

الدراسات العربية والأجنبية

АРАБИСТИКА ЕВРАЗИИ EURASIAN ARABIC STUDIES

ЕВРАЗИЯ АРАБИСТИКАСЫ

Том 7, № 1 / Vol. 7, No. 1

2024



ISSN 2619-1261

Казанский (Приволжский) федеральный университет
Египетско-российский фонд культуры и науки
Центр арабской культуры «Аль-Хадара»
Центр исследований и научного взаимодействия «Ауалим»

АРАБИСТИКА ЕВРАЗИИ

Научный журнал

2024, т. 7, № 1 ISSN 2619-1261

DOI: 10.26907/2619-1261

Издаётся с августа 2018 года – 4 выпуска в год
Размещается на платформах РУНЭБ (elibrary.ru),
БД «КиберЛенинка» и Ulrich`sPeriodicalsDirectory.

Рубрики и специальности журнала

Филологические науки:

5.9.2. Литературы народов мира
5.9.6. Языки народов зарубежных стран
5.9.8. Теоретическая, прикладная
и сравнительно-сопоставительная лингвистика

Педагогические науки:

5.8.1. Общая педагогика, история педагогики
и образования
5.8.2. Теория и методика обучения и
воспитания (по областям и уровням
образования)
5.8.7. Методология и технология
профессионального образования

Исторические науки:

5.6.2. Всеобщая история
5.6.3. Археология
5.6.7. История международных отношений
и внешней политики

Теология (по исследовательскому направлению ислам):

5.11.1. Теоретическая теология
5.11.2. Историческая теология
5.11.3. Практическая теология

Журнал включен в перечень ВАК по специальностям
«Филологические науки» (5.9.2., 5.9.6., 5.9.8.) и
«Исторические науки» (5.6.2., 5.6.7.)

Адрес редакции:

420008, Россия, Республика Татарстан, г. Казань, ул. М.
Межлаука, 3, каб. 117
Тел.: +7 (843) 221-33-21
E-mail: arabicstudies@mail.ru
Website: <http://eas20188.org>

Учредители: Хайрутдинов Рамиль Равилович, Мингазова Наиля Габделхамитовна

Зарегистрирован Федеральной службой по надзору в
сфере связи, информационных технологий и массовых
коммуникаций. Свидетельство о регистрации печатного
СМИ (журнал) ПИ № ФС 77-73332 от 24.07.2018.

Подписано в печать 26.03.2024. Бумага офсетная. Печать
цифровая. Заказ № 153/3

Формат 60x84/8. Тираж 500 экз. Отпечатано с готового
оригинал-макета, подготовленного в типографии
Издательства Казанского университета.

Дата выхода выпуска в свет: 30.03.2024.

Издатель: АНО «Институт культурного наследия»

Адрес: 420111, Россия, г. Казань, ул. Кремлёвская, 10/15.
Тел.: +7 (843) 221-33-21

Адреса типографий:

Издательство Казанского университета
420008, г. Казань, ул. Профессора Нужина, 1/37.

Тел.: (843) 233-73-59, 233-73-28.

Египетско-российский фонд культуры и науки (доп.
тираж)

11769 Nozha, 114 Joseph Tito St., Heliopolis – Cairo – Egypt
Tel.: + (202) 219 271 57 & 58

Fax: + (202) 219 271 50

www.russiannewsar.com secertary_ert@yahoo.com

Подписка и распространение: каталог «Газеты
и журналы – 2023, 2 полугодие» ГК «Урал-Пресс», 64983

Редакционно-издательская группа

Ответственный секретарь

Мингазова А.Р. (г. Казань, Россия)

Научный редактор

Субич В.Г., канд. филол. наук (г. Казань, Россия)

Оператор сайта

Витоль Е.В. (г. Казань, Россия)

Ответственный за выпуск

Мингазова Н.Г. (г. Казань, Россия)

Корректор

Мингазова А.Р. (г. Казань, Россия)

Учредитель, шеф-редактор

Хайрутдинов Рамиль Равилович – канд. ист. наук,
доцент, КФУ, г. Казань, Россия.

Учредитель, главный редактор

Мингазова Наиля Габделхамитовна – канд. филол. наук,
доцент, КФУ, г. Казань, Россия.

Заместитель главного редактора

Закиров Рафис Рафаелевич – канд. филол. наук, доцент,
РИИ, г. Казань, Россия.

Заместитель главного редактора

Аль-Аммари Мохаммед Салех – канд. пед. наук, КФУ,
г. Казань, Россия.

Председатель редакционной коллегии

Аль-Фоади Рахим Али –Ph.D., профессор, Багдадский
университет, г. Багдад, Ирак.

Сопредседатели редакционной коллегии

Аль-Джарф Рейма Саадо –Ph.D., профессор, г. Рияд,
Саудовская Аравия.

Берникова Ольга Александровна – канд. филол. наук, доцент,
СПбГУ, г. Санкт-Петербург, Россия.

Ибрагимов Ибрагим Джавпарович – канд. пед. наук,
доцент, ПГУ, г. Пятигорск, Россия.

Кириллина Светлана Алексеевна – д-р ист. наук,
профессор, ИСАА, г. Москва, Россия.

Редакционная коллегия

Алексеев Игорь Леонидович – канд. ист. наук, г. Москва,
Россия.

Аликберов Аликбер Калабекович – д-р. ист. наук, ИВ
РАН, г. Москва, Россия.

Алмазова Лейла Ильдусовна – канд. филос.наук, КФУ,
г. Казань, Россия.

Аль-Мусави Саид –Ph.D., доцент, г. Тетуан, Марокко.

Андреасян Арусяк Каляевна – канд. филол. наук,
г. Ереван, Армения.

Бельхамити Мохаммед Эль Милуд – канд. филол. наук,
г. Оран, Алжир.

Большаков Андрей Олегович – д-р ист. наук, профессор,
СПбГУ, г. Санкт-Петербург, Россия.

Газизова Альфия Ильдусовна – д-р пед. наук, профессор,
КФУ, г. Казань, Россия.

