ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.1.224.03,
СОЗДАННОГО НА БАЗЕ
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО УЧРЕЖДЕНИЯ
«ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«ИНФОРМАТИКА И УПРАВЛЕНИЕ»
РОССИЙСКОЙ АКАДЕМИИ НАУК»,
ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ
СТЕПЕНИ КАНДИДАТА НАУК

Аттестационное дело № _____

Решение диссертационного совета 24.1.224.03 от 15.10.2026, № 21

О присуждении Скачкову Николаю Андреевичу, гражданину Российской Федерации, учёной степени кандидата технических наук.

Диссертация «Вероятностные нейросетевые модели понимания и генерации текстов естественного языка» по специальности 2.3.8 – «Информатика и информационные процессы» принята к защите 25 июня 2026 г., протокол № 16, диссертационным советом 24.1.224.03 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской Академии Наук» (ФИЦ ИУ РАН), 119333, г. Москва, ул. Вавилова, д. 44/2, созданным на основании приказа Министерства науки и высшего образования Российской Федерации № 1732/нк от 13.12.2022.

Соискатель – Скачков Николай Андреевич, дата рождения 2 декабря 1997 года, в 2021 году окончил очную магистратуру Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова» (МГУ) по направлению подготовки 01.04.02 «Прикладная математика и информатика». С 2021 по 2025 год обучался в очной аспирантуре ФИЦ ИУ РАН. В настоящее время работает в ООО «Яндекс.Технологии» в должности руководителя службы качества претрейна Яндекс ДжиПиТи.

Диссертация выполнена в отделе №14 ФИЦ ИУ РАН.

Научный руководитель – Воронцов Константин Вячеславович, доктор физико-математических наук, профессор РАН, заведующий кафедрой математических методов прогнозирования факультета вычислительной математики и кибернетики МГУ.

Официальные оппоненты:

Максимов Николай Вениаминович, доктор технических наук, профессор, профессор кафедры финансового мониторинга № 75 Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ»;

Ильвовский Дмитрий Алексеевич, кандидат технических наук, научный сотрудник международной лаборатории интеллектуальных систем и структурного

анализа Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики»

дали **положительные** отзывы на диссертацию.

Ведущая организация – Федеральное государственное бюджетное учреждение науки Всероссийский институт научной и технической информации Российской академии наук в своем **положительном** отзыве, подписанном заведующим отделом интеллектуальных средств автоматизации к.ф.-м.н. Селюковым А.С. и ведущим научным сотрудником д.т.н. Хорошиловым А.А., указала, что диссертация Скачкова Н.А. «Вероятностные нейросетевые модели понимания и генерации текстов естественного языка» посвящена разработке вероятностных и архитектурных методов обучения генеративных нейросетевых моделей машинного перевода, направленных на повышение смысловой точности, лингвистической грамотности и контекстной связности автоматических переводов. В работе теоретически обоснован метод совместного обучения прямой и обратной моделей перевода на основе правдоподобия циклических переводов, предложены метод переупорядочивания переводческих гипотез с использованием контрастного обучения и попарной человеческой разметки, метод обучения моделей перевода с маскировкой входа, а также способ учёта сегментной структуры документа в аддитивно регуляризованных тематических моделях. В отзыве отмечены научная новизна, теоретическая и практическая значимость полученных результатов, их применимость в промышленных системах машинного перевода, при доменной адаптации к специализированным текстам, для малоресурсных языковых направлений, в инструментах автоматизированного перевода, при обработке длинных документов, в информационном поиске и системах научно-технической информации. Рекомендовано использование результатов в научных организациях и вузах, ведущих исследования в области обработки естественного языка, машинного перевода и информационного поиска, а также в компаниях, разрабатывающих системы машинного перевода и сервисы локализации. Сформулированы замечания к строгости экспериментального обоснования, полноте сравнений с современными методами обучения по человеческим предпочтениям, экспериментальному подтверждению возможностей маскировки входа и тематической сегментации, а также выбору метрик оценки качества перевода. Указано, что замечания носят в основном дискуссионный и рекомендательный характер, не снижают научной ценности работы и не влияют на её общую положительную оценку. Сделан вывод, что диссертация является завершённой научно-квалификационной работой, отвечает требованиям пунктов 9–14 Положения о присуждении учёных степеней, утверждённого постановлением Правительства Российской Федерации № 842 от 24 сентября 2013 г., а её автор, Скачков Николай Андреевич, заслуживает присуждения учёной степени кандидата технических наук по специальности 2.3.8 — «Информатика и информационные процессы».

Основные результаты диссертационного исследования опубликованы в 6 работах. В журналах из перечня ВАК РФ – 2, в изданиях, индексируемых в международной базе Scopus – 6.

Недостоверные сведения об опубликованных соискателем работах в диссертации отсутствуют. Диссертация соответствует п. 14 Положения о присуждении учёных степеней.

