



УДК 304.2+811.93+372.881.1

Научная статья / **Research Full Article**DOI: [10.15293/2658-6762.2401.09](https://doi.org/10.15293/2658-6762.2401.09)Язык статьи: русский / **Article language: Russian**

Возможности искусственного интеллекта в сохранении языкового разнообразия для будущих поколений

Т. В. Ермолова¹, Н. В. Савицкая¹, О. В. Дедова¹, А. В. Гузова¹¹ Московский государственный психолого-педагогический университет, Москва, Россия

Проблема и цель. Актуальная и не до конца решенная проблема сохранения культурного наследия и языкового разнообразия в мире может получить дополнительные способы решения с помощью современных технологий, в частности активно осваиваемых возможностей искусственного интеллекта.

Цель статьи – анализ полученных к настоящему моменту данных о возможности сохранения языкового разнообразия с помощью искусственного интеллекта и предварительная оценка их эффективности с учетом тенденций к языковым трансформациям, влияющим на само развитие искусственного интеллекта.

Методология. Исследование носит обзорно-аналитический характер и направлено на изучение имеющихся в научной литературе данных о возможностях искусственного интеллекта в деле сохранения языкового разнообразия и существующих ограничениях в случае его применения к редким и исчезающим языкам.

Результаты. Поиск информации по заявленной проблеме показал, что обращение к технологиям искусственного интеллекта как инструменту, способному реализовать задачу сохранения языкового разнообразия, пока остается предметом дискуссии преимущественно в медийной и общественной сферах и крайне скупо представлено в академических научных кругах. В числе вызовов, не связанных напрямую с потенциальными возможностями искусственного интеллекта, упоминается языковая пристрастность многих систем самого искусственного интеллекта, которая коренится в данных используемых для их обучения, что может способствовать дальнейшему закреплению социального и языкового неравенства отдельных этнических групп и создавать сложности в построении образовательных моделей для обучения этим языкам с помощью искусственного интеллекта.

Авторами отмечается, что системы искусственного интеллекта могут быть обучены распознавать и анализировать лингвистические паттерны языков, которые были недостаточно изучены или находятся под угрозой исчезновения, в том числе строить образовательные модели на родном языке для малочисленных этнических групп, рискующих утратить язык общения.

Библиографическая ссылка: Ермолова Т. В., Савицкая Н. В., Дедова О. В., Гузова А. В. Возможности искусственного интеллекта в сохранении языкового разнообразия для будущих поколений // Science for Education Today. – 2024. – Т. 14, № 1. – С. 184–198. DOI: <http://dx.doi.org/10.15293/2658-6762.2401.09>

✉ Автор для корреспонденции: Александра Викторовна Гузова, sashenka_0879@mail.ru

© Т. В. Ермолова, Н. В. Савицкая, О. В. Дедова, А. В. Гузова, 2024

Появляются сведения о возможности искусственного интеллекта воссоздавать утраченные языки. Основной трудностью в использовании искусственного интеллекта для сохранения языкового разнообразия является нехватка данных по многим языкам, редко используемым в медиaprостранстве или находящимся под угрозой исчезновения.

Заключение. В заключении на основе анализа результатов исследования делается вывод о том, что технологии искусственного интеллекта признаются перспективными для сохранения языкового разнообразия.

Ключевые слова: искусственный интеллект; языковое разнообразие; сохранение языков; языковая предвзятость; языковое неравенство; образовательные программы; обработка естественного языка.

Постановка проблемы

Исчезновение живых языков на нашей планете происходит постоянно, и чаще всего эти процессы связывают с сокращением численности носителей этих языков, поглощением малых этносов более многочисленными и приспособленными к современным реалиям, отсутствием письменно закрепленных источников на исчезающем языке. По оценкам Организации Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО), почти половина из 6000 языков мира находится под угрозой исчезновения¹. Это неудивительно, учитывая, что в настоящее время 97 % населения земного шара говорит на 4 % языков. Даже в странах Европейского союза (ЕС), в которых языковому многообразию придается определенное значение, насчитывается 128 языков, находящихся под угрозой исчезновения. Рабочими в ЕС официально признаны 24 языка, еще 60 региональных языков признаны языками коренных народов, пять из

которых являются полуофициальными: каталонский, галисийский, баскский, гэльский и валлийский. Остальные языки находятся под угрозой исчезновения, не имея официального статуса в ЕС, например, виламовский язык в Польше или трансильванский саксонский в Румынии².

Любой утраченный язык становится невосполнимой потерей для культурного разнообразия и лишает мир уникальных знаний, заложенных в нем. В научных и общественных кругах единодушно признается, что сохранение языкового разнообразия крайне важно не только для носителей той или иной языковой группы, но и для выживания человечества как вида. В книге «Дело и слово»³ известный публицист Дмитрий Калужный рассматривает культуру и язык (как ее неотъемлемую часть) в качестве основной технологии выживания человечества. «В нашей человеческой многообразности – залог нашей устойчивости: при любых катаклизмах и перемене внешних условий всегда найдется культура,

¹ Российский комитет программы ЮНЕСКО «Информация для всех». Усилия по сохранению многоязычия, предпринимаемые на международном уровне.

