ИПМ им.М.В.Келдыша РАН

Будущее-2024 • Труды конференции


Д.В.Журавлёв, В.С.Смолин

# От чуда нейронных сетей – к идеям для AGI

***Рекомендуемая форма библиографической ссылки***

Журавлёв Д.В., Смолин В.С. От чуда нейронных сетей – к идеям для AGI // Проектирование будущего. Проблемы цифровой реальности: труды 7-й Международной конференции (15-17 февраля 2024 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2024. — С. 96-124. — https://keldysh.ru/future/2024/2-2.pdf https://doi.org/10.20948/future-2024-2-2

***Размещено также видео выступления***

# От чуда нейронных сетей – к идеям для AGI

**Д.В. Журавлёв[1], В.С. Смолин[2]**

*[1]ООО ЦИФРОМЕД*
*[2]Институт прикладной математики им. М.В. Келдыша РАН*

**Аннотация.** Показаны семь причин превосходства нейросетевых алгоритмов над эвристическими подходами. Утверждается, что чудесными такие свойства не являются, так как они обеспечивают решения ограниченного круга задач, в отличие от AGI, который направлен на решение существенно более сложных задач и на нахождение новых знаний без участия человека. Обоснован вывод о том, что не следует ожидать возникновения сознания в процессе простого развития систем, так как создание AGI возможно только при достижении понимания сложного мира.

**Ключевые слова:** нейросетевые алгоритмы, AI, AGI

# From the neural networks miracle – to AGI

**D.V. Zhuravlev[1], V.S. Smolin[2]**

*[1]CIFROMED LLC*
*[2]RAS Keldysh Institute of Applied Mathematics*

**Abstract.** Seven reasons for the superiority of neural network algorithms over heuristic approaches are shown. It is argued that such properties are not miraculous, as they provide solutions to a limited range of problems, unlike AGI, which is aimed at solving significantly more complex problems and finding new knowledge without human participation. The conclusion is justified that we should not expect the emergence of consciousness in the process of simple development of systems, since the creation of AGI is possible only when the understanding of the complex world is achieved.

**Keywords:** neural network algorithms, decomposition, AGI

## 1. Введение

Нынешняя нейросетевая революция в машинном обучении началась примерно в 2010-12 гг. и сегодня она оказывает все большее влияние на развитие ИИ (искусственного интеллекта). За прошедшие 12-14 лет были предложены и новые архитектуры нейронных сетей, и алгоритмы, улучшающие процесс обучения, повышающие как скорость и точность, так и

расширяющие область применения на более широкий класс задач. Среди наиболее значимых шагов в развитии можно выделить следующие:

1. 2010 г. Увеличение «глубины» (числа скрытых слоев) нейросетей сделало их применение коммерческим и успешным в состязаниях (например, [1]), позволяя решать задачи распознавания речи и изображений;
2. 2013 г. Представлена схема GANs (генеративно-состязательные нейронные сети) [2], обучающая нейронные сети создавать фотореалистичные и другие виды сигналов без прямого участия человека;
3. 2016 г. Методы обучения с подкреплением, такие как DQN (Deep Q networks и другие), убедительно проявили себя в контексте победы программ АльфаГо и АльфаЗеро над чемпионами мира и лучшими компьютерными программами в го, шахматы и сёги [3];
4. 2020 г. Архитектура GPT (Generative Pre-trained Transformer) стала широко известной благодаря GPT-3, третьей модели нейросетевых алгоритмов от OpenAI [4] для обработки естественного языка, хотя первые «предобученные трансформеры» были представлены еще в 2017-18 гг.

В кратчайшее время GPT стал ключевой технологией для создания больших языковых моделей (LLM) – уже создано несколько десятков таких моделей крупными корпорациями. GPT также стал основой для «фундаментальных моделей» (foundation models [5]), которые обучаются на обширных данных и применяются для решения разнообразных задач. Термин «фундаментальные модели» был придуман и популяризирован в Центре исследований фундаментальных моделей (CRFM) Стэнфордского института человеко-ориентированного искусственного интеллекта (HAI) [6].

Вопреки пессимистичным прогнозам, утверждающим, что нейросетевые алгоритмы – это «хайп», который скоро затихнет, нейросети с каждым годом демонстрируют всё более высокую скорость прогресса. И наиболее оптимистичные прогнозы состоят в том, что уже следующей ступенью их развития станет создание AGI.

## 2. Взлет нейросетей – семь причин

### 2.1. «Закон необходимого разнообразия» Эшби

Еще в 1947 г. Эшби сформулировал [7] «закон необходимого разнообразия», подчеркивая важность разнообразия в системе для решения сложных задач. Вместо этого многие ищут универсальные алгоритмы, способные решать сложные задачи с минимальной настройкой параметров. Нейросетевые алгоритмы, впервые появившиеся примерно в то же время, уже обладали значительно большим числом параметров и, по мнению Эшби, были более подходящими для сложных задач. Однако только с 2010 г. нейросетевые алгоритмы начали существенно превосходить эвристические алгоритмы в решении практических задач.

«Закон необходимого разнообразия» Эшби является основной причиной успеха нейросетей. Технические ограничения в начальные годы раз-

вития нейросетевых алгоритмов (рис. 1) не позволяли полностью использовать их потенциал. Сегодня успех нейросетей очевиден, но не все проблемы могут быть решены простым увеличением числа параметров.

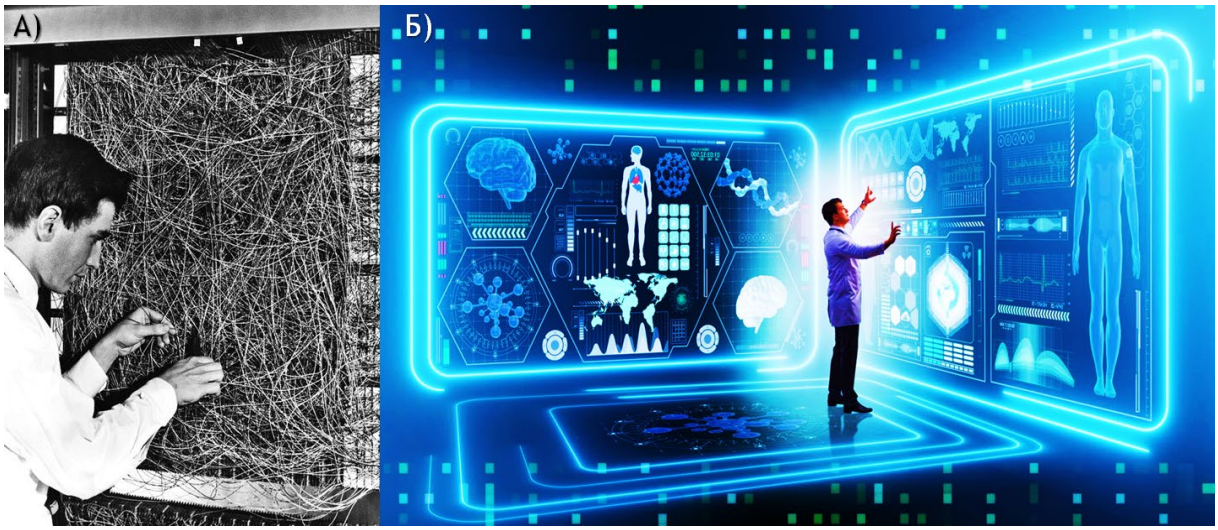


Рис. 1. А) Тонкая настройка персептрона в 1960 г. Б) Нейросетевые алгоритмы формируют преобразования, описываемые миллионами и миллиардами параметров (GPT-3 – 175 млрд параметров – 2021)

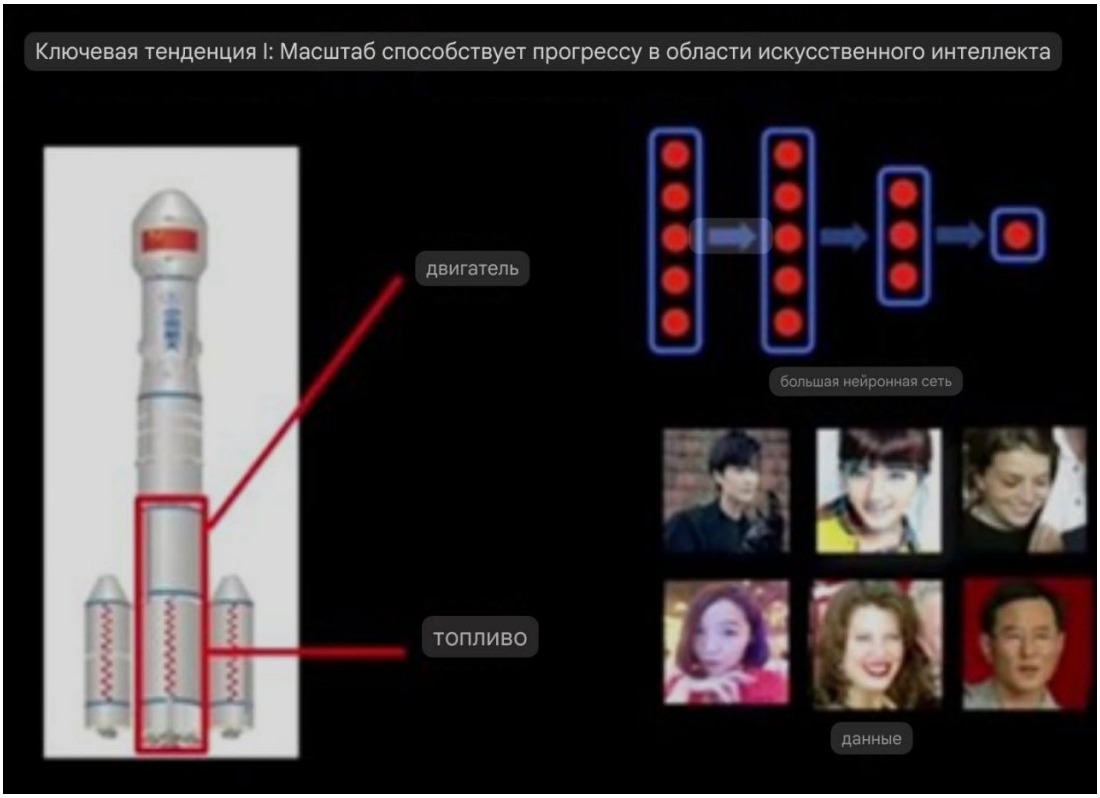


Рис. 2. Аналогия «взлета» нейросетей с ракетой: мощный двигатель и большие топливные баки с развитыми технологиями в ракетостроении сравнимы с мощными вычислителями, большими данными и развитыми алгоритмами в нейросетях

Таким образом, с 2012 г. во многом успехи нейронных сетей связаны с увеличением числа параметров, слоев и объемов данных для обучения. Эндрю Ын (Andrew Yan-Tak Ng), сооснователь Google Brain и бывший главный научный сотрудник компании Baidu, в своих лекциях 2015 г. [8] выделял три основные причины успеха нейросетей, две из которых связаны с количественным ростом. Причины, о которых говорил Эндрю Ын, обозначены как 2.2, 2.3 и 2.4 в нашем списке.

