

**Письменный отзыв официального рецензента на диссертационную работу Якунина Кирилла Олеговича на тему «Разработка моделей и методов сбора, анализа и классификации медиа-публикаций на базе методов обработки естественных языков», представленную на соискание степени «Доктор филологии» (Ph.D) по специальности 6D070400 – «Вычислительная техника и программное обеспечение»**

| № п/п | Критерии  | Соответствие критериям (необходимо отметить один из вариантов ответа)   | Обоснование позиции официального рецензента  |
|-------|---|---|--|
| 1.    | Тема диссертации (на дату утверждения) соответствует направлениям развития науки и/или государственным программам | <p>1.1 Соответствие приоритетным направлениям развития науки или государственным программам:</p> <p>1) Диссертация выполнена в рамках проекта или целевой программы, финансируемого (ой) из государственного бюджета (указать название и номер проекта или программы)</p> <p>2) Диссертация выполнена в рамках другой государственной программы (указать название программы)</p> <p>3) Диссертация соответствует приоритетному направлению развития науки, утвержденному Высшей научно-технической комиссией при Правительстве Республики Казахстан (указать направление)</p> | <p>Диссертация была выполнена в рамках проекта программно-целевого финансирования КН МОН РК BR05236839 «Разработка информационных технологий и систем для стимулирования устойчивого развития личности как одна из основ развития цифрового Казахстана» в 2018-2020 годах в РП на ПХВ «Институт информационных и вычислительных технологий» КН МОН РК.</p> |
| 2.    | Важность для науки  | <p>Работа вносит/не вносит <u>существенный вклад</u> в науку, а ее <u>важность хорошо раскрыта/не раскрыта</u></p>  | <p>Работа содержит обоснованные новые научные выводы, имеющие как научную, так и практическую ценность. Данные выводы были всесторонне раскрыты в тексте диссертации.</p> <p>Вклад в науку также обосновывается наличием публикаций в рейтинговых рецензируемых журналах, в которых были представлены основные результаты работы.</p>                      |
| 3.    | Принцип самостоятельности   | <p>Уровень самостоятельности: 1) <u>Высокий</u>;</p>  | <p>Соискатель продемонстрировал личный самостоятельный вклад во всех этапах</p>  |

|  |   |  |
|--|---|--|
|  | <p>2) Средний;<br/>3) Низкий;<br/>4) Самостоятельности нет</p>  | <p>диссертационного исследования, начиная с формулирования и обоснования темы исследования, целей и задач, проведения теоретических и экспериментальных исследований. В ходе работы была самостоятельно разработана распределённая информационная система, в рамках которой проводились все основные вычислительный эксперименты, что подтверждается наличием публично-доступного git-репозитория. Опубликовано несколько индексированных работ в рейтинговых журналах, в частности, в которых соискатель фигурирует как первый автор или как автор-корреспондент.</p> |
| <p>4. Принцип внутреннего единства</p> | <p>4.1 Обоснование актуальности диссертации:<br/>1) <u>Обоснование</u>;<br/>2) Частично обоснована;<br/>3) Не обоснована.</p> <p>4.2 Содержание диссертации отражает тему диссертации:<br/>1) <u>Отражает</u>;<br/>2) Частично отражает;<br/>3) Не отражает</p> | <p>Информационные системы для медиа-мониторинга составляют огромные сегмент рынка, который оценивается в сотни миллиардов долларов. В данной работе предложено ряд подходов, которые позволяют решать многие частные задачи анализа и классификации медиа-текстов, что обосновывает актуальность диссертационного исследования.</p> <p>Содержание диссертации полностью отражает тему диссертации – в работе предложены новые модели и методы сбора, анализа и классификации медиа-текстов.</p>  |
|  | <p>4.3 Цель и задачи соответствуют теме диссертации:<br/>1) <u>соответствуют</u>;<br/>2) частично соответствуют;<br/>3) не соответствуют</p>  | <p>Цели и задачи соответствуют теме диссертации.</p>   |

