

Артамонов В.А.

д.т.н., профессор, академик Международной академии информационных технологий, Минск
artamonov@itzashita.ru

Артамонова Е.В.

к.т.н. (PhD), член-корреспондент Международной академии информационных технологий, Минск
admin@itzashita.ru

ЭВОЛЮЦИОННЫЕ ОЖИДАНИЯ В РАЗВИТИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА¹

Ключевые слова: искусственный интеллект, машинное обучение, узкий ИИ, ИИ общего назначения, суперсильный ИИ, чат-боты, ChatGPT, угрозы, уязвимости, математические и натурные модели механизма самоорганизованного обучения нейронов в сети мозга.

Введение

Как сказал выдающийся русский учёный Д.И. Менделеев: «Всякая наука начинается с измерений и классификации». Примером тому может служить предложенная в 1869 году великим химиком и метрологом периодическая система элементов.

И с этих позиций перейдём к вопросу развития искусственного интеллекта как когнитивной науки будущего, которая кардинальным образом изменит жизнь человечества. Следуя заветам великого учёного, начнём с классификации.

В научном мире сложилась практика классификации искусственного интеллекта, основанная на требованиях национального стандарта ГОСТ Р 59277-2020² — по категориям и по виду³. По категориям были выделены три основные категории искусственного интеллекта: **узкий (слабый) ИИ** (англ. — *Artificial Narrow Intelligence, ANI*), **общий ИИ** (англ. — *Artificial General Intelligence, AGI*) и **суперсильный ИИ** (англ. — *Artificial Super Intelligence, ASI*), которые отличаются по уровню интеллектуальных возможностей и целям использования.

По виду ИИ был разделен на **поверхностный** и **глубинный**, где поверхностный ИИ основывается на правилах и заранее определенных алгоритмах, а глубинный ИИ использует машинное обучение для поиска закономерностей в данных и создания собственных алгоритмов.

Эволюция когнитивных возможностей искусственного интеллекта

Искусственный узкий интеллект (ANI) — это тип искусственного интеллекта, в котором алгоритм обучения создается для решения единственной задачи, может быть и очень большой. Любые знания, полученные в ходе этой деятельности, не будут применяться в других видах деятельности. Системы ANI настроены на работу в рамках определенного набора инструкций, для решения заранее определенных задач в той или иной области. Они могут выполнять только те функции и инструкции, которым они заранее обучены⁴. ANI не обладает интеллектом, способным мыслить нестандартно, как это сделал бы человек.

Есть четыре основные задачи искусственного интеллекта:

- классификация;
- регрессия;
- ранжирование;
- кластеризация.

¹ Данная статья и наши исследования в области ИИ были поддержаны компанией ООО «СЕКВЕНТА» (www.pcplanet.ru), г. Москва. Авторы выражают огромную благодарность руководству PCPlanet.ru за финансовую и организационную помощь в проведении научных исследований и написании данной статьи.

² Национальный стандарт Российской Федерации ГОСТ Р 59277-2020. Системы искусственного интеллекта. Классификация систем искусственного интеллекта.

³ Классификация искусственного интеллекта. Область применения ИИ. 2023. — <https://dtf.ru/s/595875-neyro-iskusstvo/1677861-na-kraiyu-sveta>

⁴ Артамонов В.А., Артамонова Е.В., Сафонов А.Е. Искусственный интеллект: когнитивное начало // Защита информации. Инсайд. 2022. — № 4. — С. 50-59.

Задача классификации – предсказание категории объекта и разделение объектов согласно определенным и заданным заранее признакам. То есть ИИ сортирует данные по нужным категориям: одежду – по цветам, сезонам или ткани, книги – по жанрам, авторам, языкам написания, информационные ресурсы – по степени их защищенности, письма – по личной или рабочей переписке и т.д.

Задача регрессии – по заданному набору признаков спрогнозировать некоторую целевую переменную. Результатом решения задачи регрессии является предсказание места на числовой прямой. Например, загруженность дорог в зависимости от времени суток и время на путь из пункта А в пункт Б в зависимости от пробок. Или каким будет объем рынка определенных товаров через два года. И даже скорость развития определенной болезни при общих показателях здоровья человека. Поскольку регрессия запрограммирована на работу с числами, ее встраивают в различные вычислительные системы, даже в классический Excel.