Филологические науки

Научная статья

УДК 821.411.21

DOI: 10.26907/2619-1261.2024.7.1.31-50

**Искусственный интеллект и анализ настроений на арабском языке:
поиск эффективных решений**

Ольга Александровна Берникова¹, Наталья Александровна Кижяева²

^{1,2}Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

¹o.bernikova@spbu.ru, <https://orcid.org/0000-0001-8397-3171>

²natalia.kizhaeva@gmail.com, <https://orcid.org/0000-0003-4343-9921>

Аннотация. В статье рассматривается использование искусственного интеллекта для анализа настроений в текстах на арабском языке. Сентимент-анализ является одним из важнейших направлений обработки естественного языка (NLP) в силу его значимости для эффективного ведения бизнеса, принятия решений в области государственного управления, открытия новых горизонтов научного поиска. Целью данной работы является сравнительное изучение инструментов для сентимент-анализа арабского языка на основе мониторинга текущего состояния технологий в соответствующей области и апробации качества их функционирования. Выборка текстов для проведения исследования учитывала вариативность арабского языка и охватывала образцы доисламской поэзии, тексты на классическом языке, египетском диалекте, языке интернет-коммуникаций. Особое внимание уделяется использованию для решения задач сентимент-анализа ChatGPT – модели, создавшей новые вызовы для современного научно-образовательного пространства. Арабский язык имеет ряд особенностей, затрудняющих применение технологий сентимент-анализа в силу его сложной морфологической структуры, богатой семантики и разнообразия диалектов. Использование формального и неформального языка, наличие сарказма и скрытых настроений затрудняют задачу точного определения тональности текста. Проведенный сравнительный анализ инструментов анализа настроений на арабском языке раскрывает текущий ландшафт доступных решений, их сильные стороны и ограничения. Хотя в разработке такого рода

технологий для арабского языка был достигнут значительный прогресс, есть возможности для их совершенствования в дальнейшем.

Ключевые слова: арабский язык, анализ настроений, искусственный интеллект, ChatGPT, большие языковые модели

Благодарности: Исследование выполнено за счет гранта Российского научного фонда № 22-28-01046, <https://rscf.ru/project/22-28-01046/>.

Для цитирования: Берникова О. А., Кижяева Н. А. Искусственный интеллект и анализ настроений на арабском языке: поиск эффективных решений // Арабистика Евразии. 2024. т. 7 № 1. С. 31–50. DOI: 10.26907/2619-1261.2024.7.1.31-50.

Philology

Original article

Artificial Intelligence and Sentiment Analysis in Arabic: Search for Effective Solutions

Olga A. Bernikova¹, Natalia A. Kizhaeva²

^{1, 2}St. Petersburg State University, St. Petersburg, Russia

¹o.bernikova@spbu.ru, <https://orcid.org/0000-0001-8397-3171>

²natalia.kizhaeva@gmail.com, <https://orcid.org/0000-0003-4343-9921>

Abstract. The paper discusses the use of Artificial Intelligence for sentiment analysis of texts in Arabic. Sentiment analysis is one of the most important areas of Natural Language Processing (NLP) due to its importance for effective business, decision-making in the field of public administration, and opening new horizons of scientific research. The purpose of this work is a comparative study of tools for sentiment analysis of the Arabic language based on monitoring the current state of technology in the relevant field and testing the quality of their functioning. The selection of texts for the study took into account the variability of the Arabic language and covered samples of pre-Islamic poetry, texts in the Classical language, the Egyptian dialect, and the language of Internet communications. Particular attention is paid to the use of ChatGPT to solve problems of sentiment analysis – a model that has created new challenges for the modern research and educational space. The Arabic language has a number of features that make it difficult to implement sentiment analysis technologies due to its complex morphological structure, rich semantics and variety of dialects. The use of formal and informal language, the

presence of sarcasm and hidden sentiments make it difficult to accurately determine the tone of a text. The conducted comparative analysis of Arabic sentiment analysis tools reveals the current landscape of available solutions, their strengths and limitations. Although significant progress has been made in developing this type of technology for Arabic, there is room for further improvement.

Keywords: Arabic, sentiment analysis, artificial intelligence, ChatGPT, large language models

Acknowledgments: The research was supported by Russian Science Foundation (project no. 22-28-01046, <https://rscf.ru/project/22-28-01046/>).

For citation: Bernikova O. A., Kizhaeva N. A. Artificial intelligence and sentiment analysis in Arabic: Search for Effective Solutions. *Eurasian Arabic Studies*. 2024;7(1):31-50. (In Russ.). DOI: 10.26907/2619-1261.2024.7.1.31-50.

Введение

«Искусственный интеллект» (Artificial Intelligence; далее ИИ) стал термином 2023 года по версии британского словаря Collins и сегодня рассматривается в качестве следующей великой технологической революции. Как отмечено на сайте издания, ИИ – термин, который описывает моделирование ментальных функций человека с помощью компьютерных программ [1].

История искусственного интеллекта восходит к середине прошлого века. В 1956 г. состоялась Дартмутская конференция, которая ознаменовала рождение этой области, наряду с введением в научный оборот понятия «Искусственный интеллект» [2].

Одной из важнейших областей ИИ является обработка естественного языка (англ. Natural Language Processing, NLP), которая направлена на распознавание, обработку и генерацию устной и письменной речи. Во многом благодаря созданию и распространению больших языковых моделей (Large Language Models - LLM), в первую очередь ChatGPT, ИИ получил новое звучание в 2022-2023 годах.

В этой связи сформировались новые вызовы для процессинга арабского языка как одного из наиболее распространенных идиомов в киберпространстве, требующего разработки специальных решений с учетом вариативности его существования во времени и пространстве. Несмотря на осуществлённый за последнее время прогресс в области процессинга арабского текста, создание новых технологий предполагает мониторинг их эффективности для решения различных задач теоретического и практического характера. Применение ИИ для обработки арабского языка важно для

обеспечения прогресса в понимании естественного языка, анализе настроений, машинном переводе, технологиях чат-ботов и распознавания речи.

В данной статье осуществляется сопоставительный анализ инструментов для анализа настроений (Sentiment Analysis) на арабском языке в контексте появления ChatGPT. Рассматриваемые решения описываются не только с точки зрения корректности получаемых результатов, но и удобства их применения непрофессиональными пользователями. Выборка материалов для проведения исследования опиралась на необходимость тестирования различных вариантов арабского языка (арабского литературно языка, классического языка, диалектов, языка интернет-коммуникации в социальных медиа). Сентимент-анализ арабского языка в исторической перспективе способствует лучшему пониманию меняющихся культурных и эмоциональных тенденций. С практической точки зрения сентимент-анализ эффективен для оценки общественного мнения, что помогает государственным органам при принятии тех или иных решений. Неоценимую помощь оказывает данный инструмент и при ведении бизнеса – для получения отзывов клиентов о товарах и услугах. Сентимент-анализ открывает новые горизонты для проведения научных исследований – от социологии и лингвистики до политологии и истории.

