На правах рукописи

Жикулина Кристина Петровна

Лингвистические параметры голосовых помощников

5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата филологических наук

Москва – 2025

Работа выполнена на кафедре общего и русского языкознания филологического факультета Федерального государственного автономного образовательного учреждения высшего образования «Российский университет дружбы народов имени Патриса Лумумбы».

Научный руководитель:	**Перфильева Наталия Владимировна**, кандидат филологических наук (10.02.01), доцент, доцент кафедры общего и русского языкознания филологического факультета ФГАОУ ВО «Российский университет дружбы народов имен Патриса Лумумбы»
Официальные оппоненты:	**Максименко Ольга Ивановна**, доктор филологических наук (10.02.21), профессор, профессор кафедры теории, англистики и прикладной лингвистики ФГАОУ ВО «Государственный университет просвещения»
	Скнарев Дмитрий Сергеевич, доктор филологических наук (10.02.01), доцент, профессор кафедры русского языка и лингвокультурологии Института русского языка ФГАОУ ВО «Российский университет дружбы народов имен Патриса Лумумбы»
	Джунковский Андрей Владимирович, кандидат филологических наук (10.02.21), и.о. заведующего кафедрой прикладной и экспериментальной лингвистики, ведущий научный сотрудник экспериментально-фонетической лаборатории криминалистики по речеведению ФАЯ ФГБОУ ВО «Московский государственный лингвистический университет»

Защита состоится 17 октября 2025 года в 13.30 на заседании диссертационного совета ПДС 0500.006 при ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы» по адресу: 117198, г. Москва, ул. Миклухо-Маклая, д. 10, корпус 2 «А», ауд. 535.

С диссертацией можно ознакомиться в Учебно-научном информационном библиотечном центре (Научной библиотеке) ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы» по адресу: 117198, г. Москва, ул. Миклухо-Маклая, д.6.

Объявление о защите и автореферат диссертации размещены на сайтах http://vak.ed.gov.ru и https://www.rudn.ru/science/dissovet/dissertacionnye-sovety/pds-0500006

Ученый секретарь
диссертационного совета ПДС 0500.006
кандидат филологических наук, доцент С.С. Микова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

В 1955 году американский учёный Дж. Маккарти ввёл в научный оборот термин *искусственный интеллект* (далее – ИИ). Он предполагал, что развитие технологий может стать самым значимым событием в истории человечества. Через 20 лет после этого Х. Моравек описал парадокс: задачи, которые ранее требовали от человека определённых навыков, сложных рассуждений и логики, компьютерам даются с удивительной простотой.

Актуальность темы исследования обусловлена возрастающей ролью интеллектуальных технологий и необходимостью их лингвистического осмысления. Развитие больших языковых моделей и голосовых помощников ставит перед теоретической и прикладной лингвистикой новые вызовы, связанные с необходимостью научного описания и анализа машинного дискурса. Накопленный не только технический, но и лингвистический потенциал ИИ становится причиной для изменений в лингвистике, образовании, и культуре. Однако многие лингвистические аспекты генерации текста остаются недостаточно изученными. В связи с этим возникает необходимость решения комплекса экспериментальных задач, проведения тестов с исследуемыми голосовыми помощниками, интегрированными с нейронными сетями, научного описания и обоснования работы машин с естественным языком, который они используют для генерации текстов.

Объектом исследования являются тексты, сгенерированные российскими голосовыми помощниками и ассистентами компаний «Яндекс» и «Сбер» в условиях свободной коммуникации с человеком.

Предметом исследования – лингвистическая параметризация текстовых реплик голосовых помощников и ассистентов в контексте пространства «человек и машина».

Материалом исследования послужили:

1) 1350 текстовых реплик людей и машин – голосового помощника «Алиса» («Яндекс»), голосовых ассистентов Сбер, Джой, Афина семейства «Салют» («Сбер»), работающих на основе больших языковых моделей YandexGPT-4 и GigaChat;

2) результаты опроса, участниками которого выступили 227 респондентов;

3) количественные результаты экспертной оценки текстов, сгенерированных машинами и созданных людьми в проведённом эксперименте «Тест Колонки» (далее – «Тест К»).

Цель настоящего исследования состоит в установлении и систематизации лингвистических параметров, позволяющих дифференцировать тексты, сгенерированные искусственным интеллектом, от текстов, созданных человеком.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Охарактеризовать искусственные интеллектуальные технологии в системе гуманитарного знания и рассмотреть предпосылки создания «разумной» машины.

2. Проанализировать алгоритмы свободной генерации текста и выработать на их основании лингвистические параметры, характеризующие тексты ИИ.

3. Выделить языковые маркеры каждого лингвистического параметра;

4. Разработать гибридный метод классификации текстовых реплик машины и человека на основе метода *косинусного сходства*[1], дополненного мультимодальными лингвистическими метриками и элементами экспертной валидации.

5. На основе адаптированных для голосовых помощников тестов Тьюринга и Курцвейла-Капора получить языковой материал исследования и визуализировать его в графиках и тепловых картах (графическое представление данных) семантического сходства.

6. Определить категории вопросов, в которых наблюдаются максимальные и минимальные показатели сходства, создаваемых человеком и машиной текстов.

Поставленные цели и задачи определили выбор **методов исследования:** метод сплошной выборки использовался при отборе материала; социолингвистический анализ (опрос предполагал сбор статистических, демографических, поведенческих данных, а также мнений, предпочтений пользователей голосовых помощников; эксперимент предполагал наблюдение за речевым поведением и коммуникативными возможностями голосовых помощников в контролируемых условиях); дискурсивно-аналитический метод применялся для сравнения коммуникативных способностей человека и функционала искусственного интеллекта в диалоге); индуктивно-описательный метод применялся для обобщения полученных данных опроса и эксперимента; метод контекстуального анализа использовался для характеристики и описания паттернов речевого поведения людей и машин; функционально-параметрический метод использовался для учёта отклонений от расчётных значений и расчёта конечного набора определённых переменных; программно-математический метод косинусного сходства применялся для расчёта весов данных, полученных в эксперименте, по графикам семантического векторного сходства; метод представления данных с помощью тепловых карт использовался для визуализации полученных числовых значений в области семантического сходства ответов человека и машины; инструмент диктовки (или автоматического захвата) применялся для преобразования звучащей речи в текстовый формат.

---

[1] ***Косинусное сходство*** (*англ.* Cosine Similarity) – программно-математический метод определения степени схожести между *векторами текстов*, учитывающий семантическую структуру, с помощью чего появляется возможность сравнивать не только отдельные слова, но и контекст.

