



АНАСТАСИЯ ПОПОВА,
научный сотрудник Института проблем лазерных
и информационных технологий
Российской академии наук (ИПЛИТ РАН)
anastasia.popova@fsmedia.ru

РОСКОШЬ ЧЕЛОВЕЧЕСКОГО МЫШЛЕНИЯ

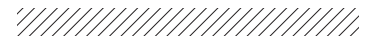
Искусственный интеллект перестает быть развлечением компьютерщиков и стремительно вторгается в повседневность: голосовое самообслуживание никого уже не удивляет, отзывы о продукции давно уже собирают и классифицируют программы, а в контакт-центрах многих организаций с клиентом в первую очередь общается робот-консультант. Попробуем разобраться, возможен ли конструктивный диалог с машиной и полноправное партнерство «искусственного» и «естественного» разума.

*Самой заветной мечтой
Самаона Гая было построить умный
корабль. Нет, не робот. Кораблей-
роботов, которые сами выбирают
курс, добираются до нужной планеты,
разгружаются и загружаются, немало
во Вселенной. Гай хотел сделать
корабль, который будет думать.*

К. Булычев, «Гай-До»

ИНТЕЛЛЕКТ? ЧТО ЭТО?

Искусственный интеллект, интеллектуальные системы, информационные технологии — эти термины настолько прочно вошли в лексикон масс-медиа, что на глазах становятся «общим местом». Между тем «искусственный интеллект», строго говоря, это раздел информатики, посвященный моделированию и воссозданию



разумных рассуждений и действий, а также (что уже ближе к обычному пониманию) свойство систем решать творческие задачи.

Однако, говоря о возможностях и угрозах искусственного интеллекта, мы имеем в виду не просто алгоритмы, которые выполняет программа (проверка правописания или перевод текстов), а нечто большее. Действительно, если программа может за секунду обработать колоссальный объем данных в соответствии с запросом, почему она не может, так сказать, «в свободное время» взяться за какую-нибудь «халтуру»? В прошлом веке ученые, впервые столкнувшись с фантастической скоростью машинного вычисления, задались вопросом: где границы возможности компьютера? Небезосновательно полагая вершиной интеллектуальных способностей человеческий разум, Алан Тьюринг (Alan Mathison Turing) предложил знаменитый тест для определения момента, когда машина сравняется в плане разумности с человеком. Предполагалось, что если компьютер сможет пройти тест Тьюринга, то мы можем уверенно заключить, что машина обладает интеллектом.

Для опровержения идеи «разумности» машины, прошедшей тест Тьюринга, Джон Серль (John Searle) поставил мысленный эксперимент, получивший название «Китайская комната» [2]. Его суть заключается в том, что при наличии входных данных на неизвестном языке (в случае Серля — китайском) и программы, составляющей ответы, человек может отвечать адекватно на этом языке, не понимая при этом самого языка. Что и моделирует программу, прошедшую тест Тьюринга, но не обладающую разумом. Этот мысленный эксперимент также породил большую волну критики и других мысленных экспериментов (в результате некоторых, путем хитро поставленных вопросов, экспериментатор, в конце концов, заставляет Серля выучить китайский). Но это опять же не доказывает неперменное наличие разума у программы, прошедшей злополучный тест Тьюринга! Упираясь в проблему понимания, некоторые ученые вообще отвергли понятие «семантики», сводя процесс мышления к оперированию синтаксисом, что, собственно, и реализуется в современных компьютерах.

СИЛЬНЫЙ ИЛИ СЛАБЫЙ?

Возвращаясь к обиходному значению слов «интеллектуальный» и «искусственный интеллект», стоит отметить, что для разделения разумности машинного интеллекта и интеллектуальности машинного разума было найдено изящное решение. Одним из философов искусственного интеллекта (тем самым Серлем, не понимающим китайского) было предложено понятие «сильного искусственного интеллекта», который, в случае существования, будет не просто моделью разума, а в буквальном смысле будет являться самим разумом — в том смысле, в котором человеческий разум — это разум. В противоположность понятию «сильного» возникло понятие «слабого искусственного интеллекта», он-то и решает прикладные задачи: становится достойным соперником в шахматной партии, подбирает рифму к словам, прокладывает маршруты для автомобиля. И пускай перед ним не стоит высокая цель самоосознания (или хотя бы прохождения набившего оскомину теста Тьюринга), зато никто не сомневается в его наличии. Особенность «слабого» интеллекта состоит в том, что он решает локальные задачи, и как решает! Надо только правильно их сформулировать. Интеллектуальные системы способны не только помочь в быту [4] и на дороге [5], но и выполнять такие непростые функции, как этический контроль (блокирование публикаций нетрезвых пользователей в Интернете) и подбор научных публикаций (выявление наиболее подходящих). Также большое будущее видится за системами поддержки принятия решений, всевозможными экспертными системами и автоматизированными системами управления. Пусть сегодня мы не можем полностью передать автоматике управление, допустим, автомобилем (учитывая непредсказуемость и зачастую алогичность ситуации на дороге), все-таки локальная задача по парковке или поддержанию движения в плотном трафике вполне ей по силам.

