

Ученый журнал

РЕЧЕВЫЕ ТЕХНОЛОГИИ



Speech technology

1

2024



Речевые

ТЕХНОЛОГИИ

1/2024

Главный редактор: Харламов Александр Александрович,
доктор технических наук, kharlamov@analyst.ru

Состав редколлегии:

Заместитель главного редактора: **Потапова Родмонга Кондратьевна,**
доктор филологических наук, профессор, rkpotarova@yandex.ru

Азаров Илья Сергеевич, доктор технических наук, профессор,
azarov@bsuir.by, Беларусь

Голенков Владимир Васильевич, доктор технических наук, профессор,
golen@bsuir.by, Беларусь

Дмитриев Владимир Тимурович, кандидат технических наук,
vol77@rambler.ru

Железны Милош, PhD,
zelezny@kky.zcu.cz, Чехия

Жигулёвцев Юрий Николаевич, кандидат технических наук,
ynzh@mail.ru

Карпов Алексей Анатольевич, доктор технических наук,
karпов@iias.spb.ru

Коренева Ольга, PhD,
quovadis36@gmail.com, Испания

Кудубаева Сауле Альжановна, кандидат технических наук,
saule_58@mail.ru, Казахстан

Кушнир Дмитрий Алексеевич, кандидат технических наук,
kushdal@yandex.ru

Лобанов Борис Мефодьевич, доктор технических наук,
lobanov@newman.bus-net.by, Беларусь

Ляксо Елена Евгеньевна, доктор биологических наук,
lyakso@gmail.com

Максимов Евгений Михайлович, доктор технических наук,
maximovet@inbox.ru

Мещеряков Роман Валерьевич, доктор технических наук, профессор,
mriv@ieee.org

Пильгун Мария Александровна, доктор филологических наук,
pilgun@yandex.ru

Ронжин Андрей Леонидович, доктор технических наук,
ronzhin@iias.spb.su

Славкова Светлана, PhD,
svetlana.slavkova@unibo.it, Италия

Смирнов Иван Валентинович, кандидат физико-математических наук,
ivs@isa.ru

Сулейманов Джавдет Шевкетович, академик Академии наук Татарстана,
профессор, alsu_73@list.ru

Чучупал Владимир Яковлевич, кандидат технических наук,
v.chuchupal@gmail.com

Содержание

Аверкин А. Н.

Объяснительный искусственный интеллект в больших речевых моделях 3

Джунковский А. В.

Искусственный интеллект как новая форма CAT. Путь к идеальной автоматизации перевода или новый вызов? 14

Катаев М. Ю.

Методика распознавания речевых команд в школьных информационных системах 18

Кулай А. А.

За пределами семантики, или Семантика Бытия и Небытия 35

Мещеряков Р. В.

Семантика на базе технологий искусственного интеллекта для инженера 43

Расходчиков А. Н.

Семантика коммуникативной среды города – информационные технологии городского дизайна 49

Харламов А. А.

ChatGPT – это интеллектуальная система? 58

XII Международная научно-практическая конференция

«Интегрированные модели и мягкие вычисления в искусственном интеллекте», ИММВ-2024 (14–17 мая 2024 г., г. Коломна)

Круглый стол «Эволюция и перспективы развития нейросетей». 69

Открытая дискуссия «Мифы и реальности искусственного интеллекта» 90

Редакция:

Редактор: *Татьяна Иванова*

Корректор: *Людмила Асанова*

Дизайн: *Анна Ладанюк*

Вёрстка: *Евгения Соболев*

Адрес редакции: 109341, Москва, ул. Люблинская, д. 157, корп. 2.

Тел.: +7 495 345 52 00

Электронная почта: narob@yandex.ru

Сайт: www.narodnoe.org

Подписано в печать 15.08.2024. Формат 60·90%. Бумага офсетная. Печать офсетная.

У.-п.л. 12. Заказ № 240817. Тираж 100 экз. Издательский дом «Народное образование».

Отпечатано в типографии ИД "Народное образование".

109341, Москва, ул. Люблинская, д. 157, корп. 2. Тел.: +7 495 345 52 00/59 00.

Объяснительный искусственный интеллект в больших речевых моделях¹

Аверкин А. Н., кандидат физико-математических наук, доцент, ведущий научный сотрудник Российского экономического университета им. Г. В. Плеханова, Москва, averkin2003@inbox.ru

Проводится краткий обзор и анализ существующих методов объяснимого искусственного интеллекта (ОИИ) в больших языковых моделях (БЯМ). ОИИ является основной частью искусственного интеллекта (ИИ) третьего поколения. Объяснимость является важным аспектом БЯМ, которые позволяют нам понять логику, лежащую в основе их решений. Для достижения объяснимости в БЯМ было разработано большое количество методов, дающих ценную информацию об их внутренней работе. Статья является продолжением работы [1], в которой были рассмотрены методы ОИИ для систем ИИ для обработки цифровых изображений.

объяснительный искусственный интеллект • большие языковые модели • трансформер • экспертные системы • искусственные нейронные сети

ВВЕДЕНИЕ

В сентябре 2019 года при принятии Стратегии ИИ президент РФ отметил, что страны с развитием искусственного интеллекта получают «преимущества, не сравнимые с ядерным оружием». И подчеркнул, что Россия имеет все шансы в этом преуспеть. Один из основных принципов развития и использования технологий искусственного интеллекта (ИИ), приведенных в Национальной стратегии развития искусственного интеллекта на период до 2030 г., является прозрачность: объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей к продуктам, созданным с использованием технологий искусственного интеллекта, к информации о применяемых в этих продуктах алгоритмах работы искусственного интеллекта. Объяснительный искусственный интеллект (ОИИ) играет ключевую роль в развитии и применении систем искусственного интеллекта в экономике данных, способствуя их прозрачности, пониманию и доверию со стороны пользователей и заинтересованных сторон. ОИИ имеет решающее значение, особенно в приложениях, где решения, принимаемые с помощью ИИ, влияют на людей или предприятия. Растущая потребность в подотчетности и этичности ИИ, особенно в таких отраслях, как финансы, здравоохранение и юриспруденция, повышает спрос на объяснимый ИИ.

¹ Исследование выполнено за счет Российского научного фонда [грант № 22-71-10112]. <https://rscf.ru/project/22-71-10112>



Каждое десятилетие в технологиях происходят революционные сдвиги, которые становятся новыми платформами, на которых строятся прикладные технологии. Так, искусственный интеллект перешел от экспертного обучения первого поколения и баз знаний, созданных вручную, к глубокому обучению второго поколения, использующего нейросети и большие обучающие выборки.

Теперь мы вступаем в третье поколение ИИ, где система искусственного интеллекта может интерпретировать и объяснить алгоритм принятия решений, даже если он имеет природу черного ящика. Объяснимый искусственный интеллект является основной частью третьего поколения ИИ. В 2030-х годах мы увидим ИИ четвертого поколения с машинами, которые сами будут обучаться и динамически накапливать новые знания и навыки. К 2040-м годам ИИ пятого поколения увидит системы искусственного интеллекта с воображением, которые больше не будут полагаться на людей в обучении.

Растущая сложность моделей ИИ и их приложений повышает спрос на объяснимость, обеспечивающую понятность и обоснованность решений. Ожидается, что постоянные инновации в области разработки новых решений в области объяснимого искусственного интеллекта поставщиками рынка создадут возможности для его роста.

Объем мирового рынка объяснимого ИИ в 2022 г. составил 6,83 млрд долларов США, и ожидается, что к 2032 г. он достигнет около 36,42 млрд долларов США, увеличиваясь в среднем на 18,22 % в течение прогнозируемого периода с 2023 по 2032 г. Объем рынка по регионам: Северная Америка — 41 %, Европа — 30 %, Азия и Тихоокеанский регион — 24 %, Латинская Америка — 3 %, Средний Восток и Африка — 2 %.

ОБЪЯСНЕНИЕ В БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЯХ

С момента появления в открытом доступе первой БЯМ, как совершенно нового класса архитектур искусственных нейронных сетей (ИНС), ChatGPT в ноябре 2022 г., произошёл экспоненциальный рост использования БЯМ в различных цифровых решениях, отразившихся на всех сферах жизни современного человека. К концу 2023 г. БЯМ стали мейнстримом в области технологий ИИ. С ростом количества моделей БЯМ, появлением версий, способных работать с мультимодальными данными, всё более предпочтительным выглядит построение гибридных интеллектуальных систем с их использованием.

БЯМ продемонстрировали впечатляющие возможности в обработке естественного языка. Однако их внутренние механизмы до сих пор плохо интерпретируются и недостаточно прозрачны, что создает нежелательные риски для последующих приложений. Поэтому понимание и объяснение этих моделей имеет решающее значение для выяснения их поведения, ограничений и социальных последствий. В этой статье мы вводим таксономию методов объяснимости и предоставляем структурированный обзор методов объяснения языковых моделей

на основе трансформеров. Мы рассмотрим основные методы, основанные на парадигмах обучения БЯМ: традиционную парадигму, основанную на тонкой настройке, и парадигму, основанную на промтах. Для каждой парадигмы мы суммируем цели и основные подходы к генерации локальных объяснений отдельных прогнозов и глобальных объяснений на основе общих знаний о модели. Мы также обсудим метрики для оценки сгенерированных объяснений и то, как их можно использовать для отладки моделей и повышения производительности. Наконец, мы рассматриваем ключевые проблемы и новые возможности для методов объяснения в эпоху БЯМ в сравнении с традиционными моделями машинного обучения.

С годами языковые модели претерпели значительные изменения. Начиная с традиционных n-граммных моделей и заканчивая современными трансформерами, мы стали свидетелями экспоненциального роста их возможностей. Внедрение архитектуры трансформеров, особенно таких моделей, как BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer) и других, раздвинуло границы обработки естественного языка до новых высот.

Потребность объяснимости в БЯМ обусловлена растущей сложностью и внедрением систем на основе ИИ. По мере того как БЯМ становятся все более изощренными, они часто действуют как «черные ящики», скрывая свои внутренние процессы принятия решений от человеческого понимания. Отсутствие прозрачности создает серьезные проблемы, особенно в таких критически важных секторах, как здравоохранение, финансы и юридические отрасли, где объяснимость имеет решающее значение для доверия и подотчетности.

Давайте подробнее остановимся на необходимости объяснимости в БЯМ с помощью таблицы 1, использующей таксономию [2].

Понимая проблемы, связанные с «черным ящиком» и влиянием предвзятости в системах ИИ, разработчики и специалисты по обработке и анализу данных могут работать над созданием более прозрачных и справедливых БЯМ. Объяснимость — это не просто техническая характеристика; она имеет основополагающее значение для создания ответственных и доверительных приложений ИИ, которые соответствуют человеческим ценностям и потребностям общества. В таблице 1 подчеркивается важность этих аспектов в критически важных секторах, таких как здравоохранение, финансы и юриспруденция, где последствия решений ИИ могут существенно повлиять на отдельных людей и сообщества. Внедряя принципы объяснимости, мы можем раскрыть весь потенциал БЯМ, сохраняя при этом справедливость, подотчетность и доверие к системам ИИ.

По мере того как БЯМ становятся все более сложными, они часто функционируют как «черные ящики», что затрудняет понимание их внутренней работы. Отсутствие прозрачности вызывает озабоченность, особенно в критически важных приложениях, таких как здравоохранение, финансы и юридический сектор, где объяснимость имеет первостепенное значение. Доверие и подотчетность имеют важнейшее значение при развертывании систем ИИ, и именно здесь большую роль играет объяснимость.

Таблица 1

Важность объяснимости в БЯМ

Задачи	Описание	Важность объяснимости в прикладных задачах
Дилемма черного ящика	БЯМ работают как «черные ящики», что затрудняет понимание их внутренних процессов	Объяснимость дает представление о процессе принятия БЯМ
	Отсутствие объяснимости может привести к недоверию, препятствуя внедрению систем ИИ в критически важных секторах	В здравоохранении, финансах и юриспруденции, где подотчетность имеет решающее значение, объяснимость имеет наиболее важное значение
	Системы ОИИ укрепляют доверие пользователей, способствуя широкому внедрению ИИ	Понимание того, что БЯМ принимают решения, способствует принятию и внедрению решений на основе ИИ
Борьба с предвзятостью	Модели ИИ, в том числе БЯМ, могут непреднамеренно увековечить предвзятость, присутствующую в обучающих данных	ОИИ помогает выявлять предвзятые решения, обеспечивая справедливость и равенство результатов
	Системы ИИ могут приводить к дискриминационным результатам, ставя в невыгодное положение определенные группы и отдельных лиц	Выявляя предубеждения, разработчики могут принимать корректирующие меры, приводя приложения ИИ в соответствие с этическими нормами
	Обеспечение справедливости в области ИИ является общественным императивом, требующим особого внимания к объяснимости и подотчетности	ОИИ способствует обеспечению равных возможностей и гарантирует, что ИИ принесет пользу всем слоям общества

Еще одним важным аспектом является предвзятость в системах ИИ. БЯМ могут непреднамеренно унаследовать предвзятость от обучающих данных, что приводит к нежелательным и дискриминационным результатам. При помощи объяснимости систем ИИ, разработчики и специалисты по обработке и анализу данных могут эффективно выявлять и устранять предвзятость, обеспечивая справедливость и доверие в приложениях ИИ.

Объяснимость является важным аспектом БЯМ, который позволяет нам понять логику, лежащую в основе их решений. Для достижения объяснимости в БЯМ было разработано несколько методов, дающих необходимую информацию об их внутренней работе. Некоторые из этих методов рассмотрены в таблице 2.

Таблица 2

Методы объяснимости в БЯМ

Техника	Описание	Преимущества
Механизмы внимания	Позволяет сосредоточиться на определенных словах или частях предложения во время обработки. Визуализация весовых коэффициентов внимания повышает прозрачность модели	Определяет важные части входных данных, помогает понять, как модель принимает решения
Задачи зондирования	Проверяет понимание лингвистических свойств моделью, демонстрируя представление о ее знаниях	Помогает интерпретировать процесс принятия решений моделью на основе понимания языка
Послойный анализ	Анализирует выходные данные каждого слоя для понимания иерархической обработки	Показывает пошаговое преобразование входных данных, помогая интерпретировать решения на каждом уровне
Объяснения, основанные на правилах	Создает интерпретируемые правила, объясняющие поведение модели	Предоставляет удобочитаемые объяснения для прогнозов модели, способствуя пониманию и доверию
Методы, основанные на возмущениях	Немного изменяет входные данные, чтобы наблюдать за изменениями в прогнозах модели	Выявляет важные признаки, понимая, насколько модель зависит от конкретных входных данных
Контрфактические объяснения	Предоставляет альтернативные входные данные для наблюдения за изменениями в прогнозах	Дает представление о границах принятия решений модели и чувствительности к изменениям входных данных

Эти методы объяснимости БЯМ предоставляют современный инструментарий для понимания и объяснения решений, принимаемых языковыми моделями ИИ. Каждый метод обладает уникальными преимуществами и дополняет другие, позволяя разработчикам и исследователям получить более глубокое представление о БЯМ и способствовать прозрачности и доверию к системам, управляемым искусственным интеллектом. Рассмотрим их немного подробнее.

Механизмы внимания в БЯМ позволяют им фокусироваться на определенных словах или частях предложения во время обработки. Визуализируя весовые коэффициенты внимания, мы можем понять, какие части входных данных модель считает наиболее релевантными для прогнозирования. Этот метод повышает прозрачность модели и помогает определить причины, лежащие в основе ее выходных данных.

Задачи зондирования включают в себя проверку понимания моделью различных лингвистических свойств, таких как синтаксис, семантика и тональность. Оценивая производительность модели в этих задачах, мы получаем представление о ее представлении знаний и узнаем о лингвистических особенностях, которые она



фиксирует. Это позволяет нам интерпретировать процесс принятия решений моделью на основе ее понимания языка.

При послойном анализе учитывается, что БЯМ состоят из нескольких слоев трансформера, каждый из которых обрабатывает входные данные по-разному. Анализ выходных данных каждого слоя может пролить свет на то, как модель обрабатывает информацию иерархически. Этот метод позволяет нам понять пошаговую трансформацию входных данных, помогая нам интерпретировать решения, принятые на каждом уровне.

Объяснения, основанные на правилах, включают в себя создание интерпретируемых правил, объясняющих поведение модели. Эти правила можно извлечь, проанализировав закономерности в данных и поняв, как модель присваивает важность различным объектам. Объяснения, основанные на правилах, предоставляют наиболее удобочитаемые объяснения прогнозов модели, способствуя лучшему пониманию и доверию.

Методы, основанные на возмущениях, включают в себя небольшое изменение входных данных и наблюдение за тем, как изменяются прогнозы модели. Возмущая входные данные и измеряя влияние на выход, мы можем определить признаки, которые существенно влияют на решения модели. Этот метод помогает понять, на какие части входных данных модель больше всего полагается при составлении прогнозов.

Контрфактические объяснения включают в себя предоставление альтернативных входных данных для модели и наблюдение за тем, как меняются прогнозы. Изучив, какие изменения во входных данных приведут к различным выходам, мы можем понять конкретные условия, при которых прогнозы модели могут измениться. Контрфактические объяснения дают ценную информацию о границах принятия решений в модели и ее чувствительности к изменениям входных данных. Используя эти методы, разработчики могут не только улучшить объяснимость БЯМ, но и решить этические и практические проблемы, связанные с внедрением моделей ИИ в реальные приложения. По мере того как ИИ продолжает внедряться в различные области, важность объяснимости становится все более значимой для укрепления доверия и обеспечения ответственного применения технологий ИИ.

ОБЪЯСНЕНИЯ ДЛЯ ДВУХ ПАРАДИГМ ОБУЧЕНИЯ БЯМ

Обучение БЯМ можно разделить на две парадигмы: традиционную тонкую настройку и подсказку, в зависимости от того, как они используются для адаптации к последующим задачам. Из-за существенных различий между двумя парадигмами в таксономии [3] были предложены различные типы объяснений (табл. 3).

Таблица 3

Общая структура механизмов объяснения в БЯМ

Объясни- мость БЯМ	Тради- ционная парадигма тонкой на- стройки	Локальное объяснение	Объяснение атрибуции объекта
			Объяснение, основанное на внимании
			Объяснение на основе примеров
			Объяснение на естественном языке
		Глобальное объяснение	Объяснение, основанное на исследовании
			Объяснение активации нейронов
			Объяснение, основанное на понятиях
			Механистическая интерпретируемость
	Использование объяснений	Отладочные модели	
		Совершенствование моделей	
	Парадигма промтов	Базовая модель	Объяснение обучения в контексте
			Объяснения в цепочке мыслей
			Инженерия представлений
		Вспомогатель- ная модель	Объяснение роли точной настройки
			Галлюцинации и неуверенность
		Модель, использующая объяснение	Повышение точности БЯМ
Приложения нижнего уровня			
Оценка объяснений	Парадигма тонкой настройки		
	Парадигма промтов		

ТРАДИЦИОННАЯ ПАРАДИГМА ТОНКОЙ НАСТРОЙКИ

Рассмотрим традиционную парадигму тонкой настройки. В этой парадигме языковая модель сначала предварительно обучается на большом корпусе неразмеченных текстовых данных, а затем настраивается на наборе размеченных данных из определенной нижестоящей области, такой как SST-2, MNLI и QQP в бенчмарке GLUE [4]. Во время тонкой настройки легко добавить полносвязные слои поверх конечного слоя кодировщика языковой модели, что позволяет ей адаптироваться к различным последующим задачам. Эта парадигма показала успех для языковых моделей среднего размера, обычно содержащих до миллиарда параметров, таких как BERT, RoBERTa, ELECTRA, DeBERTa и др. Модели объяснения этой парадигмы сосредоточены на двух ключевых областях:

- 1) понимание того, как самоконтролируемое предварительное обучение позволяет моделям приобрести базовое понимание языка (например, синтаксиса, семантики и контекстуальных отношений);
- 2) анализ того, как процесс тонкой настройки дает этим предварительно обученным моделям возможность эффективно решать последующие задачи.

Традиционная парадигма тонкой настройки делится на локальное объяснение и глобальное объяснение.

Первая категория объяснений — локальное объяснение — относится к объяснению прогнозов, генерируемых БЯМ. Рассмотрим сценарий, в котором у нас есть языковая модель, и мы вводим в нее определенный текст. Затем модель создает выходные данные классификации, например классификацию тональности или прогноз для следующего маркера. В этом сценарии роль объяснения состоит в том, чтобы прояснить процесс, с помощью которого модель сгенерировала конкретную классификацию или прогноз маркера. Поскольку цель состоит в том, чтобы объяснить, как БЯМ делает прогноз для конкретных входных данных, мы называем это локальным объяснением. Эта категория включает в себя четыре основных потока подходов к созданию объяснений, включая объяснение на основе атрибуции признаков, объяснение на основе внимания, объяснение на основе примеров и объяснение на естественном языке.

В отличие от локальных объяснений, которые направлены на объяснение индивидуальных предсказаний модели, глобальные объяснения дают представление о внутренней работе языковых моделей. Глобальные объяснения направлены на то, чтобы понять, что закодировали отдельные компоненты (нейроны, скрытые слои и более крупные модули), и объяснить знания, усвоенные отдельными компонентами. Метод рассматривает три основных подхода для девяти глобальных объяснений: методы зондирования, которые анализируют представления и параметры модели, анализ активации нейронов для определения отзывчивости модели на входные данные, и методы, основанные на концепциях.

ПАРАДИГМА ПРОМТОВ

По мере увеличения масштаба языковых моделей они демонстрируют новые способности, такие как малое обучение, то есть способность усваивать концепции всего на нескольких примерах. Они также демонстрируют способность промтов на основе цепочки мыслей, которая позволяет скормить модели последовательности промтов, чтобы направить ее генерацию в определенном направлении и заставить объяснить свои рассуждения. Учитывая эти новые свойства, исследование объяснимости преследует три основные цели: 1) понимание того, как эти большие языковые модели могут так быстро справляться с новыми задачами на ограниченных примерах, что помогает конечным пользователям интерпретировать рассуждения модели, 2) объяснение цепочек мыслей и 3) инженерия представлений.

Парадигма промтов включает в себя использование промтов, таких как предложения на естественном языке с пробелами, которые модель должна заполнить, чтобы обеспечить обучение с нулевым или небольшим количеством выстрелов без необходимости дополнительных обучающих данных. В парадигме промтов БЯМ продемонстрировали впечатляющие способности к рассуждению, включая обучение за несколько

этапов, способность к построению цепочки мыслей и такие явления, как галлюцинации, которые отсутствуют в обычной парадигме тонкой настройки. Объяснение в парадигме промтов можно разделить на две группы в соответствии с этапами разработки модели: объяснение базовой модели и объяснение модели помощника.

В базовой модели, по мере того как БЯМ увеличиваются в размерах и обучающих данных, они демонстрируют впечатляющие новые возможности, не требуя дополнительных обучающих данных. Одной из таких возможностей является обучение с помощью промтов. Этот тип парадигмы обычно работает на языковых моделях огромного размера (с миллиардами параметров), таких как GPT-3, LLaMA-1 [5], LLaMA-2 [6], Falcon [7]. Эти модели называются базовыми моделями или базовыми моделями, которые могут общаться с пользователями без дальнейшего согласования с предпочтениями человека. У базовых моделей есть два основных ограничения:

- 1) они не могут следовать инструкциям пользователя, поскольку данные перед обучением содержат мало примеров инструкций и ответов,
- 2) они склонны генерировать предвзятый и токсичный контент.

Чтобы устранить эти ограничения, базовые модели дополнительно дорабатываются с помощью контролируемой тонкой настройки для достижения способностей человеческого уровня, таких как открытый диалог. Ключевая идея состоит в том, чтобы согласовать ответы модели с обратной связью и предпочтениями человека. Наиболее типичным способом для этого процесса является настройка инструкций с помощью демонстрационных пар (подсказки, ответы) и обучение с подкреплением на основе обратной связи с человеком. Модели обучаются с обратной связью на естественном языке для ведения сложных, многоходовых разговоров. К этому семейству относятся модели GPT-3.5 и GPT4 [8] от OpenAI [9]. Эти модели, собственно, и называются вспомогательными моделями-помощниками, чат-помощниками или диалоговыми моделями. Объяснения для моделей-помощников сосредоточены на понимании того, как модели обучаются открытому интерактивному поведению из разговоров.

Благодаря широкомасштабному предварительному обучению без учителя и тонкой настройке с учителем, БЯМ, принадлежащие к этой парадигме, обладают сильными способностями к рассуждению. Тем не менее их масштаб также делает их восприимчивыми к генерации проблемных выходов, таких как галлюцинации. Исследование объяснимости здесь направлено на то, чтобы: 1) выяснить роль тонкой настройки выравнивания, 2) проанализировать причины галлюцинаций и 3) количественно оценить неопределенность.

Использование пояснений относится к использованию объяснительных возможностей БЯМ на основе подсказок для повышения прогностической производительности языковых моделей и обеспечения полезных приложений. Совершенствование БЯМ направлено на изучение того, могут ли БЯМ извлечь пользу из объяснений при изучении новых задач на ограниченных примерах. В частности, исследуется, может ли предоставление объяснений для ответов на несколько заданий повысить производительность модели в этих задачах [10]. Предусмотрены две формы объяснений: пояснения до ответа и пояснения после ответа. В [11] предлагается метод, называемый промтом цепочки мыслей, который обеспечивает промежуточные шаги рассуждения в виде объяснений в промтах перед ответами.

ОЦЕНКИ ОБЪЯСНЕНИЙ

Оценки объяснений даются с точки двух аспектов: правдоподобности и верности. Оба аспекта посвящены универсальным свойствам и метрикам, которые могут быть применены для сравнения различных подходов к объяснению. Обычно используются количественные оценочные свойства и метрики, которые, как правило, более надежны, чем качественные оценки.

Распространенные методы включают визуализацию голов внимания для одного входного сигнала с использованием двудольных графов или тепловых карт. Эти два метода представляют собой просто разные представления внимания, один в виде графика, а другой в виде матрицы. Системы визуализации различаются своей способностью отображать взаимосвязи в нескольких масштабах, представляя внимание в различных формах для разных моделей. На уровне входных данных оценки внимания для каждой пары слово/лексема/предложение между предложением-посылкой и предложением-предположением позволяют оценить достоверность предсказания модели [12]. На уровне нейронов отдельные головы внимания могут быть проанализированы для понимания поведения модели [13].

ЗАКЛЮЧЕНИЕ

Объяснимость — это важнейший аспект БЯМ, который позволяет нам понять логику, лежащую в основе их решений и прогнозов. Обеспечение объяснимости в БЯМ имеет важное значение для укрепления доверия, повышения прозрачности и обеспечения ответственности систем ИИ. Делая БЯМ объяснимыми, разработчики и специалисты по обработке и анализу данных могут представить пользователям и заинтересованным сторонам значимую информацию о том, как модели достигают конкретных результатов.

Объяснимость в БЯМ предлагают несколько существенных преимуществ, которые способствуют ответственному и эффективному использованию технологии ИИ. Объяснимость способствует доверию между пользователями и заинтересованными сторонами. Когда результаты, генерируемые искусственным интеллектом, понятны и объяснимы, пользователи чувствуют себя более комфортно, полагаясь на технологию для принятия критически важных решений. Способность понимать, как модель приходит к своим прогнозам, обеспечивает прозрачность и подотчетность, укрепляя доверие к системам ИИ. В результате пользователи с большей вероятностью примут и внедрят технологии ИИ в различных областях, что приведет к более широкому принятию и интеграции ИИ в общество.

Список использованных источников

1. Аверкин А. Н. Объяснимый искусственный интеллект как часть искусственного интеллекта третьего поколения // Речевые технологии. – 2023. – № 1. – С. 4–10. Индекс: 62203 ISSN 2305-8129

2. *Sonali, Pawar* (2023) Analyzing LLMs: Interpretability and Explainability. <https://astconsulting.in/artificial-intelligence/nlp-natural-language-processing/llm/analyzing-llms-interpretability-and-explainability/>
3. *Zhao, Haiyan & Chen, Hanjie & Yang, Fan & Liu, et al.* (2023). Explainability for Large Language Models: A Survey. (https://www.researchgate.net/publication/373686370_Explainability_for_Large_Language_Models_A_Survey)
4. *Wang A., Singh A., Michael J., et al.* Glue: A multitask benchmark and analysis platform for natural language understanding. International Conference on Learning Representations (ICLR), 2019.
5. *Touvron H., Lavril T., Izacard G., et al.* Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971, 2023a.
6. *Touvron H., Martin L., Stone K., et al.* Llama-2: Open foundation and finetuned chat models. 2023b. URL <https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models/>
7. *Almazrouei E., Alobeidli H., Alshamsi A. et al.* Falcon-40b: an open large language model with state-of-the-art performance. 2023.
8. *Bubeck S., Chandrasekaran V., Eldan R. et al.* Sparks of artificial general intelligence: Early experiments with -4. arXiv preprint arXiv:2303.12712, 2023.
9. Anthropic. Decomposing language models into understandable components? <https://www.anthropic.com/index/decomposing-language-models-into-understandable-components>, 2023.
10. *Lampinen A., Dasgupta I., Chan S. et al.* Can language models learn from explanations in context? In Findings of the Association for Computational Linguistics: EMNLP 2022, pp. 537–563, 2022.
11. *Wei J., Wang X., Schuurmans D. et al.* Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35:24824–24837, 2022.
12. *Jesse Vig.* BertViz: A Tool for Visualizing Multi-Head Self-Attention in the BERT Model. 2019.
13. *Jauret T., Kervadec C., Vuillemot R. et al.* VisQA: X-raying Vision and Language Reasoning in Transformers, July 2021. URL <http://arxiv.org/abs/2104.00926>. arXiv:2104.00926 [cs].

EXPLAINABLE ARTIFICIAL INTELLIGENCE IN LARGE LANGUAGE MODELS

Averkin A. N., Ph.D., Associate Professor, Leading researcher of Plekhanov Russian University of Economics, Moscow, averkin2003@inbox.ru

A brief review and analysis of existing methods of explicable artificial intelligence (XAI) in large language models (LLM) is carried out. XAI is a core part of the third generation artificial intelligence (AI). Explainability is an important aspect of LLM that allows us to understand the logic behind their decisions. To achieve explainability, LLM have developed a large number of methods that provide valuable information about their internal work. The article is a continuation of the work [1], in which XAI methods for AI systems for digital image processing were considered.

explanatory artificial intelligence • large language models • transformer • expert systems • artificial neural networks



Искусственный интеллект как новая форма CAT. Путь к идеальной автоматизации перевода или новый вызов?

