

Душкин Роман Викторович —
основатель Телеграм-школы по искусственному интеллекту.

Roman V. Dushkin —
Telegram-School on Artificial Intelligence.

Почему за гибридными ИИ-системами будущее



В статье [1] поставлены важные вопросы о современных достижениях в области глубокого обучения (ГО) и искусственного интеллекта (ИИ) в целом. Общий тон работы скорее пессимистичный и настраивает на переосмысление полученных результатов, даже если они и промежуточные. Маркус дает некоторые прогнозы и пишет о возможных последствиях очередной, третьей по счету, шумихи вокруг технологий ИИ [2]. В настоящей статье мне хотелось бы дать свое видение того, в каком направлении необходимо двигаться дальше при разработке искусственных интеллектуальных систем.

Оригинальная статья [1] довольно объемная, и, несмотря на то, что она переведена на русский язык, ее прочтение может занять некоторое время. Я все же рекомендую сначала про-

честь оригинальную статью, а затем вернуться к этой работе. Тем не менее вот краткие тезисы.

- Для работы ГО-систем необходимо много качественных данных, которые должны быть предварительно очищены и размечены специалистом. Чем больше слоев нейронной сети в ГО-системе, тем больше данных требуется.
- ГО-системы работают только с теми типами данных, на которых происходило обучение, и они все еще не могут обобщать и переносить найденные закономерности на данные других типов, даже очень близких.
- ГО-системам предельно сложно работать с иерархическими структурами, поэтому, к примеру, обработка языка для них — очень непро-

УДК 004.8

В статье Гари Маркуса *Deep Learning: A Critical Appraisal* [1] поставлены вопросы о современных достижениях глубокого обучения и искусственного интеллекта (ИИ). Общий тон работы пессимистичный и настраивает на переосмысление полученных результатов, даже если они и промежуточные. Маркус дает прогнозы и пишет о возможных последствиях очередной шумихи вокруг ИИ-технологий [2]. В настоящей статье дается видение того, в каком направлении следует двигаться при разработке ИИ-систем.

Ключевые слова

Искусственный интеллект, нейронные сети, глубинное обучение, гибридный подход, гибридная парадигма.



стая задача, поскольку естественный язык — это крайне глубокая иерархическая структура.

- Как следствие предыдущего ГО-системы с большим трудом воспринимают неточные и нечеткие данные, они часто не различают разницы там, где для человеческого интеллекта разница огромна.
- ГО-системы унаследовали от искусственных нейронных сетей и усугубили проблему высокой сложности (до практической невозможности) объяснения полученных результатов и произведенного вывода.
- ГО-системы не принимают во внимание уже имеющийся корпус знаний, а заново обучаются на входных данных, как-то по-своему их интерпретируя.



Why the Future is with Hybrid AI Systems

The article by Gary Marcus *Deep Learning: A Critical Appraisal* [1] dwells on the questions about the current achievements of in-depth training and artificial intelligence (AI). The overall tone of the work is pessimistic and tends to rethink the results, even if they are intermediate. Markus makes forecasts and writes about the possible consequences of another hype around AI technologies [2]. The article gives a vision on which direction to move while developing AI systems.

Keywords

Artificial intelligence, neural networks, in-depth training, hybrid approach, hybrid paradigm.

- Важной задачей является выявление причинно-следственных связей и отделение их от простых корреляций, но ГО-системам такое разделение пока дается с трудом.

- ГО-систему легко обмануть, особенно если она находится на «границе переобученности». Такая уязвимость открывает широкий простор для разнообразных атак, последствия которых не до конца осознаны. К решению этой проблемы даже не приступали.

- ГО-системы до сих пор не имеют прикладного применения, и этот пункт фактически является следствием всех предыдущих.

Далее в статье [1] автор делает неутешительные прогнозы о том, что шумиха, поднятая по поводу ГО, может привести к новой, третьей по счету, «зиме» искусственного интеллекта, а также дает свое видение того, как все это можно было бы преодолеть. Среди его советов есть такие, как применение методики спонтанного обучения, попытки решения более сложных задач, использование современных знаний в области психологии и, наконец, применение гибридных моделей с технологиями синтаксической манипуляции символами.