Задача ранжирования – решение класса задач машинного обучения с учителем, заключающихся в автоматическом подборе ранжирующей модели по обучающей выборке, состоящей из множества списков и заданных частичных порядков на элементах внутри каждого списка. Задача ранжирования (*learning to rank*) отличается тем, что ответы надо получить сразу на множестве объектов, после чего отсортировать их по значениям ответов. Она может сводиться к задачам классификации или регрессии. Часто применяется в информационном поиске и анализе текстов.

Задача кластеризации – разбиение заданной выборки данных (объектов) так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров значительно отличались друг от друга. Кластеризацию применяют для анализа и поиска признаков, по которым можно объединить объекты, для сжатия данных и поиска новизны (что не входит ни в один кластер). В чем отличие классификации от кластеризации? При классификации есть набор предопределенных классов, вы учите машину на наборе примеров и потом хотите знать, к какому классу принадлежит новый объект. При кластеризации вы используете алгоритм, который пытается сгруппировать набор объектов и определить, существует ли какая-либо взаимосвязь между ними, то есть машина учится сама. На эффективность алгоритмов влияет множество факторов вроде размера и структуры набора данных. По этой причине приходится опробовать разные алгоритмы, проверяя эффективность каждого на тестовом наборе данных, и затем выбирать лучший вариант, соответствующий решаемой задаче.

Области применения ANI: роботизированные комплексы, беспилотные транспортные системы, системы контроля управления доступом, системы обеспечения информационной безопасности (ИБ), военные системы управления оружием, медицина, образование и даже программы для настольных игр.

Искусственный интеллект общего назначения (AGI) – это гипотетический интеллектуальный агент, который может понять или научиться любой интеллектуальной задаче, которую может решить человек или какие-либо животные. AGI также определяется как автономные системы, которые превосходят человеческие возможности при выполнении большинства экономически ценных работ. Целью системы AGI является выполнение любой задачи, на которую способен человек. AGI – это, по сути, гипотетический интеллект машин. Это концепция, согласно которой машины имитируют или подражают интеллекту человека или его поведению, обладая способностью учиться и применять этот подход/интеллект для решения любых разнообразных задач. Основное отличие AGI от узконаправленного искусственного интеллекта (ANI) заключается в том, что ANI способен разрешать только задачи, определенные и заранее запрограммированные для конкретных процессов, и не способен адаптироваться к новым условиям и проблемам, а AGI – учиться и применять свои способности в широком смысле.

AGI считается следующим этапом развития искусственного интеллекта после создания ANI. Когда будет создан полноценный AGI, он сможет не только выполнять конкретные задачи, но также учиться и самостоятельно адаптироваться в большем количестве областей, не ограничиваясь только заранее зафиксированными задачами.

Примеры полноценного AGI в мире пока еще не созданы. Однако такие компании, как Open AI¹ стремятся к этому. Согласно опросам, примерно половина экспертов в области искусственного интеллекта считают, что полноценный классический общий искусственный интеллект появится к 2060 году. Однако в Open AI считают, что первый AGI пусть пока ещё несовершенный и являющийся лишь точкой на континууме интеллекта уже создан в виде чат-бота ChatGPT. Они считают, что, скорее всего, прогресс будет продолжаться и дальше, возможно, сохраняя те же темпы развития, которые мы наблюдали в последнее десятилетие. Учитывая эпохальность создания новой технологии ИИ, остановимся на ней более подробно.

ChatGPT представляет собой инновационную технологию, которая может быть использована в различных приложениях для улучшения пользовательского применения и повышения эффективности работы. Благодаря использованию глубокого машинного обучения и технологии трансформеров, ChatGPT может генерировать точные и качественные ответы на вопросы и является важным инструментом для автоматизации различных задач.

Нейросеть ChatGPT – это большая языковая модель, обученная компанией Open AI, которая использует глубокое обучение для генерации текста и ответов на вопросы. Языковая модель – это алгоритм, который предсказывает следующее слово по предыдущим. Большие языковые модели демонстрируют так называемые *эмерджентные способности* – умеют решать задачи, с которыми они раньше не встречались.

Эта модель была создана на основе технологии трансформеров, которая позволяет обрабатывать большие объемы текста и понимать связи между словами и предложениями.

Основная цель ChatGPT – предоставление пользователю качественных и точных ответов на заданные вопросы. Для этого модель использует огромный объем текстовых данных, который был подвергнут тщательной предвари-

¹ Open AI – американская компания, занимающаяся разработкой и лицензированием технологий на основе машинного обучения. Одним из основателей является предприниматель Илон Маск.