|  |   |  |
|--|---|--|
|  |   |  |
|  | <p>4.4 Все разделы и положения диссертации логически взаимосвязаны:</p> <ol style="list-style-type: none"> <li>1) <u>ПОЛНОСТЬЮ ВЗАИМОСВЯЗАНЫ;</u></li> <li>2) взаимосвязь частичная;</li> <li>3) взаимосвязь отсутствует</li> </ol>   | <p>Текст диссертационной работы обладает внутренним единством, все разделы и положения взаимосвязаны: В-первом разделе рассматривается рынок систем для медиа-мониторинга и их ограничения. Во-втором разделе представлен обзор современных подходов к классификации, которые являются технической основой для систем медиа-мониторинга, описанных в первом разделе. Третий раздел представляет собой описание разработанных докторантом моделей и методов, которые решают обнаруженные проблемы и ограничения, в частности метод ММА. Четвёртый раздел описывает архитектуру ИС, в рамках которой были реализованы предложенные модели и методы. Пятый раздел описывает результаты экспериментов, которые были необходимы для проверки эффективности и валидации предложенных моделей и методов.</p> <p>Таким образом, разделы тесно логически связаны с поставленными задачами исследования, в то же время каждый следующий раздел является логическим продолжением выводов предыдущего.</p> |
|  | <p>4.5 Предложенные автором новые решения (принципы, методы) аргументированы и оценены по сравнению с известными решениями:</p> <ol style="list-style-type: none"> <li>1) <u>критический анализ есть;</u></li> <li>2) анализ частичный;</li> <li>3) анализ представляет собой не собственные мнения, а цитаты других авторов</li> </ol> | <p>Можно выделить три основных новых метода: метод ММА, метод оценки межкорпусного дисбаланса для построения semi-supervised моделей и метод оценки социальной значимости. Эффективность этих методов была аргументирована исходя из результатов сравнения с современными «State of the art» методами (в частности BERG). Методология сравнения отвечает современным требованиям к оценке качества моделей классификации (кросс-валидация, использование метрик F1-Score и ROC AUC).</p>   |

|    |  |  |
|----|--|--|
| 5. | <p>Принцип научной новизны</p> <p>5.1 Научные результаты и положения являются новыми?<br/> 1) полностью новые;<br/> 2) частично новые (новыми являются 25-75%);<br/> 3) не новые (новыми являются менее 25%)</p> <p>5.2 Выводы диссертации являются новыми?<br/> 1) <u>полностью новые</u>;<br/> 2) частично новые (новыми являются 25-75%);<br/> 3) не новые (новыми являются менее 25%)</p> <p>5.3 Технические, технологические, экономические или управленческие решения являются новыми и обоснованными:<br/> 1) <u>полностью новые</u>;<br/> 2) частично новые (новыми являются 25-75%);<br/> 3) не новые (новыми являются менее 25%)</p> | <p>Представленные в работе научные результаты и положения, вынесенные на защиту, являются полностью новыми и были впервые опубликованы автором в рейтинговых рецензируемых международных журналах.</p> <p>Выводы диссертации являются полностью новыми.</p>  |
| 6. | <p>Обоснованность основных выводов</p> <p>Все основные выводы основаны/не основаны на весомых с научной точки зрения доказательствах либо достаточно хорошо обоснованы (для qualitative research и направлений подготовки по искусству и гуманитарным наукам)</p>  | <p>Все основные выводы являются обоснованными. Обоснования основаны на весомых с научной точки зрения доказательствах. В качестве основного подхода, использованного при валидации моделей, можно выделить методику кросс-валидации, являющуюся общепринятой при оценке моделей классификации.</p> |
| 7. | <p>Основные положения, выносимые на защиту</p> <p>Необходимо ответить на следующие вопросы по каждому положению в отдельности:<br/> 7.1 Доказано ли положение?<br/> 1) <u>доказано</u>;</p>  | <p>Все положения, вынесенные на защиту:</p> <p>1) <u>Доказаны</u><br/> 2) <u>Не являются тривиальными</u><br/> 3) <u>Являются новыми</u></p>   |

|   |   |  |
|---|---|--|
|   | <p>2) скорее доказано;<br/>3) скорее не доказано;<br/>4) не доказано</p> <p>7.2 Является ли тривиальным?<br/>1) да;<br/>2) нет</p> <p>7.3 Является ли новым?<br/>1) да;<br/>2) нет</p> <p>7.4 Уровень для применения:<br/>1) узкий;<br/>2) средний;<br/>3) широкий</p> <p>7.5 Доказано ли в статье?<br/>1) да;<br/>2) нет</p> | <p>4) Имеют ли широкие возможности для применения, не ограничивающиеся представленными в работе результатами.<br/>5) Были доказаны (продемонстрированы) в рамках публикации в международных рецензируемых рейтинговых журналах, в том числе индексируемых в Scopus и Web of Science, а также в ходе выступлений на международных научных конференциях.</p> <p>Это можно отнести ко всем основным положениям диссертации:<br/>Предложен метод векторизации текстовых документов с помощью тематической модели <b>BigARTM</b>;<br/>Предложен метод оценки тематического межкорпусного дисбаланса для самообучения классификационной модели;<br/>Предложена методика многофакторной оценки социальной значимости публикации;<br/>Предложена методика многокритериальной оценки масс-медиа ММА на базе байесовской системы агрегации, метода анализа иерархий (АНР) и тематического моделирования;</p> |
| <p>8. Принципы достоверности Доверенность источников и предоставляемой информации</p> | <p>8.1 Выбор методологии – обоснован или методология достаточно подробно описана<br/>1) да;<br/>2) нет</p> <p>8.2 Результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий:</p>                  | <p>Методология исследования обоснована, является общепринятой в научном сообществе для такого рода задач, и была достаточно подробно описана в тексте работы.</p> <p>Да, результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий.</p>  |

