

Технологии информационного общества

ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ ТЕХНОЛОГИИ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА В РАМКАХ ЦИФРОВОЙ ТРАНСФОРМАЦИИ ЭКОНОМИКИ РФ И ПОСТРОЕНИЯ ЦИФРОВОГО ОБЩЕСТВА

Статья рекомендована к публикации членом редакционного совета А.В. Богдановым 13.09.2022.

Бирюков Александр Алексеевич

*Центр компетенций НТИ по направлению «Технологии хранения и анализа больших данных» на базе МГУ имени М.В. Ломоносова, руководитель направления информационных технологий и науки о данных
Москва, Российская Федерация
alexander.biryukov@digital.msu.ru*

Терещенко Игорь Александрович

*Центр компетенций НТИ по направлению «Технологии хранения и анализа больших данных» на базе МГУ имени М.В. Ломоносова, ведущий специалист
Москва, Российская Федерация
igor.tereshchenko@digital.msu.ru*

Аннотация

Настоящая статья посвящена анализу перспектив цифровой трансформации национальной экономики Российской Федерации и роли технологии обработки естественного языка в поддержке указанной трансформации. Проведен анализ текущего состояния цифровой экономики РФ, выделены барьеры и положительные эффекты, связанные с осуществлением цифрового перехода. Описано место технологий обработки естественного языка среди новейших цифровых технологий, выделены основные возможности применения технологий обработки естественного языка, проведен анализ ключевых современных практически применимых зарубежных и отечественных моделей обработки естественного языка. Показаны возможности и положительные эффекты практического применения средств и систем обработки естественного языка в разнообразных отраслях экономики, роль применения таких средств и систем в рамках обеспечения цифровой трансформации экономики Российской Федерации.

Ключевые слова

обработка естественного языка; искусственный интеллект; машинное обучение; цифровизация; цифровая трансформация; NLP

Введение

В настоящий момент одним из приоритетных направлений внутренней политики Российской Федерации является осуществление интенсивной цифровизации значительного количества отраслей национальной экономики, перехода к более эффективным и производительным производственным процессам в самых различных областях экономической деятельности, обеспечения беспрепятственного и надежного доступа граждан страны к новейшим технологиям связи и государственным услугам, оказываемым в электронном виде. Необходимость интенсивного развития цифровой экономики подчеркивается тем, что на практике наблюдается некоторое отставание российской экономики от мировых цифровых лидеров в части процессов цифровизации, а также существенные дисбалансы в проникновении современных цифровых технологий в различных отраслях. В связи с этим разработаны и реализуются ряд национальных и отраслевых программ и проектов, направленных на фасилитацию и поддержку внедрения частными предприятиями и государственными ведомствами новых цифровых технологий, а также

© Бирюков А.А., Терещенко И.А., 2023

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2023_02_105

на обеспечение в перспективе лидирующих позиций Российской Федерации в сфере высоких технологий.

Особое значение в процессах цифровизации имеет расширение применения в разнообразных отраслях национальной экономики России комплекса технологий искусственного интеллекта, в рамках которого отдельно выделяются [6] технологии компьютерного зрения, распознавания и синтеза речи, обработки естественного языка (или NLP, от англ. Natural Language Processing), создания рекомендательных систем и интеллектуальных бизнес-систем поддержки принятия решений. Фокус внимания настоящей статьи направлен на изучение перспектив применения технологий обработки естественного языка в рамках развития цифровой экономики России, анализ ожидаемого полезного эффекта от расширения применения указанной технологии, а также существующих практик применения технологии обработки естественного языка в различных отраслях.

1 Предпосылки, преимущества и препятствия цифровой трансформации экономики Российской Федерации

Процессы цифровой трансформации в настоящее время наблюдаются на глобальном уровне, среди всех стран – участников мировой экономики. Как отмечает Басаев З. В., успешная цифровизация увеличивает конкурентоспособность как частных предприятий, так и государственного сектора экономики на региональном и мировом уровне, позволяет существенно повысить темпы экономического роста и роста производительности труда [2]. Важная роль цифровой трансформации отмечается в рамках ряда документов государственного стратегического планирования Российской Федерации; в частности, Программа «Цифровая экономика Российской Федерации» указывает, что необходимым условием дальнейшего развития экономики является использование данных в цифровой форме как фактора производства во всех сферах социально-экономической деятельности в интересах повышения конкурентоспособности страны, качества жизни ее граждан, обеспечения экономического роста и национального суверенитета [12].

Вместе с тем российская цифровая экономика в настоящее время продолжает значительно отставать в своем развитии от стран – лидеров цифровизации. Так, в Глобальном индексе инноваций 2021 года Российская Федерация находится на 45 позиции среди 132 участников (и на 29 позиции в европейском регионе среди 39 стран региона), соседствуя в рейтинге с Вьетнамом и Индией, являющимися странами с более низким уровнем экономического развития в целом [18]. К числу конкурентных преимуществ России, способствующих развитию национальной цифровой экономики, авторы Глобального индекса инноваций относят высокую долю населения с высшим образованием, высокую долю населения, занятую в отраслях с высоким уровнем требований к знаниям персонала, конкурентоспособные позиции страны в аспектах создания знаний, метрикой которых является количество регистрируемых изобретений, полезных моделей, научных исследований. В числе недостатков национальной экономики, затрудняющих интенсификацию процессов цифровизации, выделяются недостаточно высокое качество правового регулирования и защищенности прав собственности, относительно низкий уровень экологической устойчивости, а также низкий уровень соотношения объема инвестиций к ВВП и недостаточное развитие системы венчурного финансирования [18].

Еще одним барьером, несколько замедляющим цифровизацию национальной экономики Российской Федерации в целом, следует считать существенные различия в отраслевой специфике процессов цифровой трансформации [5]. Так, хотя финансовый сектор экономики является одним из лидеров по степени цифровизации в РФ и вполне сопоставим по своему уровню развития в части цифровых технологий с лучшими зарубежными аналогами, проникновение новых цифровых технологий в промышленность и сельское хозяйство остается весьма незначительным. При этом уровень цифровизации отрасли в значительной степени взаимосвязан с типичным для отрасли уровнем затрат предприятий на создание, распространение и использование цифровых технологий и предлагаемых на их основе продуктов и услуг.

Недостаточный уровень цифровизации может быть успешно преодолен в среднесрочной перспективе в случае эффективных совместных усилий государства и частного бизнеса в аспекте интенсификации инноваций и технологических инвестиций. Оптимистичный сценарий развития цифровой экономики Российской Федерации предусматривает опережающие по сравнению с мировыми темпы цифровизации, что позволит войти в число лидирующих цифровых экономик,

однако необходимым условием являются совместные скоординированные усилия государства и бизнеса при осуществлении цифровизации экономики в целом и отдельных ее отраслей [5]. Положительный эффект для российской национальной экономики от процессов цифровизации к 2025 году оценивается в 4.1 – 8.9 трлн рублей, составляя, таким образом, от 19 до 34 процентов общего объема увеличения ВВП за период с 2015 по 2025 год [17]. Большая часть обозначенного положительного эффекта будет достигнута за счет повышения эффективности рынка труда и оптимизации производственных и логистических операций. Цифровизация экономики также позволит повысить качество жизни населения в целом.