тельной обработке. Это позволяет модели учитывать контекст вопроса и генерировать ответы, которые максимально соответствуют запросу пользователя.

Чат-бот может использоваться для решения различных задач, таких как поиск информации, ответы на вопросы, генерация текста и многое другое. Эта модель может быть использована как на больших, так и на малых масштабах, она может быть адаптирована к различным типам приложений. ChatGPT обеспечивает множество преимуществ для пользователей и разработчиков.

Главные достоинства чат-бота:

- Высокая точность ответов на вопросы. Благодаря использованию глубокого обучения ChatGPT может генерировать ответы, которые точно соответствуют запросу пользователя.
- Быстрота и эффективность работы. ChatGPT может быстро обрабатывать большие объемы данных и генерировать ответы в режиме реального времени.
- Автоматическое обучение. ChatGPT может автоматически обучаться на основе новых данных, что позволяет модели постоянно улучшаться и адаптироваться к новым условиям.

ChatGPT представляет собой инновационную технологию, которая может быть использована в различных приложениях для улучшения пользовательского опыта и повышения эффективности работы. Благодаря использованию глубокого обучения и технологии трансформеров¹ ChatGPT может генерировать точные и качественные ответы на вопросы и является важным инструментом для автоматизации различных задач.

Что особенного в ChatGPT? Прежде всего, эта большая языковая модель обучалась с использованием большого количества обратных связей от пользователей, чтобы выполнять широкий спектр запросов на решение поставленного круга научных (и не только) задач. Ниже приведены основные характеристики ChatGPT:

- Содержит 175 миллиардов параметров (объем порядка 800 Gb).
- Обучающих данных – порядка 45 Tb.
- Стоимость обучения нейросети – 4,5 млн долл.
- Эксплуатационные расходы – 3 млн долл. в месяц.

Существует множество теоретических моделей и подходов к разработке AGI, которые опираются на идеи, представления и решения, которые использует человеческий мозг. На одном из таких подходов мы остановимся более подробно.

Открытие, сделанное в японском Центре исследований мозга RIKEN (2021 г.), образно говоря, вписало новую статью в «конституцию биоматематики» – принцип «свободной энергии» Карла Фристана², говорящий о том, что всей жизнью на всех масштабах организации – от отдельных клеток до человеческого мозга с миллиардами нейронов – движет универсальный императив, который можно свести к математической функции. Суть открытия заключалась в имитации механизма самоорганизованного обучения нейронов в сети мозга при поступлении нового сенсорного ввода путем минимизации свободной энергии нейронов – т.е. минимизации неопределённости в предсказаниях своей модели³.

Основные идеи исследования:

- Мозг людей и животных постоянно самооптимизируется (самоорганизованно обучается), перестраивая структуру и силу нейронных связей для адаптации к меняющимся условиям в целях сохранения гомеостаза организма;
- В результате этой постоянной самооптимизации мозга появляются и развиваются разум и сознание живого существа. Весь его жизненный опыт (в виде сенсорных внешних и внутренних ощущений) перерабатывается в самооптимизирующуюся «перепрошивку» нейронных сетей;
- Но что за базовый механизм лежит в основе самооптимизирующейся «перепрошивки» нейронных сетей мозга – великая тайна для человечества. Раскрыв ее, можно будет понять общее устройство и принципы функционирования сознания биологических существ;
- Широко распространенным мнением в околонаучных кругах (в научных, об этом говорят осторожно) является предположение, что принцип, лежащий в основе жизни и разума, станет краеугольным камнем «Единой теории всего»⁴.

Спустя полтора года после успешной имитации механизма самоорганизованного обучения нейронов в соответствии с принципом свободной энергии та же группа исследователей Центра исследований мозга RIKEN (ведущий автор Такуя Исокура) сообщила о следующем этапе экспериментальной проверки теперь уже не на модели, а на реальных нейронах (*in vitro*). Используя новую технику обратной инженерии, авторы исследования смогли подтвердить количественные прогнозы принципа свободной энергии, используя *in vitro* сети нейронов коры головного мозга крыс, которые выполняли причинно-следственный вывод информации в ходе эксперимента (см. рис. 1).

Исследователи организовали этот процесс, используя сетку электродов под нейронной сетью, чтобы стимулировать нейроны крыс по определенной схеме, которая смешивала два отдельных скрытых источника воздействия.

¹ Трансформеры – относительно новый тип нейросетей, направленный на решение последовательностей с легкой обработкой дальнедействующих зависимостей. На сегодня это самая продвинутая технология в области обработки естественной речи (NLP).