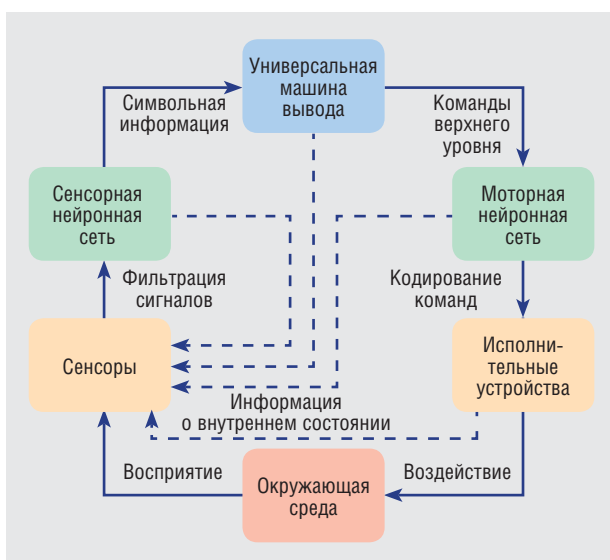
Краткое описание гибридной модели

Последняя возможность, предложенная в статье [1], является одной из наиболее предпочтительных, так как методы нисходящей парадигмы в области ИИ были разработаны много десятилетий назад [3] и сегодня незаслуженно подзабыты. Именно на слиянии двух подходов — нисходящего и восходящего — с построением ИИ-систем гибридного типа имеет смысл развивать технологии и преумножать возникшие сложности в области ГО.

Гибридная ИИ-система включает в себя лучшие качества как нисходящего, так и восходящего подхода, но при этом старается отвергнуть слабые стороны [4, 5]. В итоге получается довольно продвинутой интеллектуальной системы, которая в какой-то мере отображает устройство естественного интеллекта. Впрочем, если рассмотреть верхнеуровневую архитектуру такой системы, то будет ясно, что это довольно обычная кибернетическая система [6], взаимодействующая со средой через фильтры, поэтому неудивительно, что она напоминает биологические системы (см. *рисунк*).

Итак, как видно из *рисунка*, гибридная ИИ-система представляет собой не что иное, как

Верхнеуровневая архитектура гибридной ИИ-системы



универсальную кибернетическую машину, которая имеет *три* основных элемента: аффлекторы, подсистему управления и эффекторы. При помощи аффлекторов кибернетическая машина воспринимает сигналы окружающей среды, которые обрабатываются в подсистеме управления. Из нее сигналы поступают в эффекторы, воздействующие на окружающую среду. Это общая схема любого автономного агента, поэтому гибридная ИИ-система одновременно является и интеллектуальным агентом, реализуя агентный подход [7–9].

Гибридная ИИ-система отличается тем, что ее аффлекторы (сенсоры, датчики) и эффекторы (исполнительные устройства) связаны с подсистемой управления и принятия решений через нейронные сети. Тем самым используются сильные стороны восходящего, или «грязного», подхода. Аффлекторная нейронная сеть принимает очищенные сенсорами сигналы внешней среды и преобразует их в символы, которые подаются на вход универсальной машины вывода. Последняя осуществляет вывод на основе символьных знаний из своей базы знаний и выводит результат, который тоже представляется в виде символов. Таким образом реализуются сильные стороны нисходящего, или «чистого», подхода. Символьный результат подается на вход моторной нейронной сети, которая преобразует высокоуровневые символы в конкретные сигналы управления исполнительными устройствами.

Кроме всего прочего внутри гибридной ИИ-системы должны быть реализованы контроль-

ные связи от всех ее элементов к ее сенсорам. Тем самым реализуются адаптационные механизмы, основанные на гомеостазе внутреннего состояния системы [10]. Сенсоры фиксируют изменение внутреннего состояния каждой из подсистем, их элементов и комплексов, и в случае выхода контролируемых значений за пределы установленных гомеостатических интервалов подсистемой управления принимается решение, целью которого будет возврат измененных показателей в установочный интервал [11].

Именно система с такой архитектурой при переходе через определенный порог сложности может считаться разумной. При этом разумность в этом смысле определяется как адекватное реагирование не только на стимулы внешней среды, но и на внутренние состояния. Это также включает в себя постоянный мониторинг состояния собственной подсистемы управления, называемый саморефлексией, которая и приводит к осознанию.

Обсуждение

Далее вносятся предложения относительно того, как проблемы ГО могли бы быть преодолены при помощи гибридных ИИ-систем.

Глубинное обучение нуждается в данных

Чем больше данных, особенно качественно размеченных, тем лучше результат работы глубинных нейронных сетей [12]. Более того, они могут обобщать информацию и находить скрытые зависимости, строя что-то типа продукционных правил «если... то...» [13]. Фактически любая нейросеть при обучении делает это, так как в действительности строит правила преобразования входных данных в выходные, что на более высоком уровне абстракции можно представить в виде: «Если вход принимает такие-то значения, то выход равен тому-то» [14].

Однако проблема в том, что даже если нейронная сеть делает подобные обобщения, то их представление скрыто в ее глубинах в виде неявной информации, отображаемой только в весовых коэффициентах на связях между нейронами [15]. Интерпретация этих коэффициентов очень затруднительна, если вообще возможна, особенно с учетом того, что одна и та же межнейронная связь может делать вклад в несколько различных правил [16].

Если рассмотреть пример, приведенный в статье [1], где введено слово «шмистра», кото-

рое определяется как «сестра, возраст которой от 10 лет до 21 года», то нейронная сеть вполне может научиться определять, кто шмистра, кто просто сестра, а кто вообще ни та, ни другая. Вопрос лишь в обучающей выборке. И после того, как нейросеть обучилась, она отлично разделит шмистер от сестер, но никогда не сможет объяснить, кто такая шмистра и почему она распознает шмистру в том или ином случае.

