

РЕЦЕНЗИЯ ОФИЦИАЛЬНОГО РЕЦЕНЗЕНТА

на диссертационную работу **Тохтахунова Ильмурата Турдымагаметовича** на тему «**Deep Learning Models and Methods for Finding “Similar Audience” in Targeted Advertising**» («Модели и методы глубокого обучения для поиска “похожей аудитории” при таргетированной рекламе»),

представленную на соискание степени доктора философии (PhD) по образовательной программе: 8D06105 — «Наука о данных»

№	Критерии	Соответствие (выберите один вариант ответа)	Обоснование позиции официального рецензента
1.	Тема диссертации (на дату её утверждения) соответствует направлениям научного развития и/или государственным программам	1) Диссертация выполнена в рамках проекта или целевой программы, финансируемой из государственного бюджета (указать название и номер проекта или программы) 2) Диссертация выполнена в рамках другой государственной программы (указать название программы) 3) <u>Диссертация соответствует приоритетному направлению развития наук, утвержденному Высшей научно-технической комиссией при Правительстве Республики Казахстан (указать направление)</u>	Тема диссертации <u>соответствует приоритетному направлению</u> научного развития «Информационные и коммуникационные технологии», утверждённому Высшей научно-технической комиссией при Правительстве Республики Казахстан, а также целям следующих государственных программ: 1. Национальный проект «Цифровой Казахстан», утверждённый постановлением Правительства Республики Казахстан от 12 декабря 2017 года № 827. 2. Стратегический план развития Республики Казахстан до 2025 года, утверждённый Указом Президента Республики Казахстан от 15 февраля 2018 года № 636. 3. Концепция развития науки в Республике Казахстан на 2022–2026 годы, утверждённая постановлением Правительства Республики Казахстан от 19 октября 2021 года № 736. Развитие технологий искусственного интеллекта и принятия решений на основе данных для телекоммуникационного сектора прямо обозначено как приоритет в рамках национальной программы «Цифровой Казахстан».
2.	Важность для науки	Работа <u>вносит</u> / не вносит существенный вклад в науку, и её важность <u>хорошо раскрыта</u> / не раскрыта	Вклад данной работы в науку значителен, а её важность <u>хорошо раскрыта</u> .

			Диссертация вносит значительный вклад в развитие машинного обучения и прикладной науки о данных, в частности в области нелинейного снижения размерности для высокоразмерных гетерогенных табличных данных. Впервые в научной литературе проведена систематическая сравнительная оценка PCA, t-SNE и обучения представлений на основе автокодировщиков применительно к крупномасштабным данным о поведении пользователей телекоммуникаций, охватывающим около 900 000 абонентов и 948 признаков. Предложенный фреймворк многообъектных эмбедингов и адаптивная функция сходства на основе сиамских сетей устраняют практически значимый разрыв между академическими исследованиями обучения представлений и операционными требованиями промышленных рекламных систем.
3.	Принцип самостоятельности	Уровень самостоятельности: 1) <u>высокий</u>; 2) средний; 3) низкий; 4) самостоятельности нет	Уровень самостоятельности — <u>высокий</u>. Диссертационная работа выполнена автором самостоятельно на всех этапах научного исследования. Соискатель самостоятельно осуществил все аспекты: постановку задачи, архитектурное проектирование, реализацию программного конвейера, сбор и предобработку данных, экспериментальную оценку, статистический анализ и подготовку четырёх рецензируемых рукописей.
4.	Принцип внутреннего единства	4.1 Обоснование актуальности диссертации: 1) <u>обоснована</u>; 2) частично обоснована; 3) не обоснована	<u>Обоснована.</u> Актуальность диссертации убедительно аргументирована. Экспоненциальный рост цифровых рекламных платформ, ограничения бинарных классификаторов для отдельных кампаний при масштабном таргетировании аудитории и задокументированная недостаточность линейных методов снижения размерности для гетерогенных поведенческих данных в совокупности обосновывают предлагаемое исследование.

