НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
март–апрель 2026 Том 26 № 2 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
March–April 2026 Vol. 26 No 2 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-2-337-348
УДК 004.855.5

Многозадачный анализ психологического портрета человека на основе текстовых данных с применением полуконтролируемого обучения

Дарья Олеговна Коряковская[1], Александр Александрович Аксёнов[2]✉, Елена Витальевна Рюмина[3], Дмитрий Александрович Рюмин[4]

[1,2,3,4] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

[1] da.koryakovskaya@gmail.com, https://orcid.org/0009-0009-6207-8413
[2] a.aksenov95@mail.ru✉, https://orcid.org/0000-0002-7479-2851
[3] ryumina_ev@mail.ru, https://orcid.org/0000-0002-4135-6949
[4] dl_03.03.1991@mail.ru, https://orcid.org/0000-0002-7935-0569

Аннотация

Введение. Многозадачный анализ психологического портрета человека позволяет формировать более целостное представление о нем, что особенно востребовано в системах персонализации, HR-технологиях и человеко-машинном взаимодействии. Однако до настоящего времени подобные исследования не проводились из-за отсутствия корпусов с совместной разметкой по обеим задачам, что делает невозможным традиционное многозадачное обучение. **Метод.** Предложен метод полуконтролируемого кросс-доменного обучения, позволяющий эффективно интегрировать два раздельно аннотированных корпуса: CMU Multimodal Opinion Sentiment and Emotion Intensity (CMU-MOSEI) (для распознавания эмоций) и ChaLearn First Impressions v2 (FIv2) (для оценивания личностных характеристик), без дополнительной разметки. Экспериментальная установка включает два этапа: обучение независимых однозадачных моделей для извлечения доменно-специфичных признаков и формирование базовых прогнозов; создание совместной кросс-доменной модели с блоками перекрестного внимания, которая объединяет эмоциональные и личностные признаки. Финальное предсказание формируется путем усреднения выходов однозадачных и совместной моделей, что повышает робастность. Выполнено сравнение предобученных энкодеров (Jina-v3 и BGE-en) и контекстных моделей (трансформер и Mamba). Обучение моделей осуществлено с использованием гибридной функции потерь, сочетающей контролируемые и полуконтролируемые компоненты с псевдометками. **Основные результаты.** Эксперименты показали, что наилучшие результаты достигаются при использовании энкодера Jina-v3 и контекстной модели Mamba: средняя взвешенная точность классификации (mWACC) составила 62,52 %, а средняя взвешенная F1-мера (mMF1) — 61,03 % на корпусе CMU-MOSEI; средняя точность (mACC) составила 88,80 %, а средний коэффициент корреляции конкордации Лина (mCCC) — 25,44 % на FIv2. Модель демонстрирует устойчивую передачу знаний между задачами и превосходит современные решения. Визуализация внимания методом Gradient-weighted Class Activation Mapping подтверждает интерпретируемость прогнозов. **Обсуждение.** Представленные результаты исследования открывают возможности разработки масштабируемых систем психологического профилирования по тексту в условиях дефицита разметки. Предложенный метод применим в кадровом менеджменте, адаптивных обучающих платформах, персонализированных чат-ботах и цифровой психометрике, где требуются одновременный учет эмоционального состояния и устойчивых личностных характеристик.

Ключевые слова

распознавание эмоций, распознавание личностных характеристик, многозадачное обучение, кросс-доменное машинное обучение, полуконтролируемое машинное обучение, визуализации внимания модели

Благодарности

Раздел «Обзор литературы» выполнен в рамках бюджетной темы СПб ФИЦ РАН (№ FFZF-2025-0003), разработка компонентов полуконтролируемого машинного обучения, включая гибридную функцию потерь и механизм псевдометок, проведена при поддержке Российского научного фонда (проект № 24-71-00083), а проектирование архитектуры кросс-доменной модели с блоками перекрестного внимания и интеграцией контекстных моделей — при поддержке Российского научного фонда (проект № 24-71-00112).