Важным фактором, который может существенно способствовать быстрой и эффективной цифровой трансформации национальной экономики, является общая готовность большинства представителей российского бизнеса к цифровым преобразованиям. Цифровизация стала полноправной частью бизнес-стратегии значительного числа компаний, является одним из приоритетов бизнес-сообщества, поскольку при реализации цифровых инициатив наблюдается быстрый и значимый положительный эффект для компании. Большинство представителей бизнес-сообщества осознают необходимость стратегического подхода к цифровой трансформации для поддержки конкурентоспособности предприятия [14]. К числу цифровых технологий, которые российские компании планируют внедрить в приоритетном порядке, относятся технологии искусственного интеллекта, в том числе технологии обработки естественного языка, технологии автоматизации бизнес-процессов, процессная аналитика, промышленный интернет вещей и цифровые двойники. В качестве положительного эффекта от внедрения указанных технологий бизнес-сообщество ожидает повышение производительности труда сотрудников, сокращение издержек, улучшение опыта клиентов компании, а также повышение лояльности сотрудников [14].

Одним из ключевых компонентов современных технологий искусственного интеллекта, имеющим важное значение как с точки зрения развития алгоритмической и математической основы ИИ, так и в контексте практического применения методов ИИ, направленного на повышение эффективности бизнеса и решение общественно полезных задач, является комплекс технологий обработки естественного языка. В следующем разделе настоящей статьи описано актуальное состояние развития данного комплекса технологий, а именно представлено общепринятое определение технологий обработки современного языка, наиболее распространенные области практического применения этих технологий. Описаны наиболее продвинутые и популярные современные зарубежные и российские NLP-модели, проведен сравнительный анализ развития технологий NLP в России и за рубежом.

2 Современное состояние развития технологий обработки естественного языка в Российской Федерации и мире

Искусственный интеллект является одной из ключевых технологий, обеспечивающих процессы развития цифровой экономики как в Российской Федерации, так и в мире в целом. Важной составляющей комплекса технологий искусственного интеллекта, имеющей высокую практическую ценность, являются технологии обработки естественного языка.

В общем виде технологии обработки естественного языка можно определить как алгоритмы, использующие техники и методики искусственного интеллекта и машинного обучения для анализа произвольного текста. При помощи интеллектуальных технологий обработки естественного языка выявляется не только смысл отдельных слов, но и становится возможной идентификация идей и концепций в тексте в целом, автоматическое формирование его краткого пересказа или определяющих текст ключевых слов, выявление тональности текста, извлечение именованных сущностей, машинный перевод. Точность обработки текста при помощи технологии обработки естественного языка обеспечивается применением современных алгоритмов машинного обучения, наиболее актуальными из которых являются так называемые «трансформеры». Современные решения обработки естественного языка могут анализировать и выявлять смысл текста как на основе его семантики, так и синтаксиса. В рамках синтаксического анализа NLP-решение определяет смысл текста на основе грамматических правил языка. При помощи семантического анализа NLP-решение определяет смысл и контекст текста на основе смысла и структуры отдельных предложений [16].

Отмечается, что использование технологий обработки естественного языка получило широкое применение в программных продуктах, осуществляющих детекцию и фильтрацию спама

и фишинговых сообщений [19]. Использование NLP позволило существенно увеличить точность и эффективность фильтрации спама и фишинговой коммуникации. Технологии обработки естественного языка также широко применяются в решениях машинного перевода, позволяя учесть значение и тональность, смысловые нюансы переводимого предложения и текста и сгенерировать переведенный текст, который достаточно точно соответствует исходному фрагменту. NLP также активно применяется в рамках виртуальных голосовых помощников и чат-ботов. Наиболее продвинутые образцы виртуальных голосовых помощников и чат-ботов автоматически обучаются со временем и учитывают контекст запросов и команд пользователя для наиболее точной реакции или ответа. Еще одним важным применением технологий обработки естественного языка является анализ тональности релевантных для организации сообщений в социальных сетях и других источниках в сети Интернет. Анализ тональности с использованием NLP позволяет определить эмоциональную окраску сообщений в социальных сетях, обзорах продукции и т.п. и на основе этого внести необходимые корректировки в маркетинговую или продуктовую политику компании. Практически полезным и широко применяемым в мировой практике применением технологий NLP является автоматическое реферирование текста, позволяющее быстро обработать большие объемы текстовой информации и сделать на их основе аннотацию или резюме для научных баз данных или читателей, которым необходима краткая выжимка текста. Технологии NLP также успешно применяются для решения задач автоматической генерации текста на заданную тему.

Наиболее продвинутыми и точными средствами обработки естественного языка в настоящее время являются модели обработки естественного языка, опирающиеся на методики и алгоритмы машинного и глубокого обучения с применением нейронных сетей. Такие модели позволяют автоматически выделить и классифицировать элементы текста или устной речи, а затем оценить вероятность каждого из возможных смысловых значений для этих элементов. Использование таких моделей позволяет наиболее точно в автоматизированном режиме определять смысл больших объемов неструктурированной информации в письменном и устном виде [19]. В качестве алгоритмической основы моделей обработки естественного языка широкое распространение получили нейронные сети. Нейронные сети позволяют эффективно решать задачи регрессии, классификации текста, выделения аномалий и маркировки последовательностей элементов текста, а также могут применяться для аннотирования текстов и изображений, создания программных решений машинного перевода и диалоговых систем [15].

Для сравнительного анализа современного состояния развития технологий NLP в России и в зарубежных странах в рамках данного исследования был проведен анализ одного из наиболее популярных международных интернет-ресурсов, который является наиболее полным сборником open-source моделей и датасетов в области машинного обучения с фокусом на обработку естественного языка – Hugging Face [30]. Был проведен анализ представленных на данном ресурсе моделей и датасетов в области обработки естественного языка и задач, связанных с распознаванием речи, распознаванием текста и другими областями, которые являются близкими к анализу текстовой информации.

