

Адаптация учебных технических текстов для иностранных обучающихся с использованием модели DeepSeek V3
Гу А. (Университет ИТМО), Кукрякова Р.Р. (Университет ИТМО)
Научный руководитель – доцент ФПИн Валитова Ю.О. (Университет ИТМО)

Введение. В условиях интернационализации высшего образования российские вузы сталкиваются с необходимостью обеспечения доступности учебных материалов для иностранных студентов с разным уровнем владения русским языком как иностранным (РКИ). Существующие автоматические решения для упрощения русскоязычных текстов, как правило, решают лишь часть задачи [1, 2]: они либо снижают общую сложность текста без учёта шкалы уровней владения РКИ, либо не сохраняют профессиональную терминологию, что делает их непригодными для использования в предметном обучении иностранных студентов. Гипотеза исследования состоит в том, что с использованием модели DeepSeek V3, управляемой через текстовые промпты, возможно адаптировать учебные тексты для иностранных обучающихся достигнув оптимального баланса между тремя критериями: соответствие целевому уровню владения русским языком, сохранение профессиональной терминологии и смысловой целостности исходного текста. Основным методом выступает итеративный дизайн промптов с последующей комплексной автоматизированной оценкой результатов.

Основная часть. Исследование проходило в шесть этапов. Этап 1 — формирование исходного корпуса. На вход схемы подаётся корпус оригинальных учебных текстов с исходной сложностью C2 и выше, отобранных из учебных пособий, изданных университетом ИТМО. Этап 2 — разработка и оптимизация промптов. В результате разработки и оптимизации промптов был осуществлён переход от первой версии промпта к второй. Во второй версии промпта (v2) вместо общих требований введены конкретные лингвистические правила для каждого уровня (B1/B2). Этап 3 — автоматическая обработка и адаптация текстов через API DeepSeek V3. Этап 4 — формирование итогового корпуса. Для каждой версии промптов составляет 50 сгенерированных текстовых единиц [3]. Этап 5 — комплексная автоматизированная оценка. Этап 6 — анализ результатов и обратная связь. На этом этапе выявляем слабые места и точки роста для текущей версии промпта. При необходимости, на основе этих выводов осуществлялся возврат к этапу 2 для разработки следующей версии промпта. Ниже представлены результаты автоматической оценки.

Оценка соответствия уровню РКИ проведена через анализ лингвистических метрик сложности текста, полученных от сервиса «Текстомер» [4]. Результаты свидетельствуют о существенном прогрессе в управляемости процесса адаптации учебных текстов при использовании промпта v2. Для условия «уровень владения РКИ B1 без терминов» появились успешные адаптации учебных текстов (30% против 0% в v1). Для условия «уровень владения РКИ B2 без терминов» доля успешных адаптаций возросла с 50% до 80%. Однако задача адаптации учебных текстов с обязательным сохранением терминов осталась нерешённой для обеих итераций. Это указывает на внутреннее противоречие между глубиной требуемого упрощения и необходимостью сохранять сложную профессиональную лексику. Такие результаты связаны с тем, что профессиональная лексика относится к более высокому уровню владения РКИ. Однако следует отметить, что несмотря на 0% успеха у адаптации учебных текстов до уровня владения РКИ B1 с терминами для обеих итераций среднее отклонение фактического уровня текста от целевого диапазона сократилось с +2.21 (v1) до +1.31 (v2), что указывает на прогресс. Оценка сохранения смысловой целостности с использованием метрики BERTScore [5]

показала, что повышение управляемости упрощением в v2 было достигнуто ценой системного снижения семантического качества. Наибольшие потери смысла (падение F1-score на 0.10–0.17) наблюдались при адаптации учебных текстов до уровня владения РКИ В1, особенно в условиях без терминов. Правила для уровня владения РКИ В2 в v2 оказались более сбалансированными, обеспечив умеренные семантические потери при значительном прогрессе в достижении целевой сложности. Оценка сохранения терминологии подтвердила высокую эффективность промпта v2. По сравнению с v1, полнота сохранения терминов значительно увеличилась примерно на 16% для В1 и 28% для В2. Этот результат говорит о критической важности чёткой формулировки требования о терминах в промпте.

Выводы. В результате исследования был получен оптимизированный промпт, рекомендованный для использования в системе, особенно для адаптации учебных текстов до уровня владения РКИ В2. Разработанная комплексная методика оценки качества адаптации учебных технических текстов позволила не только констатировать качество, но и количественно проанализировать компромиссы между тремя целевыми критериями. Это позволяет утверждать, что гипотеза исследования доказана.

Список использованных источников:

1. TLDR Summarize. – URL: <https://play.google.com/store/apps/details?id=com.aya.textsummarizer> (дата обращения: 27.02.2026)
2. Text Summarizer. – URL: <https://play.google.com/store/apps/details?id=com.summarisingtool.textsummarizer> (дата обращения: 27.02.2026)
3. Корпус исходных и адаптированных текстов. – URL: <https://docs.google.com/spreadsheets/d/1qKh2nY5Cm81vcnN50uhXFIsMUT7yYXF7KlSeojJvE-g/edit?usp=sharing> (дата обращения: 27.02.2026).
4. Лапошина А. Н., Лебедева М. Ю. Текстометр: онлайн-инструмент определения уровня сложности текста по русскому языку как иностранному // Русистика. – 2021. Т. 19. – №3. – С. 331–345.
5. BertScore. – URL: <https://huggingface.co/spaces/evaluate-metric/bertscore> (дата обращения: 27.02.2026)