

Что понимают люди, но не понимает искусственный интеллект?

Казак А. В., Струтинская Н. В.

Белорусский национальный технический университет, Минск, Беларусь

Искусственный интеллект может многое, и порой это даже пугает. Однако компьютеру не хватает того, что есть практически у всех людей. Это здравый смысл, т.е. способность делать обыденные выводы, используя базовые знания о мире. Например, что спички + дрова = огонь.

Здравый смысл был назван черной материей искусственного интеллекта. Это одновременно нечто базовое и удручающе неуловимое, потому что здравый смысл состоит из неявной информации, широкого набора неписанных допущений и эмпирических правил, которые люди автоматически используют, чтобы понять происходящее. Чтобы отслеживать прогресс искусственного интеллекта, люди придумали специальные тесты, основанные на каверзных для компьютеров вопросах. Самый современный тест подобного рода имеет название «Схема Винограда». Он был предложен на замену тесту Тьюринга, который компьютеры вполне успешно прошли в 2014 году.

Схема Винограда – тест, предназначенный для испытания предельных интеллектуальных возможностей компьютера. Если машина сможет определить подразумеваемый смысловой субъект, основываясь только на подсказках из контекста, значит, машина использует рассуждения на основе здравого смысла, а следовательно, может считаться разумной.

Людям очень легко даются вопросы с подразумеваемым ими субъектом. Например: трофей не поместился в коричневый чемодан, потому что он слишком большой. Что слишком большое: трофей или чемодан? Если нужно выяснить, к чему относится данный вопрос, достаточно иметь представление, что чемодан обычно содержит вещи, а трофей обычно вещей не содержит. Также нужно знать, как работают контейнеры и что большие вещи не поместятся внутри маленьких вещей. Для людей это кажется простым, потому что

мы не только понимаем, как работает синтаксис, является ли предложение корректным, но и потому, что у нас есть опыт работы с чемоданами, трофеями и самой реальностью. Ответ на данный вопрос требует понимания.

Еще один пример: «Я расстелил скатерть на столешнице, чтобы защитить ее». Защитить что? Очевидно, у нас есть представление, что такое скатерть и столешница, и можно ли расстелить скатерть на столешнице. Но почему так сложно объяснить это компьютеру? Неявная природа большинства знаний, основанных на здравом смысле, затрудняет их явное и однозначное представление, которое так необходимо компьютерам. То, что мы узнаем в два или четыре года, на самом деле, никогда не записывается в книгу.

Тем не менее, ранние исследователи полагали, что решить проблему накопления знаний искусственным интеллектом возможно, и предложили записать все факты о мире. Они были убеждены, что их всего пара миллионов. Но собрать достаточное количество даже очевидной информации сложнее, чем кажется. Амбициозный проект под названием *Cyc Knowledge Base Visualization* начался в 1984 году со скромной цели закодировать миллионы кусочков знаний, олицетворяющих человеческий здравый смысл. И этому не было конца. Более трех десятилетий спустя база знаний *Cyc*, закодированная в сложной, специально разработанной логической форме содержит миллионы категорий и концепций и более 25 миллионов точных утверждений.

Однако эта база оказала незначительное влияние на развитие искусственного интеллекта. Последующее наполнение базы знаний и даже создание новой с помощью машинного обучения не позволило решить ту же самую проблему – проблему здравого смысла, потому что кроме очевидных фактов существует бесконечное множество неочевидных. Для каждого правила есть исключение. Пусть я слышу утверждение: «На улице идет дождь». Можно заключить, что, если я выйду на улицу, я стану мокрым. Но только, если я не буду чем-то укрыт. И это самое простое исключение.

Такие базы знаний могут содержать десятки утверждений о том, что обычно происходит, если человек бросает спичку в стопку дров. Но что

насчет потенциально бесконечного списка нечастых или необычных вещей, которые могут повлиять на сценарий? Например, спичку потушил ветер или дрова оказались сырыми? Такая проблема называется проблемой хрупкости. Если было бы возможно создать базу знаний в сто или тысячу раз больше, чем предыдущая, система все равно начала бы страдать от интеллектуальной недостаточности.

Но появились искусственные нейронные сети. Это система искусственного интеллекта, предназначенная для имитации взаимосвязанных слоев нейронов, как это происходит в биологическом мозге. Их преимущество заключается в том, что они умеют изучать структуры, не требуя от программистов их предварительного описания.

Искусственные нейронные сети умеют управлять транспортными средствами, играть в стратегические настольные и компьютерные игры, идентифицировать лица, распознавать объекты, жесты, речь, рукописный и печатный текст, визуализировать данные, создавать 3d реконструкции объектов, улучшать, восстанавливать и оживлять изображения. Такие системы уже пишут музыку, картины и даже сценарии для фильмов. Но при всей своей гибкости и очевидной интеллектуальной мощи нейронные сети имеют много слабых сторон в областях, где требуется обращение к здравому смыслу.

Принципы работы компьютерных нейронных сетей имеют фундаментальные ограничения. Люди всегда будут получать от них не точные ответы, а наиболее вероятный вариант. Ответ будет всегда приблизительный.

Нейронные сети – это черный ящик. Вероятно, это даже самая известная их проблема технического характера. Неизвестно, как и почему нейронная сеть сделала то или иное предположение. При вводе изображения кошки в нейронную сеть она может предположить, что это автомобиль. Но очень трудно понять, что заставило сделать ее неправильное заключение.

Человеку гораздо проще понять причину своей ошибки, поэтому в областях вроде медицины, банкинга, военных и космических отраслях, словом, везде, где причины выводов и ошибок критически важны, нейронные сети

используются очень ограниченно. Существует еще немало технических проблем, связанных со стоимостью создания и обучения нейронных сетей, но важнее будет сказать об ограничении, лежащем в основе технологии. Идея, стоящая за созданием искусственных нейронных сетей, подразумевает, что можно смоделировать процессы, происходящие в мозге человека. Парадокс в том, что информации в данной области мало.

Нейронные сети – это лишь попытка приблизиться к симуляции мозга, но ни в коем случае не готовое решение. Чтобы ребенок научился распознавать машины, ему достаточно показать одну из них. А в основе обучения нейронных сетей лежит изучение как можно большего количества данных. Чтобы нейронная сеть распознала автомобиль, необходимо предоставить ей изображения нескольких тысяч автомобилей.

Если компьютер выучит, что произойдет, если толкнуть стакан со стола, он не будет знать, что произойдет, если стакан не сталкивать. Нейронная сеть, обученная распознавать только канареек, бесполезна для распознавания пения птиц или человеческой речи.

Как бы активно ученые не устраняли все эти проблемы, нейронные сети навсегда останутся именно лестницей, чтобы долететь до Луны. До тех пор, пока мы создадим ракету, пройдет не один технологический прорыв. Поэтому все нынешние страхи насчет угрозы искусственного интеллекта не подкреплены реальными фактами. Возможно, некоторое беспокойство нам не повредит, но паниковать точно рано. И тут возникает моральная дилемма: все это путь в светлое и беззаботное будущее или в рабское существование под гнетом машин?

Феномен случайности как объект философского осмысления

Круголь С., Дроздович О. М.

Белорусский национальный технический университет, Минск, Беларусь

Существует ли случайность? Можно ли ее предсказать? Феномен случайности является одним из вопросов, над которым размышляли философы