URL:<http://www.ifapcom.ru/projects/multiling1/576/>

² Европейская хартия региональных языков или языковых меньшинств. Страсбург, 5 ноября 1992 г. (Издано на русском и английском языках) // URL:

http://www.coe.int/t/dg4/education/minlang/Default_en.asp

³ Калужный Д., Ермилова Е. Дело и Слово. Будущее России с точки зрения теории эволюции. Издательство: Эксмо Язык: Русский. – 2003. – 448 с. URL: <https://www.livelib.ru/author/13946/latest-dmitrij-kaluzhnyj>

которая обеспечит выживаемость популяции в целом. Сохранение национальных культур – основа совместного выживания»⁴.

Однако в современном мире диапазон возможностей для взаимного обогащения культур сужается до такой степени, что языки и культуры вымирают⁵, и, как следствие, утрачиваются многие свидетельства интеллектуальных достижений человека как представителя той или иной культуры, а также способов его мышления и бытия⁶.

В определенных условиях (политические противостояния, конфликты, военные действия) языки и их носители намеренно подвергаются дискриминации: осуществляется запрет на использование родного языка, сам язык объявляется «языком вражды» (hatespeech), его использование становится противозаконным [16].

Свой вклад в этот процесс вносят и происходящие в мире глобальные изменения, поделившие языки на «доминирующие» (широко используемые в быстро цифровизирующемся мире) и «непопулярные» (редко используемые в информационном пространстве) [2; 5]. Ученые называют это явление Language Bias (языковой предвзятостью). Например, в такой представительной организации, как ООН, в состав которой входят 193 страны, официальных языков общения только шесть: английский, арабский, испанский, китайский, русский и французский [9].

Изобретение искусственного интеллекта (ИИ) и наглядные успехи в его способности к обработке естественного языка (natural language processing, NLP), по мнению исследователей, могли бы решить проблему сохранения языкового многообразия и обеспечить инновационные решения для сохранения и возрождения языков, находящихся под угрозой исчезновения [1; 4].

Однако системы ИИ часто сами подвергаются критике за усиление языковой предвзятости, поскольку модели ИИ обучаются на больших объемах текстовых данных, которые, как правило, поступают из интернета. Эти данные охватывают широкий спектр контента, включая новостные статьи, веб-сайты, книги и посты в социальных сетях. Сам контент, доступный в интернете, не всегда является справедливым представлением всех языков, культур или точек зрения⁷ [7].

Пока не ясно, как будет развиваться информационное общество в дальнейшем. С точностью можно сказать только одно: в мировую экономику постепенно вовлекаются страны, ранее в ней не присутствовавшие, и представители лидирующих в ней стран сталкиваются с растущим разнообразием языков и их носителей. Можно отреагировать на это стандартизацией языков, как это делалось в последние годы, а можно попытаться решить эту проблему с помощью технологических инноваций, например, используя потенциал ИИ.

⁴ Калужный Д., Ермилова Е. Дело и Слово. Будущее России с точки зрения теории эволюции. Издательство: Эксмо Язык: Русский. – 2003. – 448 с. URL: <https://www.livelib.ru/author/13946/latest-dmitrij-kaluzhnyj>

⁵ Frackiewicz M. The Role of AI in Uncovering Lost Languages and Civilizations, 2023. URL: <https://ts2.pl/en/the-role-of-ai-in-uncovering-lost-languages-and-civilizations/#gsc.tab=0>

⁶ Roles for artificial intelligence in education. – 2016. URL:

<https://www.teachthought.com/the-future-of-learning/roles-for-artificial-intelligence-in-education/>

⁷ Языковое и культурное разнообразие в киберпространстве. Сборник материалов II Международной конференции (Якутск, 12–14 июля 2011 г.) / Сост. Кузьмин Е. И., Паршакова А. В. – М.: Межрегиональный центр библиотечного сотрудничества, 2013. – 384 с., 16 л. ил. URL: https://in-hub.cnews.ru/articles/2022-08-18_iskusstvennyj_intellekt_pomog_sohranit

Целью статьи является анализ полученных к настоящему моменту данных о возможности сохранения языкового разнообразия с помощью ИИ и предварительная оценка их эффективности с учетом тенденций к языковым трансформациям, влияющим на саморазвитие ИИ.

Методология исследования

Исследование носит обзорный характер и направлено на анализ существующих в отечественной и зарубежной литературе данных о возможностях ИИ в деле сохранения языкового разнообразия и существующих ограничениях в случае его применения к редким и исчезающим языкам.

Информационной базой явились опубликованные результаты лингвистических исследований, первичные источники официальных документов, публикации в СМИ, касающиеся развития ИИ и сохранения языкового разнообразия.