Научно-практическая значимость работы определяется его результатами, которые предоставляют потенциальному исследователю информацию о наиболее эффективных решениях для сентимент-анализа арабского текста и специфики их применения, а также фиксирует особенности процессинга арабского языка и эффективность соответствующих решений на данном этапе цифрового развития общества.

Подходы к решению задачи сентимент-анализа

Сентимент-анализ изучает эмоциональную окраску текстов и является частью более широкого круга задач обработки естественного языка и анализа данных. Существуют различные подходы к сентимент-анализу, которые можно классифицировать по типу алгоритма, используемого для анализа данных:

1. Классификация на основе правил и лексикона: этот подход основан на использовании заранее определенных правил и набора слов, каждое из которых ассоциировано с числом – значением тональности.

2. Машинное обучение: этот подход использует алгоритмы машинного обучения с учителем, в частности алгоритмы, основанные на деревьях решений,

линейных классификаторах, классификаторах на основе правил и вероятностных классификаторах [3], [4], [5].

3. Глубокое обучение: этот подход использует нейронные сети, такие как рекуррентные нейронные сети, сверточные нейронные сети или трансформеры. С выходом новых архитектур ведется их адаптация для арабского языка [6], [7], [8].

4. Гибридный подход: этот подход комбинирует различные методы (такие как машинное обучение и глубокое обучение) для достижения более высокой точности и устойчивости к шуму.

Реализации алгоритмов сентимент-анализа текстов на арабском языке

В рамках данного исследования для тестирования решений, направленных на анализ настроений в текстах на арабском языке, выбраны три инструмента: *SAMeL Tools*, *Mazajak* и *MonkeyLab*. Выборка обусловлена удобством использования веб-интерфейсов двух последних систем и открытостью исходного кода пакета программ *SAMeL Tools*.

Новые языковые модели на основе нейросетей зачастую доступны для загрузки на ресурсах [GitHub.io](https://github.com) и huggingface.com, что дает исследователям возможность использовать их в своих работах, а также адаптировать под различные языки и задачи. Выкладываемые модели называются предобученными, а процесс их адаптации «тонкой настройкой» (англ. *fine-tuning*).

Такой подход к разработке (использование предобученной модели и тонкая настройка) был реализован в пакете программ *SAMeL Tools* [9]. В качестве предобученных моделей были взяты модель архитектуры трансформер *mBERT* [10] и ее вариация для арабского языка *AraBERT* [6]. Для обеспечения правильной работы модели с текстами на диалектах арабского языка, авторы использовали разнообразные датасеты для дообучения и оценки. Предобработка обучающих данных (удаление диакритик, ссылок, текстов на английском языке) проводилась средствами, также предоставляемыми пакетом программ *SAMeL Tools*. Далее осуществлялась настройка на задачу классификации на три класса: положительный, негативный и нейтральный. В зависимости от варианта арабского языка были получены три модели: для классического арабского, современного арабского литературного языка и диалектов. Все они доступны для загрузки на упомянутых выше ресурсах. Чтобы воспользоваться алгоритмом сентимент-анализа в *SAMeL Tools*, необходимо обладать навыками программирования на языке *Python*. У пакета программ есть подробная документация [11] и активное сообщество разработчиков на сайте [GitHub.io](https://github.com).

Mazajak [12], онлайн-система для сентимент-анализа на арабском языке, использует две последовательные глубокие нейронные сети: сверточную нейронную сеть (CNN) и нейронную сеть архитектуры «долгая краткосрочная память» (LSTM). Предобработка обучающих данных проходила в три этапа: нормализация букв, удаление повторяющихся букв и очистка (удаление знаков пунктуации, диакритики). Для использования алгоритмов машинного обучения текст на арабском языке представляется в числовых векторов (так называемых эмбеддингов). В Mazajak для получения эмбеддингов была выбрана модель word2vec [13]. Обучающие данные для модели включают 250 млн постов в социальных сетях на арабском языке, собранных в период с 2013 по 2016 год. Такой длительный период и объем позволили учесть различные темы текстов и различные диалекты арабского. Первая нейронная сеть, CNN, находит паттерны в данных и извлекает признаки из предложений, в то время как вторая нейронная сеть, LSTM, работает с извлеченными признаками, учитывая контекст и порядок слов. На последнем слое нейронной сети происходит классификация слова на три класса: положительный, нейтральный, негативный. Mazajak доступна пользователю в виде веб-интерфейса [14].

Сравнение алгоритмов CAMeL Tools и Mazajak было проведено в работе [9] на трех датасетах SemEval [15], ASTD [16] и ArSAS [17]. Дообученная модель AraBERT от CAMeL Tools показала лучшее значение F-меры на всех трех датасетах (прирост 1% на ArSAS, 2% на ASTD и 7% на SemEval).

Одним из оказавшихся эффективным инструментов для сентимент-анализа арабского текста является MonkeyLearn [18]. К сожалению, создатели лишь описывают возможности применения сентимент-анализа (как и других задач обработки языка), но не дают каких-либо технических деталей. Вместе с тем, мы решили включить данный инструмент в данное исследование в силу простоты его использования и степени корректности полученных результатов, что продемонстрировали результаты описанного ниже эксперимента.

Апробация инструментов сентимент-анализа арабского текста

Для тестирования эффективности и корректности работы инструментов для сентимент-анализа арабского языка были отобраны фрагменты текстов с учетом темпоральных и стилистических характеристик. Феномен понятия «арабский язык» объединяет все варианты его существования в синхронии и диахронии: классический арабский язык (язык Корана, доисламской поэзии, средневековых источников и т. д.), различные диалекты арабского языка, современный литературный (стандартный) арабский язык, язык электронной коммуникации

(компьютерно-опосредованной коммуникации в социальных СМИ) и т. д. Соответственно для объективной оценки работы инструментов для сентимент-анализа требуется тестирование корректности процессинга текстов, охватывающих все перечисленные варианты языка.