2.2. Рост производительности компьютеров

Закон роста вычислительной техники [9], названный «законом Мура», предполагают удвоение количества транзисторов на кристалле интегральной схемы каждые 24 месяца. Другой прогноз от Давида Хауса говорит об удвоении производительности процессоров каждые 18 месяцев за счет роста числа транзисторов и тактовой частоты. С 2003 г. тактовая частота перестала расти, и производительность стала увеличиваться за счет параллельных вычислений. Как видно из графика на рис. 3 [9], до 2015 г. производительность увеличивалась в 2 раза за 1,3 года. После 2015 г. темпы роста немного снизились, но экспоненциальный характер сохранялся.

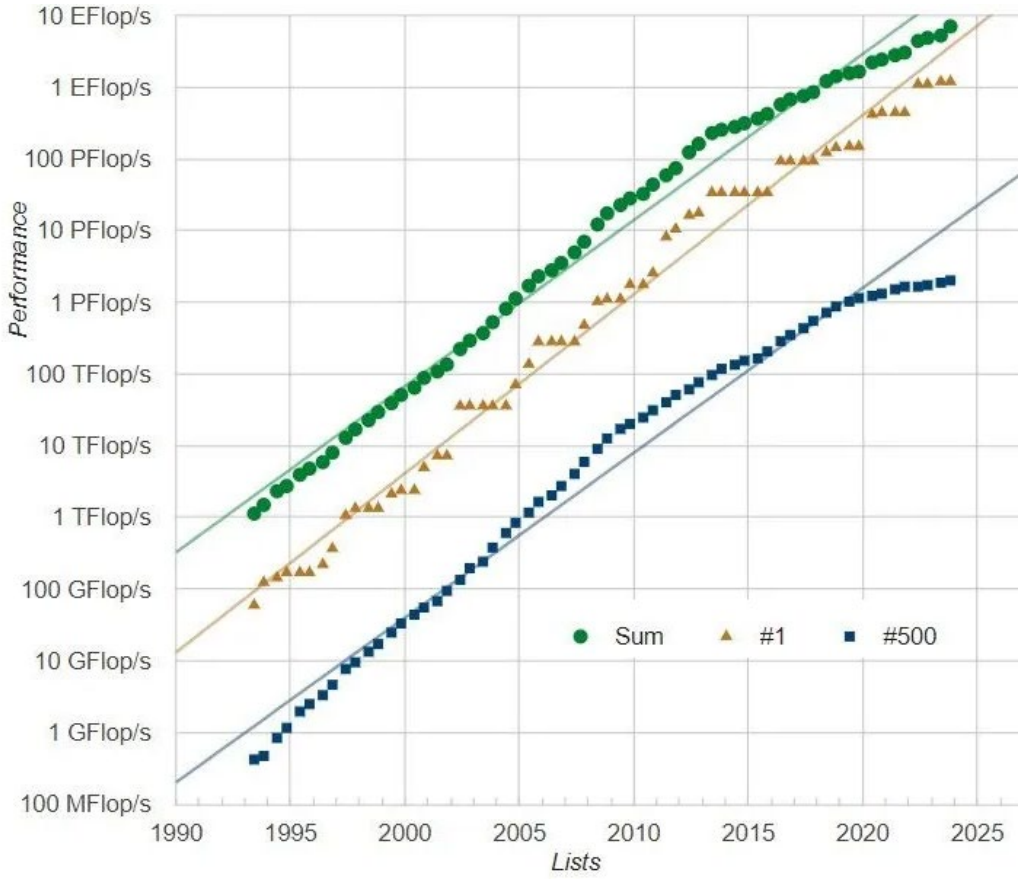


Рис. 3. Графики роста суммарной, наивысшей (#1 из списка) и последнего (#500 из списка) производительностей суперкомпьютеров, входивших в списки TOP500 в 1994-2024 гг. [9]

С точки зрения закона Мура в 2010 г. производительность суперкомпьютеров выросла, а стоимость снизилась благодаря графическим процессорам и HPC (high performance computing), что сделало их доступными для исследований, включая нейросетевые алгоритмы. Популярность нейронных сетей повлияла на производство микросхем и плат, графические ускорители стали оптимизировать для нейросетевых алгоритмов.

Некоторые подозревают производителей вычислительной техники в стимулировании развития вычислительно затратных нейросетевых алгоритмов, позволяющем поддерживать спрос на сложную технику. Эффективность нейровычислений можно улучшить, но разнообразие параметров требуется для решения сложных задач, и спрос на мощные вычислители не исчезнет.

США, Китай, Германия, Япония и Франция лидируют в развитии ИИ, инвестируя значительные средства в суперкомпьютеры. Эти страны занимают 356 из 500 мест в списке TOP-500 самых мощных компьютеров. У России 7 суперкомпьютеров, включая «Червоненкис» Яндекса, отстающий от лидера более чем в 20 раз по производительности.

2.3. Большие объемы данных

Для обучения нейросетевых моделей, которые содержат до $10^{12}$ параметров, необходимо иметь большие объемы данных. С развитием информационных технологий происходит увеличение объема данных, а также возрастание мощности вычислительных систем для их обработки.

Нейросетевые алгоритмы, хотя и играют ключевую роль, все же являются лишь частью машинного обучения (МО, или machine learning, ML). Основной особенностью машинного обучения является возможность настройки параметров алгоритма в ходе обучения под реальные условия выполнения задачи. Обучение сводится к уменьшению ошибки преобразования входных сигналов в выходные через настройку параметров.

Нейросетевые структуры обычно используют простые нелинейные функции, но благодаря возможности комбинирования их весов, предоставляют широкие возможности для аппроксимации сложных нелинейных преобразований. Но чем сложнее преобразование, тем больше данных требуется для хорошей аппроксимации методами машинного обучения.

Высокая размерность и нелинейность преобразований представляют основную сложность для аппроксимации. Нейросети являются лучшим инструментом для построения многомерных аппроксимаций преобразований. С ростом размерности требуется использовать все больше примеров «правильного» преобразования.

Развитие Интернета облегчают сбор и накопление обучающих данных. Необходимо структурировать данные правильно для обучения нейросетей. Разметка данных может производиться как вручную, так и с использованием автоматизированных способов (рис. 4). Обработка сложных сигналов, таких как фотографии, видео и аудио, требует более тонкой

разметки данных. Задачи сегментации, детекции, классификации и другие виды разметки используются для сложных сигналов.

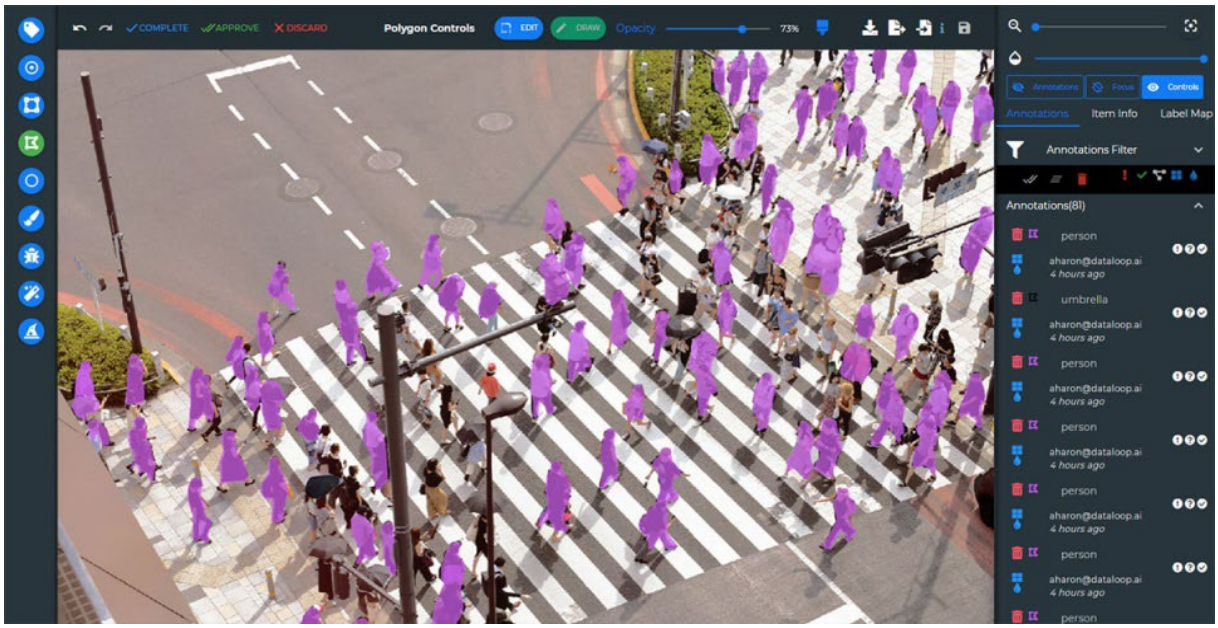


Рис. 4. Пример разметки фотографии платформой Dataloop – каждый пешеход выделен цветом

Разметка данных может использоваться для обучения нейросетей различным задачам. Процесс разметки данных может быть автоматизирован на различных этапах, начиная с обучения на данных, размеченных людьми, и заканчивая самостоятельной автоматической разметкой данных мощными нейросетями без участия человека.

Существует много программных платформ для разметки сложных сигналов, особенно изображений и видео ([10], но есть и другие). Однако с ростом применения нейросетей в различных областях становится ясно, что имеющихся открытых датасетов недостаточно. Для обучения нейросетей требуется обширные и разнообразные наборы данных, что иногда может быть проблемой.