Работы Скачкова Н.А. по теме диссертации:

1. Воронцов К.В., Скачков Н.А. Упорядочивание гипотез в моделях перевода с использовани- ем человеческой разметки // Известия Российской Академии Наук. Теория и системы управления. 2024. № 4. С. 121-128.(Перевод: Vorontsov K.V., Skachkov N.A. Reranking Hypotheses in Translation Models Using Human Markup // Journal of Computer and Systems Sciences International. 2024. V. 63. No 4. PP. 679-686.) (ВАК, Scopus).
2. Скачков Н.А., Воронцов К.В. Улучшение качества машинного перевода с использовани- ем обратной модели // Автоматика и телемеханика. 2022. № 12. С. 31-43. (Перевод: Skachkov N.A., Vorontsov K.V. Improving the Quality of Machine Translation Using the Reverse Model // Automation and Remote Control. 2022. V. 83. P.1897-1907.) (ВАК, Scopus).
3. Skachkov N. Method of Input Masking for Training Translation Models // Pattern Recognition and Image Analysis. 2025. V.35. P. 493-500. (Scopus).
4. Skachkov N.A., Vorontsov K.V. Improving Topic Models with Segmental Structure of Texts // Computational Linguistics and Intellectual Technologies: Proc. Internat. Conf. "Dialogue 2018". Moscow, 2018. V.17. P. 652-661. (Scopus).
5. Shtekh G., Kazakova P., Nikitinsky N., Skachkov N.A. Applying Topic Segmentation to Document-Level Information Retrieval // Proc. 14th Central and Eastern European Software Engineering Conf. Russia. Moscow, Russian Federation, 2018. (CEE- SECR '18). (Scopus).
6. Elshin D., Karpachev N., Gruzdev B., Golovanov I., Ivanov G., Antonov A., Skachkov N., Latypova E., Layner V., Enikeeva E., Popov D., Chekashev A., Negodin V., Frantsuzova V., Chernyshev A., Denisov K. From General LLM to Translation: How We Dramatically Improve Translation Quality Using Human Evaluation Data for LLM Finetuning // Proc. Ninth Conf. Machine Translation. Miami, Florida, USA, 2024. P.247-252. (Scopus).

На автореферат диссертации поступили отзывы:

1. Панченко Александра Ивановича, доктора компьютерных наук, PhD, заведующего лабораторией обработки естественного языка, Центр искусственного интеллекта, Сколковский институт науки и технологий.
2. Исаченко Романа Владимировича, кандидата физико-математических наук, руководителя групп базовых технологий VLM и ART, ООО «Яндекс.Технологии».

Все **отзывы положительные**, замечания носят уточняющий и рекомендательный характер, не являются принципиальными и не влияют на главные теоретические и практические результаты диссертационной работы.

Также имеется акт о внедрении результатов диссертационного исследования, полученный от ООО «Яндекс.Технологии».

Выбор официальных оппонентов и ведущей организации обосновывается их достижениями в области глубокого обучения и обработки текстов естественного языка, что подтверждается их публикациями в авторитетных рецензируемых изданиях.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

разработан вероятностный метод совместного обучения прямой и обратной моделей машинного перевода, теоретически обоснованный посредством максимизации правдоподобия циклических переводов; предложен способ его применения на этапе дообучения предобученных моделей без использования вспомогательных языковых моделей; показано, что применение метода повышает качество англо-финского и русско-казахского перевода по метрикам BLEU и CycleBLEU;

разработан метод переупорядочивания гипотез перевода, объединяющий контрастное обучение с использованием попарной человеческой разметки переводческих гипотез, порождённых самой моделью; показано, что метод сокращает долю систематических ошибок перевода, позволяет проводить доменную адаптацию без эталонных параллельных текстов и применим к дообучению современных языковых моделей, адаптированных к задаче перевода;

разработан метод обучения моделей машинного перевода с маскировкой входа, основанный на предсказании перевода по частично замаскированным исходному и целевому текстам; показано, что модификация вероятностной модели и функции потерь повышает качество англо-русского перевода по метрике BLEU и снижает число лингвистических и семантических ошибок; предложен способ применения данного подхода к использованию одноязычных данных и построению моделей постредактирования перевода;

разработан способ тематической сегментации длинных текстов на основе аддитивно регуляризованных тематических моделей, учитывающий сегментную структуру документа посредством постобработки E-шага; показано, что предложенный способ улучшает качество сегментации текстов, а его применение повышает качество контекстного перевода длинных документов по сравнению со случайной сегментацией по метрике BLEU.