Выборка текстов отличалась по тематике для объективной оценки стилистических уровней. Материал охватывал как фрагменты художественных произведений (в том числе «1000 и 1 ночь», образцы доисламской поэзии), так и классических текстов ислама. Материалы для процессинга египетского диалекта были выбраны из соответствующего раздела Корпуса арабского языка [19].

Анализ языка интернет-коммуникации осуществлялся на базе комментариев к видеосюжетам на YouTube, размещенных на официальных страницах ведущих арабских медиаресурсов. При этом общее количество комментариев к сюжету не должно было превышать 100–110 единиц. То есть обработке подвергался весь комплекс комментариев. Выборка части из них не позволила бы дать объективную оценку. Известно, что распространена технология перемещения негативных комментариев к нижней части страницы.

• «1000 и 1 ночь» [20]

قالت: بلغني أيها الملك السعيد أن حسن حمد الله تعالى على نجاته من تلك الأهوال الصعبة وشكره على نعمته وفضله ونظروا إذا هم قد لاحت لهم القبة الخضراء والفسقية والقصر الأخضر ولاح لهم جبل السحاب من بعيد فقال الشيخ عبد القدوس يا حسن أبشر بالخير فأنت الليلة ضيف عند بنات أخي ففرح حسن بذلك فرحاً شديداً وكذلك زوجته ثم نزلوا عند القبة واستراحوا وأكلوا وشربوا ثم ركبوا وساروا حتى قربوا من القصر فلما أشرفوا عليه خرجت لهم بنات أخ الشيخ عبد القدوس وتلقينهم وسلمن عليهم وعلى عمهم وسلم عليهم عمهم وقال لهم يا بنات أخي ها أنا قد قضيت حاجة أخيكم حسن وساعدته على خلاص زوجته وأولاده

Таблица 1. Оценка «1000 и 1 ночь»

Инструмент	Camel Tools (CT)	Mazajak (M)	MonkeyLearn (ML)
Тональность	положительная	положительная	положительная
Степень уверенности алгоритма в метке	0,6	-	0,97

• *Сира Пророка [21]*

وَكَانَ سَبَبُ خُرُوجِ عَمْرٍو بْنِ عَامِرٍ مِنَ الْيَمَنِ - فِيمَا حَدَّثَنِي أَبُو زَيْدٍ الْأَنْصَارِيُّ - أَنَّهُ رَأَى جُرَدًا
 يَخْفِرُ فِي سِدِّ مَأْرِبٍ، الَّذِي كَانَ يَحْبِسُ عَلَيْهِمُ الْمَاءَ - فَيَصْرَفُونَهُ حَيْثُ شَاءُوا مِنْ أَرْضِهِمْ، فَعَلِمَ أَنَّهُ لَا
 بَقَاءَ لِلسِّدِّ عَلَى ذَلِكَ، فَاعْتَزَمَ عَلَى النُّقْطَةِ مِنَ الْيَمَنِ، فَكَادَ قَوْمَهُ، فَأَمَرَ أَصْغَرَ وَوَلَدَهُ إِذَا أَعْظَمَ لَهُ وَلَطَمَهُ
 أَنْ يَقُومَ إِلَيْهِ فَيَلْطِمَهُ، فَفَعَلَ ابْنُهُ مَا أَمَرَهُ بِهِ، فَقَالَ عَمْرُو: لَا أُقِيمُ بِبَلَدٍ لَطَمَ وَجْهِي فِيهِ أَصْغَرُ وَوَلَدِي،
 وَعَرَضَ أَمْوَالَهُ. فَقَالَ أَنْشَرَأْتُ مِنْ أَنْشَرَأْفِ الْيَمَنِ: اغْتَنِمُوا غَضَبَةَ عَمْرٍو، فَاشْتَرَوْا مِنْهُ أَمْوَالَهُ.

Таблица 2. Оценка Сира Пророка

Инструмент	Camel Tools (CT)	Mazajak (M)	MonkeyLearn (ML)
Тональность	нейтральная	отрицательная	отрицательная
Степень уверенности алгоритма в метке	0,99	-	0,64

• *Доисламская поэзия: Лябид ибн Абу Раби'а [22]*

عَفَتِ الدِّبَارُ مَحَلَّهَا فَمَقَامُهَا	بِمَنْى تَأَبَّدَ عَوَّلَهَا فِرْجَامُهَا
فَمَدَافِعِ الرِّيَّانِ عُرِّي رَسْمُهَا	خَلَقًا كَمَا ضَمِنَ الوَحْيُ سِلَامُهَا
دِمْنٌ تَجَرَّمَ بَعْدَ عَهْدِ أَنْبِيسِهَا	جَجَجُ خَلُونِ حَلَالُهَا وَحَرَامُهَا
وَدَقُّ الرُّوْعِ جَوْدُهَا فَرِ هَامُ	رُزِقَتْ مَرَابِيعَ النُّجُومِ وَصَابِهَا
مِنْ كُلِّ سَارِيَةٍ وَغَادٍ مُدْجِنِ	وَعَشِيَّةٍ مُتَجَاوِبِ إِرْزَامُهَا

Таблица 3. Оценка Лябид ибн Абу Раби'а

Инструмент	Camel Tools (CT)	Mazajak (M)	MonkeyLearn (ML)
Тональность	нейтральная	отрицательная	отрицательная
Степень уверенности алгоритма в метке	0,98	-	0,9

Столь различающиеся оценки и противоречивые комментарии обусловлены характером текста. Поэзия имеет свои стилистические, синтаксические, лексические и фонетические особенности, связанные в первую очередь с необходимостью адаптировать все эти особенности рифме, отвлекаясь от всех языковых правил. В этой

связи оценка ряда инструментов вновь опиралась на так называемый лексический подход – большое количество негативной лексики определяли тональность.