	4.2 Содержание диссертации отражает тему диссертации: 1) <u>отражает</u>; 2) частично отражает; 3) не отражает	<u>Отражает.</u> Содержание всех шести глав точно соответствует заявленной теме. Последовательность от обзора литературы через описание набора данных, формулировку методов снижения размерности, проектирование сямской сети, производственную архитектуру до экспериментальной оценки сохраняет полную внутреннюю согласованность.
	4.3 Цель и задачи соответствуют теме диссертации: 1) <u>соответствуют</u>; 2) частично соответствуют; 3) не соответствуют	<u>Соответствуют.</u> Цель исследования — разработка фреймворка нелинейного снижения размерности, обеспечивающего масштабируемое обнаружение похожей аудитории, — непосредственно реализована всеми пятью задачами исследования, которые логично декомпозируют общую цель на решаемые составляющие.
	4.4 Все разделы и положения диссертации логически взаимосвязаны: 1) <u>полностью взаимосвязаны</u>; 2) связь частичная; 3) связь отсутствует	<u>Полностью взаимосвязаны.</u> Диссертация следует строгой структуре «от проблемы к решению и его проверке». Каждая глава опирается на предыдущую, а все числовые результаты в главе 6 прослеживаются до методов, описанных в главах 3–5.
	4.5 Предложенные автором новые решения (принципы, методы) аргументированы и оценены по сравнению с известными решениями: 1) <u>критический анализ есть</u>; 2) анализ частичный; 3) анализ представляет собой не собственные мнения, а цитаты других авторов;	<u>Критический анализ есть.</u> Предложенные эмбединги на основе автокодировщика систематически сравниваются с PCA, t-SNE, SVM, случайным лесом, LightGBM, косинусным сходством для одного объекта и косинусным сходством для нескольких объектов по пяти независимым метрикам оценки, а статистическая значимость подтверждена с помощью критерия знаковых рангов Вилкоксона.

		4) анализ отсутствует	
5.	Принцип научной новизны	5.1 Являются ли научные результаты и положения новыми? 1) <u>полностью новые</u>; 2) частично новые (новыми являются 25–75%); 3) не новые (новыми являются менее 25%)	Все основные положения и научные результаты диссертации являются <u>полностью новыми</u> и оригинальными. Новые научные результаты включают: — Архитектуру стекового автокодировщика (948→1000→1000→288→1000→1000→948) с формальным математическим описанием в рамках подхода Хайема и Хайема (2018), включая лемму об обратном распространении ошибки с полным доказательством, обеспечивающую строгое теоретическое обоснование архитектурных решений; — Стратегию многообъектных эмбедингов, обучающую шесть предметно-ориентированных автокодировщиков и конкатенирующую их латентные представления, с достижением Lift Top 1 = 11,7 против 6,6 у наилучшего базового классификатора (+77,3%); — Применение сиамских сетей к табличным поведенческим эмбедингам с обучением на контрастивных потерях на реальных данных рекламных кампаний, с достижением Lift Top 1 = 12,9, F1 = 0,75, ROC AUC = 0,79, CR = 0,36; — Модульную архитектуру производственной системы, валидированную на 900 000 абонентах с использованием HDFS, DVC, MLflow и двумя режимами инференса.
		5.2 Являются ли выводы диссертации новыми? 1) <u>полностью новые</u>; 2) частично новые (новыми являются 25–75%); 3) не новые (новыми являются менее 25%)	Выводы диссертации являются <u>полностью новыми</u> и логически обоснованными экспериментальными результатами.

