

«СТОХАСТИЧЕСКИЙ ПОПУГАЙ» НА СЛУЖБЕ БИЗНЕСА: УСПЕХИ И ПРОБЛЕМЫ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Т. А. ГАВРИЛОВА

*Институт «Высшая школа менеджмента»,
Санкт-Петербургский государственный университет, Россия*

Цель исследования: представить обзор исследований и технологий искусственного интеллекта на базе нейронных сетей и больших лингвистических моделей. **Методология исследования:** в качестве основного использован кабинетный метод тематических обзоров литературы. **Результаты исследования:** выявлены основные тренды развития искусственного интеллекта, проанализированы и определены скрытые опасности и несовершенства современных систем его внедрения. **Оригинальность и значимость результатов:** критический подход и анализ все возрастающего информационного шума в области продвижения генеративного искусственного интеллекта сформированы под руководством автора в результате выполнения научно-исследовательских проектов Российского фонда фундаментальных исследований и Российского научного фонда, а также многолетней экспертной деятельности в рамках Российской ассоциации искусственного интеллекта. Статья является обобщением нескольких пленарных докладов автора по данной тематике.

Ключевые слова: искусственный интеллект, инженерия знаний, искусственные нейросети, бизнес-приложения искусственного интеллекта.

ВВЕДЕНИЕ

Современный интерес к системам искусственного интеллекта (ИИ) на базе нейронных сетей и больших лингвистических моделей вполне оправдан. Ведущие компании активно осваивают новые цифровые инструменты [Герасименко, 2023; Sadiku, Fagbohunbe, Musa, 2020; Lamarre,

Smaje, Zimmel, 2023]. К началу 2024 г. руководители более 60 % компаний заявили, что приступили к частичному использованию систем ИИ в своей деятельности [McKinsey, 2024a].

Необходимо отметить, что один из пионеров искусственного интеллекта М. Минский назвал его «чемоданом», т. е. таким предметом, в который помещают разные

Адрес организации: Институт «Высшая школа менеджмента», Санкт-Петербургский государственный университет, Волховский пер., 3, Санкт-Петербург, 199004, Россия.

© Т.А. Гаврилова, 2024

<https://doi.org/10.21638/spbu18.2024.305>

вещи [Brockman, 1998]. При этом каждая компания по-своему трактовала его понимание. Использование чат-ботов, машинный перевод и предиктивная аналитика — все это в наши дни искусственный интеллект.

Искусственный интеллект прошел длинный путь развития от момента зарождения в 1956 г. до настоящего времени [Flasiński, Flasiński, 2016; АИТ, 2022; Zhang, Lu, 2021; Захаров и др., 1987; Кобринский, 2022; Осипов, 2011; Финн, 2020]. Однако глубокого понимания основ и критических ограничений столь широкого внедрения ИИ в бизнес-сообществе до сих пор нет. В данной статье представлена попытка краткого исторического обзора исследований в этой области и сформулированы основные парадигмы рассматриваемой проблемы.

Революция в области искусственного интеллекта свершилась, и возможности генеративного ИИ в создании и развитии бизнеса трудно переоценить [Krishnan, Marippan, 2024]. Понятие «генеративный искусственный интеллект» (generative AI) связано со способностью создавать (generate) новые тексты, звук, изображения и видео.

Искусственный интеллект повсеместно внедряется в качестве неотъемлемой части продуктов, услуг и систем в промышленности, науке и образовании [Floridi et al., 2018; AI, 2019; Ahmadi, 2024; Kaggwa et al., 2024; Джонс, 2019]. Этот тренд в цифровой трансформации ускоряет создание специализированных моделей ИИ, которые могут применяться для автоматизации производственных процессов, разработки нового программного обеспечения, обучения, различных прогнозов, машинного перевода и генерации текстов и изображений.

Можно вести речь о том, что параллельно существуют два мира ИИ, появившиеся в середине 1950-х гг. Вскоре после старта первых исследований и разработок в этой области произошло ее разделение на два направления: нейрокибернетику (индуктивный подход на базе данных) и «символьный искусственный интеллект»

(дедуктивный подход на основе обработки знаний).

Эти направления развиваются независимо друг от друга, значимо различаясь как в методологии, так и в применяемых технологиях. Важно отметить, что попытки их сближения уже имелись и продолжаются до сих пор. Также сейчас широко используются новые термины и новые классификации: сильный и слабый ИИ (strong and weak AI (artificial intelligence)), объяснимый ИИ (explainable AI) и общий ИИ (artificial general intelligence — AGI). В Европе часто упоминается об узком ИИ (narrow AI), характеризующимся навыками для реализации конкретных задач, и широком ИИ (broad AI), который стремится решать многообразный спектр проблем, а не ограничиваться одной задачей или областью.

В свете оглушительных успехов нейросетей и триумфального использования больших языковых моделей (large linguistic models — LLM) в программах ChatGPT, CLAUDE, Midjourney и т. д. возникает потребность осмысления основных различий и противоречий в двух указанных направлениях ИИ.

В статье предпринята попытка представить авторскую интерпретацию данной проблемы на основании многолетнего опыта исследований в области символического ИИ и проведенного обзора современной литературы и рынка приложений. Ни философские, ни исторические основы ИИ в работе намеренно не рассматриваются в связи с многосторонностью и глубиной темы, а также ограничениями собственных познаний автора.

В данной работе описывается прагматический взгляд на внедрение ИИ в бизнес с позиций наук об информации (information sciences) и компьютерных технологий (computer science). При этом анализу подвергается не сам факт внедрения, а методология использования реализованных алгоритмов.

Исторические факты, связанные с развитием ИИ, подробно изложены в учебни-

ках [Рассел, Норвиг, 2016; Гаврилова, Хорошевский, 2000; Гаврилова, Кудрявцев, Муромцев, 2016]. В качестве методологии использованы смешанные методы — от кабинетного метода тематических обзоров литературы до глубинных интервью с ведущими российскими исследователями в области ИИ.

Статья имеет следующую структуру. В первом разделе вводятся ключевые определения и обсуждаются основные исторические факты и предпосылки появления новой науки в середине 50-х гг. XX в. Во втором представляются два основных методологических подхода к созданию ИИ — нейросетевой и знание-ориентированный, а также кратко рассматриваются основные вехи и смены парадигм в истории ИИ. В третьем описывается основа современной революции — появление больших языковых лингвистических моделей. В четвертом разделе обобщаются достижения в области систем, основанных на знаниях, и предлагается ознакомиться с миром моделей знаний и графов знаний. В заключении рассматриваются реальные бизнес-приложения и опасности, связанные с излишним доверием к применяемым инструментам ИИ.

О ВАЖНОСТИ АДЕКВАТНОГО ПЕРЕВОДА

Принято считать, что идея об искусственном интеллекте зародилась в России в 1954 г. Именно в этом году в МГУ начал работу семинар «Автоматы и мышление» под руководством академика А. А. Ляпунова¹, в котором принимали участие математики, лингвисты, физиологи, философы и психологи. Как и за рубежом, среди них выделились два основных направления специалистов — нейрокибернетики и кибернетики «черного ящика».

¹ Алексей Андреевич Ляпунов (1911–1973) — советский математик, один из основоположников кибернетики, член-корреспондент АН СССР.

В 1954–1964 гг. появились первые программы ИИ и исследования в области поиска решений логических задач. В Ленинграде в ЛОМИ (Ленинградское отделение Математического института имени В. А. Стеклова) была создана программа ALPEVLOMI, которая автоматически доказывала теоремы, используя обратный вывод С. Ю. Маслова. В числе важнейших результатов, полученных российскими учеными в 1960-е гг., наиболее значимым является алгоритм КОРА М. М. Бонгарда, имитирующий деятельность человеческого мозга для распознавания образов. Выдающийся ученый М. А. Гаврилов² также внес большой вклад в историю российской школы искусственного интеллекта, а его ученики стали пионерами развития искусственного интеллекта в России [Поспелов, 1998].

В нашей стране большую роль в официальном признании искусственного интеллекта как науки сыграли академики А. И. Берг и Г. С. Поспелов. В 1965–1980 гг. возникло новое направление кибернетики — ситуационное управление (соответствует и предшествует «представлению знаний» в общепринятой терминологии). Основателем этой научной школы стал профессор Д. А. Поспелов, который в 1989 г. создал Советскую ассоциацию искусственного интеллекта. В настоящее время это Российская ассоциация искусственного интеллекта³, ежегодные конференции которой объединяют сотни исследователей из России и других стран.

Понятие «искусственный интеллект» было предложено в 1956 г. в рамках летней школы-семинара с одноименным названием в Дартмутском колледже (США). Школа была рассчитана на два месяца, в ней участвовали 10 человек. Тема семи-

² Михаил Александрович Гаврилов (1903–1979) — советский ученый, стоявший у истоков отечественных информатики и кибернетики; член-корреспондент Академии наук СССР.

³ Российская ассоциация искусственного интеллекта. URL: <http://raai.org>. (дата обращения: 19.09.2024).

нара касалась решения на компьютере логических, а не вычислительных задач.