• *Египетский диалект*

الشاب: (الراقد على الأرض وهو يستعد للنوم).. معاهم حق يا جماعة.. الشبان هنا بييجوعوا، وجزء كبير قوي منهم بينحرف، الجوع عار.. والضياح عار.. كل دي حاجات تسوأ سمعة مصر فعلاً.. بس اللي أنا مش فاهمه.. هي قضية سياسية واجتماعية؟ والا قضية جغرافية.. هو يعني احنا بنجوع هنا بس... بنحرف هنا بس...؟!.. مش مهموقين قوي كده ليه على اللي بييجوعوا هناك..؟!.. يعني مسموح لنا لسوأ سمعة مصر بس على شرط، يكون جوه الحدود.. (يتشاءب).. التاييمز النهارده بتقول ان وزير مش عارف ايه، خد سمسرة مش عارف كام، في مشروع مش عارف ايه.. تصبجوا على خير

Таблица 4. Оценка египетского диалекта

Инструмент	Camel Tools (CT)	Mazajak (M)	MonkeyLearn (ML)
Тональность	отрицательная	отрицательная	отрицательная
Степень уверенности алгоритма в метке	0,99	-	0,83

В данном случае отмечено корректное определение тональности всеми решениями, что свидетельствует о внимании, которое стало уделяться разработчиками диалектной речи. Ранее большинство решений создавалось для обработки арабского литературного языка.

• *Социальные медиа*

Для анализа корректности сентимент-анализа социальных медиа использовались комментарии пользователей на YouTube к различным видеосюжетам. В качестве примера приведем здесь результаты обработки высказываний к видео ¹100. قطعة أثرية شريرة وجدت من قبل علماء الآثار

¹ В силу ограниченного объема текста статьи приведем лишь ссылку на видеосюжет: https://www.youtube.com/watch?app=desktop&v=luxq_DuOCEE&ab_channel=%25D9%2584%25D8%25A7%25D9%2586%25D9%2587%25D8%25A7%25D9%258A%25D8%25A9

Таблица 5. Оценка египетского диалекта

Инструмент	Camel Tools (CT)	Mazajak (M)	MonkeyLearn (ML)
Тональность	нейтральная / позитивная	положительный	положительный
Степень уверенности алгоритма в метке	?	-	0,91

СТ не смог обработать все комментарии, даже после предобработки текста (т. е. после удаления никнеймов, ссылок и т. д.). В этой связи были проанализированы первые 2000 символов, продемонстрировавшие нейтральный результат с меткой 0,43 и последние 2000 символов, обработка которых показала положительную тональность с меткой 0,87. Последний показатель сопоставим с результатами ML.

Столь различающиеся оценки и противоречивые комментарии обусловлены характером текста. Поэзия имеет свои стилистические, синтаксические, лексические и фонетические особенности, связанные в первую очередь с необходимостью адаптировать их рифме, отвлекаясь от всех языковых правил. В этой связи оценка ряда инструментов вновь опиралась на так называемый лексический подход – большое количество негативной лексики определяли всю тональность фрагмента текста.

Необходимо отметить то, что Mazajak не обучен на кораническом (классическом) тексте, что говорит в пользу применения CAMel Tools для анализа классических текстов или диалектов, который позволяет подключать конкретную языковую модель, обученную на соответствующих текстах (классический арабский язык или диалект).

Новые подходы в задачах обработки естественного языка

В последние два года область искусственного интеллекта переживает смену парадигмы и небывалый расцвет в связи с появлением новых моделей, натренированных на огромном количестве данных, которые легко адаптируются к самым разным задачам (моделирование языка, компьютерное зрение, диалоговые системы, робототехника и т. п.). Такие модели принято называть фундаментальными (англ. foundation models), подчеркивая их универсальность, масштаб и эффективность. С 2021 года область обработки естественного языка продемонстрировала удивительный рост лингвистических способностей моделей, значительно расширилась область их применения в различных лингвистических ситуациях.

Языковые модели с большим количеством параметров принято называть Большими языковыми моделями (англ. Large Language Models, LLMs) [23]. Яркой демонстрацией их возможностей стала ChatGPT [24], приспособленная для ведения диалога. Развитие больших языковых моделей не только повысило эффективность решения задач обработки естественного языка, но и поменяло сам способ разработки алгоритмов искусственного интеллекта. В отличие от меньших предобученных моделей, доступ к большим языковым моделям осуществляется через интерфейс для написания инструкций для модели (англ. prompting) или в виде чат-бота.

В обучающих данных большой языковой модели, лежащей в основе ChatGPT, представлены 95 языков (точная пропорция неизвестна), в том числе арабский. Оценке производительности ChatGPT в задачах арабского NLP посвящены ряд работ [25], [26]. В [26] проведено сравнение результатов моделей GPT-3.5 и GPT-4 между собой и с state-of-the-art результатами по 7 направлениям: суммаризация, диакритизация, определение частей речи, sentiment-анализ, транслитерация, машинный перевод и перефразирование. Следует отметить специфику выбранных задач, возникающих именно при обработке арабского языка. Оценивание моделей типа ChatGPT обычно происходит в двух условиях: Zero-shot, когда модели не предъявляются примеры из обучающего корпуса, а только инструкция, где написано, что именно требуется сделать модели: суммировать, перевести, расставить диакритики, определить часть речи и Few-shot, когда в инструкции предъявлены еще и несколько примеров правильных ответов на задание. GPT-модели продемонстрировали отличные возможности в задачах обработки арабского языка в условиях zero-shot, при этом уступая специализированным узкообученным моделям, особенно в задаче диакритизации.

В рамках нашего исследования была поставлена задача апробации работы GhatGPT для sentiment-анализа тех же фрагментов текстов, которые были использованы ранее для тестирования платформ CAMeL Tools, Mazajak и MonkeyLAb. Системе был задан запрос провести sentiment-анализ представленного текста.

В результате была получена краткий анализ каждого из рассмотренных текстов с точки зрения анализа настроений. Ниже следует перевод на русский язык результатов sentiment-анализа фрагмента «1000 и 1 ночи» (оригинальный текст на арабском языке приведен выше), выполненного ChatGPT:

«Sentiment-анализ этого текста показывает наличие в целом позитивных настроений. Детали анализа:

1. *Радость и удовольствие. Из текста очевидно, что персонажи рады тому, что Хасан смог пережить тяжелые испытания. Радость появляется при употреблении таких слов, как «Хасан был этому очень рад», что указывает на состояние большого удовольствия.*

2. *Благодарность. В тексте выражается признательность и благодарность со стороны персонажей, особенно когда они говорят: «Спасибо ему за милость и благосклонность». Это указывает на Богу за спасение и благодать.*

3. *Оптимизм и надежда. Оптимизм и надежда выражаются в таких словах, как «Я обещаю хорошие новости», а также в виде зеленого купола, фонтана и зеленого дворца. Герои ожидают добра в будущем, и это отражает положительную атмосферу в тексте.*

4. *Приветствие и встреча. Он оказывает теплый прием при встрече с племянницами шейха Абдула Куддуса, создавая позитивную атмосферу и чувство близости и семьи.*

5. *Социальная интеграция. Социальная интеграция очевидна, когда шейх Абдулла Куддус помогает Хасану и удовлетворяет его потребности, что отражает сплоченность и сотрудничество в обществе.*

В целом в этом тексте создается впечатление позитивного и радостного контекста, где преобладают радость, благодарность и оптимизм»².