Теоретическая значимость данной работы заключается в разработке вероятностных методов обучения генеративных нейросетевых моделей машинного перевода и методов тематической сегментации текста, направленных на преодоление нехватки обучающих данных, ограничений авторегрессионной вероятностной

модели, систематических ошибок перевода и недостаточного учёта контекста длинных документов. Предложенный метод совместного обучения прямой и обратной моделей перевода теоретически выведен из задачи максимизации правдоподобия циклических переводов, что позволяет обосновать функцию потерь и процедуру дообучения предобученных моделей без использования вспомогательных языковых моделей. Разработанный метод переупорядочивания переводческих гипотез объединяет контрастное обучение с попарной экспертной разметкой гипотез, порождённых самой моделью, и позволяет учитывать человеческие предпочтения при обучении моделей перевода и их адаптации к специализированным текстам. Предложенный метод маскировки входа модифицирует вероятностную модель и функцию потерь за счёт предсказания перевода по частично замаскированным исходному и целевому текстам и создаёт единую основу для обучения на параллельных и одноязычных данных, а также для решения задачи постредактирования перевода. Разработанный способ учёта сегментной структуры документа в аддитивно регуляризованных тематических моделях основан на выводе формулы постобработки E-шага из требования разреженности тематических распределений сегментов и позволяет учитывать тематическую однородность фрагментов текста без усложнения EM-алгоритма.

Практическая значимость заключается в разработке методов повышения качества нейросетевого машинного перевода, применимых при нехватке параллельных данных, адаптации моделей к специализированным текстам, устранении систематических ошибок при переводе длинных документов. Предложенный метод совместного обучения прямой и обратной моделей позволяет улучшать качество перевода на малоресурсных языковых направлениях и может использоваться при расширении языковой поддержки существующих автоматических переводческих систем. Метод переупорядочивания гипотез на основе экспертной разметки позволяет адаптировать модели к требованиям конкретной предметной области без сбора эталонных параллельных текстов; его эффективность продемонстрирована в задаче перевода товарных заголовков электронной коммерции и при дообучении языковых моделей для перевода. Метод маскировки входа позволяет повышать качество автоматического перевода и создавать модели постредактирования, применимые в инструментах автоматизированного перевода для сокращения объёма ручной корректировки. Предложенный способ тематической сегментации длинных текстов позволяет выделять семантически однородные фрагменты и улучшать качество контекстного перевода документов. Разработанные методы могут использоваться в промышленных системах машинного перевода, сервисах локализации и системах обработки научно-технической информации. Результаты диссертации Н.А. Скачкова используются в программном обеспечении автоматизированных систем машинного перевода в ООО «Яндекс.Технологии».

Достоверность результатов следует из корректного использования методов, теоретического обоснования решаемых задач, а также из экспериментальных исследований, демонстрирующих эффективность предложенных методов и алгоритмов, корректность работы разработанных программных решений. Достоверность полученных результатов также подтверждена апробацией их на 4 международных и всероссийских научных конференциях, публикациями в рецензируемых журналах.

Личный вклад. Все основные результаты по теме диссертации, включающие разработанные методы, методику, техническое решение, программные реализации разработанных алгоритмов и выполненные экспериментальные исследования, получены автором самостоятельно.

В ходе защиты диссертации были высказаны следующие **замечания**:

1. При практической проверке разрабатываемых методов излишнее внимание уделяется классической метрике BLEU, хотя в последнее время активно применяются нейросетевые метрики (BERT-score, BART-score, LLM-score).
2. В большинстве случаев в качестве внешнего эталона выступает только закрытая система Google Translate (по сути, черный ящик), хотя для многих задач существуют открытые SOTA-модели (основанные, как правило, на тех же трансформерах), позволяющие добиться воспроизводимых результатов и даже провести дообучение в ограниченном масштабе.
3. При использовании маскированной функции потерь добавление одноязычных русскоязычных данных не привело к улучшению качества перевода. В работе это объясняется тем, что англо-русское направление не является малоресурсным. Представляется полезной дополнительная проверка данного объяснения на языковой паре с выраженным дефицитом параллельных данных, а также явное указание количества параллельных данных, начиная с которого, статистически значимое улучшение качества от использования одноязычных данных пропадает.
4. Предложенные методы проверяются по отдельности и на разных данных, что не позволяет говорить о синергетическом эффекте от их совместного применения в единой системе: необходимо проведение дополнительных теоретических и экспериментальных исследований. Кроме того, более уместной была бы единообразная спецификация тестовых данных.

Соискатель Скачков Н.А. ответил на заданные вопросы и согласился с замечаниями, указанными в ходе заседания.

На заседании 15 октября 2026 г. диссертационный совет принял решение присудить Скачкову Николаю Андреевичу учёную степень кандидата технических наук за создание совокупности методов обучения генеративных нейросетевых моделей машинного перевода и метода тематической сегментации длинных текстов, позволяющих повышать качество перевода в условиях нехватки параллельных

данных, учитывать человеческие предпочтения, сокращать систематические ошибки и улучшать контекстный перевод документов, которые могут быть использованы при построении и совершенствовании систем автоматического перевода.

При проведении тайного голосования диссертационный совет в количестве 22 человек, из них 5 докторов наук по специальности 2.3.8 — «Информатика и информационные процессы», участвовавших в заседании, из 29 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за – 22, против – 0, недействительных бюллетеней 0.

Заместитель председателя
диссертационного совета 24.1.224.03
д.т.н. Матвеев И.А.

Учёный секретарь
диссертационного совета 24.1.224.03
к.т.н. Рейер И.А.

15 октября 2026 г.