Джунковский А. В., кандидат филологических наук, заведующий кафедрой прикладной и экспериментальной лингвистики ФАЯ, старший научный сотрудник экспериментально-фонетической лаборатории криминалистики по речеведению Московского государственного лингвистического университета, Москва, vetinari01@gmail.com

В статье рассматривается современная проблема использования искусственного интеллекта в контексте автоматизированного перевода. Описываются проблемы, с которыми сталкиваются преподаватели-переводчики в процессе обучения студентов в контексте повсеместного распространения удобных для конечного пользователя и дешёвых интерфейсов доступа к технологиям искусственного интеллекта. Проводится демонстрация результативности перевода системой ChatGPT и делаются выводы о существовании эмпирического фундамента для опасений преподавателей-переводчиков.

прикладная лингвистика • автоматизированный перевод • искусственный интеллект • преподавание перевода • ChatGPT

В настоящее время возникла ситуация, в которой системы искусственного интеллекта стали простыми в использовании и дешёвыми для конечного пользователя. Одна из задач, которая может решаться с помощью искусственного интеллекта, — автоматизированный и автоматический перевод. Преподаватели перевода высказывают опасения, что такой простой доступ может негативно сказаться на компетенциях выпускников-переводчиков. В этой статье мы рассмотрим эту проблему подробно.

Установим терминологический аппарат. В рамках прикладной лингвистики принято подразделять автоматизированный перевод (англ. Computer-assisted translation, CAT) и автоматический перевод (англ. Machine translation, MT) [1].

Автоматизированный перевод подразумевает пост-обработку текста специалистом. Программа получает текст на исходном языке, предлагает переводчику варианты перевода сегментов (слов, предложений), альтернативы и комментарии. Часто для этой цели используют технологию

Translation Memory (рус. «переводческая память»), достаточно простую базу данных, сохраняющую сегменты, ранее переведённые специалистом вручную и предлагающие их в следующий раз, когда слово возникает в тексте. Сущность CAT заключается в упрощении работы специалиста, а не его замене.

Автоматический перевод не подразумевает пост-обработку текста. Программа получает текст на исходном языке, выдаёт текст на целевом языке.

Хотя сами переводчики положительно высказываются об ИИ-переводе [3], существует другая точка зрения, высказываемая преподавателями. Опасения преподавателей перевода заключаются в том, что современные системы искусственного интеллекта, например ChatGPT, могут быть использованы студентами-переводчиками для создания переведённых текстов, неотличимых от таковых, переведённых специалистом [5]. Опасность такого академического подлога заключается в том, что студент получает удовлетворительную (и выше) оценку, при этом не освоив навык перевода, и, выходя на рынок труда, может потенциально создать опасные ситуации при осуществлении перевода политической и дипломатической коммуникации, технических и медицинских инструкций и спецификаций, текстов правил и законов.

В связи с этим в настоящее время существует запрос на развитие технологий обнаружения признаков перевода, осуществлённого искусственным интеллектом. Интересным представляется то, что сами системы искусственного интеллекта могут обнаруживать признаки перевода, осуществленного искусственным интеллектом, однако, возможно, окажется целесообразным создание автономного специализированного лингвистического ПО для решения этой задачи.

Мы проверили качество перевода с русского языка на английский лично, используя новый ChatGPT 3.5 через интерфейс ChadGPT [4] (рис. 1).

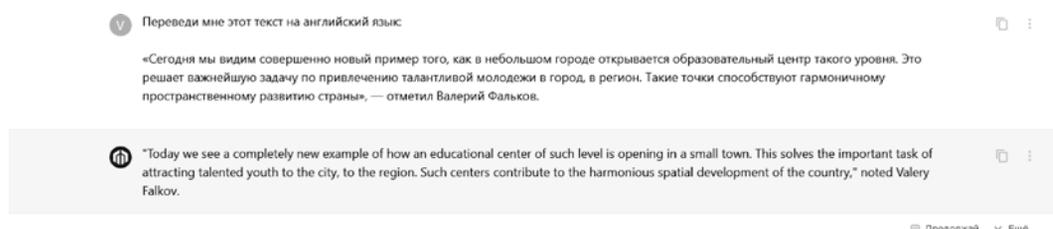


Рис. 1. Пример перевода с русского языка на английский системой ChatGPT 3.5

В качестве исходного текста мы взяли текст абзаца из новости с официального сайта Министерства науки и высшего образования Российской Федерации. На момент проверки новость была опубликована менее чем 48 ч назад и не обладала официальной версией перевода. Также текст был выбран в связи с тем, что не обладает высоким аналитическим приоритетом для иностранных think-tanks, исследовательских центров и новостных агентств.

Русскоязычная версия текста: «Сегодня мы видим совершенно новый пример того, как в небольшом городе открывается образовательный центр такого уровня. Это решает важнейшую задачу по привлечению талантливой молодежи в город,



в регион. Такие точки способствуют гармоничному пространственному развитию страны», — отметил Валерий Фальков» [2].

Перевод ИИ: Today we see a completely new example of how an educational center of such level is opening in a small town. This solves the important task of attracting talented youth to the city, to the region. Such centers contribute to the harmonious spatial development of the country, noted Valery Falkov.

С семантической точки зрения смысл передан верно, однако отметим, что на лексическом и синтаксическом уровне перевод субоптимален. Так, выбор лексем *educational center* кажется сомнительным, более удачным вариантом перевода было бы *educational facility*. *Example* слишком буквально, лучше подошёл бы *instance*. Фраза *of such level* стилистически не отвечает требованиям лаконичности и оптимального порядка слов. Фраза *to the city, to the region* была бы лучше передана через использование союза *and* вместо запятой. В рамках одного абзаца в одном случае ИИ переводит город как *town*, в другом — как *city*, что искажает смысл. Версия текста на английском смещает акцент, так как *city* обладает устойчивой коннотацией с метрополисом в английском языке.

Анализ качества перевода, осуществляемого искусственным интеллектом, позволяет сделать вывод, что опасения преподавателей перевода обладают эмпирическим фундаментом. Искусственный интеллект всё ещё допускает значительные искажения даже в случаях перевода с русского языка на английский. Если в процессе обучения переводу студенты будут использовать ИИ как систему CAT, это может привести к падению их профессиональных компетенций.

В связи с этим мы полагаем, что целесообразно создание автономного специализированного лингвистического ПО для решения задачи выявления признаков перевода, осуществленного искусственным интеллектом.

Список использованных источников

1. Баранов, А. Н. Введение в прикладную лингвистику / А. Н. Баранов. – Москва: Ленанд, 2021. – 368 с. – ISBN 978-5-9519-2145-1.
2. Валерий Фальков посетил с рабочим визитом Когалым // Министерство науки и высшего образования Российской Федерации. URL: <https://www.minobrnauki.gov.ru/press-center/news/novosti-ministerstva/79757/> (дата обращения: 27.03.2024).
3. Лосева, С. А. Перевод и искусственный интеллект: проблемы и пути развития / С. А. Лосева // Межкультурная коммуникация и профессионально ориентированное обучение иностранным языкам : материалы XVI Междунар. науч. конф., посвящ. 101-й годовщине образования Белорус. гос. ун-та, Минск, 27 окт. 2022 г. / Белорус. гос. ун-т – Минск : БГУ, 2022. – С. 162–168.
4. Chad | ChatGPT на русском. URL: <https://chadgpt.ru/> (дата обращения: 27.03.2024).
5. Mohamed, Y., Kannan, A., Bashir, M., et al. (2024). The Impact of Artificial Intelligence on Language Translation: A Review. IEEE Access. 12, 25553-25579. <http://dx.doi.org/10.1109/ACCESS.2024.3366802>.

AI AS A NEW TYPE OF COMPUTER-ASSISTED TRANSLATION. A ROAD PAVED WITH GOOD INTENTIONS

***Dzhunkovskiy A. V.**, Head of the Department of Applied and Experimental Linguistics of the FAI, Senior Researcher at the Experimental Phonetic Laboratory of Criminology for Speech Production of the Moscow State Linguistic University, Candidate of Philological Sciences, Moscow, vetinari01@gmail.com*

In this paper we examine the problem of using artificial intelligence as a CAT-tool. Nowadays access to cheap and user-friendly AI interfaces has become commonplaces and this creates concerns for interpretation and translation teachers. We test the performance of ChatGPT as a translation tool and draw conclusions on whether there is an empirical basis for concern.

applied linguistics • computer-assisted translation • artificial intelligence • teaching translation • ChatGPT



Методика распознавания речевых команд в школьных информационных системах

Катаев М. Ю., доктор технических наук, профессор кафедры автоматизированных систем управления (АСУ), Томский государственный университет систем управления и радиоэлектроники, г. Томск, профессор кафедры интеллектуальных информационных систем и технологий МФТИ, г. Долгопрудный, kty@asu.tusur.ru

В статье представлены варианты использования компьютерных звуковых технологий в информационных системах, помогающих решению разного рода задач. Рассмотрены проблемы человеко-машинного взаимодействия, приводящие к необходимости дублировать клавиатурный ввод информации с помощью речевых команд. Показаны преимущества и недостатки речевого интерфейса для организации работы в информационных системах. Особую значимость приобретают системы, которые ориентированы на голоса школьников, особенно начальной школы. Приводится описание методики и программы, реализующей речевой ввод команд с особенностями детских голосов.

голос • распознавание речи • человеко-машинное взаимодействие • информационные системы • программное обеспечение

ВВЕДЕНИЕ

Речь — это естественный способ общения, выражения мыслей большей части людей, и для некоторой группы этот способ выражения мыслей не работает ввиду заболеваний или полученных травм. Развитые в последнее десятилетие технологии позволили дать человеку уверенную возможность общаться с машинами без ввода текста руками на клавиатуре [1]. Для того чтобы реализовать такие технологии, необходимо решить задачу распознавания речи на достаточно высоком уровне точности [2]. Распознавание речи — это технология, которая позволяет общаться с вычислительными машинами на основе понимания основ человеческой речи для идентификации, понимания и соответствующего реагирования. Сущность технологии основана на информации о голосе в виде некоторой цифровой последовательности данных, обработка которых позволяет автоматически понимать человеческую речь. Реализация технологии опирается на междисциплинарные исследования в области акустики, фонетики, лингвистики, теории информации, методов преобразования временных данных, распознавания образов и др. [3].

Последние два десятилетия связаны с бурным развитием компьютерного оборудования, программного обеспечения и информационных технологий, что позволяет решать разнообразные практические задачи, в том числе и распознавания речи. Речь человека представляет собой набор слов, которые связаны между собой смыслом. Распознавание речи включает в себя измерение звуковой волны, предварительную и тематическую обработку, а также анализ преобразования голоса в шаблоны известных слов. Важным качеством систем распознавания речи является не зависящий от произношения пользователя словарный запас. В данной статье приводится описание методики распознавания речи в виде команд в учебных информационных системах школы, вузов и т. д.

РЕЧЬ ЧЕЛОВЕКА

Можно указать несколько типов речи.

1. Изолированные слова (например, команды).
2. Связанные слова (важным аспектом распознавания является длительность паузы между словами, достаточной для того, чтобы слова речи разделить как отдельные).
3. Непрерывная речь (естественная непрерывная речь человека).
4. Спонтанная речь (отличается от естественной лишь только способом произношения, например ответы на вопросы) [4].

Кратко рассмотрим эти типы речи. Формально речь можно представить как некоторый прерывистый процесс, где паузы необходимы для разделения его элементов. Таким образом, изолированное слово является выделением определенными по длительности паузами произнесенного слова с обеих сторон. Распознавание изолированных слов, произнесенных различными людьми, вне зависимости от пола и возраста позволяет реализовать направление, связанное с управлением программными приложениями, приборами. Для любого программного приложения с определенным интерфейсом, наименование функций которого можно представить в виде словаря команд (число которых конечно), состоящих из одного слова. Поэтому управление этим программным приложением можно осуществить через клавиатуру или с помощью голоса, проговаривая необходимые команды. Решение этой задачи связано с одной проблемой, связанной с высокой чувствительностью методов распознавания к фоновому шуму, часто приводящему к пониманию ложно произнесенных слов, которые понимаются словом команды.

Другим элементом речи является система связанных слов, которые состоят из отдельных слов, но временные паузы между словами являются случайными и минимальными. Это приводит к тому, что технологии не всегда могут быть применены к распознаванию изолированных слов такого типа речи. В этом случае приходится разрабатывать технологии, которые основываются на моделях словосочетаний, которые успешно позволяют распознавать отдельные слова, произнесенные связано. Такой подход позволяет быть использованным не только для произнесения команд, но и для систем голосового набора текста. Более сложный вариант связан с непрерывной речью, когда человек говорит в естественной манере. Технологии распознавания естественной речи самые сложные. Существует тип речи



спонтанный, когда, с одной стороны, это естественная речь, и, с другой стороны, речь не отрепетирована и не всегда соблюдается логическое единство структуры. Одним из проявлений такой речи является применение жаргонов, сокращений, идущих слов вместе и др. Системы спонтанного разговора являются уникальными и не всегда способны понимать смысл разговорного материала.

Рассмотрим некоторые аспекты приведения человеческого голоса от звуковой волны к цифровой форме. Такие программно-технические устройства функционируют в режиме непрерывного конвейера, который преобразует звуковую волну в цифровые сигналы, поступающие от микрофона через звуковую карту на программный модуль распознавания речи. Эти сигналы проходят несколько этапов обработки, где используются различные математические и статистические методы.

На рисунке 1 показаны этапы решения задачи распознавания речи. Сначала выполняется преобразование речи (звуковой волны) в цифровой сигнал в микрофоне. Затем идет этап начальной обработки сигнала, где удаляются шумы и начинается построение звуковых фрагментов, которые идут каждые 15 миллисекунд (этот диапазон времени может меняться), для каждого из которых выполняется извлечение признаков. Далее привлекаются языковые модели, с помощью которых выполняется моделирование изолированных слов, которые сопоставляются с элементами речи во фрагменте. Последний этап связан с распознаванием речи, выделением изолированного набора слов. Отметим, что задачей моделирования является установление параметров модели голоса в соответствии с характеристиками записанного человеческого голоса, на основе которого подбирается требуемый шаблон распознавания речи. Для речи есть несколько вариантов выбора характеристик распознавания: амплитуда, скорость перехода через нуль, спектральные коэффициенты (Фурье, Вэйвлет и т.д.) и др. Шаблоны элементов слов, которые используются при распознавании, представлены функциями времени, охватывающими все слово или набором значений характеристик, которые связаны с отдельными элементами слова.

Фактически, если бы не было множества мешающих факторов, существует возможность распознавания речи непосредственно из оцифрованного сигнала. Однако даже и в этом случае возникают проблемы, которые связаны с большой изменчивостью цифровой формы речевого сигнала. Уменьшение этой изменчивости возможно с помощью функций, которые ее уменьшают. Для этого устраняется эффект периодичности при разделении на составляющие всего спектра звуковых частот (низкие, средние и высокие или другие диапазоны), для которых выбирается центр (основная частота), амплитуда и полуширина на полувысоте. Другой вариант связан с представлением трендовой части в виде линейных, параболических или логарифмических, или их комбинаций. Следующий вариант применения акустических моделей — когда происходит представление каждого слова в виде отдельных звуков. В каждом языке есть различное по количеству сочетание звуков, позволяющее распознавать речь. Для построения акустических моделей

необходима большая база данных речи. Так, компания Google уже более 10 лет собирает различные фразы, чтобы получить статистически значимые акустические модели. Распознавание речи говорящего можно разделить на два метода: текстовый и текстово-независимый. Для первого случая поиск изолированных слов речи идет на основе шаблонов слов, а в текстово-зависимом методе говорящий произносит ключевые слова для обучения.



Рис. 1. Этапы решения задачи распознавания речи

ИССЛЕДОВАНИЯ В ОБЛАСТИ РАСПОЗНАВАНИЯ РЕЧИ

Независимо от качества полученной акустической модели, на практике все они изначально имеют довольно большой коэффициент ошибок, так как сказываются следующие факторы: особенности проговаривания слов человеком и качество входного сигнала (качество записи). Чтобы уменьшить влияние фактора особенностей произношения слов говорящим, обычно проводят адаптацию акустической модели. Для этого записываются многократно отдельные одни и те же слова из созданного набора (словаря). В итоге произнесенные слова должны быть на уровне записанных сигналов быть схожими по какому-то выбранному критерию.

Самым сложным для реализации является фактор качества входного сигнала микрофона, так как использовать профессиональную аппаратуру в большинстве случаев нет возможности, а бытовая техника вносит определенные погрешности. На записи кроме шумов технических могут присутствовать посторонние звуки, которые будут мешать правильному пониманию произнесенных слов, на уровне цифрового сигнала. Тем не менее системы распознавания речи должны учитывать эти проблемы и подстраиваться для их уменьшения.

Рассмотрим общий вид процесса распознавания речи и превращения его в текст (известное направление speech to text) (рис. 2).

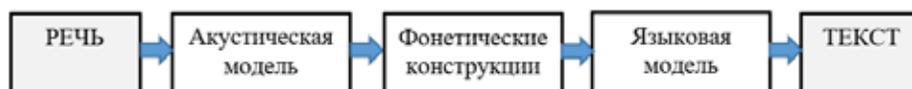


Рис. 2. Общий вид процесса распознавания речи и превращения его в текст

Алгоритм решения задачи распознавания речи и превращения его в текст следующий.

1. Обработка речи начинается с получения речевого сигнала. Сигнал преобразуется в набор акустических параметров.
2. Акустическая форма сигнала приводится в соответствие с внутренним алфавитом фонетических элементов.
3. Фонетические конструкции подаются для сопоставления с языковой моделью и её внутренними конструкциями.
4. С помощью языковой модели определяется набор слов и подбирается наиболее вероятный порядок слов, который преобразуется в текст.

ОБЗОР СИСТЕМ РАСПОЗНАВАНИЯ И БАЗЫ ДАННЫХ РЕЧИ

Для создания систем распознавания речи необходимы большие наборы записей речи различных людей и условий. В [5] проанализированы облачные системы распознавания речи с использованием таких систем, как Siri [<https://www.apple.com/siri/>], Google Speech Recognizer [<https://cloud.google.com/speech-to-text>] и Dragon [<https://voicex.com.au/dragon-speech-recognition-software.html>]. Известны системы НТК [<https://htk.eng.cam.ac.uk>] и Julius [<https://github.com/julius-speech/julius>], которые рекомендованы для использования в образовательной деятельности в области распознавания речи. Система Kaldi [<https://kaldi-asr.org>] может успешно применяться для научно-исследовательской деятельности, так как содержит необходимые для этого данные и программные инструменты. Система распознавания речи на базе CMU Sphinx [<https://cmusphinx.github.io>] также является популярной для разработки начального уровня программ и последующего их развития.

Однако рассмотренные выше системы распознавания опираются на базы данных речи на английском языке. Для создания систем распознавания речи необходимо дополнять существующие базы данных или создавать свои собственные. Русский язык принадлежит к семье языков, для которых характерна склонность к сочетанию одной или

нескольких грамматических морфем в одной словоформе [6]. Поэтому синтез слов русского языка имеет сложный механизм формирования словоформ как одного из представлений слова (например, в формировании числительных: третий, третья, третье, трети, третьей, третьем, третьих).

Известны базы данных речевых записей русской речи детей в возрасте от трех месяцев до семи лет INFANTRU и CHILDRU [7]. В базе данных INFANTRU приведена речь 187 детей в виде пятиминутных фрагментов для различных эмоциональных состояний. В базе данных CHILDRU расположены записи речи 150 детей возрастного диапазона от четырех до семи лет. Известна открытая база данных русской речи [https://github.com/snakers4/open_stt], содержащей записи чистые и с зашумлением с временем более 1000 ч. Наиболее проработанной является набор Kaldi [<http://alphacephei.com/kaldi/kaldi-ru-0.6.tar.gz>] для основных элементов русской речи. Таким образом, имеется полная возможность создать модель распознавания речи для широкого диапазона возрастов в различных условиях.

Теперь рассмотрим некоторые распространенные программные решения по распознаванию речи на сегодняшний день и их применения (таблица 1). Стоит отметить, что программные решения крупных корпораций, таких как Google, Яндекс, Apple, не будут рассмотрены ниже, так как технические решения, примененные в системах этих компаний, являются недоступными для некоммерческого использования.

Таблица 1

Общее описание и возможности программ

Название программы	ОС, язык разработки	Функции	Возможности	Источник
Voice Control System	Microsoft Windows C++	Управление воспроизведением музыки, громкостью и плейлистами	Управление музыкой	https://github.com/TeMPORaL/vcs
Microsoft Cognitive Services - Voice Assistant	Microsoft Windows, Linux C++ C#	Голосовые команды для управления, визуализация результатов	Голосовой помощник	https://github.com/Azure-Samples/Cognitive-Services-Voice-Assistant
Open-source Voice Command Interface Kit	C++		Управление роботами, управление устройствами и домашняя автоматизация	https://github.com/JunaidCS032/OpenVCIK
Speech Launcher	Microsoft Windows C++	Запуск программ с помощью голосовых команд	Голосовой запуск программ	https://github.com/JoeBiellik/speechlauncher

Название программы	ОС, язык разработки	Функции	Возможности	Источник
Effusion	Microsoft Windows C++	Управление функциями браузера с помощью	Браузер с голосовыми командами	https://github.com/tanayagar/Effusion
Victor - Voice Interactive Controller	Cross-platform C#		Голосовое управление ПК	https://github.com/allisterb/Victor
PiStudio	Microsoft Windows C#	Поворот изображения, рисование кривых, изменение яркости изображения и т.д.	Программа для редактирования фотографий с голосовым управлением	https://github.com/limo1996/PiStudio
Voice Assistant	Cross-platform Python	Управление браузером, электронными письмами, запуск системных приложений, информирует о погоде и др.	Голосовой помощник	https://github.com/nageshsinghc4/Voice-assistant
Voice_gender_detection	Cross-platform Python		Распознавание гендерной принадлежности по аудиофайлам	https://github.com/jim-schwoebel/voice_gender_detection
Leopard	Microsoft Windows, Linux, Mac OS, web browsers, Android, iOS, Raspberry Pi Python	Активация устройства с помощью голосовых команд, голосовой ввод и ведение заметок	Механизм преобразования речи в текст	https://github.com/Picovoice/leopard
Jarvis - An Intelligent AI Consciousness	Linux Python	Открытие веб-страниц, управление громкостью, создание напоминаний и т.д.	Голосовой помощник	https://github.com/ggeop/Python-ai-assistant

Окончание табл. 1

Название программы	ОС, язык разработки	Функции	Возможности	Источник
Hey Athena	Cross-platform Python	Управление умным домом, управление роботом, управление календарем и музыкой и др.	Голосовой помощник	https://github.com/rcbyron/hey-athena-client
AlexaPi (the new & awesome version)	Microsoft Windows, Linux, Raspberry Pi, Orange Pi Python		Клиент для сервиса Amazon Alexa (голосовой помощник)	https://github.com/alexa-pi/AlexaPi
Mozilla Deep Speech	Python	Автономная поддержка, низкая задержка	Голосовой помощник	https://github.com/nebulaai/voice-recognition

Можно заметить, что большинство программных решений со встроенной системой распознавания человеческой речи являются так называемыми голосовыми помощниками, действия которых направлены на облегчение взаимодействия пользователя с программой. Обычно подобные системы встраиваются в веб-браузеры, специализируемые профессиональные программы, программы для людей с ограниченными возможностями или непосредственно в операционную систему устройства, например в персональный компьютер, смартфон или автомобильную мультимедийную систему.

Кроме того, системы распознавания речи могут использоваться в службе поддержки клиентов для обработки стандартных телефонных запросов, а также в медицинских и юридических службах для документирования процессов. Программное обеспечение для распознавания голоса может помочь компаниям улучшить общение и преобразовать их в формат данных, который легко использовать и архивировать.

А теперь рассмотрим программные библиотеки, с помощью которых реализованы данные системы (табл. 2).

Таблица 2

Библиотеки, используемые программами

Название программы	Библиотеки	Источник
Voice Control System	Microsoft Speech SDK	https://github.com/TeMPORaL/vcs
Open-source Voice Command Interface Kit	Pocketsphinx	https://github.com/JunaidCS032/OpenVCIK
Microsoft Cognitive Services - Voice Assistant	Microsoft Speech SDK	https://github.com/Azure-Samples/Cognitive-Services-Voice-Assistant

Окончание табл. 2

Название программы	Библиотеки	Источник
Speech Launcher	Windows SpeechRecognitionEngine	https://github.com/JoeBiellik/speechlauncher
Effusion	Windows SpeechRecognitionEngine	https://github.com/tanayagar/Effusion
Victor — Voice Interactive Controller	Mozilla Voice Challenge	https://github.com/allisterb/Victor
PiStudio	SpeechIntegrator	https://github.com/limo1996/PiStudio
Voice Assistant	Speech Recognition*	https://github.com/nageshsinghc4/Voice-assistant
Voice_gender_detection	Speech Recognition*	https://github.com/jim-schwoebel/voice_gender_detection
Leopard	Picovoice	https://github.com/Picovoice/leopard
Jarvis — An Intelligent AI Consciousness	Google API Speech recognition	https://github.com/ggeop/Python-ai-assistant
Hey Athena	Pocketsphinx, SpeechRecognition*	https://github.com/rcbyron/hey-athena-client
AlexaPi (the new & awesome version)	Pocketsphinx	https://github.com/alexa-pi/AlexaPi
Mozilla Deep Speech	DeepSpeech	https://github.com/nebulaai/voice-recognition

ОПИСАНИЕ СИСТЕМЫ SPHINX

Анализ технологий распознавания речи, представленный в таблицах 1 и 2, приводит к выводу, что наиболее популярной системой является Sphinx. CMU Sphinx toolkit — пакет технологий распознавания речи [<https://github.com/cmusphinx>], разрабатываемый в Университете Карнеги — Меллон (CMU). Данный пакет включает набор инструментов, которые решают различные задачи и относятся к различным сферам применения, в том числе:

- 1) Pocketsphinx — «легкий» распознаватель речи на основе языка C, предназначенный для мобильных платформ;
- 2) Sphinx 4 — гибкий модифицируемый распознаватель, написанный на языке программирования Java и ориентированный на стационарные устройства;

- 3) Sphinxtrain — пакет инструментов для адаптации и обучения акустических моделей на основе скрытых моделей Маркова (СММ);
- 4) Sphinxbase — библиотека, необходимая для работы Sphinxtrain.

Важной чертой библиотеки Sphinx 4 является ее реализация исключительно на языке программирования Java, что обеспечивает мультиплатформенность написанных с ее использованием приложений. Она предоставляет программисту внешний доступ к функциям через API, предназначенный для перевода речи из аудиофайлов в текстовый формат при помощи акустических моделей CMU Sphinx. На данный момент доступны готовые модели для английского, немецкого, русского и некоторых других языков (всего 12). Библиотека поддерживает кодирование символов с помощью UTF-8, что обеспечивает поддержку множества языков, в том числе и русского. Также она предоставляет такие функции, как фильтрация фраз с неправильным грамматическим строением, фильтрация шума и распознавание речи на больших словарях. CMU Sphinx Toolkit распространяется со свободной BSDstyle лицензией, что определяет значительные экономические преимущества его использования. Sphinx 4 использует модульную архитектуру. Код внутри каждого модуля, показанного на рисунке 3, также является модульным и содержит легко заменяемые функции.

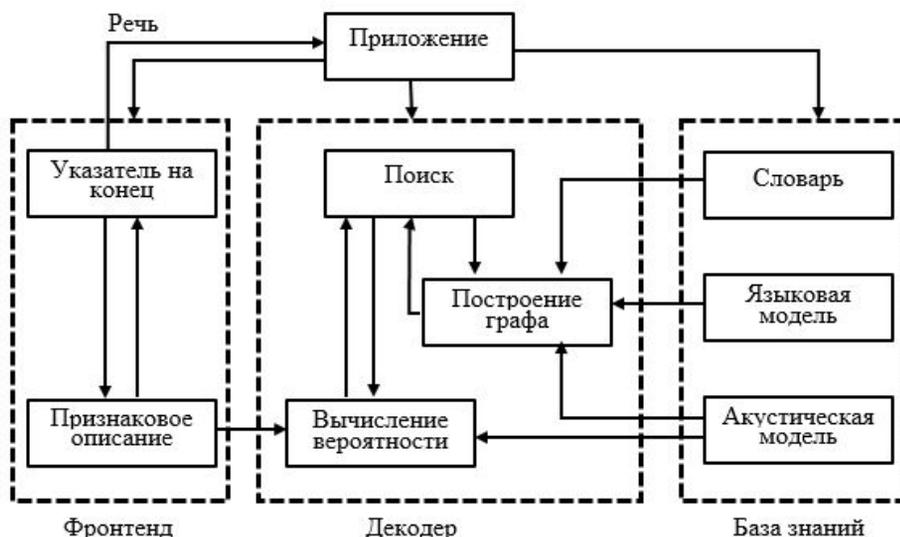


Рис. 3. Архитектура Sphinx 4

Система состоит из трех основных блоков: фронтенд (внешнее представление), декодер и база знаний. Эти части управляются внешним приложением. Блок фронтенда принимает запись речи и параметризует ее. Внутри фронтенда модуль указания на конец речи выделяет начало и конец речи, разделяет звуковой поток на речь и не-речь и удаляет области с отсутствием речи. Он может выполнять эти операции или на речевом потоке, непосредственно, или на последовательности признаковых описаний (featured vectors), вычисленных на его основе. Блок декодера непосредственно производит распознавание. Он содержит модуль конструирования графа (лингвист), который преобразует любой тип стандартной языковой модели,



предоставленной базе знаний приложением, во внутренний формат, и вместе с информацией из словаря и одной или нескольких акустических моделей строит языковую скрытую Марковскую модель. Последняя в дальнейшем используется модулем поиска для определения структуры словарной решетки, которую следует отыскать.

Модулю поиска необходимы значения вероятностей результата состояний для каждого признакового описания, чтобы определить значения последовательностей состояний. Вероятности результата состояний вычисляются модулем вычисления вероятностей состояний (акустический счетчик) — единственным модулем, имеющим непосредственный доступ к признаковым описаниям. Вычисление значений, таким образом, является задачей «по запросу», выполняющейся тогда, когда поисковый модуль сообщает значения состояний вычисляющему модулю, которому требуются значения некоторого признакового описания [8].

Акустическая модель Sphinx 4 представляет собой статистическую репрезентацию фонем, составляющих слова. Она моделирует взаимодействие пользователя (то, как он произносит слова) с системой распознавания речи, вычисляя вероятности появления различных цепочек слов среди сказанного и выбирая наиболее вероятную (то есть выдвигает гипотезу о том, что было сказано) [9]. Данная модель основана на скрытых моделях Маркова. Акустическая модель русского языка zero_ru_cont_8k_v3 построена на основе 4000 сенонов [10, 11], составляющих части трифонов, по которым можно опознать, например, их начало. Использование сенонов позволяет сократить число распознаваемых единиц в речи.