Результаты, полученные в результате анализа текста с помощью ChatGPT демонстрируют не только общую оценку «нейтральный», «положительный», «негативный», но предоставляют аргументацию соответствующей оценки, что не предусмотрено иными инструментами для сентимент-анализа. Вместе с тем, здесь отсутствует количественная метка, свойственная CAMeL TOOLS или MonkeyLAB, в связи с чем комплексное решение, касающееся одновременного использования ChatGPT и одного из указанных выше инструментов позволяет верифицировать и расширить получаемую оценку тональности текста.

В качестве другого примера сентимент-анализа текста, выполненного ChatGPT, мы использовали фрагмент касыды Лябида ибн Абу Раби'а (текст приведен выше) и получили следующее описание.

«Текст, выражающий чувства, связанные с красотой, великолепием и волшебством, особым образом подчеркивающий значимость того или иного региона. Вот анализ настроений в тексте: 1. Красота и великолепие. В тексте говорится о красоте и природном великолепии. Акцент сделан на красивом внешнем виде мест и природы, окружающей данный район. 2. Целомудрие и мир. Подчеркнуто целомудрие места, а также его безопасность, что отражает его положительное влияние на окружающую среду. Термин «целомудрие домов» используется для обозначения чистоты и мира в этом регионе. 3. Религиозное значение. В тексте указывается на важность района с религиозной точки зрения, поскольку его влияние связано с откровением и миром. Термин «созданный, поскольку откровение гарантировало его покой» используется для обозначения важности этого

² Сохранена стилистика описания, представленного ChatGPT.

места в религиозном контексте. 4. Святость. В тексте подчеркивается сакральность некоторых вещей в этом регионе, например «хадж без его дозволенного и запретного», что усиливает уважение и почитание к этим местам. В целом текст передает чувство восхищения и гордости за красоту и святость региона, с акцентом на красоту природы и ее религиозное и культурное значение».

Приведенные выше результаты сентимент-анализа данного отрывка, осуществленные другими инструментами отличаются от комментариев модели ChatGPT, которая в определенной степени приукрасил содержание, но в целом корректнее передал тональность. При этом ChatGPT уловила на месте лексики с отрицательной тональностью наличие метафор, которые способны менять тональность на противоположную. Очевидно, что одной из перспективных задач является задача сентимент-анализа поэтических произведений в диахронии. Последнее особенно важно в контексте роли доисламской поэзии в арабомусульманской культуре и развитии языка в целом.

Важно отметить, что ChatGPT предоставил в целом корректное описание тестируемых текстов, которые относились к разным временным периодам и стилистическим жанрам. Особенно отрадно, что корректную обработку получают тексты на диалекте, что до недавних пор являлось одной из задач развития арабоязычного Интернета. Ниже представлены фрагмент результатов сентимент-анализа текста на египетском диалекте, оригинал которого приведен выше:

«Анализ чувств в этом тексте показывает смесь гнева и вопросов относительно социальных и экономических условий в Египте: 1. Гнев и обида. Судя по всему, существует сильное чувство гнева по поводу социальной и экономической ситуации в Египте. Молодой человек выражает обеспокоенность по поводу влияния этих негативных вопросов на репутацию Египта. Указывает на то, что проблема голода вредит репутации страны. 2. Непонимание. Возникает непонимание и сомнение в причинах и мотивах этих негативных явлений, будь то политических, социальных или географических. Спрашивающий выражает удивление тем, что эти проблемы возникают только внутри границ, без всякого интереса к ситуации за их пределами. 3. Разочарование. Когда молодой человек зевает, он показывает сцену капитуляции и разочарования, и это, возможно, отражает отсутствие эффективного решения имеющихся проблем. В целом текст выражает чувство гнева и сомнений в отношении социальной и экономической реальности в Египте, при этом связывая эти проблемы с политикой и заявлениями средств массовой информации».

В данном контексте очевидно, что ChatGPT в определенной степени берет на себя функции литературоведческого анализа, резюмируя обрабатываемый текст и определяя настроения и идеи, которые он несет. Подобного рода решения эффективно использовать в учебном процессе. Для развития аналитического

мышления обучающихся методологически оправдано просить сделать «анализ анализа», выполненного ChatGPT. Это лишь один из многочисленных примеров потенциального применения данного решения в учебном процессе.

Выводы

Интенсификация развития искусственного интеллекта оказывает существенное влияние на методологию и задачи проведения научно-практических исследований в самых различных областях знания, включая гуманитарные науки. Сентимент-анализ текста, с его способностью анализировать и понимать настроения, выраженные в огромных объемах арабского текста, стал мощным инструментом для расшифровки дискурса киберпространства. Поскольку технологии продолжают развиваться, использование возможностей и преодоление проблем, связанных с эволюцией языка в эпоху цифровых технологий, будут иметь решающее значение для содействия эффективному общению и культурному обмену в арабских социальных сетях.

Разработка точных моделей анализа настроений на арабском языке требует обширных лингвистических ресурсов и глубокого понимания культурных нюансов. Однако достижения в области алгоритмов машинного обучения и доступность крупномасштабных наборов данных на арабском языке открыли новые возможности для решения этих проблем и повышения точности анализа настроений.

Важнейшим аргументом в пользу важности проведения исследований междисциплинарного характера, предполагающих применение технологий ИИ в гуманитарных науках, является необходимость фиксации текущего состояния возможностей и результатов подобных работ. Последнее особенно актуально в связи с тем, что системы ИИ являются самообучаемыми и порождающими по своей сути, характеризующимися потенциально непредсказуемым развитием.