➤ **Шумиха, поднятая по поводу глубинного обучения, может привести к новой, третьей по счету, «зиме» искусственного интеллекта.**

Однако если построить гибридную ИИ-систему, то ее архитектура может быть устроена следующим образом. Сенсорная нейронная сеть определяет базовые параметры входа, например, то, что на изображении женщина и ей по оценкам 18 лет. Это то, что нейронные сети ГО сегодня умеют делать очень хорошо. А вот дальше эти распознанные базовые характеристики подаются на вход универсальной машины вывода, в базе знаний которой есть правило: «Если женщина является сестрой и ей от 10 лет до 21 года, то это шмистра», после чего запускается процесс вывода — система проверит, является ли распознанная женщина сестрой, и если да, то будет абсолютно ясно, что это шмистра. Результат такого вывода может быть объяснен очень просто, спускаясь до базовых характеристик, распознанных нейронной сетью.

И обучение такой гибридной ИИ-системы может осуществляться как в процессе разработки, так и в процессе взаимодействия с ней [17]. Например, если мы общаемся с ней при помощи диалогового интерфейса (чат-бота), то при использовании в тексте слова «шмистра» ИИ-система может поинтересоваться, что это такое. И тогда можно дать ей определение.

Так что первая проблема систем ГО довольно легко решается при помощи гибридного подхода.

Глубинное обучение пока имеет небольшую глубину и плохо переносит полученные знания на другие данные

Системы ГО находят интересные закономерности, но не могут их обобщить, и отсюда происходят различные смешные казусы, описанные в статье [18, 19]. Обобщение — это операция

более высокого уровня абстракции, чем простое нахождение закономерностей, даже скрытых. И если с последним искусственные нейронные сети легко справляются и в некоторых областях уже превосходят в этом людей, то с обобщением у них серьезные проблемы.

Фактически обобщение — это символическая операция. Обобщая, можно сказать, что человек переносит найденные закономерности с отдельных экземпляров на типы и классы объектов внешнего мира или абстрактных сущностей [20]. И делается это именно на символическом уровне, когда правила вывода модифицируются при помощи метаправил. Например, одним из таких метаправил может быть следующее: «Если несколько объектов одного класса имеют одно и то же свойство, то предположить, что все объекты этого класса имеют это свойство». Классы могут объединяться в классы более высокого порядка, для них тоже можно делать обобщающие выводы и т.д. Все это можно реализовать при помощи самообучающейся универсальной машины вывода, в которой есть метаправила, в том числе и для изменения самих метаправил.

Однако для этого необходимо иметь иерархию понятий. Но у систем ГО реализация иерархий пока очень слабая, поскольку у нейросетевых моделей нет естественного способа представления иерархических понятий [21].

У глубинного обучения пока нет естественного способа работы с иерархической структурой

У систем ГО еще нет естественного способа представления иерархических структур и работы с ними, так как нейронные сети пока с трудом справляются с иерархиями понятий. Попытки представить иерархические структуры в нейросетевых моделях приводят либо к переобучению нейросетей, либо к тому, что они не могут полноценно дифференцировать объекты распознавания при спуске по иерархии вниз. Кое-каких успехов в этом вопросе достигли нейросети именно глубинного обучения, так как в них неявно и самостоятельно образуются иерархии образов, но ведь весь вопрос в том, чтобы вложить в нейросеть некоторую заданную иерархию в явном виде. Пока это вызывает определенные затруднения [22].

Хотя, к примеру, существуют нейросетевые модели, преобразующие последовательности слов в некие «векторы в трехсотмерном пространстве смыслов, суть которых человеческому со-

»» Чем больше данных, особенно качественно размеченных, тем лучше результат работы глубинных нейронных сетей.

знанию не ясна» и которые как бы группируются в какие-то кластеры. Но это опять же не иерархии, а все такие же векторные преобразования, которыми, по сути, нейросетевые модели и являются [23, 24].

С другой стороны, нисходящая парадигма имеет все средства для непосредственного представления иерархических понятий. Это можно сделать даже при помощи продукции, не говоря уже о семантических сетях и онтологии, которые именно для этого и предназначены [25]. Но здесь речь идет опять о явном представлении знаний и дедуктивном обучении с учителем, хотя вполне возможно построение искусственных интеллектуальных систем, которые обучаются не в специальном режиме обучения, а в процессе взаимодействия с окружающей средой. Главное, чтобы из среды приходило правильное подкрепление.