² Карл Джон Фристон (англ. Karl John Friston; род. 12 июля 1959, Йорк) – британский нейробиолог и специалист по визуализации мозга, изобретатель статистической параметрической разметки (SPM) и воксельной морфометрии (VBM), член Британской Медицинской Академии (1999), член Королевского научного общества (2006), в 2016 г. занял место в списке самых влиятельных учёных нейробиологов.

³ Карелов С. Сверхважный прорыв к «Единой теории всего». 2023. – <https://dzen.ru/b/ZNuVDCJqLSXxFPBg>

⁴ Образцов К. Единая теория всего. 2016.

И после всего 100 тренировок нейроны автоматически стали избирательными: некоторые стали очень сильно реагировать на источник № 1 и очень слабо — на источник № 2, а другие наоборот. Прорывным результатом этого исследования стало доказательство *in vitro*, что принцип свободной энергии является принципом самоорганизации биологических нейронных сетей. И это может означать, что принцип свободной энергии — это и есть фундамент «Единой теории всего».

Краткий вывод: Создание AGI является сложным и долгосрочным процессом, и пока что не было создано ни одной системы, которая бы полностью соответствовала определению AGI. Однако за этим стоят огромные инвестиции в научные исследования и разработку AGI, так как это может быть ключевым прорывом для развития искусственного интеллекта в многих отраслях, начиная от национальной безопасности, реального сектора экономики, промышленного производства и заканчивая здравоохранением и научными исследованиями.