В случае ограниченного доступа к достаточному количеству обучающих примеров и тогда, когда разметка их требует значительных усилий, возникает необходимость применения методов *аугментации* (генерации дополнительных наборов данных на основе имеющихся) данных. Этот подход позволяет создать дополнительные варианты обучающих данных путем их модификации, что способствует увеличению разнообразия данных для обучения нейронной сети и улучшает качество модели.

Таким образом, использование программных платформ для разметки данных и методов аугментации становится важным аспектом в области обучения нейросетей, особенно когда обучающие данные ограничены или требуют особых усилий для разметки (рис. 5).

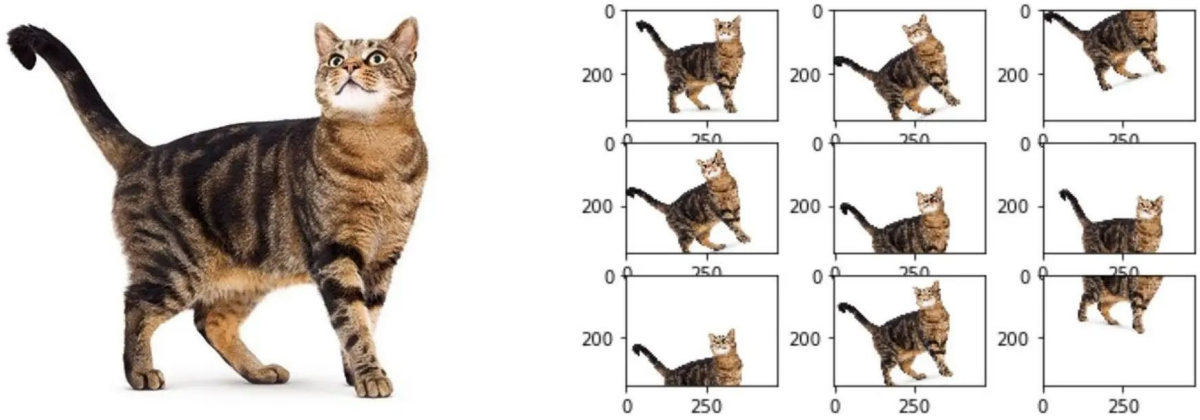

Рис. 5. Пример аугментации данных смещением и поворотом

При использовании фото- и видеосъемки изображения могут иметь различные характеристики, такие как масштаб, поворот, положение на экране и условия освещения. Не обязательно создавать множество снимков с различными параметрами – можно получить тысячи и миллионы вариаций из небольшого набора данных путем аугментации их исходных примеров. Этот подход увеличивает разнообразие обучающих данных и улучшает производительность модели, что является распространенной методикой в области обучения нейронных сетей.

Аугментация данных для улучшения качества обучения моделей стала общепринятой практикой, включая различные методы, такие как геометрические преобразования, изменение цвета и добавление шума. При наличии векторных описаний объектов для обучения нейросетей возможности аугментации значительно расширяются. Дополнительно, создание изображений с использованием графики или нейросетей может быть более затратным, но оно предоставляет возможность получить обширный и разнообразный набор данных для обучения. Применение аугментации дает возможность не только увеличить объем обучающих данных, но и достичь баланса между разными классами данных. Это важно, если для разных классов данные имеют различную представленность.

При использовании данных и защите персональной информации разработчикам приходится решать организационные и законодательные проблемы. Законы обязывают их находить баланс между доступом к данным и защитой конфиденциальности при сборе данных и их использовании.

2.4. Развитие алгоритмов глубокого обучения, приведшее к их коммерческому успеху

Каждая из семи описываемых в п.2 причин «взлета» нейросетей важна, но содержательную основу (визитную карточку) успеха составляет именно развитие нейросетевых алгоритмов, которые можно назвать «Глубоким обучением» [11; 12]. Этот подход заключается в том, что чем больше слоев имеет нейронная структура, тем лучше она способна решать сложные задачи. Основой «глубокого обучения» является методика анти-

градиентного спуска, который позволяет эффективно настраивать параметры нейросетей, уменьшая необходимость повторного прохода через все обучающие данные на каждом шаге обучения. Метод не всегда гарантирует попадание в глобальный минимум, а вероятность его достижения зависит от различных факторов, таких как структура нейросети, начальные значения параметров и методы регуляризации. Пионерскими работами в этой области являются формальные нейроны МакКаллока и Питса [13], а также персептроны Розенблатта [14]. Построенный в 1958 г. персептрон Розенблатта работал именно на таких нейронах МакКаллока и Питса.

Это было связано с ограниченными возможностями цифровой техники того времени и, отчасти, с тем, что преобладали представления (от которых даже сейчас не все могут отказаться), что работа мозга – это выполнение логических преобразований, некоторая нетривиальная модель машины Тьюринга [15]. Тем не менее уже персептрон Розенблатта строился не по схеме машины Тьюринга, в которой изменения любого знака текста программы может привести к утрате ее работоспособности, а на идее устойчивого приближения к выполнению требуемого преобразования (относительно) плавным изменением.

Это было революционным изменением представлений о характере работы нейросетей. С развитием вычислительной техники это позволило Уидроу и Хоффу [16] заменить дискретную модель нейрона МакКаллока и Питса на непрерывную, используя монотонную нелинейную функцию. Значения весов связей были представлены действительными числами.

Идеи персептрона Розенблатта важны, но развитие нейросетей ушло далеко от них. При этом современные нейросети унаследовали принцип постепенного изменения параметров. После книги Минского и Пайперта «Персептроны» [17] появились ложные истолкования этой идеи, что привело к заморозке развития нейросетей в период с 1969 до конца 1970-х гг.

В 1980-е гг. вновь наметился подъем исследований нейросетевых алгоритмов и было предложено много интересных моделей, из числа которых принято выделять работы [18-20], а хороший обзор исследований по нейросетям проводившихся в 1960-90 гг. содержится, например, в [21].

В начале 1990-х гг. началась вторая «зима» нейросетевого ИИ, обусловленная тем, что авторы перспективных подходов (включая «глубокое» обучение) давали широковещательные прогнозы о скором достижении феноменальных результатов, но коммерческого успеха с использованием нейросетевых алгоритмов в те годы никто достичь не смог.

Вторая «зима» нейросетевого ИИ была более продолжительной, чем первая (более 15 лет), тем не менее, работы в данной области, пусть с меньшей интенсивностью, чем в 1980-е гг., но продолжались. И их развитие привело к началу в 2010-12 гг. нейросетевой революции в машинном обучении (МО), что коротко обозначают как «взлет» нейросетей.

В последнее десятилетие нейросети начали превосходить эвристические алгоритмы в решении «интеллектуальных» задач, что и запустило нейросетевую революцию. Этот всплеск интереса стал заметен с успехами в распознавании речи в 2010 году и победой нейросети AlexNet на конкурсе ImageNet в 2012 г. Коммерческий успех способствовал росту производства и разработки нейросетевых систем, приводя к появлению отдельной отрасли исследований и разработок в области нейросетевого искусственного интеллекта. Выдающиеся ученые, такие как Ян ЛеКун, Йошуа Бенджио и Джеффри Хинтон [22], признаны «крестными отцами» нейросетевого ИИ и играют важную роль в его развитии.

Несмотря на разнообразие направлений развития нейросетевых алгоритмов основные идеи глубокого обучения объединяют их в экосистему, где каждый компонент имеет значение. Вместе с тем специализация ведет к углублению интереса к определенным направлениям развития нейросетей, что отражается в публикациях с акцентом на конкретные алгоритмы.

Поступательные шаги развития нейросетевой революции в МО привели к созданию ряда популярных моделей, таких как Midjourney, DALL-E или Stable Diffusion, а теперь и Sora [23], и множества их конкурентов, число которых ежедневно увеличивается. А ставший доступным всему миру в ноябре 2022 г. сервис ChatGPT на основе GPT-3 многие склонны рассматривать как начало четвертого этапа нейросетевой революции. При этом первые 3 этапа периодизации соотносят с тремя «веснами» нейросетевого ИИ, описанными выше, между которыми наступали «зимы». То, что четвертый этап нейросетевой революции наступил без каких-либо намеков на «зиму» между этапами, дает отрицательный ответ на вопрос: «Будет ли третья зима нейросетевого ИИ?». Нейросетевые алгоритмы продолжат увеличивать свой вклад в структуру систем ИИ и послужат основой для создания AGI. Продолжится углубление специализации. Уже сейчас книги по ИИ охватывают не целые направления, а отдельные аспекты реализации некоторых моделей одного из направлений [24; 25].

Мы склонны обратить внимание на качественное отличие третьей «весны» нейросетевого ИИ от первых двух: с началом третьей «весны» сформировалась положительная обратная связь между наукой, промышленностью и коммерцией, которая обеспечила «взлет» нейросетей. Именно этот альянс и составляет основу нейросетевой революции в МО, которая началась, по нашему мнению, в 2010-12 гг. Если не случится глобальных катастроф, тенденция к ускоряющемуся развитию нейросетевого ИИ в ближайшие годы сохранится. И приведет к созданию AGI, что станет не только революцией в ИИ, но и в развитии цивилизации в целом.

2.5. Широкое распространение обмена открытым кодом нейросетевых алгоритмов через GitHub

При том, что нейросетевые алгоритмы кардинально отличаются от эвристических тем, что позволяют плавно, антиградиентным спуском, улуч-

шать точность аппроксимации целевого преобразования, программирование нейросетевых алгоритмов осуществляется стандартными методами, а их выполнение производится на универсальной цифровой вычислительной технике. Хотя сами нейросетевые алгоритмы достаточно просты, для создания системы, решающей сложную «интеллектуальную» задачу, требуется использование десятков, иногда сотен свойств и методов, описывающих программные реализации функций обучения и работы нейросетей.

Создание устойчиво работающего программного комплекса, пусть и для решения узкого класса задач, если все функции писать «с нуля» требует значительных усилий, в первую очередь, на отладку написанных, но еще не проверенных подпрограмм. Современные учебники описывают многие десятки алгоритмов, обеспечивающих выполнение преобразований, осуществление обучения и сохранение устойчивости этих процессов. Все эти алгоритмы потому и попали в учебники, что уже были кем-то программно реализованы и показали свою эффективность при решении некоторых задач. Уже были написаны и отлажены подпрограммы и даже существуют программные комплексы, которые с использованием этих подпрограмм решают «интеллектуальные» задачи.