По результатам анализа, представленным в таблице 1, на момент написания статьи на ресурсе Hugging Face насчитывается 187 моделей, подходящих для работы с русским языком, что является девятым по количеству моделей результатом среди 178 языков, имеющих на платформе. Необходимо отметить, что это не значит, что российские компании и лаборатории являются разработчиками всех 187 моделей, многие модели являются зарубежными и подходят сразу для целого ряда языков (что определяется путем тестирования на соответствующих разным языкам датасетах). Вместе с тем среди моделей, подходящих для русского языка, на первом месте по количеству скачиваний со значительным опережением стоит все же отечественная модель ruBERT-base-cased-sentence от проекта DeepPavlov Московского физико-технического института [31]. Стоит отметить, что многие модели также являются недоступными для широкой публики, например GPT-3 от лаборатории OpenAI. Также не учитывается в рамках данного рейтинга модель YaLM 100B [32], которая была опубликована компанией Yandex в июне 2022 года и еще не успела заработать на платформе Hugging Face заслуженную популярность. В связи с этим данные модели будут обсуждаться далее отдельно.

Таблица 1. Количество NLP-моделей на разных языках в базе данных ресурса Hugging Face

№	Язык	Количество моделей	Количество датасетов
1	Английский	4999	801

№	Язык	Количество моделей	Количество датасетов
2	Испанский	421	110
3	Французский	397	108
4	Немецкий	303	110
5	Китайский	241	60
6	Шведский	241	57
7	Финский	214	61
8	Японский	201	62
9	Русский	187	86
10	Арабский	175	82
11	Португальский	165	90
12	Турецкий	134	76
13	Итальянский	125	87
14	Нидерландский	114	81
15	Индонезийский	109	55

Также в рамках анализа было рассмотрено наличие доступных на ресурсе датасетов для машинного обучения на соответствующих языках. Однозначной корреляции между количеством датасетов и моделей, подходящих для каждой языковой группы, выявлено не было. Следует отметить, что в рамках данного исследования нельзя сделать однозначного вывода о наличии причинно-следственных связей между количеством датасетов и количеством моделей, подходящих для конкретного языка. Для решения данной задачи необходимо отдельное исследование с учетом объемов и качества датасетов, а также с учетом происхождения разрабатываемых моделей и языковых групп, для которых они подходят.

Была рассмотрена разбивка моделей для анализа текстов на ряде языков по категориям задач. Категории задач были выбраны по единой классификации, применяемой на ресурсе Hugging Face, что позволяет проводить удобное и однозначное сравнение. Ниже представлена таблица 2, в которой представлены результаты анализа, ранжированные по убыванию по столбцу «Русский язык». Помимо русского языка в таблице также рассмотрен английский язык, который является стандартом в части международного исследования в области обработки естественного языка, а также испанский язык как лидер по количеству моделей в Западной Европе, и китайский язык как лидер в азиатской группе языков.

Таблица 2. Количество моделей NLP моделей на разных языках в зависимости от выполняемых задач

№	Задача	Моделей на русском языке	Моделей на английском языке	Моделей на испанском языке	Моделей на китайском языке
1	Перевод	62	405	218	24
2	Классификация текста	34	357	38	17
3	Трансформация текста	14	211	15	8
4	Генерация текста	13	2682	14	18
5	Выявление именованных сущностей	13	76	6	8
6	Реферирование текста	8	83	10	1
7	Предсказание слов	6	106	30	73

№	Задача	Моделей на русском языке	Моделей на английском языке	Моделей на испанском языке	Моделей на китайском языке
8	Классификация токенов	6	104	38	26
9	Автоматическое распознавание речи	6	132	17	11
10	Классификация без предварительного обучения	4	32	4	1
11	Формирование ответов на вопросы	2	105	22	6
12	Определение схожих предложений	2	25	4	1
13	Диалоговые модели	2	19	0	0
14	Конвертация текста в речь	1	62	1	18
15	Конвертация текста в изображение	1	4	0	0
16	Поиск информации в табличных данных	0	21	0	0
17	Конвертация аудиоинформации	0	16	0	0
18	Классификация аудиоинформации	0	23	0	0

Для простоты обозначения будем называть модели, подходящие для анализа русскоязычных текстов, но не обязательно разработанные в России, «русскими» и аналогично для других языков. «Русские» модели преимущественно сфокусированы на задачах перевода, классификации текстов и текстовой генерации, что является достаточно сложными верхнеуровневыми задачами, в то время как «китайские» модели сфокусированы на задачах предсказания слов и классификации токенов, таких как именованные сущности или части речи, что является относительно низкоуровневыми задачами. В то же время «испанские» модели равномерно покрывают как низкоуровневые, так и верхнеуровневые задачи с основным фокусом на переводе. Наконец, «английские» модели в основном решают задачи текстовой генерации, что является наиболее передовой задачей NLP в настоящий момент.

В результате можно заключить, что на данный момент «русские» модели решают действительно актуальные для предметной области задачи, пропуская фундаментальные, но более теоретические задачи, которые имеют научную значимость для дальнейшего развития собственных новых архитектур. В частности, наиболее актуальные и популярные модели также зачастую являются адаптациями архитектур, разработанных за рубежом (в основном в США).

Одним из наиболее популярных и технически совершенных NLP-решений на настоящий момент является модель GPT-3. Разработчики модели заявляют, что GPT-3 способна решать практически любую задачу, связанную с естественной обработкой английского языка [21]. GPT-3 представляет собой авторегрессионную языковую модель, основанную на трансформер-архитектуре [27], которая представляет собой дальнейшее развитие рекуррентных сетей, ориентированное на применение именно для решения задач обработки естественного языка [26]. GPT-3 используется для генерации связных текстов, в том числе с объемом, аналогичным объему публицистических и научных статей, формирования ответов на произвольные вопросы пользователя, применяется в рамках чат-ботов, а также используется для семантического поиска и автоматического реферирования текстов. Отмечается, что качество текста, сгенерированного при помощи GPT-3, настолько высоко, что в некоторых случаях его сложно отличить от статей, написанных профессиональными авторами [23]. Существенным ограничением для

исследовательских и практических целей применения модели, однако, является то, что полный доступ к исходному коду и всем функциональным возможностям GPT-3 на момент написания статьи эксклюзивно предоставлен корпорации Microsoft, остальные пользователи могут воспользоваться решением лишь при помощи API, имеющего некоторые функциональные и коммерческие ограничения [22].

На основе архитектуры GPT-2, предыдущей версии GPT-3, специалистами Сбербанка России было создано NLP-решение RuGPT-3, позиционируемое авторами как самая большая ИИ-модель для русского языка, которая умеет писать тексты [25]. Модель на практике используется для поддержки работы голосовых ассистентов и чат-ботов в приложениях Сбербанка для понимания запросов пользователей и формирования ответов на них, а также используется в рамках сервиса автоматического завершения программного кода на языках программирования Python, Java и JavaScript, функционирующего по принципу подсказок, предлагаемых пользователю. Отмечается, что в перспективе модель можно также будет использовать для автоматизированного перевода программного кода с одного языка программирования на другой, а также автоматически формировать исходный код программ на выбранном языке программирования на основе функциональных требований, сформулированных естественным языком [13]. Кроме того, универсальность и гибкость RuGPT-3 позволяет также использовать эту модель для решения задач анализа эмоциональной нагрузки текста, классификации текстов на положительные и отрицательные, автоматического редактирования текстов отрицательного характера, автоматического реферирования текстов, создания рекомендательных систем, перевода текста [10].