Фонетический словарь содержит транскрипции слов в фонемы. Такой список транскрипций не слишком эффективен, так как обычно содержит не больше двух-трех вариантов произношения слова, но в большинстве случаев этого достаточно. Словарь является не единственным способом представления слов в фонемы (например, может быть заменен функцией преобразования слов, основанной на машинном обучении), но на настоящий момент используется наиболее часто [12].

Лингвистическая (языковая) модель управляет поиском слова, то есть определяет, какое слово может следовать за распознанным и ограничивает процесс сопоставления фонем со словарем, отбрасывая слова, которые в данном контексте «не подходят». Часто используемыми являются два вида языковых моделей: n-граммные (содержащие статистические данные о последовательностях слов) и модели, основанные на конечных автоматах, иногда использующие веса (также называются грамматиками). Для достижения высокой точности распознавания речи лингвистическая модель должна сильно ограничивать область поиска слова, то есть хорошо предсказывать следующее слово в цепочке. Обычно языковая модель может распознать только слова из словаря. Это вызывает проблемы с распознаванием имен собственных, которые иногда решают введением слогов или даже отдельных фонем в лингвистическую модель, что, разумеется, снижает общую точность предсказания слов и распознавания [12].

Грамматики обычно служат для описания простых языков, предназначенных для отдачи каких-либо команд или управления чем-либо. Обычно грамматики пишутся вручную или генерируются алгоритмом. Чаще всего грамматики не содержат статистических вероятностей, но некоторые элементы могут иметь веса [13].

Грамматики позволяют очень точно определять структуру распознаваемой речи. Например, определенное слово может быть повторено только два или три раза. Однако такая строгость правил может и навредить, если пользователь, к примеру, пропустит слово, требуемое грамматикой в распознаваемой конструкции. В таком случае вся фраза может не распознаться. По этой причине грамматики обычно делают более свободными, чем цепочки слов в строго определенном порядке. В грамматике предпочтительнее использовать простые правила, без множества вариантов выбора, так как это замедляет распознаватель [14, 15].

АРХИТЕКТУРА ПРОТОТИПА ПРОГРАММНОЙ РЕАЛИЗАЦИИ

Для решения задач программирования с учетом вышеописанных технологий была разработана следующая архитектура системы распознавания речи, которая позволяет решать задачи школьных информационных систем. Схема архитектуры прототипа программной реализации изображена на рисунке 4. На схеме показано разбиение программной системы на компоненты (слои и модули) и их взаимосвязь.

Описание схемы, представленной на рисунке 4:

1. Пользователи — люди (учителя, школьники), управляющие программой с помощью заранее известных команд.
2. Слой представления — слой, отвечающий за взаимодействие программы и пользователя. Интерфейс приложения — интерфейс, который обеспечивает передачу информации между пользователем и программой. Пользователь дает команды программе с помощью него, а программа в свою очередь с помощью него показывает результат выполнения команды.
3. Слой бизнес-логики — слой, описывающий основные функции программы, которые направлены на достижение поставленных для нее целей. Модуль обработки данных — модуль, включающий в себя все подмодули, отвечающие за выполнение основных функций программы. Модуль бизнес-логики — модуль, который служит основой программы. На него ложатся задачи, связанные с логистикой входных и выходных данных, навигацией интерфейса программы и т.п. Sphinx 4 — модуль, отвечающий за обработку входного сигнала и за работу самого распознавания.
4. Слой доступа к данным — это слой, который предоставляет упрощенный доступ к данным, хранимым в постоянном хранилище какого-либо типа. Модуль доступа к данным — модуль, отвечающий за доступ к данным. В его функции входит чтение и запись запрошенных данных из хранилища.
5. Данные — представляют собой хранилище данных какого-либо типа. В написанной программе представлены в виде оперативной памяти персонального компьютера.



Рис. 4. Схема архитектуры программной реализации

ПРЕДЛАГАЕМОЕ РЕШЕНИЕ

Опираясь на архитектуру рисунка 4, был разработан алгоритм, блок-схема которого изображена на рисунке 5, на основе которого функционирует прототип программной реализации системы распознавания речи в разрабатываемой школьной информационной системе. По алгоритму голосовая команда пользователя поступает на вход системы через устройство входа (микрофон), после чего преобразуется в звуковой сигнал и попадает на вход нейронной сети, где происходит обработка сигнала и само распознавание речи. После чего результат распознавания (команда) с выхода нейронной сети попадает в основной блок программы, где и происходит выполнение этой команды.

Например, пользователь хочет выйти из системы и произносит команду «Выйти». Его слова через микрофон персонального компьютера (устройство входа) преобразуются в звуковой сигнал и попадают в нейронную сеть системы. В нейронной сети происходит выделение речи из всего звукового сигнала, отсекание шумов, нормализация сигнала и т.п., и в конце на выход сети попадает распознанная команда «Выйти», после чего программа выполняет функции, которые приписаны для данной команды.



Рис. 5. Блок-схема алгоритма

С использованием прототипа программной реализации был разработан функционал, который является ключевым для облегчения задачи обучения учеников и студентов, особенно с ограниченными физическими возможностями, средствами голосового интерфейса. С помощью программы ученик может с помощью голосовой команды выбрать определенный предмет из списка (рис. 5), перейти с помощью голосовых команд «дальше» и «назад» к определенному заданию, дать ответ на задание и, конечно же, выйти из программы.

Данное программное решение позволяет учащимся проходить тестирования и выполнять прочие задания, ответы на которые даются односложными шаблонными словами, так как именно на словаре с небольшим количеством слов распознавание команд диктора происходит с высокой точностью. Благодаря тому, что данная система является модулем, а консольная программа — не что иное как демонстрация работы данного модуля, в дальнейшем она будет внедрена в крупную обучающую платформу для людей с ограниченными возможностями.

Распознавание речи в реальном времени происходит с помощью библиотеки, написанной на языке программирования Python PocketSphinx [<https://pypi.org/project/pocketsphinx>], для работы которой составлен словарь из необходимых для программы слов и файл JSGF с грамматикой, а также используется всем доступная русская языковая модель zero_ru.cd_cont_4000. JSpeech Grammar Format (JSGF [<https://pypi.org/project/pyjsgf>]) — это не зависящее от платформы и поставщика текстовое представление грамматик для использования при распознавании речи. На вход библиотеки подается звуковой поток с микрофона, а на выходе получают готовые фразы из файла грамматики, по которым и происходит навигация в разрабатываемой программе.

Сама программа представляет собой небольшое консольное приложение для обучения учеников и студентов, в котором предлагается назвать одну из доступных на данном шаге команд для перехода к следующему шагу, например: «Выберите предмет из списка: физика, математика, биология». А при переходе к конкретному предмету предлагается выбрать задание или выйти и т. п.

Для проверки качества распознавания речи было составлено три словаря с количеством 15, 30 и 60 слов. Чтобы объединить слова из словаря в команды для управления персональным компьютером, был создан набор JSGF грамматик. Оценка точности работы проводилась в реальном времени с помощью подачи системе голосом команд из списка. Подача команд осуществлялась одним и тем же диктором в одинаковой последовательности.

Для численной оценки точности распознавания речи был задействован стандартный критерий WER (Word Error Rate).

Для адаптации акустической модели использовался пакет Sphinxtrain — набор утилит, входящий в CMU Sphinx Toolkit и предназначенный для тренировки и адаптации некоторых типов скрытых моделей Маркова, используемых как акустические модели в Sphinx 4. В качестве материала для адаптации акустической модели выступает список wav файлов с фразами пользователя, под которые системе следует адаптироваться, и сопоставленный ему список транскрипций того, что было произнесено в этих файлах. Файлы записаны в строго определенном формате (квантование, частота дискретизации и т.п.), который определяется форматом



акустической модели. Апробация адаптации русскоязычной акустической модели zero_ru_cd_cont_4000 была выполнена путем записи набора из одиннадцати аудиофайлов в формате 8 KHz, 16 bit, 128 Kbps, mono, WAV PCM little-endian signed, содержащих следующие команды: - назад, - выход, - химия, - показать, - математика, - дальше, - физика, - задание, - ответ, - да, - нет и других из файла грамматики. Каждому файлу была сопоставлена транскрипция произнесенного в них текста.

Запись файлов для адаптации и оценка качества распознавания проводилась в жилой комнате многоквартирного дома на встроенный микрофон персонального компьютера — для адаптации модели в нормальных условиях и в подготовленной для звукозаписи комнате на студийный микрофон — для адаптации модели в идеальных условиях.

РЕЗУЛЬТАТЫ

По результатам исследования зависимости качества системы распознавания была составлена таблица (таблица 3), по которой видно, что из-за низкого качества распознавания систем на словарях с большим количеством слов построенная на данных акустических моделях система не подходит для создания полноценного голосового интерфейса. Ведь известно, что речевой интерфейс является удобным пользователю тогда, когда ошибка распознавания (WER) голосовых команд не превышает 5%. Но исходя из того, что ошибка распознавания на словаре с количеством слов 15 у адаптированной акустической модели в нормальных условиях равна 2,78%, а у адаптированной акустической модели в идеальных условиях равна 0, можно сделать вывод, что адаптированную акустическую модель, в отличие от неадаптированной, можно использовать в создании системы распознавания с небольшим ограниченным числом команд.

Таблица 3

Результаты распознавания, %

	Количество слов в словаре		
	15	30	60
WER неадаптированной модели	31,6	40,44	73,77
WER адаптированной модели (нормальные условия)	2,78	18,37	21,28
WER адаптированной модели (идеальные условия)	0	6,12	12,76

ЗАКЛЮЧЕНИЕ

Применение разработанной системы в качестве полноценного голосового управления персональным компьютером не является рациональным, так как высокая точность распознавания возможна лишь при малом

объеме словаря, размер которого ограничивает речевой диалог, а при большом объеме словаря, набор слов которого позволяет сделать управление возможным, сильно снижается точность распознавания; создание подобной системы для дикторнезависимого голосового управления нецелесообразно, а для создания системы голосового управления заточенной под конкретного пользователя затруднено из-за громоздкой процедуры обучения, которая нуждается в создании обширной базы обучающего материала.

Но при этом ошибка распознавания на словаре с малым количеством слов у адаптированной акустической модели является приемлемой для удобного использования такой системы, из чего можно сделать вывод, что адаптированную акустическую модель, в отличие от неадаптированной, можно использовать в создании системы распознавания с небольшим ограниченным числом команд.

Список использованных источников

1. Ронжин, А. Л. Речевой и многомодальный интерфейсы / А. Л. Ронжин, А. А. Карпов, И. В. Ли. – М.: Наука, 2007. – 173 с.
2. Сорокин, В. Н. Распознавание личности по голосу: аналитический обзор / В. Н. Сорокин, В. В. Вьюгин, А. А. Тананыкин // Информационные процессы. – 2012. – Т. 12. – № 1. – С. 1–30.
3. Газин, А. И. Методики формирования баз голосовых образов для оценки стойкости систем голосовой аутентификации личности: дисс. ... канд. техн. наук. – Пенза: ПГУ, 2011. – 184 с.
4. Тампель, И. Б. Автоматическое распознавание речи — основные этапы за 50 лет // Научно-технический вестник информационных технологий, механики и оптики. – 2015. – Т. 15. – № 6. – С. 957–968.
5. Assefi, M., Liu, G., Wittit, M. P., & Izurieta, C. (2016). Measuring the Impact of Network Performance on Cloud-Based Speech Recognition Applications. *International Journal of Computer Applications-IJCA*, 23, 19–28.
6. Тампель, И. Б. Автоматическое распознавание речи / И. Б. Тампель, А. А. Карпов. – СПб: Университет ИТМО, 2016. – 138 с.
7. Ляксо, Е. Е. Базы данных речи русских детей INFANTRU и CHILDRU / Е. Е. Ляксо, О. В. Фролова, А. Д. Громова, Ю. С. Гайкова // Речевые технологии. – 2009. – № 2. – С. 14–30.
8. Lamere P., Kwok P., Walker W., Gouva E., Singh R., Raj B., Wolf P. Design of the CMU Sphinx-4 Decoder // Mitsubishi electric research laboratories, 2003, Report TR2003-110. – 5 p.
9. Jelinek, F. *Statistical Methods for Speech Recognition*. – The MIT Press, 2022. – 306 p.
10. Overview of the CMUSphinx toolkit // [cmusphinx.github.io](https://cmusphinx.github.io/wiki/tutorialoverview/): Open source speech recognition toolkit resource. <https://cmusphinx.github.io/wiki/tutorialoverview/> (дата обращения: 29.03.2024).
11. Остроухов, З. Creating Russian voice model for cmu-sphinx. <https://github.com/zamiron/gu4sphinx> (дата обращения 30.03.2024).
12. Юлдашев, Р. Ю. Разработка системы голосового управления роботами на базе CMU Sphinx Toolkit / Р. Ю. Юлдашев, С. Д. Шелудько, В. Ю. Радыгин, А. М. Науменко // Научно-практический электронный журнал Аллея Науки. – 2017. – № 10. – С. 1–13.
13. Knight S., Gorrell G., Rayner M., Milward D., Koeling R., Lewin I. Comparing grammar-based and robust approaches to speech understanding: a case study // *Proceedings of Eurospeech*, 2001. – P. 1779–1782
14. Кипяткова, И. С. Автоматическая обработка и статистический анализ новостного текстового корпуса для модели языка системы распознавания русской речи / И. С. Кипяткова, А. А. Карпов // Информационно-управляющие системы. – 2010. – № 4(47). – С. 2–8.
15. Захаров, В. П. Корпусная лингвистика / В. П. Захаров, С. Ю. Богданова. – СПб.: Изд-во С.-Петербург. ун-та, 2020. – 234 с.



METHODS OF SPEECH COMMAND RECOGNITION IN SCHOOL INFORMATION SYSTEMS

Kataev M. Yu., Doctor of Technical Sciences, Professor of the Department of Automated Control Systems (ACS), Tomsk State University of Control Systems and Radioelectronics, Tomsk, Professor of the Department of Intelligent Information Systems and Technologies at MIPT, Dolgoprudny, e-mail: kmy@asu.tusur.ru

The article presents options for using computer sound technologies in information systems that help solve various kinds of problems. The problems of human-machine interaction leading to the need to duplicate the keyboard input of information using speech commands are considered. The advantages and disadvantages of the speech interface for organizing work in information systems are shown. Of particular importance are systems that focus on the voices of schoolchildren, especially primary schools. The description of the methodology and the program implementing the speech input of commands with the characteristics of children's voices is given.

voice • speech recognition • human-machine interaction • information systems • software

За пределами семантики, или Семантика Бытия и Небытия

Кулай А. А., член совета Костромского ВОИР (Всероссийского общества изобретателей и рационализаторов), г. Кострома

В статье представлен необычный и во многом философский доклад автора с элементами мистики и метафизической направленности. Вопросы по теме можно отправить автору по адресу: 31pk09ksq67t@mail.ru

философия • квантовая семантика • Бытие • Небытие • интерференция

*Космическая философия начинается с больших чисел...
Посвящается участникам клуба «Сразу ВСЁ»*

В статье мы рассмотрим *μαντική*¹ в продолжение сложившихся взглядов автора и направлений на семантику с двух точек зрения: **философии, квантовой семантики**².

В области философского размышления дихотомии Бытия и Небытия³ давно служит фундаментальной осью, вокруг которой вращаются запросы о существовании. Тем не менее за этой бинарной рамкой скрывается богатая ткань онтологического исследования, которая превосходит простую семантику и в глубине открывает сущность **реальности**⁴, сознания и загадочной природы самого человеческого и машинного существования. Однако то, что существует для человека, пока не всегда существует для мира машин...

В сердце этого философского лабиринта стоит неуловимый вопрос о том, что по-настоящему означает существовать во Вселенной. За пределами концептуальных противопоставлений, таких как Бытие и Небытие, лежит обширное поле

¹ Мантика (др.-греч. *μαντική, μαντευτική* — искусство прорицания, пророческий дар) — гадания в Древней Греции и Риме, которые проводились с целью установления воли богов на основании различных ниспосылаемых ими предзнаменований.

² Направления: религия, космология и физика звука в данный доклад не вошли по причине значительной доработки текста и его осмысления в новых условиях и, возможно, будут представлены позднее (прим. авт.)

³ Небытие (абсолютное) — это все абсолютно несуществующее либо всё то, что абсолютно отрицает Бытие в физике и метафизике.

Например, с позиций квантовой физики Небытие (часть абсолютного Небытия) можно рассматривать как состояние, которое не имеет определенного физического существования или не может быть точно измерено. В квантовой механике существует понятие вероятностного существования частиц и состояний, которые могут быть как в суперпозиции состояний, так и в квантовом вакууме.

В квантовой теории поля пустота не является полностью пустой, а наполнена квантовыми флуктуациями и виртуальными частицами, которые могут возникать и исчезать на краткое время. Эти флуктуации представляют собой некоторое «относительное Небытие», поскольку они не являются постоянными или стабильными состояниями.

Таким образом, с точки зрения автора доклада и квантовой физики, Небытие может быть истолковано как состояние неопределенности, вероятностного существования или временного проявления частиц или полей, которые не имеют постоянного или определенного физического и метафизического существования.

⁴ Reality Principle — The real universe contains all and only that which is real. Teleologic Evolution. Chris Langan, Dr. Gina Langan.



онтологического изучения, вызывающее нас усомниться в традиционных представлениях о реальности. Мы приглашаем вас исследовать области мышления, где обычный язык и категоризация не способны удовлетворить и мир человека, и мир машин, и вычислений и самого существования человека.

Существование в своем самом банальном смысле не поддается упрощенной категоризации. Оно охватывает не только конкретный мир объектов и сущностей, но также область идей, чувств и сознания. Оно приглашает нас задуматься о природе реальности за пределами ограничений языка и восприятия, настоятельно призывая нас исследовать взаимосвязь всех вещей и неизведанные тайны, формирующие наше собственное существование.

В поисках понимания существования за пределами языка звёздного неба, семантики Бытия и Небытия мы вынуждены столкнуться с глубокими вопросами о природе реальности пределах человеческого понимания и симбиотических отношений между восприятием и истиной.

Это путешествие в глубины метафизического исследования открывает двери к новым перспективам, стимулируя нас преодолевать дихотомическое мышление и принимать сложность во взаимосвязи со вселенной в области видимого и невидимого спектра.

Путешествуя по сложному миру философских дискуссий, касающихся существования, мы сталкиваемся с мозаикой точек зрения, каждая из которых предлагает уникальный взгляд на тайну Бытия и Небытия. От древних мудростей до современных философских исследований исследование существования призывает нас преодолеть языковые границы и погрузиться в области, где слова блекнут, а интуиция царствует.

Само исследование существования за пределами семантики Бытия и Небытия — это путешествие в сердце глубоких тайн, лежащих в основе нашей реальности и за её пределами. Это приглашение переосмыслить наше понимание вселенной, принять взаимосвязь всех вещей и погрузиться в глубины сознания, где скрыта суть существования, ожидающая своего раскрытия.

Какими нам видятся пределы этого путешествия? В первом приближении наше семантическое путешествие выглядит на числовой оси⁵ по минимуму от $-13,799 \pm 0,021^6$ миллиарда лет до 7000×10^6 седечиллионов⁷ лет по максимуму.

⁵ По сути, шкале времени жизни Вселенной (прим. авт.).

⁶ Возраст Вселенной — время, прошедшее с начала расширения Вселенной. Не настаиваю, есть и другие представления и о Вселенной, и о возрасте (прим. авт.).

⁷ Название периода времени дано по длинной шкале именованных единиц (прим. авт.).

Для нашего небольшого мысленного путешествия в мире чисел, где представления о семантике⁸ изменятся не раз, нам придётся посмотреть на семантику с разных точек, а именно: а) классической, б) неклассической, с) постнеклассической.

Кратко дадим их основные определения и толкования:

- а) семантика⁹ — раздел науки о выявлении системы знаков, естественно или искусственно созданных и соотносящихся с их понятийным содержанием и отображением (звучание, написание и т.д.), где «смысл» есть частный случай»;
- б) семантика (от др.-греч. *σημαντικός* «обозначающий») — это область лингвистики, которая изучает смысловое значение единиц языка, слов, фраз и предложений, а также способы, которыми люди интерпретируют их. Она занимается тем, как слова и языковые единицы связаны с концепциями и реальными объектами в мире, а также как эти связи влияют на коммуникацию и понимание. Семантика — это своего рода «танец» слов, где каждое слово выступает как танцор, передавая свою уникальную роль и значение, чтобы создать великолепное хореографическое представление, которое мы называем языком;
- с) семантика в постнеклассической терминологии может быть определена как изучение значения, интерпретации и воздействия символов, знаков и языковых конструкций в контексте их использования в различных культурных и социальных практиках. Это также включает анализ того, как значения формируются, передаются и воспринимаются в различных контекстах и ситуациях общения.

Семантика — это область лингвистики, изучающая значения слов, фраз и предложений, а также способы, которыми люди, машины понимают и интерпретируют языковые выражения. Она занимается тем, как слова сочетаются друг с другом для передачи конкретных значений и как контекст влияет на их интерпретацию.

ОСНОВНЫЕ РАЗЛИЧИЯ В ПРИВЕДЕННЫХ ОПРЕДЕЛЕНИЯХ И ТОЛКОВАНИЯХ

Классическое определение семантики относится к изучению значения слов и их комбинаций в языке. Оно фокусируется на том, как слова передают значения и как эти значения связаны с реальным миром.

Неклассическое определение семантики шире и включает в себя изучение значения не только слов, но и более сложных языковых единиц, таких как фразы, предложения и тексты в реальном и нереальном мире. Определение также может включать аспекты когнитивной лингвистики, психолингвистики и других областей, которые изучают, как люди и машины воспринимают и интерпретируют язык.

⁸ Деление принято условно, для облегчения понимания эволюции «семантики» на временной шкале (прим. авт.).

⁹ Семантика — это раздел лингвистики, изучающий значение слов, фраз, предложений и их взаимосвязь с реальным миром. Семантика занимается тем, как люди понимают значения слов и как эти значения сочетаются для образования более сложных конструкций смысла (прим. авт.).



Постнеклассическое определение семантики еще более широко и может включать в себя анализ влияния контекста, культурных аспектов, социальных факторов и других аспектов на понимание значения языковых единиц. Определение также может затрагивать проблемы многозначности, многомирности и многомерности в семантической динамике¹⁰, неоднозначности и различий в толковании значений слов и выражений разными людьми и машинами в разных мультикультурных контекстах.

Заканчивая лишь малую часть философского взгляда на семантику и её формы на шкале времени, а по сути, жизни Вселенной, в которой придётся жить и человеку, и машинам, семантика Вселенной заставляет нас выделять представление о врожденном смысле, организации и коммуникации в космическом обществе. Она будет обязательно исследовать фундаментальные принципы, лежащие в основе структуры и взаимосвязей всего живого и неживого во Вселенной. Представление о семантике Вселенной будет включать в себя рассмотрение сложной системы физических и космических законов, космических отношений в физическом и метафизическом измерениях и человека, и машин, формировать наше понимание к Бытию и Небытию, формам реальности и существования на шкале времени.

Вероятно, из-за частого упоминания слова «машины» у вас возник закономерный вопрос: что общего может быть между семантикой и миром интеллектуальных машин¹¹ Generative Pre-trained Transformer (GPT)?

По мнению автора, общим пространством для семантики и GPT является **понимание** языковой информации и её обработка и интерпретация.

Модели GPT (Generative Pre-trained Transformer), такие как GPT-3, являются передовыми моделями языка, использующими глубокое обучение для генерации текстовых ответов, похожих на человеческие. Они обучены на больших объемах данных для понимания семантики языка и генерации связных и контекстуально значимых текстов. И здесь мы снова столкнёмся с расширенными толкованиями семантики, где семантика может быть рассмотрена как не только изучение языкового значения, но также как анализ подлежащего значения и отношений в системах коммуникации, интерпретации символов и обработке информации.

Она охватывает целостное понимание того, как значение конструируется, передается и интерпретируется в различных контекстах помимо языка, затрагивая области символики, когнитивистики и представления.

¹⁰ Семантическая динамика — это область лингвистики в недалёком будущем, изучающая значения слов и выражений в контексте изменяющихся условий, содержания семантических ядер и полей, ситуаций во времени и пространстве. Эта область исследует, как значения слов и фраз могут меняться в зависимости от контекста или времени, а также как эти изменения влияют на понимание сообщений, коммуникаций, сигналов (прим. авт.).

¹¹ Такие машины называют в области компьютерных наук как искусственный интеллект (ИИ) (прим. авт.).

Суть семантики превосходит простой лингвистический анализ, превращаясь в многомерное исследование процессов создания значения в различных областях человеческого взаимодействия и концептуализации. Вероятно, на этом этапе в лингвистике будут скорее всего пересмотрены теоретические взгляды на *direct reference theory*¹² и воззрения философов, таких как Джон Стюарт Милль, Бертран Рассел, Сол Крипке, придерживающихся направления логического позитивизма, а также теории верификационизма¹³, теории использования¹⁴ и семантику инференциализма¹⁵.

Отдельно заслуживает внимания вопрос о производстве семантики с помощью семантики, но эта тема выходит за рамки философии и близка к теме информационной экономики, где семантика выступает в качестве товара...

Продолжая тему философии, я хотел бы вас познакомить с таким ракурсом, как философия в квантовой семантике. Что это такое?

Квантовая семантика — это концепция, которая объединяет идеи квантовой физики с теорией семантики. Эта концепция предполагает, что смысл слов и языка может быть описан и объяснен с использованием квантовых понятий, таких как суперпозиция, интерференция и измерение.

Одним из ключевых аспектов квантовой семантики является понятие суперпозиции, которое говорит о том, что слова и понятия могут существовать одновременно в нескольких состояниях или значениях. Когда мы измеряем или задаем вопрос о значении, происходит «коллапс» суперпозиции, и мы получаем конкретное значение или интерпретацию.

Квантовая семантика также привносит идею интерференции, где различные значения или интерпретации могут взаимодействовать и влиять друг на друга, создавая сложные комбинации смыслов.

В целом квантово-семантический подход предлагает новый взгляд на проблемы, связанные с пониманием смысла и значения слов и языка. Он вызывает интересные вопросы о природе смысла, взаимодействии значений и о том, как мы можем точно и однозначно общаться друг с другом. Однако следует отметить, что квантовая семантика является относительно новым направлением в лингвистике и активно исследуется и обсуждается в научно-исследовательском и философском сообществе, где традиционные представления о значении и существовании претерпевают парадигмальное изменение, поднимая глубокие вопросы о природе реальности, причинности и наблюдении.

¹² Direct reference theory — теория прямой ссылки Людвига Витгенштейна.

¹³ Верификационизм, также известный как принцип проверки, или критерий проверяемости значения, — это философская доктрина, которая утверждает, что утверждение имеет смысл только в том случае, если оно либо эмпирически проверяемо, то есть подтверждено через чувства (прим. авт.).

¹⁴ См. *Philosophische Untersuchungen* («Философские исследования») — известный труд философа Людвига Витгенштейна, 1953.

¹⁵ Известен как *Inferential role semantics* — подход к теории значения, который идентифицирует значение выражения с его отношением к другим выражениям (прим. авт.).



В квантовой философии семантика превосходит традиционные границы, погружаясь в таинственный мир суперпозиции, запутанности и корпускулярно-волнового дуализма. Здесь простое наблюдение влияет на сами наблюдения, ставя под сомнение наши представления об объективности, языке, истине.

Такой подход, по мнению философов, предлагает посмотреть и, возможно, переосмыслить саму суть Бытия и Небытия, реальности и нереальности, приглашая нас принять неопределенность, нелокальность и взаимосвязь всех вещей и побуждает задать вопросы о фундаментальных принципах нашего понимания, размывая границы между наблюдателем и наблюдаемым, между реальностью и восприятием. Задать свои вопросы о природе существования сознания во Вселенной.

Он ставит перед исследователями вызов преодолеть ограничения классической мысли и отправиться в путешествие по реальностям вероятности, потенциала и сущности Бытия и попытаться ответить, каковы наши мысли в квантовой философии о семантике. Как эти идеи резонируют с нашим пониманием реальности и существования человека и мира машин во Вселенной?

Для простоты понимания давайте представим себе, что в основаниях философии квантовой семантики как области, находятся объединяющие принципы квантовой физики и лингвистики. В её основе лежит простая идея-гипотеза о том, что язык и мышление могут быть рассмотрены как квантовые явления.

Представьте себе, что слова и смыслы, которыми мы оперируем, могут существовать во всех состояниях одновременно, пока не произойдет акт наблюдения или измерения.

В контексте квантовой семантики каждое слово может быть каким-то образом связано с другими словами в суперпозиции значений. Это открывает удивительные перспективы для понимания того, как мы и машины строим свою реальность через язык (символы, знаки, протознаки, признаки и числа) и как эти концепции могут быть интерпретированы в рамках квантовой теории.

Другим из удивительных аспектов квантовой семантики является идея-гипотеза о том, что наше и машинное понимание мира формируется не только через язык, но и через сам процесс взаимодействия с информацией. Это открывает новые горизонты для исследования наших когнитивных способностей и способов восприятия окружающей Вселенной сквозь мир слов и значений.

Погружение в мир квантовой семантики может изменить ваш взгляд на язык, мышление и саму природу реальности и виртуальности.

Готовы ли мы и к такому путешествию и нужно ли оно?

Предвижу вопросы из области: Приведите примеры комбинации смыслов и объяснений с точки зрения квантовых понятий как **суперпозиция, интерференция, измерение?**

Хорошо, давайте рассмотрим пример комбинации смыслов и объяснений, используя квантовую семантику с **понятиями суперпозиции, интерференции и измерения.**

Предположим, у нас есть понятие «любовь». В классической семантике мы можем дать определение этого понятия, например, как глубокое чувство привязанности и притяжения между людьми.

Однако в квантовой семантике мы можем рассматривать понятие «любовь» как суперпозицию различных значений и интерпретаций. Например, мы можем представить, что в суперпозиции есть аспекты страсти, доверия, взаимного понимания и преданности.

Когда мы начинаем обсуждать или изучать понятие «любовь», происходит интерференция различных значений и интерпретаций. Это означает, что эти аспекты взаимодействуют друг с другом и могут создавать новые комбинации смыслов. Например, страсть и преданность могут взаимодействовать таким образом, что создается сильная взаимная привязанность.

В конечном счете, чтобы получить определенное значение или интерпретацию понятия «любовь», мы должны произвести измерение. Измерение может быть представлено в виде конкретных действий или выражений, которые позволяют нам определить, какую комбинацию смыслов и интерпретаций мы выбираем.

В данном примере квантовая семантика позволяет нам увидеть, как различные аспекты и значения понятия «любовь» могут взаимодействовать и формировать более сложные комбинации смыслов. Она также подчеркивает важность измерения, чтобы определить конкретное значение или интерпретацию в контексте конкретной ситуации или отношений.

Предвижу вопрос из области: Приведите пример интерференции в квантовой семантике?

В квантовой семантике интерференция проявляется взаимодействием различных значений или интерпретаций слов или понятий.