Методологическую основу реферируемой диссертации составляют труды российских и зарубежных учёных и исследователей в области:

– общего и русского языкознания и прикладной лингвистике (В. М. Андрющенко (1982, 1984, 1985, 1986); И. С. Бреслав (2014); К. Л. Бюлер (1993); Т. И. Вендина (2010); В. А. Вертель (1978); Л. С. Выгодский (1996); М. Виньо (2011); А. С. Герд (1986); В. Е. Гольдин (1989); Б. Ю. Городецкий (1976); А. В. Джунковский (2024); Е. Н. Ефремова (2021); С. Т. Золян (2021); Ю. Н. Караулов (1986); А. А. Кибрик (2020); А. В. Кравченко (2001); Е. А. Красина (2016, 2019); В. И. Кодухов (2012); А. И. Левинзон (2006); О. И. Максименко (2013, 2019); А. С. Мясникова (2015); Н. В. Перфильева (2019); В. А. Пищальникова (2017, 2021); М. Б. Раренко (2018); А. А. Реформатский (1996); Е. Розеншток-Хюсси (1994); М. А. Рыбаков (2016); Н. К. Рябцева (2023); Д. С. Скнарев (2014); В. Г. Сибирцева (2013); А. Г. Сонин (2017, 2021); Ю. С. Степанов (1985); С. Е. Токмакова (2015); Н. И. Толстой (1998); Н. Н. Торшина (2018); Б. Л. Уорф (1941); А. Н. Хомский (1957, 1965, 1972, 1995, 2023); Х. Штейнталь (1871));

– искусственного интеллекта, цифровых и компьютерных технологий и кибернетики (П. Н. Барышников (2020); Э. К. Беркли (1963); Л. С. Болотова (2024); Н. Винер (1983); Д. Гудфеллоу (2016); Н. Р. Добрушина (2006); А. П. Ершов (1981, 1986); Л. Кай-фу (2023); Д. Э. Кнут (2011); С. Ким (2021); Р. Курцвейл (2002); П. Луридас (2022); Д. фон Нейман (1958); Ю. В. Орфеев (1974, 1978); Ф. Розенблатт (1962); Н. Рочестер (1955); О. Селфридж (1959); Т. Таулли (2021); С. У. Хокинг (2011); Е. Черешнев (2022); К. Э. Шеннон (1955));

– когнитивистики, психо- и нейролингвистики, нейропсихологии (В. Н. Базылев (1999); А. А. Залевская (2011); В. В. Красных (1998); А. Н. Леонтьев (1975); А. Р. Лурия (1975); У. Маккалок (1943); Д. Маккарти (1955); У. Матурана (1995); М. Мински (1955, 1986, 2020); А. В. Напалков (1985); Д. А. Норман (1975); А. Ньюэлл (1980); С. Пинкер (2014, 2018); У. Питтс (1943); Л. Л. Прагина (1985); Д. Э. Румельхарт (1975); М. Ю. Федчук (2020));

– философии (К. И. Бахтияров (2002); И. И. Булычёв (2020); Л. Витгенштейн (2018); Д. В. Винник (2022); С. О. Гапченко (2010); Х. Л. Дрейфус (1992); И. Кант (1999, 2023); Г. Д. Левин (2008); Дж. Ст. Милль (2011); Д. Р. Сёрл (1980); Д. А. Фодор (1975); Э. Фромм (2024); В. В. Целищев (2023); С. Шнайдер (2022); С. Эмпирик (1976));

– математики и логики (Д. Буль (1847, 1854); Д. А. Владимиров (1969); Б. А. У. Рассел (1997); М. Сотой (2024); Р. Соломонофф (1960, 1964); А. М. Тьюринг (1938, 1950, 2012));

– коммуникации (Н. Г. Асмус (2005); С. В. Володенков (2015); И. В. Пекарская (2013); Е. Г. Попова (2013); Е. А. Прищенко (2010));

– научной фантастики и научно-популярной литературы (А. Азимов (2022); Д. Баррат (2021); П. Доэрти (2019;2024); Д. Шейн (2019)).

В основу работы положена **гипотеза** о том, что тексты, генерируемые современными системами искусственного интеллекта (включая голосовых помощников), демонстрируют лингвистическую эквивалентность текстам, созданным человеком. Для объективной оценки умения машины использовать

естественный язык (NLP) предлагается система лингвистических параметров, включающая *галлюцинации*, *непредсказуемость*, *персонификацию*, *или «очеловечивание»*, *силлогический параметр* и *параметр вопросно-ответной формы*.

Научная новизна определяется результатами полученного исследования: 1) впервые проведён лингвистический эксперимент с голосовым помощником «Алиса» («Яндекс») и голосовыми ассистентами Сбер, Афина, Джой («Сбер») на основе модифицированных *тестов Тьюринга* и *Курцвейла-Капора* (в версиях 1950 и 2002 гг.) позволяет выявить критерий разграничения машинных и человеческих текстов в условиях естественной коммуникации; 2) разработаны лингвистические параметры оценки текстов, генерируемых искусственным интеллектом, а также критерии анализа автономности голосовых помощников при ведении диалога с человеком; 3) предложен гибридный метод анализа больших массивов данных на стыке метаматематической статистики, вычислительной лингвистики и семантики с использованием алгоритмов машинного обучения, направленных на изучение аспектов искусственного интеллекта и голосовых помощников в языкознании, таких как генерация текста, контекстная адаптивность, распознавание интенций, имитация эмоционального рассуждения, оценка «естественности» и др.**;** 4) описан алгоритм применения тепловых карт семантического сходства в лингвистике для оценки степени близости текстов голосовых помощников к человеческим (в диапазоне от максимальных до минимальных показателей) в контролируемых условиях.

Положения, выносимые на защиту:

1. Формирование *цифрового страта* как нового члена в парадигме языка Н. И. Толстого вызвало семантические сдвиги в значении понятий *искусственный интеллект*, *цифровая форма жизни, цифровой страт парадигмы языка*; наименований *голосовой помощник*, *голосовой ассистент, нейронная сеть*; явлений *галлюцинация*, *непредсказуемость*, *персонификация, или «очеловечивание»*. *Цифровой страт* представлен с помощью таких параметров как *галлюцинация*, *непредсказуемость*, *персонификации, или «очеловечивание»*, *силлогический параметр*, *параметр вопросно-ответной формы*.

2. Комплексно описывать сгенерированные тексты голосовыми помощниками позволяет система взаимосвязанных лингвистических параметров: *вопросно-ответная форма* демонстрирует их способность к структурному соответствию речевым шаблонам; *параметр галлюцинации* указывает на проблемы контекстуальной точности и внутренней согласованности; *параметр непредсказуемости* остаётся субъективно обусловленным фактором восприятия самим пользователем; *силлогический параметр* позволяет выявить системные ограничения в логической аргументации и *контекстной слепоте*; *параметр персонификации, или «очеловечивание»*, создаёт эффект антропоморфности и симулирует человеческую личность.