Так что при использовании гибридного подхода опять можно взять лучшее от обеих парадигм. На нижнем уровне распознающие нейросети будут определять конкретные объекты и понятия, с которыми приходится взаимодействовать системе, а на уровне выше эти распознанные понятия могут укладываться в семантические сети для определения иерархических взаимоотношений и получения обобщенных выводов на основе классов высших порядков, в которые входит распознанный объект [26]. Например, визуальная нейросеть распознает лицо мужчины на фотографии, а на верхнем уровне к конкретным



характеристикам распознанного лица добавляются не только персональные данные и физические характеристики, но и такие понятия, как «взрослый», «самец», «человек», «примат», «млекопитающее» и все остальные классы, в которые входят все нижележащие объекты иерархии. А далее к ним могут быть применены все правила и метаправила вывода для получения как заключений по конкретной ситуации, так и новых знаний.

Глубинное обучение до сих пор с трудом работает с неточно определяемыми понятиями

В процессе жизнедеятельности люди постоянно сталкиваются с ситуациями, когда данные неполны, информация неточна, знания противоречивы, нечетки и неопределенны. Эти так называемые НЕ-факторы знания пронизывают всю информационную ткань нашей реальности уже в силу самой ее природы, поскольку точность данных невозможно повышать до бесконечности в силу фундаментальных ограничений [21]. Человеческий разум научился обрабатывать такие ситуации, но системы глубинного обучения — нет, поскольку они требуют четко размеченных данных и максимально полного набора значений входных параметров для обучения. Только в последнее время начали появляться комбинированные механизмы для обработки при помощи нейронных сетей знаний с нечеткостью или неопределенностью [27, 28].

Вместе с тем символичный подход в искусственном интеллекте прекрасно справляется с НЕ-факторами. Для этого предназначены разнообразные формализмы от метода Демпстера — Шафера [29, 30], нечеткой логики Л. Заде и до мягких и лингвистических вычислений [31]. Эти формализмы позволяют осуществить вывод

даже в случае неопределенности высокой степени, получив вполне приемлемые результаты.

Тем не менее, знания обладают еще и НЕ-факторами другого рода, которые относятся не к конкретной информации, но к знаниям в целом. Это уже упомянутая неполнота, это противоречивость и некорректность, это неадекватность и некоторые другие свойства знания [32], которые проявляются у людей в разной степени, но характеризуют каждую личность по степени ее профессионализма в той или иной области (можно быть экспертом в одной области, обладая всей доступной полнотой знания в ней, и профаном в другой, не имея в ней знаний вообще). Гибридные ИИ-системы могут объединить мощь глубинного обучения для выявления новых знаний и формальные методы обработки НЕ-факторов для работы в условиях неопределенности и неполноты информации.

Глубинное обучение до сих пор недостаточно прозрачно

О невозможности интерпретировать результаты обучения нейронных сетей уже сказано очень много. Тут вопрос не в непознаваемости, а в слишком высокой вычислительной потребности, необходимой, чтобы понять результаты настройки весовых коэффициентов на связях между нейронами, полученных в результате обучения [33]. Если сеть состоит из сотен тысяч нейронов и тысяч слоев, то количество весовых коэффициентов на связях просто циклопическое, и «дисассемблирование» результатов обучения невыполнимо с вычислительной точки зрения.

Эта проблема уже обусловила появление новой парадигмы — XAI, *eXplainable Artificial Intelligence* (объяснимый искусственный интеллект) [34]. XAI должен объяснять свои решения и результаты, а это вполне возможно для гибридных ИИ-систем, поскольку в рамках нисходящего подхода объяснение — это одна из ключевых особенностей. Символьные системы всегда могут объяснить результаты вывода, так как известны как правила вывода, так и входные данные [35]. Даже в случае обработки НЕ-факторов знаний вполне можно объяснить, как получены те или иные результаты или приняты те или иные решения.

В случае создания и использования гибридных ИИ-систем сети глубинного обучения готовят базовую информацию для принятия решений универсальным решателем. Фактически именно на этом уровне и будет заканчиваться «необъяснимое» в работе таких систем.



Но человеческое сознание работает примерно так же. Например, если человек видит кошку, то он обычно не может объяснить, как он распознал в животном именно кошку (все рационализирующие объяснения на тему «это маленькое животное с острыми ушками и полосатым хвостом» не годятся, так как это именно рационализация постфактум). Но символ «кошка» и все его над-символы могут участвовать в дальнейшем выводе, и результаты такого вывода могут быть вполне объяснены.

Глубинное обучение пока плохо интегрируется с уже существующими знаниями

Эта проблема является следствием всех перечисленных ранее. Чтобы интегрировать искусственные нейросети с уже накопленными и даже формализованными знаниями, необходимо проделать большую работу по приведению таких знаний в тот вид, который могут воспринять системы ГО. Однако многие знания уже приведены к виду, в котором их могут воспринять символьные системы, а значит, и гибридные. И работа по формализации знаний ведется именно в этом направлении — знания представляются в рамках символьных формализмов, а не размечаются для потребления нейронными сетями. Размечаются данные в каждом конкретном случае, а общие знания формализуются.