Давайте рассмотрим пример суперпозиции и интерференции в контексте квантовой семантики.

Представим, что у нас есть понятие «истина» (И) и понятие «ложь» (Л). В классической логике мы можем сказать, что каждое утверждение может быть либо истинным, либо ложным. Однако в квантовой семантике мы можем считать, что эти значения существуют одновременно в суперпозиции.

Если мы рассмотрим утверждение: «Это утверждение истинно», то в квантовой семантике мы можем представить его в виде суперпозиции (И) и (Л), где они существуют одновременно до измерения или задания значения.



Когда мы измеряем или задаем значение этого утверждения, происходит интерференция между (И) и (Л). Возможны два варианта интерференции.

1. **Конструктивная интерференция:** (И) и (Л) могут взаимодействовать таким образом, что смысл утверждения может быть интерпретирован как «это утверждение истинно и ложно одновременно». Это может привести к парадоксальной ситуации, известной как парадокс лжеца.
2. **Деструктивная интерференция:** (И) и (Л) могут взаимодействовать таким образом, что смысл утверждения может быть интерпретирован как «это утверждение не имеет определенного значения». Это может привести к ситуации, когда невозможно определить, истинно ли утверждение или ложно.

Это всего лишь пример, и реальные применения квантовой семантики требуют более сложных анализов и моделей. Однако он помогает нам понять, как интерференция может играть роль в определении значения и интерпретации понятий в контексте квантовой семантики.

А что дальше?

В итоге семантика Бытия и Небытия представляет собой сложную проблему, требующую глубокого философского анализа. Определение этих понятий и их взаимодействие вызывают множество вопросов и вызовов для понимания и мышления и у машин, и у человека.

Через исследования и обсуждения мы сможем приблизиться к пониманию природы Бытия и Небытия, расширяя наши и машинные знания, глубину мысли и понимание смысла.

BEYOND SEMANTICS, OR THE SEMANTICS OF BEING AND NON-BEING

Kulai A. A., member of the Council of Kostroma VOIR (All-Russian Society of Inventors and Innovators), Kostroma

This article presents an unusual and largely philosophical report by the author with elements of mysticism and metaphysical orientation. Questions on the topic can be sent to the author at: 31pk09ksq67t@mail.ru

philosophy • quantum semantics • Being • Non-being • interference

Семантика на базе технологий искусственного интеллекта для инженера¹

Мещеряков Р. В., доктор технических наук, профессор, главный научный сотрудник, заведующий Лабораторией «Киберфизических систем», Федеральное государственное бюджетное учреждение науки «Институт проблем управления имени В. А. Трапезникова» Российской академии наук, Москва

В статье представлен практический подход к использованию семантики. В качестве методологической базы используются наивные понятия семантического треугольника Фреге и простейших схем обработки информации в программных и робототехнических системах. Приводятся примеры использования семантики для функционирования систем и делаются выводы о перспективах развития.

семантика • программа • робот • интерфейс

ВВЕДЕНИЕ

Развитие технологий искусственного интеллекта неуклонно расширяет сферу применения реальных систем обработки различной слабоструктурированной информации. При этом инженеры и технологи получают в руки инструментарий для решения прикладных задач. Вместе с тем объемы обрабатываемых данных существенно превышают возможности человека и возникают вопросы их агрегирования с формированием соответствующих выводов и решений — как это происходит при моделировании языковой деятельности [1].

Для работы в слабоструктурированных данных недостаточно иметь большую выборку и надежные методы классификации и кластеризации — необходима новая парадигма агрегации и интерпретации данных, которая основана на знаниях. В этой связи методы семантического анализа представляют интерес для практических работ.

ЧТО ТАКОЕ СЕМАНТИКА?

Известные определения термина «семантика» в достаточной мере отражают все многообразие смыслов и ориентированы на различные уровни представления многомодальных сигнальных конструкций (текста, речи, жестов, цветов и др.).

¹ Работа поддержана грантом РФФ 24-11-00340 «Исследование и разработка методов обработки слабоструктурированной информации на естественных языках в условиях сильных шумов для решения задач безопасности».

При реальном же использовании в инженерных прикладных системах приходится ориентироваться на более узкий набор сигнальных конструкций, в которых преобладает текст и который наиболее популярен для обработки. С другой стороны, следует отметить, что для обработки важен второй составляющий элемент семантики — предметная среда и ситуация [2–4].

Существующее одно из наиболее известных представлений взаимосвязей в виде семантического треугольника Фреге [5] лишь в некоторой части покрывает потребности инженеров и проектировщиков (рис. 1). Более сложные представления семантики могут использоваться на этапах исследования, однако для прикладных решений часто избыточны.



Рис. 1. Семантический треугольник Фреге

Представление составных частей семантики на рисунке 1 недостаточно и отражает собой работу функций отображения из одних представлений информации в другие. В указанном представлении отсутствуют подходы к представлению процессов, а также синтеза и декомпозиции предметов.

Таким образом, получаем надежную формализацию для работы, но невысокий функционал. Так, в работе [6] предлагается развитие прикладных технологий Интернета вещей с использованием семантического подхода по обработке информации в сложных системах для разработки интеллектуальных методов и предоставления интеллектуальных услуг для производства.

Можно отметить, что для ряда прикладных разработчиков семантика представляет собой выбранную категорию, смысл, значение, информацию о содержании и их связи(ях), а также его представлении в системе.

В качестве примеров прикладных систем выберем программные и робототехнические системы.

ПРОГРАММНАЯ СИСТЕМА

Очевидный вопрос о наделении машины мыслительными процессами задается очень давно и связан с попыткой передать машине функций человека [7]. Вместе с тем в реальной жизни имеющиеся ресурсы и технологии искусственного интеллекта позволяют проводить выборку и сопоставление различных прикладных понятий на различных языках и проводить сложные семантические отношения. Примером такой системы является система по реализации семантической реляционной модели международного китайского основного словаря в сценариях искусственного на основе лингвистической теории и технологии компьютерной обработки [8].

Практически в любой программной системе информация обрабатывается по схеме, представленной на рисунке 2.



Рис. 2. Схема обработка данных в программной системе

Представленная система наивна и не показывает возможность наличия значительного количества обратных связей и более сложных архитектурных вычислительных конструкций, а также возможность наличия нескольких входов, выходов и внутренних состояний (по аналогии с работой формальных автоматов).

В процессе обработки данных важна не только семантика собственно данных — входных, выходных и внутренних, но и собственно семантики процессов обработки этих данных. Для реализации семантики в программной системе по указанной модели достаточно иметь представление, которое будет однозначно интерпретировано по модели, представленной на рисунке 1.

Наличие семантических коллизий, как правило, решается на уровне интерпретации обработчика программной системы. Следует отметить, что исходные данные имеют небольшое количество коллизий, которые также решаются на уровне предобработки. Выходные данные однозначно идентичны и представляют собой уже значения, которые имеют однозначный смысл, который определен логикой работы алгоритмов программного обеспечения и входными данными.

РОБОТОТЕХНИЧЕСКАЯ СИСТЕМА

Более сложной системой, которая включает в себя программную часть, является робототехническая система или просто робот. Как правило, у робота имеется широко

развитая сенсорная подсистема и они могут работать в группе [9, 10]. По аналогии с программной системой можно представить обработку информации в виде схемы, представленной на рисунке 3.

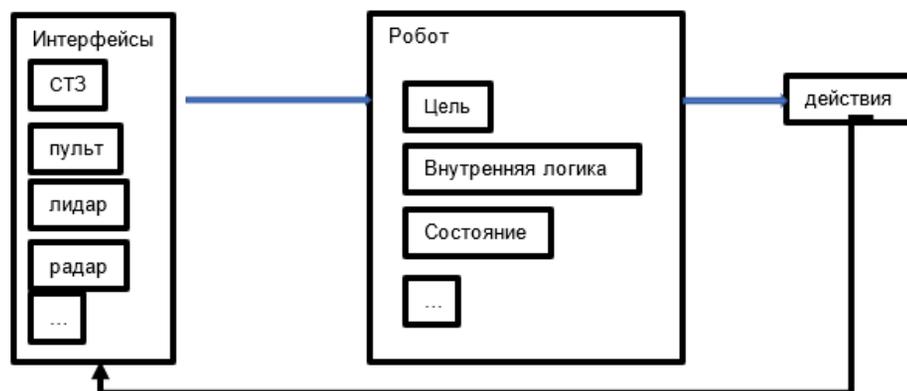


Рис. 3. Схема обработки данных в робототехнической системе

В отличие от чисто программной системы, поступающая информация еще более слабо формализована, источники имеют различную природу, а установление семантического соответствия между предметом и предполагаемым смыслом данных важно для функционирования робота. Представленные на рисунке 3 источники данных имеют различную точность, пространственное разрешение, скорость поступления, законы распределения ошибок и пр.

Другой вопрос, который задается при работе робототехнической системы, — это обнаружение и классификация объектов интереса. Например, для робототехнических систем, которые перемещаются в пространстве, важно определить, является ли препятствие критическим (наезд на него остановит робота или будут нанесены повреждения) или нет (можно перемещаться через него). Для роботов-манипуляторов важно иметь информацию об окружении — в том числе наличие человека в зоне действия для ненанесения вреда [11].

Информация, поступающая с систем технического зрения и другой сенсорики, должна быть предварительно обработана и верифицирована [12, 13], иначе получаемые для работы робота данные приведут к несоответствующим алгоритмам обработки.

Отдельной задачей стоит взаимодействие человека с роботом с использованием интерфейсов [14, 15], и это требует однозначного определения семантики выдаваемых команд роботу. Неверная команда или ее семантическая интерпретация может привести к непоправимым последствиям. Как правило, в указанной задаче дополнительно требуют от робота подтверждение о принятой команде.

ЗАКЛЮЧЕНИЕ

Предлагаемый инженерный подход к понятию семантики позволяет ставить задачи исследователям на уровне конкретных прикладных решений, что является чрезвычайно важным в настоящее время в связи с увеличением программных и робототехнических систем с искусственным интеллектом.

Для прикладных решений следует уделить большее внимание представлению семантики внутри технических систем и разрешению возникающих коллизий на различных уровнях систем обработки информации.

Перспективными представляются направления учета многообразия сторон информации и его генерации для применения, в том числе с использованием генеративных моделей технологий искусственного интеллекта.

Список использованных источников

1. Моделирование языковой деятельности в интеллектуальных системах. Серия: Проблемы искусственного интеллекта / Ред. Кибрик А.Е. – М.: Наука, 1987. – 280 с.
2. Харламов, А. А. Семантика текста как модель ситуации // Речевые технологии. – 2019. – № 1. – С. 14.
3. Михайлова, М. Ю. Семантика невыразимого и смежные явления // Известия Самарского научного центра Российской академии наук. – 2015. – Т. 17, № 1. – С. 963–966.
4. Мещеряков, Р. В., Бондаренко, В. П. Диалог как основа построения речевых систем // Кибернетика и системный анализ. – 2008. – Т. 44. № 2. – С. 30–42.
5. Чулахин, Н. П. Семантический треугольник Готлоба Фреге и семантический тетраэдр в философии смысла // Вестник Новосибирского государственного педагогического университета. – 2012. – Т. 2, № 1. – С. 59–67.
6. Patel P., Ali M. I. and Sheth A. "From Raw Data to Smart Manufacturing: AI and Semantic Web of Things for Industry 4.0," in IEEE Intelligent Systems, vol. 33, no. 4, pp. 79–86, Jul./Aug. 2018, doi: 10.1109/MIS.2018.043741325.
7. Тьюринг, А. Может ли машина мыслить? / Пер. с англ. Ю.В. Данилова. – М.: ГИФМЛ, 1960.
8. Zhu Y. "Semantic Relationship Model of Chinese International Core Vocabulary in Artificial Intelligence Scenarios" 2023 IEEE 3rd International Conference on Social Sciences and Intelligence Management (SSIM), Taichung, Taiwan, 2023, pp. 312–315, doi: 10.1109/SSIM59263.2023.10468710.
9. Кутахов, В. П., Мещеряков, Р. В. Управление групповым поведением беспилотных летательных аппаратов: постановка задачи применения технологий искусственного интеллекта // Проблемы управления. – 2022. – № 1. – С. 67–74.
10. Chueshev A., Melekhova O., Meshcheryakov R. Cloud robotic platform on basis of fog computing approach // Lecture Notes in Computer Science. 2018. Т. 11097 LNAI. С. 34–43.
11. Galin R. R., Meshcheryakov R. V. Human-robot interaction efficiency and human-robot collaboration // Studies in Systems, Decision and Control. 2020. Т. 272. С. 55–63.
12. Wagenpfeil S. , Hemmje M. "Towards AI-based Semantic Multimedia Indexing and Retrieval for Social Media on Smartphones," 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA, Zakynthos, Greece, 2020, pp. 1–9, doi: 10.1109/SMA49528.2020.9248445.
13. Wang H., Jiang Z., Xie L., Jiang D., Shen W. , Tian Q. "Domain-Adaptive Semantic Segmentation Emerges From Vision-Language Supervised Domain-Debiased Self-Training," ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 3930–3934, doi: 10.1109/ICASSP48485.2024.10447308.



14. Харламов, А. А., Ермашин, К. В. Речевой диалог с коллаборативным роботом на основе многомодальной семантики // Вестник Московского государственного лингвистического университета. Языкознание. Прагмалингвистическое многоголосие. – 2016. – № 15. – С. 206–215.
15. Мещеряков, Р. В., Туровский, Я. А. Перспективные направления развития человеко-машинных интерфейсов // В сборнике: Информационные технологии в науке, образовании и производстве (ИТНОП-2023). Сборник трудов IX Международной научно-технической конференции. – Белгород, 2023. – С. 23–29.

SEMANTICS ON ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR AN ENGINEER

***Meshcheryakov R. V.**, Doctor of Technical Sciences, Professor, Chief Researcher, Head of the Laboratory of Cyberphysical Systems, Federal State Budgetary Institution of Science V. A. Trapeznikov Institute of Management Problems of the Russian Academy of Sciences, Moscow*

This article presents a practical approach to using semantics. The naive concepts of Frege's semantic triangle and the simplest information processing schemes in software and robotic systems are used as a methodological basis. Examples of the use of semantics for the functioning of systems are given and conclusions are drawn about development prospects.

semantics • program • robot • interface

Семантика коммуникативной среды города — информационные технологии городского дизайна

Расходчиков А. Н., кандидат социологических наук, председатель правления фонда «Московский центр урбанистики «Город», silaslowa@mail.ru, ORCID: 0000-0001-6814-9029

В статье представлены теоретические основания и систематизированы практические методы исследования коммуникационной среды городов. Продемонстрированы возможности изучения текстовой информации социальных сетей и анализа Больших данных для оценки ключевых компонентов городских коммуникативных процессов. В качестве ключевых компонентов, позволяющих формировать представление о коммуникационной среде города и отслеживать динамику ее изменений, выделены публичные коммуникации и процессы группообразования в социальных сетях, местные топонимы, городские события и символные объекты.

информационные технологии • цифровизация • информационно-коммуникационная среда города • семантика городской среды • дизайн коммуникаций • городской дизайн

Современный город — это пространство коммуникаций, особенно это утверждение справедливо, если речь идет о крупных городах, мегаполисах и мировых центрах, где пересекаются международные потоки людей, информации и капиталов. В то время когда старые функции городов: обороны, торговли, промышленного производства, финансовых центров — постепенно утрачивают свое градообразующее значение [1], а сами города теряют форму упорядоченной интуитивно понятной структуры, формирующей миропонимание [2], коммуникативная среда, с развитием информационных технологий, наоборот, получает стремительное развитие, превращая коммуникативность в одну из главных характеристик городской среды. Происходящие изменения требуют теоретического осмысления этой составляющей городской жизни, а также новых методов изучения каналов коммуникации, значения отдельных элементов городской среды, символов и знаков, используемых для взаимодействия различных акторов и сообществ. Что, в свою очередь, актуализирует необходимость разработки такого направления, как семантика городской среды.

О ключевой роли социальной коммуникации в цивилизационном развитии человека достаточно подробно писал Вернадский. Появление городов привело к более интенсивному взаимодействию людей, росту сложности систем и коммуникаций, появлению правил совместной жизни и деятельности, торговли, административных структур управления, развитию науки и искусств [3]. Эти идеи получили свое развитие в научных концепциях Т.М. Дридзе — семиосоциопсихологии [4]. Характерно, что параллельно этому направлению, связанному с языком, Тамара Моисеевна развивала свою экоантропоцентрическую парадигму городского развития [5].



Такое сочетание неслучайно. Так, именно на примере городской среды мы можем наблюдать и исследовать коммуникативные процессы в самом концентрированном и многомерном проявлении. Большое внимание к коммуникационной составляющей городской среды в своих работах проявляют и современные исследователи, такие как Р. Флорида. На примерах масштабных статистических исследований городов США Флорида демонстрирует как уже не концентрацию финансов или административной власти, а именно коммуникативные характеристики городов становятся определяющими для привлечения наиболее ценных специалистов и формирования креативного класса [6]. Это направление продолжают в своих работах Винод Кумар, определяющий важность коммуникационной предпринимательской среды для мировых городов [7], и В. Алексеев, который обращает внимание на значения качества коммуникативной среды города для привлечения инвестиций [8]. Инфраструктура коммуникаций и качество взаимодействий выделяются рядом исследователей как ключевые факторы в стратегиях «умных городов» [9], для целей туристического развития территорий [10], в проектах оздоровления городской среды [11].

Теоретические основания для исследования коммуникационной среды города заложены в работах Т.М. Дридзе — в описании информационной составляющей среды жизнедеятельности горожан, А.В. Тихонова — посвященных реализации механизмов социального участия и субъект-субъектной модели управления [12], Л. Мановича — в описании механизмов визуального анализа городских коммуникаций [13], Д. Хенсон и Б. Хиллера — в теоретической модели синтаксиса городских пространств (Space syntax) [14]. Однако представленные основания базируются на очень разных теоретических подходах и относятся к мало связанным между собой областям знания. В результате целостная теоретическая картина понимания коммуникационной среды города, ее семантики не формируется, равно как и практические методы исследования не складываются в единую систему.

Цель настоящей работы — обозначить концептуальные рамки исследовательской программы семантического анализа коммуникационной среды городов. Для этого необходимо описать область, входящих в это направление объектов, явлений и фактов; выделить ключевые компоненты, позволяющие формировать наиболее значимые характеристики коммуникационной среды города и отслеживать динамику ее изменений, а также предложить наиболее подходящие методы для изучения данного направления.

Коммуникативная среда города охватывает широкий круг информационных сообщений, символов, знаков, значений, окружающих человека в городе и использующихся для взаимодействия горожан и городских сообществ. Коммуникативная среда включает в себя все виды коммуникации между людьми, которые происходят в городе: устную и письменную речь, межличностное общение (в том числе в мессенджерах и социальных сетях), сообщения СМИ, рекламу, а также символичные сообщения через архитектурные формы, памятники, историческое наследие, названия улиц и городских районов, культурные мероприятия,

городские праздники и фестивали, уличное искусство местные бренды и символы. Иногда даже местная мода и кухня могут нести в себе определенные послания для приезжих.

Т. М. Дридзе описывала данную область исследования как информационную составляющую среды жизнедеятельности человека в городе — одну из четырех составляющих сложного процесса обмена веществом и энергией между человеком и создаваемой им средой. Данный подход также перекликается с современными направлениями дизайна городской среды (Urban design), такими как синтаксис пространства (Space syntax), основой которого стали работы Джулиан Хенсон и Билла Хиллера «Социальная логика пространства» [14] и другие, активно использующиеся по сей день, например в работах Майкла Бетти, исследующего города как системы сетей и потоков [15] А также предложенной Скоттом Кэмпеллом моделью «треугольника планирования», основанной на взаимодействии целей экономического развития, сохранения природной среды и социальной справедливости [16].

Однако современный город — система систем, где одновременно взаимодействуют сотни процессов и происходят тысячи интеракций. Даже если рассматривать отдельно коммуникативную среду как одну из городских систем, количество взаимодействий здесь настолько велико и разнородно, что охватить всю эту область в рамках исследования не представляется возможным. Отсюда возникает важная задача выделения ключевых компонентов коммуникативной среды, анализ которых позволяет составить представление об основных ее характеристиках и происходящих изменениях. К таким ключевым компонентам или узлам можно отнести:

- **социальные сети** — в части публичных коммуникаций и процессов группообразования как открытый и доступный источник данных межличностных и межгрупповых речевых взаимодействий;
- **местные топонимы** (названия улиц, районов, жилых комплексов и деловых центров) — позволяющие понять, какие исторические события, личности и идеи влияют на конкретный город, а также изучать закономерности взаимодействия новых названий и социальных изменений;
- **городские события**, праздники и ивент-мероприятия — как наиболее концентрированные проявления культурной и деловой общественной жизни города, места коллективного взаимодействия и репрезентации;
- **символьные объекты** — являющиеся значимой частью городской идентичности и даже самоидентичности коренных жителей.

Возможно, этот список нуждается в дополнении и будет в последующем скорректирован по результатам исследований. Однако на данный момент на этапе формулировки концептуальных рамок исследовательской программы он выглядит необходимым и достаточным. Для иллюстрации стоит привести примеры исследований перечисленных основных компонентов коммуникативной среды городов. Эти примеры демонстрируют наличие и разработанность методов анализа, доступность данных, а также практическую применимость полученных выводов.

Исследования коммуникаций в социальных сетях можно условно разделить на два основных направления: сетевой анализ процессов группообразования в сети

Интернет и семантический анализ текстов. Первый метод хорошо представлен в работах Г.В. Градосельской, использующую собственные авторские методики зернового приращения, анализа информационных волн и связей между группами в социальных сетях [17]. На представленном ниже рисунке 1 можно наблюдать сетевую проекцию города Томска в социальной сети «ВКонтакте2». Различные по целям создания, составу участников и тематике обсуждения группы и публичные страницы здесь структурируются в отдельные тематические кластеры на основе взаимного участия пользователей. Исследование раскрывает картину межличностных и межгрупповых онлайн-коммуникаций жителей Томска позволяет выявить структуру интересов горожан, изучать тенденции связанности и разобщенности городских сообществ и другие аспекты межличностных и межгрупповых взаимодействий. Контент-анализ публикаций и комментариев в городских онлайн-сообществах позволяет изучать актуальные городские проблемы, качество взаимодействий органов власти с городскими сообществами, различия в восприятии города и прочие аспекты городской жизни.



Рис. 1. Результаты картирования групп в социальной сети «ВКонтакте» города Томска

Исследования речевых коммуникаций в сети Интернет представлены в работах М.А. Пильгун, где классические методы лингвистического анализа текстов сочетаются с результатами обработки больших массивов текстовых коммуникаций при помощи нейросетевой технологии TextAnalyst . В качестве примера можно привести исследование «Образ будущего по данным социальных медиа» (рис. 2), где на основе анализа эксплицитной и имплицитной информации в текстовой коммуникации пользователей изучались социальные ожидания в отношении будущего городов. Данный метод позволяет выявлять наиболее значимые мнения и оценки пользователей в отношении различных тем городской жизни, а также исследовать закономерности коллективного восприятия горожанами отдельных проблем и событий [18].



Рис. 2. Результаты анализа ассоциативной сети

Довольно интересные подходы к изучению городских событий, праздников и ивент-мероприятий содержатся в работах А.Н. Языкеева, где на основе анализа Больших данных (Big Data) демонстрируются возможности оценки аудиторий городских мероприятий, отношения к ним со стороны местных жителей и туристов, а также доказываются влияние событийных программ с притоком инвестиций и активизацией инфраструктурного развития городов [19]. Представленная ниже диаграмма (рис. 3) демонстрирует масштабность проведенных в Калининграде мероприятий, соотношение местных жителей и туристов среди их участников, а также уровень внимания к событиям со стороны СМИ.

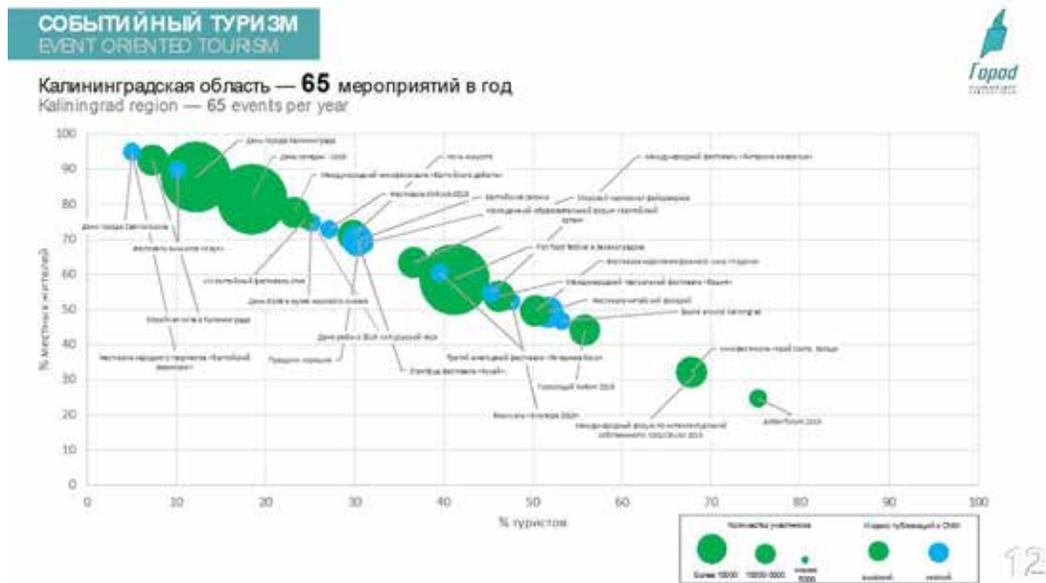


Рис. 3. Оценка соотношения туристов и местных жителей — участников массовых мероприятий в Калининградской области

Возможности изучения символического поля городов можно проиллюстрировать на результатах исследования в исторических районах Москвы Лефортово-Басманный [20]. На представленных ниже диаграммах (рис. 4) можно видеть, как даже в небольших районах мегаполиса можно выявлять значимые для жителей объекты и исторические персоны. Влияние символического поля на идентичность горожан здесь можно проследить в разнице ответов коренных жителей и недавно переехавших в районы. Исследование, в частности, продемонстрировало, что «новые горожане» при выборе исторически значимых личностей чаще ориентируются на название улиц, местные топонимы, а коренное население — на реальных персон, связанных с историей городских районов. В практическом смысле такие исследования позволяют не только работать с имиджем и брендингом территории, но и предотвращать градостроительные ошибки, связанные с непродуманными проектами обновлений и реновацией значимых для жителей объектов и пространств.



Рис. 4. Результаты исследования символических объектов районов Лефортово и Басманный Москвы

Примеры исследования взаимодействия топонимов и происходящих в городах социальных изменений пока не встречаются в научной литературе. В то же время это интересное направление исследований выглядит вполне осуществимым благодаря методической модели А.А. Харламова, реализованной в программном продукте TextAnalyst [21–23]. Разработанная им программная система позволяет проводить семантический анализ больших объемов текстов. Для целей данного исследовательского направления большой интерес представляет способность программы выстраивать сеть, отображающую частоту употребления топонимов, сочетаемость их с другими элементами и отслеживать частоту таких сочетаний. Это очень важный аспект для анализа коммуникационной среды города как сложного и многоуровневого явления с большим количеством взаимодействий. Именно связи между отдельными элементами городской среды. Эти связи

на практике довольно сложно выявить и еще сложнее анализировать, особенно когда такие связи образуются между объектами различной природы, например архитектурными комплексами и рекламными посланиями или историческими персонами и местной идентичностью. Именно связи между различными элементами городской среды представляют наименее изученную область в городских исследованиях.

ВЫВОДЫ

Семантика городской среды представляется значимым направлением исследований, позволяющим выявлять наиболее значимые характеристики коммуникативной среды городов, оценивать влияние историко-культурного наследия, местной символики, ценностей и идей на формирование городской идентичности, условия социализации приезжих. Одна из основных задач данного направления — создание каналов и пространств, которое способствуют эффективной коммуникации и взаимодействию между людьми и городскими сообществами. Область интересов данного направления охватывает широкий круг информационных сообщений, знаков и символов, окружающих человека в городе и использующихся для взаимодействия горожан и городских сообществ.

Примеры исследований коммуникативной среды городов, приведенных в работе, демонстрируют наличие и разработанность методов анализа, доступность данных, а также практическую применимость полученных выводов. Одним из важных ориентиров исследования коммуникационной среды города становится оценка тенденций к сплочению городских сообществ или, наоборот, формированию разрывов, неравенства и конфликтов. В практическом плане эти аспекты могут быть использованы в процессе городского и территориального планирования, создания и обновления городских пространств.

Семантика городской среды как исследовательское направление открывает возможности для гармонизации городской среды через изучение сочетаемости топонимов, качества связей между различными явлениями городской жизни. Результаты исследований коммуникационной среды городов могут использоваться как основание для культурно и исторически обоснованного брендинга территорий, оценки новых названий с точки зрения их соответствия городской идентичности.

Список использованных источников

1. *Расходчиков, А. Н.* Контекст и ключевые вопросы современного осмысления образов городов будущего: обзор конференции / А. Н. Расходчиков // *Abyss* [Вопросы философии, политологии и социальной антропологии]. – 2023. – № 2. – С. 193–205. – DOI 10.33979/2587-7534-2023-2-193-205. – EDN IGGDYC.
2. *Холодная, А. А.* Семантическое пространство города в современной российской культуре / А. А. Холодная // *Культура и цивилизация*. – 2017. – Том 7. – № 2А. – С. 497–503.
3. *Вернадский, В. И.* Биосфера и ноосфера / В. И. Вернадский. – М.: Айрис-пресс, 2012. – 576 с. – (Библиотека истории и культуры). – ISBN 978-5-8112-4512-3.
4. *Адамьянц, Т. З.* Семиосоциопсихологическая концепция социальной коммуникации в социальном познании (материалы круглого стола) / Т. З. Адамьянц // *Социологические исследования*. – 2015. – № 12. – С. 142–144. – EDN VGMGAJ.



