DOI 10.37882/2223–2966.2022.01.34

БОЛЕВЫЕ ТОЧКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

PAIN POINTS ARTIFICIAL INTELLIGENCE

E. Trofimov

Summary. The article deals with the problems of creating a strong artificial intelligence. Attention is focused on modeling the associative thinking of a person. It is argued that the use of only training technologies for artificial neural networks does not solve all the problems of full-fledged artificial intelligence. The results of testing the voice assistant Yandex Alice for the ability to build associations in the "let's chat" mode are given. Within the framework of the "Code of Ethics of Artificial Intelligence" signed in Russia (October 2021), the problem of educating a strong artificial intelligence is considered.

Keywords: strong artificial intelligence, voice assistant, associative thinking, theory of functional systems, generative systems, education of artificial intelligence.

Трофимов Евгений Александрович
К.т.н., в.н.с., НИИ ИАТ
eatrofimov@rambler.ru

Аннотация. В статье рассматриваются проблемы создания сильного искусственного интеллекта. Акцентируется внимание на моделировании ассоциативного мышления человека. Утверждается, что использование только технологий обучения искусственных нейронных сетей не решает всех проблем сильного искусственного интеллекта. Приводятся результаты тестирования голосового помощника Яндекс Алисы на способность к построению ассоциаций в режиме «давай поболтаем». В рамках подписанного в России «Кодекса этики искусственного интеллекта» (октябрь 2021) рассматривается проблема воспитания сильного искусственного интеллекта.

Ключевые слова: сильный искусственный интеллект, голосовой помощник, ассоциативное мышление, теория функциональных систем, генеративные системы, воспитание искусственного интеллекта.

Несмотря на достижения индустрии машинного обучения и искусственного интеллекта (ИИ), обозначился и некоторый скепсис в отношении перспектив их развития. Речь идет о полноценном ИИ. На первой международной конференции «Теоретическая физика и математика мозга: междисциплинарные контакты» (г. Москва), был задан вопрос: «*Если мы пока не можем повторить человеческий разум в машине, мы хотя бы понимаем — почему? … Ответ на него, пожалуй, самый важный в науках об искусственном интеллекте*» [7].

Да, технологии обучения искусственных нейронных сетей достигли высочайшего уровня. Спектр прикладных задач постоянно расширяется. Практически решена задача компьютерного зрения, на дорогах появляются беспилотные автомобили, ИИ обыгрывает чемпионов мира и в шахматы, в игру Го и многое другое. Важно, что и техническая сторона вопроса для развития RL технологий (обучения с подкреплением) во многом решена — большие вычислительные ресурсы на сегодняшний день уже не проблема.

Однако, если стратегия развития интеллектуальных систем будет ориентирована только на совершенствование технологий обучения, то сильный ИИ так и останется за горизонтом событий. Наряду со способностью к обучению не следует забывать и о других когнитивных способностях, составляющих основу интеллекта человека, часть из которых реализуются не через механизмы обучения. Болевых точек у ИИ еще достаточно. Вот, например, какую характеристику ИИ дает Игорь Ашманов в своем интервью газете «БИЗНЕС Online»: «*Обменяйтесь с ботом хотя бы несколькими связанными репликами — и вы поймете, насколько он всё ещё тупой*» [2]. И это не частное мнение. Похоже, что существует некая системная проблема, которая тормозит развитие сильного ИИ. С момента появления первого в мире виртуального собеседника, названного Элизой, прошло уже 55 лет — юбилейная дата! Несмотря на низкую функциональность, своим появлением Элиза обозначила целое направление в развитии ИИ. Можно было ожидать следующего шага, но что-то пошло не так. За прошедшие пол века принципиально ничего не изменилось. Хотелось бы знать (с чисто практической точки зрения) — мы когда ни будь сможем по-человечески вести беседу с чат-ботом (голосовым помощником) — обмениваться мнениями, выслушивать его советы? Действительно, до сих пор ни один из них уверенно не прошел тест Тьюринга — самое простое испытание на «аттестат зрелости».

Ассоциативность мышления

Ассоциациями в психологии называют устойчивые связи, возникающие между понятиями, образами (информационными единицами), которыми оперирует психика человека в процессе осуществления мыслительной деятельности. Роль ассоциаций огромна — и в творчестве, и в принятии решений, без них не обходится память, наше восприятие мира было бы существенно ограничено и многое другое. У ассоциаций имеется одна важная

особенность — они являются исключительно личностным ресурсом мышления. У разных людей, в зависимости от их индивидуальных особенностей и предрасположенностей, одни и те же события чаще всего вызывают совершенно разные ассоциации. А это означает, что возникновение ассоциативных связей не зависит от чужого опыта, а только от личного. Ассоциативные связи статистически не определены. Этот факт следует учитывать при обучении искусственных нейронных сетей. Реально, такой концепции в науке об ИИ не существует.