Значит, имеет смысл сосредоточиться именно на совмещении двух подходов, а это путь к гибридизации нисходящей и восходящей парадигмы. Гибридная ИИ-система по своей природе будет сразу же интегрирована со всем корпусом имеющихся знаний.

Глубинное обучение пока не способно автоматически отличать причинно-следственную связь от корреляции

Даже люди с их естественным интеллектом часто не могут отличить причинно-следственную связь от корреляции. Естественный отбор способствовал тому, что неокортекс человека склонен находить причинно-следственные связи там, где их может и не быть [36]. С точки зрения эволюции это вполне разумно, так как лучше избежать мнимой угрозы и выжить, чем долго размышлять о том, является ли одно событие прямым следствием другого, и в результате погибнуть в пасти тигра или крокодила. Так что излишняя склонность видеть причинно-следственные связи у людей, скорее всего, является инструментом уменьшения числа ошибок второго рода в распознавании образов [37].

Вместе с тем формальная логика дает все необходимые инструменты для определения причинно-следственных связей в результатах наблюдений. Чаще всего приходится использовать аппарат нечеткой логики и теории доверия, поскольку наблюдения или эксперименты обычно никогда не бывают чистыми. Так что приходится закладывать некоторую нечеткость, допуская, что в результатах есть примеси невыявленных факторов. Поэтому одновременное применение методов математической статистики (нахождение корреляций) и формальной логики (нахождение причинно-следственных связей) позволяет получить достоверный результат.

Технологии ГО с легкостью имплементируют статистические методы, но формальная логика — это вотчина символьного подхода. Поэтому для разрешения этой проблемы нейронных сетей опять требуется гибридизация парадигм.

Глубинное обучение пока хорошо работает в роли аппроксимации, но его результатам часто нельзя полностью доверять

Эта проблема является следствием сложности интерпретации результатов, получаемых на выходе нейронных сетей ГО, и предыдущей проблемы. Да, ГО неплохо справляется со статистическими моделями, как уже было показано. Аппроксимация — это упрощение, замена сложного вычислительного процесса более простым [38]. Если некоторый вычислительный процесс можно хотя бы приблизительно представить в виде произведения матриц, то методами ГО вполне можно аппроксимировать такой процесс.

Вполне возможно, что это сродни тем процессам, которые происходят внутри естественного интеллекта. Нейронные сети в нервной системе человека (и даже многих животных) вполне успешно аппроксимируют довольно непростые вычислительные задачи. Как поймать рукой бро-



➤ ➤ Лучше избежать мнимой угрозы и выжить, чем долго размышлять о том, является ли одно событие прямым следствием другого, и в результате погибнуть в пасти тигра или крокодила.

шенный мяч, быстро решить систему дифференциальных уравнений второго порядка? Есть большие сомнения в том, что какой-то биохимический процесс в нейронах решает подобные уравнения. Напротив, нейронные сети вполне успешно аппроксимируют решение таких задач после многочисленных актов обучения. Однако при этом любому человеку будет крайне сложно объяснить, как он ловит брошенный мяч.

И опять на помощь приходит гибридная парадигма. Не важно, как настроились весовые коэффициенты в нейронной сети. Если сеть обучилась успешно аппроксимировать какой-либо процесс, то на более высоком уровне абстракции такие результаты можно объяснить, применяя те или иные правила (как минимум объяснить можно так же, как эти результаты объяснил бы человек). И тогда далее можно было бы использовать аппарат какой-либо теории доверия (например, той же теории Демпстера — Шафера) для получения определенной степени доверия к выводу и его объяснению.

Глубинное обучение тяжело использовать в прикладных целях

Ну и, наконец, в статье [1] описана проблема в некотором роде упаднического характера. На самом деле проблемы тут нет. Системы ГО вполне можно использовать в различных прикладных задачах, просто сама отрасль еще слишком молода, чтобы получить прикладное применение. Но процессы уже идут. Системы машинного зрения давно используются в таких отраслях, как обеспечение безопасности и организация дорожного движения. Распознавание инцидентов на автомобильных дорогах и в зонах контроля видеонаблюдения — самые что ни на есть прикладные задачи. Рекомендации по результатам покупок или голосования в социальных сетях — тоже прикладные задачи, успешно решаемые методами глубинного ГО.

Конечно, методы ИИ еще не так широко применяются, как хотелось бы. Но ИИ войдет в жизнь человека и общества и изменит ее, часто до неузнаваемости. В частности, его влияние испытывают на себе следующие сферы [39]:

- государственное и муниципальное управление;

- обеспечение личной и общественной безопасности;
- транспорт и логистика;
- промышленность;
- образование;
- наука;
- здравоохранение;
- культура.

Можно предположить, что во всех этих сферах появятся и будут развиваться мощные системы поддержки принятия решений (СППР), которые по своей сути являются яркими примерами реализации нисходящей парадигмы [40]. Но эти СППР далее будут впитывать в себя методы ГО и другие технологии восходящей парадигмы и в конечном итоге станут гибридными СПР — системами принятия решений, из которых пресловутый человеческий фактор зачастую исключен в принципе.