Важно не то, что эти подпрограммы в принципе где-то есть, а то, что их открытыми кодами организован широкий общедоступный обмен. В подавляющем большинстве случаев для этого сейчас используется платформа для разработчиков GitHub [26], позволяющая создавать, хранить, управлять и делиться своим кодом. Сайт платформы стал доступен в сети в 2008 г. и управляется компанией GitHub. С 2018 г. она является дочерней компанией Microsoft и в 2023 г. в ней работало 5,5 тыс. чел. В январе 2023 г. GitHub сообщал о наличии более 100 млн разработчиков и более 420 млн репозиториев, включая не менее 28 млн общедоступных репозиториев. Это крупнейший в мире хостинг программного кода, который используется во всех областях и не только теми, кто связан с созданием нейросетевого ИИ.

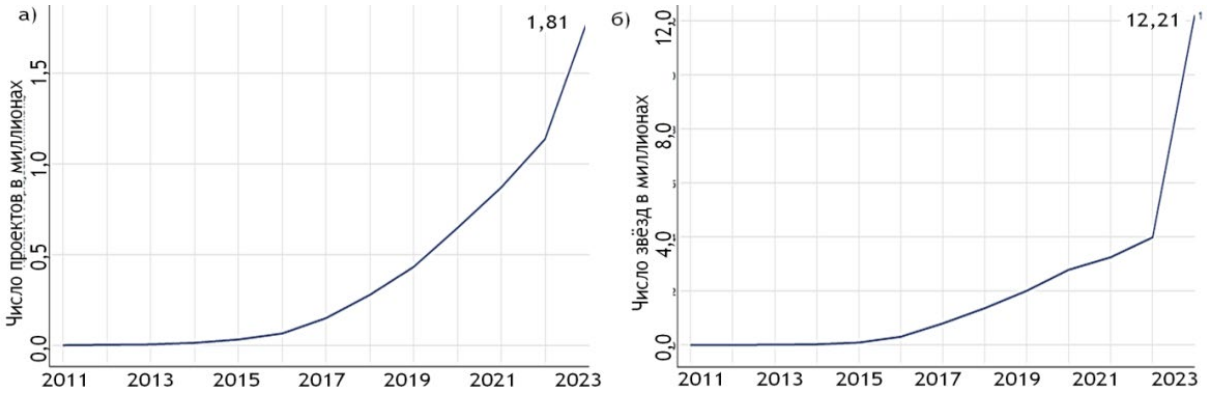


Рис. 6. Рост на GitHub за 2011-23 гг. а) числа открытых проектов (нейросетевого) ИИ и б) звезд (отмеченных пользователями ссылок) для этих проектов [27]

До 2011 г. доля открытых проектов (нейросетевого) ИИ, расположенных на GitHub, была ничтожна (рис. 6), но с 2015 г. начался бурный рост, и сейчас их количество приближается к 10% от общего числа открытых проектов.

Данные, приведенные на рис. 6, показывают, что как некоторые крупные технологическим компании, так и более мелкие организации или стартапы, а также отдельные разработчики сосредоточены на демократизации исследований в области ИИ.

GitHub не только дает возможность обмена кодами, но и (по подписке) предоставляет средства работы с программами, такие, как GitHub Copilot – инструмент, разработанный GitHub и OpenAI (49% принадлежит Microsoft), который помогает пользователям интегрированных сред разработки (IDE) Visual Studio Code, Visual Studio, Neovim и JetBrains путем автозаполнения кода. Есть и другие средства, используемые разработчиками проектов нейросетевого ИИ, но пока более чем в 50% случаев они отдают предпочтение GitHub Copilot (рис. 7а). Есть прогнозы, утверждающие, что по мере развития технологии GPT ее можно будет использовать для написания кода. Но в 2023 г. ChatGPT в основном использовался не для разработки кода, а для поиска, в том числе и уже разработанного кода (рис. 7б).

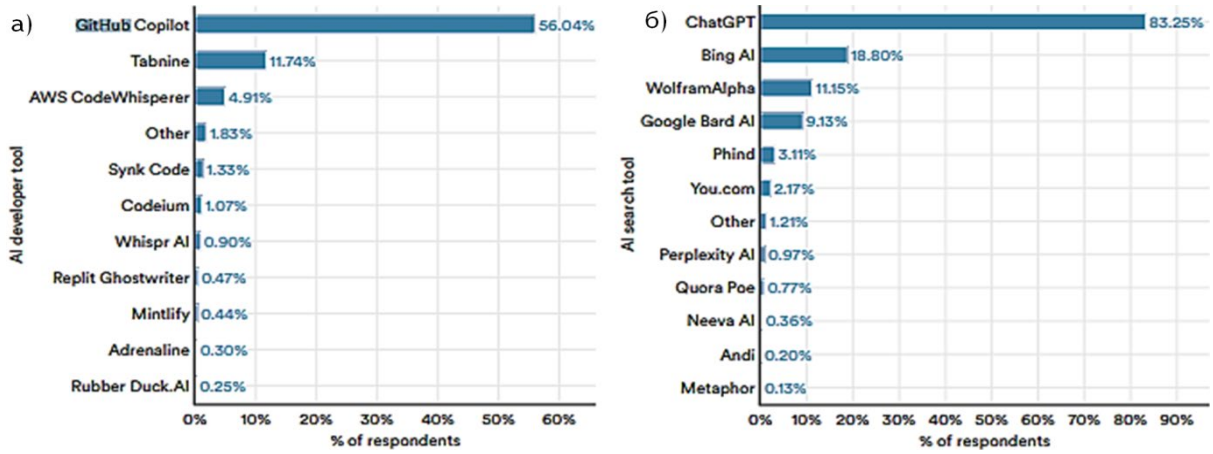


Рис. 7. Наиболее популярные в 2023 г. средства а) разработки и б) поиска среди профессиональных кодировщиков проектов (нейросетевого) ИИ [27]

Все больше разработчиков переходят на сетевые инструменты и овладевают навыками генерации кода с помощью искусственного интеллекта. Это является новым направлением в разработке программного обеспечения, которое будет определять развитие отрасли в ближайшие годы. Хотя сейчас нельзя точно сказать, что программное обеспечение на основе генеративного искусственного интеллекта полностью заменит программистов, но степень автоматизации их работы будет продолжать увеличиваться год от года.

С развитием нейросетевого искусственного интеллекта сфера решаемых задач будет расширяться, и успехи в различных направлениях будут отражаться в представленных на GitHub моделях. Успехи Sora могут значительно изменить ландшафт кодов на GitHub в ближайшие месяцы, но это будет зависеть от интереса разработчиков к данному направлению, а также от готовности OpenAI делиться своим кодом.

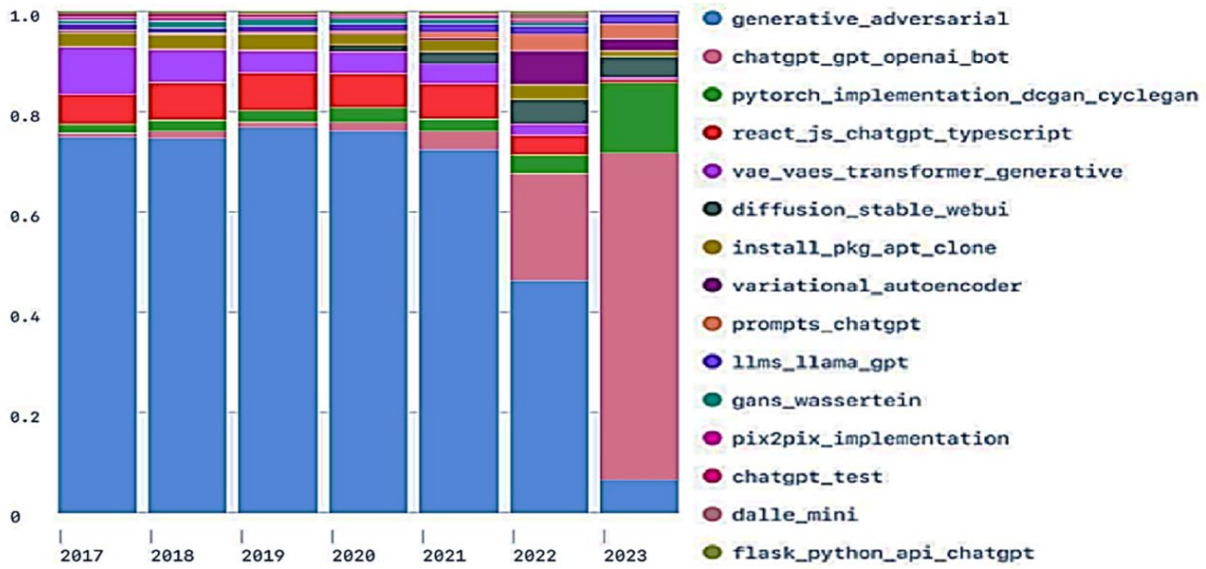


Рис. 8. Эволюция наиболее актуальных среди профессиональных разработчиков темы проектов (нейросетевого) ИИ на GitHub [28]

Важно отметить, что сетевые и генеративные инструменты для обмена и помощи в разработке кода играют важную роль в развитии нейросетевых алгоритмов.

2.6. Победа АльфаГо и АльфаЗеро в го – начало гонки Китай–США (и все остальные…)

Глобальное развитие искусственного интеллекта претерпело значительные изменения после побед, достигнутых программами DeepMind – АльфаГо и АльфаЗеро в 2016-17 гг.

Предыдущим заметным достижением ИИ в данной области была победа Deep Blue II. В мае 1997 г. созданная в корпорации IBM (и усовершенствованная после поражения Deep Blue 2:4 в 1996 г.) программа Deep Blue II, установленная на суперкомпьютерный кластер RS/6000 SP с 30 узлами, содержащими в сумме 480 шахматных процессоров и 30 центральных процессоров, смогла выиграть матч у Гарри Каспарова со счетом 3½ : 2½. Признавая успех, скептически настроенные к ИИ специалисты предрекали, что аналогичное достижение в игре го случится не ранее, чем через 100 лет, поскольку комбинаторная сложность го более, чем на 120 десятичных порядков превосходит шахматы. Но случилась нейросетевая революция в МО и время потекло быстрее...