При анализе данных был использован системный метод, позволивший выявить взаимосвязь развития ИИ и современных языковых трансформаций, а также их синергетическое воздействие на тренды межъязыковой коммуникации.

Результаты исследования

Поиск информации по заявленной проблеме показал, что необходимость поддерживать языковое разнообразие и обращение к технологиям ИИ как инструменту, способному реализовать эту задачу, пока остается предметом дискуссии преимущественно в медийной и общественной сферах и крайне скупо представлено в академических научных кругах. Однако картина постепенно меняется, о чем, в частности, свидетельствуют дебаты на научной платформе ResearchGate, в которых ученые опытным путем пытаются определить

наиболее многообещающие приложения, способные сохранить языковое разнообразие для будущих поколений.

По имеющимся сведениям, одним из наиболее многообещающих применений ИИ является машинное обучение, подмножество ИИ, которое включает в себя обучение компьютеров на основе загруженных данных и возможность получения решения без дополнительного программирования. Алгоритмы машинного обучения могут научиться распознавать и анализировать лингвистические паттерны, помогая исследователям документировать и понимать языки, которые были плохо изучены или находятся под угрозой исчезновения.

Например, проект Google Euphonia использует ИИ для улучшения распознавания речи у людей с речевыми нарушениями. Данная технология может быть оптимизирована и обучена понимать различные речевые модели, акценты и диалекты, необходимые при коммуникации с представителями «непопулярных» и редких языков.

Еще одним значительным вкладом ИИ является область обработки естественного языка (NLP) – отрасль ИИ, которая делает акцент на взаимодействии между компьютерами и человеческим языком. NLP может быть использована для создания языковых конструкций, которые представляют собой математические модели, «описывающие» грамматику и лексику языка, редко используемого в онлайн-коммуникации. Эти модели можно также использовать для создания приложений, необходимых в процессе изучения иностранных языков, автоматических переводчиков и других инструментов, которые делают языки, находя-

щиеся под угрозой исчезновения, более доступными для учащихся и носителей языка⁸ [15].

Кроме того, ИИ может также помочь в создании цифровых языковых архивов. Эти архивы служат хранилищами лингвистических данных, включая аудиозаписи, транскрипции и переводы. Искусственный интеллект способен автоматизировать процесс расшифровки и перевода этих ресурсов, сделав их более доступными для исследователей, носителей языка и общественности⁹. В качестве примера можно привести идею создания Ауана (Айана) – голосового переводчика с русского на эвенкийский язык на базе ИИ. Его целью является максимальная архивация лексических единиц эвенкийского диалекта для сохранения национального языка эвенков – одного из малочисленных народов России.

Однако использование ИИ для сохранения языков сопряжено с определенными трудностями. Одним из главных препятствий является нехватка данных по многим языкам, находящимся под угрозой исчезновения. Алгоритмы машинного обучения и NLP требуют больших объемов данных для эффективной работы, и эти данные часто недоступны для языков с небольшим количеством носителей. Кроме того, необходимо тщательно учитывать этические последствия использования ИИ для документирования и сохранения языков, особенно когда речь идет об уважении прав коренных общин [1].

В то же время все чаще поднимается вопрос о том, какие вызовы могут стоять перед учеными при решении проблемы языкового

разнообразия с помощью ИИ. В числе этих вызовов, не связанных напрямую с потенциальными возможностями ИИ, упоминается языковая предвзятость (Language Bias) многих систем самого ИИ, которая коренится в данных, используемых для их обучения, что может способствовать закреплению социального неравенства отдельных этнических групп [17].

Модели ИИ обучаются на огромных объемах текстовых данных, поступающих из интернета. Основная часть онлайн-контента представлена на английском или нескольких доминирующих языках. Непопулярные языки или диалекты представлены в сети недостаточно. Модели ИИ, обученные на этих данных, лучше понимают и обрабатывают информацию на этих доминирующих языках и могут испытывать трудности с менее представленными языками. Это создает предвзятое отношение к определенным языкам и культурам, что может ограничить полезность ИИ для людей с разным языковым происхождением.

Данные, используемые для обучения моделей ИИ, часто содержат предубеждения и стереотипы, которые распространены в обществе. Например, гендерные, расовые и культурные предрассудки можно найти в языке, используемом в новостных статьях, книгах и постах в социальных сетях. Обученные на основе этих предвзятых данных модели ИИ воспроизводят их в своих выходных данных, что приводит к закреплению этих стереотипов системами ИИ и дискриминации определенных групп населения. Например, исследование, проведенное в 2020 г. Лабораторией этики

⁸ Artificial Intelligence and the Preservation of Linguistic Diversity by Marcin Fraćkiewicz (DA, *NN, *RS, Artificial intelligence, Artificial intelligence, News on 5 September 2023) URL: <https://ts2.space/en/artificial-intelligence-and-the-preservation-of-linguistic-diversity/>

⁹ KoigiBob, Embracing artificial intelligence to preserve dying languages (News on 17 November, 2020). URL: <https://www.fairplanet.org/story/embracing-artificial-intelligence-to-preserve-dying-languages/>

ИИ, показало, что языковые модели часто увековечивают гендерные и расовые стереотипы, что приводит к предвзятым результатам и ограничивает эффективность тестирования уровня владения языком с помощью ИИ [8].

