

REVIEW

on the PhD thesis by **Tokhtakhunov Il'murat Turdymagametovich** entitled «Deep Learning Models and Methods for Finding "Similar Audience" in Targeted Advertising» submitted for the degree of Doctor of Philosophy (PhD) in the EP 8D06105 – «Data Science»

This PhD thesis investigates deep learning-based representation learning methods for automated look-alike audience detection in targeted digital advertising systems. The research addresses a practically significant problem faced by modern telecommunications operators: the need to identify users who exhibit behavioral characteristics similar to a predefined reference audience, enabling efficient and scalable B2B advertising services without requiring campaign-specific model retraining.

The core computer science methods applied in this work — stacked autoencoders, Siamese neural networks, contrastive metric learning, and distributed data processing — are state-of-the-art tools in the field of machine learning and data science. Their systematic application to heterogeneous tabular telecommunications data, comprising approximately 900,000 subscriber records and 948 features organized across six behavioral entity domains, demonstrates both methodological rigor and practical relevance.

Il'murat Tokhtakhunov carried out a comprehensive series of experiments covering all major stages of the research pipeline: data preprocessing and feature engineering, autoencoder architecture design and training, embedding quality evaluation using clustering metrics, look-alike audience retrieval benchmarking, and production system architecture validation. The experimental results — including a Lift Top 1 score of 12.9, ROC AUC of 0.79, and Conversion Rate of 0.36 validated across 17 independent advertising campaign datasets — convincingly demonstrate the effectiveness and practical applicability of the proposed framework.

Throughout the research process, Il'murat Tokhtakhunov demonstrated a high level of scientific maturity, systematic thinking, and genuine commitment to producing results of industrial-grade quality. His ability to bridge the gap between theoretical machine learning research and the operational constraints of production data engineering environments is particularly noteworthy and reflects well on the depth of his domain understanding.

The doctoral candidate demonstrated a high degree of academic independence at every stage of the work. He was not only able to identify and clearly articulate the research problem but also to independently design and implement the full computational pipeline — from raw data extraction through anonymization-compliant preprocessing to the deployment of a dual-mode inference architecture using Apache Spark and Pandas. All research stages were completed in strict accordance with scientific standards, within the agreed timeline, and without requiring detailed guidance on methodological choices.

Il'murat Tokhtakhunov consistently showed initiative in exploring new methodological directions. In particular, the decision to apply Siamese network contrastive learning to autoencoder-derived tabular embeddings — adapting an approach established in image and text domains to the specific characteristics of telecommunications behavioral data — reflects

original scientific thinking. He independently designed the experimental evaluation protocol, including the multi-task validation procedure across 17 prospective advertising campaigns, which provides a more rigorous and unbiased performance estimate than is typically found in comparable applied machine learning research.

The doctoral candidate has consistently maintained high standards of scientific ethics and academic professionalism. His approach to research is distinguished by careful attention to reproducibility — evidenced by the use of data version control (DVC), experiment tracking (MLflow), and artifact management (MinIO) — as well as by a commitment to formal mathematical rigor, including a formally proved Backpropagation Lemma for the stacked autoencoder architecture. He has published the results of this research as the lead author in internationally ranked journals indexed in Scopus, including one Q1 publication (Engineered Science, CiteScore 12.5, Percentile 97%), two Q2 publications (Procedia Computer Science; Engineering Technology & Applied Science Research), and a fourth manuscript currently under review. The results have also been presented at the Department of Mathematical and Computer Modelling at the International Information Technology University and at the School of Digital Technologies, Narxoz University.

Collaboration with Il'murat Tokhtakhunov has been highly productive throughout the doctoral programme. He combines strong technical skills with clear scientific communication and a consistent orientation toward results that are simultaneously novel, rigorous, and practically deployable. I am confident that he possesses all of the qualities necessary for the successful completion of this doctoral work and for a productive subsequent scientific and professional career.

I believe that the PhD thesis «Deep Learning Models and Methods for Finding "Similar Audience" in Targeted Advertising», submitted for the degree of Doctor of Philosophy (PhD) in the EP 8D06105 «Data Science» is a completed independent research work that meets all the requirements for a PhD thesis. The author **Tokhtakhunov Il'murat Turdymagametovich** deserves to be awarded the academic degree of Doctor of Philosophy (PhD) in the EP: 8D06105 – «Data Science».

Scientific consultant

Associate Professor, PhD

Department of Mathematical and Computer Modelling,
International Information Technology University


Nurtas Marat