Еще одним современным отечественным NLP-решением является модель YaLM 100B, разработанная компанией Яндекс. Как GPT-3 и RuGPT-3, YaLM основана на трансформер-архитектуре. Для обучения модели были использованы 1,7 терабайт текстовых источников на английском и русском. Обучение проходило в течение 65 дней на 800 видео ускорителях A100 [32]. Обученная модель обладает 100 миллиардами параметров, что делает ее крупнейшей в мире опубликованной в открытом доступе GPT-подобной моделью. Кроме того, модель обладает способностью решать новые задачи без дополнительного обучения [3]. YaLM на практике используется в поисковой системе Яндекс для автоматической генерации подзаголовков и краткого описания объектов, которые ищет пользователь. Кроме того, YaLM служит основой разработанного компанией Яндекс голосового ассистента «Алиса» и используется в том числе как технологическая основа при поддержании диалога пользователя и голосового ассистента. В качестве основных направлений будущего развития модели ее авторы выделяют увеличение числа параметров модели для дальнейшего повышения уровня качества генерируемого текста, а также расширение использования внешних источников информации моделью.

Таким образом, хотя Российская Федерация в настоящий момент не входит в число стран – лидеров цифровой трансформации, наиболее продвинутые и передовые российские технологические NLP-решения являются конкурентоспособными на мировом уровне и сравнимы по своим функциональным возможностям и качеству решения задач обработки естественного языка с лучшими зарубежными аналогами. Расширение практического применения технологий NLP даже на текущем уровне их развития для решения актуальных социально-экономических задач способно оказаться важным фактором, способствующим цифровой трансформации различных отраслей экономики Российской Федерации, создающим ряд полезных эффектов в данных отраслях и национальной экономике в целом. При этом следует отметить, что часть ожидаемых полезных эффектов в целом соответствует мировой практике применения NLP-технологий, в то время как другую часть возможно охарактеризовать как специфичную именно для российской экономики. Детализированное описание положительных эффектов применения и развития технологий обработки естественного языка для отдельных отраслей и для процесса цифровой трансформации экономики РФ в целом представлено в следующем разделе настоящей статьи.

3 Перспективы технологий NLP в цифровой трансформации Российской Федерации

В соответствии с Дорожной картой развития «сквозной» цифровой технологии «Нейротехнологии и искусственный интеллект» к 2024 году ожидается увеличение доли Российской Федерации на мировом рынке решений, применяющих технологии искусственного интеллекта, до 1.7%. Это станет существенным ростом по сравнению с аналогичным показателем 2018 года, составлявшим около 0.2%. Одним из важных драйверов активного развития технологий искусственного

интеллекта в России при этом будут являться технологии и решения, связанные с обработкой естественного языка: объем российского рынка NLP-решений в 2018 году составлял лишь 400 млн рублей, к 2024 году ожидается его рост до 29.9 млрд рублей [6].

В ряде отраслей национальной экономики следует ожидать проявления позитивных эффектов, возникающих как следствие проводимых процессов цифровизации с применением существующих технологий обработки естественного языка и наблюдаемых в мировой практике [6] [16] [20]. Добывающая и обрабатывающая промышленность смогут осуществить оптимизацию разведки и извлечения запасов полезных ископаемых за счет автоматизации анализа накопленных геофизических данных в форме текстов, оптимизацию производственных процессов за счет снижения производственных ошибок и минимизации влияния человеческого фактора, смогут более эффективно и точно анализировать производственную документацию и осуществлять конкурентный анализ. В строительстве произойдет сокращение сроков и затрат на проектирование и строительство объектов за счет автоматизации анализа данных, улучшение качества реализации проектов за счет автоматизированного обнаружения ошибок строительства и строительной документации. Возрастет эффективность предприятий торговли вследствие минимизации влияния человеческого фактора, расширения возможностей оплаты товаров и услуг голосом, который идентифицируется системами искусственного интеллекта, а также оптимизации маркетинговых решений на основе автоматизированного анализа тональности и содержания отзывов клиентов. Кроме того, следует ожидать развития отрасли туризма вследствие увеличения иностранного туристического потока и количества клиентов за счет появления систем мгновенного перевода речи.

Технологии NLP имеют важное значение в финансовом секторе экономики. NLP-решения способны существенно увеличить эффективность борьбы коммерческих банков и финансовых регуляторов с отмыванием денежных средств и мошенничеством. Технологии NLP уже сейчас активно применяются для создания чат-ботов и голосовых ассистентов, позволяющих пользователям решать значительное число задач при взаимодействии с банками без необходимости привлечения банковских сотрудников. Представляется, что такие чат-боты и голосовые ассистенты станут еще более продвинутыми в ближайшей перспективе, их функциональные возможности продолжат расширяться. Это приведет к увеличению лояльности клиентов банка вследствие скорости и качества решения их задач и запросов, а также снизит рутинную нагрузку на банковских сотрудников, существенно снизит издержки, связанные с документооборотом [7]. В сфере здравоохранения технологии обработки естественного языка способны анализировать большие объемы медицинской документации и осуществлять автоматическую категоризацию и резюмирование медицинских записей, что позволит улучшить качество и увеличить оперативность постановки диагноза, принятия решений о методе и порядке лечения, госпитализации больных и увеличит вероятность благоприятного исхода заболевания. Специализированные NLP-системы также могут оказать поддержку в рамках программ предотвращения суицида при помощи автоматизации анализа содержания и характера записей в социальных сетях, публикуемых пациентами. В фармацевтике NLP-решения позволяют автоматически анализировать большой объем специализированной документации, формируемой при разработке новых препаратов и проведении клинических испытаний, что ускорит и сделает более качественным процесс принятия релевантных решений как самими фармацевтическими компаниями, так и отраслевой регулирующей инстанцией. NLP также находит неожиданные области применения, например, такие как обработка геномных данных, представленных в буквенном выражении. Применение NLP в данной области может значительно ускорить процесс анализа генома [28].

Исследователи также отмечают потенциал применения методов обработки и генерации естественного языка при создании научных, публицистических статей, художественной литературы. Ожидается, что к 2027 году значительная часть всего объема создаваемого текста будет написана авторами совместно с системами обработки естественного языка, а для некоторых видов текста эта доля может превысить 90% [24]. Это будет способствовать повышению уровня оперативности, достоверности и качества новостного информирования населения средствами массовой информации, увеличит число и разнообразие публикуемых публицистических и научных статей, художественной литературы, а также поможет привлечению новых авторов.