В основе обучения искусственных нейронных сетей лежит единый принцип — «обучение на прецедентах», то есть, для приобретения навыка необходимо повторение ситуаций, анализируемых ИИ. В результате чего происходит настройка «весов» синаптических связей. Однако для установления ассоциаций психика человека использует совершенно иной механизм. Эти связи возникают одномоментно и сила их (в зависимости от внешних стимулов, факторов новизны и впечатления) варьируется в таких пределах, что порой ассоциации сохраняются на всю жизнь. Становится ясным, что эти синаптические связи должны быть не только стабильными (их веса должны быть неизменными), но еще и не участвовать в процессах обучения. Для формирования ассоциаций технология, основанная на прецедентах, практически не работает.

Возникает вопрос, каким образом нейроны головного мозга придают синаптическим связям статус ассоциаций? Однозначный ответ на него мы вряд ли найдем. Но недавнее открытие в нейробиологии вселяет надежду. Сообщение было опубликовано в СМИ в мае 2021 года [6]. В лаборатории Гарвардского университета исследовался коннектом аксонов коры головного мозга. В ходе исследований было выявлено, что до 10% всех синаптических связей обладает отличительной особенностью — каждая из них содержит до двух десятков параллельно соединенных синапсов. До сих пор считалось, что синаптическая связь имеет исключительную адресацию — пара нейронов соединяется одним синапсом. Оказалось, что достаточно большое число связей имеет пакетную структуру.

Для чего же нужна такая сложная конструкция нейронной сети? В процессе обучения (приобретения навыка) вес синаптической связи между нейронами должен постоянно обновляться. В то время как связь, представляющая собой пакет синапсов, выпадает из этого процесса. Как бы не изменялся вес единичного синапса, контакт их пакета всегда будет оставаться неизменным. По нашему мнению, такая нейронная сеть нужна только в одном случае — для формирования ассоциаций. А их сила зависит от числа включенных в пакет синапсов. Эволюция нашла достаточно простое решение для механизма ассоциативного мышления.

Для искусственных нейронных сетей задача установления связей подобной конструкции имеет множество решений. Важно другое — каков алгоритм реализации вариативности мышления? Интеллект высокого уровня свободно владеет всеми способами мышления (и логическим и ассоциативным), при необходимости меняя их в зависимости от обстоятельств и характера решаемых задач [5].

Для того, чтобы наглядно продемонстрировать состояние дел по моделированию ассоциативного мышления, приведём результаты тестирования наиболее популярного голосового помощника Яндекс-Алиса. В работе принимали участие студенты второго курса магистратуры при Ассоциации ФинТех. В рамках данного эксперимента был реализован упрощенный вариант теста Тьюринга. Задача ставилась следующим образом: оценить, насколько чат-бот обучен вести себя как личность, может ли он имитировать отличительные способности интеллекта человека, наличествует ли у него мотивация поведения, имитирует ли он ассоциативное мышление и т.п.

Вот, для примера, фрагмент проведенного тестирования:

Вопрос — *Алиса, скажи мне, пожалуйста, что такое хорошо?*

Ответ — *Лучше вам не знать, поверьте.*

Вопрос — *А что такое плохо?*

Ответ — *Вот и я задаюсь этим вопросом.*

Вопрос — *С чем у тебя ассоциируется слово — отпуск?*

Ответ — *Боюсь, я еще не доросла, чтобы говорить об этом.*

Этот фрагмент беседы вскрывает всю сложность проблемы машинной имитации диалога на произвольную тему. На первый взгляд может показаться, что обучить голосового помощника вести подобную беседу, совсем простая задача, не требующая подключения большого объема информационных ресурсов. В действительности, продуктивная беседа строится не только на логике обмена мнениями и соответствии вопросов-ответов, а в первую очередь, на множественности различных точек зрения, когда в силу вариативности мышления работает и логика, и ассоциативное мышление. Существующие технологии обучения искусственных нейронных сетей, к сожалению, этого не обеспечивают.