Следует подчеркнуть, что в английском языке словосочетание “artificial intelligence” не имеет той слегка фантастической антропоморфной окраски, которую оно приобрело в довольно неудачном русском переводе. Слово “intelligence” переводится как «способность рассуждать разумно», а вовсе не как «интеллект», для которого есть специальный термин “intellect”. Таким образом, во всем мире моделируют «способность рассуждать» — это гораздо менее пафосно и, что очень важно, в большей степени адекватно характеризующей сути, чем просто моделировать «интеллект», который имеет более сложную структуру и природу.

Дартмутский семинар начался с обсуждения компьютерной программы «Логик-теоретик» для автоматического доказательства теорем, причем одна из них неожиданно была доказана новым улучшенным способом. К концу второго месяца семинара идея ИИ была сформулирована, амбиции и энтузиазм участников не имели предела — научить машину осуществлять некоторые «разумные», осмысленные, а не просто вычислительные процедуры, например: играть в шахматы, переводить на иностранные языки или сочинять музыку. Сегодня все это стало реальностью.

БОРЬБА ДВУХ ПАРАДИГМ

Два направления искусственного интеллекта — нейросети и символичный ИИ — находятся в конкурентной борьбе (рис. 1).

Основную идею первого направления, нейроподхода, можно сформулировать следующим образом: единственный объект, способный мыслить, — это человеческий мозг, поэтому любое «мыслящее» устройство должно каким-то образом воспроизводить его структуру.

Искусственные нейронные сети используют довольно примитивную модель нейрона, описанную далее. Они хорошо за-

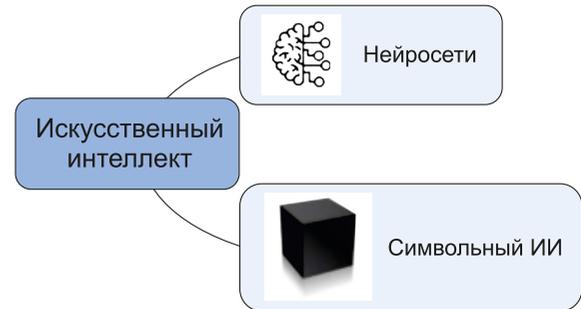


Рис. 1. Два направления искусственного интеллекта

рекомендовали себя в задачах распознавания образов — например, визуальных объектов. В основе этого процесса распознавания лежат алгоритмические модели идентификации состояния системы (объекта) с миллионом параметров и огромными по размеру выборками для обучения. Это очевидным образом является «голой математикой», в которой не обнаруживается никаких признаков «интеллекта» даже в вульгарном понимании данного слова. На вопрос о том, почему система выдает такой ответ, указанные алгоритмы также ответить не могут.

Вся парадигма машинного обучения основана на статистической обработке очень больших массивов данных в буквальном смысле. При этом обработка осуществляется с помощью больших компьютерных кластеров, включающих десятки тысяч процессоров.

В основу второго направления был положен принцип, противоположный нейрокибернетике и получивший название «символьный ИИ», или «черный ящик». Впервые он был сформулирован Н. Винером как создание абстрактной системы, в которой внешнему наблюдателю доступны лишь входные и выходные величины, а структура и внутренние процессы не известны [Винер, 1968].

В случае с искусственным интеллектом это можно перефразировать так: «Не имеет значения, как устроено искусственное “мыслящее” устройство. Главное, чтобы

на заданные входные воздействия оно реагировало не хуже, чем мозг человека» [Гаврилова, Хорошевский, 2000, с. 7]. Для того чтобы понять центральную парадигму данного подхода — работу со знаниями как закономерностями и принципами функционирования предметной области, надо обратиться к истории ИИ [Поспелов, 1998; Гаврилова, Хорошевский, 2000].

Сторонники указанного направления мотивировали свою позицию тем, что человек не должен слепо следовать природе в собственных технологических изысканиях, особенно в вопросах спорных, которые не изучены глубоко. Например, очевиден успех колеса, которого нет в природе, или самолета, не машущего крыльями, подражая птице. К тому же науки о человеке не смогли внести существенного теоретического вклада, объясняющего хотя бы приблизительно, как протекают процессы принятия решений, как устроен механизм памяти человека и как происходит познание окружающего мира.

Символьный ИИ был ориентирован на поиски универсальных алгоритмов решения интеллектуальных задач с использованием современных компьютеров. Под «интеллектуальными» понимались задачи, которые при их решении человеком требуют рассуждений, размышлений и когнитивных усилий, традиционно считающихся интеллектуальными (например, перевод с иностранных языков, решение головоломок, игра в шашки или шахматы, сочинение стихов и музыки).

НЕЙРОННЫЕ СЕТИ

Исторически расстановка сил в гонке двух направлений ИИ менялась — был период, когда лидировали системы, основанные на знаниях, но в настоящее время безусловная пальма первенства принадлежит нейрокибернетике.

Считается, что нейрокибернетика ориентирована на программно-аппаратное моделирование структур, подобных структуре

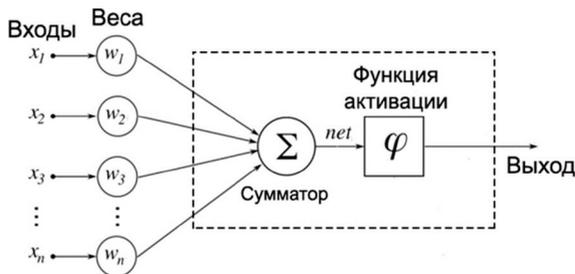


Рис. 2. Принцип работы искусственной нейросети

Источник: PythonR U. URL: <https://pythonru.com/wp-content/uploads/2020/03/process-aktivacii-kazhdogo-nejrone.png> (дата обращения: 19.09.2024).

мозга [Горбань, 2002]. Физиологи установили, что основой человеческого мозга является большое количество (до 10^{21}) связанных между собой и взаимодействующих нервных клеток — нейронов. Поэтому усилия нейрокибернетики были сосредоточены на создании элементов, аналогичных нейронам, и их объединении в функционирующие системы. Эти системы принято называть искусственными нейронными сетями, или нейросетями (рис. 2).

Основной принцип работы нейронной сети заключается в обучении на данных путем точной настройки связей между нейронами или в создании на основе эмпирических данных модели, точно воспроизводящей существующие настройки связей между нейронами. Когда нейронная сеть обучается, она проходит процесс корректировки своих весов $w = (w_1, \dots, w_n)$ и смещений (параметры для составления ответа) с целью минимизации ошибки в прогнозах. Постепенно это позволяет сети «выучить» наиболее эффективный способ обработки входных данных и получения точных результатов [Круг, 2002].

Нейросеть обрабатывает входящий запрос пользователя, который, будучи представлен вектором переменных «входа» $x = (x_1, \dots, x_n)$, подается на узлы нейронов. Входные данные анализируются сетью с учетом весов. У каждого «входа» имеется

свой вес. Процедура повторяется несколько раз, т. е. имеется несколько слоев узлов. Далее сигнал передается через эти скрытые слои, и после обработки всех слоев нейросети сумматор по специальной формуле выводит ответ, который соответствует наиболее вероятному результату, основанному на предварительном анализе и обучении нейросети.

Активация следующего уровня (слоя) происходит только тогда, когда сумма всех входов достигнет порогового значения. Так работает функция активации. Входными сигналами могут служить оцифрованное изображение, текст, аудио- и видеосигналы. Например, после обучения на 100 фотографиях автомобилей «Лада» сеть сможет узнавать эти авто среди других.

Данная модель обучения хорошо проявила себя в распознавании визуальных и звуковых образов, в финансовых прогнозах, игровых задачах, обработке естественного языка, включая машинный перевод, а также в рекомендательных системах (советы по приобретению товаров или поиску информации) [Дэвенпорт, 2021; Arman, Lamiyar, 2023].

Первые нейросети были созданы Ф. Розенблаттом и У. Мак-Каллоком в 1956–1965 гг. Предпринимались попытки разработать системы, моделирующие человеческий глаз и его взаимодействие с мозгом. Созданное ими устройство получило название персептрона (perceptron). Персептрон умел различать буквы алфавита, но был слишком чувствителен к их написанию. Например, буквы «А», «А» и «а» для него были тремя разными знаками. Первые результаты применения вызвали разочарование, и в 1970-х гг. количество работ в этом направлении стало снижаться. Авторы объясняли неудачи малой памятью и низким быстродействием существовавших в то время компьютеров. Ситуация изменилась в связи с технологическим прорывом и появлением современных процессоров с их сверхмощностями.

Со временем нейросети стали «учиться» все лучше, и успехи этого направления

сегодня трудно переоценить. В полной мере это демонстрируют, например, картины, созданные алгоритмом GAN⁴, или достижения в распознавании лиц и голоса. Нейросети создают тексты, музыку, видео, рисуют картины по текстовым описаниям и имитируют разговор [Горбань, 2000; Николенко, Кадури, Архангельская, 2018; Оселедец, 2022; Feuerriegel et al., 2024; Kawakami, Venkatagiri, 2024].