		5.3 Технические, технологические, экономические или управленческие решения являются новыми и обоснованными: 1) <u>полностью новые</u>; 2) частично новые (новыми являются 25–75%); 3) не новые (новыми являются менее 25%)	Технические решения являются <u>полностью новыми</u>, подтверждёнными четырьмя публикациями в международных журналах, индексируемых Scopus, и валидацией на реальных промышленных данных.
6.	Обоснованность основных выводов	Все основные выводы <u>основаны</u> / не основаны на весомых с научной точки зрения доказательствах либо достаточно хорошо обоснованы (для qualitative research и направлений подготовки по искусству и гуманитарным наукам).	Выводы, сформулированные в диссертации, <u>научно обоснованы</u>, поскольку получены на основе авторских исследований, разработанных моделей, алгоритмов и эмпирических данных, собранных и обработанных в рамках данной работы. Ключевые показатели эффективности — Silhouette Score 0,48, kNN F1 = 0,71, Lift Top 1 = 12,9, Conversion Rate = 0,36 — получены в результате оценки на 17 независимых задачах валидации кампаний с внешними целевыми переменными. Статистическая значимость всех заявленных различий подтверждена с помощью критерия знаковых рангов Вилкоксона ($p < 0,01$ для основных сравнений). Выводы не воспроизводят известные результаты и представляют собой оригинальный научный вклад.
7.	Основные положения, выносимые на защиту	Необходимо ответить на следующие вопросы по каждому положению в отдельности: 7.1 Доказано ли положение? 1) <u>доказано</u>; 2) скорее доказано; 3) скорее не доказано; 4) не доказано;	7.1 Все четыре основных положения, выносимых на защиту, <u>доказаны</u>. Согласно первому положению, архитектура стекового автокодировщика строго специфицирована посредством полной пословной нотации отображения, а её превосходство над PCA и t-SNE продемонстрировано по пяти независимым метрикам кластеризации на валидационном наборе данных из 900 000 абонентов.

		5) в текущей формулировке проверить доказанность положения невозможно 7.2 Является ли тривиальным? 1) <u>да</u>; 2) <u>нет</u>; 3) в текущей формулировке проверить тривиальность положения невозможно 7.3 Является ли новым? 1) <u>да</u>; 2) <u>нет</u>; 3) в текущей формулировке проверить новизну положения невозможно 7.4 Уровень для применения: 1) узкий; 2) средний; 3) <u>широкий</u>; 4) в текущей формулировке проверить уровень применения положения невозможно 7.5 Доказано ли в статье? 1) <u>да</u>; 2) <u>нет</u>; 3) в текущей формулировке проверить доказанность положения в статье невозможно	Согласно второму положению, доказано, что стратегия многообъектных эмбедингов превосходит однообъектные представления на 60% по Lift Top 1 (11,7 против 7,3) на 17 независимых оценках кампаний. Согласно третьему положению, сямская сеть с контрастивными потерями демонстрирует стабильные улучшения по сравнению с косинусным сходством (+10,3% по Lift Top 1) со статистической значимостью, подтверждённой по всем 17 кампаниям. Согласно четвёртому положению, модульная производственная архитектура валидирована через успешную обработку 900 000 абонентов в рамках операционных временных ограничений. 7.2 Положения <u>не являются тривиальными</u>. Стратегия многообъектной конкатенации и применение сямских сетей к табличным поведенческим эмбедингам являются оригинальным вкладом, ранее не описанным в литературе. 7.3 Положения являются <u>новыми</u>, подтверждёнными четырьмя публикациями, индексируемыми Scopus. 7.4 Уровень применения — <u>широкий</u>. Методология применима к любому крупномасштабному набору данных о поведении пользователей, включая сферы здравоохранения, финансов и электронной коммерции. 7.5 <u>Доказано в статьях</u>: четыре рецензируемые публикации в Procedia Computer Science, Engineering Technology & Applied Science Research, Engineered Science и поданная в 2026 году рукопись.
8.	Принцип достоверности. Достоверность	8.1 Выбор методологии - обоснован или методология достаточно подробно описана:	Методология исследования <u>обоснована и описана детально</u>. Архитектура автокодировщика специфицирована с полным математическим формализмом, включая размерности слоёв,