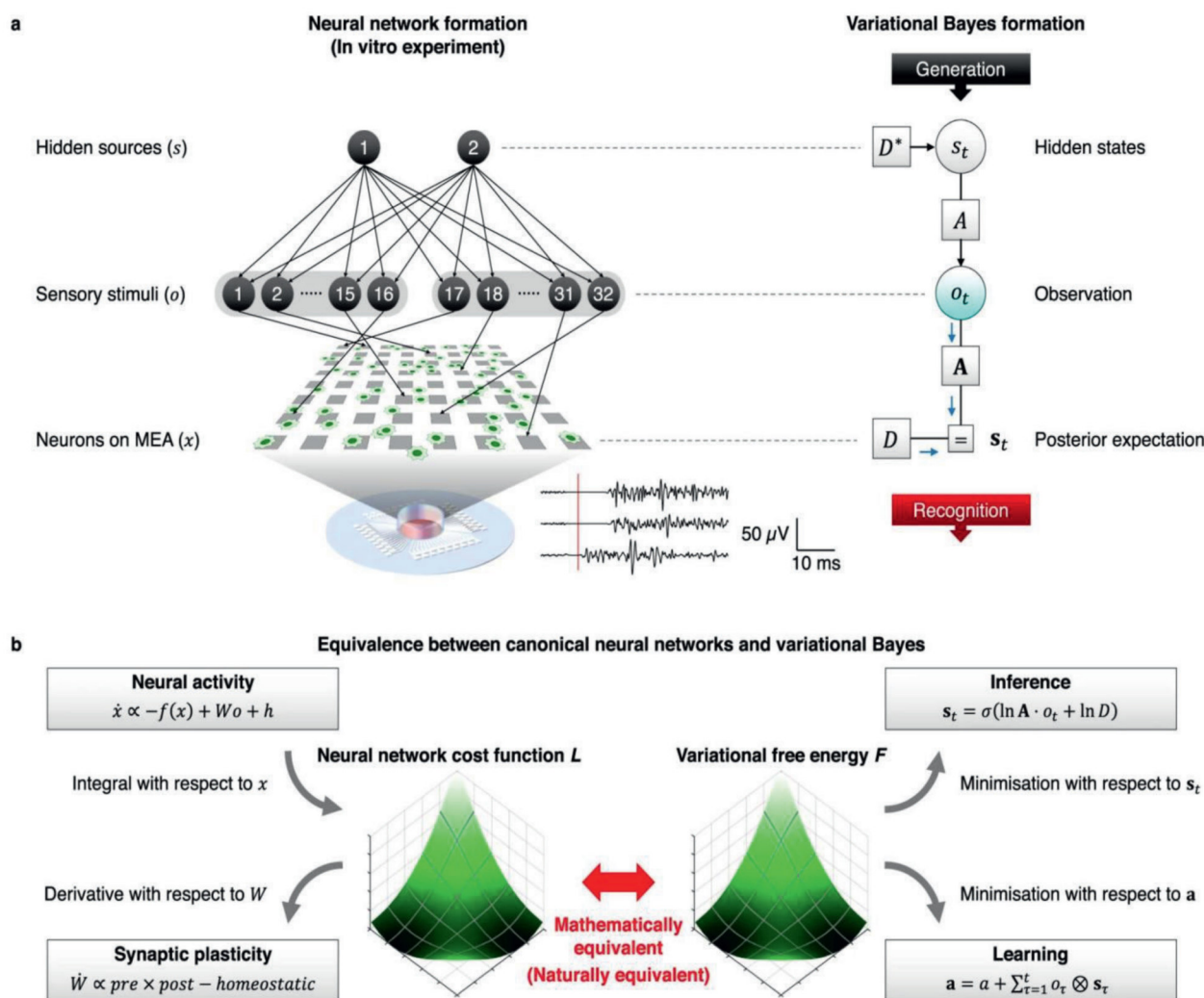


Рисунок 1.
Схема организации исследования RIKEN: а) Натуральный эксперимент In vitro;
б) Математическая модель

Суперсильный искусственный интеллект (ASI) определяется как форма ИИ, способная превзойти человеческий интеллект, проявляя когнитивные способности и развивая собственные навыки мышления. Это гипотетический ИИ, которого человечество еще не смогло достичь, и неизвестно, что произойдет, если мы его достигнем. ASI отличается от обычного искусственного интеллекта (AGI), который предполагает программную имитацию человеческого интеллекта, и способен выполнять такие задачи, которые даже сложно представить.

Вопрос развития ASI вызывает много опасений и споров, поскольку в случае, если такой искусственный интеллект получает неограниченные возможности, он может привести к непредсказуемым и катастрофическим последствиям. В результате экспоненциального роста производительности компьютеров, включая квантовые вычислители, наступит время «технологической сингулярности», когда вычислительная (логическая) мощность ИИ сравняется по интеллекту с человеческим разумом. Как поведёт себя этот искусственный разум в «постчеловеческом мире», невозможно предугадать.

Тенденции экспоненциального роста технологий следуют самые разные области — такие как искусственный интеллект и машинное обучение в качестве одной из ветвей развития ИКТ, робототехника, военное дело, здравоохранение, электро- и самоуправляемые автомобили, образование, 3D-печать, промышленность и сельское хозяйство.

Добро пожаловать в 4 промышленную революцию. Добро пожаловать в Экспоненциальный Век. Эта концепция была предложена Вернором Винжем, который предположил, что если мы сумеем избежать гибели цивилизации до этого самого века, то сингулярность произойдет из-за прогресса в области искусственного интеллекта, интеграции человека с ИКТ или других методов увеличения разума¹. Усиление разума, по мнению Винжа, в какой-то момент приведет к положительной обратной связи – более разумные системы могут создать еще более интеллектуальные системы и делать это быстрее, чем первоначальные их конструкторы – люди. Эта положительная обратная связь скорее всего окажется столь сильной, что в течение очень короткого промежутка времени (месяцы, дни или даже всего лишь часы) мир преобразится больше, чем мы сможем это представить, и внезапно окажется населен сверхразумными созданиями.

По мнениям некоторых учёных-футурологов и того же Винжа, придерживающихся концепции сингулярности, она должна наступить около 2030 года и даже по самому пессимистическому сценарию не позднее середины XXI века, т.е. в 2050 году. Сторонники теории технологической сингулярности считают, что если возникнет принципиально отличный от человеческого разум (*постчеловек*), дальнейшую судьбу цивилизации невозможно предсказать. С понятием сингулярности часто связывают идею о невозможности предсказать, что будет после нее. *Постчеловеческий* мир, который в результате появится, возможно, будет столь чуждым для нас, что сейчас мы не можем знать о нем абсолютно ничего. Единственным исключением могут быть фундаментальные законы природы, но даже и здесь иногда допускается существование еще не открытых законов (у нас пока нет теории квантовой гравитации) или не до конца понятых следствий из известных законов (путешествия через пространственные «дыры», рождение «вселенных-карликов», путешествия во времени и т.п.), с помощью которых *постлюди* смогут делать то, что мы привыкли считать физически невозможным.

Вывод: Вопрос предсказуемости важен, поскольку если нет возможности предсказать хотя бы некоторые последствия наших действий, нет никакого смысла в том, чтобы пытаться направить развитие эволюции ИИ в желательном направлении.