Ссылка для цитирования: Коряковская Д.О., Аксёнов А.А., Рюмина Е.В., Рюмин Д.А. Многозадачный анализ психологического портрета человека на основе текстовых данных с применением полуконтролируемого обучения // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 2. С. 337–348. doi: 10.17586/2226-1494-2026-26-2-337-348

Multi-task human's psychological profile analysis based on text data using semi-supervised learning

Darya O. Koryakovskaya[1], Alexandr A. Axyonov[2]✉, Elena V. Ryumina[3], Dmitry A. Ryumin[4]

[1,2,3,4] St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation

[1] da.koryakovskaya@gmail.com, https://orcid.org/0009-0009-6207-8413
[2] a.aksenov95@mail.ru✉, https://orcid.org/0000-0002-7479-2851
[3] ryumina_ev@mail.ru, https://orcid.org/0000-0002-4135-6949
[4] dl_03.03.1991@mail.ru, https://orcid.org/0000-0002-7935-0569

Abstract

Multi-task analysis of a human's psychological profile enables a more holistic representation of the individual, which is particularly valuable in personalization systems, HR technologies, and human–Artificial Intelligence interaction. However, such studies have not been conducted to date due to the lack of datasets jointly annotated for both emotion and personality traits, rendering conventional multi-task learning infeasible. We propose a semi-supervised cross-domain learning method that effectively integrates two separately annotated corpora, CMU-MOSEI (for emotion recognition) and ChaLearn First Impressions v2 (FIv2) (for personality trait assessment), without requiring additional labeling. The experimental setup comprises two stages: first, independent single-task models are trained to extract domain-specific features and generate baseline predictions; second, a joint cross-domain model with cross-attention blocks fuses emotional and personality-related representations. Final predictions are obtained by averaging the outputs of the single-task and joint models, enhancing robustness. We compare pre-trained encoders (Jina-v3 and BGE-en) and contextual decoders (Transformer and Mamba), using a hybrid loss function that combines supervised and semi-supervised components with confidence-based pseudo-labeling. Experiments show that the best performance is achieved with the Jina-v3 encoder and the Mamba contextual model: mWACC = 62.52 % (Mean Weighted Accuracy Classification) and mMF1 = 61.03 % (Mean Weighted F1-Measure) on CMU-MOSEI (Multimodal Opinion Sentiment and Emotion Intensity) corpus; mACC = 88.80 % (Mean Accuracy) and mCCC = 25.44 % (Mean Concordation Correlation Coefficient) on FIv2. The model demonstrates stable knowledge transfer across tasks and outperforms current state-of-the-art methods. Attention visualization via Grad-CAM confirms the interpretability of predictions. The proposed method enables the development of scalable text-based psychological profiling systems under realistic annotation scarcity. It is applicable in recruitment, adaptive learning platforms, personalized chatbots, and computational psychometrics where simultaneous consideration of emotional states and stable personality traits is essential.

Keywords

emotion recognition, personality trait recognition, multitask learning, cross-domain learning, semi-supervised learning, model attention visualization

Acknowledgements

The "Related work" was prepared within the framework of the budget topic of the St. Petersburg Federal Research Center of the Russian Academy of Sciences (No. FFZF-2025-0003). The development of semi-supervised machine learning components, including a hybrid loss function and a pseudo-labeling mechanism, was supported by the Russian Science Foundation (project No. 24-71-00083). The design of the cross-domain model architecture featuring cross-attention blocks and integration of contextual models was supported by the Russian Science Foundation (project No. 24-71-00112).

For citation: Koryakovskaya D.O., Axyonov A.A., Ryumina E.V., Ryumin D.A. Multi-task human's psychological profile analysis based on text data using semi-supervised learning. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 2, pp. 337–348 (in Russian). doi: 10.17586/2226-1494-2026-26-2-337-348

Введение

Под психологическим портретом понимается целостное представление о личности, объединяющее личностные и реляционные особенности, проявляющиеся через эмоции и социальное взаимодействие [1]. В компьютерных науках для описания психологического портрета человека преимущественно используются данные о текущем эмоциональном состоянии и устойчивых личностных характеристиках, что позволяет существенно повысить эффективность интеллектуальных систем. Такие системы находят применение в задачах персонализации, подбора кадров, мониторинга здоровья и адаптивного обучения [2–5]. Эмоциональные и личностные характеристики взаимосвязаны, поскольку личностные характеристики определяют интенсивность и валентность эмоциональных реакций, а их частотность проявлений со временем формирует устойчивые поведенческие шаблоны [6]. Дополнительные подтверждения этому представлены в современных научных работах, посвященных их совместному анализу [7, 8]. Тем не менее в области машинного обучения методы распознавания эмоциональных и личностных характеристик разрабатываются изолированно [9, 10].