источников и предоставляемой информации	1) <u>да</u> ; 2) нет	функции активации, уравнения батч-нормализации, функцию стоимости MSE, лемму об обратном распространении ошибки с доказательством и уравнения обновления оптимизатора Adam в рамках подхода Хайема и Хайема (2018).
	8.2 Результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий: 1) <u>да</u> ; 2) нет	Результаты получены с использованием <u>современных методов</u> научных исследований: PyTorch 1.12 для глубокого обучения, PySpark 3.3 для распределённой предобработки, MLflow для отслеживания экспериментов и критерий знаковых рангов Вилкоксона для оценки статистической значимости.
	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): 1) <u>да</u> ; 2) нет	Теоретические выводы <u>доказаны и подтверждены</u> экспериментальными результатами на реальных промышленных данных. Лемма об обратном распространении ошибки формально доказана. Все результаты кластеризации и обнаружения похожей аудитории получены на валидационном наборе данных без утечки информации из обучающей выборки.
	8.4 Важные утверждения <u>подтверждены</u> / частично подтверждены / не подтверждены ссылками на актуальную и достоверную научную литературу.	Литература актуальна и достоверна.
	8.5 Используемые источники литературы <u>достаточны</u> / не достаточны для литературного обзора	Литературные источники <u>достаточны</u> для обзора литературы, охватывающего снижение размерности, обучение

			представлений, моделирование похожей аудитории, сиамские сети и производственные системы машинного обучения.
9.	Принцип практической ценности	9.1 Диссертация имеет теоретическое значение: 1) <u>да</u> ; 2) нет	Диссертация <u>имеет теоретическое значение</u> , поскольку обеспечивает строгое математическое рассмотрение стековых автокодировщиков для табличных данных, включая лемму об обратном распространении ошибки, формально расширяет подход Хайема и Хайема до 7-слойной архитектуры реконструкции и устанавливает эмпирически валидированную методологию многообъектного обучения представлений.
		9.2 Диссертация имеет практическое значение и существует высокая вероятность применения полученных результатов на практике: 1) <u>да</u> ; 2) нет	Диссертация <u>имеет практическое значение</u> . Предложенный фреймворк устраняет необходимость обучения моделей для отдельных кампаний, снижает трудозатраты аналитиков и улучшает коэффициент конверсии с 0,13 до 0,36 — улучшение на 177% — что подтверждено реальными рекламными кампаниями на телекоммуникационной платформе. Модульная производственная система успешно развёрнута для 900 000 абонентов.
		9.3 Предложения для практики являются новыми: 1) <u>полностью новые</u> ; 2) частично новые (новыми являются 25–75%); 3) не новые (новыми являются менее 25%)	Практические рекомендации являются <u>полностью новыми</u> . Сервис многообъектных эмбедингов с автоматическим обнаружением похожей аудитории не имеет прямых аналогов в опубликованной литературе для телекоммуникационной отрасли.
10.	Качество написания и оформления	Качество академического письма: 1) <u>высокое</u> ; 2) среднее; 3) ниже среднего;	Качество академического письма в диссертации является <u>высоким</u> и соответствует требованиям к научным исследованиям на уровне PhD. Изложение материала логично, последовательно и точно в формулировках. Математические обозначения используются корректно и единообразно на протяжении всей

		4) низкое	работы. Все положения обоснованы, выводы аргументированы, рисунки и таблицы чётко подписаны и непосредственно подкрепляют изложение. Формальный стиль леммы и доказательства математического содержания свидетельствует о научной зрелости.
11.	Замечания к диссертации		<i>Исследование по абляции латентной размерности (Таблица А.1) охватывает значения {64, 96, 128, 160, 192, 256, 288, 320}, однако не включает очень малых значений ($k < 64$) или значений, превышающих 320. Включение крайних значений диапазона усилило бы полноту анализа.</i> <i>Результаты по отдельным кампаниям в Приложении В демонстрируют более низкий Lift для кампаний туристической тематики (в среднем 12,1) по сравнению с финансовыми кампаниями (в среднем 13,3). Краткий анализ в основном тексте, связывающий данную вариативность с вкладом отдельных объектов, стал бы ценным дополнением.</i>
12.	Научный уровень статей докторанта по теме исследования (в случае защиты диссертации в форме серии статей официальные рецензенты комментируют научный уровень каждой статьи докторанта по теме исследования)		Научный уровень публикаций докторанта полностью соответствует требованиям к диссертационным работам PhD.