5. Дридзе, Т. Градоустройство: от социальной диагностики к конструктивному диалогу заинтересованных сторон / Т. Дридзе. – М.: Изд-во «Институт психологии РАН», 1998.
6. Florida, R. (2017). *The new urban crisis: How our cities are increasing inequality, deepening segregation, and failing the middle class – And what we can do about it*. New York, Basic Books, ISBN: 978-0-46-507974-2
7. Vinod Kumar, T. M. *Indo-Pacific Smart Megacity System*. Series: *Advances in 21st Century Human Settlements*, Year: 2023, Page 429, DOI: 10.1007/978-981-99-6218-1_8
8. Алексеев, В. Н. Формирование инвестиционного проекта и оценка его эффективности: учебно-практическое пособие / В. Н. Алексеев, Н. Н. Шарков. – 4-е изд. – Москва: Издательско-торговая корпорация «Дашков и К», 2020. – 176 с.
9. Ильина, И. Н. Трансформация подходов к развитию «умного города» / И. Н. Ильина, М. Коно. – Москва: Национальный исследовательский университет «Высшая школа экономики», 2023. – 248 с. – ISBN 978-5-7598-2579-1. – DOI 10.17323/978-5-7598-2579-1. – EDN WCUQKP.
10. Языкеев, А. Н. Возможности анализа больших данных для оценки ивент-мероприятий / А. Н. Языкеев, С. В. Дусенко // *Актуальные проблемы развития туризма: Материалы VII Международной научно-практической конференции*, Москва, 15–16 марта 2023 г. – Москва: РУС «ГЦОЛИФК», 2023. – С. 127–133. – EDN CVMFKQ.
11. Концевая, А. В. Формирование городской здоровьесберегающей среды / А. В. Концевая, А. А. Анциферова, Д. К. Муканеева // *Оздоровление городской среды*. – М.: Фонд «Московский центр урбанистики “Город”», 2022. – С. 88–95.
12. Тихонов, А. В. Социология управления / А. В. Тихонов. – Издание второе, дополненное и переработанное. – Москва: Канон+, 2007. – 427 с.
13. Манович, Л. Язык новых медиа / Лев Манович. – М.: АД Маргинем Пресс, 2018. – 400 с.
14. Hillier B., Hanson J. (1984). *The Social Logic of Space*. Cambridge University Press. p. 281. doi:10.1017/CBO9780511597237
15. Batty, M. (2013). *The New Science of Cities*. Cambridge, MA, The MIT Press, 496 pp. ISBN 978-0-262-01952-1, DOI: 10.7551/mitpress/9399.001.0001
16. Кэмпбелл С. Зеленые города, растущие города или справедливые города? Городское планирование и противоречия устойчивого развития (5 сентября 2021 г.). URL: <https://архитекторы.рф/articles/skott-kempbell-zelenye-goroda-rastuschie-goroda-ili-spravedlivye-goroda-gorodskoe-planirovanie-i-protivorechiya-ustoychivogo-razvitiya>
17. Градосельская, Г. В. Два сценария будущего молодежи: результаты картирования групп социальной сети «ВКонтакте» на примере Томска / Г. В. Градосельская, А. Н. Расходчиков // *Вестник Московского университета. Серия 12: Политические науки*. – 2020. – № 4. – С. 50–68. – EDN GFHULB.
18. Пильгун, М. А. Образ будущего по данным социальных медиа / М. А. Пильгун // *Вторые Шанявские чтения – «Образ общего будущего в дискурсах гендерных исследований стран СНГ и России»: доклад на международной научно-практической конференции*. URL: <https://urbanru.ru/events/obraz-goroda-budushhego/>
19. Языкеев, А. Н. Анализ текстовых сообщений в социальных сетях для оценки эффективности ивент-мероприятий / А. Н. Языкеев // *Речевые технологии*. – 2023. – № 4. – С. 66–72. – EDN XCPZII.
20. Расходчиков, А. Н. «Новые горожане»: инклюзивная политика и социокультурная динамика исторических городов / А. Н. Расходчиков // *Визуальная антропология – 2022. Исторический город: актуализация прошлого в перспективе*

будущего: Материалы IV Международной научной конференции, Великий Новгород, 22–23 сентября 2022 г. / Под ред. С. С. Аванесова, Е. И. Спешиловой. – Великий Новгород: Новгородский государственный университет имени Ярослава Мудрого, 2022. – С. 158–164. – DOI 10.34680/visant-2022-158-164. – EDN MQECRS.

21. Харламов, А. А. Анализ текстов с использованием искусственных нейронных сетей на основе нейроподобных элементов с временной суммацией сигналов (часть 1) / А. А. Харламов // Речевые технологии. – 2023. – № 1. – С. 87–99. – EDN: KXYSCE
22. Харламов, А. А. Анализ текстов с использованием искусственных нейронных сетей на основе нейроподобных элементов с временной суммацией сигналов (часть 2) / А. А. Харламов // Речевые технологии. – 2023. – № 2. – С. 22–39. – EDN: UEEXXM
23. Харламов, А. А. Анализ текстов с использованием искусственных нейронных сетей на основе нейроподобных элементов с временной суммацией сигналов (часть 3) / А. А. Харламов // Речевые технологии. – 2023. – № 4. – С. 3–21. – EDN YTTUAY.

SEMANTICS OF THE COMMUNICATIVE ENVIRONMENT OF THE CITY — INFORMATION TECHNOLOGIES OF URBAN DESIGN

Raskhodchikov A. Ni., PhD (sociological sciences), Chairman of the Board, The Moscow Center of Urban Studies “City”, silaslova@mail.ru, ORCID: 0000-0001-6814-9029

The article presents theoretical foundations and systematizes practical methods of studying the communication environment of cities. The possibilities of studying textual information of social networks and analyzing big data to evaluate key components of urban communication processes are demonstrated. Public communications and group formation processes in social networks, local toponyms, urban events and symbolic objects are highlighted as key components that allow to form an idea of the city communication environment and monitor the dynamics of its changes.

information technologies • digitalization • information and communication environment of the city • semantics of urban environment • communication design • urban design



ChatGPT — это интеллектуальная система?

Харламов А. А., доктор технических наук, старший научный сотрудник Института высшей нервной деятельности и нейрофизиологии РАН, профессор кафедры прикладной и экспериментальной лингвистики МГЛУ, профессор департамента программной инженерии ВШЭ, профессор кафедры интеллектуальных информационных систем и технологии Московского физико-технического института, Москва, kharlamov@analyst.ru

В статье предпринята попытка сформулировать представление об интеллектуальных системах с учетом представлений в структуре и функциях мозга человека. Показано, что интеллектуальные системы моделируют две основные интеллектуальные функции человека: синтез осмысленных связных текстов и квази-текстов и анализ текстов и квази-текстов. Это системы, основанные на представлении знаний. Формализация понятия «интеллектуальная система» позволила оценить структуру и функциональность GPT 3.5, на основе которой была реализована система ChatGPT, что, в свою очередь, позволило оценить функциональность этой последней системы. Показано, что ChatGPT реализует функцию генерации текстов, присущую интеллектуальным системам, но эта генерация осуществляется на основе исходного содержательного материала, который вносится в систему извне. То есть если говорить об интеллектуальности ChatGPT содержательно, то можно сказать, что ChatGPT реализует синтез текстов на уровне ниже прагматики, если под семантикой понимать допустимую сочетаемость образов объектов и событий в модели мира.

интеллект • интеллектуальные системы • лингвистический процессор • экстралингвистический процессор • ChatGPT • TextAnalyst

ВВЕДЕНИЕ

Когда мои студенты затрудняются ответить на вопрос: «Что такое интеллектуальные системы?», я даю им наводящую подсказку: «Интеллектуальные системы — это системы, реализующие интеллектуальные функции». И поскольку ответ на вопрос, что такое интеллектуальные функции, они знают, то с легкостью отвечают на вопрос: «Что такое интеллектуальные системы?».

На самом деле ответ на вопрос, какие интеллектуальные функции присущи человеку, не столь прост, как может показаться на первый взгляд. Давайте рассмотрим его подробно. Более или менее стандартная

отсылка к интеллектуальным функциям человека включает две основные функции: **понимание текстов естественного языка и синтез таких текстов.**

1. ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

Чтобы говорить об **интеллектуальных системах**, надо понять, что из себя представляет интеллект. С разных точек зрения он имеет разные определения.

Определение 1. В БСЭ интеллект определяется как **способность мышления, рационального познания**, в отличие от таких, например, душевных способностей, как чувство, воля, интуиция, воображение и т.п. [1].

Определение 2. В англо-русском толковом словаре по искусственному интеллекту и робототехнике 2022 г. издания [2] интеллект определяется как **способность мозга решать интеллектуальные задачи путем приобретения, накопления и упорядочивания знаний, их адаптации в процессе обучения и на опыте к различным ситуациям.**

Определение 3. Таким образом, интеллект — это **способность рационального познания или свойство выполнять творческие функции, которые базируются на знаниях.**

Формулировки нейроинформатики позволяют продемонстрировать структуру и функциональность естественного интеллекта в части обработки специфической информации [3, 4, 5].

Определение 4. С точки зрения нейроинформатики (с точки зрения обработки информации в мозге человека) под **интеллектуальной системой** можно понимать систему, обладающую **свойствами интеллекта**: (1) реализовать две основные интеллектуальные функции человека — понимание текстов языка и синтез осмысленных связных текстов — **вторая сигнальная система**; и (2) реализовать две другие интеллектуальные функции человека — понимание квазитекстов и синтез осмысленных связных квазитекстов — **первая сигнальная система.**

Определение 5. По этой причине в структуру интеллектуальной системы (рис. 1) входят: (1) **лингвистический** и (2) **экстралингвистический процессоры**, имеющие в своем составе знания, представленные в виде словарей образов событий разной сложности разных модальностей, а также правил их использования, как поуровневых, так и межуровневых.

Таким образом, интеллектуальность функции системы (и самой системы) определяется моделью мира, которую эта функция использует, то есть представленными в ней знаниями.

1.1. Функциональность интеллектуальных систем

Модель мира, которая формируется и используется в процессе функционирования интеллектуальных систем, состоит из двух частей: языковой и экстралингвистической.

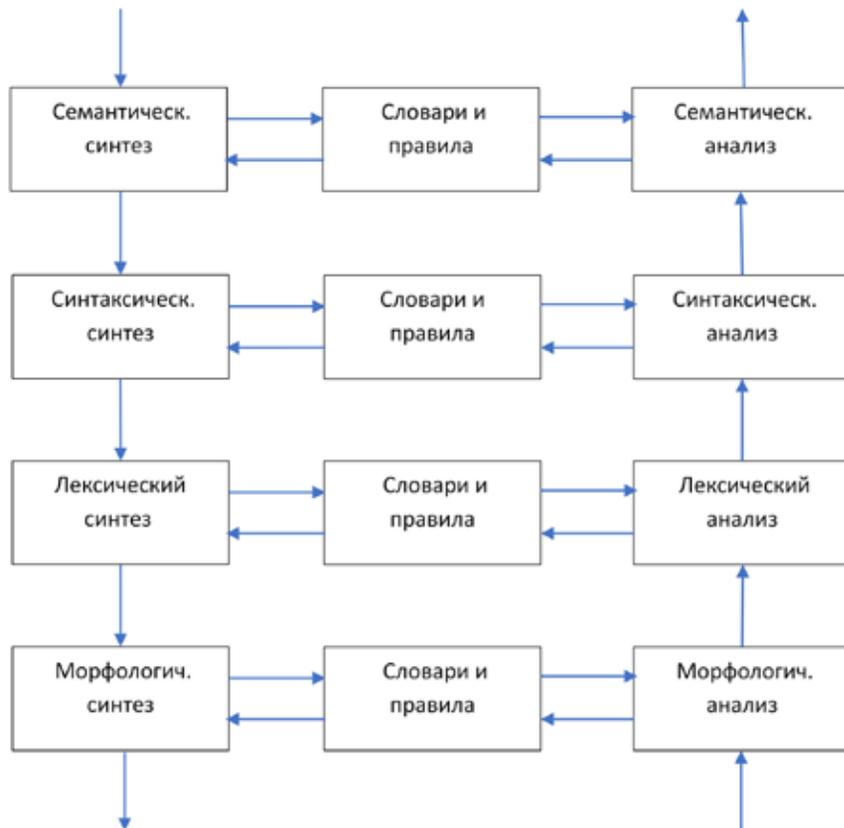


Рис. 1. Лингвистический процессор. Здесь справа — поуровневый анализ (всех уровней языка, снизу-вверх), на входе текст, на выходе — представление ситуации в терминах шаблонов ситуаций, слева — поуровневый синтез (от семантики до морфологии сверху-вниз), вверху — представление ситуации, внизу — текст, ее описывающий, в центральной колонке система поуровневых словарей и правил их использования как внутри уровня, так и между уровнями

Определение 6. Языковая часть модели мира (лингвистический процессор) представляет собой: (1) иерархию словарей уровнеобразующих элементов языка, верхний уровень которой представляет словарь допустимой сочетаемости слов (точнее, корневых основ слов, описывающих объекты и события мира), который характеризует допустимую сочетаемость объектов и событий мира в терминах экстралингвистических представлений, а нижние уровни учитывают лингвистические особенности языка, используемого для этого описания; (2) систему правил манипулирования словами словарей как на уровнях, так и в процессе взаимодействия уровней.

Лингвистический процессор называется «процессором» не случайно. В нем в процессе обучения происходят поуровневые процессы формирования поуровневых словарей и правил манипулирования словами этих словарей, а также правил взаимодействия между словарями разного уровня [3] и их использование в процессе анализа-синтеза

текстов. Состав словарей анализирующей части симметричен составу словарей синтезирующей части процессора.

В процессе анализа формируются последовательно словари всех уровней: от морфемного до семантического (самые нижние уровни — акустико-фонетический и графематический — здесь не рассматриваются, так как они различны для звучащей и написанной речи).

Перечень допустимых пар корневых основ слов (на семантическом уровне) виртуально объединяется в направленную однородную взвешенную семантическую сеть. В случае анализа конкретного текста эта семантическая сеть описывает содержание этого текста. Когда модель мира строится на большом корпусе текстов, эта сеть описывает подязык, представленный в этом корпусе текстов.

Семантическая сеть, построенная на очень большой текстовой выборке, является моделью языка (расширяющуюся вниз иерархией представлений лингвистического процессора), где конкретная цепочка на этой сети соответствует конкретному тексту.

Определение 7. Экстралингвистическая часть модели мира представляет собой: (1) иерархию словарей уровнеобразующих элементов экстралингвистических языков, описывающих мир, верхний уровень которой (второй сверху уровень) представляет словарь допустимой сочетаемости квази-слов, который характеризует допустимую сочетаемость объектов и событий мира в терминах экстралингвистических представлений, а нижние уровни учитывают особенности квази-языков, используемых для этого описания; (2) систему правил манипулирования словами квази-словарей как на уровнях, так и в процессе взаимодействия уровней.

В случае экстралингвистической семантической сети мы имеем модель квази-языка, аналогичную языковой модели, представляющую потенциально реализацию всех мыслимых квази-текстов этого квази-языка.

Определение 8. Квазитексты — последовательности, описывающие мир в терминах экстралингвистических модальностей, где уровни анализа-синтеза определяются специфическими уровнеобразующими элементами, характерными для конкретных неязыковых сенсорных и эффекторных модальностей. Так, для зрительной модальности это: (1) на нижнем уровне — элементарные представления — точка, сегмент дуги, отрезок, пересечения разных типов; (2) уровнем выше — фрагменты объектов; (3) далее — объекты; на верхнем уровне — сцены в составе объектов.

Верхние уровни анализирующих частей лингвистического и экстралингвистического процессоров представляют собой словари попарной сочетаемости образов событий мира в экстралингвистическом представлении и описывающего эту попарную сочетаемость словаря пар корневых основ слов в лингвистическом представлении, причем оба словаря виртуально могут быть представлены как однородные направленные взвешенные семантические сети.

Определение 9. Модель мира представляется двумя семантическими сетями (представлениями верхних уровней лингвистического и квази-лингвистического процессоров), работающими как единое целое.



2. CHATGPT

Попробуем понять, что из описанного выше, касающегося интеллектуальных систем, представлено в системе ChatGPT, то есть насколько ChatGPT является интеллектуальной системой. ChatGPT включает в свой состав в качестве компонента модель языка GPT-3.5, которая используется с целью синтеза связных текстов.

GPT (от англ. Generative Pre-trained Transformer — «**генеративный предварительно обученный трансформер**») — генеративная искусственная нейронная сеть, разработанная компанией OpenAI. GPT-3.5 — это языковая модель, реализованная на основе архитектуры трансформера. В этом случае языковая модель обучается без учителя на неаннотированном тексте. Трансформер при генерации вывода слева направо обучается максимизировать вероятность, назначенную следующему слову в обучающих данных, с учётом предыдущего контекста [6].

Доработанный вариант GPT-3, получивший название GPT-3.5, стал общедоступным через веб-интерфейс под названием ChatGPT в 2022 г. GPT-3.5 — это большая языковая модель на основе нейронной сети, обученной на большом объеме размеченного текста с использованием обучения без учителя. Несмотря на обучение на простых задачах, таких как предсказание следующего слова в предложении, нейронные языковые модели, обученные на большом объеме текстового материала, включают в свое представление большую часть синтаксиса и семантики естественного языка.

2.1. Архитектура системы GPT

Используемая в системе GPT архитектура типа Transformer включает в себя: (1) трансформерный энкодер — нейронную сеть, которая получает на вход последовательность токенов и преобразует ее в последовательность векторов признаков. Это делается с помощью множества слоев, каждый из которых выполняет множество операций, таких как **многоканальное внимание и одномерная свертка**; и (2) многоуровневый декодер. Декодер имеет несколько уровней, каждый из которых обрабатывает предыдущую часть предложения и использует ее для генерации следующего слова. Каждый уровень декодера состоит из слоев, которые выполняют множество операций, таких как внимание к прошлым токенам и многомерная свертка.

В состав модели Transformer входит механизм внимания, который позволяет сети взвешивать важность различных частей входных данных при составлении прогнозов, позволяя сети сосредоточиться на том, что наиболее важно для получения точных прогнозов. Действуя как когнитивный фильтр, механизм внимания помогает сети обрабатывать и понимать сложные данные, выявляя и выделяя наиболее релевантную информацию.

2.2. Функциональность системы

Функциональность системы сводится к моделированию языка на множестве текстов в виде сети слов, что позволяет за счет использования механизма внимания выбирать из нескольких близких цепочек слов наиболее подходящую для представления содержания, с точки зрения, с одной стороны, проецирования на эту сеть каких-то последовательностей входных реперных объектов (например, слов или последовательностей слов), а с другой стороны — выбора наиболее оптимальной (с учетом оценки весов получившихся цепочек) окончательной цепочки.

2.3. Обучение системы

Процесс обучения GPT включает в себя технику машинного обучения, называемую тонкой настройкой, которая направлена на повышение производительности предварительно обученной модели на конкретной задаче.

2.4. ChatGPT — интеллектуальная система

Из рассмотрения функциональности системы, можно заметить, что ChatGPT реализует (см. рис. 1, левая часть) синтез текста, используя модель языка в ней зашитую, включая допустимую попарную сочетаемость слов в тексте (допустимую комбинаторику сочетаемости событий и объектов в мире), то есть семантику текста. Однако формирование содержания текста не входит в функциональность системы. Из чего можно сделать вывод, что **ChatGPT реализует функцию синтеза текста по некоторой содержательной заготовке (которая в систему вносится извне)**, причем синтез оказывается тем лучше, чем более подробно эта содержательная заготовка. Это хорошо можно объяснить на примере перевода с языка на язык: чем лучше представлено содержание в тексте на входном языке, тем лучше получается качество текста на выходном языке.

Качество же синтезированного текста (форма, а не содержание этого текста) определяется мощностью выборки, использованной при обучении ChatGPT. Причем обучение системы происходит не в один этап — уже предобученная модель языка дообучается на предметную область.

3. TEXTANALYST

Чтобы оценка технологии ChatGPT не показалась субъективной, попробуем рассмотреть с тех же позиций другую технологию — автоматического анализа текстов TextAnalyst [7], чтобы оценить степень ее интеллектуальности.

Технология для автоматической смысловой обработки текстовой информации TextAnalyst основана на использовании структурных свойств текста, которые могут быть выявлены с помощью статистического анализа, реализованного на основе иерархических структур из динамических ассоциативных запоминающих устройств (ДАЗУ) — искусственных нейронных сетей на основе нейроподобных элементов с временной суммацией сигналов [3, 4, 5]. В технологии реализовано



автоматическое выявление семантики текста в виде семантической сети текста, и реализуются функции организации текстовой базы в гипертекстовую структуру, автоматического реферирования, кластеризации и классификации текстов, а также функция смыслового поиска.

Иерархические структуры из ДАЗУ являются удобным инструментом для выявления структурных свойств текстовой информации. Использование указанного инструмента позволяет автоматически, на основе анализа статистики слов и их связей в тексте, реконструировать внутреннюю структуру текста.

Статистический анализ выявляет наиболее часто встречающиеся элементы текста: слова или устойчивые словосочетания. При выявлении связей учитывается статистика попарного появления слов в предложениях текста. Далее статистические показатели пересчитываются в семантические с помощью алгоритма, подобного алгоритму перерасчета весов в сетях Хопфилда. Идея подобных алгоритмов заключается в том, что при расчете какой-то характеристики элемента сети учитываются подобные характеристики элементов в сети, с ним связанных, а также учитываются численные показатели связей. После пересчета статистических характеристик в семантические, понятия, которые мало соответствуют анализируемой предметной области, получают малый вес, а наиболее представительные наделяются высокими показателями.

Семантические веса элементов сети используются при расчете близости (релевантности) фрагментов текста к запросу со стороны пользователя системы. На их основе возможно выделение наиболее информативных участков текста. Использование ассоциативных связей элементов сети позволяет расширять поле поиска информации. Ответ на запрос пользователя в этом случае может содержать информацию, явно не указанную в запросе, но связанную с ней по смыслу.

3.1. Архитектура системы

В состав ядра системы, помимо блока первичной обработки, входят следующие блоки: лингвистический процессор, блок выделения понятий текста, блок формирования семантической сети, блок хранения семантической сети.

Блок первичной обработки извлекает текст из входного потока данных, очищает от нетекстовой информации, а также обрабатывает аббревиатуры, инициалы, заголовки, адреса, номера, даты, указатели времени.

Лингвистический процессор осуществляет предобработку входного текста на основе априорных лингвистических знаний, общих для выбранного языка (в настоящий момент поддерживаются несколько европейских языков, помимо русского и английского), и выполняет следующие функции: сегментацию предложений текста на основе знаков пунктуации и специальных грамматических слов, нормализацию слов и словосочетаний — удаление флексий (окончаний) с сохранением только корневых

основ, а также — удаление в тексте семантически несущественных, вспомогательных слов.

Блок выделения ключевых понятий предметной области (слов и словосочетаний) создан на базе программной модели иерархических структур из ДАЗУ и реализует алгоритмы автоматического формирования частотного словаря текста.

Блок формирования семантической сети реализован как база данных, в которой представляются семантические связи понятий предметной области. Поскольку типы семантических связей в системе не определяются, такие связи представляют собой просто ассоциативные связи. В качестве критерия для определения наличия семантической связи между парой понятий используется частота их совместной встречаемости в предложениях анализируемого текста.

Элементы семантической (ассоциативной) сети и их связи имеют числовые характеристики, отражающие их смысловой вес в данной предметной области. Для более точной оценки семантических весов понятий используются веса всех связанных с ними понятий, то есть веса целого семантического сгущения. В результате такого анализа наибольший вес приобретают понятия, обладающие мощными связями и находящиеся как бы в центре семантических сгущений.

3.2. Функциональность программы TextAnalyst

В системе реализованы следующие функции обработки текстовой информации. Функции: (1) формирования гипертекстовой структуры; (2) навигации по базе знаний; (3) формирования тематического дерева; (4) реферирования текстов; (5) автоматической кластеризации множества текстов; (6) сравнения текстов (автоматической классификации текстов); и, наконец, (7) функция формирования ответа на смысловой запрос пользователя — формирование тематического реферата.

3.3. Интеллектуальность технологии

Из рассмотрения функциональности системы можно заметить, что технология TextAnalyst реализует (см. рис. 1) первый, второй и четвертый квадратики правой части функциональности системы, то есть в ней отсутствует анализ формы текста (третий квадратик). Но семантическая сеть, которая реализуется в результате работы системы, характеризует не допустимую попарную сочетаемость в рамках всего языка (модель языка), а только ту часть, которая касается конкретного проанализированного текста (или группы текстов). То есть сформированная семантическая сеть не ограничивает допустимую сочетаемость, а характеризует конкретный текст — описывает семантику этого текста. Что касается интеллектуальности, можно сделать вывод, что TextAnalyst реализует функцию выявления содержательного компонента текста, и если присмотреться к анализу текстов человеком [8], то этот содержательный компонент текста очень напоминает представление содержания текста в мозге человека [9, 10], которое внес автор текста.

TextAnalyst реализует семантический анализ текста как временной срез содержания, представленного в тексте. Из чего можно сделать вывод, что TextAnalyst также



является интеллектуальной системой, так как опирается на модель языка, которая представлена иначе, чем в ChatGPT.

ЗАКЛЮЧЕНИЕ

В докладе представлены два основных предмета: (1) интеллектуальные системы как частичные модели естественного интеллекта; и (2) система ChatGPT, ее структура и функциональность. Рассмотрение этих предметов показало, что **система ChatGPT реализует функцию синтеза текста по цепочке заданных извне реперных данных**. Причем качество текста как текста языка зависит от объема обучающей выборки, использованной при предобучении системы ChatGPT. А **содержательная часть сгенерированного текста зависит от объема и качества исходных данных, внесенных в систему пользователем в процессе генерации текста**.

Качество синтезированного текста как текста языка получается хорошим, поскольку для обучения системы (для формирования модели языка) использована колоссальная по объему обучающая выборка, элементы которой в системе представляются очень большим числом параметров. Однако исходный содержательный материал вносится в систему извне. **То есть ChatGPT реализует интеллектуальную функцию генерации осмысленного текста с опорой на модель языка**.

Список использованных источников

1. Большая советская энциклопедия (БСЭ). Интеллект (niv.ru).
2. *Пройдаков, Э., Теплицкий, Л.* Англо-русский толковый словарь по искусственному интеллекту и робототехнике. – М.: А. Сандлер, 2022. – 608 с.
3. *Харламов, А.* Ассоциативная память — среда для формирования пространства знаний. От биологии к приложениям. – Dusseldorf: Palmarium Academic Publishing, 2017. – 109 с.
4. *Neuroinformatics and Semantic Representations: Theory and Applications.* Alexander A. Kharlamov & Maria Pilgun eds. – Newcastle upon Tyne: Cambridge Scholars Publishing, 2020. – 317 pp.
5. *Integral Robot Technologies and Speech Behavior.* Alexander A. Kharlamov & Maria Pilgun eds. – Newcastle upon Tyne: Cambridge Scholars Publishing, 2023. – 407 pp.
6. ChatGPT: Optimizing Language Models for Dialogue (англ.). OpenAI (30 ноября 2022). Дата обращения: 13 января 2023 г.
7. TextAnalyst Microsystems, Ltd (analyst.ru)
8. *Косиченко, Е. Ф.* Имя собственное в семиотическом пространстве культуры и художественного текста: монография. – М.: ФГБОУ ВО МГЛУ, 2017. – 294 с.
9. *Леонтьева, Н. Н.* Автоматическое понимание текстов. – М.: Академия, 2006. – 304 с.
10. *Лурия, А. Р.* Язык и сознание. – СПб: Питер, 2019. – 336 с.

IS CHATGPT AN INTELLIGENT SYSTEM?

Kharlamov A. A., Doctor of Technical Sciences, Senior Researcher at the Institute of Higher Nervous Activity and Neurophysiology of the Russian Academy of Sciences, Professor of the Department of Applied and Experimental Linguistics at MGLU, Professor of the Department of Software Engineering at the Higher School of Economics, Professor of the Department of Intelligent Information Systems and Technology at the Moscow Institute of Physics and Technology, Moscow, kharlamov@analyst.ru

The article attempts to formulate an idea of intelligent systems, taking into account the ideas in the structure and functions of the human brain. It is shown that intelligent systems model two main intellectual functions of a person: synthesis of meaningful coherent texts and quasi-texts and analysis of texts and quasi-texts. These are systems based on knowledge representation. The formalization of the concept of “intelligent system” allowed us to evaluate the structure and functionality of GPT 3.5, on the basis of which the ChatGPT system was implemented, which, in turn, allowed us to evaluate the functionality of this latter system. It is shown that ChatGPT implements the text generation function inherent in intelligent systems, but this generation is carried out on the basis of the original content material that is introduced into the system from the outside. That is, if we talk about the intellectuality of ChatGPT in a meaningful way, then we can say that ChatGPT implements the synthesis of texts at a level below pragmatics, if semantics is understood as the permissible compatibility of images of objects and events in the world model.

intelligence • intelligent systems • linguistic processor • extralinguistic processor • ChatGPT • TextAnalyst



XII Международная научно-практическая конференция «Интегрированные модели и мягкие вычисления в искусственном интеллекте», ИММВ-2024 (14–17 мая 2024 г., г. Коломна)

В старинном русском городе Коломне с 14 по 17 мая 2024 г. прошла **XII Международная научно-практическая конференция «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (ИММВ-2024)**, посвященная памяти Виктора Владимировича Емельянова — крупного ученого в области информатики, управления и искусственного интеллекта, ведущего специалиста по интеллектуальному имитационному моделированию и управлению производством, интегрированным интеллектуальным технологиям, многоагентным системам и эволюционному моделированию.

Организаторами конференции ИММВ-2024 явились: *Российская ассоциация искусственного интеллекта, Российская ассоциация нечетких систем и мягких вычислений, Федеральный исследовательский центр «Информатика и управление» РАН, администрация Коломенского городского округа, Коломенский институт (филиал) Московского политехнического университета, Институт компьютерных технологий и информационной безопасности Южного федерального университета.*

Тематика конференции охватывала следующие актуальные научные и практические направления в области искусственного интеллекта:

- нечеткие модели, мягкие вычисления, измерения и оценки;
- машинное обучение, нейросетевые технологии и вероятностные модели в искусственном интеллекте;
- биоинспирированные подходы, эволюционные модели, генетические алгоритмы;
- когнитивные модели в искусственном интеллекте;
- интеллектуальный анализ данных, инженерия знаний и онтологии;
- гибридные интеллектуальные системы;
- интеллектуальные агенты, киберфизические системы, интеллектуальное производство.

В рамках конференции ИММВ-2024 был организован *круглый стол «Эволюция и перспективы развития нейросетей»*, а также проведена *открытая дискуссия «Мифы и реальности искусственного интеллекта»*.

В качестве пленарных докладчиков и модераторов были приглашены ведущие отечественные ученые в области интегрированных моделей, мягких вычислений и измерений, искусственного интеллекта.

В конференции приняли участие более 130 ученых, исследователей и специалистов из академических институтов, университетов, отраслевых институтов и организаций Москвы, Санкт-Петербурга, Белгорода, Брянска, Донецка, Дубны, Железногорска, Калининграда, Красноярска, Новосибирска, Ростова-на-Дону, Самары, Смоленска, Таганрога, Твери, Тулы, Ульяновска, Челябинска, из Мехико (Мексика) и Могилева (Республика Беларусь).

Вашему вниманию предлагаются материалы **круглого стола «Эволюция и перспективы развития нейросетей»** и **открытой дискуссии «Мифы и реальности искусственного интеллекта»**.