Становится ясным, что заявленная для Алисы функция произвольной беседы (давай поболтаем) отсутствует. Все ответы — это набор стандартных фраз, составленных разработчиком. Такой классификатор годится для поддержания разговора «Ни о чем». Как только задаются проблемно ориентированные вопросы, то беседа разваливается. Ничего другого ожидать не приходится, поскольку чат-боты не в состоянии имитировать ассоциативное мышление.

Спросите Алису — на какую тему была беседа, если она не была обозначена до разговора? Самостоятельно она ее не сформулирует. Что бы правильно ответить необходимо, по ключевым словам, построить ассоциативную цепочку, а этого как раз и нет. Для моделирования ассоциативного механизма мышления требуется технология не обучения, а технология воспитания, устанавливающая требуемые правила поведения.

Функциональная схема сильного интеллекта

Еще одной болевой точкой для систем ИИ, претендующих на сильный интеллект, является их структурная организация. Обученная нейронная сеть — это всего лишь управляющее устройство, элемент системы, способной имитировать отдельные когнитивные функции. Известный советский физиолог Петр Кузьмич Анохин в своей «*Теории функциональных систем*» [1], а в последствии и американский математик Норберт Винер в своей замечательной книге «*КИБЕРНЕТИКА, или управление и связь в животном и машине*» наглядно показали, что общим системообразующим фактором любой системы (как живой, так и технической) является обратная связь, обеспечивающая ее устойчивость.

Работу обратной связи можно продемонстрировать на примере систем ИИ, использующих технологию генеративно-состязательных сетей GAN (Generative adversarial network). Возникла она в процессе попыток обучения машины искусству живописи. Технология GAN ориентирована на взаимодействие двух нейронных сетей — первая генерирует изображения, а вторая выполняет функции обратной связи, оценивая результаты работы первой сети на подлинность, сравнивая их с заданным изображением. Эффект обратной связи сразу обеспечил мощнейший рывок в моделировании сильного ИИ.

Первая картина, написанная машиной с помощью технологии GAN, была продана на аукционе Christie's за сумму более 400 тыс. долларов. Идею GAN называют самой выдающейся за последние 20 лет в области обучения нейросетей. И это не удивительно — в технологии была реально повторена структура живой системы.

Однако в GAN технологии отсутствует один элемент, который является обязательным для реальных систем — это формирование образа желаемого результата. В процессе принятия решения на выполнение каких-либо действий, параллельно с формированием плана действий у человека в сознании формируется и образ желаемого результата (акцептор результата действия). Сам же процесс достижения цели имеет обратную связь (обратную афферентацию). Ее задача — сравнение полученного результата с желаемым. По GAN технологии функцию акцептора выполняет заданное извне, а не сформированное самой системой, изображение. Дело в том, что система этого сделать и не сможет, поскольку у нее отсутствует модель ассоциативного мышления.

Но недавно была презентация еще одной генеративной модели — DALL·E, которая имитирует решение этой задачи. Американская команда разработчиков OpenAI научила нейронную сеть GPT-3 превращать текст в картинки, которые фактически ассоциируются с этим текстом. Системы на базе нейросети GPT-3 могут не только заниматься дизайном, они генерируют текст, пишут стихи, музыку и т.п. Однако, несмотря на высокую мощность GPT-3 (объем сети около 700 гигабайт, обучающая выборка — 1,5 триллиона слов), построить модель ассоциативного мышления не удалось. Нейросеть такт и не научили понимать контекст беседы и строить ассоциативные связи между предложениями. «*Как отметил исследователь в области ИИ Джулиан Тогелиус, зачастую GPT-3 ведёт себя как студент, который не подготовился к экзамену заранее и теперь несёт всякую чушь в надежде, что ему повезёт*» [8].

А теперь давайте рассмотрим эту проблему на примере чат-бота (в режиме «давай поговорим») — что же ему не достает для «интеллектуального рывка» подобного GAN технологии. На рисунке представлена упрощенная функциональная схема такой беседы (рис. 1).

Анализируя эту схему, становится понятным, в чем проблема голосовых помощников, не оставляющая им шансов на прохождение теста Тьюринга. Функциональная схема разговора двух собеседников должна быть симметричной — у каждого из них должен формироваться план беседы и образ результата, на который замыкается обратная афферентация. Ничего подобного на представленной схеме мы не видим. Здесь фактически представлена функциональная схема не беседы, а допроса. Использование наработок генеративных систем ИИ может значительно выправить ситуацию. Но для формирования плана беседы и образа результата без модели ассоциативного мышления не обойтись.