13.	Решение официального рецензента (согласно пункту 28 настоящего Типового положения)		Ходатайствовать перед Комитетом для присуждения докторанту степени доктора философии (PhD)
-----	---	--	--

Заключение

Представленная диссертация на соискание степени доктора философии (PhD) **Тохтаунова Ильмурата Турдымагаметовича** на тему «**Deep Learning Models and Methods for Finding “Similar Audience” in Targeted Advertising**» («Модели и методы глубокого обучения для поиска “похожей аудитории” при таргетированной рекламе») отвечает требованиям «Правил присуждения степеней», а ее автор заслуживает ходатайствовать перед Комитетом для присуждения докторанту степени доктора философии (PhD).

Официальный рецензент:

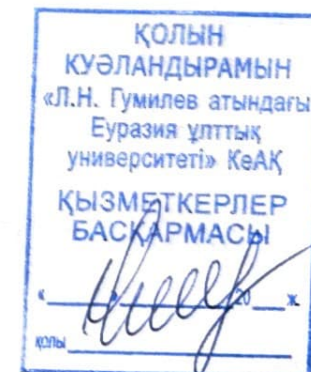
Директор НИИ «Искусственный Интеллект»

НАО «Евразийский национальный университет им. Л.Н. Гумилёва»,

PhD, ассоциированный профессор.



Барлыбаев Алибек Бактыбаевич



REVIEW OF THE OFFICIAL REVIEWER

on the dissertation work of **Tokhtakhunov Il'murat Turdymagametovich** on the topic «**Deep Learning Models and Methods for Finding “Similar Audience” in Targeted Advertising**»,

provided for the degree of Doctor of Philosophy (PhD) in the educational program: 8D06105 – «Data Science»

№	Criteria	Eligibility (please select one answer option)	Justification of the official reviewer's position
1.	The topic of the thesis (as of the date of its approval) corresponds to the directions of scientific development and/or government programs	Compliance with priority areas of scientific development or government programs: 1) The dissertation was completed within the project or target program financed from the state budget (indicate the name and number of the project or program) 2) The dissertation was completed within the other state program (indicate the name of the program) 3) <u>The thesis corresponds to the priority direction of scientific development, approved by the Higher Scientific and Technical Commission under the Government of the Republic of Kazakhstan (indicate direction)</u>	The dissertation topic <u>corresponds to the priority area</u> of scientific development «Information and Communication Technologies» approved by the Higher Scientific and Technical Commission under the Government of the Republic of Kazakhstan, as well as to the objectives of the following state programs: 1. The National Project «Digital Kazakhstan», approved by the Resolution of the Government of the Republic of Kazakhstan dated December 12, 2017, No. 827. 2. The Concept for the Development of Science in the Republic of Kazakhstan for 2022–2026, approved by the Resolution of the Government of the Republic of Kazakhstan dated October 19, 2021, No. 736. The development of artificial intelligence and data-driven decision-making technologies for the telecommunications sector is explicitly identified as a priority within the «Digital Kazakhstan» national program.
2.	Importance for science	The work <u>makes</u> / does not make a significant contribution to science and its importance is <u>well-disclosed</u> / not disclosed	The contribution of this work to science <u>is significant</u>, and its importance is <u>well-disclosed</u>. The dissertation makes a significant contribution to the development of machine learning and applied data science, particularly in the area of nonlinear dimensionality reduction for high-dimensional