Угрозы ИИ и кризис человечества. Человечество стоит на пороге не только технологического, но и философского кризиса, считает историк Юваль Харари, автор книги «*Sapiens: Краткая история человечества*»². Новые технологии формируют новые формы антиутопии. И общество пока не понимает, как адаптироваться к меняющейся реальности.

Харари вывел формулу предстоящего глобального кризиса:

$$\mathbf{B} \times \mathbf{C} \times \mathbf{D} = \mathbf{HH}$$

В данном случае **B** – это познания в биологии, **C** – это вычислительная мощность, а **D** – это данные. Если помножить их друг на друга, появится возможность взламывать людей (**HH** – hack humans).

Под взломом исследователь подразумевает возможность управлять человеком на глубинном уровне, то есть контролировать его желания и стремления. Харари опасается, что власти и корпорации скоро изучат людей настолько, что смогут с легкостью регулировать их мысли.

Технологии отдаленно будут напоминать таргетированную рекламу, только их действие будет более точным, а эффект стопроцентным.

Ранее исследователь отмечал, что в сложившихся обстоятельствах привычные философские концепции отмирают. Это касается свободы воли и свободы выбора. Люди ошибочно полагают, что контролируют ситуацию, но на самом деле это не так.

Главное следствие масштабного внедрения искусственного интеллекта – *это утрата человеком автономии и авторитета*. При этом ИИ не обязательно выходить на один интеллектуальный уровень с людьми и обладать сознанием. Алгоритмам МО достаточно будет изучить личность досконально, чтобы найти самую слабую точку и запустить процесс манипуляций.

Общество подвержено взлому на всех уровнях, но больше всего Харари пугает биологический: «Эксперты по ИИ могут общаться с философами. С историками – да, пожалуйста. С литературными критиками – замечательно. Но меня пугает их общение с биологами», признал он в интервью изданию Wired. Тем не менее, исследователь подчеркивает, что ИИ обладает и массой преимуществ. Особенно это касается медицины. Харари подчеркивает, что никто не станет препятствовать внедрению технологии – ведь она способна принести столько пользы людям.

Распространение ИИ в комплексе с биотехнологическими открытиями породит два возможных сценария антиутопии.

Первый: надзор-капитализм – даст алгоритмам полную власть над людьми. Машинный интеллект и МО решат за нас, где жить, работать, с кем встречаться и за кого голосовать.

Второй: укрепление тоталитаризма и диктатуры, при котором каждый житель Земли — это объект непрерывной слежки. Особую роль в этом процессе сыграют биометрические и видеосистемы, которые не дадут гражданину скрыться от всевидящего ока государства.

Историк подчеркивает, люди могут даже не заметить, как оказались во власти ИИ и МО. Большинство не сможет понять, как работают механизмы алгоритмов и как именно нами манипулируют. Человечество привыкло к традиционным формам объяснения и повествования, но машинный интеллект работает со статистическими данными и оперирует другими понятиями.

¹ Vinge V. The Coming Technological Singularity. 1993. – <http://www-han.sdsu.edu/faculty/vinge/misc/singularity.html>; Новоселов А. Технологическая сингулярность как ближайшее будущее человечества. – <http://andrzej.virtualave.net/Articles/singularity.html>

² Харари Ю. Sapiens: Краткая история человечества. – <http://www.labirint.ru/books/498309>

Харари считает, что *чрезмерное усложнение систем* – одна из главных актуальных проблем. Из-за этого, например, ученым все сложнее объяснять свои теории и доносить до аудитории суть открытий.

Важный побочный эффект этого – расцвет теорий заговора. По этой причине сейчас возникает все больше антиглобалистов и тех, кто не верит в глобальное потепление. То же касается и сферы финансов – с каждым годом она усложняется, и некоторые концепции можно объяснить, только если потратить 10 лет на изучение экономики и математики. «В этом тоже выражается философский кризис», отмечает Харари.

Он также считает, что сегодня человек борется не с отдельными людьми, а с государствами и корпорациями. Перед лицом таких мощных соперников шансов на успех мало. Более того, влияние некоторых стран выходит за географические рамки. Историк обвиняет развитые государства и крупные корпорации в цифровом колониализме.

Вывод: По мнению некоторых учёных-футурологов, после наступления технологической сингулярности, человечество ожидает технологический и философский кризис. Социум погрузится в эпоху цифрового колониализма.