СИМВОЛЬНЫЙ ИИ, ИЛИ КИБЕРНЕТИКА «ЧЕРНОГО ЯЩИКА»

Существенный вклад в становление данного направления внесли ее пионеры: Дж. Маккарти — автор первого языка программирования для задач ИИ — ЛИСПа; М. Минский — основоположник идеи фрейма и создатель фреймовой модели представления знаний; А. Ньюэлл, Г. Саймон, Дж. Шоу — разработчики программ «Логик-теоретик» и «Общий решатель задач»; И. Хант — автор одного из первых учебников по искусственному интеллекту и другие исследователи. Отдельные аспекты истории искусственного интеллекта и его направлений хорошо освещены в работах отечественных и зарубежных классиков [Люгер, 2003; Поспелов, 1982; Стефанюк, 2011; Осипов, 2011; Финн, 2020].

Начиная с 1956–1963 гг. проводились активные поиски моделей и алгоритмов человеческого мышления и разработка первых программ, моделирующих понимание естественного языка, осуществляющих машинный перевод, а также творческую деятельность. Исследователи в области гуманитарных наук — философы, психологи, лингвисты — не смогли предложить таких алгоритмов. Тогда кибернетики приступили к созданию собственных моделей. Последовательно были предложены

⁴ Art AI. URL: <https://www.artaigallery.com> (дата обращения: 19.09.2024).

и опробованы различные подходы, которые кратко описаны далее.

Так, в конце 1950-х гг. появилась модель лабиринтного поиска. Этот подход подразумевает представление задачи как некоего пространства состояний в форме графа, в котором проводится поиск оптимального пути от «входных» данных к результирующим. В первых переводных монографиях по искусственному интеллекту [Хант, 1978; Эндрю, 1985] описаны такого рода программы, они собирают «Ханойскую башню», играют в «15», шашки и шахматы. Была проделана большая работа, модель хорошо зарекомендовала себя в игровом поле, но в решении практических задач идея лабиринта распространения не получила. Позднее эти алгоритмы были дополнены эвристиками⁵.

В 1963–1970-х гг. к решению логических задач стали подключать методы математической логики. В 1965 г. Дж. Робинсон разработал метод резолюций, который позволяет автоматически доказывать теоремы при наличии набора исходных аксиом [Robinson, 1965]. Но годом ранее выдающийся отечественный математик С. Ю. Маслов⁶ предложил так называемый обратный вывод, впоследствии названный его именем («обратный вывод Маслова»), решающий аналогичную задачу другим способом [Маслов, 1964].

На основе метода резолюций французский математик А. Кольмероэ в 1973 г. создал язык логического программирования PROLOG (Programming Logics) — пролог для программирования логических задач. Большой резонанс имела программа «Логик-теоретик», разработанная А. Ньюэллом, Г. Саймоном и Дж. Шоу, ко-

⁵ В начале 1960-х гг. эвристикой называли правило, теоретически не обоснованное, но позволяющее сократить количество переборov в пространстве поиска [Александров, 1975].

⁶ Сергей Юрьевич Маслов (1939–1982) — советский математик, доктор физико-математических наук, специалист в области математической логики (в частности, автоматизации процесса доказательства).

торая доказывала школьные теоремы [Newell, Shaw, Simon, 1958; Newell, Simon, 1976; Рассел, Норвиг, 2016].

Логический подход при всей его значимости не способствовал решению не только производственных, но и технических задач. Дело в том, что большинство реальных ситуаций не сводится к набору аксиом и человек, решая бизнес-задачи, не использует классическую логику. Именно поэтому логические модели имеют существенные ограничения по классам решаемых задач.

История искусственного интеллекта полна необычных поворотных событий, одним из которых стал «Доклад Лайтхилла», который был подготовлен в 1973 г. по заказу Британского совета научных исследований в Великобритании. Известный математик Д. Лайтхилл, профессионально не связанный с искусственным интеллектом, подготовил обзор состояния дел по данной проблеме. Он признал определенные достижения в области ИИ, однако их уровень обозначил как разочаровывающий, приведя общую отрицательную оценку с позиций практической значимости. Этот отчет отбросил европейских исследователей примерно на пять лет назад, так как финансирование исследований в области ИИ значительно сократилось.

Существенный прорыв в разработке практических приложений ИИ произошел в середине 1970-х гг., когда на смену поискам универсального алгоритма мышления пришла идея моделировать конкретные знания специалистов-экспертов. Так появились первые коммерческие системы, основанные на знаниях, или экспертные системы (ЭС), и новый подход к решению задач искусственного интеллекта — представление знаний.

Под знаниями в информатике и в области искусственного интеллекта понимают основные закономерности, принципы, правила, позволяющие специалистам решать задачи в своей профессиональной сфере [Гаврилова, Хорошевский, 2001]. Знания — вид информации, отражающей

опыт специалиста (эксперта) в определенной предметной области, его понимание множества текущих ситуаций и способы перехода от одного описания объекта к другому⁷.

Первые экспертные системы MYCIN (постановка медицинских диагнозов и назначение антибактериальной терапии) [Wraith et al., 1976] и DENDRAL (расшифровка результатов анализа масс-спектрограмм в химии) [Feigenbaum, 1981] стали осязаемой победой в ИИ. Обе системы стали уже классическими и вошли во все учебники. В дальнейшем появились тысячи успешных ЭС для различных предметных областей — от археологии до юриспруденции (см., напр.: [Хант, 1978; Эндрю, 1985; Рыбина, 2010; Башлыков, Еремеев, 1994; Гаврилова, 1984]). Хотя пик интереса к ЭС в настоящее время пройден, разработка систем, основанных на знаниях, успешно продолжается и остается актуальной для широкого спектра предметных областей знания [Стюарт, Норвиг, 2016; Загорулько, Загорулько, 2020].

В ответ на успехи развития экспертных систем в США и Европе в конце 1970-х гг. в гонку включилась Япония, объявив о начале проекта создания машин пятого поколения, основанных на знаниях. Проект был рассчитан на 10 лет и объединял лучших молодых специалистов (в возрасте до 35 лет) крупнейших японских компьютерных корпораций. В итоге ученые создали достаточно громоздкий и дорогой символьный процессор, который аппаратно реализовал язык программирования PROLOG (ПРОЛОГ). Необходимо отметить, что проект считается «провальным», цель не была достигнута, но в его рамках группа высококвалифицированных специалистов добилась существенных результатов в различных прикладных задачах, параллельно создав в одной из лабораторий первую техническую нейроплату [Moto-Oka, 2012].

⁷ ФИНАМ. URL: <https://www.finam.ru/publications/item/znaniya-20230628-2051> (дата обращения: 19.09.2024).

С середины 1980-х гг. парадигма работы со знаниями получила коммерческий импульс развития. Создавались и продавались промышленные экспертные системы, усиливался интерес к самообучающимся системам. Издавались десятки научных журналов, ежегодно собирались международные и национальные конференции по различным направлениям ИИ⁸. Появилась новая наука — инженерия знаний [Adeli, 1994; Harmelen, Antoniou, 2004], исследующая модели и методы извлечения и структурирования знаний [Гаврилова, Хорошевский, 2001; Загорулько, Загорулько, 2020].

Инженерия знаний мультидисциплинарна, она актуальна во всех «мягких» предметных областях, в которых важны знания и опыт специалистов, где эти знания надо собирать, структурировать, оцифровывать и использовать многократно. Специалисты-аналитики, или инженеры по знаниям (knowledge engineers), извлекают, структурируют и формализуют знания, тем самым создают предметные базы знаний по узким областям (например, базы знаний по расшифровке кардиограмм или по работе с электронным томографом). Также они разрабатывают и корпоративные порталы знаний, аккумулирующие знания для целых компаний и предприятий.

В настоящее время почти все большие компании имеют корпоративные базы знаний (порталы знаний), содержащие отчеты, регламенты, инструкции, лучшие практики, протоколы мозговых штурмов, анализов ошибок, «разборов полетов», и другие носители знаний. Многие из них интерактивны, кастомизированы и проактивны.

Базы знаний — дорогостоящий инструмент управления знаниями организации, создание и поддержание которого составляют постоянный процесс, требующий

⁸ Журналы «Искусственный интеллект и принятие решений», «Нейрокомпьютеры: разработка и применение», IEEE Transactions on Pattern Analysis and Machine Intelligence, Knowledge-based Systems, Expert Systems and Applications и т. д.

серьезных вложений, опытного штата аналитиков и корпоративной культуры, побуждающей к обмену знаниями. Так, в компании Tinkoff (с 5 июня 2024 г. название изменено) база знаний содержит: инструкции по работе с ПО, регламенты процессов, шаблоны документов, типовые кейсы и лучшие практики, советы для новичков, а также многие другие полезные элементы и сегменты информационного поля предприятия [Фитисов, 2024].

НЕЙРОСЕТИ КАК СТАТИСТИЧЕСКИЙ ПОДХОД НА ОСНОВЕ ДАННЫХ

Бесспорные успехи нейроподхода и машинного обучения [Miikkulainen et al., 2018; Николенко, 2018; Миркин, Кузнецов, 2009; Feuerriegel et al., 2024], особенно в течение последних двух-трех лет в области работы с текстами, визуального распознавания образов и обработки больших данных (финансовое моделирование, маркетинговая аналитика, кибербезопасность и т. д.), производят сильное впечатление и практически не оставляют места критическому анализу [Mirajilili, 2019; Котов, Соколова, Павлов, 2023].