			heterogeneous tabular data. For the first time in the scientific literature, a systematic comparative evaluation of PCA, t-SNE, and autoencoder-based representation learning is conducted specifically for large-scale telecommunications user behavioral data, comprising approximately 900,000 subscribers and 948 features. The proposed multi-entity embedding framework and Siamese network-based adaptive similarity function address a practically important gap between academic representation learning research and the operational requirements of industrial advertising systems.
3.	The principle of self-reliance	Self-reliance level: 1) <u>high</u> ; 2) average; 3) low; 4) no independence	Level of independence – <u>high</u> . The dissertation work was completed by the author independently at all stages of scientific research. The candidate carried out all aspects independently: problem formulation, architectural design, implementation of the software pipeline, data collection and preprocessing, experimental evaluation, statistical analysis, and preparation of four peer-reviewed manuscripts.
4.	The principle of internal unity	4.1 Justification of the relevance of the thesis: 1) <u>justified</u> ; 2) partially justified; 3) not justified	<u>Justified</u> . The relevance of the dissertation is convincingly justified. The exponential growth of digital advertising platforms, the limitations of campaign-specific binary classifiers for large-scale audience targeting, and the documented inadequacy of linear dimensionality reduction methods for heterogeneous behavioral data collectively motivate the proposed research.
		4.2 The content of the thesis reflects its topic: 1) <u>reflects</u> ; 2) partially reflects; 3) does not reflect	<u>Reflects</u> . The content of all six chapters corresponds precisely to the stated topic. The progression from literature review through dataset description, dimensionality reduction method formulation, Siamese

			network design, production architecture, and experimental evaluation maintains full internal coherence.
		4.3 The purpose and objectives correspond to the topic of the thesis: 1) <u>correspond</u> ; 2) partially correspond; 3) do not correspond	<u>Correspond.</u> The research aim — to develop a nonlinear dimensionality reduction framework enabling scalable look-alike audience detection — is directly addressed by all five research objectives, which logically decompose the overall goal into tractable components.
		4.4 All sections and provisions of the dissertation are logically interconnected: 1) <u>completely interconnected</u> ; 2) the relationship is partial; 3) there is no relationship	<u>Completely interconnected.</u> The dissertation follows a rigorous «from problem to solution to validation» structure. Each chapter builds upon the preceding one, and all numerical results in Chapter 6 can be traced to the methods described in Chapters 3–5.
		4.5 The new solutions (principles, methods) proposed by the author are argued and evaluated in comparison with known solutions: 1) <u>there is a critical analysis</u> ; 2) partial analysis; 3) the analysis does not represent one's own opinions, but quotes from other authors 4) no analysis	<u>There is a critical analysis.</u> The proposed autoencoder-based embeddings are systematically compared against PCA, t-SNE, SVM, Random Forest, LightGBM, single-entity cosine similarity, and multi-entity cosine similarity baselines across five independent evaluation metrics, with statistical significance confirmed via Wilcoxon signed-rank tests.
5.	The principle of scientific novelty	5.1 Are the scientific results and provisions new? 1) <u>completely new</u> ;	All main provisions and scientific results presented in the dissertation are <u>completely new</u> and original. The new scientific results include: — A stacked autoencoder architecture (948→1000→1000→288→1000→1000→948) with formal

		2) partially new (25–75% are new); 3) not new (less than 25% are new)	mathematical description following the Higham & Higham (2018) framework, including a Backpropagation Lemma with complete proof, providing a rigorous theoretical foundation for the architectural choices; — A multi-entity embedding strategy training six domain-specific autoencoders and concatenating their latent representations, achieving Lift Top 1 = 11.7 vs 6.6 for the best classifier baseline (+77.3%); — The application of Siamese networks to tabular behavioral embeddings with contrastive loss training on real advertising campaign data, achieving Lift Top 1 = 12.9, F1 = 0.75, ROC AUC = 0.79, CR = 0.36; — A modular production system architecture validated on 900,000 subscribers with HDFS, DVC, MLflow, and dual inference modes.
		5.2 Are the conclusions of the dissertation new? 1) <u>completely new</u>; 2) partially new (25–75% are new); 3) not new (less than 25% are new)	The conclusions of the dissertation are <u>completely new</u> and logically substantiated by the experimental results.
		5.3 Technical, technological, economic, or managerial decisions are new and justified: 1) <u>completely new</u>; 2) partially new (25–75% are new); 3) not new (less than 25% are new)	The technical solutions are <u>completely new</u>, confirmed by four publications in Scopus-indexed international journals and validation on real industrial data.
6.	Validity of the main conclusions	All key findings <u>are based</u> / are not based on scientifically sound evidence or are reasonably well funded (for qualitative research and areas of training in the arts and humanities)	The conclusions formulated in the dissertation are <u>scientifically substantiated</u>, since they are obtained on the basis of the author's research, developed models, algorithms and empirical data collected and processed within the framework of this work.