Отметим, что независимо от задачи в методах преимущественно используются трансформерные нейросетевые модели [11, 12] и векторные модели [13, 14] для извлечения признаков. Тогда как для моделирования задачно-специфичных признаков и предсказаний чаще применяются контекстные нейросетевые модели [15, 16]. Хотя разрабатываемые методы имеют пересечение по используемым моделям, они изолированы по причине отсутствия общедоступных корпусов, аннотированных для обеих задач, что затрудняет совместное машинное обучение.

Для решения данной проблемы в настоящей работе предлагается метод полуконтролируемого кросс-доменного машинного обучения на текстовых данных, объединяющий задачи распознавания эмоций и оценивания личностных характеристик без необходимости дополнительной аннотации данных. Метод интегрирует два специализированных корпуса: CMU Multimodal Opinion Sentiment and Emotion Intensity (CMU-MOSEI) [17], аннотированный по эмоциональным категориям, и ChaLearn First Impressions v2 (FIv2), содержащий аннотации личностных характеристик [18]. Метод реализует последовательную схему машинного обучения. В результате на первом этапе формируются однозадачные модели, после чего обучается совместная модель, использующая общее текстовое представление и механизм перекрестного внимания, обеспечивающий обмен признаками между задачами.

Ключевым элементом метода является полуконтролируемое машинное обучение, зарекомендовавшее себя как эффективный способ повышения точности в условиях нехватки обучающих данных [19, 20]. В предлагаемом методе полуконтролируемое машинное обучение компенсирует недостаток аннотированных данных за счет объединения размеченной выборки из одного корпуса с неразмеченной выборкой из другого. Таким образом, предлагаемый метод устраняет недостаток современных решений, связанных с совместным распознаванием эмоций и личностных характеристик по тексту, за счет применения общего текстового энкодера и гибридной функции потерь, объединяющей контролируемую и полуконтролируемую составляющие и обеспечивающей кросс-доменное согласование. Дополнительная аннотация не требуется.

Обзор литературы

В задачах распознавания эмоций по тексту ключевую роль играют контекстно-зависимые трансформерные энкодеры, такие как Bidirectional Encoder Representations from Transformers (BERT) и его производные, включая Robustly optimized BERT approach (RoBERTa). Их дообучение на текстовых корпусах, предназначенных для распознавания эмоций, обеспечивает устойчивый учет контекста высказываний [11, 21]. В качестве простых базовых моделей по-прежнему применяются предобученные векторные представления слов, такие как Global Vectors for Word Representation (GloVe), в сочетании с нейросетевыми моделями на основе Convolutional Neural Network (CNN), Gated Recurrent Units (GRU) и Bidirectional Long Short-Term Memory (BiLSTM) [22–24]. Такие решения эффективны в условиях ограниченных вычислительных ресурсов, однако при наличии достаточной аннотации трансформерные модели демонстрируют более высокую точность [13, 25]. В работе [26] текстовая модальность реализована в виде каскада BERT и Bidirectional GRU (BiGRU). Так, BERT формирует контекстные представления, BiGRU агрегирует их, после чего применяется полносвязный классификатор. Современные большие языковые модели демонстрируют высокий потенциал в задачах распознавания эмоций благодаря способности улавливать сложные контекстные и семантические зависимости [27].

К ключевым ограничениям метода относятся: различия в методологиях аннотаций (одна эмоциональная метка или несколько эмоциональных меток на предложение), что затрудняет сопоставление результатов; снижение точности при переносе нейросетевых моделей на другие корпуса и типы текстов [28, 29]. Частично эти ограничения компенсируются многокорпусными методами машинного обучения, использующими контекстно-независимые механизмы внимания, которые повышают способность нейросетевой модели к обобщению на данных, не входивших в обучающую выборку [30].