Президент Российской ассоциации
искусственного интеллекта
профессор **В.В. Борисов**

Круглый стол «Эволюция и перспективы развития нейросетей»

**(14 мая 2024 г., пресс-центр конькобежного центра МО
«Коломна»)**

Модератор:

Визильтер Ю.В., д.ф.-м.н., проф. РАН (Москва, ГоиНИИАС),

Дискуссанты круглого стола:

Забейхайло М.И., д.ф.-м.н., проф. (Москва, ФИЦ ИУ РАН),
Елистратов В.В., д.т.н., проф. (Москва, УРТИИ МО РФ),
Кобринский Б.А., д.м.н., проф. (Москва, ФИЦ ИУ РАН),
Борисов В.В., д.т.н., проф. (Смоленск, филиал НИУ «МЭИ»),
Еремеев А.П., д.т.н., проф. (Москва НИУ «МЭИ»),
Аверкин А.Н., к.ф.-м.н., доц. (Москва, ФИЦ ИУ РАН),
Симонов Н.А., к.ф.-м.н. (Москва, ФТИАН РАН),
Добрынин Д.А., к.т.н. (Москва, ФИЦ ИУ РАН).

Ю. В. Визильтер:

Коллеги, я предварительно сформулировал ряд вопросов по тематике нашего круглого стола, которые предлагаю объединить в основные группы. Во-первых, насколько мы вообще понимаем то, что сейчас происходит в области искусственных нейронных сетей? То есть насколько у нас есть теоретическая основа для понимания того, как нейросети работают? Есть ощущение, что, несмотря на все замечательные успехи, практика в сфере нейросетей опережает теорию. И в отсутствие теоретического обоснования есть опасение, что мы не знаем, куда движемся и куда дальше придем.

Во-вторых, следует ли нам ожидать дальнейшего бурного развития нейросетей или потенциал нейросетей сейчас близок к насыщению? Я, например, сторонник этой точки зрения. Считаю, что через два-три года мы выйдем на плато развития нейросетевых технологий и ничего нового от нейросетей не получим. Хотя это непопулярная точка зрения, и большинство исследователей считают, что нейросети будут активно развиваться.

В-третьих, несколько лет назад говорили, что вся «история» про нейросети — это «история» про большие данные. Обобщая, можно сформулировать вопрос, насколько справедлива точка зрения, что искусственный интеллект — это «наука о данных», а насколько — что это «наука о методах»? Лежат ли по-прежнему в основе успешных нейросетевых решений сбор и обработка больших данных? Мешает ли сегодня недостаток данных для обучения решению практических задач? Как бороться с этой проблемой?

В-четвертых, вопросы о больших языковых моделях (LLM, Large Language Model). Действительно ли они способны рассуждать или просто воспроизводят то, что в них заложили? Это противопоставление гипотезы Large World Model против гипотезы Large Savant Idiot (Savant Idiot — психическое состояние, когда люди с особым развитием всё запоминают, но при этом неспособны рассуждать, создавать новое и т. д.). Так вот вопрос: что же такое эти генеративные нейросетевые и большие языковые модели — действительно внутри них имеется модель мира, или они просто запомнили всё на свете и спроси у них, они всё расскажут.

В-пятых, насколько в дальнейшем изменится роль и функции специалистов в области искусственного интеллекта? Вообще всех специалистов? В том числе насколько будет востребована профессия программиста, насколько вообще программирование уйдёт в историю (если верить обещаниям разработчиков GitHub Copilot и других систем автоматического программирования). Что будет основным инструментом разработчика интеллектуальных систем в ближайшие годы — машинное обучение или инженерия запросов? Какие нейросетевые технологии наиболее сильно повлияют на нашу жизнь в ближайшие пять лет? 10 лет? 15?

Далее, насколько появление больших языковых моделей повлияет на развитие науки в ближайшие годы? Сегодня Александр Валерьевич Бухановский говорил про использование этих моделей в качестве помощников учёных, которые в значительной степени автоматизируют не только рутинную обработку данных, но и рутинные научные задачи, например планирование и подготовку рутинных отчётов об экспериментах.

Насколько вообще мы предвидим вклад больших языковых моделей в науку? Недавно в Nature были опубликованы работы, где приведено решение математических проблем, которые выполнили большие генеративные модели. Причем эти результаты опубликованы именно как математические, а не как работы по машинному обучению. Имеются

также новые результаты использования больших генеративных моделей и в других областях науки – химии, биологии.

В-шестых, обладают ли генеративные нейросети творческими способностями, или это грубая их имитация. Сейчас, как правило, в этом ключе говорят про большие языковые модели, про которые генерируют тексты. А ведь есть и другие генеративные модели: которые генерируют и изображения, и музыку, и звук. Вроде бы это два разных вида генеративных моделей, но вот насколько они различаются, а что в их природе общего, и как это использовать — тоже интересная тема для обсуждения.

В-седьмых, есть целая группа вопросов о том, как обеспечить доверие и безопасность при практическом применении нейросетевых решений, особенно в автономной робототехнике? Нейросетевые решения являются «непрозрачными», ведут себя как «черный ящик», и в этом смысле мы не можем им полностью доверять. Насколько важна проблема непрозрачности нейросетевых решений? Каковы пути решения этой проблемы? Кроме того, важный вопрос — это их уязвимость к атакам. Нейросети могут неправильно работать, если, например, изменить несколько пикселей на изображении, в отличие от человека, который не станет по-другому интегрально воспринимать изображение (что еще раз является подтверждением того, что механизм функционирования нейросетей отличен от того, как это делают люди). И это характерно, как показали последние исследования, не только для сверточных нейросетей.

М. И. Забейло (реплика):

И зрительную систему человека можно ввести в заблуждение зрительными иллюзиями... Более простой пример: классический средневековый портрет «Знатная дама», ей подрисовали усики — и нейросеть уже ничего не понимает. А человек понял бы сразу, что это фейк.

Ю. В. Визильтер (продолжая):

Вопрос про фейки — это еще один вопрос про безопасность и токсичность нейросетей в социальной сфере. Группа вопросов обеспечения доверия и безопасности при практическом применении нейросетевых решений также охватывает проблемы эмбодимента (embodiment) — физического взаимодействия с роботами.

В-восьмых, говоря о перспективах — насколько важна аппаратная реализация нейросетей? В настоящее время базовая аппаратная реализация нейросетевых решений — это либо графические карты (например, аппаратные решения Nvidia), либо спецпроцессоры. Но все они в основном эффективно реализуют тензорные вычисления, обеспечивая аппаратное ускорение нейросетей. Достаточно давно развиваются нейроморфные решения, более приближенные к их биологическим аналогам. Возникает вопрос: насколько важно их развивать, какова перспектива их внедрения и практического использования? Также развиваются аппаратные нейросетевые решения, основанные на оптических и голографических эффектах и принципах. Эти технологии также считаются очень перспективными.

Если смотреть ещё дальше, то нас ожидают, возможно, квантовые компьютеры, хотя это совсем не нейросети. Однако имеются подходы для реализации нейросетевых моделей на квантовых компьютерах.

Я считаю, что аппаратная реализация нейросетей — это очень важная часть проблемы, и, думаю, что Василий Васильевич Елистратов затронет тему о том, насколько важно иметь стек отечественных доверенных аппаратно-программных решений в сфере нейросетевых технологий.

В-девятых, важный вопрос — про общий (сильный) искусственный интеллект. Несколько лет назад я обычно говорил, что этот вопрос пока не находится на повестке дня. Сейчас так сказать уже нельзя.

При этом, конечно, нужно различать, говорим ли мы о методах и технологиях общего искусственного интеллекта (то есть о создании больших моделей, способных решать уже не отдельные узкофункциональные, а широкие наборы универсальных задач в открытых средах) или о создании искусственной личности со своими чувствами, желаниями и волей. Это, мне кажется, тоже важный вопрос для обсуждения.

Наконец, вопрос, который больше всего волнует в СМИ — возможно ли создание такого нейросетевого искусственного интеллекта, который был бы по-настоящему Super-Human? То есть, не на проценты превышал качество распознавания человеком визуальных образов или лучше писал бы статьи и диссертации (что уже стоит на повестке дня), а такой, уровень решений которого будет настолько превосходить человеческий, что его замыслы, цели и решения будут людям совершенно непонятны? Насколько мы верим, что это возможно?..

Давайте вместе подумаем над этими вопросами.

А сейчас я хочу передать микрофон Василию Васильевичу Елистратову, чтобы он осветил те вопросы, которые его интересуют.

В. В. Елистратов:

Уважаемые коллеги, чтобы наше обсуждение было более конструктивным, хотелось бы определить состав аудитории. Прошу поднять руки разработчикам, заказчикам, эксплуатантам систем с искусственным интеллектом, а также тех, кто готовит кадры. Так как здесь присутствуют представители по всем названным номинациям, уверен, что решения нашего круглого стола будут конструктивными. Для обсуждения других вопросов приглашаю всех в конгрессно-выставочный центр «Патриот» (г. Кубинка), где с 12 августа 2024 г. пройдет цикл научно-деловых мероприятий, в том числе затрагивающих вопросы развития технологий искусственного интеллекта.

Хотелось бы поднять несколько интересующих меня проблем.

Первая проблема — подготовка отраслевого кадрового резерва, так как это очень важный аспект развития технологий искусственного интеллекта, в том числе с учетом динамики их развития.

Другая проблема, будем ли мы нейросетевые технологии использовать в дальнейшем? Как долго они проживут? На данный момент, мне кажется, этого никто не скажет. Тем более что любые интеллектуальные технологии, в том числе нейросетевые, как правило, ориентированы на решение специфических задач.

В этих условиях отраслевые специалисты, ориентированные на решение специализированных задач, сейчас «на вес золота». В условиях дефицита специалистов необходимо привлечение в оборонную сферу специалистов в области передовых нейросетевых технологий из различных отраслей народного хозяйства. Для этого требуется обеспечить им такие условия труда и быта, чтобы при переходе в оборонный сектор они практически не замечали изменений (за исключением соблюдения режимных требований).

Для создания необходимого кадрового резерва на предприятиях-партнёрах и в вузах нами сейчас запущена активная фаза подготовки специалистов двойного (в гражданской и оборонной сфере) назначения. Мы прекрасно понимаем, что специалист выйдет из стен университетов только через четыре-пять лет. Возникает противоречие: с одной стороны, для подготовки классных специалистов нужны высококвалифицированные преподаватели, имеющие необходимые компетенции, которыми, в свою очередь, обладает достаточно узкий круг специалистов, глубоко погруженных в прагматику технологий искусственного интеллекта, но не всегда и не в полной мере обладающих возможностью передачи своих знаний и умений. Поэтому в ряде университетов остро встаёт проблема по обеспечению качества преподавания дисциплин по искусственному интеллекту. И пока этот вопрос до сих пор не решен.

В этот переходный период для решения озвученной кадровой проблемы мы используем базу ведущих университетов, научно-исследовательских центров и институтов, систему дополнительного профессионального образования, то есть готовим необходимых нам специалистов «во внешнем контуре». Мы совместно с нашими партнерами проводим системную работу по подготовке специалистов по номинациям: заказчики, разработчики, приемщики, пользователи систем с искусственным интеллектом.

Например, мы активно используем научно-методические и учебно-методические разработки, возможности программно-инструментальных средств, созданных в ГосНИИАС под руководством Юрия Валентиновича Визильтера, в образовательных целях при подготовке необходимых нам специалистов.

Также исследовательский центр «Сильный искусственный интеллект в промышленности» (Санкт-Петербург, Университет ИТМО), один из руководителей которого, Александр Валерьевич Бухановский, принимает участие в работе нашего круглого стола, готовит для нас группу специалистов по номинации «Квалифицированный заказчик систем с искусственным интеллектом».

Еще один существенный вопрос — сертификация систем (прежде всего, доверенных систем) искусственного интеллекта.

И это только небольшая часть проблем и вопросов, которые я хотел бы (и могу) обсудить на данном круглом столе.

Мы достаточно открыты для сотрудничества, большинство полученных вами результатов в гражданской сфере может быть востребовано для обеспечения и повышения обороноспособности страны, безопасности государства. Для этого на площадке Минобороны России под моим управлением организована экспертиза проектов, и многие из вас уже представляли свои проекты и участвовали в экспертизе проектов. Мы формируем научно-технологический задел, реестр проектов, рекомендуемых к использованию в работах по госзаданиям. Эти апробированные результаты можно проверить на различных реальных задачах.

Поэтому я попрошу Вадима Владимировича Борисова обеспечить «точку входа» для ваших предложений в Управление развития технологий искусственного интеллекта МО РФ.

Ю. В. Визильтер:

Мне хотелось бы передать слово специалистам, которые на примере собственной многолетней работы могут оценить динамику развития искусственного интеллекта как научного направления. И мне хотелось бы попросить осветить во временной перспективе сопоставление и противостояние классических методов искусственного интеллекта и нейросетевых методов. Насколько одно побеждает другое, или возможно их совместное развитие и использование. Сначала хотелось бы передать микрофон Михаилу Ивановичу Забейло, а затем Борису Аркадьевичу Кобринскому.

М. И. Забейло:

В своем завтрашнем выступлении я постараюсь несколько сбалансировать сегодняшнее обсуждение проблематики нейросетей, рассмотрев развитие исследований в области искусственного интеллекта, начиная с Алана Тьюринга и группы Code-Breakers, работавших в Bletchley Park, до нынешних времен, в том числе до третьей (DARPA-й) волны искусственного интеллекта.

Сегодня же я бы хотел вернуться к проблеме подготовки кадров, которую поднял в своем выступлении Василий Васильевич Елистратов. Накопленный в этой сфере опыт показывает, что необходимо обеспечить дополнительное образование либо курсы повышения квалификации в области искусственного интеллекта, в том числе для должностных лиц, принимающих ответственные решения. Это привело бы к совершенно иной динамике внедрения этих методов и технологий в различные отрасли народного хозяйства, в том числе и в оборонную сферу.

Другой аспект — специфика области обороны и безопасности такова, что не про всё можно говорить. И соответственно, результаты на выходе изделий, как правило, закрыты разными требованиями, грифами и т.д. А вопросы не то что доверия и доказуемого описания области, в которой предлагаемое решение надежно работает, а хотя бы верификации этого решения, как были, так и остаются. Что делать? Фактически проблема. Наверное, если бы это был Open Source,

то про ошибки было бы всё ясно. А что делать в противном случае? Для таких случаев имеется простой подход, используемый в ряде задач различного назначения, но опирающийся на одни и те же математические модели и методы, — интерполяционные и экстраполяционные модели (по Юрию Ивановичу Журавлёву). То есть можно отработать ряд задач «в открытой постановке» для последующей их «аранжировки», отладки и верификации этих решений специалистами для «закрытого» применения.

В. В. Елистратов:

Михаил Иванович, я сразу же отвечу на Ваши пожелания. Да, мы в той или иной степени выполняем и поддерживаем ряд таких работ (при участии различных государственных заказчиков и фондов), которые могут быть масштабированы, транслированы в оборонную сферу. К нам обращаются за консультациями представители по линии Минпромторга, Фонда перспективных исследований, Агентства стратегических инициатив, фонда Национальной технологической инициативы и других фондов с просьбой экспертной оценки эффективности и полезности результатов различных проектов, а зачастую и выработки рекомендаций по их повышению.

Ю. В. Визильтер:

Поскольку у нас все-таки дискуссия, хотелось бы озвучить несколько соображений против высказанной выше позиции по поводу возможности и целесообразности конверсии успешных гражданских решений в оборонную сферу. Это очень популярная идея, которая совсем не представляется мне бесспорной.

Где такая конверсия работает? Предположим, у нас есть успешная система обработки документов. Ну конечно, её можно быстро превратить в военную систему обработки документов, так как разницы здесь нет, функциональная задача одна и та же. Но проблема в следующем: ни у кого (на мой взгляд, не надо здесь заблуждаться) в России нет сегодня никаких особых ноу-хау в искусственном интеллекте, которые были бы в мире неизвестны.

Область открытая, всё публикуется и в основном все приложения специфически зависят от данных. Это один из тех вопросов, которые я выше пытался обозначить для дискуссии. Таким образом, если у вас есть хорошие данные, на которых вы можете выучить определённого класса алгоритм (а алгоритм этот всё равно известен), то сила вашей системы именно в тех данных, на которых система обучена, и в подборе правильных алгоритмов именно под эти данные. И если некая компания научила, например, замечательную гражданскую систему видеонаблюдения определять, носят ли работники каски и защитные костюмы, то это совершенно не означает, что теперь эти же разработчики вот эту же систему должны срочно адаптировать под какую-нибудь систему обнаружения целей для оружия. Дело в том, что там совсем другие данные, другого характера, и, возможно, они потребуют вообще других алгоритмов. В таких случаях идея конверсии никак не работает, потому что она не облегчает суммарно решение задачи.

Прошу прощения за длинный комментарий. Теперь хотелось бы передать слово Борису Аркадьевичу Кобринскому, чтобы он нам высказал свою точку зрения.

Б. А. Кобринский:

Юрий Валентинович, так как у нас сейчас завязалась дискуссия, то я буквально несколько слов скажу по ее поводу, а затем продолжу ответы на вопросы, поставленные Вами в начале круглого стола.

То, что касается подготовки кадров, мне приходится работать совершенно с разными категориями студентов и специалистов: и на ВМК МГУ (с магистрантами технического профиля), и в отделении медицинской кибернетики в РНИМУ им. Пирогова (со специалистами в области медицины, знающими высшую математику, но при этом не являющимися «технарями»). Нам действительно надо вычленять необходимое образовательное «ядро» в сфере искусственного интеллекта, а к нему формировать дисциплины дополнительного образования.

При этом нужно понимать, что все охватить нельзя. И конечно же, при формировании дисциплин необходимо учитывать специфику подготовки заказчиков, разработчиков и пользователей систем искусственного интеллекта.

А то, что Юрий Валентинович сказал по поводу конверсии гражданских решений в оборонную сферу, то я с этим и согласен, и не согласен.

От данных, конечно же, многое зависит. Однако в оборонную сферу могут быть перенесены отдельные оригинальные идеи и подходы, а не решения целиком. И у многих из нас такой положительный опыт есть.

Теперь, возвращаясь к одному из вопросов Юрия Валентиновича Визильтера о сопоставлении/противопоставлении классических методов искусственного интеллекта и нейросетей. Например, в 1980-е годы мы делали экспертные системы. Они давали объяснение, позволяли работать как с типичными, так и с атипичными ситуациями, поскольку эти ситуации были в явном виде описаны в базе знаний.

А для нейросетей значимость нетипичных данных теряется. Это так называемая проблема меньшинства заключающаяся в том, что, например, если каких-либо данных, характеризующих 1-й класс больше, чем данных, задающих 2-й класс, то нейросеть выдаёт худшие результаты на данных для 2-го класса.

Этот вопрос связан с доверием к модели и к результатам, а также с их объяснением. Эти вопросы особенно остро стоят при использовании нейросетей для критических приложений.

Согласиться или нет с решением, которое предлагает нейросеть? Если человек согласился, а произошла ошибка, то его накажут. Если же человек не согласился, то ему скажут «тебе дали нейросеть, обученную специалистами, а ты не согласился, и это привело к ошибке»... Причем в зависимости ситуации управления, в которой находится человек (в контуре, над контуром или вне контура управления), нужны разные типы объяснений.

И еще об одном поставленном вопросе: «Обладают ли нейросети способностью к рассуждению или это имитация»? Я бы на этот вопрос ответил так: нейросети имитируют рассуждения. Потому что идёт имитация в той области, информацию о которой нейросеть освоила.

А как только нейросеть выходит за рамки предметной области, появляются псевдорассуждения, галлюцинации. Приведу пример: мы начали новый проект, аналогичные работы по которому мы не смогли найти. Обратились к ChatGPT, который нам подобрал пять наиболее близких по тематике публикаций, одну из которых он приписал Вадиму Владимировичу Борисову. Я к нему обратился за материалами этой статьи. Однако, как оказалось, Вадим Владимирович такую статью не писал. То есть ChatGPT подобрал и названия журналов, и авторов, а также сгенерировал названия статей и их аннотации. Все было подобрано разумно, только это были «чистой воды» галлюцинации!

В. В. Борисов (реплика):

А я затем посмотрел и подумал, что, в принципе, неплохая тематика и, очевидно, стоило бы написать статью на тему, сгенерированную ChatGPT!

Б. А. Кобринский (продолжая):

Но ещё хуже, когда происходит встраивание ложных данных в достоверную информацию. На это сначала можно не обратить внимание. И затем уже мы сами будем генерировать галлюцинации, основанные на частично недостоверной информации. Над решением этой проблемы стоит серьезно задуматься.

В. В. Борисов:

Наша дискуссия вышла за рамки оговоренной темы, но я считаю, это неплохо. Тем не менее я хотел бы вернуть нашу дискуссию к поставленным на круглом столе вопросам, дополнив их следующими, по моему мнению, важными вопросами:

во-первых, какие модели (в контексте существующих и рассмотренных классов нейросетевых моделей) отмирают;

во-вторых, правда ли, что сейчас «выживают» только так называемые чемпионские решения по принципу «победитель забирает все», «угнетая» развитие всех других моделей (которые стоило бы развивать и которые, возможно, могли бы «выстрелить» в средне- или долгосрочной перспективе)?

Ю. В. Визильтер:

По поводу судьбы разных научно-исследовательских направлений, коллеги, давайте попробуем вспомнить опыт последних десятилетий. Я, например, первый раз столкнулся с нейросетевыми технологиями достаточно давно, и, например, будучи еще студентом, участвовал в молодёжной нейросетевой олимпиаде, которую проводил Александр Николаевич Горбань. Это было время очередного всплеска интереса к искусственным нейронным сетям, который был спровоцирован распространением алгоритма их обучения BackPropagation. И казалась, ну вот, сейчас всё решится — мы научились эффективно обучать нейросети, и теперь, наконец-то, начали получать значимые конкретные результаты. Был очень большой энтузиазм. Однако, придя на работу в практический НИИ, поскольку требовалось

решать практические задачи технического зрения, я, как и многие коллеги, в тот момент был вынужден от нейростей отказаться и надолго обратиться к совершенно другим методам, поскольку нейросети в то время не были лучшим практическим инструментом. Другие методы в области технического зрения были лучшими, именно ими мы и пользовались.

К счастью, нейросетевые методы продолжали развиваться, и это на новом витке развития привело к большим успехам и новому взрыву интереса к нейросетевым архитектурам и алгоритмам, начиная с 2011 г.

М. И. Забейло (уточняющий вопрос):

А почему же все-таки состоялся этот взрыв?

Ю. В. Визильтер (продолжая):

Современный искусственный интеллект, как мне представляется, выстроен вокруг бенчмарков (Benchmark). Мировое сообщество исследователей делает так: появилась новая задача, для неё создаётся бенчмарк, то есть некая база, на которой нужно продемонстрировать результаты, и идет соревнование. На методы, которые победили, особенно, победили с большим отрывом, сразу все смотрят и говорят: «О, как интересно!». Именно такие события произошли в 2011 г. Было соревнование по распознаванию визуальных образов на ImageNet, которое представлялось очень сложным, результаты были очень слабыми, и вдруг появляется нейросеть AlexNet, которая с огромным отрывом побеждает все предыдущие решения и достигает уровня человека. С этого момента свёрточные нейросети стали активно развиваться. Точно также никто не ждал появления трансформеров (Transformer), которые вдруг стали показывать совершенно потрясающие результаты в анализе естественного языка (NLP), хотя до этого были огромные накопленные онтологии, словари, методы лингвистического, семантического анализа языка и т. д. И вдруг трансформеры начали всех побеждать в задачах NLP, опять же, на бенчмарках.

Сейчас именно так и развивается наука. Никто, к счастью, не может гарантировать, что завтра не появится что-то совсем новое и не начнёт обгрызать нейросети. Иначе наука закончится через три года. Поэтому мне кажется, что ответ всё-таки такой: для того чтобы привлечь к себе внимание, новый метод должен побить предыдущий на какой-нибудь конкретной задаче, а лучше – на многих разных задачах. Другого способа в нынешней науке я не вижу.

Но на вопрос, стоит ли развивать ли методы, которые сегодня ещё не бьют лидеров, ответ – конечно, да, иначе лидеры никогда не сменятся и мы не увидим новых методов.

М. И. Забейло (уточняющий вопрос):

Юрий Валентинович, и всё-таки вопрос: метод градиентного спуска — это вторая половина 1960-х годов. Метод обучения нейросетей Back Propagation — это 1974 г. (Александр Иванович Галушкин). До середины 1990-х годов если и вспоминали Мак-Каллока и Питтса, то очень

«локально». А вот в начале 2000-х и далее, к 2011 г., — вдруг взрыв. С Вашей точки зрения, почему?

Ю. В. Визильтер (продолжая):

Действительно, первые свёрточные нейросети появились еще в начале 1980-х годов. Почему же они выстрелили только в 2011 г.? Как мы сейчас знаем, если вы делаете свёрточную нейросеть небольшой и учитесь на небольшом объеме данных, то она совсем не обгоняет по качеству другие методы. В 2000-х главными методами были случайные деревья (Random Forest). То есть на всех соревнованиях по машинному обучению, в том числе в задачах компьютерного зрения, побеждали алгоритмы типа Random Forest или метода опорных векторов (SVM Владимира Наумовича Вапника) или бустинга (вспомните знаменитые детекторы AdaBoost). И вдруг в 2011 году выяснилось, что начиная с определённого масштаба данных на миллионах изображений большие свёрточные нейросети (ImageNet и др.) начинают качественно превосходить все существовавшие методы, потому что все предыдущие методы, включая метод опорных векторов, случайные деревья, неспособны обучаться на сверхбольших массивах данных. И там, где эти методы выходят на насыщение, свёрточные нейросети способны продолжать учиться. И на этих сверхбольших объёмах обучающих данных свёрточные нейросети раскрыли свои возможности.

После прорыва 2010-х прошло почти 15 лет. Сложились новые условия, и появились трансформеры, которых, так же как и свёрточные нейросети, никто не ждал...

М. И. Забежайло:

Хочу сделать короткое дополнение. Отмеченный прорыв в сфере нейросетей оказался возможным благодаря прорыву в инфраструктуре: High Performance Computing (HPC), распределённые вычисления и широкополосная связь (смотри отчёт комиссии Э. Шмидта — Б. Урока Конгрессу США).

Ю. В. Визильтер:

Михаил Иванович, я продолжу Вашу мысль. Прорыв в больших языковых моделях был также связан с тем, что человечество взяло и создало своего цифрового двойника в Интернете. Если не было бы всех этих терабайт цифровых данных, текстов, накопленных в Интернете за предыдущее время, никакого бы прорыва не было бы. Даже если бы трансформеры были придуманы в 1960-е годы, так как их просто нечем было бы «накормить».

М. И. Забежайло:

Вы сказали «трансформер» и упомянули Владимира Наумовича Вапника (размерность Вапника — Червоненкиса)... Трансформеры и механизм фокусировки внимания — как вы к этому относитесь с точки зрения размерности Вапника — Червоненкиса? И не идём ли мы к подмене задачи классификации, использования обобщающей способности, к задаче информационного поиска?

Ю. В. Визильтер:

Это очень важный вопрос. Я его перефразирую следующим образом: а что определяет сложность модели? Когда говорят «большая модель», обычно говорят про миллионы, миллиарды настраиваемых параметров, коэффициентов. Но на самом деле в современной нейросети, помимо коэффициентов, есть нелинейности. И еще вопрос, что важнее: иметь больше обучаемых коэффициентов или иметь больше

нелинейностей? И вот почему, известны способы представления нейросети как дерева решений. Просто дерево будет очень-очень большим.

М. И. Забейло (реплика):

... для рекуррентных нейросетей не все так чисто...

Ю. В. Визильтер (продолжая):

Нет, конечно, для фидфорвард-нейросетей (Feedforward, нейросети прямого распространения сигналов) типа многослойного персептрона или свёрточной нейросети с определенными видами нелинейностей можно построить эквивалентные деревья решений. Так вот, количество ветвлений в этих деревьях связано не с количеством коэффициентов, а с количеством нелинейностей. То есть получается, что нужно при оценке сложности и емкости решающего правила нужно учитывать и количество обучаемых коэффициентов, и количество нелинейностей.

При этом чем трансформер отличается от классической нейросети? У трансформера сразу встречается нелинейность, потому что там в модуле внимания берётся произведение двух входных векторов и потом дальше ещё идет обработка. Поэтому нельзя просто взять и сказать: у этой свёрточной нейросети столько-то коэффициентов, и поэтому она имеет такую-то сложность, а трансформер с таким же количеством коэффициентов имеет такую же сложность. Потому что у трансформера гораздо больше нелинейностей в эквивалентном пересчете. То есть сложность трансформера обусловлена нелинейной многоэтапной обработкой.

М. И. Забейло:

Можно ещё раз зацепиться за сложность нейросетей? Известно с 1996 года, что рекуррентные нейросети полны по Тьюрингу. То есть любой алгоритм можно описать средствами рекуррентных нейросетей. Но при этом, как только стали строить реальные рекуррентные нейросети, оказалось, что у них с точки зрения вычислительной эффективности всё совершенно плохо. Возникают доказуемо трудноразрешимые задачи со всеми прелестями на этот счёт.

Итак, по поводу сложности: как вы видите проблему сложности в перспективе развития нейросетей по тем трендам, о которых вы говорили?

Ю. В. Визильтер:

Это действительно интересно, хотя мало кто об этом задумывается. Давайте посмотрим, как работает трансформер. Например, ему подадут на вход 1000 токенов, и он учится формировать 1001-й. Потом он берёт этот 1001-й токен и по ним предсказывает наиболее вероятный 1002-й. Но, по сути, этот процесс напоминает работу машины Тьюринга, машину Поста, конечные автоматы... Здесь в самом деле есть текстовая лента, символьная последовательность.

М. И. Забежайло:

Даже лямбда-конверсии Чёрча, тоже помним...

Ю. В. Визильтер:

Да, совершенно верно. Как ни странно, очень многие современные идеи реализуют базовые идеи кибернетики, которые были сформированы еще на заре ее возникновения и развития. Эти удивительные большие языковые модели, о которых никто не мог и подумать во времена Чёрча и Тьюринга, на самом деле реализуют предложенные ими идеи. То есть, на вход подается лента символов, машина смотрит на все предыдущие символы, генерирует следующий символ, смотрит, генерируют следующий — это математическая машина, которая когда-то была использована в чисто теоретических рассуждениях, для того чтобы оценивать сложность, полноту, замкнутость математических систем.

Или давайте рассмотрим диффузные нейросетевые модели (Diffusion Models). Идея их работы такова: берем некоторое изображение и его зашумляем, потом еще раз зашумляем, потом еще раз зашумляем, и так много раз, до тех пор, пока изображение полностью не превратится в шум.

