# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ОБРАБОТКА РУССКОГО ЯЗЫКА: ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

***Асанов Руслан Арсенович, преподаватель***
***Гулистанского государственного педагогического института***
***Меджитова Эльмира Дляверовна,***
***Студентка Гулистанского государственного***
***педагогического института***

**Аннотация.** В данной статье рассматривается развитие искуственного интеллекта тесно связано с обработкой естественного языка. Несмотря на значительные успехи в области обработки английского языка, задачи связанные с русским языком по-прежнему представляют собой ряд научных и технических вызовов. Также рассматривается основные проблемы обработки русского языка в системах, анализируются существующие подходы и предлагаются направление для дальнейших исследований.

**Ключевые слова:** искусственный интеллект, обработка естественного языка, русский язык, морфология, семантика, машинное обучение.

**Annotation.** This article discusses the development of artificial intelligence closely related to natural language processing. Despite significant advances in the field of English language processing, tasks related to the Russian language still pose a number of scientific and technical challenges. It also examines the main problems of Russian language processing in systems, analyzes existing approaches, and suggests directions for further research.

**Keywords:** artificial intelligence, natural language processing, Russian language, morphology, semantics, machine learning.

Современные технологии искусственного интеллекта всё активнее внедряются в повседневную жизнь, в том числе в сферу обработки естественного языка. Русский язык как один из крупнейших по числу носителей и сложности грамматической структуры представляет собой интересный и одновременно трудный объект для компьютерной обработки. Однако он относится к числу языков со сложной морфологией, высшей степенью синтаксической вариативности и богатой лексической системой, что усложняет его автоматическую обработку. В отличии от английского языка, для которого создано огромное количество обучающих корпусов и программных решений, русскоязычные ресурсы развиты в меньшей степени, что ограничивает потенциал искусственного интеллекта при работе с русским языком.

Развитие ИИ-технологий, таких как нейросетевые модели (например, - BERT, GPT), открыло новые возможности для понимания. Однако при их применении к русскому языку возникают уникальные проблемы, требующие корректировки и внесения дополнений к выданным результатам. Русский язык – это важный элемент национальной идентичности. Качественная обработка русской речи с помощью ИИ

необходима для создания голосовых помощников, поисковых систем, систем машинного перевода, автоматического реферирования и анализа больших текстов.

1. Лингвистические проблемы обработки русской устной и письменной лексики. Русский язык характеризуется высоким уровнем флективности слова (изменение по родам, числам, падежам и временам), что усложняет задачи токенизации, морфологического анализа и лемматизации. Например, одно слово может иметь десятки различных форм, что требует высокой способности к обобщению и контекстному анализу.

2. Свободный порядок слов. Порядок слов в русском языке сравнительно свободен, что создаёт проблемы для моделей, обученных на фиксированном (фиксированном) порядке как английский язык. Например, предложение «мальчик дал девочке книгу» может быть переведено без учёта интонационного выделения, что изменяет смысл. Важно учитывать семантические зависимости, все зависящие от порядка слов. Для этого всё чаще используются синтаксические деревья зависимостей и механизмы внимания.

3. Многозначность и омонимия. Русский язык богат на омонимы и многозначные слова, что требует семантического анализа. Значение играет ключевую роль в интерпретации значений. Например, слово «ключ» может обозначать как источник воды, так и шифр, или способ решения задачи.

4. Фразеология и идиомы. Фразеологизмы часто не интерпретируются буквально, но это требует от системы способности к контекстному и прагматическому анализу. Например, фраза "бить баклуши" не имеет отношения к физическому действию и не интерпретируется корректно без знания культурного контекста. Это особенно актуально для задач машинного перевода и генерации текста.

5. Орфографические и пунктуационные вармации. Разнообразие орфографии, используемые разговорные формы и сокращения (например, в социальных сетях) также создают сложности. Люди сдают боты (устройства) и программному обеспечению написанное и проговариваемое содержимое значительно сложнее.

Современные подходы и достижения. С 2019 года началась активная разработка нейросетевых моделей для понимания письменной речи и родного языка для обучения модели. Появились подходы, совмещающие статистические и нейронные методы, работающие на границах различных дисциплин (например, грамматический разбор, правила орфографии, синтаксический анализ). Такие подходы используются, например, в системах грамматической проверки и морфосинтаксического анализа.

Перспективы развития: при использовании ИИ модели необходимо выявить и определить стиль речи и сферу использования, включая специализированные тексты (например, медицинские, правовые, научные).

Поддержка региональных групп и диалектов: русский язык имеет множество региональных вариаций, а также используется как второй язык в странах СНГ.

Интеграция с другими ИИ-системами: обработка языка всё чаще становится частью комплексных ИИ-систем: роботов, голосовых ассистентов, аналитических платформ.

Несмотря на значительное количество ручного труда и ограниченность ресурсов в последние годы наблюдается значительный прогресс в области их автоматизации и обработки благодаря усилиям исследователей и развитию технологий глубокого обучения, а также доступна мощные инструменты и модели.

Совместная работа ИИ энтузиастов, инженеров, программистов и научного сообщества в целом может сыграть важную роль в эффективности, точности и качественной обработке русского языка в будущем.

**Список использованной литературы**

1. Devlin J., Chang M. W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019.
2. Kuratov Y., Arkhipov M. Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language. 2019.
3. Национальный корпус русского языка: www.ruscorpora.ru
4. DeepPavlov: open-source library for NLP tasks in Russian.
5. Tikhonov A., et al. RuGPT: Russian Generative Pre-trained Transformer Models. 2020.
6. Sidorov G. et al. Lingvo-Driven Approaches for Russian NLP. 2021.