Для преодоления языковой предвзятости и поощрения языкового разнообразия разработчики и исследователи ИИ выдвигают ряд необходимых инициатив.

Диверсификация обучающих данных: использование более разнообразных и репрезентативных наборов данных, включая текстовые данные из различных языковых слоев и культурных контекстов

Устранение систематических ошибок: более тонкая настройка моделей ИИ с использованием внешних входных данных от рецензентов, учет их рекомендаций по преодолению предвзятости.

Сотрудничество с экспертами: взаимодействие с лингвистами, антропологами и культурологами, чтобы лучше понимать нюансы различных языков, диалектов и культур и целенаправленно разрабатывать системы ИИ, учитывающие потребности различных сообществ.

Инвестиции в языки с ограниченными ресурсами: инвестирование в исследование и разработку языков, носители которых ограничены в финансах и не могут обеспечить надлежащую представленность в соцсетях и, как следствие, в системах ИИ.

Поощрение многоязычных ИИ-систем: разработка моделей ИИ, способных обрабатывать и понимать несколько языков одновременно, что способствует развитию языкового разнообразия.

Создание этических и справедливых моделей ИИ: привлечение внимания к этическим

аспектам систем ИИ и соблюдение принципов справедливости, подотчетности и прозрачности, что позволяет свести к минимуму языковые предпочтения и создать системы, уважающие языковое разнообразие.

Инициативы с открытым исходным кодом: выпуск моделей ИИ в виде ресурсов с открытым исходным кодом, что позволяет исследователям и разработчикам по всему миру вносить свой вклад в борьбу с языковой предвзятостью.

В настоящее время оптимальными в преодолении языковой дискриминации и сохранении языкового разнообразия считаются технологии OpenAI, ChatGPT и GPT-4.

GPT-3 от OpenAI считается передовой моделью ИИ в вопросе сохранения многоязычия при онлайн-коммуникации. Она может понимать и генерировать текст на нескольких языках, включая английский, испанский, французский, немецкий, итальянский, голландский и др. Эта многоязычная поддержка позволяет пользователям из разных языковых групп взаимодействовать с сервисами и ресурсами на своих родных языках и способствует языковому разнообразию за счет снижения языковых барьеров.

Преодолению языковых барьеров способствовали и услуги перевода на основе ИИ. Такие сервисы, как Google Translate и DeepL, позволяют пользователям переводить тексты, понимать необходимый контент на иностранных языках и общаться онлайн. Это во многом способствовало межкультурному общению, расширению ареала активно используемых в сети языков и преодолению языковых барьеров¹⁰.

В сфере образования широко используются такие платформы ИИ, как Duolingo и

¹⁰ Semantris от Google: как ИИ помогает учить английский играючи // Сообщество IT-специалистов Хабр [официальный сайт]. URL:

<https://habr.com/ru/company/english-dom/blog/520618/>

Rosetta Stone, которые особенно популярны при изучении иностранных языков. Эти и другие платформы значительно упрощают процесс обучения второму и третьему языку людей разных возрастных категорий и социальных слоев и легко подстраиваются к актуальному уровню языковых компетенций учащихся. Предлагая курсы на достаточно широком спектре языков, эти модели способствуют языковому разнообразию и делают изучение языков более доступным для людей из разных слоев общества¹¹ [14].

Инструменты OpenAI для определения уровня владения языком также обеспечили более равный доступ к образованию и возможностям трудоустройства для тех, для кого английский язык не является родным. Кроме того, системы языкового обучения с помощью искусственного интеллекта, такие как iTalki и Preply, связывают изучающих язык с носителями языка, способствуя дальнейшему языковому разнообразию и межкультурному обмену.

В последнее время звучат опасения о полной замене преподавателя в образовательных учреждениях на цифровые устройства с доступом к ИИ. Однако и зарубежные, и отечественные исследователи единодушны во мнении, что с развитием искусственных нейронных сетей и цифровых технологий функциональные возможности педагога меняются, но не меняется его значимость [3; 10]. Педагог в учебном процессе продолжает исполнять роль наставника на основе своих профессиональных компетенций и опыта. Технологические программы формируются из блоков базы данных, хранения, поиска и преобразования информации, необходимой для реше-

ния задач [6]. Искусственный интеллект на основе мгновенного анализа большого числа возможных вариантов выбирает наиболее рациональный сценарий, но алгоритм отбора не предполагает креативного или интуитивно правильного действия, свойственного человеку, а также не может заменить живого общения преподавателя и учеников [12; 17].