Сопоставительный анализ решений для обработки текста на арабском языке продемонстрировал эффективность использования ChatGPT совместно с одним из инструментов для сентимент-анализа текста. При этом в качестве преимуществ SAMeL Tools стоит отметить наличие показателя степени уверенности в метке и факт того, что он обучен на разнообразных текстовых данных. При этом сложность использования данного решения состоит в необходимости владения языком программирования. В этом отношении MonkeyLab кажется предпочтительнее, так как не предполагает каких-либо компетенций программиста, но при этом демонстрирует корректную работу и приводит метку уверенности. Кроме того, важно отметить, что MonkeyLab является мультязычным инструментом, использующим глубокое обучение. Что касается Mazajak, то он также предоставляет

пользователю веб-интерфейс для работы с системой. Но при этом скрыта реализация, не вполне очевидно, как осуществляется предобработка текста. Данный инструмент обучен только на социальных сетях, что делает сомнительным его использования на текстах на классическом арабском языке. Также по сравнению с SAMeL Tools, в Mazajak используется менее современная модель представления текста в виде численных векторов – word2vec.

Настоящую революцию в методологии проведения сентимент-анализа текста произвело появление ChatGPT. Использование данной модели в различных сферах деятельности в последнее время все больше волнует научное сообщество. Осуществляются попытки написания научных статей с помощью ChatGPT с последующим критическим анализом. Отмечается, что на первый взгляд полученный текст кажется точным, логичным. Он написан простым языком, поэтому может быть хорошо воспринят самой широкой аудиторией. Отмечена и быстрота работы инструмента. Вместе с тем, в качестве негативных моментов выделяются ошибки при формировании ссылок, стилистические особенности текста, которые, как отмечают некоторые исследователи, достаточно поверхностны («роботизированы») [27].

ChatGPT, как языковая модель, использует свою способность понимать и генерировать текст, похожий на человеческий, что делает его удобным для анализа настроений. Его способность генерировать контекстуально релевантные ответы может способствовать точному анализу настроений за счет контекстуализации языковых сложностей и культурных отсылок.

Проведенный сравнительный анализ инструментов анализа настроений на арабском языке раскрывает текущий ландшафт доступных решений, их сильные стороны и ограничения. Хотя в разработке такого рода технологий для арабского языка был достигнут значительный прогресс, есть возможности для их совершенствования, особенно в отношении разнообразных лингвистических особенностей, выраженных в арабском тексте.

Список источников

1. Word of the Year. URL: <https://www.collinsdictionary.com/woty>. (Дата обращения: 1 ноября 2023).
2. Gold E. The History of Artificial Intelligence from the 1950s to Today. Artificial Intelligence. URL: <https://www.freecodecamp.org/news/the-history-of-ai/>. (Дата обращения: 12 ноября 2023).

3. Biltawi M., Etaiwi W., Tedmori S., Hudaib A., and Awajan A. Sentiment Classification Techniques for Arabic Language: A Survey // 7th International Conference on Information and Communication Systems (ICICS). Irbid, Jordan. 2016. P. 339–346. <https://doi.org/10.1109/IACS.2016.7476075>.
4. Shoukry A. and Rafea A. A Hybrid Approach for Sentiment Classification of Egyptian Dialect Tweets. Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing), ACLING'15, (Washington, DC, USA). IEEE Computer Society. 2015. P. 78–85.
5. Khalifa K. and Omar N. A Hybrid Method Using Lexicon-Based Approach and Naive Bayes Classifier for Arabic Opinion Question Answering // Journal of Computer Science. 2014. Vol. 10, no. 10. P. 1961–1968. <https://doi.org/10.3844/jcssp.2014.1961.1968>.
6. Antoun W., Baly F., Hajj H. Arabert: Transformer-based Model for Arabic Language Understanding. arXiv preprint arXiv:2003.00104. 2020. <https://doi.org/10.48550/arXiv.2003.00104>.
7. Abdul-Mageed M., Elmadany A. R., Nagoudi E. M. B. ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. arXiv preprint arXiv:2101.01785. 2020. <https://doi.org/10.48550/arXiv.2101.01785>.
8. Nagoudi E. M. B., Elmadany A. R., Abdul-Mageed M. AraT5: Text-to-text Transformers for Arabic Language Generation. arXiv preprint arXiv:2109.12068. 2021. <https://doi.org/10.48550/arXiv.2109.12068>.
9. Obeid O. et al. CAMEL tools: An Open Source Python Toolkit for Arabic Natural Language Processing. Proceedings of the Twelfth Language Resources and Evaluation Conference. 2020. P. 7022–7032.
10. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Association for Computational Linguistics. 2019. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
11. CAMEL Tools Documentation URL: <https://camel-tools.readthedocs.io/en/stable/>. (Дата обращения: 1 октября 2023).
12. Farha I. A., Magdy W. Mazajak: An Online Arabic Sentiment Analyser, Proceedings of the Fourth Arabic Natural Language Processing Workshop. 2019. P. 192–198.

13. Mikolov T., Chen k., Corrado G., and Dean J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. 2013.
14. Mazajak: An Online Arabic Sentiment Analyser. URL: <http://mazajak.inf.ed.ac.uk:8000/>. (Дата обращения: 12 ноября 2023).
15. Rosenthal S., Farra N., Nakov P. SemEval-2017 task 4: Sentiment Analysis in Twitter. arXiv preprint arXiv:1912.00741. 2019.
16. Nabil M., Aly M., Atiya A. Astd: Arabic Sentiment Tweets Dataset. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015. 2515-2519.
17. Elmadany A., Mubarak H., Magdy W. Arsas: An Arabic Speech-act and Sentiment Corpus of Tweets //OSACT. 2018. URL: http://lrec-conf.org/workshops/lrec2018/W30/pdf/22_W30.pdf (Дата обращения: 12 ноября 2023).
18. MonkeyLearn. URL: <https://monkeylearn.com/> (Дата обращения: 12 ноября 2023).
19. ArabiCorpus. URL: <https://arabicorpus.byu.edu/> (Дата обращения: 12 ноября 2023).
20. ألف ليلة وليلة دار القلم. بيروت، لبنان. 1993، ص. 74
21. ابن هشام السيرة النبوية مؤسسة علوم القرآن. ص. 13
22. لبيد بن ربيع URL: <https://www.aldiwan.net/poem21196.html> (Дата обращения: 12 ноября 2023).
23. Shanahan M. Talking about large language models. arXiv:2212.03551 [cs.CL]. 2022. <https://doi.org/10.48550/arXiv.2212.03551>.
24. OpenAI. URL: <https://openai.com/blog/chatgpt/>. (Дата обращения: 12 ноября 2023).
25. Alyafeai Z. et al. Taqyim: Evaluating Arabic NLP Tasks Using ChatGPT Models //arXiv preprint arXiv:2306.16322. 2023.
26. Khondaker M. T. I. et al. GPTAraEval: A Comprehensive Evaluation of ChatGPT on Arabic NLP //arXiv preprint arXiv:2305.14976. 2023.
27. Hill-Yardin E. L., Hutchinson M. R., Laycock R., Spencer S. J. A Chat(GPT) about the Future of Scientific Publishing. // Brain Behavior and Immunity. Vol. 110. 2023. P. 152–154. <https://doi.org/10.1016/j.bbi.2023.02.022>.