Нарисованная выше учёными-футурологами довольно пессимистическая картина мира, после достижения человеческой цивилизацией временной точки технологической сингулярности, скрашивается последними исследованиями учёных-математиков: *возможности ИИ оказались небеспредельными*¹. Подобно человеческому разуму ИИ ограничен парадоксами теории множеств.

До сих пор считалось, что самой фундаментальной проблемой в развитии технологий ИИ является *необъяснимость* принимаемых им решений.

В январе 2019 к этой проблеме добавилась еще одна фундаментальная проблема – это принципиальная *непредсказуемость*, какие задачи ИИ может решить, а какие нет.

На пути триумфального развития технологий машинного обучения, как казалось, способных при наличии большого объема данных превзойти людей в чем угодно: в играх, распознавании, предсказаниях и т.д. встала первая из 23 проблем, поставленных в докладе Давида Гильберта на международном математическом конгрессе в Париже еще в 1900 году².

Первой в списке этих 23 проблем, решение которых до сих пор считается высшим достижением для математики, была так называемая *гипотеза континуума* (*континуум-гипотеза* или *1 проблема Гильберта*), которую выдвинул и пытался решить (но потерпел неудачу) еще сам создатель теории множеств Георг Кантор.

И вот сейчас, на исходе второго десятилетия XXI века, гипотеза континуума, будучи примененная к задачам машинного обучения, стала холодным отрезвляющим душем для всех техно-оптимистов ИИ.

Машинное обучение оказалось не всесильно. Даже в широком спектре сценариев обучаемость ИИ не может быть ни доказана, ни опровергнута.

Первая же научная сенсация 2019 года оказалась совершенно неожиданной. Опубликованная 7 января того же года в Nature Machine Intelligence статья «Обучаемость может быть неразрешимой»³ устанавливает предел возможностей машинного обучения – ключевого метода вычислений, на котором стоит весь современный ИИ.

Этот научный вывод столь важен, что журнал Nature сопроводил статью еще двумя популярно её разъясняющими статьями «Недоказуемость приходит в машинное обучение» и «Машинное обучение приводит математиков к неразрешимой задаче».

Суть всех этих статей в следующем. Обнаружены сценарии, в которых невозможно доказать, может ли алгоритм машинного обучения решить конкретную проблему. Этот вывод может иметь огромное значение как для существующих, так и для будущих алгоритмов обучения.

Обучаемость ИИ не может быть ни доказана, ни опровергнута с использованием стандартных аксиом математики, поскольку это связано с парадоксами, открытыми австрийским математиком Гёделем в 1930-х годах⁴.

Парадоксы – это формально-логические противоречия, которые возникают в теории множеств и формальной логике при сохранении логической правильности рассуждения. Парадоксы возникают тогда, когда два взаимоисключающих (противоречащих) суждения оказываются в равной мере доказуемыми.

С точки зрения математики, вопрос «обучаемости» сводится к тому, сможет ли алгоритм извлечь шаблон из ограниченных данных. Ответ на этот вопрос связан с парадоксом, известным как вышеупомянутая континуум-гипотеза (проблема континуума или 1 проблема Гильберта) и разрешенным в 1963 г. американским математиком Полом Кёэном [там же].

Решение оказалось весьма неожиданным: то, что утверждается в гипотезе континуума, нельзя ни доказать, ни опровергнуть, исходя из аксиом теории множеств. Гипотеза континуума логически независима от этих аксиом. Неспециалисту довольно трудно понять, почему утверждения такого рода играют для математики столь большую роль и ставятся на первое место в списке важнейших проблем. Отметим лишь, что на самом деле речь идет о вещах принципиальных и фундаментальных, так как континуум – это, по сути, базовая математическая модель окружающей нас физической, пространственно-временной реальности (частью которой являемся и мы сами), а в математике континуум

¹ Ben-David S., Hrubeš P., Moran S., Shpilka A., Yehudayoff A. Learnability can be undecidable // Nature Machine Intelligence. 2019. – N 1. – P. 44-48.

² Демидов С. С. «Математические проблемы» Гильберта и математика XX века // Историко-математические исследования. – М., 2001. – № 6 (41). – С. 84-99.

³ Ben-David S., Hrubeš P., Moran S., Shpilka A., Yehudayoff A. Learnability can be undecidable // Nature Machine Intelligence. 2019. – N 1. – P. 44-48.

⁴ Демидов С. С. «Математические проблемы» Гильберта и математика XX века // Историко-математические исследования. – М., 2001. – № 6 (41). – С. 84-99.

еще и синоним совокупности всех действительных чисел, также центрального понятия математики и ее рабочего инструмента.