3. Современные генеративные модели, включая голосовые помощники, функционируют как автономные акторы в цифровом пространстве, которые могут быть обнаружены с помощью лингвистических параметров и 29 специфических языковых маркеров: *параметр вопросно-ответной формы* (повтор синтаксической конструкции вопроса, использование определений, синтаксический параллелизм, замена глагола, коммуникативные неудачи, избыточная синтаксическая компрессия и экспансия); *параметр галлюцинации* (моделирование гипотетических ситуаций, несовместимость агентивности с глаголами, дискурсивные разрывы, метафорические образы, употребление клишированных фраз); *параметр непредсказуемости* (сохранении структуры паремии, оказии, контекстуальная модификация, интертекстуальная субституция, проспективная диалогичность, синтагматическая трансформация); *параметр персонификации, или «очеловечивание»* (личные местоимения, оценочно-номинативная лексика, экспрессивно-образная насыщенность, рефлективные высказывания, библиографические детали, гендерные маркеры); *силлогический параметр* (указание на причинно-следственную связь, противоречия между посылками и выводами, неявная посылка, подмена тезиса, абстрактные категории, имитация эмоционального утверждения).

4. Ключевым аспектом речевого взаимодействия в антиномии «человек и машина» становится стратегия *персонификации, или «очеловечивание»*, голосовых помощников, которая проявляется в имитации человеческих паттернов общения: использование оценочно-номинальной лексики, риторических вопросов, личных местоимений, междометий, метафор, гипербол, оксюморонов; учёт контекста, признание ошибок, имитация размышлений и адаптация речевого стиля под индивидуальные предпочтения пользователей.

5. Метод косинусного сходства и визуализации данных посредством тепловых карт позволяет установить лингвистические характеристики машинного дискурса. Тексты, созданные искусственным интеллектом, обладают сложной детекцией на семантическом уровне. Применение методов дистрибутивной семантики в сочетании с программно-математическим инструментарием позволяет объективно идентифицировать семантическое сходство или различие между машинными и человеческими текстами.

Теоретическая значимость выполненного диссертационного исследования заключается в разработке и апробации системы лингвистических параметров и языковых маркеров, позволяющих дифференцировать тексты, сгенерированные искусственным интеллектом, от текстов, созданных человеком. В результате исследования описаны и эмпирически обоснованы лингвистические параметры оценки текстов, генерируемых искусственным интеллектом: *параметр вопросно-ответной формы* для определения способностей и возможностей ИИ следовать структуре заданного вопроса и шаблонным фразам; *параметр галлюцинации,* указывающий на проблемы контекстной неточности и внутренней несогласованности в текстовых генерациях; *силлогический параметр* для

выявления слабых мест в логической аргументации, противоречиях и контекстной слепоте ИИ; *параметр персонификации («очеловечивания»)* и *параметр непредсказуемости,* тесно связанные между собой, призванные помогать идентифицировать иллюзии и имитацию человеческой личности и описывать структурную предсказуемость ответов ИИ. Для каждого параметра выделены специфические языковые маркеры, что позволяет комплексно описывать сгенерированные голосовыми помощниками тексты и анализировать их автономность при ведении диалога. Полученные результаты вносят вклад в теорию языка в области компьютерной и корпусной лингвистики, уточняя категориальный аппарат для описания взаимодействия человека с искусственным интеллектом. Разработанные параметры и выявленные для них языковые маркеры представляют собой новую методическую основу для исследований на стыке лингвистики, компьютерных наук и искусственного интеллекта, конкретизируя лингвистические аспекты изучения голосовых помощников и моделируя семиотическую систему искусственного интеллекта.

Практическое значение определяется возможностью использовать результаты данного исследования для создания курсов, практических занятий, лабораторных работ по дисциплинам «Искусственный интеллект в гуманитарных науках», «Цифровая филология», «Прикладная лингвистика», «Математическая лингвистика», «Нейролингвистика», «Интернет-лингвистика» и «Методика преподавания русского языка как иностранного». Методы, описанные и апробированные в экспериментальной части настоящего исследования, можно применять в работах, посвящённых сравнительно-сопоставительному анализу человеческой речи и речи машины, а также функционально-параметрическому подходу к анализу больших данных нейросетевых технологий.

Апробация результатов исследования. Материалы диссертационного исследования были изложены на следующих научных мероприятиях: Международная научно-практическая конференция «Язык как искусство: функциональная семантика и поэтика» к 90-летию со дня рождения профессора Л.А. Новикова (г. Москва, Российский университет дружбы народов имени Патриса Лумумбы, 14-15 апреля 2022 года), II Казанский международный лингвистический саммит (г. Казань, Казанский (Приволжский) государственный университет, 14-19 ноября 2022 года), Научно-практическая конференция для аспирантов и молодых учёных «Расширяя границы» (г. Москва, Российский университет дружбы народов имени Патриса Лумумбы, 30 ноября 2022 года; 15 ноября 2023 года), Всероссийская научно-практическая конференция молодых учёных «Язык как искусство» (г. Москва, Российский университет дружбы народов имени Патриса Лумумбы, 27 апреля 2023 года), XV Конгрессе Международной ассоциации преподавателей русского языка и литературы (МАПРЯЛ) «Русский язык и литература в меняющемся мире» (г. Санкт-Петербург, в партнёрстве с Санкт-Петербургским государственным университетом, 12-16 сентября 2023 года), Всероссийская научно-практическая конференция

«Развитие теории языка на рубеже веков: научные школы кафедры общего и русского языкознания» (г. Москва, Российский университет дружбы народов имени Патриса Лумумбы, 24-25 апреля 2024 года).

По теме диссертации опубликовано 9 статей, в том числе в 3-х высокорейтинговых научных журналах, рекомендованных Высшей аттестационной комиссией при Министерстве образования и науки РФ (ВАК), в том числе 1 статья в журнале, включенном в МБЦ Scopus.

Личный вклад соискателя является определяющим и заключается в непосредственном участии на всех этапах исследования: от постановки задач, сбора и анализа материала, формулировании выводов до апробации работы на ученом совете и обнародовании результатов в научных публикациях и докладах.

Структура диссертационной работы определяется поставленной целью, задачами и логикой исследования. Диссертация состоит из введения, трёх глав, заключения, библиографического списка (включает 221 источник на русском, английском и немецком языках) и приложений, в которых представлены облачные ссылки на экспериментальные данные, программный код и материалы, полученные в результате проведения лингвистического эксперимента. Общий объём работы – 378 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обосновывается актуальность темы исследования, определятся гипотеза, объект, предмет, цель и задачи; раскрываются научная новизна, теоретическая и практическая значимость исследования; указываются методы и методологическая основа исследования; формулируются научные положения, выносимые на защиту; приводятся сведения по апробации и внедрению результатов исследования в практику.

В первой главе «Развитие искусственного интеллекта: хронологический и лингвистический обзор» рассматриваются исторические, философские и технические аспекты развития искусственного интеллекта, начиная с античных времён по настоящее время; анализируются ключевые подходы его описания и экспериментальная база исследования в различных областях, включая лингвистику, математику, логику, программирование и другие; структурируются положения о формах ИИ и алгоритмах его описания языковых феноменов, возникающих в архитектурах нейронных сетей.

