чем констатировать свою неосведомленность относительно семантики языка, человек из китайской комнаты уже должен задать определенную синтаксическую интерпретацию! Он должен понять вот эти чернильные пятна на листе бумаги именно как знаки, которые, возможно, объединены какой-либо системой правил функционирования, составляя при этом единое целое – язык. Он должен понять, что вот эти чернильные пятна в принципе могут что-то обозначать. При рассмотрении того, как тот или иной субъект – неважно, человек или машина – овладевает и пользуется языком, синтаксис не должен возникать по принципу Deus ex machina. На физическом уровне в среде материальных носителей языковых структур нет никакого синтаксиса. Для того чтобы тот или иной материальный объект оказался знаком, ему следует задать не только семантическую интерпретацию, которая покажет, что, собственно, этот знак обозначает, но и, прежде всего, интерпретацию синтаксическую, которая покажет, что данный материальный объект в принципе может что-то обозначать, т. е. является знаком.

Известно, что в основание информатики была положена гениальная в своей простоте идея, которую продуцировали американские математики и техники, в частности Клод Шеннон, – объединение логики и электричества. К тому времени – 30-м годам XX в. – в логике уже прочно зарекомендовал себя новый подход, основанный на слиянии формально-логической символики и языка математики. Сначала на основе алгебры Дж. Буля, которая представляла собой формализацию арифметических действий, было предложено воспользоваться алгебраическим языком и для формализации логического процесса рассуждения. Были введены символы для всех возможных логических констант, характеризующих формальные элементы в суждениях, – логических союзов. Отрицание «–», дизъюнкция «V», конъюнкция «&», импликация «⊃», тождество «≡». Затем было предложено заменить логические референты «истина» и «ложь» арифметическими символами 1 и 0. Далее Г. Фреге и

другие философы аналитической традиции построили систему референций для каждого логического союза – это были так называемые таблицы истинности. Таким образом, появилась возможность оценивать истинность какого-либо сложного высказывания на основе анализа составляющих его простых высказываний и их истинностных функций. Все это, в свою очередь, привело к возможности формального контроля над системой дискурсивного рассуждения вообще, т.е. к оценке логической последовательности и необходимости выводов из посылок какой угодно степени сложности.

Новизна идеи информатики заключалась лишь в том, что было предложено интерпретировать наличие или отсутствие напряжения в электрической цепи как знаки арифметических символов 1 и 0 соответственно. Так и возник цифровой компьютер. Теперь все логические наработки по анализу рассуждений можно было бы смоделировать на электрическом уровне путем создания соответствующих элементов цепи – сначала электрических ламп, затем транзисторов, и, в конце концов, электронных микросхем, закодированных на выполнение-имитацию истинностной функции какого-либо логического союза.

Еще раз обратим внимание на ключевые элементы данного процесса интерпретации. Сначала цифры 1 и 0 были поняты как знаки логических референтов «истина» и «ложь». Арифметика в данном случае оказывалась синтаксисом для логической семантики. Но составляет ли этот синтаксис «онтологию машины»? То есть действительно ли hardware компьютера – это нагромождение нулей и единиц? Если бы это было так, то идея информатики не представляла бы какой-либо ценности. В том-то и дело, что физика не имеет синтаксиса вообще. Наличие напряжения в электрической цепи – это еще не единица. Это лишь наличие напряжения в электрической цепи. Представим себе по ходу ситуацию, что мы вкручиваем лампочку в патрон настольной лампы с плохим качеством контакта проводников тока. Лампочка то загорается, то гаснет – будем ли мы считать данные события физического уровня передачей, скажем, какого-либо зашифрованного кода? В данном случае, конечно же, нет. В этом как раз и состоит первичный интерпретативный шаг – понять определенный уровень электрического напряжения как знак. И пока неважно знак чего: логического референта или, скажем, предупреждения об опасности пожара. Физический уровень в качестве материального носителя языковых выражений прежде семантической должен вначале получить синтаксическую интерпретацию, которая задается внешним образом, через пользователя данной знаковой системой.

Если при использовании электронной энциклопедии я задаю вопрос о столице Непала и получаю надлежащий ответ, то машина не только не

понимает значений символов, с которыми она оперирует в соответствии с определенным алгоритмом, она не представляет собой даже и формальную синтаксическую систему. На физическом уровне вслед за одним случаем высокого уровня напряжения в определенном участке цепи возникает другой случай – только и всего. Для того чтобы эти факты высокого напряжения понять как знаки, которые могут подчиняться определенным операциональным правилам их сочетания, необходимо задать первичную синтаксическую интерпретацию, на которую в дальнейшем и будет опираться программист при формулировке соответствующего алгоритма операций.

Интересно, что в связи с появлением данного аргумента в отношении синтаксиса системы ИИ в когнитивной науке с новой силой вспыхнули дискуссии вокруг так называемой проблемы гомункулуса. Параллельно развитию информатики и AI-философии в когнитивной науке, основанной на современных достижениях нейрофизиологии, очень активно стали проявлять себя исследования, основанные на интерпретации мозга как цифрового компьютера. Известно, что на физическом уровне движение нейронов представляет собой, в определенном смысле, движение электрических зарядов, а нейронные цепи можно уподобить цепям электрическим. Если понимать компьютер как скопление информации, закодированной в цифровом виде, то точно так же можно было бы отнестись и к работе головного мозга. В таком представлении мозг оказывался подобным машине, где “hardware” представляет собой физическое наличие нейронных связей, на которые накладываются синтаксическая и семантическая интерпретации. Однако если даже синтаксис не свойствен физике, а привносится на материальный уровень внешним образом, то здесь возникает проблема. Кто задает подобные синтаксические интерпретации? С компьютером, который мы покупаем в магазине, все проще – здесь интерпретацию задает пользователь. Но как быть с головным мозгом? Получается, что кроме того, кто является физическим носителем нейронных связей, должен существовать еще и тот, кто интерпретирует эти нейронные соединения как синтаксические элементы системы. Так возникает проблема гомункулуса – своеобразного «разума в разуме», того, кто интерпретирует физику. По сути же, с точки зрения материалистической онтологии, в мозге, как и в компьютере, нет никаких нулей и единиц.

В целом проблема гомункулуса заслуживает большего внимания и может стать предметом исследования отдельной работы. Мы не будем далее развивать здесь эту тему. В соответствии с намеченной целью нам следовало разобраться в вопросе различий в использовании языка системой ИИ и человеком, с тем, чтобы зафиксировать существенные при-

знаки разумности. Выводы, к которым мы приходим на основании изложенной выше суммы аргументов, являются следующими.

1. Существенными признаками разумности с точки зрения лингвистического подхода в AI-философии следует считать способность человеческого сознания к заданию как семантической, так и, прежде всего, синтаксической интерпретаций каким-либо материальным образованиям мира природы так, чтобы эти образования получали статус знаковой системы.

2. Система искусственного интеллекта вышеизложенными способностями не обладает.

ЛИТЕРАТУРА

1. *Searle J.* Minds, Brains, and Programs // The Philosophy of Artificial Intelligence, in M. Boden, ed. New York: Oxford University Press, 1990.

2. *Hauser L.* Why Isn't My Pocket Calculator a Thinking Thing? // Minds and Machines. 1993. Vol. 3, №. 1. February.

3. *Dennett D.* Evolution, Error and Intentionaly // Sourcebook on the Foundations of Artificial Intelligence, in Y. Wilks and D. Partridge, eds. New Mexico University Press, 1988.

4. *Searle J.* Is the Brain a Digital Computer? // Proceedings and Addresses of the American Philosophical Association. 1991. 64.