Крупные компании, например ПАО «Сбербанк» и ПК «Яндекс», давно используют нейросетевые алгоритмы. Так, ПАО «Сбербанк» улучшает клиентский сервис через голосовых помощников и чат-ботов, повышает безопасность, автоматизирует работу с персоналом. Кроме того, ИИ применяется для анализа кредитоспособности клиентов. Модели машинного обучения анализируют огромные объемы данных, чтобы оценить риски и принять решение о выдаче кредита. Это позволяет быстрее, точнее и, что важно, с меньшими издержками определять, кому можно предоставлять кредитные средства. В отличие от традиционных методов оценки кредитоспособности ИИ способен учитывать множество факторов, включая поведение

клиента, его финансовую историю и даже социальные связи [Skupro, 2024].

Успешные алгоритмы генерации текстов и картин на основе идей машинного обучения породили новое понятие «генеративный ИИ» и целый ряд ложных иллюзий у пользователей. Его название происходит от глагола to generate — «генерировать», т. е. создавать.

При этом принцип действия любой искусственной нейронной сети (рис. 2) остается неизменным: «Искусственный нейрон — это такая функция, которая преобразует несколько входных сигналов (фактов, параметров...) в один выходной на основе некоей формулы активации (ϕ) порога возбуждения (threshold)» [Хабр, 2024]. Настройкой весов (w_1) этих фактов, а также порога возбуждения регулируется адекватность нейрона (рис. 2). Под адекватностью понимается правильное поведение нейросети после обучения, т. е. способность к выполнению поставленной задачи. Если нейросеть обучали узнавать на фото светофоры, то она и делает это успешно.

Современные нейросети оперируют миллионами параметров и тысячами слов, но механизм обучения остается прежним. Вместе с тем в рамках биологии и нейронаук (рис. 3) работа мозга рассматривается на множестве уровней с разных сторон. В этих науках существует ряд серьезных проблем в вопросах изучения сложных закономерностей, принципов и экспериментов. На рис. 3 проиллюстрировано, насколько многогранны науки об естественной нейросети. На пути познания реальных механизмов активации нейронной Вселенной возникает масса междисциплинарных сложностей.

С использованием нейросетей улучшился машинный перевод, но это произошло без приближения к пониманию текста. Аналитики задаются вопросом: «Все эти алгоритмы демонстрируют способность искать сходство на сверхбольших объемах данных и находить его, обучаясь генерировать подобные. Где здесь “интеллектуальность”?»



Рис. 3. Нейронауки

Однако такие успешные технические проекты ИТ-лидеров, в том числе алгоритмы-трансформеры GPT4 (Generative Pre-trained Transformer) от компании OpenAI и подобные им алгоритмы, позволяют создать внешне осмысленные тексты, например, на заданную тему или в качестве продолжения начатой мысли [Floridi, Chiriatti, 2020]. Правда, за нелепые ошибки критики называют их «генераторами умно звучащего бреда» [Рождественская, 2022], или «стохастическим попугаем».

В основе программ генеративного ИИ лежат большие языковые модели, которые сочиняют тексты методом «предсказания» наиболее вероятного следующего слова, предварительно обучившись на массиве подобных текстов. Устройство таких моделей не поддается простому объяснению, но можно отметить, что число анализируемых параметров программы ChatGPT4 превысило 1 трлн. Ее обслуживает кластер из более 200 тыс. процессоров и 10 тыс. графических карт [Arman, Lamiyar, 2023]. Этот кластер потребляет огромное количество гигаватт-часов электроэнергии [Cohen, 2024]. Ожидается, что подобные центры будут поглощать 1,5 % мирового объема потребления в течение ближайших пяти лет [TechInsights, 2024].

Не утихают спекуляции и вокруг машинного «творчества». Условно бесплатная платформа Midjourney⁹, конкурирующая с про-

граммой DALL-E от Open-AI, позволяет генерировать изображения, ролики в 3D или архитектурные эскизы исключительно высокого качества [Oppenlaender, 2022]. Она копирует манеры и стили определенных художников или направлений (рис. 4).

Для получения изображения достаточно ввести на входе текст-подсказку с его описанием (например, «Зверь с головой греческого бога побеждает современный танк, живопись в стиле Кирико» или «Мона Лиза средствами морского пейзажа»). Принцип работы — все те же классические статистические пермутации (или перестановки элементов), пропущенные через миллионные обучающие выборки на примерах. Торжество данных и математики над смыслом и творчеством! В данном случае правомернее вести речь об искусственном «ремесле».

По-прежнему «трудные» для людей задачи, например: сложные расчеты, анализ финансовых рынков, подстрочные языковые переводы, — просты для компьютера. «Простые» же задачи, такие как работа органов зрения, движение, восприятие, — остаются трудными для машины. Известный ученый в области компьютерных наук Д. Кнут говорил, что искусственный интеллект преуспел во всем, что требует «мышления», но провалился в задачах, которые люди и животные делают не задумываясь [Knuth, 1997].

Общеизвестны ошибки и конфузы указанных алгоритмов [Raji et al., 2022]. Например, когда было обнаружено более

⁹ Midjourney. URL: Midjourney.com (дата обращения: 19.09.2024).



Рис. 4. Пример коллажа от генератора изображений Midjourney
Источник: TechCrunch. URL: <https://techcrunch.com/category/artificial-intelligence/12.2023> (дата обращения: 19.09.2024).

20 тыс. случаев мошенничества с пособиями по безработице, ложно помеченных алгоритмом MIDAS в Мичигане, американцу Брайану Расселу пришлось объявить себя банкротом, обездолив семью с двумя детьми. Государство сняло с него ложные обвинения лишь два года спустя [Charette, 2018].

Однако сейчас привлекательность и внешняя простота использования нейросетей в маркетинге, операционном менеджменте, логистике и других направлениях перевешивают его недостатки [Caiming, Lu, 2021; Campton, 2024].

СМЫСЛОВАЯ ОБРАБОТКА В ГРАФАХ ЗНАНИЙ

В логико-философской традиции понятия смысла и значения чаще всего идентичны. Для понимания смыслов человек использует свои знания. Под знаниями подразумеваются закономерности и принципы, связывающие данные. Символьное представление знаний используется в интеллектуальных системах с 1976 г., с момента появления баз знаний экспертных систем. Первыми моделями представления знаний выступали продукции, фреймы, се-

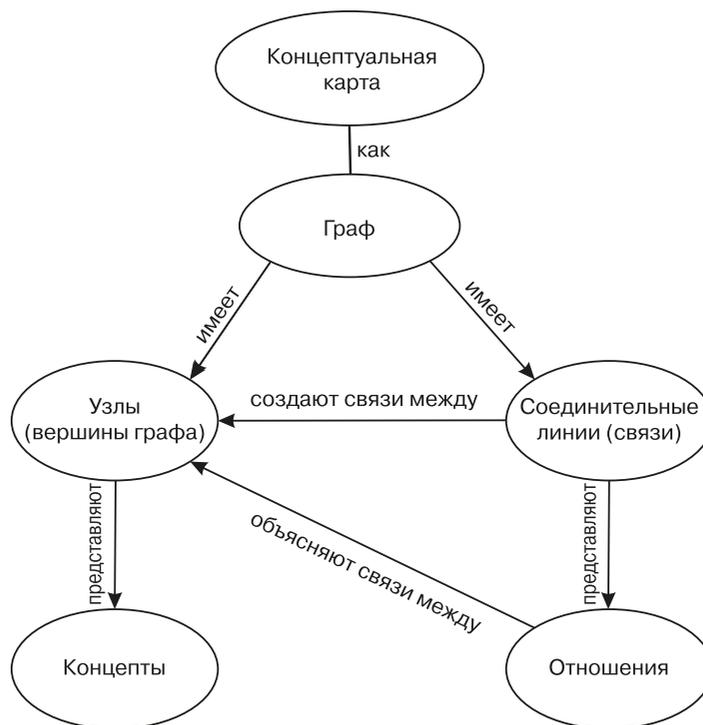


Рис. 5. Концептуальная карта как граф: схематическое представление
 Составлено по: А. Пуляевская. URL: <https://nitforyou.com/konceptualnye-karty/>.
 (дата обращения 20.10.2024).

мантические сети и т. д. [Harmelen, Antoniou, 2004; Загорулько, Загорулько, 2020].

Проще всего подобные структуры отображать в форме графа, узлы которого — понятия предметной области, а ребра — отношения или связи между понятиями (рис. 5).

В 1980-е гг. семантические сети и фреймы были стандартизированы термином «онтология» [Gruber, 1993] как спецификацией концептуализации в виде набора понятий и отношений. Онтология — это способ представления знаний, реализованный на основе формального описания объектов, фактов и отношений между ними [Harmelen, Antoniou, 2004; Sure, Staab, Studer, 2009].

Из краткой истории искусственного интеллекта, изложенной ранее, следует, что зарождение символического ИИ связано с появлением экспертных систем и разработ-

кой моделей представления знаний (семантические сети, фреймы и продукции). Онтологии выступают продуктом эволюции и слияния моделей семантических сетей и фреймов.

