

ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

на диссертационную работу Ватолина Алексея Сергеевича
«Обучение и оценивание мультязычных нейросетевых моделей
семантического векторного представления научных текстов», представленную
на соискание ученой степени кандидата технических наук по специальности

2.3.8 «Информатика и информационные процессы»

В условиях стремительного роста объемов научной информации, актуальность диссертационной работы Ватолина Алексея Сергеевича не вызывает сомнений. Задача эффективной обработки, поиска и анализа научных текстов на разных языках становится все более сложной, что требует разработки новых, более совершенных методов и инструментов. Работа А.С.Ватолина посвящена именно этой важной проблеме и предлагает решения в области семантического векторного представления научных текстов.

А.С.Ватолин начал свою работу в этой области под моим руководством, будучи аспирантом, и с самого начала продемонстрировал глубокий интерес и способности к научной деятельности. За время работы над диссертацией он проявил себя как зрелый и самостоятельный исследователь, способный ставить и решать сложные научные задачи.

Центральным результатом диссертационного исследования является разработка и обучение семейства компактных и эффективных двуязычных моделей SciRus. А.С.Ватолиным была применена двухэтапная методология обучения, включающая маскированное языковое моделирование на большом мультязычном корпусе и последующее контрастивное дообучение на парах «заголовков-аннотация» и на парах с использованием информации о цитировании. Этот подход позволил создать модели, которые, несмотря на значительно меньшее число параметров по сравнению с аналогами, демонстрируют высокий уровень качества на задачах обработки научных текстов, особенно на русском языке. Практическая ценность моделей SciRus подтверждается их успешным внедрением в новый режим «нейропоиска» на

портале eLIBRARY.RU, что упрощает и ускоряет анализ научной информации для широкого круга исследователей.

Для объективной оценки разработанных моделей и их сравнения с существующими решениями А.С.Ватолин создал специальные инструменты. Им была разработана методология и 9 задач для комплексного русско-английского набора задач RuSciBench, который включает в себя различные сценарии использования, такие как классификация, регрессия и поиск. Этот набор задач, основанный на данных российской научной электронной библиотеки, стал первым в своем роде и позволил провести всесторонний анализ качества моделей векторного представления для научных текстов. Высокая значимость этого инструмента подтверждается его интеграцией в международную систему оценки МТЕВ.

Кроме того, А.С.Ватолин создал RuSciFact – первый русскоязычный набор данных для оценки качества верификации научных фактов с помощью моделей векторизации текстов. Разработанная им полуавтоматическая методика формирования данных, которая сочетает генерацию утверждений с помощью больших языковых моделей и последующую экспертную валидацию, представляет собой современный подход к созданию ресурсов для столь сложной и специфической задачи.

А.С.Ватолин активно участвовал в апробации полученных результатов, представляя их на ведущих отечественных и международных конференциях. Его публикации в рецензируемых научных изданиях, индексируемых в Scopus и Web of Science, свидетельствуют о высоком научном уровне исследования.

Диссертационная работа «Обучение и оценивание мультязычных нейросетевых моделей семантического векторного представления научных текстов» является законченным, самостоятельным научным исследованием, содержащим новые, теоретически и практически значимые результаты.

Говоря о личных качествах Ватолина Алексея Сергеевича, следует отметить его инициативность, целеустремленность, ответственность, умение

вести научный поиск, большую работу с источниками на иностранном языке и самое главное — результативность.

Диссертационная работа Ватолина Алексея Сергеевича «Обучение и оценивание мультязычных нейросетевых моделей семантического векторного представления научных текстов» является законченным, оригинальным научным исследованием, отличающимся научной новизной, достаточной теоретической и практической значимостью и может быть представлена к защите в диссертационный совет по специальности 2.3.8 «Информатика и информационные процессы».

Научный руководитель:

д.ф.-м.н., главный научный сотрудник,
руководитель отдела №27 ФИЦ ИУ РАН.

Абгарян Каринэ Карленовна

« 27 » июня 20 25 г.

Подпись Абгарян Каринэ Карленовны заверяю



Р.В. РАЗУМЧИК
по доверенности
1968-105 от 26.09.2024 г.