

**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»**

Диссертационная работа Н. А. Скачкова посвящена разработке методов обучения моделей машинного перевода. Актуальность темы определяется тем, что современные нейросетевые системы по-прежнему допускают пропуски, избыточную генерацию, смысловые и грамматические ошибки, нестабильно работают при дефиците параллельных данных и могут нарушать связность длинных текстов. Поэтому методы повышения точности, естественности и контекстной согласованности перевода имеют существенное научное и прикладное значение.

Цель работы состоит в разработке новых методов обучения генеративных моделей машинного перевода, учитывающих нехватку и недостаточное качество данных, ограничения авторегрессионной постановки и проблему сегментации длинного текста. Представленные в автореферате результаты соответствуют этой цели.

Первый результат относится к совместному обучению прямой и обратной моделей. Метод обоснован через максимизацию правдоподобия циклического перевода и применяется при дообучении готовой модели, что повышает устойчивость процедуры. Эксперименты на англо-финском и русско-казахском направлениях подтверждают полезность подхода в условиях ограниченности параллельных корпусов.

Второй результат связан с переупорядочиванием переводческих гипотез по парным оценкам человека. В работе для этого используются реальные ошибки самой модели. Показано уменьшение числа недостаточных, избыточных и неточных переводов, а также улучшение автоматических метрик качества перевода. Показано, что метод переносится на большие языковые модели и может быть применен в задаче доменной адаптации.

Третий результат состоит в применении маскировки входных и целевых последовательностей. Эксперименты показывают преимущество над авторегрессионной базовой моделью по автоматическим метрикам качества. Подход также допускает добавление в обучение одноязычных данных и обобщается на задачу постредактирования.

Четвёртый результат посвящён тематической сегментации документов. Автор учитывает сегментную структуру при построении аддитивной тематической модели посредством постобработки E-шага. Эффективность метода исследована по метрикам WindowDiff и Pk, а также в задачах информационного поиска и документного перевода. В задаче перевода экспериментально показано, что качество документного перевода по метрике BLEU становится лучше по сравнению с использованием случайной сегментации.

Отмечая научную новизну, значимость и достоверность результатов, хочется отметить наличие теоретического обоснования используемых моделей и экспериментальных подтверждений роста качества с использованием общепринятых

Автореферат последовательно отражает содержание и основные результаты диссертации. 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 по указанной специальности.

Сколковский институт науки и технологий



подпись Панченко А.И.
 подтверждено
 менеджером
 по персоналу Р. Мур Е.В.