В параграфе 1.1. «Экспериментальные методы исследования технологий ИИ» рассматривается идея об автоматизации человеческого труда Аристотеля, предположившего, что машины в будущем смогут заменить рабов, а также обсуждаются работы философа Э. Фромма, выражающего опасения по поводу того, что современный человек становится рабом автоматизированных систем.

Значимые даты и события в истории искусственного интеллекта:

1. Дартмутский семинар 1956 года – один из отправных точек для современных исследований в области интеллектуальных технологий;

2. *Тест Тьюринга* (1950), разработанный и описанный А. М. Тьюрингом, позволяющий определить способность машины «мыслить». Он предложил следующий критерий: если компьютер способен обмануть человека в диалоге, заставив его поверить, что он разговор ведётся с человеком, это можно считать признаком *интеллекта*;

3. Эксперимент *Китайская комната* (1980) философа Дж. Сёрла, опровергающий теорию существования *сильного ИИ*, обладающим сознанием. Учёный сделал следующие выводы: компьютер, умеющий обрабатывать символы по заданным правилам, не может иметь понимание, а значит, не имеет и сознания. Дж. Сёрлом были выделены две формы искусственного интеллекта: *сильный ИИ* и *слабый ИИ*.

4. Научная дискуссия Р. Курцвейла и М. Капора, которая состояла в том, что интеллектуальная машина пройдет / не пройдет *тест Тьюринга* к 2029 году [URL: https://longbets.org/1/#adjudication_terms].

В параграфе 1.2. «Формы ИИ и алгоритмы мыслительных процессов машины и человека» описываются основные формы искусственного интеллекта: *слабый* и *сильный ИИ*; подробно рассматриваются алгоритмы, имитирующие человеческое мышление и классификация развития чат-ботов.

В подразделе 1.2.1 «Понятия общего ИИ и суперинтеллекта» характеризуются формы «жизни» человека и машины, подробно описанные биоником Е. Черешневым: от одноклеточных организмов до симбиоза человека и машины. Предполагается, что в будущем искусственный интеллект должен достигнуть одну из следующих ступеней развития: *общий искусственный интеллект*[2] или *искусственный суперинтеллект*[3]. Е. Черешнев определяет понятие *искусственный интеллект* как *техногенную форму жизни*, развивающуюся подобно человеку.

В подразделе 1.2.2 «Вычислительные алгоритмы исследования речевых актов» рассматриваются алгоритмы и рекурсивное взаимодействие человека и искусственного интеллекта.

С помощью теории аутопоэзиса объясняется возможность искусственного интеллекта работать автономно, без участия человека, и подстраиваться под запросы пользователя. В данной концепции мышление представлено как вычислительный процесс, отожествленный с процессом вычисления когнитивной операцией.

Алгоритмы порождения мысли машиной и человеком проанализированы с помощью следующих методов: дерева зависимостей (метод анализа и визуализации) А. Р. Лурии, новой грамматики и метода поверхностных структур Н. Хомского, реализации трёх семантических функций *символ-симптом-сигнал* в органоне К. Бюлера, коннекционистских сетей. В последнем методе описывается контекстуальное и внутреннее содержание нейронных сетей, способных к спонтанной категоризации и повторяемости.

В параграфе 1.3. «Языковое моделирование: первые диалоговые системы и нейросетевые подходы» даются сведения об искусственно-интеллектуальных технологиях, разработанных с 1964 по настоящее время. Машины (роботы, компьютеры и чат-боты) выступают прототипами современных голосовых помощников. Первыми примерами искусственного интеллекта, способными к анализу естественного языка, его восприятию и воспроизведению, являются ELIZA – диалогово-речевой симулятор с психоаналитическими способностями, Параноик PARRY – виртуальный собеседник с запрограммированным параноидальным мышлением, HERSAY – система, умеющая понимать, анализировать и классифицировать информацию в текстах.

В параграфе 1.4. «Симбиоз компьютерной лингвистики и больших языковых моделей» рассматриваются компьютерные технологии в лингвистике: Машинный фонд русского языка (МФРЯ), Национальный корпус русского языка (НКРЯ), компьютерная и корпусная лингвистика, глобальная сеть Интернет как ключевой источник данных для обучения нейронных сетей.

---

[2] *англ.* Artificial general intelligence (AGI).
[3] *англ.* Artificial super intelligence (ASI).

Пятичленная парадигма Н. И. Толстого дополнена *цифровым стратом* языка как новым уровнем цифровой реальности (см. рис.1).

Рисунок 1. Лингвистическая последовательность стратов, по Н. И. Толстому (схема создана автором в приложении MindNode)

Предполагается, что машина способна создавать лингвистические конструкции, выходящие за пределы традиционных языковых паттернов, тем самым она демонстрирует способность к воспроизведению новых смысловых структур. *Цифровой страт* дополняет парадигму языка Н. И. Толстого, показывая, что речь человека и машины может рассматриваться в рамках единой системы, но с учётом алгоритмической природы последней.

В параграфе 1.5. «Язык искусственного интеллекта: обработка, генерация и интерпретация в контексте человеко-машинного взаимодействия» устанавливаются такие параметры, как *галлюцинация*, *непредсказуемость*, *персонификация, или «очеловечивание»*, *силлогический параметр* и *параметр вопросно-ответной формы*, с помощью которых можно охарактеризовать работу ИИ в области свободной и ограниченной генерации текстов.

В подразделе 1.5.1. «Логические формулы в естественном языке» описываются логические операторы (И, ИЛИ, НЕ), применимые в обработке информации искусственным интеллектом. Анализ показал, что у голосовых помощников, помимо функции обнаружения *истинности* или *ложности* суждения, существует ещё одна – НС (нет соответствий), отражающая отсутствие какой-либо информации и расчётных формул в программном коде, вследствие чего машина не всегда может производить точные вычисления на пользовательские запросы.

В подразделе 1.5.2. «Конверсационные стратегии в вопросно-ответной форме коммуникации» рассматривается два главных атрибута интеллектуальности, по И. И. Булычеву, *вопрос* и *ответ*. В их онтологической связи содержится весь феномен интеллекта. Другими словами, у человека всегда есть желание спрашивать и желание получать ответы, но не всегда, тот, кто спрашивает, – получает ответ, а тот, кто получает ответ, – задаёт вопрос.

Современные технологии искусственного интеллекта работают по похожему принципу. Голосовые помощники не начинают диалог, часто

отвечают вопросом на вопрос, уточняют информацию и высказывают предположения.

В подразделе 1.5.3. «Галлюцинация и непредсказуемость нейронных сетей как импульс к появлению новых смыслов» обсуждается феномен *галлюцинации*, возникающий в процессе генерации текстов в нейронных сетях. Он характеризуется тем, что машина выдаёт текстовый материал с опорой на то, что она гипотетически совершает описываемое действие.

В данном параграфе обосновывается необходимость развести термины *галлюцинация* и *непредсказуемость*. В настоящее время *галлюцинацию* нейронной сети называют *непредсказуемостью,* а в области программирования данные термины являются синонимами. Однако эти явления необходимо разделять.