Задача оценивания личностных характеристик по тексту обычно рассматривается в рамках модели «Большой пятерки», также известной как OCEAN, и формулируется как регрессионная задача на основе контекстных векторных представлений, извлеченных с помощью нейросетевых моделей BERT или RoBERTa [31, 32]. Такие модели, как правило, обеспечивают более высокую точность по сравнению с классическими признаковыми методами на основе Linguistic Inquiry and Word Count и GloVe [15, 33]. Параллельно развиваются гибридные методы, в которых контекстные векторные представления текста дополняются психолингвистическими и тональными признаками. Например, использование нейросетевой модели на основе графов [34] повышает чувствительность модели к таким дополнительным признакам, а включение аффективной информации при формировании векторных представлений текста дополнительно повышает точность прогнозирования личностных характеристик [35].

К ключевым ограничениям методов относятся: различия в методологиях аннотирования и применяемых опросниках; ограниченное количество крупных согласованно размеченных корпусов по характеристикам OCEAN; снижение качества нейросетевых моделей при переносе на другие языки без этапа дополнительной адаптации [36].

Многозадачное распознавание психологического портрета человека в текстах остается малоизученной областью [37]. В работе [38] используется общий текстовый энкодер с раздельными «головами», предназначенными для распознавания эмоций и оценивания личностных характеристик. Однако машинное обучение на едином корпусе с совместной аннотацией снижает переносимость модели как на внешние корпуса (с иной лингвистико-демографической структурой), так и на смежные задачи. В работе [39] эмоции и личностные характеристики распознаются независимыми нейросе-

тевыми моделями, что исключает совместное использование единого контекстного представления текста и затрудняет многозадачный анализ психологического портрета человека.

Основная трудность состоит в извлечении общих латентных представлений при отсутствии корпуса с совместной аннотацией. Предлагаемый метод направлен на преодоление этого ограничения.

Метод

Функциональная схема предлагаемого метода многозадачного анализа психологического портрета человека на основе текстовых данных с применением полуконтролируемого машинного обучения представлена на рис. 1. Метод формирует прогнозы эмоциональных состояний и личностных характеристик, обрабатывая текстовые данные посредством кросс-доменной нейросетевой модели, обученной на задачно-специфичных корпусах. Полуконтролируемое машинное обучение позволяет использовать данные корпуса совместно, обеспечивая кросс-доменную передачу знаний без дополнительной аннотации корпусов по другим психологическим состояниям.

Компоненты кросс-доменной модели. Кросс-доменная нейросетевая модель включает два компонента: энкодер, формирующий признаковые представления текста; классификатор, выполняющий контекстно-зависимое прогнозирование. Для извлечения лингвистических признаков в настоящей работе выполнено сравнение нескольких энкодеров. Jina-v3 [40], основан на XLM-RoBERTa [41], представляет собой многоязычный трансформер, предварительно обученный на 30 языках с применением контрастного машинного обучения. Он формирует 1024-мерные векторные представления, эффективно отражающие кросс-лингвистические семантические связи благодаря 24-слойному двунаправленному механизму внимания. BAAI General Embedding (BGE-en) [42] представляет собой англо-ориентированную нейросетевую модель, построенную на BERT-подобной архитектуре и прошедшую трехэтапное машинное обучение: предварительное обучение с маскированием языка; контрастное обучение с извлечением сложных негативных примеров; многозадачное дообучение на аннотированных корпусах. BGE-en формирует компактные 384-мерные векторные представления, оптимизированные для задач семантического поиска и анализа текста в рамках на одном языке.

Для моделирования контекстных, эмоциональных и личностных шаблонов в высказываниях, а также многозадачного прогнозирования дополнительно применяются и сравниваются две контекстные модели: трансформер [43] и Mamba [44]. Трансформер основан на механизме внимания, что позволяет эффективно моделировать долгосрочные зависимости, однако сопровождается квадратичной вычислительной сложностью $O(n^2)$ по длине последовательности. Mamba, напротив, основана на рекуррентно-усиленной нейросетевой архитектуре с механизмом селективного обновления состояния, обеспечивая линейную сложность $O(n)$ и потенциально более эффективную обработку длинных последовательностей. Сравнение контекстных нейросетевых моделей позволяет оценить значимость глобальных контекстных связей (улавливаемых вниманием) по сравнению с эффективной последовательной обработкой (характерной для Mamba) в задачах совместного прогнозирования эмоций и личностных характеристик. Поскольку длина высказы-