Все перечисленное подводит к главной идее — для разработки систем искусственного интеллекта необходимо использовать обе парадигмы, позволяющие получить наиболее эффективные результаты. Это значит, что должны быть взяты положительные стороны обеих парадигм, а отрицательные максимально отвергнуты. Такой подход позволит создать ИИ-системы, которые будут обладать мощью глубинного обучения для распознавания образов и решения других задач машинного обучения, но при этом будут иметь возможность объяснять полученные результаты и принятые решения. В свою очередь это позволит снять многочисленные вопросы этического характера, возникающие при использовании ИИ-систем.

При этом можно считать достоверным, что другие подходы, в том числе упор на повсеместное использование искусственных нейронных сетей, могут привести к третьей «зиме» искусственного интеллекта.

□

ПЭС 18090 / 29.05.2018

Источники

1. Marcus G. Deep Learning: A Critical Appraisal [Электронный ресурс]. New York University, 2017 // Cornell University Library. URL: <https://arxiv.org/abs/1801.00631>.

2. Голованов Г. Gartner: «Хайп вокруг ИИ идет во вред технологии и потребителям» [Электронный ресурс] // Хайтек. 2017. 19 июля. URL: <https://hightech.fm/2017/07/19/ai-hype>.
3. Minsky M. *Computation: Finite and Infinite Machines*. Englewood Cliffs, N.J.: Prentice-Hall. 1967.
4. Колесников А.В. Гибридные интеллектуальные системы: Теория и технология разработки. СПб.: Изд-во СПбГТУ, 2001. 711 с.
5. Medsker L.R. *Hybrid Intelligent Systems*. Boston: Kluwer Academic Publishers, 1995. 298 p.
6. Эшби У.Р. Введение в кибернетику. М.: Изд-во иностранной литературы, 1959. 432 с.
7. Weiss G. (ed.). *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*. MIT Press, 1999.
8. Wooldridge M. *An Introduction to MultiAgent Systems*, John Wiley & Sons Ltd, 2002, paperback, 366 p.
9. Зайцев И.М., Федяев О.И. Агентно-ориентированный подход к моделированию интеллектуальных распределенных систем: Сб. Донецк: ДонНТУ, 2008. С. 337–338.
10. José Antonio Martin H.; Javier de Lope; Dario Maravall. *Adaptation, Anticipation and Rationality in Natural and Artificial Systems: Computational Paradigms Mimicking Nature*. Natural Computing. Springer, 2009. 8 (4): 757–775. doi:10.1007/s11047-008-9096-6.
11. Бир С. Мозг фирмы. М.: Радио и связь, 1993. 416 с.
12. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. М.: ДМК Пресс, 2017. 652 с.
13. Miller G.A., Chomsky N. *Pattern conception*. Paper for Conference on pattern detection, University of Michigan, 1957.
14. Каллан Р. Основные концепции нейронных сетей. М.: Издательский дом «Вильямс», 2001. 288 с.
15. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. М.: Горячая линия — Телеком, 2001. 382 с.
16. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018. 480 с.
17. Флах П. Машинное обучение. М.: ДМК Пресс, 2015. 400 с.
18. Huang S., Papernot N., Goodfellow I., Duan Y., Abbeel P. *Adversarial Attacks on Neural Network Policies*, 2017. arXiv, cs.LG.
19. Kansky K., Silver T., Mély D. A., Eldawy M., Lázaro-Gredilla M., Lou X. et al. *Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics*, 2017. arXiv, cs.AI.
20. Кондаков Н.И. Обобщение // Логический словарь-справочник. 2-е изд. М.: Наука, 1975. С. 395.
21. Душкин Р.В. Методы получения, представления и обработки знаний с НЕ-факторами. М., 2011. 115 с., ил.
22. Marcus G. *The Algebraic Mind: Integrating Connectionism and cognitive science*. Cambridge, Mass.: MIT Press, 2001.
23. Mikolov T., Chen K., Corrado G., Dean J. *Efficient Estimation of Word Representations in Vector Space*. arXiv, 2013.