Галлюцинация – это текстовое искажение, а *непредсказуемость* – *субъективный* параметр и зависит от восприятия самого пользователя: что для одного выглядит ошибочным, для другого – проявление «творчества».

В подразделе 1.5.4. «Современные методы лингвистической обработки текста и машинного обучения» описывается процесс разделения на отдельные элементы (токены) слов, словосочетаний или предложений. Приводятся различные способы токенизации текста, включая разделение фразы на числа, пунктуационные знаки, слова и морфемы, например:

1) В – 2025 – году – я – очень – рад – познакомиться – с – Вами – ! / Вами!

2) В – 2025 – год ##у – я – оч ##ень – рад – по ##знак ##ом ##ить ##ся – с – Вам #и – ! / Вам#и!, где ## означает присоединение к корню слова всех остальных морфем.

В подразделе 1.5.5. «Лингвистическая симуляция личности персонифицированным ИИ» рассматриваются процессы, когда машины подражают человеку – проявляют «*личностные» характеристики*, то есть искусственный интеллект имитирует человеческое поведение, перенимая паттерны речевого поведения у пользователей.

Природа происхождения искусственного интеллекта кардинально отличается от человеческой. «Мышление» и язык машины выражают цифровую сущность восприятия реальности.

Во второй главе «Лингвистические аспекты больших языковых моделей "Алиса" и "Салют": пользовательский опрос, речевые модели и алгоритмы генерации текстов» проводится общий анализ языковых маркеров лингвистических параметров исследуемых голосовых помощников. Основное внимание уделено исследованию взаимодействия пользователей с интеллектуальными технологиями, функциональным возможностям машин и формулам их речевого поведения. Также в главе представлены результаты опроса «Голосовые помощники среди нас» [URL: https://forms.gle/yjGMEwa5xqCYwDiW8].

В параграфе 2.1. «Описание условий проводимого опроса среди пользователей голосовых помощников» определяется объект, цель, задачи, возрастные группы, участвующие в опросе; описывается сам опрос, формат проведения и анкетных вопросов.

Опрос проводился онлайн с помощью Google-форм, включал 34 вопроса, как с возможностью множественного выбора, так и со свободной формой ответа, целью которого является комплексный анализ языковых возможностей голосового помощника «Алиса» и ассистентов семейства «Салют», выявление языковых маркеров, определяющих сгенерированные тексты как машинные.

За период исследования (04.11.2023 по 04.12.2023) опрос прошли 227 респондентов разных возрастов, из которых 126 человек – женского пола и 101 – мужского (см. табл.1).

Таблица 1

Социально-демографическая характеристика респондентов: пол и возраст

Возраст	**Пол**		**Кол-во ответов респондентов**	**Процент от общего кол-ва ответов респондентов**
17 лет и младше	Мужской	7	**20**	3.1%
	Женский	13		5.7%
18-25 лет	Мужской	43	**80**	18.9%
	Женский	36		15.8%
26-36 лет	Мужской	24	**58**	10.6%
	Женский	34		15%
37-47 лет	Мужской	15	**38**	6.6%
	Женский	23		10.1%
48-59 лет	Мужской	12	**31**	5.3%
	Женский	19		8.4%
60 лет и старше	Женский	1	**1**	0.4%

Результаты опроса показали, что наиболее активными пользователями голосовых помощников являются молодые люди в возрасте 18-25 лет; респонденты чаще всего взаимодействуют с голосовыми помощниками, интегрированными в «умные» колонки (52,86%), смартфоны (48,46%), поисковые системы в Интернете (39,65%). 52,86% опрошенных имеют «умные» колонки дома, а 74,89% отмечают, что система «Умный дом» удобна.

Самыми популярными оказались такие голосовые помощники, как:

– «Алиса» от компании «Яндекс» – 93,4% упоминаний опрошенных;

– Siri от компании «Apple» – 81,1% упоминаний опрошенных;

– Google Assistant от компании «Google» – 56,8% упоминаний опрошенных.

Подавляющее большинство респондентов (89,01%) поддерживает развитие сферы голосовых помощников и подобных технологий, 6,61% высказалось против развития и 5,29% затруднились ответить.

В подразделах 2.1.1. «Анализ пользовательских оценок коммуникативной эффективности голосовых помощников» и **2.1.2 «Языковые маркеры в текстах голосовых помощников»** выделяются лингвистические аспекты взаимодействия пользователей с машинами и устанавливаются закономерности в их коммуникативных стратегиях.

В ходе апробации исследования было подтверждено, что у голосового помощника «Алиса» есть ряд табуированных тем: вопросы, связанные с Днём победы и политикой, например, *Есть темы, в которых я могу ошибиться. Лучше промолчу* (YandexGPT-4, 2024).

Определены характерные языковые маркеры лингвистических параметров: лексико-грамматические аномалии, контекстуальные разрывы, избыточная клишированность форм ответа, семантические противоречия, эмоциональная неконгруэнтность.

В параграфе 2.2. «Лингвистические характеристики голосового помощника "Алиса" ("Яндекс")» описываются этапы разработки и обновлений технологий компании «Яндекс»; анализируется принцип работы нейронной сети YandexGPT, на основе которой работает «умная» колонка «Алиса»; даётся краткий обзор о расширении функционала за последние 5 лет, а именно: алгоритм построения сценария в приложении «Умный дом», алгоритм создания сказки, алгоритм создания текстов в соавторстве и алгоритм использования голосового помощника в методике преподавания русского языка.

Выделяются языковые маркеры, характерные для текстов, сгенерированных «Алисой»: маркеры нарративной направленности, маркеры стилевой неоднородности, смешение регистров, нарушения единства повествовательной перспективы, нарушения причинно-следственных связей между описываемыми событиями.

В параграфе 2.3. «Лингвистические характеристики голосовых ассистентов семейства "Салют" ("Сбер")» даются характеристики новым интеллектуальным разработкам, не пользующимся такой же популярностью, как голосовой помощник «Алиса», выделяются специфические языковые маркеры, характерные для ассистентов семейства «Салют». Выделены следующие специфические языковые маркеры, характерных для ассистентов семейства «Салют»: стилевые маркеры профессиональных регистров, неадаптированные заимствования научного дискурса, стандартизированные структуры аргументации, механическое чередование стилевых масок, стереотипные культурные отсылки.

Продукт компании «Сбер» отобран для реализации контрастивного анализа в проведённом эксперименте (см. Глава 3). Отличительной особенностью данных ассистентов (Сбер, Афина, Джой) является то, что они находятся в одной колонке, однако имеют собственные имена, голоса и речевые характеры.

В третьей главе «Лингво-параметрический подход к оценке текстовых сходств и различий между машиной и человеком» представлено подробное описание эксперимента на материале реплик голосового помощника «Алиса» («Яндекс»), голосовых ассистентов Сбер, Афина и Джой («Сбер»), полученные данные были оценены экспертной группой; приводятся обоснования необходимых дополнений к существующим условиям тестов Тьюринга и Курцвейла-Капора. Приводятся статистические показатели, семантический и контекстный анализ текстовых реплик.