Значительные перспективы, которые в настоящий момент остаются практически не реализованными, имеются у применения систем обработки естественного языка, в том числе решений понимания естественного языка и генерации естественного языка, в сфере образования.

В существующих исследованиях отмечается [9], что разработка, внедрение и использование интеллектуальных систем, применяющих NLP-технологии, позволит осуществить персонализацию обучения, применять инклюзивный подход к обучающимся, предоставит возможность оперативно внедрять новые дисциплины в программу обучения в зависимости от требований рынка труда. Применение специализированных NLP-систем в сфере образования способно оказать значительное положительное влияние на эффективность и конкурентоспособность национальной системы образования в целом, способность национальной системы образования осуществлять качественную подготовку квалифицированных специалистов для востребованных и активно развивающихся отраслей цифровой экономики.

NLP-решения также имеют достаточно высокий потенциал для применения в рамках решения задач стратегического развития компаний в самых разнообразных секторах экономики. Использование технологий обработки естественного языка в рамках корпоративных интеллектуальных систем прогнозирования позволяет в автоматическом режиме осуществлять анализ больших объемов структурированных и неструктурированных текстовых данных, извлекать из них важные сведения и определять закономерности, которые остались бы незамеченными ранее, что позволяет сконструировать более точную статистическую прогностическую модель предприятия, которая обеспечивает более высокую точность принимаемых управленческих решений и может служить основой для реорганизации бизнес-процессов и методов управления в рамках предприятия, что способно в дальнейшем обеспечить большую эффективность его работы [4] и позволит перенаправить финансовые и человеческие ресурсы компании с решения рутинных задач на более сложные, творческие задачи тактического и стратегического развития [11].

Применение технологий искусственного интеллекта и обработки естественного языка также может привести к положительным эффектам в аспекте модернизации национальной системы нормативно-правового регулирования и законодательного процесса. Исследователи отмечают [8], что использование технологий искусственного интеллекта и обработки естественного языка позволит существенно повысить эффективность законодательного процесса и национального правового регулирования за счет снижения организационных и процедурных издержек, выявления полного спектра общественных отношений, воздействие на которые фактически окажет законопроект, и оценки степени влияния и рисков такого воздействия. В дополнение к этому технология NLP может использоваться для анализа договорных обязательств, подготовки стандартных юридических документов и автоматизированного формирования разнообразных правовых заключений. Закономерным продолжением применения технологий NLP в праве, когда тексты, написанные изначально на человеческом языке, анализируются с помощью технологий обработки естественного языка, является приведение существующей нормативно-правовой базы к изначальному написанию в машиночитаемом виде, что приведет к машиноисполняемому праву [29].

Отраслевые эксперты выделяют ряд основных актуальных трендов в развитии технологий NLP, которые могут оказать существенное положительное влияние на практики применения этой технологии в краткосрочной и среднесрочной перспективе [1].

Ожидается более широкое распространение low-code решений обработки естественного языка, обладающих более низкими по сравнению с традиционными моделями требованиями к техническим и ИТ-компетенциям специалистов предприятия при их настройке и внедрении, что может способствовать распространению технологии NLP в консервативных с точки зрения цифровизации отраслях.

Появление более совершенных многоязычных моделей NLP позволит усовершенствовать методики и системы машинного перевода, упростит и сделает точнее и качественнее автоматизированную обработку и генерацию текста на языках, отличных от английского, что может стать положительным фактором для ускорения темпов внедрения технологий NLP в российской цифровой экономике.

С технической точки зрения ожидается совершенствование методик трансферного обучения, упрощающих переобучение модели NLP и делающих возможным качественное обучение модели на меньшем объеме данных, расширение использования в рамках одной модели одновременно методик обучения с учителем и обучения без учителя, а также более широкое применение в решениях NLP методик обучения с подкреплением.

Как следствие, на основе усовершенствованной технологической базы станет возможным создание еще более продвинутых чат-ботов и голосовых ассистентов, способных решить даже нестандартные и сложные запросы пользователя, увеличение эффективности и степени автоматизации анализа коммуникаций с пользователем, разработка более совершенных решений мониторинга публикаций в социальных сетях, СМИ и медиа, а также программного обеспечения, автоматически выявляющего недостоверную информацию в новостных заметках и других источниках.

С учетом описанных выше ожидаемых позитивных эффектов в различных отраслях экономики представляется последовательным вывод о существенном положительном влиянии технологий NLP на процесс цифровой трансформации Российской Федерации в целом. Технология обработки естественного языка повысит эффективность, производительность и конкурентоспособность национальной экономики, увеличит производительность труда, позволит предоставить потребителям более качественные и удобные товары и услуги, будет способствовать решению задач цифровизации российской экономики опережающими темпами по сравнению с общемировыми. Кроме того, применение технологий обработки естественного языка может заложить основу для создания и применения в будущем разнообразных систем искусственного интеллекта в тех отраслях и направлениях деятельности, которые в настоящий момент принято относить к консервативным и обладающим низким уровнем готовности к цифровой трансформации.

Заключение

В настоящее время цифровизация экономики является всеобщим мировым трендом. Успехи и неудачи цифровой трансформации отдельных стран уже в среднесрочной перспективе могут определить положение и конкурентоспособность национальных экономик на мировой арене. Российская Федерация на настоящий момент не является членом группы стран – лидеров цифрового перехода, в связи с чем особую важность приобретает задача обеспечения опережающего развития инновационных секторов российской экономики. В случае реализации оптимистического сценария развития российской цифровой экономики становится возможным существенный рост валового внутреннего продукта, производительности труда, будет обеспечена конкурентоспособность национальной экономики на мировом уровне, а российская цифровая продукция и инновационные компании, в первую очередь сектор информационных технологий, могут стать мировыми лидерами в своих рыночных сегментах.

Важным компонентом развития цифровизации является разработка и внедрение широкого спектра информационных систем, использующих инновационные технологии искусственного интеллекта, во всех отраслях экономики. В свою очередь, особое место среди средств искусственного интеллекта занимает технология обработки естественного языка, обеспечивающая обработку, анализ, генерацию компьютером неструктурированных и слабоструктурированных текстов, написанных на естественном языке. Проведенный анализ показал, что существующие на данный момент модели обработки естественного языка, представленные российскими разработчиками, являются конкурентоспособными на мировом уровне по отношению к наиболее продвинутым зарубежным аналогам. Также проведенный обзор перспектив применения технологий NLP показывает, что они имеют значительный потенциал применения в различных областях и отраслях экономики РФ.