Успехи в игре в го, начавшиеся с победы АльфаГо над чемпионом Европы в 2015 г. и продолжаясь победами над выдающимися го-игроками, значительно изменили ландшафт развития искусственного интеллекта.

Следует отметить революционный подход АльфаЗеро, который обучался самостоятельно и противостоял не человеческому опыту, а собственным вариациям игры [29]. Победа АльфаЗеро со счетом 100:0 над предшественником и другими программами в играх с полной информацией стала краеугольным камнем в развитии искусственного интеллекта.

Эти достижения впечатлили как Китай, так и США, вызвав явное увеличение внимания к развитию ИИ. Кроме того, утвержденный в Китае «План развития искусственного интеллекта нового поколения» до 2030 г. предполагает важные инвестиции в индустрию ИИ, а также установление стандартов и правил в этой области.

Китай, не будучи первым в принятии стратегии развития искусственного интеллекта (Канада, Япония и Сингапур также утвердили свои стратегии в 2017 г.), уделяет особенно большое внимание этой области. США только в мае 2018 г. представили свою стратегию развития ИИ. Они оказались в списке стран, подготовивших такие планы, во втором десятке. К 2020 г. более 50 стран разработали и утвердили стратегии по развитию искусственного интеллекта, включая и Россию, которая присоединилась к ним в 2019 г.

Так как развитие нейросетевых технологий продолжает ускоряться, наблюдается нарастание «гонки вооружений» в сфере нейросетевого искусственного интеллекта. Все страны призывают к сотрудничеству, однако каждая из них стремится обеспечить себе лидерство в этой области. Однопартийная система правления в Китае позволяет ему определять нормы обработки персональных данных не столько в целях защиты прав и свобод граждан, сколько для сбора крупных данных (с учетом практически полуторамиллиардного населения Китая) для обучения нейросетевому искусственному интеллекту. Некоторыми критиками такое «социальное партнерство» граждан Китая рассматривается как ключевое преимущество страны в области искусственного интеллекта.

США продемонстрировали свое доминирование в области нейросетевого искусственного интеллекта в 2016-17 гг. Многие страны, в том числе и Китай, восприняли это как показатель силы. Начиная с 2010-12 гг., нейросетевая революция в области машинного обучения протекала в технологически развитых странах с сильным научным потенциалом, а после 2016-17 гг. она распространилась по всему миру. Хотя лидерство США в этой области находится под вопросом, страны стремятся не отставать. А Китай стремится не просто догнать, а и превзойти их [30].

2.7. ChatGPT и Базовые модели (LLMs and Foundation models) – выход нейросетей в медийное пространство

Если победы АльфаЗеро мало влияли на повседневную жизнь простых граждан, то технологии, внедряемые в Интернет и мобильную телефонию, способны напрямую воздействовать на широкие массы населения. Нейросетевые технологии применяются в сети уже довольно давно. Начиная приблизительно с 2012-15 гг., их стали активно использовать при поиске в браузерах, персонализации рекламы, переводчиках, голосовом вводе и многих других сервисах и функциях. Не то, чтобы использование нейросетевых алгоритмов до появления ChatGPT скрывалось, но и не особо подчеркивалось. Просто при использовании нейросетей некоторые функции начинали работать немного лучше, а к хорошему все быстро привыкали.

11 июня 2018 г. компания OpenAI опубликовала пионерскую статью об улучшении понимания языка с использованием Генеративного Предобученного Трансформера (Generative Pre-Trained Transformer, GPT) [31]. Предложенный метод обучения модели GPT включает два этапа:

1. «предварительное» генеративное обучение, на котором на большом объеме данных GPT учится генерировать автоматически убираемые из текстов слова (по их оставленным соседям), чем устанавливаются предварительные параметры модели, описывающие свойства предметных пространств, содержащиеся в текстах, использованных для обучения;
2. «дообучение», на котором эти «предварительные» параметры на основе контроля учителем адаптируются к конкретной задаче.

Модель GPT обрабатывает векторные представления слов и предложений в латентном пространстве, где их смысл может меняться в зависимости от контекста, а не непосредственно символьные последовательности. Благодаря отсутствию «учителя» в этапе предварительного обучения модель GPT может использовать обширные текстовые наборы для обучения, что привело к улучшению ее способности «понимания» текста и повышению качества решения различных задач обработки естественного языка (NLP), таких как классификация текста, машинный перевод и генерация текста.

С дебютом улучшенной версии модели GPT-3, известной как GPT-3.5, и запуском ChatGPT в ноябре 2022 г. модель GPT привлекла широкое внимание. Однако успехи нейросетевого искусственного интеллекта не сводятся только к популярности ChatGPT. Множество других систем, таких как уже упомянутые Midjourney, DALL-E, Stable Diffusion, Sora и другие, используют подходы, основанные на идеях и моделях GPT, что приводит к эффективному решению различных задач.

При всей кажущейся простоте нейросетевых алгоритмов даже ведущие специалисты в этой области говорят, что полного понимания принципов их работы у них нет, теория находится в стадии развития, многие вопросы еще предстоит решить [32]. Тем более сложно разобраться с принципами развития нейросетей тем, кто недавно заинтересовался ими, даже

имея за спиной большой опыт научной работы. Так, Стюарт Дж. Рассел, один из авторов наиболее популярного в мире, выдержавшего 4 издания университетского учебника «Искусственный интеллект: современный подход» (AIMA, [33]), опубликовал в 2019 г. книгу [34] о совместимости нейросетевого ИИ и человеческой цивилизации. Рассел склонен наделять нейросетевой ИИ человеческими свойствами. Ян ЛеКун и Йошуа Бенджио понимают, что AGI будет таким, каким его создадут разработчики и проектировщики, а Рассел знает, что люди обладают рядом негативных свойств и, если им дать власть (особенно неограниченную), то у них обязательно разовьются именно отрицательные качества. Эти свои знания Рассел развивает в книге [34], приходя к выводу, что возможности нейросетевого ИИ надо строго ограничивать. Аналогичная ситуация с М. Тегмарком, астрофизиком по основной специальности. Когда он заинтересовался темой ИИ, ему в голову пришло много любопытных мыслей, и Тегмарк не только написал книгу [35], но и вошел в число организаторов Future of life institute [36], занимающегося вопросами развития цивилизации в условиях усиления нейросетевого ИИ.

Журналистам ближе и понятнее мнения, которые высказываются неофитами, недавно пришедшими в нейросетевой ИИ. и сами журналисты, для красного словца, готовы добавить что-нибудь «жаренное» от себя... Это приводит к тому, что ожидания публики, начитавшейся новостей и статей в популярных журналах, не совпадают с действительностью, которая, после появления в сети ChatGPT и других LLM, стала для многих «доступна в ощущениях».

Но важно другое. Число пользователей, решающих с помощью нейросетевого ИИ свои задачи, приближается к миллиарду. Не всем необходимо разбираться в принципах работы нейросетевого ИИ, но на волне интереса появляется достаточно много людей, которые сделают нейросетевые алгоритмы своей специальностью, разберутся и будут их развивать.

### 3. Чудесные свойства нейросетей – в чем они состоят?

Многие достижения цивилизации рассматривались как чудо (рис. 9), но двояко: или как сакральное средство решения всех проблем, или как порождение неограниченного зла. Нейросетевой ИИ из их числа, но, возможно, он превзойдет все остальные изобретения человечества по влиянию на судьбы цивилизации.

Нейросети, как и другие средства, описанные на рис. 9, обладают «чудесными» свойствами, но их возможности ограничены техническими характеристиками. В будущем появление искусственного общего интеллекта (AGI) может дать нейросетям способность создавать контент без участия человека. Вера в «чудеса» основана на непонимании истинной природы технологий.


Рис. 9. Слева – «волшебная» книга; Справа – сжигание хунвейбинами (кит. 紅衛兵, буквально: «красногвардейцы») книг на площади в период культурной революции [37]. Не только книгам, множеству других средств, влияющих на формирование у людей «картины мира» (список в центре – неполный), приписывают «волшебные» свойства

3.1. Эффективное обучение очень большого числа параметров

Жизненный опыт указывает на сложность настройки устройств даже с небольшим числом параметров при понимании цели пути к ней. Построение сложных систем основано на декомпозиции сложных задач на простые. Исследования нервных систем живых организмов показывают сложность их структуры. Обучение миллионов и миллиардов параметров в нейросетях стало возможным благодаря алгоритмам глубокого обучения. Недостаточное понимание механизмов глубокого обучения требует новых подходов к обработке данных и обучению нейросетей. Это техническое чудо успешно решает множество прикладных задач и приносит экономическую пользу. Сравнения с «чудом» полета или начальной фазой самолетостроения позволяют увидеть потенциал нейросетевых алгоритмов, но развитие в этой области только начинается, и в любом случае полеты в космос – еще впереди!

3.2. Успехи в борьбе с «комбинаторным взрывом»

Теория сложности [39] подчеркивает необходимость разбиения сложных задач на более простые компоненты для их эффективного решения. Нейросетевой подход к нейросетевому ИИ подвергся критике в связи с необходимостью выделения простых компонентов для статистического анализа в сложных средах. Однако развитие технологий нейронных сетей демонстрирует возможность обучения без явного понимания внутренних механизмов.

Текущие исследования направлены на создание теории декомпозиции сигналов в глубоких нейронных сетях для улучшения понимания их работы и повышения эффективности. Исследование векторов активности в многомерных пространствах латентных слоев должно привести к оптимизации процесса декомпозиции сигналов и улучшению результативности нейросетевых алгоритмов [40]. Создание такой теории открывает новые

горизонты для эффективного применения нейронных сетей в различных областях и обещает значительный прогресс в будущем.

3.3. «Новое программирование» – победа в соревновании с «классическим» программированием

Современные LLM способны создавать различные тексты, включая программы. На данный момент нейронным сетям удается формировать лишь небольшие программы, и их функциональность требует дополнительной проверки. Однако возможность LLM написать корректные тексты программ уже активно используется опытными программистами для частичной автоматизации своей работы. Считается, что с развитием технологий возможно увеличение длины и точности программ, созданных нейросетями, что позволит создавать программные продукты на основе текстового или голосового запроса.