Обсуждение. Заключение

Технологии ИИ признаются наиболее перспективными для сохранения языкового разнообразия. Системы ИИ могут быть обучены распознавать и анализировать лингвистические паттерны, помогать исследователям документировать и понимать языки, которые были недостаточно изучены или находятся под угрозой исчезновения. Появляются сведения о возможности ИИ воссоздавать утраченные языки. Основной трудностью в использовании ИИ для сохранения языкового разнообразия является нехватка данных по многим языкам, редко используемым в медиапространстве или находящимся под угрозой исчезновения. Алгоритмы машинного обучения и NLP (natural language processing – обработка естественного языка) требуют большого объема данных для эффективной работы, и эти данные часто недоступны для языков с небольшим количеством носителей, отсутствием оцифрованных письменных источников на данном языке и недостаточной активностью в социальных сетях носителей языка [11]. Путь к борьбе с языковой предвзятостью и поощрению языкового разнообразия на основе ИИ пока еще только намечается, но усилиями создаваемых инструментов, подобных OpenAI, особенно с ChatGPT и GPT-4, прокла-

¹¹ Школа английского языка Skyeng: Гугл в помощь: 7 сервисов от поисковика, которые помогают учить

английский. – 2023. URL: <https://skyeng.ru/magazine/gugl-v-pomosh-7-servisov-ot-poiskovika-kotorye-pomogayut-uchit-anglijskij/>

дываются пути к более инклюзивному и справедливому будущему, способствуют большему языковому разнообразию и борются с языковой предвзятостью в глобальном масштабе.

Создание образовательных платформ с использованием ИИ позволяет эффективно решать проблему обучения иностранным языкам людей разных возрастных категорий и социальных слоев, поскольку продуцируемые ИИ программы легко подстраиваются к актуальному уровню языковых компетенций учащихся.

Завершая данный обзор, можно сделать вывод о том, что инвестиции в исследования и разработку инклюзивных технологий на основе ИИ позволяют добиваться значительных успехов на пути к созданию более лингвистически разнообразного и инклюзивного мира. Благодаря постоянному совершенствованию моделей ИИ эти передовые технологии обладают потенциалом для преодоления языковых барьеров, решают образовательные задачи, способствуют глобальному взаимопониманию и обеспечению того, чтобы ни один язык не был забыт.

СПИСОК ЛИТЕРАТУРЫ

1. Белов С. Д., Зрелова Д. П., Зрелов П. В., Кореньков В. В. Обзор методов автоматической обработки текстов на естественном языке // Системный анализ в науке и образовании. – 2020. – № 3. – С. 8–22. DOI: <https://doi.org/10.37005/2071-9612-2020-3-8-22> URL: <https://elibrary.ru/item.asp?id=44288349>
2. Гойчета Л. И., Климова Д. А., Мазурова Д. А., Семенова А. Р., Позднякова А. А., Пономаренко А. С. Дискриминационная коммуникация или дискриминирующее речевое поведение. (обзор) // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. – 2023. – № 1. – С. 41–51. DOI: <https://doi.org/10.31249/ling/2023.01.02> URL: <https://elibrary.ru/item.asp?id=50082764>
3. Есионова Е. Ю. Искусственный интеллект как альтернативный ресурс для изучения иностранного языка // Гуманитарные и социальные науки. – 2019. – № 3. – С. 155–166. DOI: <https://doi.org/10.23683/2070-1403-2019-74-3-155-166> URL: <https://elibrary.ru/item.asp?id=39138327>
4. Кармова М. Р., Логина М. В. Роль искусственного интеллекта в вопросе сохранения миноритарных языков (на примере абазинского языка) // Мир науки, культуры, образования. – 2023. – № 4. – С. 251–254. DOI: <https://doi.org/10.24412/1991-5497-2023-4101-251-254> URL: <https://elibrary.ru/item.asp?id=54504954>
5. Опарина Е. О. Сосуществование языков в условиях глобализации // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. – 2020. – № 2. – С. 44–49. URL: <https://www.elibrary.ru/item.asp?id=42854769>
6. Berdasco A., López G., Diaz I., Quesada L., Guerrero L. A. User Experience Comparison of Intelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana // 13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019. – 2019. – Vol. 31 (1). – P. 51. DOI: <https://doi.org/10.3390/proceedings2019031051>
7. Beukeboom C. J. Mechanisms of linguistic bias: How words reflect and maintain stereotypic expectancies // Social Cognition and Communication / J. Laszlo, J. Forgas, & O. Vincze (Eds.). –