Таким образом, важной задачей становится тиражирование данных средств, анализ наилучшего опыта их применения и широкое внедрение в различных отраслях. Системы обработки естественного языка могут успешно применяться и повышать эффективность работы предприятий в широком спектре отраслей, в том числе в промышленности, строительстве, на транспорте, в финансовом секторе экономики, в здравоохранении и фармацевтике, на предприятиях розничной и оптовой торговли, а также имеют высокий потенциал полезного применения в рамках законодательного процесса. Успешное широкое внедрение систем обработки естественного языка в этих областях повысит эффективность их работы, производительность труда, финансовые и производственные показатели предприятий, что, в свою очередь, станет немаловажным фактором в обеспечении развития национальной экономики в целом и обеспечении ее цифровой трансформации.

Благодарности

Исследование выполнено при финансовой поддержке в рамках реализации программы Центров компетенций Национальной технологической инициативы на базе Московского государственного университета имени М.В.Ломоносова (договор о предоставлении средств юридическому лицу, индивидуальному предпринимателю на безвозмездной и безвозвратной основе в форме гранта, источником финансового обеспечения которых полностью или частично является субсидия, предоставленная из федерального бюджета № 70-2021-00252 от 15.12.2021).

Литература

1. 9 Natural Language Processing (NLP) Trends in 2022 // URL: <https://monkeylearn.com/blog/nlp-trends/> (дата обращения: 28.06.2022).
2. Басаев З.В. Цифровизация экономики: Россия в контексте глобальной трансформации // Мир новой экономики. Том 12, № 4 (2018). С. 32 – 38. URL: https://wne.fa.ru/jour/article/view/206?locale=ru_RU (дата обращения: 28.06.2022).
3. Блог компании Яндекс. Как Яндекс применил генеративные нейросети для поиска ответов // URL: <https://habr.com/ru/company/yandex/blog/561924/> (дата обращения: 28.06.2022).
4. Брускин С.Н. Методы и инструменты продвинутой бизнес-аналитики для корпоративных информационно-аналитических систем в эпоху цифровой трансформации // Современные информационные технологии и ИТ-образование. 2016. №3-1. С. 234 – 239. URL: <https://cyberleninka.ru/article/n/metody-i-instrumenty-prodvinutoy-biznes-analitiki-dlya-korporativnyh-informatsionno-analiticheskikh-sistem-v-epohu-tsifrovooy> (дата обращения: 28.06.2022).
5. Доклад НИУ ВШЭ. Цифровая трансформация отраслей: стартовые условия и приоритеты // URL: <https://conf.hse.ru/mirror/pubs/share/463148459.pdf> (дата обращения: 28.06.2022).
6. Дорожная карта развития «сквозной» цифровой технологии «Нейротехнологии и искусственный интеллект» // URL: <https://digital.gov.ru/uploaded/files/07102019ii.pdf> (дата обращения: 13.01.2022).
7. Ештокин С.В. Проблемы внедрения и использования чат-ботов, робоэдвайзинга и других продуктов эпохи Индустрии 4.0 в деятельности российских банков // Проблемы рыночной экономики. № 4 (2020). С. 151 – 164. URL: <http://www.market-economy.ru/archive/2020-04/2020-04-151-164-eshtokin.pdf> (дата обращения: 28.06.2022).
8. Зенин С.С., Кутейников Д.Л., Япрынцева И.М., Ижаев О.А. Технология обработки естественного языка (nlp) в законодательном процессе // Вестник ЮУрГУ. Серия: Право. 2020. №3. С. 76 – 80. URL: <https://cyberleninka.ru/article/n/tehnologiya-obrabotki-estestvennogo-yazyka-nlp-v-zakonodatelnom-protseesse> (дата обращения: 28.06.2022).
9. Калимуллин Н.Р. Перспективы применения искусственного интеллекта в образовании // Актуальные проблемы юридического образования: междунар. науч.-метод. конф. (Минск, 9 дек. 2021 г.). Тез. докл. С. 70 – 72. URL: https://elib.amia.by/bitstream/docs/6233/1/aktyal_obrazovanie2021-70.pdf (дата обращения: 28.06.2022).
10. Краткий экскурс в ruGPT-3. Инструкция и демонстрация // URL: <https://habr.com/ru/post/589663/> (дата обращения: 28.06.2022).
11. Невмывако В.П. Всеобщая цифровизация и искусственный интеллект в системе управления стратегическим развитием России в условиях перехода в Индустрию 4.0 // Экономика и социум: современные модели развития. 2020. Том 10. № 4. С. 331-344. URL: <https://creativeconomy.ru/lib/111513> (дата обращения: 28.06.2022).
12. Программа «Цифровая экономика Российской Федерации», утверждена распоряжением Правительства Российской Федерации от 28.06.2017 № 1632-р // URL: <http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf> (дата обращения: 28.06.2022).
13. Сбер ruGPT-3 (нейросеть) // URL: [https://www.tadviser.ru/index.php/Продукт:Сбер_ruGPT-3_\(нейросеть\)](https://www.tadviser.ru/index.php/Продукт:Сбер_ruGPT-3_(нейросеть)) (дата обращения: 28.06.2022).