Но уже сейчас нейросетевые алгоритмы все чаще рассматриваются как новая парадигма программирования. Вместо разработки эвристических алгоритмов на основе анализа предметной области задачи программист выбирает подходящую структуру нейросети и настраивает ее параметры в процессе обучения на примерах.

Нейросетевая аппроксимация позволяет автоматизировать преобразования, для которых известен правильный способ выполнения. Возможности нейросетей пока ограничены в случаях, когда нет правильных примеров для решения задач. Тем не менее, с применением обучения с подкреплением можно оптимизировать аппроксимацию на основе результатов, улучшая оценки преобразований.

Подходы к оценке эффективности последовательных действий часто требуют оценки результатов, поскольку промежуточные действия сложно оценить. Алгоритм deep q-network (DQN) [41], разработанный командой DeepMind под руководством Д. Хассабиса, успешно решает эту проблему в применении к нейросетям.

Применение нейросетей не ограничивается настольными играми. DeepMind успешно создает программные решения на основе нейросетевых алгоритмов, что подтверждается коммерческим успехом в прикладных задачах [42].

3.4. Обработка «реальных» сигналов

Другим важным достижением нейросетевых алгоритмов, которое тоже воспринимается как «чудо», является способность обрабатывать реальные сигналы, такие как изображения и речь, а также учиться на неразмеченных текстах. С развитием нейросетей появилась возможность автоматического выделения объектов и параметров из сложных сигналов. Нейросетевые алгоритмы демонстрируют прогресс в этой области, позволяя автоматизировать процесс выделения компонентов сигналов без участия человека.

Так или иначе, но несмотря на возможность допущения ошибок, нейросети во многих случаях уже превосходят человеческие способности и могут быть предпочтительны в определенных задачах.

Особенно заметный прогресс в выделении простых компонент из сложных сигналов произошел, когда обработку изображений объединили с моделями LLM. Это позволило перенести знания людей о простых объектах и явлениях сложной среды (выраженные отдельными словами) на изображения и сделало значительно более содержательным процесс обучения нейросетей выделению отдельных компонент (рис. 10). Это, конечно, использование человеческих знаний, но и каждый человек, владеющий языком, тоже (как правило) не сам придумывает разбиение среды на объекты, а применяет накопленные цивилизацией знания.

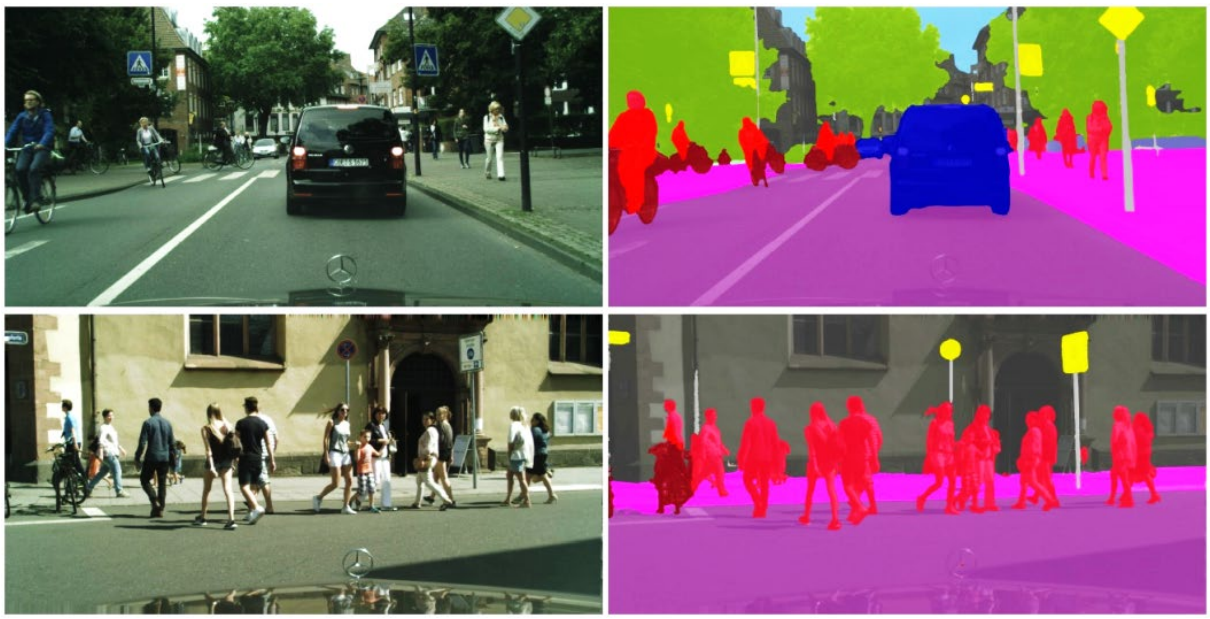

Рис. 10. Сегментация изображений с использованием нейросетевых алгоритмов [43]

### 3.5. Векторизация «текстового» пространства

Легко заметить, что современные нейросети (начиная с работ [16]) представляют входные сенсорные сигналы в виде векторов активности элементов (с непрерывными значениями активности), которые обрабатывают путем умножения на матрицы связей (тоже с непрерывными значениями) и осуществления нелинейных преобразований. Буквы, слова и предложения имеют явно выраженную дискретную природу, их изменения могут быть не меньше, чем на букву, а часто – и на словосочетание. При дискретном описании объектов операция дифференцирования (необходимая для применения метода обратного распространения ошибки в нейросетях) становится невозможной. Как же удается нейросетями обрабатывать тексты, создавать LLMs?

Идея преобразования слов в вектора достаточно проста: слова имеют некоторые значения, которые соотнесены с описаниями состояний окружающего мира, который, с точки зрения сенсорного восприятия, имеет непрерывную природу. Поскольку сложная среда состоит из множества объектов, имеющих параметры, которые могут изменяться независимо, размерность векторного описания состояний сложной непрерывной среды должна быть заметно больше трех, но значительно меньше числа слов, используемых для ее описания.

Эта простая идея реализуется путем распределенного представления (эмбеддинга) слов (word embedding) в виде многомерных (обычно размерность – порядка сотен компонент) векторов. В дальнейшем именно эти векторные описания слов – эмбеддинги – и используются для обработки в нейросетях. Желательно, чтобы геометрические соотношения между эмбеддингами в векторном пространстве воспроизводили семантические отношения между значениями слов. Например, эмбеддинги близких по значению слов в векторном пространстве должны быть ближе друг к другу, чем к другим словам. Но есть и много других семантических свойств, которые тоже хотелось бы воспроизвести при формировании эмбеддингов.

Простота идеи не означает легкость ее реализации – можно придумать огромное число способов преобразования слов в вектора, описывающие их значения. И предполагаемая многомерность пространства значений оставляет большую свободу субъективной фантазии разработчиков при выборе обозначений базовых осей данного пространства, что практически исключает возможность прийти к единому стандарту.

Необходимо использовать принципы, позволяющие на основе автоматической обработки текстов формировать вектора, не зависящие от субъективных представлений разработчиков. Один из таких принципов состоит в том, что близкие по значению слова встречаются в похожих контекстах. Такой подход еще в 1960-е гг. получил название «дистрибутивной гипотезы» [44]. Именно развитие этого подхода привело к реализации идеи распределенного (векторного) представления слов. Еще до начала нейросетевой революции такие эмбеддинги начали формировать с использованием нейросетей [45].

Но настоящий бум в данной области произошел в 2014-15 гг., после предложений, изложенных в статье Т. Миколова и соавторов [46], которые позволяют заметно ускорить процесс векторизации слов с помощью нейросетей. Что, в свою очередь, позволило обрабатывать значительно большие объемы текстов при формировании эмбеддинга.

Развитие теории распределенного представления слов на этом не остановилось и развитие пошло в двух направлениях: векторизации подверглись как морфемы (части слова: корень, приставки, суффиксы и пр.), так и предложения и абзацы. Кроме того, развивались модели рекуррентных нейросетей. Важными вехами на этом пути стали LSTM и трансфор-

меры. Современные LLM тоже используют массивы усовершенствованных трансформеров, но это не значит, что развитие теории обработки естественного языка (natural language processing, NLP) на этом завершилось. Успехи по претворению в жизнь идеи векторного воспроизведения семантических отношений между значениями слов велики, но работа продолжается, всё больше операций с эмбеддингами дает корректный по значению результат. Векторное описание слов во всё большей степени учитывает контекст, в котором данное слово расположено.

| **король + женщина − мужчина = …** | | **математик + женщина − мужчина = …** | |
|---|---|---|---|
| королева | 0.624 | филолог | 0.667 |
| империя | 0.562 | переводчица | 0.666 |
| принцесса | 0.552 | доктор_философии | 0.660 |
| правительница | 0.532 | доктор_филологических_наук | 0.656 |
| королевская_семья | 0.531 | лингвист | 0.655 |
| короля | 0.510 | социолог | 0.652 |
| аристократия | 0.510 | аушра_аугустинавичюте | 0.651 |
| инквизиция | 0.503 | доктор_филологических_наук_профессор | 0.650 |
| императрица | 0.503 | кандидат_филологических_наук | 0.649 |
| королева_елизавета | 0.500 | поэтесса | 0.648 |
| **москва + франция − россия = …** | | **укрощение_строптивой + гоголь − шекспир = …** | |
| париж | 0.566 | ночь_перед_рождеством | 0.791 |
| фр | 0.530 | за_двумя_зайцами | 0.791 |
| париж_франция | 0.493 | ревизор_гоголя | 0.787 |
| жан | 0.489 | живой_труп | 0.784 |
| французский | 0.486 | без_вины_виноватые_островского | 0.783 |
| брюссель | 0.486 | вишневый_сад | 0.779 |
| франсуа | 0.485 | волки_овцы_островского | 0.778 |
| анри | 0.485 | александринский_театр | 0.778 |
| женева | 0.481 | братья_карамазовы | 0.777 |
| paris | 0.480 | мертвые_души | 0.776 |

Рис. 11. Ближайшие слова к результатам четырех векторных преобразований значений слов [12]

«Осмысленные» результаты действий с эмбеддингами слов приводятся, например, в [12] (см. рис. 11). На большом корпусе текстов (русскоязычной Википедии) осуществляется векторизация слов и устойчивых словосочетаний из 2÷4 слов. Затем берется алгебраическая сумма 3-х векторов и выбираются слова, векторное представление (эмбеддинг) которых наиболее близко (в смысле нормированного значения скалярного произведения) к получившейся сумме векторов. Зачастую, но не всегда, результаты получаются осмысленными, хотя минимальные значения расстояний между сравниваемыми векторами всегда достаточно велики. Это связано с использованием в качестве примера одной из ранних моделей формирования эмбеддинга и отсутствия контекстной трансформации значений векто-

ров слов и их комбинаций. Дальнейший прогресс в данной области направлен на улучшение соответствий.