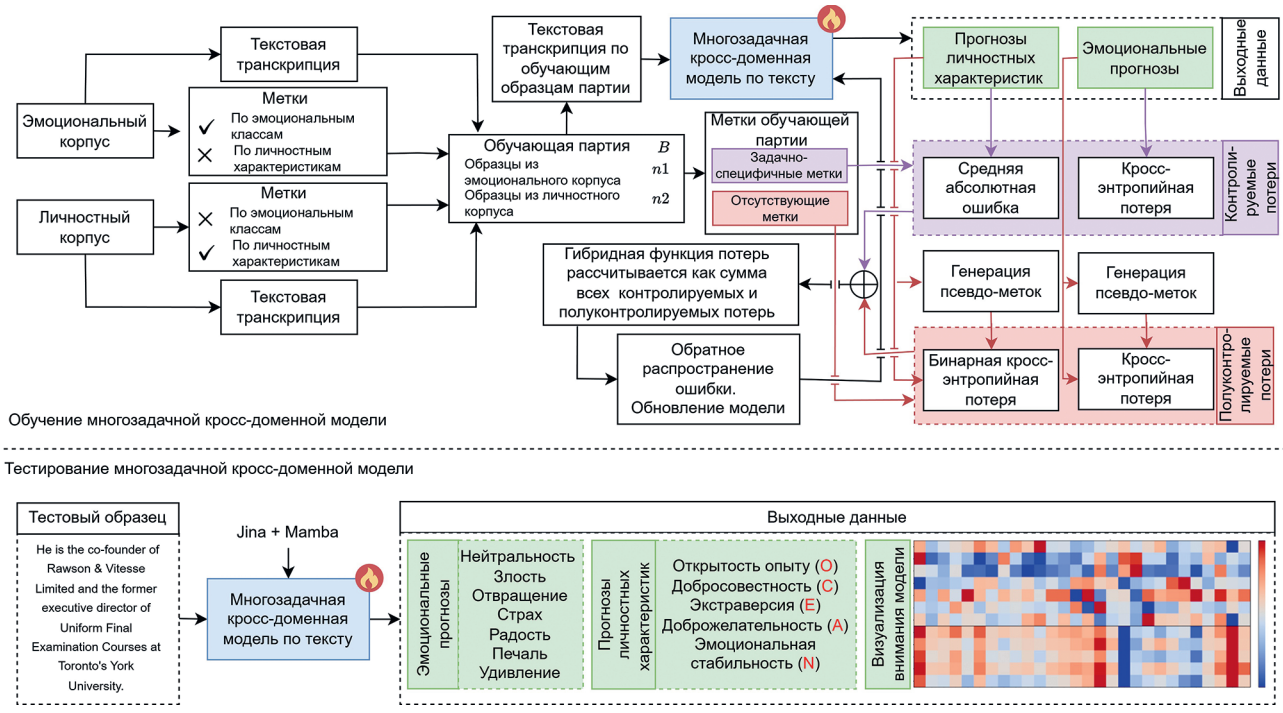

Рис. 1. Функциональная схема предлагаемого метода

Fig. 1. Pipeline of the proposed method

ваний существенно варьируется между примерами текстовых высказываний в используемых корпусах, для унификации данных все векторные представления дополняются нулями до длины самой длинной последовательности в каждой обучающей выборке. Такой метод обеспечивает совместимость последовательностей и корректность машинного обучения нейросетевых моделей, позволяя полноценно учитывать контекст и структуру текста при анализе психологического портрета человека.

Архитектура мультизадачной кросс-доменной модели. Функциональная схема предложенной кросс-доменной модели представлена на рис. 2. На этапе 1 входными данными задачно-специфичной модели являются текстовые транскрипции, проходящие предобработку — токенизацию текста. На этапе 2 следует извлечение признаков с использованием предобученных энкодеров (включая Jina-v3 и BGE-en). Модель обрабатывает последовательности признаков обучающей выборки, сформированной из файлов эмоционального ($X^{EM} \in R^{Bs\times T\times D^E}$) или личностного ($X^{PT} \in R^{Bs\times T\times D^E}$) корпусов, где Bs, T и D^E — размер обучающей выборки, длина последовательности и размерность признаков установленной энкодером E. Входные данные Bs, T, D^E передаются в две отдельные задачно-специфичные нейросетевые модели, предназначенные для контекстного моделирования признаков. Каждая модель предварительно обучается отдельно, что обеспечивает устойчивое формирование признаковых представлений внутри своего домена до начала совместного многозадачного кросс-доменного обучения.