14. Совместное исследование PwC и АBBYY. Digital IQ 2020 в России // URL: <https://www.pwc.ru/ru/publications/digital-iq-2020/pwc-abbyu-digital-iq-2020.pdf> (дата обращения: 17.01.2022).
15. Суровягин Д.П. Интеллектуальная обработка текстов с помощью глубоких нейронных сетей: основные проблемы и результаты // Проблемы и вызовы цифрового общества: тенденции развития правового регулирования цифровых трансформаций. Сборник научных трудов по материалам I Международной научно-практической конференции. 2019. С. 87 – 91. URL: <https://elibrary.ru/item.asp?id=42356445> (дата обращения: 28.06.2022).
16. A. Leigh. 10 Examples of Natural Language processing (NLP) and How to Leverage Its Capabilities // URL: <https://global.hitachi-solutions.com/blog/natural-language-processing> (дата обращения 28.06.2022).
17. Исследование Digital McKinsey. Цифровая Россия: новая реальность // URL: https://www.mckinsey.com/ru/~/_media/mckinsey/locations/europe%20and%20middle%20east/russia/our%20insights/digital%20russia/digital-russia-report.pdf (дата обращения: 28.06.2022).
18. Global Innovation Index 2021 // URL: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2021.pdf (дата обращения: 28.06.2022).
19. IBM Cloud Learn Hub Natural Language Processing (NLP) // URL: <https://www.ibm.com/cloud/learn/natural-language-processing> (дата обращения: 28.06.2022).
20. Natural language processing is changing these 5 industries // URL: <https://fastdatascience.com/natural-language-processing-is-changing-these-5-industries/> (дата обращения: 28.06.2022).
21. OpenAI API // URL: <https://openai.com/blog/openai-api/> (дата обращения: 28.06.2022).
22. OpenAI is giving Microsoft exclusive access to its GPT-3 language model // URL: <https://www.technologyreview.com/2020/09/23/1008729/openai-is-giving-microsoft-exclusive-access-to-its-gpt-3-language-model/> (дата обращения: 28.06.2022).
23. OpenAI Releases GPT-3, The Largest Model So Far // URL: <https://analyticsindiamag.com/open-ai-gpt-3-language-model/> (дата обращения: 28.06.2022).
24. R. Dale. Natural language generation: The commercial state of the art in 2020 // Natural Language Engineering. № 26(4). С. 481-487. URL: <https://www.cambridge.org/core/journals/natural-language-engineering/article/natural-language-generation-the-commercial-state-of-the-art-in-2020/BA2417D73AF29F8073FF5B611CDEB97F> (дата обращения: 28.06.2022).
25. RuGPT-3 // URL: <https://developers.sber.ru/portal/tools/rugpt-3> (дата обращения: 28.06.2022).
26. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Re´mi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, Alexander M. Rush. Transformers: State-of-the-Art Natural Language Processing // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020. С. 38 – 45. URL: <https://aclanthology.org/2020.emnlp-demos.6/> (дата обращения: 28.06.2022).
27. Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christo-pher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei. Language Models are Few-Shot Learners // arXiv preprint. arXiv:2005.14165. 2020. URL: <https://arxiv.org/abs/2005.14165> (дата обращения: 28.06.2022).
28. Wahab, A., Tayara, H., Xuan, Z. et al. DNA sequences performs as natural language processing by exploiting deep learning algorithm for the identification of N4-methylcytosine // Sci Rep 11, 212 (2021). URL: <https://www.nature.com/articles/s41598-020-80430-x> (дата обращения: 28.06.2022).

29. Понкин И.В. Концепт машиночитаемого и машиноисполняемого права: актуальность, назначение, место в РегТехе, содержание, онтология и перспективы // International Journal of Open Information Technologies. 2020. №9. URL: <https://cyberleninka.ru/article/n/kontsept-mashinochitaemogo-i-mashinoispolnyaemogo-prava-aktualnost-naznachenie-mesto-v-regtehe-soderzhanie-ontologiya-i-perspektivy> (дата обращения: 28.06.2022).
30. Hugging Face // URL: <https://huggingface.co> (дата обращения: 28.06.2022).
31. DeepPavlov: rubert-base-cased-sentence. Model card // URL: <https://huggingface.co/DeepPavlov/rubert-base-cased-sentence> (дата обращения 28.06.2022).
32. Блог компании Яндекс. Яндекс выложил YaLM 100B – сейчас это крупнейшая GPT-подобная нейросеть в свободном доступе. Вот как удалось её обучить // URL: <https://habr.com/ru/companу/yandex/blog/672396/> (дата обращения 03.07.2022).

PROSPECTS OF APPLICATION OF NATURAL LANGUAGE PROCESSING TECHNOLOGY IN THE DIGITAL TRANSFORMATION OF THE RUSSIAN ECONOMY AND THE BUILDING OF A DIGITAL SOCIETY

Biryukov, Alexander Alexeyevich

*Centre for storage and analysis of big data at Lomonosov Moscow State University, head of information technologies and data science department
Moscow, Russian Federation
alexander.biryukov@digital.msu.ru*

Tereschenko, Igor Alexandrovich

*Centre for storage and analysis of big data at Lomonosov Moscow State University, leading specialist
Moscow, Russian Federation
igor.tereshchenko@digital.msu.ru*

Abstract

This article is devoted to the analysis of the prospects of digital transformation of the national economy of the Russian Federation and to the role of natural language processing technology in supporting this transformation. The analysis of the current state of the digital economy of the Russian Federation is carried out, the barriers and positive effects associated with the implementation of the digital transition are highlighted. The place of natural language processing technologies among the latest digital technologies is described, the main possibilities of using natural language processing technologies are highlighted, the analysis of key modern practically applicable foreign and domestic models of natural language processing is carried out. The possibilities and positive effects of the practical application of natural language processing tools and systems in various sectors of the economy, the role of the use of such tools and systems in ensuring the digital transformation of the economy of the Russian Federation are shown.

Keywords

natural language processing; artificial intelligence; machine learning; digitalization; digital transformation; NLP

References

1. 9 Natural Language Processing (NLP) Trends in 2022 // URL: <https://monkeylearn.com/blog/nlp-trends/> (accessed on 28.06.2022).
2. Basaev Z.V. Cifrovizaciya ekonomiki: Rossiya v kontekste global'noj transformacii // Mir novoj ekonomiki. Tom 12, № 4 (2018). S. 32 – 38. URL: https://wne.fa.ru/jour/article/view/206?locale=ru_RU (accessed on 28.06.2022).
3. Blog kompanii YAndeks. Kak YAndeks primenil generativnye nejroseti dlya poiska otvetov // URL: <https://habr.com/ru/company/yandex/blog/561924/> (accessed on 28.06.2022).
4. Bruskin S.N. Metody i instrumenty prodvnutoj biznes-analitiki dlya korporativnyh informacionno-analiticheskikh sistem v epohu cifrovoj transformacii // Sovremennye informacionnye tekhnologii i IT-obrazovanie. 2016. №3-1. S. 234 – 239. URL: <https://cyberleninka.ru/article/n/metody-i-instrumenty-prodvnutoy-biznes-analitiki-dlya-korporativnyh-informacionno-analiticheskikh-sistem-v-epohu-tsifrovoy> (accessed on 28.06.2022).
5. Doklad NIU VSHE. Cifrovaya transformaciya otraslej: startovye usloviya i priority // URL: <https://conf.hse.ru/mirror/pubs/share/463148459.pdf> (accessed on 28.06.2022).
6. Dorozhnaya karta razvitiya «skvoznoj» cifrovoj tekhnologii «Nejrotekhnologii i iskusstvennyj intellekt» // URL: <https://digital.gov.ru/uploaded/files/07102019ii.pdf> (accessed on 13.01.2022).
7. Eshtokin S.V. Problemy vnedreniya i ispol'zovaniya chat-botov, roboedvajzina i drugih produktov epohi Industrii 4.0 v deyatel'nosti rossijskikh bankov // Problemy rynochnoj ekonomiki. № 4 (2020). S. 151 – 164. URL: <http://www.market-economy.ru/archive/2020-04/2020-04-151-164-eshtokin.pdf> (accessed on 28.06.2022).
8. Zenin S.S., Kutejnikov D.L., YApryncev I.M., Izhaev O.A. Tekhnologiya obrabotki estestvennogo yazyka (nlp) v zakonodatel'nom processe // Vestnik YUUrGU. Seriya: Pravo. 2020. №3. S. 76 –

