

ОТЗЫВ

**на автореферат диссертации Скачкова Николая Андреевича
«Вероятностные нейросетевые модели понимания и генерации текстов
естественного языка»,
представленной на соискание учёной степени кандидата технических
наук
по специальности 2.3.8 «Информатика и информационные процессы»**

Машинный перевод относится к важным направлениям обработки естественного языка. Современные системы могут утрачивать, добавлять или искажать информацию при переводе. Дополнительные трудности возникают при дефиците параллельных данных, работе в специальной предметной области и переводе документов, контекстные связи в которых выходят за границы предложения. Решению этих проблем посвящена диссертация Н. А. Скачкова, что определяет её актуальность.

Автореферат даёт целостное представление о цели, задачах, новизне и практической значимости работы. После обзора развития машинного перевода и базовой вероятностной постановки последовательно рассматриваются четыре предложенных решения.

Для перевода на малоресурсных направлениях соискатель выводит функцию обучения из идеи циклического правдоподобия, что позволяет использовать знания обратной модели. Переход к дообучениям готовых моделей снижает вычислительные ограничения подхода. Результаты на англо-финской и русско-казахской языковых парах показывают повышение BLEU и CycleBLEU при использовании предложенного соискателем подхода.

Для учёта предпочтений человека предложено формировать пары из гипотез самой переводческой модели и применять попарное ранжирование в контрастной функции потерь. Такой способ не ограничивает множество ошибок заранее заданными правилами и не требует полного корпуса эталонных переводов. Применение метода к большой языковой модели позволяет добиться улучшения BLEURT и COMET. Практическую значимость результата подтверждают доменная адаптация и участие системы в WMT24.

В качестве усложнения используемой в обучении функции потерь соискателем предлагается использовать метод маскировки частей исходного текста и перевода. Автор показывает возможность объединить в одной архитектуре перевод, одноязычное обучение и постредактирование. В экспериментах при добавлении русскоязычных одноязычных данных прироста не обнаружено; при этом в задаче постредактирования предложенный автором метод позволяет улучшить качество переводов, полученных с помощью более

слабой модели.

В задаче обработки длинных документов предложена модификация аддитивной тематической модели, учитывающая сегментную структуру при постобработке E-шага. Получено улучшение WindowDiff и Pk по сравнению с TopicTiling и более простыми видами разреживания. Сегментация документов с помощью тематической модели приводит к улучшению BLEU относительно случайного разбиения.

К автореферату имеется следующее замечание. При описании обучения на предпочтениях недостаточно подробно описаны требования к согласованности оценок экспертов и зависимость результата от объёма разметки. Эти замечания имеют дискуссионный характер и не влияют на общую высокую оценку исследования.

Автореферат отражает содержание диссертации и позволяет оценить личный вклад соискателя. Работа «Вероятностные нейросетевые модели понимания и генерации текстов естественного языка» является завершённой научно-квалификационной работой, соответствует специальности 2.3.8 «Информатика и информационные процессы» и требованиям, предъявляемым к кандидатским диссертациям. Скачков Николай Андреевич заслуживает присуждения учёной степени кандидата технических наук.

Роман Владимирович Исаченко,

Кандидат физико-математических наук,

ООО «Яндекс.Технологии»

руководитель групп базовых технологий VLM и ART

Подпись

Дата: 29.09.2026

Подпись Исаченко Романа Владимировича заверяю. Представитель по доверенности б/н от 15.01.2025 / К.К. Вологжанина /