Теперь возьмем самое левое и следующее изображение, то есть чистое и чуть зашумлённое. Как известно, мы легко можем дать нейросети выучить обратное преобразование для избавления от шума – это называется задачей денойзинга (denoising) или шумоподавления. Дальше мы можем научить нейросеть обратному преобразованию и между третьим и вторым изображением для его восстановления. И так – между всеми следующими смежными изображениями. В итоге мы можем, таким образом, научить нейросеть поэтапно восстанавливать изображение из шума.

Вроде получается ерунда – ну, как такое может быть, ведь в шуме информации нет, а в изображении есть. Но все, кто изучал теорию информации, знают, что на самом деле всё наоборот. Белый шум — это смесь всей возможной информации. Известный эффект: если в театре нужно, чтобы актёры производили шум толпы, надо, чтобы каждый говорил что-то своё, и тогда мы услышим шум.

Еще на заре кибернетики Эшби сказал, что можно сделать «усилитель смысла», которому на вход будет подаваться шум, а из него будут отфильтровываться осмысленные сигналы. Получается, что диффузные нейросетевые модели — это и есть эшбиевские усилители смысла!

Именно так работают все современные Midjourney, которые рисуют нам волшебные картинки по текстовым запросам. То есть опять базовая идея, высказанная на заре кибернетики, сегодня практически реализуется с использованием современных нейросетевых технологий. Мне кажется, это само по себе чудесно, удивительно и интересно!

М. И. Забежайло:

Но только для задач, которые корректны по А. Н.Тихонову...

Ю. В. Визильтер:

Из шума изображение выделить — это совершенно некорректная задача!

М. И. Забейло:

Не буду здесь спорить. Вопрос в машине Тьюринга: я могу построить дедуктивное доказательство корректности заключения. А как быть с доверием к результатам искусственных нейронных сетей, не только трансформеров, с вашей точки зрения?

Ю. В. Визильтер:

С моей точки зрения, всё проверяется экспериментом. То есть если у нас есть тестовые задачи, а сейчас тестовых задач очень много (некоторые бенчмарки – для очень сложных задач), и если на них нейросети показывают результаты, которые близки к 100 %, значит, результат похож на правду.

Позвольте небольшое отступление. Нам кажется, что наша цивилизация всегда требовала логических обоснований. Но многие антропологи считают, что это не так (см., например, *Джозеф Хенрик. «Секрет нашего успеха. Как культура движет эволюцией человека, одомашнивает наш вид и делает нас умнее»*). На самом деле традиционная культура устроена так: ты спрашиваешь эскимоса, почему он строит свою иглу именно из этих частей, именно с такой сложной последовательностью действий, которую европеец выучить не может. Эскимос этого обосновать не может, а просто говорит, что «так принято делать». То есть всю жизнь люди учились так же, как нейросети. Был процесс накопления в культуре неких ноу-хау, которые один человек придумать не может. Это называется кумулятивной культурной эволюцией.

Б. А. Кобринский (ремарка):

Методом проб и ошибок?

Ю. В. Визильтер:

Да, но это происходит именно на протяжении поколений. Почему так долго, сотни тысяч лет, развивались каменные орудия? Потому что это очень сложно происходило, с нуля. Проверяли много раз, если теряется, допустим, европеец в австралийской пустыне, то он не может употреблять в пищу некоторые растения, потому что он не знает, как их надо приготовить: некоторые из них ядовитые, некоторые вызывают расстройство желудка... То есть в большей части в человеческой культуре бытовали «непрозрачные» алгоритмы, которые нельзя объяснить. И мне кажется, что такая же история сейчас воспроизводится в нейронных сетях. То есть то, что нейросети «непрозрачные», — это по современным меркам плохо, а с практической точки зрения — хорошо!

Б. А. Кобринский:

Юрий Валентинович, я хочу вернуться к Вашему ответу на вопрос Михаила Ивановича Забейло. Вы сказали, на тестовой выборке мы получаем некое доказательство. Тестовая выборка, мы все хорошо знаем, это не доказательство. Ведь есть валидация, и есть верификация? Валидация выполняется на близких данных, а верификация осуществляется в других условиях. Для этих задач данные могут отличаться, и тогда можно ли говорить о доверии к модели, к результату только на основе использования тестовых данных.

Ю. В. Визильтер:

Коллеги, у нас дискуссия, кто бы еще хотел взять слово?

А. П. Еремеев:

Мы занимаемся проблематикой искусственного интеллекта уже лет 40. На нашей кафедре прикладной математики НИУ «МЭИ» преподавал Дмитрий Александрович Поспелов, его первые ученики — Вадим Николаевич Вагин и Юрий Иванович Клыков.

Наша проблематика — системы поддержки принятия оперативных решений реального времени. Нами создан ряд решений и в атомной энергетике, и в оборонной сфере. Одними из основных требований для критических приложений являются надежность, устойчивость и объяснимость результатов. И с большим трудом удавалось убедить пользователей таких критических приложений, когда система управлялась на основе классических моделей (на основе дифференциальных и конечно-разностных уравнений), что причиной возникновения внештатных ситуаций являлась либо недостоверная информация, либо неучтенные факторы (наглядный пример — авария на Фукусимской АЭС).

Возвращаясь к нейросетям. Имеются известные теоремы А.Н. Тихонова, статьи академика В.Б. Бетелина, а также Ваши статьи, Юрий Валентинович. С одной стороны, у нейросетей нет объяснимой компоненты, а с другой, чем сложнее нейросеть, тем меньше ее надежность и устойчивость.

Если же говорить о вкладе отечественных ученых в теорию и практику искусственных нейронных сетей, то следует отметить, что недавно был предложен новый тип нейронных сетей KAN (Kolmogorov-Arnold Network) на основе фундаментальной теоремы Колмогорова — Арнольда, в которых, в отличие от многослойных персептронов, обучаются функции активации. Эти сети проще, чем известные сети глубокого обучения, а значит, для них проще решаются проблемы надежности, устойчивости и обучения. Это очень важно для бортовых решений в условиях ограничения вычислительных мощностей. Хотелось бы понять, насколько эффективны нейросети этого типа и для каких задач?

Ю. В. Визильтер:

Относительно недавно Макс Тегмарк (автор замечательной книги «Наша математическая Вселенная. В поисках фундаментальной природы реальности») предложил для графа нейросети (типа многослойного персептрона) нелинейности перенести на ребра. Эта идея звучит совершенно по-новаторски, хотя автор опирается на известные теоремы. Но на самом деле такие нейросетевые модели уже были известны и ранее. Подобные решения многие предлагали лет пять-шесть назад. Например, ReLU является элементом нелинейности, который традиционно используется в нейросетях. Если вы возьмете от одного выхода несколько ReLU с разными сдвигами и дальше поставите сумматор с коэффициентами, то получите такую же управляемую нелинейность. То есть ничего супернового здесь нет. И мы это пробовали, и не только мы. Просто на этом фактотрубаемой нелинейности особо не акцентировали внимание.

А идея понятна, вы можете любую зависимость аппроксимировать небольшим количеством таких элементов. При этом эту нелинейность можно сделать как функцию: синус, косинус, экспонента и т. д., получив любые сложные нелинейности.

Но практика показывает, что такие нейросети очень медленно обучаются, и для сложных примеров их пока выучить не удалось. Поэтому я бы сказал, что перспективы использования таких нейросетей пока достаточно неопределенные.

А по поводу вопроса, какие нейросети мы сможем «запихнуть на борт» (то есть как нейросеть сделать наименее вычислительно затратной), то тут сегодня есть много идей и приемов, а именно: во-первых, можно выполнить «квантование» нейросетей. Сейчас все трансформеры работают уже не на шестнадцать и даже не на восьми, а на четырех битах. Притом что качество их работы более-менее сохраняется; во-вторых, часть коэффициентов нейросетей можно просто обнулить; в-третьих, можно выполнить дистилляцию нейросетей. То есть берем относительно большую нейросеть, например с 70 млрд параметров, и преобразуем ее в относительно меньшую модель, например с 7 млрд параметров; в-четвертых, можно вообще изменить архитектуру нейросетей. Так, самое вычислительно затратное в трансформерах — это модули внимания (квадратичные вычисления, все со всеми сравниваются). Соответственно, модуль внимания можно заменить на что-либо альтернативное, менее вычислительно затратное, например на рекуррентные сети, либо запускать в работу только часть блоков, либо вообще найти альтернативу трансформерам;

в-пятых, можно разбить нейросети на блоки, каждый из которых будут работать по своему назначению, и будет роутер, который запускает эти отдельные блоки (это называется Mixture-of-Experts, то есть смесь экспертов).

И, как мне кажется, именно сочетанием этих подходов мы скорее сейчас получим практический эффект, нежели с использованием той архитектуры, которую предложили Тагмарк с коллегами.

А. Н. Аверкин:

Был задан вопрос по поводу объяснимого искусственного интеллекта, я считаю, что это ключевой вопрос в контексте проблемы «непрозрачности» нейросетей.

По программе DARPA с 2017 г. было потрачено более 10 млрд долларов на разработку методов объяснимого искусственного интеллекта (Explainable Artificial Intelligence) для объяснения сначала сверточных нейросетей, а затем и трансформеров.

Сейчас эти подходы используются не только для «подсветки» наиболее важных элементов данных, пикселей для нейросетей, но и для их лингвистического описания, а также решения задач защиты нейросетей.

Вопрос обеспечения устойчивости нейросетей — тоже очень серьезный вопрос. Фактически все нейросети неустойчивы.

Все нейросети, к которым имеется доступ через Интернет, можно «положить», «отравив» соответствующим образом обучающую выборку. Для этого даже не нужно знать архитектуру этих нейросетей, просто достаточно примерно знать, на каких данных они обучаются.

Методы же объяснимого искусственного интеллекта позволяют выявить наиболее уязвимые места в обучающей выборке и генеративно их «закрасить», существенно снизив эффективность вредоносных атак на нейросети.

Предложен также ряд методов объяснимого искусственного интеллекта для больших языковых моделей, ориентированных на объяснение их поведения, на выявление галлюцинаций.

Подводя итог сказанному, хочу еще раз отметить, что непрозрачность нейросетей — это ключевой момент, преодолев который (для третьего или, может, даже четвертого поколения развития нейросетей), нейросети научатся объяснять сами себе (естественно, на своем языке), как они получили то или иное решение, и вступать в диалог. Если же нейросеть сможет на логическом, символическом языке самой себе объяснить, как она получила то или иное решение, то мы получим гораздо более сильный искусственный интеллект, чем тот, который имеется сейчас.

В. В. Борисов:

Алексей Николаевич, у меня к Вам вопрос: Вы действительно имеете в виду, что нейронные сети должны научиться объяснить самим себе, как они получили то или иное решение?

А. Н. Аверкин:

Да, это так. Это не я придумал. Об этом говорит и Ян Лекун (Yann LeCun), то есть действительно нейросеть сама себе должна объяснить, как она получила то или иное решение. Об этом говорил и Юрий Валентинович Визильтер. И я с этим полностью согласен.

Ю. В. Визильтер:

Абсолютно верно было сказано. По международной классификации Self-Reflection — это один из инструментов общего искусственного интеллекта, примеры которого я приводил в своем пленарном докладе.

М. И. Забежайло:

Жизнь показывает, что «дьявол скрывается в деталях». А не могли бы вы прокомментировать термин «объяснение»? Что именно Вы имеете в виду под этим термином? Что значит «объяснить»?

А. Н. Аверкин:

Ну, во-первых, чтобы система к любому своему решению давала хоть какое-то объяснение, во-вторых, чтобы это объяснение было понятно всем: и разработчикам, и неподготовленным пользователем, то есть чтобы она могла ориентироваться на того человека, которому объясняет. Дальше, объяснение должно быть достаточно точным. Объяснение предполагает наличие разработанной аксиоматики,

ряда метрик точности объяснения и других оценочных параметров, которые антропоморфны, ориентированы на человека.

Сейчас уже разработан с десяток стандартов для объяснимого искусственного интеллекта.

Б. А. Кобринский:

Алексей Николаевич, Вы сказали, что должно быть объяснение, понятное всем. Однако объяснение для разных уровней всегда разное...

А. Н. Аверкин:

Пункт первый — объяснение должно быть всегда; пункт второй — оно должно быть понятно всем. Это не противоречит друг другу, это просто разные позиции одного и того же суждения.

Н. А. Симонов:

Есть еще одна проблема нейросетей — проблема случайных ошибок. Например, для нейросети ImageNET ошибки распознавания составляют 10%, и за последние четыре года этот процент не уменьшается. Я исследовал, почему подобные ошибки возникают, и пришел к выводу, что причина ошибок носит случайный характер.

Ю. В. Визильтер:

Мне кажется, что это очень ценное замечание, которое нас возвращает к теме о том, что нельзя считать, что нейросети «думают» так же, как люди. Нейросети реагируют совершенно на другое.

Вот захотели мы, например, разработать нейросетевую систему распознавания объектов на модельных данных, которые выдают современные симуляторы. Для человека эти модельные данные выглядят похожими на реальные. Но если мы обучим нейросеть на этих модельных данных, то выяснится, что на реальных данных она будет работать неправильно. Начинаем искать причины, и выясняется, и текстура не та, и трава другая, и кирпичи не похожи.

Реальные же данные выглядят немного по-другому. Так вот, нейросети замечательно выучивают не то, чего мы от них хотели (мы же хотели, чтобы они выучили типы объектов, которые нам понятны), а они выучат текстуры, которые создает симулятор.

Нейросети идут по пути наименьшего сопротивления — просто выучивают то, что им дают на примерах. А мы никак не можем им объяснить, чего же мы от них хотим (в отличие от больших языковых моделей, которым мы можем дать соответствующие указания и объяснить, почему они ошибаются).

М. И. Забежайло:

Небольшой комментарий к тому, что вы только что сказали.

Для больших языковых моделей при работе с текстами имеется прямая связь между синтаксисом и семантикой. А для нейросети

в рассмотренном Вами примере для текстур прямой связи между синтаксисом и семантикой нет.

Возможно, что в аудитории есть специалисты, которые сталкивались с реакцией больших языковых моделей на предъявление им кода. Когда им код предлагаешь и спрашиваешь, есть ошибки или нет ошибок, то в ответ приходит некоторый другой код, который даже не компилируется. Проблема в том, что семантика текста программы другая. Языковая модель работает синтаксически, она преобразовывает символы. И для нее является проблемой, как связать баги с текстом через индексы и символы.

Поэтому в задачах, где синтаксис и семантика связаны между собой, есть надежда на положительный эффект, а где они не связаны — над этим нужно работать.

Ю. В. Визильтер:

У трансформеров есть токенизатор, который разбивает текст на токены и предварительно его кодирует. И если токенизатор научен, например, на английский текст, то он совершенно не оптимально работает для программного кода. И те, кто создает нейросети для программирования, используют другие токенизаторы.

Д. А. Добрынин:

Юрий Валентинович, такой простой, но одновременно сложный вопрос: как Вы считаете, какие основные ограничения для нейросетей прекратят их развитие в ближайшем будущем?

Ю. В. Визильтер:

В значительной степени всё зависит от данных, на которых нейросети обучаются, а также от объема этих данных. Если говорить про большие языковые модели, то уже на сегодня их разработчики использовали для обучения почти весь цифровой след человечества в Интернете (по крайней мере, его текстовую часть). Откуда же им добывать новые знания?

Это, может быть, один из самых главных вопросов: могут ли нейросети выйти за пределы того, чему их научили? Мне кажется, скорее всего, нет. С точки зрения качества решений они ограничены тем, что они узнали при обучении. Если обучающая выборка ограничена, то ограничены и возможности обучаемых на ней алгоритмов.

Другое дело – если новые знания приобретаются не пассивно, а активно, в постоянных экспериментах, которые производятся в открытом внешнем мире. То есть интеллектуальный агент в процессе обучения сам пополняет обучающую выборку. Это то, что называется обучением с подкреплением. Особенно интересно, когда такое активное обучение проводится с помощью больших языковых моделей.

Например, DeepMind в последнее время развивает идею «генеративного расширения выборки», которая заключается в следующем. Если мы обучаем нейросеть на лучших примерах действий человека из упомянутого «цифрового следа», то нейросетевая модель в лучшем случае сможет воспроизвести именно лучшее поведение человека. Можно ли достичь большего?

Вот что предлагают коллеги из DeepMind. У нейросетей (трансформеров) есть параметр «температуры». Нейросеть в процессе обучения собирает некое распределение, на основе которого с какой-то вероятностью предсказывает следующий токен. Повышая «температуру» сети, можно размывать это распределение и выбирать все менее вероятные ответы. То есть, повышая температуру, мы будем получать не самый вероятный ответ, а второй по вероятности, третий, десятый и т. д. Когда мы «выкручиваем» температуру на максимум, нейросеть начинает как бы бредить при высокой температуре: повышается ее креативность, но при этом ухудшается качество. Используя это, мы можем усилить генерацию бредовых ответов. Среди этих бредовых ответов большая часть никуда не годится, а оставшаяся часть – хуже «нормальных» выученных ответов. Но случайно среди этого бреда могут оказаться и несколько гениальных ответов. Если у нас есть экспериментальный стенд или контрольный алгоритм, работающий как фильтр, который проверяет качество полученных ответов (он-то на самом деле и является в этой схеме источником новых знаний!), то отобранные им гениальные ответы отделяются от «бреда» и добавляются в обучающую выборку, на которой выполняется дальнейшее дообучение нейросети. Потом эта дообученная нейросеть снова в режиме «бреда» генерирует множество примеров, и они снова фильтруются. И так, итеративно за счёт дообучения путём искусственной генерации примеров и их тестирования, выученное распределение «знаний» нейросети постепенно смещается в сторону того, что нас интересует, и чего никогда не было в исходном множестве примеров. Это, собственно, напоминает историю о том, откуда эскимосы знают, как правильно построить иглу. Такой непростой и небыстрый путь эволюции знаний нейронных сетей.

Поэтому мой прогноз, что на сегодня быстрый прогресс качества работы больших языковых моделей уже остановился или остановится в течение нескольких лет, поскольку просто не останется новой информации, которую они быстро могли бы «сожрать», «переварить» и использовать. А дальше будет относительно медленный экспериментальный процесс добычи новых знаний из окружающего мира. Он все равно может оказаться намного быстрее, чем это можем делать мы, люди, но возможности нейросетей уже никогда не будут расти так быстро, как они росли до сих пор за счет готовой к употреблению цифровой информации, собранной и подготовленной людьми.

Конечно, данный путь эволюции нейросетей потребует гигантских вычислительных ресурсов. Но если к этому добавить вербальное обучение с подкреплением (Reinforcement Learning), методы рефлексии и саморефлексии, которые приближают нас к более объяснимому искусственному интеллекту, направленный поиск решений, то не такое уж это и гигантское количество попыток... И тогда есть шанс, что мы все-таки получим сверхчеловеческий (Superhuman) искусственный интеллект. Однако я не думаю, что он будет развиваться очень быстро. Наиболее вероятный прогноз заключается в том, что

в ближайшие два-три года нейросетевые технологии выйдут на «пологое плато», после чего будут продолжать развиваться и дальше, но уже существенно медленнее.

Коллеги, у нас заканчивается время. Спасибо огромное всем, кто принял участие в дискуссии и в работе круглого стола...



XII Международная научно-практическая конференция «Интегрированные модели и мягкие вычисления в искусственном интеллекте», ИММВ-2024 (14–17 мая 2024 г., г. Коломна)

Открытая дискуссия «Мифы и реальности искусственного интеллекта» (15 мая 2024 г., пресс-центр конькобежного центра МО «Коломна»)

Модераторы:

Забейло М.И., д.ф.-м.н., проф. (Москва, ФИЦ ИУ РАН),
Кобринский Б.А., д.м.н., проф. (Москва, ФИЦ ИУ РАН),
Борисов В.В., д.т.н., проф. (Смоленск, филиал НИУ «МЭИ»)

Дискуссанты:

Соловьев А.А. (Коломна, Коломенский институт Московского политеха),
Рыськин В.О. (Коломна, Коломенский институт Московского политеха),
Чиченина С.Г. (Коломна, Коломенский институт Московского политеха),
Рухов С.А. (Коломна, Коломенский институт Московского политеха),
Киселева С.А. (Коломна, Коломенский институт Московского политеха),
Мымриков Н.А. (Коломна, Коломенский институт Московского политеха),
Бутенко Н.С. (Коломна, Коломенский институт Московского политеха),
Поляков И.В. (Коломна, Коломенский институт Московского политеха),
Коннов В.С. (Коломна, Коломенский институт Московского политеха),
Подвесовский А.Г., к.т.н., доц. (Брянск, БГТУ),
Стаин И.А. (Коломна, Коломенский институт Московского политеха),
Макаркин Н.С. (Коломна, Коломенский институт Московского политеха),
Грачева П.Р. (Коломна, Коломенский институт Московского политеха),
Волкова Ю.Р. (Коломна, Коломенский институт Московского политеха),
Егоркин А.А. (Коломна, Коломенский институт Московского политеха),
Виноградов Д.В. д.ф.-м.н. (Москва, ФИЦ ИУ РАН),
Неопознанный дискуссант № 1, (Коломна, Коломенский институт Московского политеха),
Неопознанный дискуссант № 2, (Коломна, Коломенский институт Московского политеха),
Неопознанный дискуссант № 3 (Коломна, Коломенский институт Московского политеха).

М. И. Забейко:

Предлагается начать с вводных слов модераторов, а затем перейдем к ответам на вопросы и к дискуссии.

Б. А. Кобринский (вступительное слово):

Почему мы дискутируем о мифах и реальностях искусственного интеллекта? Сегодня, как говорится, только ленивый не говорит об искусственном интеллекте. Все имеют какое-либо представление об искусственном интеллекте: правильное, неправильное, мифологическое, футуристическое...

Поэтому обсудим, что такое мифы, а что — реальности искусственного интеллекта.

У людей, не занимающихся профессионально проблематикой искусственного интеллекта, но трактующих его с теоретических, общеполитических позиций, зачастую складывается мифологическое представление, что искусственный интеллект является всемогущим, вездесущим, универсальным, покрывающим все сферы жизнедеятельности человека, позволяет решать все вопросы.

Именно такая точка зрения является основанием для появления понятий «сильный» (*Artificial General Intelligence, AGI*) и «суперсильный» (*Artificial Super Intelligence, ASI*) искусственный интеллект.

Однако представить себе, что даже в отдаленной перспективе появится суперсильный искусственный интеллект, обладающий свойствами, превышающими человеческий интеллект, практически невозможно, несмотря на все сегодняшние достижения.

Для пояснения мифов и реальностей в области искусственного интеллекта обратимся к важным историческим фактам возникновения и эволюции трактовки понятия «искусственный интеллект».

Джон Маккарти, впервые предложивший термин *Artificial Intelligence* в 1955 г., трактовал его, прежде всего, с точки зрения «разумных рассуждений».

Один из основателей движения в сфере искусственного интеллекта в Советском Союзе, академик Гермоген Сергеевич Поспелов отмечал, что прямой перевод термина *Artificial Intelligence* на русский язык — «искусственный интеллект» — является метафорой (хотя, по его мнению, и неудачной), означающей широкое применение компьютеров для решения творческих задач.

Основатель Российской ассоциации искусственного интеллекта Дмитрий Александрович Поспелов, говорил, что искусственный интеллект — это технология решения задач, опирающаяся, прежде всего, на экспертные знания о предметной области.

Джордж Люгер представлял искусственный интеллект в качестве области компьютерной науки, занимающейся автоматизацией разумного поведения.

Отметим, что автоматизация разумного поведения является ключевым свойством, учитываемым в различных трактовках понятия «искусственный интеллект».

Искусственный интеллект как область науки является ответом на вызов при решении трудных задач в плохо формализованных слабоструктурированных предметных областях (социологии, медицине, биологии и других), и в областях с нестандартными, атипичными ситуациями. Особенностью искусственного интеллекта как области науки является то, что его развитие инициируется, как правило, постановкой и необходимостью решения сложных задач, не поддающихся решению другими подходами.

В процессе развития искусственного интеллекта создавались системы на знаниях, экспертные системы, интеллектуальные системы. Появились новые подходы в инженерии знаний, мультиагентные системы с биоинспирированным взаимодействием между интеллектуальными агентами при решении комплексных проблем.

Дмитрий Александрович Поспелов предложил революционную теорию ситуационного управления сложными техническими и организационными системами, основанную на: логико-лингвистических моделях для представления знаний; обучении и обобщении для построения процедур управления по текущим ситуациям; дедуктивном подходе для многошаговых решений. Динамические системы управления, построенные на ситуационном подходе, учитывали нетипичные ситуации, работали с данными в реальном масштабе времени, реализовывали эффективное человеко-машинное взаимодействие.

Подходы и системы, ориентированные на извлечение знаний из больших массивов данных, основаны на дальнейшем развитии технологий искусственного интеллекта, на машинном обучении.

Отметим, что стоит разделять методы так называемого традиционного искусственного интеллекта (в его классическом представлении) и методы машинного обучения (прежде всего, искусственные нейронные сети), для которых ключевую роль играют алгоритмы и технологии настройки, обучения, самоорганизации.

Кстати, когда говорят о нейронах (и искусственных нейронных сетях в целом), обычно подразумевают так называемые формальные нейроны, реализующие достаточно простые вычислительные операции и существенно отличающиеся от нейронов мозга. Чем больше мы узнаем от нейрофизиологов о работе мозга, тем яснее становится, насколько метафористичны и далеки искусственные нейронные сети от их биологических аналогов.

При этом искусственные нейронные сети представляют собой модели типа «черного ящика», которые могут интерпретировать динамику весов обученных сетей, но не позволяют получить объяснение выдаваемых решений.

Имеется также целый ряд проблем, связанных с качеством и спецификой подготовки датасетов для обучения искусственных нейронных сетей.

Существует ряд проблем, являющихся вызовом для современных подходов и методов искусственного интеллекта.

Важная проблема — обеспечение эффективного человеко-машинного взаимодействия для различных ситуаций взаимодействия (человек внутри и вне контура управления), разных уровней и степени осведомленности, объяснимости, интерпретируемости.

Еще одна проблема — оценка и обеспечение доверия к решениям, формируемым на основе методов, моделей и технологий искусственного интеллекта для критических приложений, в которых ответственность за решения крайне высока (например, медицина, оборона, АЭС).

Сейчас часто обсуждают *ChatGPT* и другие большие языковые модели (*Large Language Model, LLM*), которые наряду с полезными функциями генерируют «галлюцинации», создавая правдоподобные, но вымышленные ответы. Это может включать создание искусственных названий и подбор подходящих авторов, если в доступных хранилищах нет нужных данных. Такие «галлюцинации» могут быть сложными, сочетающими достоверные и ложные сведения, что затрудняет их идентификацию. Опасны и многочисленные фейки, поступающие на наши устройства, в которых крайне трудно выявить искаженные данные.

Еще одна проблема, поднимаемая в связи с написанием большими языковыми моделями текстов, подобных человеческому, — проблема «понимания-имитации». Большие языковые модели генерируют (имитируют) тексты на основе тех текстов, которым они обучены. Даже те рассуждения, которые на этом фоне формируются, носят характер имитации, хотя выглядит это как понимание моделью текстов и умение создать новый текст. Но это — именно имитация («под А.С. Пушкина», «под М.Ю. Лермонтова»), никакого понимания здесь нет. Создать что-то принципиально новое (например, новую постэйнштейновскую физику) — это, к сожалению (или к счастью), невозможно.

Проблема эмоций — нужны ли эмоции системам искусственного интеллекта? К чему может привести наличие у них эмоций? С одной стороны, казалось бы, почему бы роботу не обладать эмоциями? Нам это было бы, может быть, приятно. Но представим себе, что робот решил улыбнуться человеку, а человек может отреагировать по-разному. Робот эту реакцию адекватно поймёт? И как он отреагирует на эмоцию человека? Поэтому проблема эмоций очень сложная и многоликая. Мы должны задуматься над тем, чего мы хотим и чего мы ждем от решения данной проблемы.

Таким образом, общий вопрос ко всей нашей дискуссии: всё-таки искусственный интеллект — это метафора, устройство или это область исследований?

М. И. Забейайло (вступительное слово):

Благодаря усилиям организаторов, мы просмотрели ваши вопросы несколько заранее. Пока ещё не комментируя каждый из них в отдельности, хочу обратить ваше внимание на то, что «разброс» поставленных вами вопросов очень широк. Поэтому для придания конструктивного характера нашей дискуссии предлагается определить границы области, которой мы занимаемся, и относительно которой будут сформулированы ответы на ваши вопросы.

Например, заклинания о том, что нас всех ждет безработица в связи с развитием искусственного интеллекта, войны роботов и людей — есть большое количество людей, которые на обсуждении таких тем просто зарабатывают деньги. А к исследованиям и разработкам это имеет минимальное отношение.

Исторически сложилось так, что можно выделить три приоритетных трактовки термина «искусственный интеллект».

Первая трактовка: «искусственный интеллект» как метафора (подобная, например, такой метафоре, как искусственная кожа — натуральная кожа, искусственные материалы — естественные материалы). В этом смысле термин «искусственный интеллект», предложенный Джоном Маккарти в 1955 г., предполагает такую метафору.

При этом требуется уточнение того, что такое «интеллект». И если мы понимаем, что собираемся моделировать, и понимаем, какими средствами мы собираемся моделировать (а мы занимаемся компьютерным моделированием) некоторые функции естественного интеллекта с помощью компьютерных технологий.

Какие это функции? Прежде всего, познавательные.

Поэтому хотя эта метафора (как утверждал Гермоген Сергеевич Поспелов) и неудачна, тем не менее она существует.

Другой подход направлен на трактовку понятия «искусственный интеллект» как автоматического устройства, превосходящего возможности человека. Эта трактовка часто связывается с деятельностью национального центра компетенций в области искусственного интеллекта Сбербанка, где Герман Оскарович Греф регулярно говорит об искусственном интеллекте, например, как о существующем уже здесь и сейчас, и реализованном на их суперкомпьютере «Кристофари».

К сожалению, данная трактовка искусственного интеллекта в виде некоего устройства, так же как и трактовка искусственного интеллекта как метафоры, оспаривается и уязвима.

В этом смысле устройства, например такие, как роботы, нужно, прежде всего, разделять на обычные и интеллектуальные. И нужно отделять возможности сенсорики и мехатроники от их способности к рассуждению, анализу данных, планированию и принятию решений. Поэтому не всякое устройство может быть системой искусственного интеллекта.

Я бы предпочел рассматривать искусственный интеллект как область исследования и разработок, направленных на имитацию и усиление познавательных возможностей естественного интеллекта компьютерными средствами.

В этой трактовке имеется три основополагающих элемента: имитация и усиление; познавательные функции естественного интеллекта;

компьютерное моделирование. При использовании этих рамок многие конструкции, отнесенные Борисом Аркадьевичем Кобринским к мифам искусственного интеллекта, окажутся за пределами нашей дискуссии.