В общем, если не заманивать на мыслительные способности, а говорить лишь о компьютерной эмуляции интеллектуальных процессов, то кажется, что возможности «слабого» искусственного интеллекта поистине безграничны. В настоящее время, чтобы не пускаться в экзистенциальные споры, именно это



понятие и подразумевают под словами «искусственный интеллект».

ТЫ МЕНЯ... ПОНИМАЕШЬ?

Так почему же, несмотря на все успехи интеллектуальных программ, год за годом все новые поколения исследователей стремятся создать именно «сильный» интеллект? Что это, попытка осмыслить природу разума и сознания или честолюбивая мечта сравниться с Создателем,

кем бы тот ни был? А может, все-таки существуют задачи, которые под силу решить только настоящему искусственному разуму? Вопрос остается открытым, но бесспорный факт заинтересованности общественности в «сильном» искусственном интеллекте подтверждается объявленным в начале года конкурсом проектов, непосредственно связанных с этой темой (<http://futureoflife.org/grants/large/initial>).

Существует несколько вариантов теста Тьюринга. Современная версия представляет собой следующее задание. Группа экспертов общается с неизвестным существом. Они не видят своего собеседника и могут общаться с ним только через какую-то изолирующую систему — например клавиатуру. Им разрешается задавать собеседнику любые вопросы, вести разговор на любые темы. Если в конце эксперимента они не смогут сказать, общались ли они с человеком или с машиной, и если на самом деле они разговаривали с машиной, можно считать, что эта машина прошла тест Тьюринга. Постановка задачи в этом тесте во многом отражает неопределенность самого термина «разумность» и «интеллект». Неудивительно, что он вызвал такую волну критики. Как альтернатива тесту Тьюринга была разработана Схема винограда (Winograd schema) [1], включающая набор вопросов, которые кажутся простыми для разумного человека, но очень сложны для машин. Они представляют трудность, так как требуют здравого смысла, который все еще не доступен машинам, а также в них встречаются понятия и факты, которые не часто упоминаются на веб-страницах и потому не попадают в базы данных.

Мы привыкли воспринимать технические объекты как отдельные вещи, которые покупаются и выбрасываются, ломаются и подвергаются ремонту, входят в моду и морально устаревают. Мы не видим, что отдельно взятый технический объект есть лишь некая окаменелость техногенеза — непрекращающегося процесса эволюции технической жизни. Познать технический объект — значит увидеть его в странной, нечеловеческой оптике: как мгновенный срез этой жизни, заключающий в себе всю свою предшествующую историю.

Михаил Куртов, кандидат философских наук

И, что интересно, в это же время был открыт прием заявок на гранты, посвященные исследованию безопасности сильного интеллекта. Также о неугасающем интересе к проблеме сильного искусственного интеллекта свидетельствует ежегодно проводимая международная конференция, полностью посвященная этой проблематике и объединяющая сообщество энтузиастов этой области (<http://agi-conf.org/2015/>).

Может быть, ответ на вопрос, зачем человеку нужен искусственный разум, кроется в простых словах: «счастье — это когда тебя понимают». Но в таком случае, что же такое понимание? Сидя в своей «китайской комнате», Серль говорил, что понимание сути вопроса при известном алгоритме ответа невозможно потому, что любая

программа имеет синтаксический характер, а человеческое понимание семанлично. Значит, мы не сможем увидеть понимание в последовательностях символов, составляющих программу, но, может быть, его можно обнаружить как-то иначе? В книге *Data Mining: Practical Machine Learning Tools and Techniques* [3] авторы предлагают проверять понимание испытанным веками способом — экзаменом. Как и студентам, компьютеру дают некоторые конкретные задачи: перевести текст, выделить факты, дать конкретное значение многозначного слова. Получается, вовсе не обязательно существование «понимания» в каком-то трансцендентном значении, это может быть просто определенное состояние слушателя, в соответствии с которым он «правильно» решает задачи. Значит, обучаясь языку, человек учится употреблять слова правильно в данном контексте, и чем, в таком случае, этот процесс отличается от грамотно написанного алгоритма?..