По сути, Гёдель и Коэн доказали, что континуум-гипотеза не может быть доказана ни как истинная, ни как ложная, начиная со стандартных аксиом, утверждений, принятых как истинные для теории множеств, которые обычно принимаются за основу всей математики. Иными словами, утверждение не может быть ни истинным, ни ложным в рамках стандартного математического языка.

Вывод: Математически доказано, что возможности ИИ не беспредельны. И какими бы огромными вычислительными ресурсами не обладал человек, машинное обучение никогда не приведёт к победе искусственного разума над человеческим.

В пользу данного доказательства говорят и последние исследования нейробиологов в области исследования структуры и возможностей человеческого мозга, о которых мы рассказывали выше.

Здоровый человеческий мозг содержит около 200 млрд нервных клеток, которые соединяются друг с другом сотнями триллионов синапсов. От каждой нервной клетки могут отходить десятки тысяч синапсов. В одной только коре больших полушарий человека находится около 125 трлн синапсов – в 1500 раз больше, чем звёзд в нашей галактике. По результатам визуальной реконструкции данных учёные обнаружили, что каждый синапс содержит около 1000 молекулярных «переключателей» наподобие аналоговых транзисторов. То есть отдельный синапс можно сравнить с микропроцессором. Получается, что количество «транзисторов» в человеческом мозге теперь нужно увеличить на три порядка. Их больше, чем транзисторов во всех компьютерах на планете вместе взятых¹.

Вывод: Получается что один человеческий мозг по сложности примерно равен всей мировой ИТ-инфраструктуре, а учитывая тот факт, что возможности человеческого мозга задействованы человеком максимум на 20%, говорить о победе ИИ над человеческим разумом не приходится даже в отдалённой перспективе.

Заключение

В рамках эволюции ИИ проводятся исследования возможности имитации процессов человеческого мышления, таких как понимание, интуиция, создание беспричинных прозрений. Ученые ищут способы формирования логических объяснений результатов работы нейронных сетевых систем для повышения уверенности и достоверности выводов, сделанных с помощью таких систем. Ведутся работы в области когнитивного моделирования и крупномасштабные естественные вычисления для целенаправленного и ускоренного решения обратных задач на концептуальных моделях ИИ, принимая их с учетом неформализованной когнитивной семантики.

При обработке информации ИИ использует новые ИТ на основе новых вычислительных систем, которые представлены в виде квантовых и оптических компьютеров. Гибридная киберфизическая система или нейроморфная парадигма используются пока без погружения на атомарном уровне в биологическое вещество.

В новой парадигме человек явно вовлечен в работу системы ИИ, и рефлекторно и когнитивно способен воздействовать на ситуацию. Это включение делает возможным постановку задач в обратных, некорректных и в когнитивных пространствах. Для их решения используется задачи категорирования, когнитивное моделирование, крупномасштабные эволюционные вычисления и генетические алгоритмы. Исследуется обработка семантически интерпретируемых данных с учетом неформализованной и беспричинной когнитивной семантики и когнитивной архитектуры фреймворков². Вместо логико-лингвистической или нейросетевой архитектуры представления предметов и событий, происходит углубление до атомарного уровня их семантической интерпретации. Семантика моделей на основе ASI становится квантово-релятивистской. Сила такой семантической интерпретации на десятки порядков выше, чем у традиционной нейронной сети и логико-лингвистического ИИ. При разработке методов ASI учитываются явления человека и природы в их совокупности – коллективная бессознательная потенциальная сингулярность, квантовая нелокальность, корпускулярно-волновой дуализм, термодинамический и релятивистских эффект, связь низкотемпературной плазмы с веществами мозга и тела, а также спонтанные колебания естественных нейронов. Возможность решения сложных междисциплинарных задач исследовательские группы уже определяют. Будем надеяться на то, что данные направления исследований не засекретят в угоду военно-промышленному лобби или державам, стремящимся к мировому господству.³

¹ В человеческом мозге столько же «транзисторов», сколько их в мировой ИТ-инфраструктуре. – [https://www.cell.com/neuron/fulltext/S0896-6273\(10\)00766-X](https://www.cell.com/neuron/fulltext/S0896-6273(10)00766-X)

² Фреймворк (англ. framework — «каркас, структура») – заготовка, готовая модель в ИТ для быстрой разработки, на основе которой можно дописать собственный код.

³ Дополнительные материалы доступны на онлайн-странице <http://itzashita.ru>