- New York, NY: Psychology Press., 2014. – P. 313–330. URL: <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203744628-22/mechanisms-linguistic-bias-camiel-beuke-boom>
8. Cowls J., Morley J. The 2020 Yearbook of the Digital Ethics Lab // Digital Ethics Lab Yearbook / J. Cowls & J. Morley (Eds.). – Springer International Publishing, 2021. DOI: <https://doi.org/10.1007/978-3-030-80083-3> URL: <https://philpapers.org/rec/COWTY>
 9. Gerrand P. Estimating Linguistic Diversity on the Internet: A Taxonomy to Avoid Pitfalls and Paradoxes // Journal of Computer-Mediated Communication. – 2007. – Vol. 12 (4). – P. 1298–1321. DOI: <https://doi.org/10.1111/j.1083-6101.2007.00374.x>
 10. Kim D., Ruecker D., Kim D.-J. Mobile Assisted Language Learning Experiences // International Journal of Mobile and Blended Learning. – 2017. – Vol. 9 (1). – P. 49–66. DOI: <https://doi.org/10.4018/IJMBL.2017010104>
 11. Lenci A., Padó S. Editorial: Perspectives for natural language processing between AI, linguistics and cognitive science // Frontiers in Artificial Intelligence. – 2022. – Vol. 5. – P. 1059998. DOI: <https://doi.org/10.3389/frai.2022.1059998> URL: https://mic.org.ru/phocadownload/yazyk-mpr_2011.pdf
 12. Murphy R. F. Artificial Intelligence Applications to Support K–12 Teachers and Teaching: A Review of Promising Applications, Challenges, and Risks // RAND Corporation. – 2019. DOI: <https://doi.org/10.7249/PE315>
 13. Reagan T. Why language endangerment and language death matter: “Took away our native tongue ... And taught their English to our young” // Linguistic Legitimacy and Social Justice. – 2019. – P. 285–314. DOI: https://doi.org/10.1007/978-3-030-10967-7_9
 14. Siu S. Ch. ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation // SRN Electronic Journal. – 2023. DOI: <http://dx.doi.org/10.2139/ssrn.4448091> URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4448091
 15. Weinstein J. Viewpoint discrimination, hate speech, and political legitimacy: A reply // Constitutional Commentary. – 2017. – Vol. 32. – P. 715. URL: <https://constitutionalcommentary.lib.umn.edu/article/viewpoint-discrimination-hate-speech-and-political-legitimacy-a-reply/>
 16. Yano T., Resnik P., Smith N. A. Shedding (a thousand points of) light on biased language // *Proceedings of the ACM on Human-Computer Interaction*. – 2010. – Vol. 6. – P. 152–158. DOI: <https://doi.org/10.5555/1866696.1866719> URL: <https://dl.acm.org/doi/10.5555/1866696.1866719>
 17. Zhang D. T., Mishra S., Brynjolfsson E., Etchemendy J., Ganguli D., Grosz B., Lyons T., Manyika J., Niebles J. C., Sellitto M., Shoham Y., Clark J., Perrault R. The AI Index 2021 Annual Report. – AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March, 2021. – P. 110. DOI: <https://doi.org/10.48550/arXiv.2103.06312>

Поступила: 02 декабря 2023

Принята: 10 января 2024

Опубликована: 29 февраля 2024



Заявленный вклад авторов:

Ермолова Татьяна Викторовна: сбор эмпирического материала, выполнение статистических процедур.

Савицкая Наталья Васильевна: сбор материалов, литературный обзор.

Дедова Ольга Витальевна: сбор материалов, оформление текста статьи.

Гузова Александра Викторовна: организация исследования, интерпретация результатов.

Все авторы ознакомились с результатами работы и одобрили окончательный вариант рукописи.

Информация о конфликте интересов:

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов в связи с публикацией данной статьи

Информация об авторах

Ермолова Татьяна Викторовна

кандидат психологических наук, профессор,
кафедра зарубежной и русской филологии,
Московский государственный психолого-педагогический университет,
ул. Сретенка, д. 29, 127051, г. Москва, Россия.
ORCID ID: <https://orcid.org/0000-0003-4260-9087>
E-mail: yermolova@mail.ru

Савицкая Наталья Васильевна

кандидат педагогических наук, доцент,
кафедра зарубежной и русской филологии,
Московский государственный психолого-педагогический университет,
ул. Сретенка, д. 29, 127051, г. Москва, Россия.
ORCID ID: <https://orcid.org/0000-0002-1769-5553>
E-mail: n.sawa@yandex.ru

Дедова Ольга Витальевна

старший преподаватель,
кафедра зарубежной и русской филологии,
Московский государственный психолого-педагогический университет,
ул. Сретенка, д. 29, 127051, г. Москва, Россия.
ORCID ID: <https://orcid.org/0000-0001-6321-4127>
E-mail: olgadedova2007@gmail.com



Гузова Александра Викторовна

кандидат педагогических наук, доцент,

кафедра зарубежной и русской филологии,

Московский государственный психолого-педагогический университет,



ул. Сретенка, д. 29, 127051, г. Москва, Россия.