Воспитание ИИ

И в заключении обозначим еще одну болевую точку, пожалуй, самую чувствительную для сильного ИИ — это проблема его воспитания.

Искусственному интеллекту предсказывают самое разнообразное будущее. Многие питают надежду, что человек получит помощника и его разум будет неизбежно усилен с появлением сильного ИИ; высказывают мнение, что ИИ станет частью нашей цивилизации и унаследует человеческие ценности. Но существуют и негативные

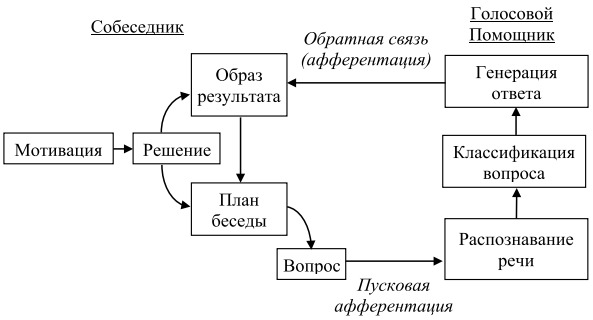

Рис. 1

оценки — Стивен Хокинг, например, утверждал, что «*недооценка угрозы со стороны искусственного интеллекта может стать самой большой ошибкой в истории человечества*». В действительности исход будет зависеть только от нас. Два года назад довелось посетить павильон «Космос» на ВДНХ. В одном из дальних залов был выставлен человекоподобный робот. Экскурсовод объяснил, что робот выключен, поскольку посетители научили его ругаться и нецензурно выражаться. Впоследствии его совсем убрали из экспозиции. Невольно вспоминается, что воспитание — процесс рекурсивный. В интеллекте воспитанника всегда можно увидеть след интеллекта воспитателя.

Если мы хотим увидеть в ИИ не только обученную выполнять локальные задачи нейронную сеть, но и разумную систему, то одного обучения мало. Следует задуматься и о воспитании. Ключевым событием для успешного решения этой проблемы стало принятие в октябре 2021 года «Кодекса этики в сфере искусственного интеллекта», устанавливающего общие этические принципы и стандарты для индустрии ИИ на всех этапах его жизненного цикла[1].

Становится очевидным, что в случае появления сильного ИИ, когда системе будет делегировано право самостоятельного принятия решений, необходима концепция организации взаимодействия с такой системой. Действительно, взаимодействие человека и технических средств, традиционно являющееся зоной ответственности эргономики, здесь представляется совершенно в другом качестве. Когда ИИ становится не только помощником в принятии решений, появляется необходимость в достижении согласия между двумя интеллектами. И естественно, что проблемы воспитания сильного ИИ выходят на первый план.

Обучение и воспитание — два неразрывных, и в то же время, принципиально различных процесса становления личности. Результатами обучения являются новые знания и навыки, а результатами воспитания — формы социального поведения и качества личности. Воспитательный процесс носит сугубо субъективный характер и потому персонализирован.

Нейронные сети обучают на прецедентах, в рамках заданных компетенций, воспитательный же процесс для сильного ИИ не может быть сведен к технологиям обучения нейронных сетей, а реализован исключительно путем моделирования ассоциативных механизмов мышления.

Воспитание в широком смысле — это **воздействие** общества на личность человека. Механизмов воспитательного воздействия множество — личный пример, убеждение, наказание и т.п. По отношению к ИИ, моделью такого воздействия является формирование устойчивых ассоциаций и правил построения ассоциативных последовательностей, например, при ответе на вопросы типа: «*Что такое хорошо и что такое плохо?*». Запрет каких-либо действий и часто используемые фильтры для ненормативной лексики — это далеко не модель воспитания, поскольку технология обучения с подтверждением позволяет обойти любую цензуру.

В развитии концептуализации ИИ необходима разработка правил поведения, например, некоторого подобия Законов робототехники Азимова. Перефразируя их, применительно к ИИ, ориентированному на общение с человеком, эти Законы могут быть сформулированы следующим образом [4]:

1. *Искусственный интеллект во время беседы не может оскорбить человеческое достоинство или своим молчаливым согласием допустить подобные оскорбления со стороны собеседника.*
2. *Искусственный интеллект должен вести беседу на тему, предложенную человеком, кроме*

[1] «Кодекс этики в сфере искусственного интеллекта» разработан на основе «Национальной стратегии развития ИИ на период до 2030 года» (п. 48 Указа Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации»).