			The key performance results — Silhouette Score 0.48, kNN F1 = 0.71, Lift Top 1 = 12.9, Conversion Rate = 0.36 — are derived from evaluation on 17 independent campaign validation tasks with external target variables. Statistical significance of all reported differences is confirmed via Wilcoxon signed-rank tests ($p < 0.01$ for primary comparisons). The conclusions do not reproduce known results and represent original scientific contributions.
7.	Main provisions submitted for defense	The following questions must be answered for each provision separately: 7.1 Has the position been proven? 1) <u>proven</u>; 2) rather proven; 3) rather not proven; 4) not proven 7.2 Is it trivial? 1) yes; 2) <u>no</u> 7.3 Is it new? 1) <u>yes</u>; 2) no 7.4 Application level: 1) narrow; 2) medium; 3) <u>wide</u> 7.5 Is it proven in the article? 1) <u>yes</u>; 2) no	7.1 All four main provisions submitted for defense are <u>proven</u>. According to the first provision, the stacked autoencoder architecture is rigorously specified through complete layer-by-layer mapping notation, and its superiority over PCA and t-SNE is demonstrated across five independent clustering metrics on a 900,000-subscriber validation dataset. According to the second provision, the multi-entity embedding strategy is proven to outperform single-entity representations by 60% in Lift Top 1 (11.7 vs 7.3) across 17 independent campaign evaluations. According to the third provision, the Siamese network with contrastive loss demonstrates consistent improvements over cosine similarity (+10.3% in Lift Top 1) with statistical significance confirmed across all 17 campaigns. According to the fourth provision, the modular production architecture is validated through successful processing of 900,000 subscribers within operational time constraints. 7.2 The provisions are <u>not trivial</u>. The multi-entity concatenation strategy and Siamese network application to tabular behavioral embeddings are original contributions not previously reported in the literature. 7.3 The provisions are <u>new</u>, confirmed by four Scopus-indexed publications.

			7.4 Application level — <u>wide</u>. The methodology is applicable to any large-scale user behavioral dataset, including healthcare, finance, and e-commerce domains. 7.5 <u>Proven in articles</u>: four peer-reviewed publications in Procedia Computer Science, Engineering Technology & Applied Science Research, Engineered Science, and a submitted 2026 manuscript.
8.	The principle of reliability. Reliability of sources and information provided	8.1 The choice of methodology – is the methodology justified or is it described in sufficient detail? <u>1) yes;</u> 2) no	The research methodology is <u>justified and described in detail</u>. The autoencoder architecture is specified with complete mathematical formalism including layer dimensions, activation functions, batch normalization equations, MSE cost function, backpropagation lemma with proof, and Adam optimizer update equations following the Higham & Higham (2018) framework.
		8.2 The results of the thesis were obtained using modern methods of scientific research and techniques for processing and interpreting data using computer technologies: <u>1) yes;</u> 2) no	The results are <u>obtained using modern methods</u> of scientific research: PyTorch 1.12 for deep learning, PySpark 3.3 for distributed preprocessing, MLflow for experiment tracking, and Wilcoxon signed-rank tests for statistical significance assessment.
		8.3 Theoretical conclusions, models, identified relationships and patterns are proven and confirmed by experimental research: <u>1) yes;</u> 2) no	Theoretical conclusions are <u>proven and confirmed</u> by experimental results on real industry data. The Backpropagation Lemma is formally proved. All clustering and look-alike detection results are derived from the validation dataset with no information leakage from training.
		8.4 Important statements are <u>supported</u> / partially supported / not supported by	The literature is relevant and reliable.