ORCID ID: <https://orcid.org/0000-0001-8335-6528>

E-mail: sashenka_0879@mail.ru



The potential of artificial intelligence in preserving linguistic diversity for future generations

Tatiana V. Ermolova¹, Natal'ya V. Savitskaya¹, Ol'ga V. Dedova¹, Alexandra V. Guzova  ¹

¹ Moscow State University of Psychology and Education, Moscow, Russian Federation

Abstract

Introduction. *The significant and only partially solved problem of preserving cultural heritage and linguistic diversity in the world can obtain additional ways of solving by means of using modern technologies, in particular, the Artificial Intelligence (AI).*

The purpose of this article is to analyze up-to-date research data on the possibility of preserving linguistic diversity by means of Artificial Intelligence technologies and evaluate their effectiveness taking into account linguistic transformations determining the development of AI.

Materials and Methods. *This review is aimed at analyzing the data available in scholarly literature on the potential of artificial intelligence in preserving linguistic diversity and the existing limitations in the case of its application to rare and endangered languages.*

Results. *The search for data on the research problem revealed that using AI technologies as tools for implementing the task of preserving linguistic diversity is still the subject of discussion mainly in the media and public spheres and is extremely poorly represented in academic literature. Among the challenges partially related to the potential of AI is the linguistic bias of many AI systems, which is rooted in the data used to train them, which can further entrench social and linguistic inequalities among certain ethnic groups and create difficulties in building educational models for teaching these languages with the help of AI.*

The authors argue that AI systems can be trained to recognize and analyze linguistic patterns of languages that have been understudied or are endangered, including building educational models in native languages for small ethnic groups at risk of losing the language of communication. There is evidence that AI can recreate lost languages. A major challenge in using AI to preserve linguistic diversity is the lack of data on many languages that are rarely used in the media space or are endangered.

Conclusions. *Finally, the authors conclude that AI technologies are recognized as promising for preserving linguistic diversity.*

For citation

Ermolova T. V., Savitskaya N. V., Dedova O. V., Guzova A. V. The potential of artificial intelligence in preserving linguistic diversity for future generations. *Science for Education Today*, 2024, vol. 14 (1), pp. 184–198. DOI: <http://dx.doi.org/10.15293/2658-6762.2401.09>

  Corresponding Author: Alexandra V. Guzova, sashenka_0879@mail.ru

© Tatiana V. Ermolova, Natal'ya V. Savitskaya, Ol'ga V. Dedova, Alexandra V. Guzova, 2024

**Keywords**

Artificial intelligence; Linguistic diversity; Language preservation; Language bias; Language inequality; Educational programs; Natural language processing.

REFERENCES

1. Belov S. D., Zrelva D. P., Zrelv P. V., Koren`kov V. V. Overview of methods for automatic natural language text processing. *Sistemny`j analiz v nauke i obrazovanii*, 2020, no. 3, pp. 8–22. (In Russian) DOI: <https://doi.org/10.37005/2071-9612-2020-3-8-22> URL: <https://elibrary.ru/item.asp?id=44288349>
2. Goychita L. I., Klimova D. A., Mazurova D. A., Semenov A. R., Pozdnyakova A. A., Ponomarenko A. S. Discriminatory communication or discriminatory speech behavior (review). *Social and Humanitarian Sciences. Domestic and Foreign Literature. Series 6: Linguistics*, 2023, no. 1, pp. 41-51. (In Russian) DOI: <https://doi.org/10.31249/ling/2023.01.02> URL: <https://elibrary.ru/item.asp?id=50082764>
3. Esionova E. Yu. Digital intelligence as an alternative resource for foreign language learning. *Gumanitarny`e i social`ny`e nauki*, 2019, no. 3, pp. 155–166. (In Russian) DOI: <https://doi.org/10.23683/2070-1403-2019-74-3-155-166> URL: <https://elibrary.ru/item.asp?id=39138327>
4. Karmova M. R., Logina M. V. The role of artificial intelligence in the preservation of minority languages (on the example of Abazian language). *Mir nauki, kul`tury`, obrazovaniya*, 2023, no. 4, pp. 251-254. (In Russian) DOI: <https://doi.org/10.24412/1991-5497-2023-4101-251-254> URL: <https://elibrary.ru/item.asp?id=54504954>
5. Oparina E. O. The coexistence of languages in the context of globalization. *Social and humanitarian sciences. Domestic and foreign literature. Series 6: Linguistics*, 2020, no. 2, pp. 44–49. (In Russian) URL: <https://www.elibrary.ru/item.asp?id=42854769>
6. Berdasco A., López G., Diaz I., Quesada L., Guerrero L. A. User experience comparison of intelligent personal assistants: Alexa, google assistant, Siri and Cortana. *13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019*, 2019, vol. 31 (1), pp. 51. DOI: <https://doi.org/10.3390/proceedings2019031051>
7. Beukeboom C. J. Mechanisms of linguistic bias: How words reflect and maintain stereotypic expectancies. In: J. Laszlo, J. Forgas, & O. Vince (Eds.), *Social Cognition and Communication*. New York, NY: Psychology Press., 2013, pp. 313–330. URL: <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203744628-22/mechanisms-linguistic-bias-camiel-beukeboom>
8. Cows J., Morley J. The 2020 yearbook of the digital ethics lab. In: J. Cows & J. Morley (Eds.), *Digital Ethics Lab Yearbook*. Springer International Publishing, 2021. DOI: <https://doi.org/10.1007/978-3-030-80083-3> URL: <https://philpapers.org/rec/COWTY>
9. Gerrand P. Estimating linguistic diversity on the internet: A taxonomy to avoid pitfalls and paradoxes. *Journal of Computer-Mediated Communication*, 2007, vol. 12 (4), pp. 1298–1321. DOI: <https://doi.org/10.1111/j.1083-6101.2007.00374.x>
10. Kim D., Ruecker D., Kim D.-J. Mobile assisted language learning experiences. *International Journal of Mobile and Blended Learning*, 2017, vol. 9 (1), pp. 49–66. DOI: <https://doi.org/10.4018/IJMBL.2017010104>