В СЕТЯХ РАЗУМА

Исследования в данной области ведутся по разным направлениям. Одно из мнений, связанных с созданием искусственного разума, заключается в том, что невозможно выделить одни функции разума (решение не формализуемых творческих задач, поддержание конструктивного диалога) от других (таких как самосознание, эмоциональные реакции на события), как невозможно создать корову, состоящую только из мяса и молока. В этом вопросе никак не отделить мух от котлет, потому что

разум — это не мухи и не котлеты, а цельная структура, до сих пор до конца не понятая. А если вместилище разума — мозг, то логично сделать следующее предположение: все, что должно обладать разумом, должно воспроизводить структуру мозга. Таким образом, мы вплотную подходим к сложной и интересной теме искусственных нейронных сетей.

В общем случае искусственная нейронная сеть представляет собой совокупность простых процессоров с управляемым взаимодействием. На вход этой системы подается сигнал, он пробегает по сети от нейрона к нейрону (зачастую по очень сложной траектории), изменяясь, усиливаясь и ослабевая, и на выходе формируется правильно обработанный сигнал: например, из рукописного текста формируется текст печатный. Но для того чтобы выходной сигнал был правильным, сеть необходимо настроить. К этому процессу обычно применяют термин «обучение», его суть состоит в нахождении коэффициентов связей между нейронами (или в выборе наиболее удачной конфигурации этих связей). Обученная нейронная сеть выявляет сложные зависимости между входными данными и выходными, выполняет обобщение, распознавание, систематизацию, аппроксимацию. Но, несмотря на эти впечатляющие возможности, в настоящее время нейронные сети используются в решении весьма ограниченного круга задач. Это связано не только со сложностью обучения (чем больше сеть, тем сложнее ее обучить, и для разных топологий нейронных сетей подходят одни и противоположны другие алгоритмы обучения), но и с тем, что даже правильно обученная сеть не всегда будет выдавать правильное решение.

Однако возможность создания искусственного интеллекта на основе нейронных сетей все-таки существует. Может быть, правда, для этого понадобится сделать их сложными, многоуровневыми и с изменяющейся топологией. И может быть, для создания такой сложной системы понадобится интеллект, существенно превышающий по возможностям человеческий.

МЫ В ОТВЕТЕ ЗА ТЕХ, КОГО ВОПЛОТИЛИ

Еще одна интересная идея заключается в том, что для создания полно-



ценного искусственного разума необходимо наличие таких, казалось бы, человеческих черт, как эмоциональность, чувственность, духовность. То есть речь идет не просто об интеллекте, но уже о настоящей личности. Конечно, в первую очередь встает вопрос целесообразности: зачем эмулировать духовную жизнь и эмоции, если они ценны постольку, поскольку мы верим в их искренность? Ведь нельзя сказать, что компьютерная программа что-либо чувствует, так же, как нельзя описать чувства дифференцируемой функции. Однако этические проблемы создания искусственного разума остаются.

Массовое внедрение даже «слабого» искусственного интеллекта может привести к сокращению количества рабочих мест, потере права на анонимность и личную жизнь, снижению уровня ответственности.

И, конечно, самые сильные опасения вызывает предположение о том, что успех в создании сколько-нибудь сильного искусственного интеллекта может стать началом конца человеческой расы. В разговоре о развитии информационных технологий крайне сложно обойти эту тему стороной, поэтому практически каждому известному лицу в мире компьютерных технологий так или иначе пришлось выразить свою позицию. Например, Билл Джой (William Nelson Joy), один из основателей исследовательского отдела Sun Microsystems, в своем эссе «Почему мы не нужны будущему» сравнивает разработку искусственного интеллекта с созданием атомной бомбы. А глава компаний SpaceX и Tesla Motors Илон Маск (Elon Reeve Musk) считает, что «Прогресс в сфере развития искусственного интеллекта невероятен, и если вы не являетесь членом специальной исследовательской группы, вы даже не представляете, насколько быстро все происходит. И эти слова — не паника человека, не посвященного в проблему. Ведущие компании предпринимают серьезные шаги для обеспечения безопасности, но можно лишь надеяться, что им удастся не допустить выхода этого сверхразума во Всемирную сеть». Билл Гейтс (William Henry Gates III) также озабочен развитием искусственного интеллекта, видя опасность не только в возможности возникновения сверхразума, но и в том, что даже на современном уровне развития технологий человек может