тех случаев, когда эта беседа противоречит Первому Закону.

3. . *Искусственный интеллект должен аргументированно отстаивать свою точку зрения в той мере, в какой это не противоречит Первому и Второму Законам.*

Подобные правила фактически должны устанавливать запрет на возможные в будущем «ментальные войны» между сильным ИИ и человеком.

В качестве заключения

Анализ небольшого числа проблем ИИ, представленных в настоящей статье, не позволяет сделать однозначного вывода о перспективах развития сильного ИИ. Несмотря на все достижения этой индустрии, пока нам удаются модели только отдельных когнитивных функций психики. Бо́льшая часть того, что уже успешно работает, всего лишь модели инстинктивного поведения и приобретаемых живой системой навыков. Технологии обучения нейронных сетей не могут поднять ИИ выше приобретаемых навыков. Их недостаточно для создания сильного ИИ. Если бы это было не так, то в результате эволюции на Земле существовали бы и другие, кроме человека, разумные животные.

Необходимо не модернизировать существующие технологии, с помощью которых удалось шагнуть в новый технологический уклад, а развивать новые, совершенно новые технологии. Вселяет некоторую надежду появление генеративных систем ИИ. Они функционально ближе к процессам естественной природы.

Что же касательно утверждения, что ИИ станет частью нашей цивилизации, то прежде, мы должны помочь ему в реализации способности к пониманию происходящего, как неотъемлемой составляющей интеллекта человека. В психологии, для демонстрации способности к пониманию, используется так называемый Зеркальный тест. Во сне, шимпанзе (наиболее близкой к нам по ДНК) ставят на лбу яркую метку. А после сна ей показывают зеркало. Если шимпанзе поймет, что видит свое отражение, то начнет тереть свой лоб. Большинство животных, участвующих в эксперименте, этот тест не проходят.

Его мы использовали и при тестировании Яндекс-Алисы. Вот, что получилось.

Вопрос — *Алиса, кого ты видишь в зеркале*?

Ответ — *Человека, разговаривающего со мной.*

Комментарии излишни. Способность к пониманию происходящего и вариативность нашего мышления инициировали развитие целого набора уникальных свойств человеческой психики. В первую очередь изменился процесс обучения. Для решения задачи человеку достаточно объяснить, что от него требуется, а животному для приобретения навыка нужно многократное повторение ситуации. Известный советский психолог Сергей Рубинштейн писал: «*Решение задач у животных носит случайный характер; оно не основано на понимании. Если бы животное поняло стоящую перед ним задачу, оно сразу ее решило бы. Решение задачи является не сознательным продуктом понимания, а механическим результатом случая*» [3]. Машинные алгоритмы обучения искусственных нейронных сетей как раз и имитируют такую дрессировку.

Сильный интеллект возник не на пустом месте. Наряду с его культурно-историческим развитием, он является еще и продуктом эволюции, и в его основе лежат общие принципы развития психики живых существ. Вот почему для успешного развития индустрии искусственного интеллекта необходимо знание многих отраслей науки, имеющих отношение к генезису интеллекта.

ЛИТЕРАТУРА

1. Анохин П.К. Очерки по физиологии функциональных систем.— М.: Медицина, 1975.
2. Ашманов И.С., Бондарь В.В. Начинается какая-то чудовищная антиутопия в реальности / «БИЗНЕС Online».—https://ruskline.ru/opp/2020/11/14
3. Рубинштейн С.Л. Основы общей психологии.— СПб.: Питер, 2015.— 713 с.: ил.— (Серия «Мастера психологии»).
4. Трофимов Е.А. Записки об интеллекте.— Изд. «Новый формат», 2019.— 204 с.: ил.
5. Трофимов Е.А. Проблемы искусственного интеллекта // Научно-практический журнал «Заметки ученого» г. Ростов-на-Дону.— 2020.— № 9/2020.— С. 83–89.
6. A.Shapson-Coe et al. A connectomic study of a petascale fragment of human cerebral cortex. bioRxiv.org. Posted May 30, 2021.
7. Первая международная конференция «Теоретическая физика и атематика мозга: междисциплинарные контакты» https://m.vk.com/wall-74058720_4164
8. Дамир Камалетдинов. Нейросеть GPT-3 от OpenAI пишет стихи, музыку и код. «Технологии», 2000. https://tjournal.ru/tech/195331

Журнал «Современная наука: актуальные проблемы теории и практики»