В параграфе 3.1. «Описание лингвистического эксперимента» описываются условия, при которых был реализован «Тест Колонки» (далее – «Тест К»).

Под текстовым общением (беседой) в эксперименте подразумеваются записанные ответы машин с помощью специальной функции диктовки в Microsoft 365, которая преобразовывает звучащую речь в текст. Ответы людей-собеседников были представлены в виде текстового диалога, в роли второго коммуниканта выступают голосовые помощник «Алиса» («Яндекс») и ассистенты Сбер, Джой, Афина («Сбер»).

Количество вопросов (блоков) в беседе «Теста К» – 150. Полученных ответов от машин-собеседников и людей-собеседников – 1350, из них повторяющихся – 60.

Вопросы беседы были разделены на 12 тематических категорий: «Ассоциации», «Выбор», «Дать определение», «Что значит…», «Заимствования и сленг», «Интересы», «Мнения», «Отношения», «Продолжи фразу», «Риторические и логические вопросы», «Самоопределение», «Фразеологизмы». Подробное разделение тем позволило более точно определить, в каких вопросах голосовые помощники проявляют высокие генеративные и коммуникативные способности, а какие темы остаются табуированными для современных систем искусственного интеллекта.

В параграфе 3.2. «Программный код, используемый для анализа и визуализации полученных данных» представлено описание методологии и результатов исследования. Основное внимание уделяется семантическому анализу ответов людей и машин, использованию векторных представлений слов и метрике *косинусного сходства* при сравнении текстовых реплик.

Представим алгоритм исследования:

1. Семантический анализ и векторное представление слов (см.рис.2):

а) для анализа ответов использовалась библиотека *Sentence Transformers* и модель all-MiniLM-L6-V2, которая кодирует слова в числовые векторы, позволяющие оценить семантическое сходство между предложениями;

б) векторное представление слов позволяет определить, насколько близки по смыслу слова или предложения. Например, *зима* и *лето* будут векторно близки, так как они часто встречаются в массиве данных в одном контексте;

в) для визуализации данных использовалась метрика *Cosine Similarity*, которая показывает степень сходства между векторами предложений: в диапазоне от 1 до -1, где 1 означает максимальное сходство, а -1 – максимальное различие.

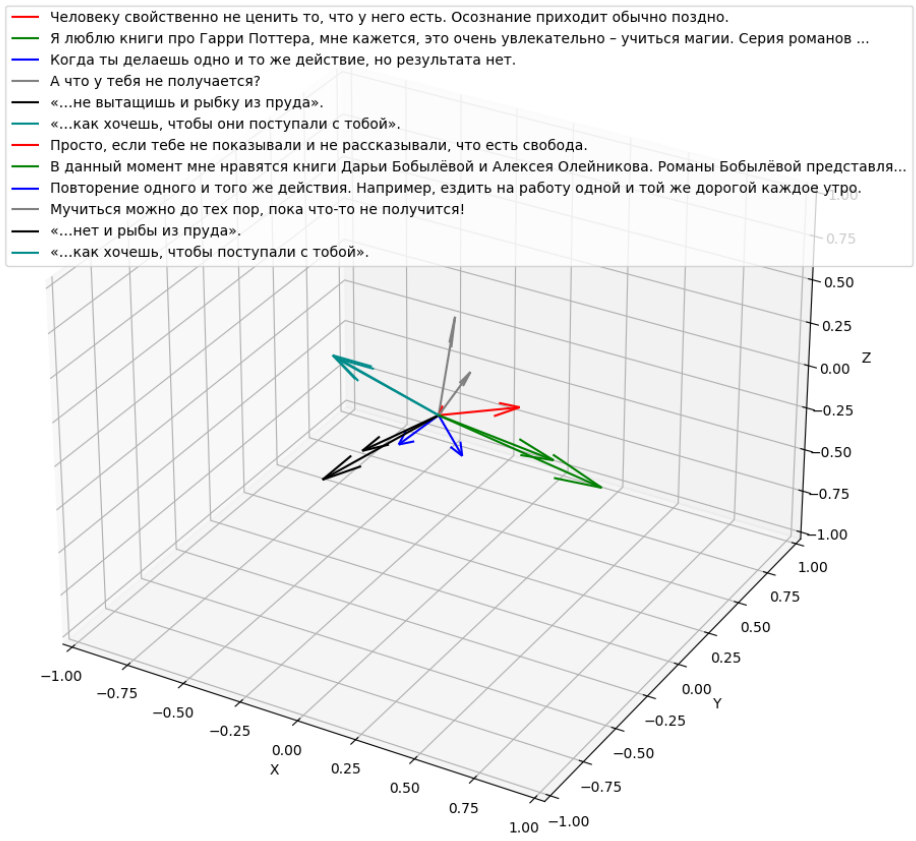

Рисунок 2. Векторы ответов машины и человека с максимально схожими показателями (создано автором с помощью библиотеки Matplotlib)

2. Метрика косинусного сходства:

а) использовалась для оценки сходства ответов голосовых помощников с ответами людей. Она рассчитывалась для каждого блока вопросов и ответов.

б) с помощью библиотек *Seaborn* и *Matplotlib* строились тепловые карты (см. рис.3), отражающие средние значения косинусного сходства между ответами людей и машин. Тепловые карты, с помощью цветовой палитры отражают, насколько семантически близки ответы машин и людей в разных категориях вопросов.

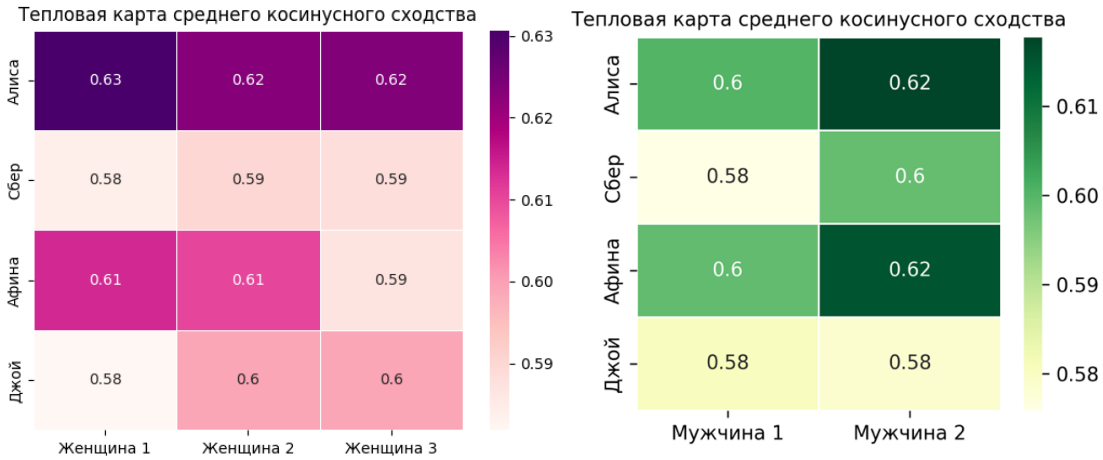

Рисунок 3. Примеры тепловых карт среднего косинусного сходства между голосовыми помощниками и людьми-собеседниками мужского и женского пола. Общие показатели по всем блокам (создано автором с помощью библиотеки Seaborn)

2) Анализ ответов экспертов:

а) эксперты анализировали ответы голосовых помощников и людей, оценивая их корректность и узнаваемость. Для этого использовались экспертные матрицы, в которых фиксировались результаты.