References

1. Marcus G. *Deep Learning: A Critical Appraisal*. New York University, 2017. Cornell University Library, available at: <https://arxiv.org/abs/1801.00631>.
2. Golovanov G. Gartner: "Khayp vokrug II idet vo vred tekhnologii i potrebitelyam" [Gartner: Hype Around AI is Harmful for Technology and Consumers]. Khaytek, 2017, July, 19, available at: <https://hightech.fm/2017/07/19/ai-hype>.
3. Minsky M. *Computation: Finite and Infinite Machines*. Englewood Cliffs, N.J.: Prentice-Hall. 1967.
4. Kolesnikov A.V. *Gibridnye intellektual'nye sistemy: Teoriya i tekhnologiya razrabotki* [Hybrid Intelligent Systems: Theory and Technology of Development]. Saint Petersburg, Izd-vo SPbGTU, 2001, 711 p.
5. Medsker L.R. *Hybrid Intelligent Systems*. Boston: Kluwer Academic Publishers, 1995. 298 p.
6. Eshbi U.R. *Vvedenie v kibernetiku* [Introduction to Cybernetics]. Moscow, Izd-vo inostrannoy literatury, 1959, 432 p.
7. Weiss G. (ed.). *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*. MIT Press, 1999.
8. Wooldridge M. *An Introduction to MultiAgent Systems*. John Wiley & Sons Ltd, 2002, paperback, 366 p.
9. Zaytsev I.M., Fedyayev O.I. *Agentno-orientirovanny podkhod k modelirovaniyu intellektual'nykh raspredelennykh system* [Agent-Oriented Approach to Modelling Intelligent Distributed Systems]. Donetsk: DonNTU, 2008, pp. 337–338.
10. José Antonio Martin H.; Javier de Lope; Dario Maravall. *Adaptation, Anticipation and Rationality in Natural and Artificial Systems: Computational Paradigms Mimicking Nature*. *Natural Computing*. Springer, 2009. 8 (4): pp. 757–775. doi:10.1007/s11047-008-9096-6.
11. Bir S. *Mozg firmy* [Company Brain]. Moscow, Radio i svyaz', 1993, 416 p.
12. Gudfellov Ya., Bendzhio I., Kurvill' A. *Glubokoe obuchenie* [Deep Learning]. Moscow, DMK Press, 2017, 652 p.
13. Miller G.A., Chomsky N. *Pattern conception*. Paper for Conference on pattern detection, University of Michigan, 1957.
14. Kallan R. *Osnovnye kontseptsii neyronnykh setey* [Basic Concepts of Neural Networks]. Moscow, Izdatel'skiy dom "Vil'yams", 2001. 288 s.
15. Kruglov V.V., Borisov V.V. *Iskusstvennye neyronnye seti. Teoriya i praktika* [Artificial Neural Networks. Theory and Practice]. Moscow, Goryachaya liniya — Telekom, 2001, 382 p.
16. Nikolenko S., Kadurin A., Arkhangel'skaya E. *Glubokoe obuchenie* [In-depth Training]. Saint Petersburg, Piter, 2018, 480 p.
17. Flakh P. *Mashinnoe obuchenie* [Machine Learning]. Moscow, DMK Press, 2015, 400 p.
18. Huang S., Papernot N., Goodfellow I., Duan Y., Abbeel P. *Adversarial Attacks on Neural Network Policies*, 2017. arXiv, cs. LG.
19. Kansky K., Silver T., Mély D. A., Eldawy M., Lázaro-Gredilla M., Lou X. et al. *Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics*, 2017. arXiv, cs.AI.
20. Kondakov N.I. *Obobshchenie* [Generalization]. *Logicheskiy slovar'-spravochnik* [Logical Dictionary-Directory]. 2-e izd. Moscow, Nauka, 1975, p. 395.
21. Dushkin R.V. *Metody polucheniya, predstavleniya i obrabotki znaniy s NE-faktorami* [Methods for Obtaining, Presenting and Processing Knowledge with Non-Factors]. Moscow, 2011, 115 p.
22. Marcus G. *The Algebraic Mind: Integrating Connectionism and cognitive science*. Cambridge, Mass.: MIT Press, 2001.

24. Socher R., Huval B., Manning C.D., Ng A.Y. Semantic compositionality through recursive matrix-vector spaces. *Proceedings from Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, 2012.
25. Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д. *Онтологии и тезаурусы: модели, инструменты, приложения*. М.: Бином. Лаборатория знаний, 2009. 173 с.
26. Sowa J.F., Borgida A. *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, 1991.
27. Bart K. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewood Cliffs, NJ: Prentice Hall, 1992.
28. Abraham A. *Adaptation of Fuzzy Inference System Using Neural Learning*, *Fuzzy System Engineering: Theory and Practice*, Nadia Nedjah et al. (Eds.), *Studies in Fuzziness and Soft Computing*, Springer Verlag Germany, Chapter 3, pp. 53–83, 2005.
29. Dempster A.P. A generalization of Bayesian inference, *Journal of the Royal Statistical Society, series B*, vol. 30, pp. 205–247, 1968.
30. Shafer G. *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
31. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976. 166 с.
32. Душкин Р.В., Рыбина Г.В. Об одном подходе к автоматизированному извлечению, представлению и обработке знаний с НЕ-факторами. В кн.: *Известия РАН. Теория и системы управления*. № 5. М., 1999. С. 34–44.
33. Хайкин С. *Нейронные сети: Полный курс*. 2-е изд. М.: Издательский дом «Вильямс», 2006. 1104 с.
34. Knight W. DARPA is funding projects that will try to open up AI's black boxes. *MIT Technology Review*. Retrieved 2017–11–02.
35. Regina B., McCullough D., Rambow O., DeCristofaro J., Korelsky T., Lavoie B. *A New Approach to Expert System Explanations*. USAF Rome Laboratory Report.
36. Haselton M. Error Management Theory. *Journal of Personality and Social Psychology*. 2000, January.
37. Кэрролл Р.Т. *Энциклопедия заблуждений: собрание невероятных фактов, удивительных открытий и опасных поверий*. М.: Издательский дом «Вильямс», 2005. 672 с.
38. Лоран П.Ж. *Аппроксимация и оптимизация*. М.: Мир, 1975. С. 496.
39. Душкин Р.В., Жарков А.Д., Иванов Д.А. От безопасного к умному городу — Нижний Новгород: ИТ-форум – 2020 «Развитие цифрового государства. Создание систем весогабаритного контроля, интеллектуальных транспортных систем». Нижний Новгород, 2017, 12–14 апреля.
40. Терелянский П.В. *Системы поддержки принятия решений. Опыт проектирования*. Волгоград: ВолгГТУ, 2009. 127 с.