11. Lenci A., Padó S. Editorial: Perspectives for natural language processing between AI, linguistics and cognitive science. *Frontiers in Artificial Intelligence*, 2022, vol. 5, pp. 1059998. DOI: <https://doi.org/10.3389/frai.2022.1059998> URL: https://mic.org.ru/phocadownload/yazyk-mpr_2011.pdf
12. Murphy R. F. Artificial intelligence applications to support K–12 teachers and teaching: A review of promising applications, challenges, and risks. *RAND Corporation*, 2019. DOI: <https://doi.org/10.7249/PE315>
13. Reagan T. Why language endangerment and language death matter: “Took away our native tongue ... And taught their English to our young”. *Linguistic Legitimacy and Social Justice*, 2019, pp. 285–314. DOI: https://doi.org/10.1007/978-3-030-10967-7_9
14. Siu S. Ch. ChatGPT and GPT-4 for professional translators: Exploring the potential of large language models in translation. *SRN Electronic Journal*, 2023. DOI: <http://dx.doi.org/10.2139/ssrn.4448091> URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4448091
15. Weinstein J. Viewpoint discrimination, hate speech, and political legitimacy: A reply. *Constitutional Commentary*, 2017, vol. 32, pp. 715. URL: <https://constitutionalcommentary.lib.umn.edu/article/viewpoint-discrimination-hate-speech-and-political-legitimacy-a-reply/>
16. Yano T., Resnik P., Smith N. A. Shedding (a thousand points of) light on biased language. *Proceedings of the ACM on Human-Computer Interaction*, 2010, vol. 6, pp. 152–158. DOI: <https://doi.org/10.5555/1866696.1866719> URL: <https://dl.acm.org/doi/10.5555/1866696.1866719>
17. Zhang D. T., Mishra S., Brynjolfsson E., Etchemendy J., Ganguli D., Grosz B., Lyons T., Manyika J., Niebles J. C., Sellitto M., Shoham Y., Clark J., Perrault R. *The AI Index 2021 Annual Report*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March, 2021, P. 110. DOI: <https://doi.org/10.48550/arXiv.2103.06312>

Submitted: 02 December 2023

Accepted: 10 January 2024

Published: 29 February 2024



This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (CC BY 4.0).

The authors' stated contribution:

Tatiana Victorovna Ermolova

Contribution of the co-author: collection of empirical material, implementation of statistical procedures.

Natal'ya Vasil'evna Savitskaya

Contribution of the co-author: literary review.

Ol'ga Vital'evna Dedova

Contribution of the co-author: design of the text of the article.

Alexandra Victorovna Guzova

Contribution of the co-author: organization of the study, interpretation of the results

All authors reviewed the results of the work and approved the final version of the manuscript.





Information about competitive interests:

The authors declare no apparent or potential conflicts of interest in connection with the publication of this article

Information about the Authors

Tatiana Victorovna Ermolova

PhD in Psychology, Professor,
Moscow State University of Psychology and Education,
Sretenka st., 29, 127051, Moscow, Russian Federation.
ORCID: <https://orcid.org/0000-0003-4260-9087>
E-mail: yermolova@mail.ru

Natal'ya Vasil'evna Savitskaya

PhD in Pedagogy, Associate Professor,
Department of Foreign and Russian Philology,
Moscow State University of Psychology and Education,
Sretenka st., 29, 127051, Moscow, Russian Federation.
ORCID ID: <https://orcid.org/0000-0002-1769-5553>
E-mail: n.sawa@yandex.ru

Ol'ga Vital'evna Dedova

Senior Educator,
Department of Foreign and Russian Philology,
Moscow State University of Psychology and Education,
Sretenka st., 29, 127051, Moscow, Russian Federation.
ORCID ID: <https://orcid.org/0000-0001-6321-4127>
E-mail: olgadedova2007@gmail.com

Alexandra Victorovna Guzova

PhD in Pedagogy, Associate Professor,
Department of Foreign and Russian Philology,
Moscow State University of Psychology and Education,
Sretenka st., 29, 127051, Moscow, Russian Federation.
ORCID ID: <https://orcid.org/0000-0001-8335-6528>
E-mail: sashenka_0879@mail.ru