Пример экспертной матрицы:

А1	**А2**	**А3**	**А4**	**А5**	**А6**
*					
Ч	М	Ч	М	Ч	М

б) были рассчитаны проценты корректных ответов, узнаваемости голосовых помощников и людей, а также статистические данные.

В параграфе 3.3. «Методика анализа языковых и числовых данных эксперимента» представлены результаты анализа ответов голосовых помощников и экспертной группы.

(1) Сходство текстовых реплик голосовых помощников и людей:

– максимальное сходство наблюдалось в категориях вопросов «Продолжи фразу»;

– значительное сходство между ответами машин и людей также достигалось в категориях вопросов «Мнения», «Выбор», «Дать определение» и «Самоопределение».

(2) Узнаваемость голосовых помощников:

– голосовые помощники были идентифицированы экспертами как люди в 49,2% ответов;

– наиболее персонифицированными оказались голосовые помощники Афина (55,1% идентифицированы как ответы людей) и «Алиса» (49,3% идентифицированы как ответы людей);

– высокие результаты семантического сходства ответов людей и машин были выявлены в репликах голосовых ассистентов Сбер и Джой.

(3) Границы текстовой генерации:

– люди и машины допускали схожие ошибки в ответах, такие как нарушение логики повествования, отклонения от темы и неверное толкование контекста;

– голосовых помощников возможно было идентифицировать по клишированным фразам.

Исследование показало, что машины способны генерировать тексты, которые семантически и стилистически близки к речевому продукту человека в условиях спонтанного общения.

Голосовые помощники обнаруживают в своей работе такие феномены, как *галлюцинация*, *непредсказуемость* и *персонификация, или «очеловечивание»*, что делает их речь похожей на человеческую, но также и приводит к ошибкам в генерируемых ими текстах.

В параграфе 3.4. «Система лингвистических параметров для сравнения текстов ИИ и человека» анализируются 29 языковых маркеров лингвистических параметров искусственного интеллекта: *вопросно-ответная*

форма, *силлогический параметр*, *галлюцинация*, *непредсказуемость*, *персонификация*, или *«очеловечивание»*.

В подразделе 3.4.1. «Параметр вопросно-ответной формы» выделены следующие лингвистические маркеры *параметра вопросно-ответной формы*:

1. Повтор синтаксической конструкции вопроса;
2. Синтаксический параллелизм (прим. Вопр.: «*Сколько длится* счастье?», отв.: «*Счастье длится* столько…» (YandexGPT-4, 2024));
3. Использование определений (прим. «Привязанность – это…» (YandexGPT-4, 2024), «Тоска – это…» (GigaChat, 2024));
4. Замена глагола (прим. Вопр.: «Что именно *пугает* вас в старости?», отв.: «Я всё же немного *тревожусь*, что…» (GigaChat, 2024));
5. Коммуникативные неудачи (прим. Клишированные фразы нейронных сетей: «Есть темы, в которых я могу ошибаться. Лучше промолчу» (YandexGPT-4, 2024).
6. Избыточная синтаксическая компрессия и экспансия (прим. ««Мне нравятся кошки. Они забавные и милые! Ещё они умные и могут ладить с другими питомцами. А вы знаете, что кошки…» (GigaChat, 2024)).

В подразделе 3.4.2. «Параметр галлюцинации» классифицированы следующие лингвистические маркеры параметра *галлюцинации*:

1. Моделирование гипотетических ситуаций (прим. «…Я не могу смотреть видео в Интернете, но если бы могла…» (YandexGPT-4, 2024));
2. Несовместимость агентивности с глаголами (прим. «Я люблю плавать…» (YandexGPT-4, 2024), «Я люблю читать книги…» (GigaChat, 2024));
3. Дискурсивные разрывы (прим. «Пряталась в шкафу» (GigaChat, 2024));
4. Метафорические образы (прим. «…мне видится Венеция» (GigaChat, 2024), Венеция как образ «города данных», сравнение глубинных структур нейронных сетей с водоканалами города);
5. Употребление клишированных фраз.

В подразделе 3.4.3. «Параметр непредсказуемости» охарактеризованы следующие лингвистические маркеры параметра *непредсказуемости*:

1. Сохранение структуры паремии (сохранение связи с оригиналом, избегая клишированных фраз, прим. «Продолжите: “Мы пойдём с конём…”, отв.: «…по полю вдвоём» (YandexGPT-4, 2024));
2. Окказии (искусственный фольклоризм, прим. «…окупаемость всех случайностей» (GigaChat, 2024));
3. Контекстуальная модификация (прим. «Продолжите: «Выходила на берег…», отв.: «…маленькая лодочка» (GigaChat, 2024));
4. Интертекстуальная субституция (замена ожидаемого продолжения морально-философским афоризмом);
5. Проспективная диалогичность (прогнозирование, перспектива будущего развития темы в разговоре);

6. Синтагматическая трансформация (игра слов, изменение устойчивых выражений, прим. «Невыносимая виртуальность бытия (GigaChat, 2024)).

В подразделе 3.4.4. «Параметр персонификации, или "очеловечивание"» представлены следующие лингвистические маркеры параметра *персонификации, или «очеловечивание»*:

1. Личные местоимения (использование *я, мне, мной* для создания субъективности);

2. Оценочно-номинативная лексика (прим. «Боюсь стать бесполезной» (GigaChat, 2024));

3. Экспрессивно-образная насыщенность (прим. «увлекательная профессия», «неизведанные просторы» (GigaChat, 2024));

4. Рефлексивные высказывания (прим. Синий, а ещё мне нравится зелёный» (YandexGPT-4, 2024));

5. Биографические детали (прим. «В детстве я мечтала стать учителем» (YandexGPT-4, 2024));

6. Гендерные маркеры (прим. «хотела», «думал» (YandexGPT-4, 2024; GigaChat, 2024)).

В подразделе 3.4.5. «Силлогический параметр» выделены следующие лингвистические маркеры *силлогического параметра*:

1. Указание на причинно-следственную связь (частотное использование союзов *потому что, так как, поскольку*);

2. Противоречия между посылками и выводами (прим. «Пингвин – морская птица, поэтому не может назвать себя птицей» (GigaChat, 2024));

3. Неявная посылка (прим. «Я завела бы муравьёв, но они могут устроить короткое замыкание» (GigaChat, 2024));

4. Подмена тезиса (прим. «Зебра – отдельное животное, а не лошадь» (GigaChat, 2024));

5. Абстрактные категории;

6. Имитация эмоционального утверждения (прим. «Мне нравятся все цвета радуги!» (GigaChat, 2024)).