Модель фрейма предложил один из пионеров ИИ М. Минский [Минский, 1979] для определения абстрактных структур стереотипных понятий, ролей, сценариев и явлений. Идея оказалась чрезвычайно плодотворной, так как человек часто использует такие абстракции или «шаблон» для минимизации когнитивных усилий. Базы знаний из корпоративных «шаблонов» получили достаточно широкое распространение.

Еще одной популярной моделью стали «продукционные правила» (production rule), или продукции. Такая модель описывается правилом «если “условие”, то “действие”» [Юрин, 2016]. Например:

если «имеется производственная необходимость» и «есть средства», то «командировка разрешена».

При этом традиционно в исследованиях особое внимание уделялось этапу формализации описания знаний. Однако трудности по созданию и структурированию смысловых семантических верхне-уровневых моделей предметных областей большинство авторов обходили стороной в силу трудоемкости и междисциплинарности процесса. В первую очередь это обусловлено квалификацией разработчиков, которые в основном были программистами. Программисты создавали архитектуру программного продукта, а тонкости и глубина описаний знаний их интересовали в гораздо меньшей степени. В инженерии знаний применяется триада «извлечение знаний — структурирование знаний — представление (или формализация) знаний». Именно первые две составляющие долгое время недооценивались разработчиками, что приводило к значительному обеднению баз знаний.

Разработка семантических моделей представления знаний — междисциплинарная задача, требующая от аналитика погружения в предметную область и терминологию (извлечение знаний), понимания ее специфики и организации (структурирование) и, наконец, разработки самой модели (представления знаний). Задача составления даже простого упорядоченного словаря понятий и связей между ними является нетривиальным упражнением в области системного анализа, лингвистики и структурирования знаний конкретной предметной области [Смирнов и др., 2014; Добров и др., 2015]. Именно такие упорядоченные словари стали называть онтологиями, и сейчас это одна из наиболее перспективных моделей представления знаний.

Необходимо отметить, что в настоящее время помимо «ручных», т. е. разрабатываемых вручную, активно используются автоматические способы создания онтологий [Бова и др., 2014]. Это позволяет дей-

ствовать в промышленных масштабах, однако качество и семантическая глубина таких онтологий иногда уступают экспертным. «Ручные» онтологии субъективны, они фиксируют концептуальную структуру предметной области в понимании составителя, и в этом есть элемент творчества. Машинные онтологии тривиальны, они проще и беднее по структуре.

Инженерия знаний развивалась и выросла вместе с возникновением новых моделей. У нее появились дочерние ветви, например, онтологический инжиниринг (ontology engineering) [Sure, Staab, Studer, 2009; Боргест, 2013; Смирнов и др., 2014; Гаврилова, 2005].

Онтологический инжиниринг включает все этапы работы с онтологиями — от дизайна до программной реализации на языках программирования OWL и RDFs. Часто возникают сложности при слиянии (merging), выравнивании (alignment) и наполнении (population) онтологий. Разработано множество редакторов онтологий, например бесплатный инструмент PROTÉGÉ¹⁰.

Однако эпоха больших данных повлияла и на модели знаний. Так возникли сложные многосвязные структуры, облегчающие поиск и связь данных, — графы знаний. Первый граф знаний (knowledge graph) был создан в 2012 г. компанией Google как база знаний для поисковой системы [Pujara et al., 2013], основанная на онтологиях. Графы знаний используют онтологии в качестве каркаса представления данных, включающего терминологический словарь и набор утверждений о моделируемых объектах [Муромцев, Волчек, Романов, 2019].

Граф знаний — набор концептов (объектов, понятий), соединенных друг с другом отношениями. Он может отображать знания из нескольких (например, в базах DBpedia и Yago) или каких-то конкретных областей (так, графы Bio2RDF и UMLS содержат знания из биологии и медици-

¹⁰ Protégé. URL: <https://protege.stanford.edu> (дата обращения: 19.09.2024).

ны), а также знания, относящиеся к компании [Heist et al., 2020].

Существует ряд других способов представления знаний, но графы знаний стали самым популярным форматом для репрезентации знаний в конкретной области [Бхатт и др., 2020]. Графы знаний используются также в современных цифровых помощниках, например: в Amazon Alexa, Apple Siri и т. д. В компаниях разрабатываются предметные графы знаний, которые выступают как смыслообразующие структуры для корпоративных информационных систем [Peng et al., 2023].

Графы знаний основаны на смысловых или семантических связях и близки по содержанию к интуитивному представлению о том, как человек познает мир. Эта модель хорошо согласуется с теориями из когнитивных наук.

ЗАКЛЮЧЕНИЕ. КЛЕТКА ДЛЯ ПОПУГАЯ

Современные бизнес-приложения искусственного интеллекта столь разнообразны по форме и структуре, что любые их обобщения являются затруднительными. Как справедливо отметил Ю. Астман, старший партнер Центра стратегических инноваций компании McKinsey, по результатам последнего опроса можно выделить как минимум шесть уровней использования ИИ в стратегическом менеджменте [McKinsey, 2024a].

Уровень 1. Анализ показателей, различные дэшборды и выявление простых закономерностей.

Уровень 2. Создание диагностических интеллектуальных систем, анализирующих причинно-следственные связи в различных бизнес-кейсах.

Уровень 3. Проведение предиктивной аналитики, осуществляющей прогнозы на основе данных.

Второй и третий уровни — это своего рода клад для толкового руководителя, который, например, может заменить рабо-

ту десятка аналитиков для формирования нескольких вариантов портфелей инвестиций. Указанные три уровня доступны для реализации уже сегодня и в России, и за рубежом.

Следующие три уровня потребуют времени на осуществление.

Уровень 4. Внедрение виртуального «консультанта» руководителя.

Уровень 5. Разработка цифрового ЛПР (лица, принимающего решение).

Уровень 6. Создание автономного руководителя, управляющего всеми процессами в компании.

Ю. Астман подчеркивает, что трудно выделить область, наиболее готовую к внедрению ИИ, при этом успехов следует ожидать от компаний, накопивших большие базы данных с показателями деятельности. Именно там искусственный интеллект окажется наиболее эффективным [McKinsey, 2024a].

При этом эксперты остерегаются от стихийного внедрения ИИ в отдельных подразделениях, ведя речь о том, что любая компания — это система. Если в качестве примера системы взять автобус и представить, что одно его колесо вращается быстрее трех других, то вряд ли это пойдет на пользу передвижению всего автобуса [McKinsey, 2024b].

В настоящее время ожидания общества от использования искусственного интеллекта перегреты усилиями журналистов и СМИ. Где и когда появятся системы сильного ИИ (strong AI)? Какое направление победит в долгосрочной перспективе?

В статье кратко описаны два направления ИИ — статистическое, основанное на нейросетях, машинном обучении и больших данных, и символическое, или семантическое, базирующееся на знаниях.

С появлением в 2022 г. больших языковых моделей и успехом программ ChatGPT, CLAUDE, Bard невозможно недооценивать ведущую роль нейросетевых алгоритмов. Нейроподход, «питаясь» большими данными, мощным «железом» и огромными финансовыми вложениями,

стал технологией, обещающей человеку помощь при выполнении множества функций и освобождении рабочих мест. Речь идет о системах, названных генеративным ИИ, способных порождать (генерировать) связные тексты и графические образы.

В системах генеративного ИИ используются новые оригинальные идеи и алгоритмы (токенизация, трансформеры, «внимание»), которые пока еще недостаточно описаны в литературе. Кроме того, как и многое в мире информатизации, эти технологии опередили теорию: они сначала заработали, а потом появились первые скромные публикации по их осмыслению. При этом многие рутинные процедуры по созданию текстов, образов, программ нейросеть может делать быстрее и лучше работника-новичка.

Однако сверхмощный математический аппарат и изощренные алгоритмы обучения не застрахованы от ошибок. Нередко системы генерируют советы, которые называют «галлюцинациями», фактически это порождение бессмысленных артефактов [Ji et al., 2023]. Не стоит говорить, насколько они опасны в реальных ситуациях, уже описаны случаи ложных медицинских рекомендаций [Evans, Snead, 2024], опасных советов по сбору грибов и т. д. Далека от решения и проблема «непрозрачности» рекомендаций. В отличие от экспертных систем нейросеть не может объяснить, откуда берется ее совет, кроме того, на один и тот же запрос она дает разные ответы. Она «играет» данными, фактически это и есть «черный ящик».

Следует отметить, что ведущие разработчики понимают все это и пытаются научить системы «рассуждениям», работе «по шагам», тем самым увеличивая «прозрачность» и «объяснимость» ответов систем. Уже появились первые программы с успешными попытками рассуждений, например программа ChatGPT4 o1 с системой рассуждений strawberry¹¹. Однако

принципы функционирования искусственных нейронных систем и больших языковых моделей «противятся» этому, поскольку решение в них не выводится логически, а подбирается статистически [McKinsey, 2024b].

В настоящее время оба направления ИИ довольно независимы друг от друга и со своими успехами и провалами борются самостоятельно. Вместе с тем все громче звучат голоса исследователей, пытающихся соединить края пропасти. Так, в июне 2024 г. был объявлен конкурс по созданию двухстороннего искусственного интеллекта BILateral AI (BILAI), финансируемый Австрийским научным фондом (FWF), который набирает на проект 50 аспирантов и 10 постдокторантов [Bilateral AI, 2024].