То, что касается имитации и усиления эмоций. Мы не очень понимаем, что такое эмоции и высшая нервная деятельность, поэтому давайте ориентироваться на те функции, которые мы понимаем, а именно: память, рассуждение, оценки, рефлексия. Эти функции можно переложить на компьютер, хотя при этом мы теряем часть глубины в понимании, поскольку компьютер работает с синтаксисом значков, а семантику мы «привязываем» сами.

Развивая эту линию, вы сможете ответить на многие вопросы, например о перспективах искусственных нейронных сетей. Призываю вас к критическому мышлению в отношении к этой предметной области. Задавайте вопросы, анализируйте проблемы и ищите ответы.

Очень часто задают вопрос о том, что же такое система искусственного интеллекта. Если в общем, то это система, основанная на различных методах, моделях, механизмах и устройствах и порождающая новые знания из имеющихся данных и знаний.

Интерполяционно-экстраполяционные модели рассуждений (здесь использованы формулировки, предложенные академиком Юрием Ивановичем Журавлевым) предполагают работу с определенными примерами-прецедентами, которые анализируются и интерполируются (регрессиями, искусственными нейронными сетями и др.), а затем после того, как проверено качество интерполяции, анализируются новые объекты (например, строится прогноз или диагноз) и проверяется экстраполируемость выявленных зависимостей на этих новых объектах. Вот и получается интерполяционно-экстраполяционная модель.

Однако важно также понимать, какие данные и знания мы вкладываем в компьютер, как они должны быть подготовлены, чтобы их можно было использовать в интерполяционно-экстраполяционных моделях.

Проблема представления данных и знаний для применения в интерполяционно-экстраполяционных моделях нетривиальна. Это можно проиллюстрировать на классической проблематике для системы искусственного интеллекта, поставленной в начале Второй мировой войны перед командой *Code-Breakers* в *Bletchley Park* с целью взлома кодов «Энигмы». Как упаковать тексты перехваченных радиogramм немецких подводников в такую форму, чтобы можно было проводить перебор (или перевод?)?

Но это только первая часть проблемы. Другая часть заключается в том, можем ли мы доверять результату тех синтаксических преобразований, которые выполнила система искусственного интеллекта. Например, когда у *ChatGPT* спросили, если большой симфонический оркестр в составе 120 музыкантов исполняет Девятую симфонию Бетховена в течение 80 мин, то сколько времени будет ее исполнять сокращенный состав оркестра в 60 музыкантов? *ChatGPT* ответил, что 40 мин, и привел определенную пропорцию. Но мы-то понимаем, что ответ неправильный.

Оценка доверия к результатам и достаточности оснований для принятия решений — это ещё один очень важный аспект использования системы искусственного интеллекта.

Если есть замкнутая предметная область, то ее можно аксиоматизировать. Например, переписать все теоремы предметной области и задать между ними нужные правила вывода, указав на конкретную теорему. В принципе, можно сжать это множество теорем до небольшого множества аксиом, добавить к ним правила вывода и получить все теоремы. И все истинные утверждения этой предметной области могут быть получены.

Однако, как правило, мы работаем в открытых предметных областях, для которых не можем выполнить исчерпывающую аксиоматизацию. Например, может появиться новый факт (который Нассим Талеб очень удачно назвал «чёрным лебедем»), который в наши предыдущие представления о предметной области никак не вписывается. И как в этом случае использовать техники дедуктивных доказательств — неизвестно. В каждом конкретном случае нужно понимать, что можно, а что нельзя. Поэтому оценка достаточности оснований для принятия результатов очень важна.

Гендиректор ВНИКТИ профессор Валерий Семенович Коссов задал вопрос: можно ли для студентов по-простому объяснить, что связано с проблематикой искусственного интеллекта, как системы искусственного интеллекта могут использоваться для практических задач? Я предлагаю вам такой ответ: у вас есть прецеденты, у вас есть техника интерполяции имеющихся прецедентов, и вы можете получать новые заключения. Но нужно позаботиться о том, чтобы знания были представлены, а оценка доверия к результатам была неоспариваемой.

Еще одна важная проблема. Известно «заклинание»: чем больше данных, тем лучше.

Те, кто сталкивался с проблематикой вычислительной сложности, скажут, что данных много, но как считать будем? Для примера, меланома кожи — онкологическое заболевание, запускаемое мутациями примерно 700 генов. Представьте себе, что у нас есть бинарный вектор длины 700: есть мутации или нет по каждому из генов. Оказывается, что некоторые сочетания мутаций безвредны, а некоторые — вредоносны. Вопрос: можно ли выделить те сочетания генов, которые вредоносны? Перебирать все 2^{700} вариантов невозможно. Даже перебор 2^{100} вариантов на самом мощном компьютере занял бы 30 тыс. лет.

Прямым перебором сделать это нельзя, но специалисты же решают эти задачи. Если мы имеем *Big Data* и нужно за ограниченное время принять решение, то следует посмотреть на то, как это делает специалист. Если удастся понять способ рассуждения, которым он анализирует данные, то можно переложить этот способ на компьютер, который будет делать то же самое быстрее и в больших объемах. Вот это — область

применения искусственного интеллекта в техническом проектировании, анализе материалов и других задачах.

По поводу определений сильного искусственного интеллекта.

Борис Аркадьевич Кобринский уже говорил про определение искусственного интеллекта, данное Джоном Сёрлом (*искусственный интеллект — автоматическое устройство, превосходящее возможности человека во всех областях, которые традиционно характеризуются как интеллектуальная деятельность*).

Даже нобелевский лауреат Роджер Пенроуз, замечательный физик, попытался дать определение сильному искусственному интеллекту, хотя он практически не имеет отношения к *Computer Science*.

Можно ли дать операционально корректное определение сильного искусственного интеллекта, позволяющее отделить «сильный» искусственный интеллект от «не-сильного»? Констатирую, что операционально корректного определения сильного искусственного интеллекта (по крайней мере — на данный момент) не существует!

Поэтому миф о том, что сильный искусственный интеллект нас поработит, я предлагаю оставить журналистам. А исследователям следует заниматься другими, более интересными и актуальными проблемами, которых достаточно много.

Это, прежде всего, озвученная ранее проблема Больших данных, принятия решений за ограниченное время и невозможности привести дедуктивное доказательство для уверенности в результате.

Как поступает человек, принимающий решение? Он ищет объяснение и интерпретацию. Сказать «давайте искать объяснение» — хорошо, но этого мало. Нужно понять, что такое объяснение. Один из вариантов уточнения понятия «объяснение» — это ответ на вопрос «почему». Вот для нынешней культуры современного поколения, «выросшего в гаджетах», вопросы «зачем» и «почему» не всегда актуальны. Тем не менее, если задуматься, как строить ответ на вопрос «почему», простое решение — использование причинно-следственных связей.

А как сделать так, чтобы причинно-следственные связи, извлекаемые из данных, были устойчивыми? Известно, если взять выборку и построить ее регрессионную интерполяцию, затем добавить к выборке ещё один пример, то интерполяция может измениться. Как сделать так, чтобы интерполяции были устойчивыми? Один из ответов — анализировать причинность. Например, если занимаемся медицинской диагностикой и ищем причины возникновения патологий, извлекая из выборки набор факторов, характеризующих причины, то в итоге будем терапией влиять на эти причины. Мы должны сделать это так, чтобы при добавлении к выборке описания новых пациентов той же нозологии причинность сохранялась. Какой же математикой (моделями, методами и алгоритмами) искать причины?

Проблем достаточно много, вам есть чем заниматься и в плане фундаментальных исследований, и в плане разработки эффективных алгоритмов для разрешения моделей, и в плане прикладных систем.

Так что вперед, дорогу осилит идущий!

В. В. Борисов (вступительное слово):

Коллеги, так как мы отчасти в курсе поставленных вами дискуссионных вопросов, то с учетом выступлений Бориса Аркадьевича Кобринского и Михаила Ивановича Забежайло я бы хотел вернуть часть этих вопросов в адрес аудитории.

Предварительно отмечу, что практически все вопросы, которые вы планируете задать модераторам, я бы отнес к мифологической трактовке искусственного интеллекта, в соответствии с которой он противопоставляется естественному интеллекту.

Мы, конечно же, постараемся ответить на все вопросы в предложенной вами постановке. Но, помимо наших ответов, я бы посоветовал вам самим попытаться ответить на эти вопросы.

При этом надеюсь, что эти вопросы, будучи рефлексивно заданными самим себе, позволят дискуссантам в большей степени осознать рассматриваемую проблематику искусственного интеллекта в тех границах, которые определили модераторы сегодняшней дискуссии. Тем более что рассматриваемые проблемы являются достаточно сложными и требуют системных усилий как для их понимания, так и для осознания путей их решения. Причем наших ответов на ваши вопросы в этом случае все равно будет явно недостаточно. Но конечный результат представляется вполне достойным — формирование вашей собственной системы взглядов в сфере искусственного интеллекта.

И еще несколько в большей степени методических вопросов и рекомендаций для помощи вам в понимании проблематики искусственного интеллекта.

Во-первых, задайте себе вопрос: всё ли то, что вы считаете искусственным интеллектом, таковым является? Не относятся ли те методы, модели и технологии, которые вы по каким-либо причинам причисляете к искусственному интеллекту, к другим, более традиционным областям, например к автоматическому управлению, кибернетике, компьютерным наукам, информатике?

Во-вторых, не кажется ли вам, что взрывной эффект от использования методов искусственного интеллекта и машинного обучения в определенной мере обязан синергетическому эффекту, в том числе от применяемых цифровых технологий и созданной инфотелекоммуникационной инфраструктуры? А весь успех такого синергетического объединения этих различных подходов приписывается в основном искусственному интеллекту в угоду неких тенденциозных причин?

В-третьих, ответьте себе на вопрос: что вы имеете в виду под «носителем» искусственного интеллекта? Так, когда задаете вопрос про искусственный интеллект, вы, как правило, спрашиваете, может ли искусственный интеллект сделать то-то или то-то. Для конкретизации вопросов настоятельно рекомендую использовать термин «искусственный

интеллект» в связке с терминами «метод», «модель», «алгоритм», «система», «технология».

Относительно же мифологической трактовки искусственного интеллекта с точки зрения его противопоставления естественному интеллекту (прослеживающейся в большинстве вопросов) я бы хотел отослать дискуссантов к замечательной серии бесед Виктора Константиновича Финна и Михаила Михайловича Забейло, которая называется «Чертова дюжина идеального интеллекта». В этой серии бесед рассматриваются признаки идеального интеллекта. Делается вывод, что далеко не всё то, что может делать человек в процессе своей познавательной деятельности, может быть автоматизировано.

А. А. Соловьев, Коломенский институт Московского политеха (вопрос):

Искусственный интеллект — хороший инструмент для программиста или его замена? Если инструмент, то сможет ли он заменить программиста?

М. И. Забейло (ответ):

Оказалось, что достаточно много участников, которые задавали этот вопрос. Программирование — это искусство, но в каждом искусстве существуют свои методы. Например, внутри программирования давно существуют методы тестирования программ. Те, кто более глубоко интересовался этой предметной областью, знают про технику *Model Checking*. Посмотрите замечательную книгу Кларк Э.М., Грамберг О., Пелед Д. «Верификация моделей программ. *Model Checking*», по прочтении которой можно сделать вывод, что ни в программировании, как, впрочем, и в любом другом искусстве, «железка бездушная» вряд ли заменит человека.

Тем не менее есть целый ряд задач, где не требуется особо изощренный процесс написания кода, где такие порождения кода могут быть автоматизированы. В этом смысле современные системы автоматизации программирования движутся в этом направлении. К сожалению, здесь не всё идет гладко. Те, кто интересуется деталями, зайдите на сайт Института системного программирования РАН, возглавляемого академиком А.И. Аветисяном. Там есть информация о том, что возможно делать сейчас и какие проблемы еще остаются открытыми. Ответ на вторую часть вопроса: сможет ли искусственный интеллект заменить программиста? В общем случае — нет, а в рутинном программировании — да. Примеры и без искусственного интеллекта достаточно хорошо известны. Те, кто знаком или по крайней мере слышал о системе *SAP*, знают, что в компонентно-интеграционной версии *SAP* на шине *NetWeaver* есть механизм, позволяющий в графическом редакторе собрать из фрагментов бизнес-процессов свой собственный процесс, а автоматически из репозитория фрагментов программ соберется исполняемая программа, которая будет его реализовывать. Это программирование без программирования. Но в чём проблема? Вы никогда не будете считать эту программу эффективной и оптимальной по ресурсам и скорости вычисления. Поэтому обычно после того, как программа была сгенерирована, правильные люди садились и переписывали этот *proof of concept*, например, в среде разработки *Oracle*.

В. О. Рыськин, Коломенский институт Московского политеха (вопрос):

В чем опасность принятия мифа о том, что искусственный интеллект может стать самосознательным и обладать человеческими эмоциями?

М. И. Забейло (ответ):

Борис Аркадьевич Кобринский уже говорил, что мы не очень понимаем, что такое эмоции. Поэтому, если удастся дать неоспоримое определение, можно будет продолжить эту дискуссию. А пока полагаю, что нет предмета для обсуждения в смысле угрозы.

С. Г. Чиченина, Коломенский институт Московского политеха (вопрос):

Есть ли у искусственного интеллекта способность к творчеству? Если нет, то может ли он её обрести в будущем?

М. И. Забейло (ответ):

Как и в предыдущем вопросе, моё личное оценочное суждение, мы не очень понимаем, что такое творчество. Способность генерировать некоторые объекты — да, генеративный искусственный интеллект демонстрирует много интересного. Но иногда он демонстрирует то, над чем мы просто смеёмся. Желающие могут открыть Интернет и посмотреть картинку по тексту «лосось, идущий вверх по течению». Те, кто не хочет открывать, скажу, там замечательный ландшафт с изгибом реки, а внизу тушка лосося, как на развале в супермаркете. И утверждается, что это лосось, идущий вверх по течению реки.

Б. А. Кобринский (продолжение ответа):

Проблему творчества здесь обсуждать не будем. Если говорить о создании каких-либо новых произведений, то уже продемонстрировано, что искусственный интеллект может что-то создавать. Недавно музыкантов и композиторов беспокоил вопрос, что их заменит искусственный интеллект. Но это так же как и ответ по поводу программистов: и да и нет. Если нужно подобрать музыку к фильму и режиссёр хочет, чтобы это было похоже на музыку к какому-либо другому фильму, то *ChatGPT* может помочь это сделать. И это будет дешевле, чем приглашать композитора. Но создать принципиально новое произведение, новую оперу — вот это высокое творчество, которое не по силам искусственному интеллекту.

С.А. Рухов, Коломенский институт Московского политеха (вопрос):

Стоит ли в ближайшем будущем вводить законы, нормативные акты, связанные с искусственным интеллектом, регламентирующие его деятельность? Или же нужно будет начинать это делать, когда наступит переломный момент?

М. И. Забейло (ответ):

Данная проблематика очень глубокая, и можно долго обсуждать эту тему.

Я предлагаю воспользоваться следующей аналогией. Сейчас искусственный интеллект имеет множество применений в области обороны и безопасности. В частности, системы вооружений с элементами искусственного интеллекта активно используются по всему миру. В своё время, когда в 1970-е годы две сверхдержавы, СССР и США, поняли, что могут уничтожить друг друга в результате обмена ядерными ударами, то в отсутствие какого бы то ни было законодательного регулирования в этой сфере они начали договариваться друг с другом.

В Женеве проходили консультации наших и американских экспертов. В частности, с нашей стороны участвовали специалисты Института физики Земли АН СССР, чьей задачей было предложить методы, позволяющие идентифицировать факт подземного ядерного взрыва, уточнить координаты, оценить мощность этого взрыва. С американской стороны Тед Кеннеди (третий по старшинству из братьев Кеннеди) возглавил группу американских экспертов. В результате долгих переговоров была выработана система соглашений, позволяющая регулировать эту область. Была построена производственная (основанная на правилах) экспертная система, которая методом триангуляции позволяла точно определить место взрыва, оценить мощность и по дополнительным характеристикам определить материал, используемый для взрыва.

Думаю, что аналогичные шаги будут сделаны и в отношении искусственного интеллекта. По крайней мере, замминистра иностранных дел РФ Сергей Александрович Орджоникидзе совместно с Дмитрием Геннадьевичем Грибковым возглавляют с нашей стороны структуру, которая пытается вести такие переговоры. Но, к сожалению, противная сторона всячески уклоняется от совместных действий в этом направлении. Тем не менее регулировать эту область придется, потому что ответственность за действия систем искусственного интеллекта все равно будут нести люди, а какие люди — это предмет обсуждения.

С. А. Киселева, Коломенский институт Московского политеха (вопрос):

Что, по вашему мнению, должно сделать человечество при наступлении технологической сингулярности? И что произойдет раньше: технологическая сингулярность или момент, когда уровень генной инженерии достигнет такого пика, что созданные человеком существа (например, генетически модернизированные животные) достигнут способности мышления и степени развития, приближенной к человеку, если не превзойдут его?

М. И. Забейло (ответ):

Я бы уклонился от обсуждения этой темы и вот почему. Мы пока что не можем сказать, что такое мышление. Для тех, кто интересуется, посмотрите на сайт академика Константина Владимировича Анохина, который был одним из первых адептов нейросетевой парадигмы не только в нашей стране, но и в мире. Эволюция его взглядов на эту предметную область показывает, что существует много проблем. Мы очень многого не понимаем, как устроена высшая нервная деятельность, интуиция и инсайт. Поэтому, что касается разговоров о технологической сингулярности и успехах генной инженерии, давайте подождём и посмотрим, как всё будет развиваться. А там, может быть, появятся какие-либо новые данные и знания, которые позволят строить неоспариваемые суждения на этот счет.

Н. А. Мыриков, Коломенский институт Московского политеха (вопрос):

В чем заключается основной миф об искусственном интеллекте, который вызывает наибольшее непонимание среди общества?

М. И. Забейло (ответ):

Есть различные мифы об искусственном интеллекте. Вот лишь некоторые из них: «искусственные нейронные сети могут всё»; «сильный искусственный интеллект способен на всё» (см. выше определения Дж. Сёрла и Р. Пенроуза); «чем больше данных, тем лучше».

К сожалению, мифов об искусственном интеллекте много. Если мы занимаемся естественнонаучной или инженерной деятельностью, то давайте двигаться от точных однозначных определений к тем следствиям, которые могут получаться из этих определений корректными способами рассуждения.

Н. С. Бутенко, Коломенский институт Московского политеха (вопрос):

Правда ли, что искусственный интеллект в скором времени будет генерировать контент под запрос?

М. И. Забежайло (ответ):

Уже сейчас генерируется контент под запрос. У меня был случай, когда я, будучи председателем ГЭК в одном уважаемом вузе, вынужден был санкционировать неудовлетворительную оценку при защите магистерской диссертации некоему персонажу, потому что в списке литературы его магистерской диссертации были фейковые источники, сгенерированные *ChatGPT*.

И. В. Поляков, Коломенский институт Московского политеха (вопрос):

Какой прогресс в области искусственного интеллекта может произойти в ближайшие 5, 10, 20 лет?

М. И. Забежайло (ответ):

Когда-то на Физтех, когда я учился, приезжал Владимир Высоцкий. В ответ на одну записку он сказал: «При наличии времени на ваш вопрос я ответил бы брошюрой». Смотрите, что говорят визионеры, а от себя отмечу, доверие, доверенность объяснения, объяснимость, причинность, причинно-следственные связи, математика причинности — вот тот круг вопросов, которые, как я считаю, будет активно развиваться в ближайшие 5, 10, 20 лет.

В качестве обоснования этому для тех, кто захочет добраться до материалов Американского оборонного агентства *DARPA*, посмотрите на их программы *Explainable Artificial Intelligence* и *Artificial Intelligence Exploration*. Вот это более-менее очевидные тренды развития в области искусственного интеллекта.

В. С. Коннов, Коломенский институт Московского политеха (вопрос):

Возможно ли, что искусственный интеллект будет взаимодействовать с людьми ради какой-нибудь своей личной выгоды?

М. И. Забежайло (ответ):

В определении, которое я вам ранее предлагал (*искусственный интеллект — как особая область исследования и разработок, направленных на имитацию и усиление познавательных возможностей естественного интеллекта компьютерными средствами*), понятие выгоды отсутствует. Можно строить какие-либо конструкции оптимизационного характера типа «лучше-хуже», но должен быть критерий оптимизации. Поэтому, с моей точки зрения, ситуаций про личные выгоды системы искусственного интеллекта ожидать не приходится, это вымыслы фантастов.

А. Г. Подвесовский (комментарий):

Совершенно не ясно, как может компьютерная программа преследовать выгоду для себя лично?..

М. И. Забейло (продолжение ответа):

И вообще, где там может быть это «лично»?!

А. Г. Подвесовский (комментарий):

Да, модели решения не принимают!

Б. А. Кобринский (комментарий):

Ведь в масс-медиа этот вопрос возник в основном в связи с появлением темы о человекоподобных роботах, которые якобы будут умнее нас. Тогда и возникает эта мифическая мысль, что если робот станет умнее нас и может заменить людей, тогда он и выгодой может руководствоваться. А так как первое (то, что системы искусственного интеллекта могут быть умнее нас) нереализуемо, то и говорить о выгоде систем искусственного интеллекта не приходится.

А. Г. Подвесовский (комментарий):

Причем этот вопрос, который сейчас был задан, абсолютно не нов. Тема выгоды для систем искусственного интеллекта эксплуатировалась еще в произведениях писателей-фантастов с 40–50-х годов XX в. Есть такой, очень показательный рассказ писателя-фантаста Роберта Шекли «Страж-птица», который очень хорошо все эти проблемы ставит и описывает.

М. И. Забейло (продолжение ответа):

Посмотрите, в Интернете сейчас много высказываний по поводу этого рассказа. Произошла реинкарнация интереса к «Страж-птице».

И. А. Стаин, Коломенский институт Московского политеха (вопрос):

Когда современные нейросети перерастут в настоящий искусственный интеллект?

М. И. Забейло (ответ):

Моё личное оценочное суждение: не очень понятно, что такое «настоящий» искусственный интеллект. Была замечательная история с тем, когда готовился указ Президента России об искусственном интеллекте, и эксперты из одного вуза вписывали термин «сильный искусственный интеллект» в указ Президента. Но им было объяснено, «вы рискуете, потому что, когда указ будет подписан, то Контрольно-ревизионное управление Администрации Президента поставит его на контроль. И если там стоит срок к 31 августа такого-то года предъявить систему сильного искусственного интеллекта, то нужно будет не только ее предъявить как устройство, нужно будет предъявить результаты экспертизы, что это действительно сильный искусственный интеллект». А так как непонятно, чем «сильный» искусственный интеллект отличается от «несильного» (особенно для сотрудников Контрольно-ревизионного управления), то будут неприятности. Представители этого вуза сказали, «Да нет, всё хорошо — у них [на той стороне земного шара] есть сильный, и у нас должен быть сильный искусственный интеллект». А дальше произошло ожидаемое — даме, которая была замминистра экономики РФ и курировала этот проект, пришлось поменять место работы (не потому, что ее уволили, а потому, что она, как человек думающий, поняла, что пора уходить, и перешла в «Роснефть» на спокойную административную должность).

Н. С. Макаркин, Коломенский институт Московского политеха (вопрос):

Правда ли, что код искусственного интеллекта состоит из команд *If* и *Else*?

М. И. Забежайло (ответ):

Нет, не правда! Те, кто хочет в этом убедиться, откройте, например, руководства фреймворков *PyTorch*, *Tensorflow*, и вы в этом убедитесь.

П. Р. Грачева, Коломенский институт Московского политеха (вопрос):

Насколько реально предположение, что искусственный интеллект обладает абсолютной непогрешимостью и всегда принимает оптимальные решения (например, в банковской сфере и в экономике)?

М. И. Забежайло (ответ):

Не всегда. В узких кругах известна история о применении искусственного интеллекта в банковской сфере, когда искусственный интеллект в одном крупном банке из-за ошибки в коде поменял «+» на «-», и это привело к убыткам с девятью нулями. Но в силу того, что прибыль в этой организации очень большая, то эти убытки удалось компенсировать. Поэтому относительно абсолютной непогрешимости искусственного интеллекта — как говорил классик, «не верю»!

Ю. Р. Волкова, Коломенский институт Московского политеха (вопрос):

Если искусственный интеллект научится «думать» и «чувствовать» как человек, то каким образом предполагается решать этические вопросы его применения? Например, если использовать его для ведения боевых действий или заставлять выполнять сложные, монотонные задачи? Насколько вообще этично будет его «эксплуатировать», зная, что он не человек, но тоже испытывает какие-то чувства? Также если учитывать, что человек не может долго выполнять монотонную работу, а война оказывает негативное на психику. То есть разумный искусственный интеллект может просто отказаться выполнять задачу? Как планируется решать такие вопросы?

М. И. Забежайло (ответ):

Отдаю должное интенции автора этого вопроса к гуманистической направленности.

Но смотрите, для тех, кто изучал математическую логику, известно, что импликация (связка) «Если-То» обладает даже в двузначном случае очень странным свойством — если посылка импликации ложна, то вне зависимости от того, какое следствие, вся импликация истинна. Например, знаменитая формула «Если существуют ведьмы, то луна состоит из швейцарского сыра». Это утверждение истинно, потому что посылка ложна.

В этом смысле, посылка «если искусственный интеллект научится «думать» и «чувствовать» как человек» ... А если не научится?.. Давайте подумаем, посмотрим, как всё будет развиваться.

А пока то, что проблемы этики искусственного интеллекта очень существенны, имеет массу подтверждений. В частности, в Администрации

Президента РФ имеется подразделение, которое курирует эту проблематику. И основные вопросы — этические, в том числе в части отстаивания позиции Российской Федерации на международном уровне.

А. А. Егоркин, Коломенский институт Московского политеха (вопрос):

Насколько реально, что искусственный интеллект начнет сам развиваться и, как итог, человечество не будет способно понимать новые технологии, создаваемые искусственным интеллектом?

М. И. Забейло (ответ):

С точки зрения фантастов такая ситуация, наверное, возможна. А с точки зрения тех, кто ежедневно пишет код, разрабатывает алгоритмы или доказывает теоремы относительно систем искусственного интеллекта, боюсь, что это фантастическая ситуация, которая вряд ли реальна, потому что известно, что в компьютере есть кнопка включения и выключения. Если что-то пошло не так, то важно на ранней стадии поймать этот момент и воспользоваться данной кнопкой. И запустить всё заново...

Неопознанный дискуссант № 1, Коломенский институт Московского политеха (вопрос):

В чем основная проблема искусственного интеллекта, она связана с непониманием человеком самого себя или из-за недостатков существующих технических средств?

М. И. Забейло (ответ):

Я думаю, и то и другое. Но техника быстро эволюционирует. Посмотрите, за последние четверть века: *High-performance computing*, широкополосная связь, распределённые вычисления, облачные вычисления, масса других технических «причуд».

Неопознанный дискуссант № 2, Коломенский институт Московского политеха (вопрос):

По вашему мнению, закон Мура будет соблюдаться в ближайшем времени?

М. И. Забейло (ответ):

Хороший вопрос. Не знаю. Возможно, что новые физические принципы для развития как элементной базы, так и алгоритмики вычислений, может быть, они позволят нам уйти от этого ограничения.

Имеются естественные ограничения (так же как и в лёгкой атлетике, мы вряд ли дождёмся, чтобы легкоатлеты прыгали в высоту на 3 м).

Неопознанный дискуссант № 3, Коломенский институт Московского политеха (продолжение вопроса):

А искусственный интеллект может с этим как-то помочь?

М. И. Забейло (продолжение ответа):

Да, он уже помогает, потому что вот то, что я ранее говорил про задачу диагностики меланомы кожи и разделения сочетанных мутаций на безвредные и вредоносные. Никакими стандартными переборными методами эта задача не решается. А вот методами искусственного интеллекта, точнее, интерполяционно-экстраполяционными методами, интеграцией индукции, аналогии и абдуктивного объяснения удалось получить вполне осмысленные и признаваемые медиками результаты.

Д. В. Виноградов (реплика из зала):

Основной миф в области искусственного интеллекта, который американцы внедряют в наши мозги, это то, что искусственный интеллект можно создать методом грубой силы — тренировкой больших языковых моделей, за счет того, что огромные вычислительные кластеры молотят данные методом перебора и упаковывают в нейросети текущее состояние знаний. Эти нейросети работают как сжиматели информации для того, чтобы просто выдавать быстрые ответы.

Неопознанный дискуссант № 4, Коломенский институт Московского политеха (вопрос):

Возвращаясь к теме фантастов, может ли искусственный интеллект быть воплощен с помощью молекул ДНК, а не взаимосвязями в нейросетях?

М. И. Забейло (ответ):

Я не знаю.

CONTENTS

Averkin A. N.

Explainable Artificial Intelligence in Large Language Models.....14

Dzhunkovskiy A. V.

AI as a new type of computer-assisted translation. A road paved with good intentions.....14

Kataev M. Yu.

Methods of speech command recognition in school information systems.....18

Kulai A. A.

Beyond Semantics, or the Semantics of Being and Non-Being..... 35

Meshcheryakov R. V.

Semantics on artificial intelligence technologies for an engineer 43

Raskhodchikov A. Ni.

**Semantics of the communicative environment of the city – information technologies
of urban design..... 49**

Kharlamov A. A.

Is ChatGPT an intelligent system? 58

*XII International Scientific and Practical Conference "Integrated Models and Soft Computing
in Artificial Intelligence", IMMB-2024 (May 14-17, 2024, Kolomna)*

Round table "Evolution and prospects for the development of neural networks" 69

Open discussion "Myths and realities of artificial intelligence" 90

Система анализа поведения людей на соответствие установленным правилам и регламентам SecurOS SBA (Semantic Behavior Analyzer)

Система собирает, обрабатывает и анализирует видеоданные, полученные от стереокамер; выдает оповещения при детекции действий или бездействии, нарушающих регламенты, а также опасных ситуаций; преобразует результаты анализа в интерактивные карточки событий и табличные отчеты, а также отображает статистику инцидентов в виде дашбордов (графиков и диаграмм).

Интерактивные карточки событий дают возможность специалистам служб безопасности и оперативного контроля незамедлительно реагировать на происшествия. Отчеты и дашборды предназначены для управляющих сотрудников и позволяют делать выводы о ситуации на конкретных участках или на предприятии в целом.

Архитектура решения



Ключевые преимущества решения

- SecurOS SBA дает возможность учитывать поведенческие характеристики людей с привязкой к элементам сцены, а также причинно-следственные связи во взаимодействиях между людьми или человеком и объектом (-ами). Таким образом, глубина и точность анализа при использовании технологий SecurOS SBA значительно превышает возможности традиционной видеоаналитики, что существенно расширяет круг применений решения.
- Лингвистический процессор преобразует текст регламентов в работающие правила на встроенной экспертной системе, т.е. без необходимости создания специализированного программного обеспечения под каждый проект.