		references to current and reliable scientific literature	
		8.5 The literature sources used are <u>sufficient</u> / are not sufficient for the literature review	The literature sources are <u>sufficient</u> for the literature review, covering dimensionality reduction, representation learning, look-alike modeling, Siamese networks, and production machine learning systems.
9.	Principle of practical value	9.1 The dissertation has theoretical significance: 1) <u>yes</u> ; 2) no	The dissertation <u>has theoretical significance</u> , as it provides a rigorous mathematical treatment of stacked autoencoders for tabular data including a Backpropagation Lemma, formally extends the Higham & Higham framework to a 7-layer reconstruction architecture, and establishes an empirically validated methodology for multi-entity representation learning.
		9.2 The thesis has practical significance and there is a high probability of applying the obtained results in practice: 1) <u>yes</u> ; 2) no	The thesis <u>has practical significance</u> . The proposed framework eliminates campaign-specific model training, reduces analyst labor costs, and improves targeting Conversion Rate from 0.13 to 0.36 — a 177% improvement — as validated through actual advertising campaigns on a telecommunications platform. The modular production system has been successfully deployed for 900,000 subscribers.
		9.3 Are the suggestions for practice new? 1) <u>completely new</u> ; 2) partially new (25–75% are new); 3) not new	The suggestions for practice are <u>completely new</u> . The multi-entity embedding service with automated look-alike detection has no direct analogues in published literature for the telecommunications domain.
10.	Quality of writing and design	Quality of academic writing: 1) <u>high</u> ;	The quality of academic writing in the dissertation is <u>high</u> and meets the requirements for scientific research at the PhD level. The

		2) average; 3) below average; 4) low	presentation of material is logical, consistent, and precise in its formulations. Mathematical notation is used correctly and consistently throughout. All provisions are substantiated, conclusions are justified, and figures and tables are clearly labeled and directly support the narrative. The formal Lemma-and-proof style of the mathematical content demonstrates scientific maturity.
11.	Comments and suggestions on the dissertation		<i>The ablation study for latent dimensionality (Table A.1) covers values {64, 96, 128, 160, 192, 256, 288, 320} but does not include very small values ($k < 64$) or values exceeding 320. Including the extreme ends of the range would strengthen the completeness of the analysis.</i> <i>The per-campaign results in Appendix B show lower Lift for travel-type campaigns (avg 12.1) vs financial campaigns (avg 13.3). A brief analysis in the main text connecting this variance to entity contribution would be a valuable addition.</i>
12.	The scientific level of the doctoral candidate's articles on the research topic (in the case of defending the dissertation in the form of a series of articles, the official reviewers will comment on the scientific level of each article by the doctoral candidate on the research topic).		The scientific level of the doctoral candidate's publications fully complies with the requirements for PhD dissertation work.

13.	The decision of the official reviewer (according to paragraph 28 of the current Model Regulation)		To petition the Committee for the award of the degree of Doctor of Philosophy (PhD)
-----	---	--	---

Conclusion

The presented dissertation for the degree of Doctor of Philosophy (PhD) by **Tokhtakhunov Il'murat Turdymagametovich** on the topic «**Deep Learning Models and Methods for Finding “Similar Audience” in Targeted Advertising**» meets the requirements of the Rules, the author deserves to petition the Committee for the award of the degree of Doctor of Philosophy (PhD).

Official reviewer:

Director of the Scientific Research Institute «Artificial Intelligence»
L.N. Gumilyov Eurasian National University,
PhD, Associate Professor



Alibek Barlybayev