Каждый вход проецируется в общее латентное пространство $\mathbf{H}^{EM}$ и $\mathbf{H}^{PT}$ с размерностями D^{EM} и D^{PT} согласно формулам:

$$\mathbf{H}^{EM} = \text{Dropout}(\text{LayerNorm}(\mathbf{X}^{EM}\mathbf{W}^{EM})), \quad \mathbf{W}^{EM} \in R^{D^E\times D^{EM}}, \tag{1}$$

$$\mathbf{H}^{PT} = \text{Dropout}(\text{LayerNorm}(\mathbf{X}^{PT}\mathbf{W}^{PT})), \quad \mathbf{W}^{PT} \in R^{D^E\times D^{PT}}, \tag{2}$$

где $\mathbf{W}^{EM}$ и $\mathbf{W}^{PT}$ — матрицы весов проекции эмоциональных и личностных признаков; LayerNorm(·) — выполняет нормализацию масштабов признаков; Dropout(·) — осуществляет стохастическое обнуление нейронов в процессе машинного обучения.

Каждая последовательность обрабатывается контекстной моделью (трансформер или Mamba), состоящей из N контекстных слоев $\text{TEL}_n(\cdot)$, предназначенных для захвата контекстных зависимостей:

$$\mathbf{H}^{EM}_{n+1} = \text{TEL}^{EM}_n(\mathbf{H}^{EM}_n),$$
$$\mathbf{H}^{PT}_{n+1} = \text{TEL}^{PT}_n(\mathbf{H}^{PT}_n).$$

Финальные скрытые состояния агрегируются по временному измерению посредством контекстного усреднения:

$$\overline{\mathbf{H}}^{EM} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{H}^{EM}_N)[:, t, :], \tag{3}$$

$$\overline{\mathbf{H}}^{PT} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{H}^{RT}_N)[:, t, :]. \tag{4}$$