|  |  |  |
|--|--|--|
|  | <p>1) да;<br/>2) нет</p>   |  |
|  | <p>8.3 Теоретические выводы, модели, выявленные взаимосвязи и закономерности доказаны и подтверждены экспериментальным исследованием (для направлений подготовки по педагогическим наукам результаты доказаны на основе педагогического эксперимента):<br/>1) да;<br/>2) нет</p>                                 | <p>Да, в работе представлены результаты и анализ результатов нескольких серий вычислительных экспериментов (кросс-валидация моделей классификации), которые всесторонне подтверждают вынесенные на защиту положения и выводы диссертации.</p>  |
|  | <p>8.4 Важные утверждения подтверждены/частично подтверждены/не подтверждены ссылками на актуальную и достоверную научную литературу</p>   | <p>Все важные нетривиальные утверждения подтверждены ссылками на актуальную, достоверную и релевантную научную литературу.</p>   |
|  | <p>8.5 Исползованные источники литературы достаточны/не достаточны для литературного обзора</p>  | <p>В работе представлено 92 ссылки на релевантные научные работы. Исползованные ссылки являются достаточными для литературного обзора.</p>   |
| <p>9. Принципы практической ценности</p> | <p>9.1 Диссертация имеет теоретическое значение:<br/>1) да;<br/>2) нет</p> <p>9.2 Диссертация имеет практическое значение и существует высокая вероятность применения полученных результатов на практике:<br/>1) да;<br/>2) нет</p> <p>9.3 Предложения для практики являются новыми?<br/>1) полностью новые;</p> | <p>9.1 Диссертацией имеет большое теоретическое значение: в ней была доказана возможность использования автоматического тематического моделирования для получения эффективных векторных представлений (embeddings) текстов.</p> <p>9.2 Результаты, полученные в ходе работы, имеют большое практическое значение, и были внедрены в МОН РК в 2020 году. Разработанная в ходе работы ИС фактически является готовым для внедрения продуктом, которые может использоваться в различных областях экономики.</p> |

|  |  |   |
|--|--|---|
|  | <p>2) <u>частично новые</u> (новыми являются 25-75%);<br/>3) не новые (новыми являются менее 25%)</p>              | <p>9.3 Разработанная система Media Analytics позволяет решать большинство типовых задач медиа-мониторинга, которые решаются и другими аналогичными продуктами. Также работа включает предложения новых направлений практического применения: применение автоматического тематического моделирования, оценка межкорпусного дисбаланса для получения (полу-)автоматической разметки и оценка вклада отдельных тем (топиков) на средний показатель заданного критерия (например тональности)</p> |
| <p>10. Качество написания и оформления</p> | <p>Качество академического письма:<br/>1) <u>высокое</u>;<br/>2) среднее;<br/>3) ниже среднего;<br/>4) низкое.</p> | <p>Диссертация написана грамотным научно-техническим, доступным для чтения языком. В диссертационной работе имеются незначительные стилистические ошибки не снижающие качества диссертационной работы.</p>  |

**Заключение.** Проведена качественная научно-исследовательская работа, по результатам которой получены актуальные, новые результаты, имеющие теоретическую и прикладную значимость, разработана методика мультифакторной оценки социальной значимости на базе многокритериальной методики оценки масс-медиа ММА с использованием тематической векторизации текстовых документов. Разработана архитектура и программная реализация распределенной информационно-системы для сбора, обработки, оценки и визуализации текстовых данных масс-медиа и социальных сетей Media Analytics, включающей предложенную методику оценки, а также ее оценку и верификацию. На основании вышеизложенного рекомендую присудить Якунину Кириллу Олеговичу степень доктора философии (Ph.D) специальности 6D070400 – «Высшаяшая техника и программное обеспечение»

PhD, профессор,

АО «Казахстанско-Британский технический университет»

Подпись *А.А. Акжолова*

Акжолова А.Ж.



Департамент  
кадровой работы