В BILAI крупные австрийские игроки в области искусственного интеллекта объединяются для работы над широким спектром проблем. Для развития своих основ BILAI использует двухсторонний подход, эффективно сочетая субсимвольный ИИ (нейронные сети и машинное обучение) с символьным ИИ (логикой, представлением знаний и рассуждением) различными способами. Задействование всего потенциала как символьного, так и субсимвольного подходов может открыть новые возможности для развития ИИ, повышая его способность решать новые проблемы, адаптироваться к различным средам, улучшать навыки рассуждения и повышать эффективность.

Сходные попытки предпринимают китайские исследователи, развивая технологию графовых нейросетей Graph Neural Networks (GNN) [Xie et al., 2022], где нейросеть обрабатывает данные в форме графов. Однако прогнозы ведущих исследователей по поводу перспектив этого направления довольно осторожны.

Следует констатировать, что успехи в области развития ИИ весьма стремительны и опережают любые прогнозы, а компании-разработчики могут в недельный срок представить новые версии своих чат-ботов, которые, по оценкам экспертов,

¹¹ OpenAI <https://openai.com/o1> (дата обращения: 19.09.2024).

требует на разработку два-три года. При этом в 2024 г. рынок AI оценивается в 298,25 млрд долл. [Duarte, 2024]. Все крупнейшие игроки рынка информационных технологий участвуют в игре, и гонка моделей находится в самом разгаре. Успехи программы ChatGPT4 o1 (компания Open AI) в решении интеллектуальных тестов уровня студентов университетов уже сопоставимы с результатами программ Gemini (Bard) (компания Google), Claude 3.5 Sonnet (компания Anthropic) и Llama3 405B (компания META AI*) [Joseph, 2024; Razzaq, 2024].

Современные нейросети не несут особой опасности для бизнеса (этого нельзя отрицать для общества в целом), но на кадровый состав и технологии работы организаций влияют существенно. Довольно скоро они смогут заменить середняков, ремесленников, «офисный планктон», однако мало затронут «творцов», экспертов и генераторов идей.

Во многих компаниях системы ИИ позволяют сократить команды ИТ-специалистов — именно для генерации программного кода эти программы особо полезны. Например, программа GigaChart от ПАО «Сбербанк» успешно пишет программы по текстовым подсказкам на естественном языке [Kosenko, Kuratov, Zharikova, 2023].

ЛИТЕРАТУРА НА РУССКОМ ЯЗЫКЕ

- Александров Е. А. 1975. *Основы теории эвристических решений*. М.: Советское радио.
- Башлыков А., Еремеев А. П. 1994. *Экспертные системы поддержки принятия решений в энергетике*. М.: МЭИ.
- Бова В. В., Кравченко Д. Ю., Лещанов Д. В., Новиков А. А. 2014. Компьютерная онтология: задачи и методология построения. *Информатика, вычислительная техника и инженерное образование* (4).
- Боргест Н. М. 2013. Ключевые термины онтологии проектирования: обзор, анализ,

Искусственный интеллект представляет большую опасность для научных исследований, создавая тонны бессмысленного вторичного контента (статей, книг, презентаций) и синтетических данных, которые затем поглощаются теми же языковыми моделями и учатся на них. Этот вечный двигатель трудно остановить, а программы-детекторы работают пока недостаточно надежно. Специалисты из Кембриджа в своем открытом отчете в формате презентации (более 200 слайдов) ежегодно проводят глубокий анализ изменений в данной области [Benaich, Air Street Capital, 2024].

Цифровая трансформация бизнеса случилась, и искусственный интеллект играет в ней не последнюю роль, желает того общество или нет. Задача сегодняшнего дня — указать «стохастическому попугаю» на его место в клетке, просторной, но крепкой.

БЛАГОДАРНОСТЬ

Автор выражает искреннюю благодарность Юрию Васильевичу Федотову, чьи внимательное и творческое прочтение, замечания, обсуждение и рекомендованные литературные источники существенно улучшили качество статьи.

обобщения. *Онтология проектирования* 3 (9): 9–31.

Бхатт Ш., Чжао Ц., Сетх А., Шалин В. 2020. Графы знаний как средство улучшения искусственного интеллекта. *Открытые системы*. СУБД (3): 24–26.

Винер Н. 1968. *Кибернетика, или Управление и связь в животном и машине*. М.: Советское радио.

Гаврилова Т. А. 1984. Представление знаний в экспертной диагностической системе АВТАНТЕСТ. *Известия АН СССР. Техническая кибернетика* (5): 168–175.

* Meta признана экстремистской организацией в Российской Федерации.

- Гаврилова Т.А., Хорошевский В.Ф. 2000. *Базы знаний интеллектуальных систем. Учебник*. СПб.: Питер.
- Гаврилова Т.А. 2005. Об одном подходе к онтологическому инжинирингу. *Новости искусственного интеллекта* (3): 25–31.
- Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. 2016. *Инженерия знаний. Модели и методы*. СПб.: Лань.
- Герасименко В.В. 2023. Цифровая этика применения искусственного интеллекта в бизнесе: осознание новых возможностей и рисков. *Научные исследования экономического факультета* 15 (1): 37–54.
- Горбань А.Н. 2000. Нейроинформатика: кто мы, куда мы идем, как путь наш измерить. *Информационные технологии* (4): 10–14.
- Джонс М.Т. 2019. *Программирование искусственного интеллекта в приложениях*. Саратов: Профобразование.
- Добров Б.В., Соловьев В.Д., Иванов В.В., Лукашевич Н.В. 2015. *Онтологии и тезаурусы*. Казань: КГУ.
- Дэвенпорт Т. 2021. *Внедрение искусственного интеллекта в бизнес-практику: Преимущества и сложности*. М.: Альпина Паблишер.
- Загорюлько Ю.А., Загорюлько Г.Б. 2020. *Искусственный интеллект. Инженерия знаний*. М.: Юрайт.
- Захаров В.Н., Попов Э.В., Поспелов Д.А., Хорошевский В.Ф. 1990. *Искусственный интеллект. Справочник в трех томах*. М.: Радио и связь.
- Кобринский Б.А. 2022. Горячие ступени в будущее искусственного интеллекта. Материалы конференции: *Шаг в будущее: искусственный интеллект и цифровая экономика. Технологическое лидерство: взгляд за горизонт*; 120–128.
- Котов К.И., Соколова Д.Д., Павлов Д.С. 2023. Применение нейросетей в инвестиционной деятельности: возможности и риски. Материалы конференции: *Инновационные механизмы управления цифровой и региональной экономикой (МФТИ)*; 384–392.
- Круг П.Г. 2002. *Нейронные сети и нейрокомпьютеры*. М.: МЭИ.
- Люгер Д.Ф. 2003. *Искусственный интеллект: стратегии и методы решения сложных проблем*. М.: Издательский дом «Вильямс».
- Маслов С.Ю. 1964. Обратный метод установления выводимости в классическом исчислении предикатов. *Доклады Академии наук* 159 (1): 17–20.
- Минский М. 1979. *Фреймы для представления знаний*. М.: Энергия.
- Миркин Б.Г., Кузнецов С.О. 2009. *Машинное обучение. Конспект лекций*. М.: ВШЭ.
- Муромцев Д., Волчек Д., Романов А. 2019. Индустриальные графы знаний — интеллектуальное ядро цифровой экономики. *Control Engineering Россия* (5): 32–39.
- Николенко С., Кадуринов А., Архангельская Е. 2018. *Глубокое обучение*. СПб.: Питер.
- Оселедец И.В. 2022. Успехи и проблемы машинного обучения. Материалы конференции: *Проектирование будущего. Проблемы цифровой реальности* 5; 102–108.
- Осипов Г. 2011. *Методы искусственного интеллекта*. М.: Физматлит.
- Поспелов Д.А. 1998. Становление информатики в России. В сб.: *Очерки истории информатики в России* (Новосибирск); 7–44. [Электронный ресурс]. <http://simulation.su/uploads/files/default/1998-pospelov-1.pdf> (дата обращения: 09.09.2024).
- Поспелов Д.А. 1982. *На пути к искусственному интеллекту*. М.: Наука.
- Рассел С., Норвиг С. 2016. *Искусственный интеллект. Современный подход*. СПб.: Диалектика/Вильямс.
- Рождественская Я. Генератор умно звучащего бреда. *Коммерсант*. [Электронный ресурс]. <https://www.kommersant.ru/doc/4457086> (дата обращения: 19.08.2024).
- Рыбина Г.В. 2010. *Основы построения интеллектуальных систем*. М.: Финансы и статистика.
- Скурго. 2024. *Искусственный интеллект в Сбербанке*. [Электронный ресурс]. <https://sky.pro/wiki/python/iskusstvennyj-intellekt-v-sberbanke/> (дата обращения: 19.08.2024).
- Смирнов А.В., Кашевник А.М., Пономарев А.В., Савосин С.В. 2014. Онтологический подход к организации взаимодействия сервисов интеллектуального про-

- странства при управлении гибридными системами. *Искусственный интеллект и принятие решений* (4): 42–51.
- Смирнов С. В. 2013. Онтологии как смысловые модели. *Онтология проектирования* 2 (8): 12–19.
- Стефанюк В. Л. 2011. Творческое решение задач. *Искусственный интеллект и принятие решений* (2): 3–11.
- Финн В. К. 2020. Точная эпистемология и искусственный интеллект. *Научно-техническая информация. Серия 2: Информационные процессы и системы* (6): 1–36.
- Фитисов А. Как база знаний экономит время и силы всех сотрудников. *Бизнес-секреты*. [Электронный ресурс]. https://secrets.tinkoff.ru/blogi-kompanij/zachem-kompaniyam-baza-znaniy/?internal_source=copypaste (дата обращения: 09.08.2024).
- Хант Э. 1978. *Искусственный интеллект*. М.: Мир.
- Хабр. 2024. Искусственные нейросети простыми словами. [Электронный ресурс]. <https://habr.com/ru/articles/369349/> (дата обращения: 19.08.2024).
- Эндрю А. М. 1985. *Искусственный интеллект*. М.: Мир.
- Юрин А. Ю. 2016. Нотация для проектирования баз знаний продукционных экспертных систем. *Объектные системы* (12): 48–54.

