



Автономная некоммерческая организация
«Институт Искусственного интеллекта»

121170, Москва, Кутузовский пр-т, 32, к. 1, 4.В.08
E-mail: info@airi.net, Web: airi.net
ОГРН 1207700493978 ИНН 7730261209 КПП 773001001

ОТЗЫВ

на автореферат диссертации Ватолина Алексея Сергеевича на тему «Обучение и оценивание мультязычных нейросетевых моделей семантического векторного представления научных текстов», представленной на соискание ученой степени кандидата технических наук по специальности 2.3.8 «Информатика и информационные процессы».

Диссертационная работа А. С. Ватолина посвящена актуальной теме. В условиях быстрого роста числа научных публикаций разработка эффективных инструментов для семантического поиска, анализа и верификации информации становится важной задачей для развития науки. Работа предлагает комплексное решение для создания и оценки специализированных мультязычных моделей для русскоязычного и англоязычного научного домена, что вносит значительный вклад в данную область исследования.

Научный вклад автора состоит в том, что им решен ряд научных и практических задач. К наиболее существенным результатам можно отнести:

1. Разработку семейства компактных двуязычных моделей SciRus для векторного представления научных текстов. Исследована двухэтапная методика обучения, включающая маскированное языковое моделирование и контрастивное дообучение на парах «заголовок–аннотация». Для вариантов SciRus-tiny-cite и SciRus-small-cite дополнительно используются пары «цитирующая статья — цитируемая статья». Эксперименты показывают, что использование данных о цитировании повышает качество представлений, а модель SciRus-small-cite на бенчмарке SciDocs достигает результатов, сопоставимых с более крупной специализированной англоязычной моделью SciNCL.
2. Создание и апробацию первого мультиязычного русско-английского эталонного набора данных RuSciBench для оценки качества моделей на научных текстах. Данный инструментарий, включающий задачи классификации, регрессии и поиска, заполняет существующий пробел в средствах оценки для русскоязычной научной среды и, благодаря интеграции в международную систему оценки МТЕВ, способствует повышению воспроизводимости и стандартизации исследований в данной области.
3. Предложение оригинальной полуавтоматической методики формирования наборов данных для верификации научных фактов на русском языке. На основе этой методики создан первый русскоязычный набор данных для проверки научных фактов RuSciFact, который позволяет оценивать способность моделей к логическому выводу.

Практическая значимость работы подтверждается успешным внедрением модели SciRus-tiny на портале eLibrary.ru, где реализован режим «нейропоиск», что упрощает анализ научной информации для широкого круга исследователей. Разработка и публикация открытых наборов данных и моделей также имеет большую практическую ценность, поскольку стимулирует дальнейшие исследования и разработки в научном сообществе.

Автореферат последователен, структурирован, отличается научным стилем и логичностью изложения. Содержание автореферата и публикаций в полной мере отражает ключевые выводы диссертации.

В качестве замечания, носящего рекомендательный характер, можно отметить, что в автореферате было бы полезно подробнее представить различия в качестве разработанных моделей между отдельными научными дисциплинами. Это позволило бы точнее оценить границы их применимости при неоднородности научного корпуса. Однако данное замечание не снижает общей высокой оценки проделанной работы.

В заключение диссертационная работа Ватолина Алексея Сергеевича представляет собой законченное научное исследование, содержит решение важной научной задачи, обладает научной новизной и практической ценностью. Работа соответствует требованиям Положения о присуждении ученых степеней, а ее автор, Ватолин Алексей Сергеевич, безусловно, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.8 «Информатика и информационные процессы».

Ph.D., ведущий научный сотрудник,
руководитель группы «ИИ в
промышленности» Управления по
исследованиям АНО «Институт
Искусственного интеллекта», доцент
департамента анализа данных и
искусственного интеллекта НИУ ВШЭ
107207, Москва, Чусовская, 11-7-37
iamakarov@hse.ru
+79151524532

 Макаров Илья Андреевич

10 сентября 2026

Подпись рецензента, сотрудника АНО «Институт искусственного интеллекта»,
удостоверяю

Бизнес-партнер по управлению персоналом
АНО «Институт Искусственного интеллекта»



 Трехлетова Ю. А.