23. Mikolov T., Chen K., Corrado G., Dean J. *Efficient Estimation of Word Representations in Vector Space*, arXiv, 2013.
24. Socher R., Huval B., Manning C.D., Ng A.Y. *Semantic compositionality through recursive matrix-vector spaces. Proceedings from Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, 2012.
25. Dobrov B.V., Ivanov V.V., Lukashovich N.V., Solov'ev V.D. *Ontologii i tezaurusy: modeli, instrumenty, prilozheniya* [Ontologies and Thesauri: Models, Tools, Applications]. Moscow, Binom. Laboratoriya znaniy, 2009, 173 p.
26. Sowa J.F., Borgida A. *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, 1991.
27. Bart K. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewood Cliffs, NJ: Prentice Hall, 1992.
28. Abraham A. *Adaptation of Fuzzy Inference System Using Neural Learning*, *Fuzzy System Engineering: Theory and Practice*, Nadia Nedjah et al. (Eds.), *Studies in Fuzziness and Soft Computing*, Springer Verlag Germany, Chapter 3, pp. 53–83, 2005.
29. Dempster A.P. A generalization of Bayesian inference, *Journal of the Royal Statistical Society, series B*, vol. 30, pp. 205–247, 1968.
30. Shafer G. *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
31. Zade L. *Ponyatie lingvisticheskoy peremennoy i ego primenenie k prinyatiyu priblizhennykh resheniy* [The Concept of a Linguistic Variable and Its Application to Adopting Approximate Solutions]. Moscow, Mir, 1976, 166 p.
32. Dushkin R.V., Rybina G.V. *Ob odnom podkhode k avtomatizirovannomu izvlecheniyu, predstavleniyu i obrabotke znaniy s NE-faktorami* [One Approach to Automated Retrieval, Presentation and Processing of Knowledge with Non-Factors]. *Izvestiya RAN. Teoriya i sistemy upravleniya* [RAS Proceedings. Theory and Control Systems]. No 5. M., 1999, pp. 34–44.
33. Khaykin S. *Neyronnye seti: Polnyy kurs* [Neural Networks: Full Course]. 2-е изд. Moscow, Izdatel'skiy dom "Vil'yams", 2006, 1104 p.
34. Knight W. DARPA is funding projects that will try to open up AI's black boxes. *MIT Technology Review*. Retrieved 2017–11–02.
35. Regina B., McCullough D., Rambow O., DeCristofaro J., Korelsky T., Lavoie B. *A New Approach to Expert System Explanations*. USAF Rome Laboratory Report.
36. Haselton M. Error Management Theory. *Journal of Personality and Social Psychology*, 2000, January.
37. Kerroll R.T. *Entsiklopediya zabluzhdeniy: sobranie neveroyatnykh faktov, udivitel'nykh otkrytiy i opasnykh poveriy* [Encyclopedia of Misconceptions: Collection of Incredible Facts, Amazing Discoveries and Dangerous Beliefs]. Moscow, Izdatel'skiy dom "Vil'yams", 2005, 672 p.
38. Loran P.Zh. *Approksimatsiya i optimizatsiya* [Approximation and Optimization]. Moscow, Mir, 1975, p. 496.
39. Dushkin R.V., Zharkov A.D., Ivanov D.A. *Ot bezopasnogo k umnomu gorodu — Nizhniy Novgorod: IT-forum – 2020. "Razvitiye tsifrovogo gosudarstva. Sozdanie sistem vesogabaritnogo kontrolya, intellektual'nykh transportnykh sistem"* [From Safe to Smart City — Nizhny Novgorod: IT forum – 2020. "Digital State Development. Creating Systems of Weight-Dimensional Control, Intelligent Transport Systems"]. Nizhniy Novgorod, 2017, April, 12–14.
40. Terelyanskiy P.V. *Sistemy podderzhki prinyatiya resheniy. Opyt proektirovaniya* [Decision Support Systems. Design Experience]. Volgograd, VolgGTU, 2009, 127 p.