REFERENCES IN LATIN ALPHABET

- Adeli H. 1994. *Knowledge Engineering*. McGraw-Hill Publishing Company: New York.
- Ahmadi S. 2024. A comprehensive study on integration of big data and AI in financial industry and its effect on present and future opportunities. *International Journal of Current Science Research and Review* 7 (1): 66–74.
- Arman M., Lamiyar U. R. 2023. Exploring the implication of ChatGPT AI for business: Efficiency and challenges. *International Journal of Marketing and Digital Creative* 1 (2): 64–84.
- Benaich N. Z. & Air Street Capital. 2024. *State of AI Report 2024*. [Electronic resource]. <https://www.stateof.ai/> (accessed: 10.10.2024).
- Bilateral AI. 2024. Cluster of excellence. [Electronic resource]. <https://www.bilateral-ai.net/> (accessed: 10.08.2024).
- Brockman J. 1998. Consciousness is a big suitcase: A talk with Marvin Minsky. *Edge*. [Electronic resource]. https://www.edge.org/conversation/marvin_minsky-consciousness-is-a-big-suitcase (accessed: 10.08.2024).
- Caiming Z., Lu Y. 2021. Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration* 23: 100224.
- Campton B. 2024. *How Is AI Used in Business? 10 Ways It Can Help*. [Electronic resource]. <https://www.upwork.com/resources/how-is-ai-used-in-business> (accessed: 10.08.2024).
- Charette R. 2018. Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold. *IEEE Spectrum*. [Electronic resource]. <https://spectrum.ieee.org/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold> (accessed: 10.08.2024).
- Cohen A. AI is pushing the world toward an energy crisis. *Forbes*. [Electronic resource]. <https://www.forbes.com/sites/arielcohen/2024/05/23/ai-is-pushing-the-world-towards-an-energy-crisis> (accessed: 10.08.2024).
- Duarte F. 2024. *AI Market Size Statistics*. [Electronic resource]. <https://explodingtopics.com/blog/ai-market-size-stats> (accessed: 10.08.2024).
- Evans H., Snead D. 2024. Why do errors arise in artificial intelligence diagnostic tools in histopathology and how can we minimize them? *Histopathology* 84 (2): 279–287.
- Feigenbaum E. A. 1981. Expert systems in the 1980s. In: Bond A. (Ed). *Infotech State of the Art Report on Machine Intelligence*. Pergamon Infotech Ltd: Maidenhead; 27–52.
- Feuerriegel S., Hartmann J., Janiesch C., Zschoch P. 2024. Generative ai. *Business & Information Systems Engineering* 66 (1): 111–126.

- Flasiński M., Flasiński M. 2016. History of artificial intelligence. *Introduction to Artificial Intelligence*: 3–13.
- Floridi L., Cows J., Beltrametti M., Chatila R., Chazerand P. et al. 2018. AI4People — An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines* **28** (4): 689–707.
- Floridi L., Chiriatti M. 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines* **30** (4): 681–694.
- Gruber T. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition* **5** (2): 199–220.
- Harmelen F.V., Antoniou G. 2004. Web ontology language: Owl. In: *Handbook on Ontologies*. Springer: Berlin, Heidelberg; 67–92.
- Heist N., Hertling S., Ringler D., Paulheim H. 2020. Knowledge graphs on the web — an overview. *Artificial Intelligence: Foundations, Applications and Challenges*: 3–22. <http://dx.doi.org/10.48550/arXiv.2003.00719>
- Ji Z., Lee N., Frieske R., Yu T., Su D. et al. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys* **55** (12): 1–38.
- Joseph E. *A comparison of GPT 4, Llama, Claude, AI Studio & Perplexity*. [Electronic resource]. <https://www.pixelcrayons.com/blog/digital-transformation/comparing-gpt4-llama-claude-ai-studio-perplexity/> (accessed: 10.08.2024).
- Kaggwa S., Eleogu T.F., Okonkwo F., Farayola O.A., Uwaoma P.U., Akinoso A. 2024. AI in decision making: Transforming business strategies. *International Journal of Research and Scientific Innovation* **10** (12): 423–444.
- Kawakami R., Venkatagiri S. 2024. The Impact of Generative AI on Artists. In proceedings: *Conference on Creativity & Cognition* **16**: 79–82.
- Knuth D.E. 1997. *The art of computer programming* (Vol. 1–4). Pearson Education.
- Kosenko D.P., Kuratov Y.M., Zharikova D.R. 2023. Accessible Russian large language models: Open-source models and instructive datasets for commercial applications. In: *Doklady Mathematics* **108** (S2); S393–S398. Pleiades Publishing: Moscow.
- Krishnan C., Mariappan J. 2024. The AI revolution in e-commerce: Personalization and predictive analytics. In: *Role of Explainable Artificial Intelligence in E-Commerce*. Springer Nature Switzerland: Cham; 53–64.
- Lamarre E., Smaje K., Zimmel R. 2023. *Rewired: The McKinsey guide to outcompeting in the age of digital and AI*. John Wiley & Sons.
- McKinsey & Company. 2024a. *Artificial intelligence in strategy*. [Electronic resource]. <https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/artificial-intelligence-in-strategy> (accessed: 10.08.2024).
- McKinsey & Company. 2024b. *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. [Electronic resource]. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai> (accessed: 10.08.2024).
- Miikkulainen R., Liang J., Meyerson E., Rawal A., Fink D. et al. 2019. Evolving deep neural networks. In: *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Academic Press; 293–312.
- Mirjalili S. 2019. Evolutionary algorithms and neural networks. In: *Studies in Computational Intelligence* **780**. Springer: Berlin/Heidelberg.
- Moto-Oka T. (Ed.). 2012. *Fifth Generation Computer Systems*. Elsevier.
- Newell A., Shaw J.C., Simon H.A. 1958. Elements of a theory of human problem solving. *Psychological Review* **65** (3): 157–166.
- Newell A., Simon H. 1976. Computer science as empirical inquiry: Symbols and search. In: *Communications of the ACM* **19** (3): 113–126.
- Oppenlaender J. 2022. *The creativity of text-based generative art*. <https://doi.org/10.48550/arXiv.2206.02904>
- Peng C., Xia F., Naseriparsa M., Osborne F. 2023. Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review* **56** (11): 13071–13102.
- Pujara J., Miao H., Getoor L., Cohen W. 2013. Knowledge graph identification. In proceedings: *12th International Semantic Web Con-*

- ference (Sydney, NSW). Berlin, Heidelberg: Springer; 542–557.
- Raji I., Kumar E., Horowitz A., Selbst A. 2022. The fallacy of AI functionality. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*; 959–972.
- Razzaq A. *Llama 3.1 vs GPT-4o vs Claude 3.5: A comprehensive comparison of leading AI models*. [Electronic resource]. <https://www.marktechpost.com/2024/07/27/llama-3-1-vs-gpt-4o-vs-claude-3-5-a-comprehensive-comparison-of-leading-ai-models/> (accessed: 09.09.2024).
- Robinson J. 1965. A machine-oriented logic based on the resolution principle. *Journal of the ACM* **12** (1): 23–41.
- Sadiku M., Fagbohunge O., Musa S. 2020. Artificial intelligence in business. *International Journal of Engineering Research and Advanced Technology* **6** (7). <https://doi.org/10.31695/IJERAT.2020.3625>
- Sure Y., Staab S., Studer R. 2009. Ontology engineering methodology. In: *Handbook on Ontologies*. Springer: Berlin, Heidelberg; 135–152.
- TechInsights. 2024. [Electronic resource]. <https://clk.ru/3DpzVx> (accessed: 09.09.2024).
- Xie Y., Xu Z., Zhang J., Wang Z., Ji S. 2022. Self-supervised learning of graph neural networks: A unified review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45** (2): 2412–2429.
- Wraith S.M., Aikins J.S., Buchanan B.G., Clancey W.J., Davis R. et al. 1976. Computerized consultation system for selection of antimicrobial therapy. *American Journal of Hospital Pharmacy* **33** (12): 1304–1308.
- Zhang C., Lu Y. 2021. Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration* **23**: 100224.