лишиться работы, которая придает смысл его жизни. Схожую позицию имеет и известный футуролог Эрик Шмидт (Eric Emerson Schmidt). Он считает, что нужно беспокоиться не о появлении глобальной угрозы, а о том, чтобы улучшить систему образования и системы стимулирования в глобальном масштабе, чтобы подготовить людей к этому новому миру.

Орен Этциони (Oren Etzioni), директор Института искусственного интеллекта им. Аллена, призывает не опасаться мифической угрозы чужеродного разума: «Беспокоиться по поводу того, что искусственный интеллект начнет делать то, что он хочет, — все равно, что беспокоиться о том, что калькулятор начнет делать свои собственные вычисления. Искусственный интеллект, как и калькулятор, нуждается в том, чтобы человек управлял им и ставил ему задачи». Профессор психологии и неврологии Нью-Йоркского университета Гэри Маркус (Gary Marcus) также утверждает, что не стоит беспокоиться по поводу появления порабощающего сверхразума, по крайней мере в ближайшее время. Но при этом подчеркивает, что опасность заключается в том, что искусственному интеллекту не обязательно быть «сильным», чтобы натворить бед, огромный вред может принести самый обычный автомат из-за не исправленных вовремя ошибок. Весьма тревожно высказался в этом отношении великий физик Стивен Хокинг (Stephen William Hawking): «Успешное создание искусственного интеллекта станет самым большим событием в истории человечества, и, если мы не научимся избегать рисков, оно также может оказаться последним».

Появится ли достаточно мощный интеллект и как скоро это произойдет, доподлинно неизвестно, в любом случае, хотелось бы, чтобы человечество было к этому готово. Подводя итог, хочется привести слова знаменитого футуролога Рэя Курцвейла (Raymond Kurzweil): «Технологии всегда были палкой о двух концах, с тех пор как огонь начал обогревать наши дома и сжигать другие. Но искусственный интеллект — это не палка о двух концах, которая находится в руках одного-двух людей; ее держат миллиард или два миллиарда рук. В конечном счете, самое важное, что мы должны сделать, — это обезопасить работу искусственного интеллекта на уровне

Эволюция способности к моделированию, очевидно, привела в конечном итоге к субъективному осознанию. Нет оснований полагать, что электронные компьютеры действуют осознанно, когда они что-нибудь моделируют, хотя нам приходится допускать, что в будущем они, возможно, будут сознавать свои действия. Быть может, осознание возникает тогда, когда модель мира, создаваемая мозгом, достигает такой полноты, что ему приходится включать в нее модель самого себя.

Ричард Докинз (Richard Dawkins), доктор наук, этолог, эволюционный биолог

человеческих, гуманных и социальных институтов. Мы уже человеко-машинная цивилизация. Лучший путь избежать конфликта в будущем — это продолжить совершенствование наших социальных идеалов».

Возможно, стремление наделять вычислительный прибор разумом — всего лишь отражение потребности понять его, воплотившейся ранее в аниматических верованиях наших предков. Никто не знает, как далеко до конца, но тем интереснее. Предлагаю держать глаза открытыми и смотреть на прекрасное полотно под названием «будущее», которое разворачивается пред нами прямо сейчас. ●

Иллюстрации автора

Желающим продолжить размышление на тему будущего человечества в условиях развития не только естественного, но и искусственного интеллекта предлагаем ознакомиться с философией трансгуманизма, а специалистов в этой области приглашаем высказать свое мнение на сайте журнала в виде комментариев к статье или написать автору.

ЛИТЕРАТУРА

1. Levesque H. J., Davis E., Morgenstern L. The Winograd schema challenge // KR. 2012.
2. Searle, John. R. Minds, brains, and programs // Behavioral and Brain Sciences. 1980.
3. Ian H. Witten, Eibe F. Data Mining: Practical Machine Learning Tools and Techniques // MKS. 2005.
4. А. Попова. Мой ласковый и нежный дом // Control Engineering Россия. 2014. № 5.
5. А. Попова. Крылья, колеса, мозги // Control Engineering Россия. 2014. № 6.