Bibliographic References

1. Word of the Year. URL: <https://www.collinsdictionary.com/woty>. [Accessed: 1.11.2023].

2. Gold E. The History of Artificial Intelligence from the 1950s to Today. Artificial Intelligence. URL: <https://www.freecodecamp.org/news/the-history-of-ai/>. [Accessed: 12.11.2023].
3. Biltawi M., Etaiwi W., Tedmori S., Hudaib A., and Awajan A. Sentiment Classification Techniques for Arabic Language: A Survey. *2016 7th International Conference on Information and Communication Systems (ICICS)*. Irbid, Jordan. 2016:339-346. [https://doi: 10.1109/IACS.2016.7476075](https://doi.org/10.1109/IACS.2016.7476075).
4. Shoukry A. and Rafea A. A Hybrid Approach for Sentiment Classification of Egyptian Dialect Tweets. Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing), ACLING'15, (Washington, DC, USA). IEEE Computer Society. 2015. P. 78–85.
5. Khalifa K. and Omar N. A Hybrid Method Using Lexicon-Based Approach and Naive Bayes Classifier for Arabic Opinion Question Answering. *Journal of Computer Science*. 2014;10(10):1961-1968. <https://doi.org/10.3844/jcssp.2014.1961.1968>.
6. Antoun W., Baly F., Hajj H. Arabert: Transformer-based Model for Arabic Language Understanding. arXiv preprint arXiv:2003.00104. 2020. <https://doi.org/10.48550/arXiv.2003.00104>.
7. Abdul-Mageed M., Elmadany A. R., Nagoudi E. M. B. ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. arXiv preprint arXiv:2101.01785. 2020. <https://doi.org/10.48550/arXiv.2101.01785>.
8. Nagoudi E. M. B., Elmadany A. R., Abdul-Mageed M. AraT5: Text-to-text Transformers for Arabic Language Generation. arXiv preprint arXiv:2109.12068. 2021. <https://doi.org/10.48550/arXiv.2109.12068>.
9. Obeid O. et al. CAMEL tools: An Open Source Python Toolkit for Arabic Natural Language Processing. Proceedings of the Twelfth Language Resources and Evaluation Conference. 2020. 7022–7032.
10. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics. 2019. 4171–4186. [https://doi: 10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423).
11. CAMEL Tools Documentation. URL: <https://camel-tools.readthedocs.io/en/stable/>. [Accessed: 1.10.2023].

12. Farha I. A., Magdy W. Mazajak: An Online Arabic Sentiment Analyser. Proceedings of the Fourth Arabic Natural Language Processing Workshop. 2019. 192–198.
13. Mikolov T., Chen k., Corrado G., and Dean J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. 2013.
14. Mazajak: An Online Arabic Sentiment Analyser. URL: <http://mazajak.inf.ed.ac.uk:8000/>. [Accessed: 12.11.2023].
15. Rosenthal S., Farra N., Nakov P. SemEval-2017 task 4: Sentiment Analysis in Twitter. arXiv preprint arXiv:1912.00741. 2019.
16. Nabil M., Aly M., Atiya A. Astd: Arabic Sentiment Tweets Dataset. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015. 2515–2519.
17. Elmadany A., Mubarak H., Magdy W. Arsas: An Arabic Speech-act and Sentiment Corpus of Tweets. OSACT. 2018. URL: http://lrec-conf.org/workshops/lrec2018/W30/pdf/22_W30.pdf. [Accessed: 12.11.2023].
18. MonkeyLearn. URL: <https://monkeylearn.com/>. [Accessed: 12.11.2023].
19. ArabiCorpus. URL: <https://arabicorpus.byu.edu/>. [Accessed: 12.11.2023].
20. 1000 and 1 night. Lebanon. 1993. P. 74. (In Arab.).
21. Ibn Hisham. [Biography of the Prophet]. Foundation for Qur'anic Sciences. P. 13. (In Arab.).
22. Labid bin Rabi'. URL: <https://www.aldiwan.net/poem21196.html>. [Accessed: 12.11.2023]. (In Arab.).
23. Shanahan M. Talking about large language models. arXiv:2212.03551 [cs.CL]. 2022. <https://doi.org/10.48550/arXiv.2212.03551>.
24. OpenAI. URL: <https://openai.com/blog/chatgpt/>. [Accessed: 12.11.2023].
25. Alyafeai Z. et al. Taqyim: Evaluating Arabic NLP Tasks Using ChatGPT Models. arXiv preprint arXiv:2306.16322. 2023.
26. Khondaker M. T. I. et al. GPTAraEval: A Comprehensive Evaluation of ChatGPT on Arabic NLP. arXiv preprint arXiv:2305.14976. 2023.
27. Hill-Yardin E. L., Hutchinson M. R., Laycock R., Spencer S. J. A Chat (GPT) about the Future of Scientific Publishing. *Brain Behavior and Immunity*. 2023;110:152-154. <https://doi.org/10.1016/j.bbi.2023.02.022>.

Информация об авторе

Берникова Ольга Александровна – канд. филол. наук, доцент Санкт-Петербургского государственного

Information about the author

Olga A. Bernikova – PhD in Philology, Associate Professor of St Petersburg State University, St Petersburg,