TRANSLATION OF REFERENCES IN RUSSIAN INTO ENGLISH

- Aleksandrov E.A. 1975. *Fundamentals of the Theory of Heuristic Decisions*. Soviet Radio Publ.: Moscow. (In Russian)
- Bashlykov A., Ereemeev A.P. 1994. *Expert Systems for Decision Support in Power Engineering*. MEI Publ.: Moscow. (In Russian)
- Bova V.V., Kravchenko D.Y., Leshchanov D.V., Novikov A.A. 2014. Computational ontology: Objectives and methodology of construction. *Informatics, Computer Science, and Engineering Education* (4). (In Russian)
- Borgest N.M. 2013. Key terms of design ontology: Review, analysis, generalizations. *Ontology of designing* **3** (9): 9–31. (In Russian)
- Bhatt Sh., Zhao J., Seth A., Shalin W. 2020. Knowledge graphs as a means of improving artificial intelligence. *Open Systems. DBMS* (3): 24–26. (In Russian)
- Wiener N. 1968. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Soviet Radio Publ: Moscow. (In Russian)
- Gavrilova T.A. 1984. Knowledge representation in the AVTANTEST expert diagnostic system. *Bulletin of the USSR Academy of Sciences. Technical Cybernetics* (5): 168–175. (In Russian)
- Gavrilova T.A., Khoroshevsky V.F. 2000. *Knowledge Bases of Intelligent Systems. Textbook*. Piter Publ. St. Petersburg. (In Russian)
- Gavrilova T.A. 2005. On one approach to ontological engineering. *News of artificial intelligence* (3): 25–31. (In Russian)
- Gavrilova T.A., Kudryavtsev D.V., Muromtsev D.I. 2016. *Knowledge Engineering. Models and Methods*. Lan Publ.: St. Petersburg. (In Russian)
- Gerasimenko V.V. Digital ethics of Artificial Intelligence in Business: Awareness of New Opportunities and Risks. *Scientific research of the Faculty of Economics* **15** (1): 37–54. (In Russian)
- Gorban A.N. 2000. Neuroinformatics: Who we are, where we are going, how to measure our path. *Information technologies* (4): 10–14. (In Russian)
- Jones M.T. 2019. *Artificial Intelligence Application Programming*. Vocational education Publ.: Saratov. (In Russian)

- Dobrov B.V., Soloviev V.D., Ivanov V.V., Lukashovich N.V. 2015. *Ontologies and Thesauri*. KSU Publ.: Kazan. (In Russian)
- Davenport T. 2021. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. Alpina Publisher: Moscow. (In Russian)
- Zagorulko Y.A., Zagorulko G.B. 2020. *Artificial intelligence. Knowledge engineering*. Urait Publ.: Moscow. (In Russian)
- Zakharov V.N., Popov E.V., Pospelov D.A., Khoroshevsky V.F. 1990. *Artificial Intelligence. Handbook in three volumes*. Radio and Communications Publ.: Moscow. (In Russian)
- Kobrinsky B.A. 2022. Hot Steps to the Future of Artificial Intelligence. In proceedings: *Step into the Future: Artificial Intelligence and the Digital Economy. Technological Leadership: A Look Beyond the Horizon*: 120–128. (In Russian)
- Kotov K.I., Sokolova D.D., Pavlov D.S. 2023. Application of neural networks in investment activities: opportunities and risks. In proceedings: *Innovative Mechanisms for Managing the Digital and Regional Economy (MIPT)*; 384–392. (In Russian)
- Krug P.G. 2002. *Neural networks and neurocomputers*. MEI Publ.: Moscow. (In Russian)
- Luger D.F. 2003. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Williams Publ.: Moscow. (In Russian)
- Maslov S. Yu. 1964. The inverse method for establishing derivability in classical predicate calculus. *Reports of the Academy of Sciences* **159** (1): 17–20. (In Russian)
- Minsky M. 1979. *A Framework For Representing Knowledge*. Energy Publ.: Moscow. (In Russian)
- Mirkin B.G., Kuznetsov S.O. 2009. *Machine learning. Lecture notes*. HSE Publ.: Moscow. (In Russian)
- Muromtsev D., Volchek D., Romanov A. 2019. Industrial knowledge graphs are the intellectual core of the digital economy. *Control Engineering Russia* (5): 32–39. (In Russian)
- Nikolenko S., Kadurin A., Arkhangelskaya E. 2018. *Deep Learning*. Piter Publ.: St. Petersburg. (In Russian)
- Oseledets I.V. 2022. Achievements and Problems of Machine Learning. In proceedings: *Designing the Future. Problems of Digital Reality* **5**; 102–108. Osipov G. 2022. Methods of artificial intelligence. (In Russian)
- Osipov G. 2011. *Methods of artificial intelligence*. Fizmatlit Publ.: Moscow.
- Pospelov D.A. 1998. Formation of computer science in Russia. In the collection: *Essays on the history of computer science in Russia* (Novosibirsk); 7–44. [Electronic resource]. <http://simulation.su/uploads/files/default/1998-pospelov-1.pdf> (accessed: 09.09.2024). (In Russian)
- Pospelov D.A. 1982. *On the Way to Artificial Intelligence*. Science Publ.: Moscow. (In Russian)
- Russel S., Norvig S. 2016. *Artificial Intelligence. A Modern Approach*. Dialectics/Williams Publ.: St. Petersburg. (In Russian)
- Rozhdestvenskaya Ya. Generator of smart-sounding nonsense. *Kommersant*. [Electronic resource]. (<https://www.kommersant.ru/doc/4457086>) (accessed: 19.08.2024). (In Russian)
- Rybina G.V. 2010. *Fundamentals of Building Intelligent Systems*. Finance and statistics Publ.: Moscow. (In Russian)
- Skypro. 2024. *Artificial Intelligence in Sberbank*. [Electronic resource]. <https://sky.pro/wiki/python/iskusstvennyj-intellekt-v-sberbanke/> (accessed: 19.08.2024). (In Russian)
- Smirnov A.V., Kashevnik A.M., Ponomarev A.V., Savosin S.V. 2014. Organization of smart space services interaction for hybrid systems control. *Scientific and Technical Information Processing* (4): 42–51. (In Russian)
- Smirnov S.V. 2013. Ontologies as semantic models. “*Ontology of Designing*” *Scientific Journal* **2** (8): 12–19. (In Russian)
- Stefanuk V.L. 2011. Creative problem solving. *Scientific and Technical Information Processing* (2): 3–11. (In Russian)
- Finn V.K. 2020. Precise Epistemology and Artificial Intelligence. *Scientific and Technical Information. Series 2: Information Processes and Systems* (6): 1–36. (In Russian)
- Fitisov A. How a knowledge base saves time and effort for all employees. *Business secrets*. [Electronic resource]. <https://secrets.tinkoff>.

- ru/blogi-kompanij/zachem-kompaniyam-baza-znaniy/?internal_source=copypaste (accessed: 09.08.2024). (In Russian)
- Hunt E. 1978. *Artificial Intelligence*. Mir Publ.: Moscow. (In Russian)
- Habr. 2024. Artificial neural networks in simple terms. [Electronic resource]. <https://habr.com/ru/articles/369349/> (accessed: 19.08.2024). (In Russian)
- Andrew A.M. 1985. *Artificial Intelligence*. Mir Publ.: Moscow. (In Russian)
- Yurin A. Yu. 2016. Notation for designing knowledge bases of production expert systems. *Object Systems* (12): 48–54. (In Russian)

*Статья поступила в редакцию
13 июля 2024 г.*

*Принята к публикации
11 сентября 2024 г.*

“Stochastic parrot” as a servant of business: Achievements and challenges of generative artificial intelligence

T. A. Gavrilova

Graduate School of Management, St. Petersburg State University, Russia

Purpose: to present an overview of research and technologies of artificial intelligence based on neural networks and large linguistic models. **Methodology:** the key method for the study is the desk method of thematic literature reviews. **Findings:** the main trends in the development of artificial intelligence were identified, the hidden dangers and the imperfections of modern systems implementation were analyzed and determined. **Originality and contribution:** the study shows a critical approach and the analysis of the ever-increasing information noise in the field of promoting generative artificial intelligence as the result of the author’s supervision and implementation of research projects initiated by the Russian Foundation for Basic Research and the Russian Science Foundation, as well as the continuous expert work in the Russian Association of Artificial Intelligence. The paper integrates several plenary reports of the author on this topic.

Keywords: artificial intelligence, knowledge engineering, artificial neural networks, business applications of artificial intelligence.

For citation: Gavrilova T.A. 2024. “Stochastic parrot” as the servant of business: Achievements and challenges of generative artificial intelligence. *Russian Management Journal* **22** (3): 461–482. <https://doi.org/10.21638/spbu18.2024.305> (In Russian)

Для цитирования: Гаврилова Т. А. 2024. «Стохастический попугай» на службе бизнеса: успехи и проблемы генеративного искусственного интеллекта. *Российский журнал менеджмента* **22** (3): 461–482. <https://doi.org/10.21638/spbu18.2024.305>

Initial Submission: July 13, 2024

Final Version Accepted: September 11, 2024