80. URL: <https://cyberleninka.ru/article/n/tehnologiya-obrabotki-estestvennogo-yazyka-nlp-v-zakonodatelnom-protssesse> (accessed on 28.06.2022).
9. Kalimullin N.R. Perspektivy primeneniya iskusstvennogo intellekta v obrazovanii // Aktual'nye problemy yuridicheskogo obrazovaniya: mezhdunar. nauch.-metod. konf. (Minsk, 9 dek. 2021 g.). Tez. dokl. S. 70–72. URL: https://elib.amia.by/bitstream/docs/6233/1/aktyal_obrazovanie2021-70.pdf (accessed on 28.06.2022).
 10. Kratkij ekskurs v ruGPT-3. Instrukciya i demonstraciya // URL: <https://habr.com/ru/post/589663/> (accessed on 28.06.2022).
 11. Nevmyvako V.P. Vseobshchaya cifrovizaciya i iskusstvennyj intellekt v sisteme upravleniya strategicheskim razvitiem Rossii v usloviyah perekhoda v Industriyu 4.0 // Ekonomika i socium: sovremennye modeli razvitiya. 2020. Tom 10. № 4. S. 331-344. URL: <https://creativeconomy.ru/lib/111513> (accessed on 28.06.2022).
 12. Programma «Cifrovaya ekonomika Rossijskoj Federacii», utverzhdena rasporyazheniem Pravitel'stva Rossijskoj Federacii ot 28.06.2017 № 1632-r // URL: <http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf> (accessed on 28.06.2022).
 13. Sber ruGPT-3 (nejroset') // URL: [https://www.tadviser.ru/index.php/Produkt:Sber_ruGPT-3_\(nejroset'\)](https://www.tadviser.ru/index.php/Produkt:Sber_ruGPT-3_(nejroset')) (accessed on 28.06.2022).
 14. Sovmestnoe issledovanie PwC i ABBYY. Digital IQ 2020 v Rossii // URL: <https://www.pwc.ru/ru/publications/digital-iq-2020/pwc-abbyy-digital-iq-2020.pdf> (accessed on 17.01.2022).
 15. Surovyagin D.P. Intellektual'naya obrabotka tekstov s pomoshch'yu glubokih neyronnyh setej: osnovnye problemy i rezul'taty // Problemy i vyzovy cifrovogo obshchestva: tendencii razvitiya pravovogo regulirovaniya cifrovyyh transformacij. Sbornik nauchnyh trudov po materialam I Mezhdunarodnoj nauchno-prakticheskoy konferencii. 2019. S. 87 – 91. URL: <https://elibrary.ru/item.asp?id=42356445> (accessed on 28.06.2022).
 16. A. Leigh. 10 Examples of Natural Language processing (NLP) and How to Leverage Its Capabilities // URL: <https://global.hitachi-solutions.com/blog/natural-language-processing> (data obrashcheniya 28.06.2022).
 17. Issledovanie Digital McKinsey. Cifrovaya Rossiya: novaya real'nost' // URL: https://www.mckinsey.com/ru/~/_media/mckinsey/locations/europe%20and%20middle%20east/russia/our%20insights/digital%20russia/digital-russia-report.pdf (accessed on 28.06.2022).
 18. Global Innovation Index 2021 // URL: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2021.pdf (accessed on 28.06.2022).
 19. IBM Cloud Learn Hub Natural Language Processing (NLP) // URL: <https://www.ibm.com/cloud/learn/natural-language-processing> (accessed on 28.06.2022).
 20. Natural language processing is changing these 5 industries // URL: <https://fastdatascience.com/natural-language-processing-is-changing-these-5-industries/> (data obrashcheniya: 28.06.2022).
 21. OpenAI API // URL: <https://openai.com/blog/openai-api/> (accessed on 28.06.2022).
 22. OpenAI is giving Microsoft exclusive access to its GPT-3 language model // URL: <https://www.technologyreview.com/2020/09/23/1008729/openai-is-giving-microsoft-exclusive-access-to-its-gpt-3-language-model/> (accessed on 28.06.2022).
 23. OpenAI Releases GPT-3, The Largest Model So Far // URL: <https://analyticsindiamag.com/open-ai-gpt-3-language-model/> (accessed on 28.06.2022).
 24. R. Dale. Natural language generation: The commercial state of the art in 2020 // Natural Language Engineering. № 26(4). S. 481-487. URL: <https://www.cambridge.org/core/journals/natural-language-engineering/article/natural-language-generation-the-commercial-state-of-the-art-in-2020/BA2417D73AF29F8073FF5B611CDEB97F> (accessed on 28.06.2022).
 25. RuGPT-3 // URL: <https://developers.sber.ru/portal/tools/rugpt-3> (accessed on 28.06.2022).
 26. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Re´mi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, Alexander M. Rush. Transformers: State-of-the-Art Natural Language Processing // Proceedings of the 2020 Conference on Empirical Methods in Natural

- Language Processing: System Demonstrations. 2020. S. 38 – 45. URL: <https://aclanthology.org/2020.emnlp-demos.6/> (accessed on 28.06.2022).
27. Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei. Language Models are Few-Shot Learners // arXiv preprint. arXiv:2005.14165. 2020. URL: <https://arxiv.org/abs/2005.14165> (accessed on 28.06.2022).
 28. Wahab, A., Tayara, H., Xuan, Z. et al. DNA sequences performs as natural language processing by exploiting deep learning algorithm for the identification of N4-methylcytosine // Sci Rep 11, 212 (2021). URL: <https://www.nature.com/articles/s41598-020-80430-x> (accessed on 28.06.2022).
 29. Ponkin I.V. Koncept mashinochitaemogo i mashinoispolnyaemogo prava: aktual'nost', naznachenie, mesto v RegTekhe, sodержanie, ontologiya i perspektivy // International Journal of Open Information Technologies. 2020. №9. URL: <https://cyberleninka.ru/article/n/kontsept-mashinochitaemogo-i-mashinoispolnyaemogo-prava-aktualnost-naznachenie-mesto-v-regtehe-soderzhanie-ontologiya-i-perspektivy> (accessed on 28.06.2022).
 30. Hugging Face // URL: <https://huggingface.co> (accessed on 28.06.2022).
 31. DeepPavlov: rubert-base-cased-sentence. Model card // URL: <https://huggingface.co/DeepPavlov/rubert-base-cased-sentence> (accessed on 28.06.2022).
 32. Blog kompanii YAndeks. YAndeks vylozhil YaLM 100B – seychas eto krupnejshaya GPT-podobnaya nejroset' v svobodnom dostupe. Vot kak udalos' eyo obuchit' // URL: <https://habr.com/ru/company/yandex/blog/672396/> (accessed on 03.07.2022).