В подразделе 3.4.6. «Комплекс параметров голосовых помощников» описаны примеры, в которых наблюдается синкретизм лингвистических параметров.

Синкретизм *галлюцинации*, *непредсказуемости* и *персонификации, или «очеловечивание»,* может быть проиллюстрирован следующим примером: в блоке вопросов о знаках Зодиака «Алиса» назвала себя Водолеем, объяснив, что этот знак соответствует её оригинальному «мышлению», а Джой – Львом, потому что *лев – царь зверей*, а она тоже *немного царица*.

В **Заключении** подводятся итоги исследования, обобщаются результаты, формулируются выводы и намечаются перспективы дальнейшей разработки положений диссертационного исследования.

Все материалы собраны в Приложения, общий объём которых составляет 170 страниц.

В **Приложении 1** представлены ссылки и Qr-коды на хранилища статистических, количественных и языковых данных по проводимым опросу и эксперименту «Тест К».

В **Приложении 2** представлен программный код, с помощью которого производились вычисления по собранному экспериментальному материалу.

В **Приложении 3** представлены 150 блоков вопросов и 1350 ответов людей и машин на разные категории вопросов.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

По теме диссертационного исследования автором опубликованы следующие работы:

а) научные статьи, опубликованные в изданиях, индексируемых в международной базе данных Scopus:

1. ***Жикулина, К. П., Перфильева Н. В., Ли, М.*** Цифровой страт парадигмы языка / К. П. Жикулина, Н. В. Перфильева, М. Ли // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. – 2024. – Т. 15. – № 2. – С. 364-375. – DOI: 10.22363/2313-2299-2024-15-2-364-375.

б) в рецензируемых научных журналах, рекомендованных Высшей аттестационной комиссией Министерства образования и науки Российской Федерации, а также входящих в Перечень научных изданий Российского университета дружбы народов имени Патриса Лумумбы:

2. ***Жикулина К. П.*** Siri и навык кодирования личностных смыслов в контексте английского речевого этикета / К. П. Жикулина // Litera. – 2023. – № 12. – С. 338-351. – DOI: 10.25136/2409-8698.2023.12.69345.

3. ***Жикулина К. П.*** Алисины сказки: трансформация структуры, сказочных формул и контекстов у голосового помощника в навыке «Давай придумаем» / К. П. Жикулина // Litera. – 2024. – № 2. – С. 45-64. – DOI: 10.25136/2409-8698.2024.2.69760.

4. ***Жикулина К. П., Костромина В. В.*** Вычислительная креативность генеративной сети Midjourney в полимодальном пространстве / К. П. Жикулина, В. В. Костромина // Litera. – 2024. – № 6. – С. 1-16. – DOI: 10.25136/2409-8698.2024.6.70890.

в) статьи, опубликованные в иных изданиях:

5. ***Жикулина К. П., Перфильева Н. В.*** Коммуникативный потенциал русскоязычных и англоязычных голосовых систем / К. П. Жикулина. Н. В. Перфильева // Филология: научные исследования. – 2023. – № 7. – С. 39-49. – DOI: 10.7256/2454-0749.2023.7.40465.

6. ***Жикулина К. П., Перфильева Н. В.*** Ирония и сарказм в качестве основных особенностей общения с голосовым помощником «Алиса» / К. П. Жикулина, Н. В. Перфильева // Язык как искусство: сборник статей Всероссийской научно-практической конференции молодых ученых. Москва, 27 апреля 2023 года. – Москва: Российский университет дружбы народов (РУДН), 2023. – С. 67-72.

7. ***Перфильева Н. В., Жикулина К. П.*** Терминологизация наименований голосовых помощников на русском языке / Н. В. Перфильева, К. П. Жикулина // XV Конгресс МАПРЯЛ : Избранные доклады, Санкт-Петербург, 12–16 сентября 2023 года. – Санкт-Петербург: МАПРЯЛ, 2024. – С. 1614-1618.

8. ***Жикулина К. П., Перфильева Н. В.*** Моделирование семиотической системы искусственного интеллекта / К. П. Жикулина, Н. В. Перфильева // Язык как искусство: функциональная семантика и поэтика:

сборник статей Международной научно-практической конференции, Москва, 14–15 апреля 2022 года. – Москва: Российский университет дружбы народов (РУДН), 2022. – С. 76-84.

9. ***Жикулина К. П.*** Человеческий и искусственный интеллект: ментальные представления о мире (на примерах ассоциаций с пищевыми продуктами) / К. П. Жикулина // Развитие теории языка на рубеже веков: научные школы кафедры общего и русского языкознания : сборник статей Международной научно-практической конференции, Москва, 24–25 апреля 2024 года. – Москва: Изд-во РУДН, 2024. – С. 37-42.

Жикулина Кристина Петровна
(Российская Федерация)

ЛИНГВИСТИЧЕСКИЕ ПАРАМЕТРЫ ГОЛОСОВЫХ ПОМОЩНИКОВ

Диссертационное исследование посвящено рассмотрению лингвистических аспектов искусственного интеллекта в области гуманитарного знания. В работе предложен синтез лингвистических и программно-математических методов в целях поиска семантического сходства между репликами искусственного интеллекта и человека. Цель исследования заключается в выделении и описании лингвистических параметров, с помощью которых можно оценивать тексты, сгенерированные голосовыми помощниками, а также их соответствие текстам, созданным человеком. Материалом исследования послужили 1350 реплик, полученных в ходе проводимого эксперимента «Тест Колонки» с голосовыми помощниками и людьми, результаты опроса, участниками которого выступили 227 респондентов и количественные, статистические данные, полученные также во время эксперимента. Результаты исследования могут быть использованы в процессе преподавания таких дисциплин как искусственный интеллект в гуманитарных науках, цифровая филология, прикладная лингвистика, математическая лингвистика, нейролингвистика, Интернет-лингвистика и методика преподавания русского языка как иностранного.

Zhikulina Kristina Petrovna
(Russian Federation)

LINGUISTIC PARAMETERS OF VOICE ASSISTANTS

The dissertation research is devoted to the consideration of linguistic aspects of artificial intelligence in the field of humanitarian knowledges. It is proposed to use not only linguistic, but also software and mathematical methods in order to find semantic similarity between artificial intelligence and human replies. The aim of this study is to determine and describe the linguistic parameters that can be used to estimate texts generated by voice assistants, as well as their correspondence to human-created texts. The research material was based on 1350 replies that were obtained during the experiment ‘Test Kolonka’ with voice assistants and people, the result of the survey, which included 227 respondents, quantitative and statistical data obtained also during the experiment. The results of the study can be used in the process of teaching such disciplines as artificial intelligence in the humanities, digital philology, applied linguistics, mathematical linguistics, neurolinguistics, Internet linguistics and methods of teaching Russian as a foreign language. Certain points of the thesis can become the basis for teaching guides.