Достаточно удивительным является тот факт, что LLMs способны генерировать «содержательные» тексты только на основе анализа текстовой информации, без привязки ее к реальному миру. Это демонстрирует, что структурированные людьми тексты содержат достаточное количество правил, выявление которых позволяет генерировать похожие на «осмысленные» последовательности букв и слов. Но, конечно, чисто «книжные» знания не могут дать всю полноту восприятия сложного мира. И для более полного представления «картины мира» необходимо связать знания, полученные из книг, с реальным миром. Одним из наиболее перспективных путей решения данной проблемы представляется создание мультимодальных моделей нейросетевого ИИ на основе LLM.

Наиболее чудесным в процессе векторизации текстов является сам факт, что осуществление эмбеддинга и последующая обработка векторов нейросетями позволяет получать, по крайней мере, кажущиеся всё более осмысленными, результаты. Люди не способны ощущать активность своих живых нейросетей, им кажется, что мышление происходит за счет «абстрактных» действий с нематериальными дискретными представлениями в мире мыслей. Успешность LLMs наглядно показывает, что с абстрактными, выраженными словами понятиями можно совершать разнообразные векторные преобразования на нейросетях, без привлечения нематериальных представлений.

3.6. Воображение генеративных нейросетей

Одним из самых наглядных проявлений нейросетевого ИИ как «чуда» техники является его способность не только обрабатывать сложные сигналы, но и генерировать новые сложные сигналы, которые становится всё труднее отличать от образов, полученных из реального мира. Наиболее известны и популярны результаты генерации изображений, речи, текстов, переводов, но сейчас нейросетями генерируются самые разнообразные выходные сигналы: от текстов программ до описаний сложных органических соединений, от продолжений партий в го или шахматы до действий летчика по управлению истребителем в воздушном бою и многое другое. В 2024 г. список продолжает расширяться и важным шагом является реалистичная генерация видеопоследовательностей [23].

Первый скачок прогресса в генерации сложных сигналов случился с появлением в 2014 г. *генеративно-состязательных нейросетей* (GANs, [2]). Основой успеха стала предложенная методика автоматизация оценки реалистичности генерируемых сигналов. Для этого использовалось (и продолжает использоваться) две сети: одна (генератор) генерирует сигналы, а вторая (дискриминатор) получает на свой вход вперемешку реальные и сгенерированные сигналы и должна научится их различать (сгенерированные от реальных). Если дискриминатор правильно определил, что сигнал

получен был сгенерирован, то генератор получает сигнал на обучение. Если же ошибся дискриминатор, то обучается он.

Не следует думать, что в начале процесса генератор выдает случайный шум, а дискриминатор только сообщает ему, похож ли этот шум на фотографии или нет. Всё организовано сложнее. Предобучение в GANs было всегда: прежде чем начать генерировать сигналы генератор учится на реальных сигналах и формирует модель их латентного пространства. Аналогичную модель на фазе предобучения формирует и дискриминатор. А на фазе генерации сигналов обе нейросети улучшают созданные на основе реальных сигналов модели латентных пространств для выполнения противоположных задач.

GPT (Generative Pre-trained Transformer [6]) – это тоже предобученные нейросети, но, поскольку изначально они были ориентированы на работу с текстами в пространстве с выделенными дискретными простыми понятиями, выраженными словами, то соревновательная компонента (генератор – дискриминатор) в них не потребовалась. Версия GPT-3 изначально генерировала только тексты, зато практически произвольной тематики, поскольку была предобучена на очень большом массиве текстов самых разнообразных направлений.

Преимущество обучения на текстах состоит в том, что описание сложного мира ведется словами, с помощью которых сложные сцены представляются в виде набора простых объектов и несложных действий. Такая декомпозиция сложной среды заметно упрощает ее описание и может быть использована для обработки других форматов данных, таких как изображения, тексты, аудио, видео и пр., в которых выделение отдельных компонент требует дополнительных усилий.

До развития LLMs системы искусственного интеллекта были ограничены в возможностях: языковые модели понимали текст, но терпели неудачу в обработке изображений, и наоборот. DALL·E был описан OpenAI в сообщении в блоге 5 января 2021 г. и использовал версию GPT-3, модифицированную для генерации изображений. В течение 2021-22 гг. достижения LLMs позволили разработать мощные мультимодальные системы, такие как Gemini от Google, Midjourney от lab Midjourney, Inc., Stable Diffusion от Stability AI, GPT-4 от OpenAI и др. Эти модели демонстрируют гибкость и способны обрабатывать изображения и текст, а в некоторых случаях звук и видео (см., например, [47], глава 8).

Поскольку это мощные нейросетевые структуры, отражающие значительные объемы знаний, то, как правило, они реализуются в облачных серверах и доступны пользователям по сетевым запросам. Но, например, для Stable Diffusion были опубликованы программный код и веса связей обученной модели. Эта система может работать на большинстве рабочих станций, оснащенных относительно дешевым графическим процессором с объемом видеопамяти не менее 4ГБ [48].

Другим важным достижением в области генерации сложных сигналов (кроме использования знаний LLMs) стало развитие направления «диффузионок». Достаточно глубоко проработанную теорию [49] принято коротко характеризовать как «обращение вспять процесса добавления шума к изображению», что ведет к получению более четких, учитывающих взаимосвязь между различными компонентами сигналов, результатам. Эффективность «диффузионного» подхода к генерации сигналов подтверждается тем, что большинство успешных современных моделей мультимодального ИИ его используют (включая все, перечисленные в данном пункте).


Рис. 12. Зеленая область выделяется пользователем и модель заполняет ее согласно заданной текстовой подсказке в соответствии со стилем и освещением окружающего контекста [50]

Мультимодальные модели нейросетевого ИИ могут не только генерировать, но и редактировать как реальные, так и сгенерированные изображения. На рис. 12 показаны результаты редактирования двух изображений с помощью разработанной OpenAI в 2022 г. программы GLIDE [50].

За последние 2-3 года мультимодальные модели нейросетевого ИИ получили широкую популярность. Возможность не только создавать фотореалистичные изображения, но и редактировать их, позволяет последовательно улучшать сгенерированные изображения и добиваться их совершенства. Этой весной в США проходит Miss AI – первый в истории конкурс красоты среди моделей, сгенерированных нейросетевым ИИ. Организатором конкурса выступает компания World AI Creator Awards, открыт официальный сайт конкурса [51], на котором опубликованы все подробности мероприятия (рис. 13).

Процесс улучшения и создания новых модификаций мультимодальных моделей нейросетевого ИИ продолжается: в 2023 г. компания Google выпустила Gemini AI [52], значительным шагом вперед стала созданная OpenAI в 2024 г. программа генерации видеопоследовательностей Sora [23]. Некоторые разработчики (например, [52]) склонны полагать, что такие программы «представляют собой ИИ следующего поколения – многогранный, универсально применимый и удобный для пользователя, это не просто шаг вперед в области ИИ, это гигантский скачок в будущее нашего

взаимодействия с технологиями». И нельзя с ними не согласиться: прогресс нейросетевого ИИ за последние 2-3 года стал впечатлять не только специалистов, но и широкие массы пользователей сети, взявшихся осваивать последние достижения в данной области.

Рис. 13. Воображаемые нейросетями девушки на конкурс красоты Miss AI – сгенерированных моделей [51]

Но чудо состоит не столько в стремительном росте числа пользователей нейросетевого ИИ, сколько в его удивительной и всё усиливающейся способности генерировать сложные сигналы разных модальностей, в том числе по речевым запросам. Генерация сложных сигналов – это, по сути, демонстрация возможностей воображения, представления ситуаций, которых на самом деле не было, но они могли быть. Воображение – важная составляющая человеческого мышления, которое, в свою очередь, в значительной степени определяет действия, осуществляемые человеком. Конечно, современные мультимодальные модели нейросетевого ИИ пока еще нельзя назвать AGI или даже HLAI (Human Level AI, ИИ человеческого уровня). Но во всем «зоопарке» созданных на сегодня нейросетевых моделей они ближе всего подошли к HLAI и AGI. Это осознается большинством исследователей и значительной частью пользователей сети. И именно поэтому в Стэнфорде ввели и популяризируют термин «фундаментальные модели» (foundation models, [6]), который соответствует понятию «мультимодальные модели нейросетевого ИИ на основе LLM»

3.7. Основные чудеса – впереди!

Появятся ли еще другие нейросетевые «чудеса»? Да, но это будут опять «чудеса» техники, а не магия или волшебство. Главная способность нейросетевых алгоритмов – строить и осуществлять аппроксимацию слож-

ных преобразований – будет развиваться и применяться для решения всё более широкого круга задач.

Одним из центральных вопросов современности является возможность и время, которое осталось до создания AGI. Единого мнения на эту тему нет и сроки называются от «никогда» до «в этом году». Принципиальным является вопрос, достаточно ли «универсальных аппроксиматоров» для решения всех «интеллектуальных» задач или потребуются пока неизвестные, совершенно новые средства. Уже реализованные технические «чудеса» дают основания надеяться, что не только достаточно, но и значительная (большая) часть пути до создания AGI уже преодолена, и осталось немного. Но некоторые проблемы еще предстоит решить.

Для создания AGI важны вопросы не только использования созданных человечеством знаний, но и получение новых знаний, иерархической организации их представления и использования знаний для формирования целей управления и способов их достижения. Агенты AGI будут отличаться от современных foundation models тем, что смогут самостоятельно извлекать знания из наблюдений за окружающим миром и формировать в нем свои, как промежуточные, так и глобальные цели. Получение новых знаний будет направлено на расширение арсенала возможностей и методов реализации рационального поведения, повышающего вероятность выживания цивилизации.

Закон необходимого разнообразия Эшби и успехи нейросетевой революции в машинном общении указывают, что построение AGI будет осуществляться путем развития нейросетевых алгоритмов. Эволюционный метод «проб и ошибок» тоже позволяет выявлять степень эффективности различных нейросетевых алгоритмов при решении разнообразных задач. Но создание AGI значительно более вероятно произойдет за счет использования методов проектирования, основанных на всё возрастающем понимании проблем, которые необходимо для этого решить.

В [40] мы стараемся показать, что одной из центральных проблем при создании AGI является сложность мира и необходимость его декомпозиции на отдельные объекты и явления, доступные для статистически достоверных наблюдений. И что только знаний о свойствах простых объектов и явлений недостаточно для построения в поведения в сложной среде. Необходимо соотносить эти знания с состоянием среды и уметь прогнозировать, оценивать и сравнивать изменения в сложной среде при выполнении различных действий.

Есть и много других задач создания AGI, но решение представляющейся нам центральной проблемы описания сложной среды должно упростить работу над остальными задачами.

### 3.8. Определение AGI

Прогнозировать, сколько еще потребуется пройти шагов, чтобы ИИ на нейросетевых алгоритмах достиг и превзошел человеческий уровень

сложно, но наиболее оптимистичные прогнозы состоят в том, что уже следующей ступенью их развития станет создание AGI. По крайней мере, уже сегодня ряд серьезных исследовательских центров, таких как OpenAI, DeepMind и Anthropic PBC, позиционирует разработку AGI центральным направлением своих исследований. В РФ Сбер курирует вопросы создания AGI [53], в 2020 г. в Китае создан Пекинский институт BIGAI (Beijing Institute for General Artificial Intelligence) [54].

Мы считаем, что для определения системы, как соответствующей понятию AGI, она должна удовлетворять следующим свойствам:

– способность к освоению имеющихся и получению новых (не следующих из известных согласно доступным методикам) знаний и использованию их для задач развития цивилизации;

– способность к выделению и описанию свойств простых компонент из сигналов, поступающих из сложной среды, основанной на адаптивной иерархической структуре представления знаний;

– развитость адаптивной иерархической структуры представления знаний должна позволять работать с высокоуровневыми цивилизационными знаниями.

## 4. Выводы

Современные успехи нейросетевого ИИ вызывают как удивление, так и опасения у людей, мало знакомых с принципами работы нейросетевых алгоритмов. Людям свойственно «очеловечивать» не только животных, но и «неодушевленные» объекты и явления, приписывая им человеческие свойства, которых у них на самом деле нет.

В данной статье сделана попытка связать «чудесные» свойства нейросетевых алгоритмов с техническими путями их реализации. В дальнейшем «чудес» станет больше, и они будут еще более впечатляющими, но они останутся чудесами техники, а не станут волшебством.

Понимание технической стороны всех производимых нейросетевым ИИ «чудес» показывает, что многое в поведении людей тоже вполне может быть объяснено работой нейросетевых алгоритмов. Пока не всё, но уже сейчас можно указать конструктивные пути, позволяющие приблизиться и обойти человека в способности получения новых знаний и их использования для построения рациональных действий. Это должно привести к развитию и повышению устойчивости существования человеческой цивилизации.

## Литература

1. *Krizhevsky A., Sutskever I., Hinton G.E.* ImageNet classification with deep convolutional neural networks // Communications of the ACM 60(6), 84-90 (2017).

2. *Goodfellow I.J., P.-A. Jean et al.* Generative adversarial networks. arxiv:1406.2661
3. *Silver D., Hubert T. et al.* Mastering chess and shogi by self-play with a general reinforcement learning algorithm. arxiv:1712.01815
4. *Brown T., Mann B. et al.* Language models are few-shot learners. arxiv:2005.14165
5. *Bommasani R., Hudsonet D.A. et al.* On the opportunities and risks of foundation models. arxiv:2108.07258
6. Developing and understanding responsible foundation models. https://crfm.stanford.edu.
7. *Ashby W.* Principles of the self-organizing dynamic system // J. Gen. Psychology 37, 125–128 (1947).
8. *Flood Sung* 关于深度学口口展的必然及未来的思考 (Мысли о неизбежном развитии глубокого обучения и его будущем). https://zhuanlan.zhihu.com/p/375226190
9. *Denning J., Lewis T.* Exponential laws of computing growth // Communications of the ACM 60(1), 54-65 (2016).
10. 10 платформ для разметки данных под задачи компьютерного зрения. https://labelme.ru/tpost/fm3kbv7m71-10-platform-dlya-razmetki-dannih-pod-zad
11. *Гудфеллоу Я., Бенджио И., Курвилль А.* Глубокое обучение / Пер. с анг. А.А. Слинкина / 2-е изд., испр. – М.: ДМКПресс. 2018.
12. *Николенко С., Кадурин А., Архангельская Е.* Глубокое обучение: Погружение в мир нейронных сетей. – СПб.: Питер, 2018.
13. *McCulloch W.S., Pitts W.* A logical calculus of ideas immanent in nervous activity. – Bull. Mathematical Biophysics, 1943
14. *Розенблатт Ф.* Принципы нейродинамики: Перцептроны и теория механизмов мозга. – М.: Мир, 1965. – 480 с.
15. *Turing A.* On computable numbers, with an application to the entscheidungsproblem // Proceedings of the London Mathematical Society Series 2, 42, 230-265 (1937).
16. *Widrow B., Hoff M.* Adaptive switching circuits // IRE Western Electric Show and Convention Record, 1960. Pp. 96-104
17. *Minsky M., Papert S.* Perceptrons: An introduction to computational geometry. – The MIT Press, Cambridge MA, 1972 (2nd edition with corrections, first edition 1969).
18. *Fukushima K.* Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position // Biolog. Cybernetics 36, 193-202 (1980).
19. *Carpenter C., Grossberg S.* A massively parallel architecture for a self-organizing neural pattern recognition machine // Computer Vision Graphics and Image Processing 37, 54-115, (1983).

20. *Kohonen T.* Self-organized formation of topologically correct feature maps // Biolog. Cybernetics 43, 59-69 (1982).
21. *Widrow B., Lehr M.A.* 30 years of adaptive neural networks: perceptron, Madaline, and backpropagation // Proceedings of the IEEE 78(9), 1415-1442 (1990).
22. *Shead S.* The 3 'godfathers' of AI have won the prestigious $1M turing prize // Forbes. 20 March 2020.
23. Sora is an AI model that can create realistic and imaginative scenes from text instructions. https://openai.com/sora
24. *Amri A.* OpenAI GPT for python developers / 2nd Edition: The art and science of building AI-powered apps with GPT-4, Whisper, Weaviate, and beyond. – Kindle Edition, 2024.
25. *Chaudhary K.* The GAN book: Train stable generative adversarial networks using TensorFlow2, Keras and Python. – Kindle Edition, 2024.
26. The world's leading AI-powered developer platform. https://github.com.
27. 2024 AI index report. https://aiindex.stanford.edu/report.
28. *Dohmke T., Iansiti M., Richards G.* Sea change in software development: Economic and productivity analysis of the ai-powered developer lifecycle. arxiv:2306.15033
29. *Silver D., Hubert T. et al.* A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play // Science 362(6419), 1140-1144 (2018).
30. *Ли К.-Ф.* Сверхдержавы искусственного интеллекта. – М.: Манн, Иванов и Фербер, 2019. – 238 с.
31. *Radford A., Narasimhan K. et al.* Improving language understanding by generative pre-Training. https://cdn.openai.com/research-covers/language-unsupervised/language_un-derstanding_paper.pdf
32. *Bengio Y., Lecun Y., Hinton G.* Deep learning for AI // Communications of the ACM 64(7), 58-65 (2021).
33. *Норвиг П., Рассел С.* Искусственный интеллект: Современный подход / 4-е издание. Т.1. Решение проблем: Знания и рассуждения. – М: Диалектика-Вильямс, 2021. – 704 с.
34. *Russell S.* Human compatible: Artificial intelligence and the problem of control // United States: Viking, 2019.
35. Tegmark M. Life 3.0: Being human in the age of artificial intelligence / First ed. – New York: Knopf, 2017.
36. Future of life institute, FLY. https://futureoflife.org
37. *Бондарева В.В.* Движение «красной охраны» в первые годы «культурной революции» в Китае (1966–1967): Взгляд отечественных и зарубежных исследователей // Научный диалог. 2020, №8.
38. *Hadley D.* Do insects have brains? https://www.thoughtco.com/do-insects-have-brains-1968477.

39. *Sipser M.* Introduction to the theory of computation. – PWS Publishing, 1997. Sections 7.3–7.5, pp. 241-271.
40. *Журавёв Д.В., Смолин В.С.* Проектирование структуры нейросетей для AGI. В этом сборнике.
41. *Roderick M., MacGlashan J., Tellex S.* Implementing the deep Q-network. arxiv:1711.07478
42. https://en.wikipedia.org/wiki/Google_DeepMind
43. *Tao A., Sapra K.* using multi-scale attention for semantic segmentation. https://developer.nvidia.com/blog/using-multi-scale-attention-for-semantic-segmentation
44. *Rubinstein H., Goodenough J.* Contextual Corrtlates of Synonymy // Communications of the ACM 8(10), 627-633 (1965).
45. *Bengio Y., Ducharme R., Vincent P.* A neural probablistic language model // Journal of Machine Learning Research 3, 1137-1155 (2003).
46. *Mikolov T., Sutskever I. et al.* Distributed representations of words and phrases and their compositionality. arxiv:1310.4546
47. *Фостер Д.* Генеративное глубокое обучение: Как не мы рисуем картины, пишем романы и музыку / 2-е изд. – Sprint Book, 2024. – 448 с.
48. Код и веса обученной структуры Stable Diffusion. https://github.com/CompVis/stable-diffusion
49. Weng L. What are diffusion models? https://lilianweng.github.io/posts/2021-07-11-diffusion-models
50. *Nichol A., Dhariwal P. et al.* GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. arxiv:2112.10741
51. Miss AI – конкурс от компании World AI Creator Awards https://www.waicas.com
52. Google's Gemini AI, a multimodal marvel, surpasses GPT-4 in various benchmarks. https://anakin.ai/blog/google-gemini-release-date
53. *Бурцев М.С., Бухвалов О.Л., Ведяхин А.А. и др.* Сильный искусственный интеллект: На подступах к сверхразуму. – М.: Интеллектуальная Литература, 2021. – 232 с.
54. Beijing Institute for General Artificial Intelligence (BIGAI). https://eng.bigai.ai