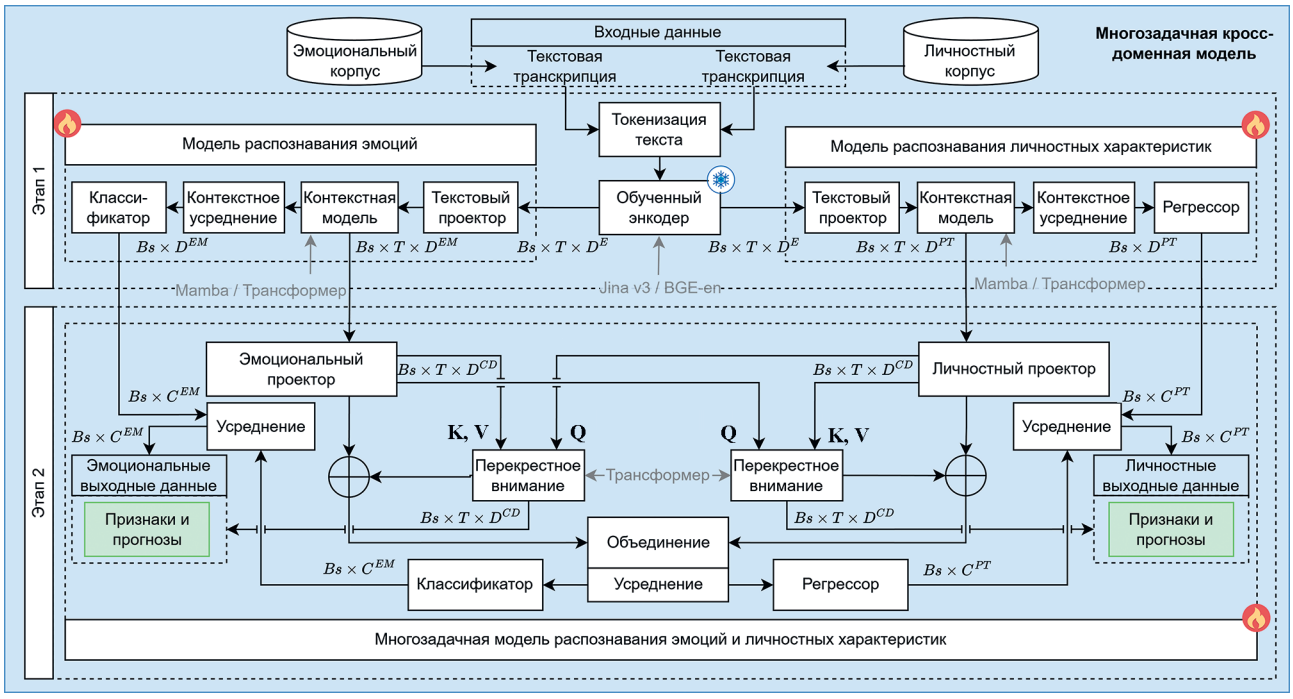

Рис. 2. Функциональная схема предложенной кросс-доменной модели.
DEM — эмоциональные признаки; DPT — признаки личностных характеристик; **Q** — матрица запросов; **K** — матрица ключей; **V** — матрица значений

Fig. 2. Pipeline of the proposed cross-domain model. Notation: DEM denotes emotional features, DPT denotes personality-trait features, **Q** is the query matrix, **K** is the key matrix, and **V** is the value matrix

Далее формируются доменно-специфичные «головы» для задачи классификации и регрессии, состоящие из полносвязных слоев согласно формулам:

$$\mathbf{y}^{\widehat{EM}} = \mathbf{W}_2^{EM}\text{Dropout}\Big(\varphi\Big(\text{LayerNorm}\Big(\mathbf{W}_1^{EM}\overline{\mathbf{H}}^{EM}\Big)\Big)\Big), \quad (5)$$

$$\mathbf{y}^{\widehat{RT}} = \sigma\Big(\mathbf{W}_2^{PT}\,\text{Dropout}\Big(\varphi\Big(\text{LayerNorm}\Big(\mathbf{W}_1^{PT}\,\overline{\mathbf{H}}^{PT}\Big)\Big)\Big)\Big), \quad (6)$$

где $\varphi(\cdot)$ — функция активации GELU; $\sigma(\cdot)$ — сигмоидная функция активации, используемая для нормализации оценок личностных характеристик, ограниченных в [0, 1]. Размеры весов: $\mathbf{W}_1^{PT} \in R^{D^E \times D^{PT}}$, $\mathbf{W}_1^{PT} \in R^{D^E \times D^{PT}}$, $\mathbf{W}_2^{EM} \in R^{D^{EM} \times C^{EM}}$ и $\mathbf{W}_2^{PT} \in R^{D^{PT} \times C^{PT}}$, где C^{EM} и C^{PT} — количество классов эмоций и личностных характеристик.

На этапе 2 (рис. 2) для обеспечения признакового кросс-взаимодействия между задачно-специфичными моделями вводятся два блока перекрестного. Входными данными модели являются неусредненные однодоменные скрытые состояния, полученные из задачно-специфичных моделей. Скрытые состояния проходят через доменно-специфичные проекторы. Каждый проектор работает аналогично формулам (1) и (2), преобразуя признаки различной размерности ($\mathbf{H}_N^{EM} \in R^{Bs \times T \times D^{EM}}$ и $\mathbf{H}_N^{PT} \in R^{Bs \times T \times D^{PT}}$) в общее пространство признаков ($\mathbf{H}^{\widetilde{EM}}, \mathbf{H}^{\widetilde{PT}} \in R^{Bs \times T \times D^{CD}}$, где D^{CD} — размерность скрытого состояния) для их последующей агрегации. Скрытые состояния $\mathbf{H}^{\widetilde{EM}}$ и $\mathbf{H}^{\widetilde{PT}}$ подаются в два блока перекрестного внимания, каждый из которых включает N трансформерных слоев $TL_n(\cdot)$. Работа слоев описывается следующими формулами:

$$\mathbf{H}_{n+1}^{\widetilde{EM}} = \text{TL}_n^{EM \to PT}(\mathbf{Q} = \mathbf{H}_n^{\widetilde{EM}}, \mathbf{K} = \mathbf{H}_n^{\widetilde{PT}}, \mathbf{V} = \mathbf{H}_n^{\widetilde{PT}}),$$

$$\mathbf{H}_{n+1}^{\widetilde{PT}} = \text{TL}_n^{PT \to EM}(\mathbf{Q} = \mathbf{H}_n^{\widetilde{PT}}, \mathbf{K} = \mathbf{H}_n^{\widetilde{EM}}, \mathbf{V} = \mathbf{H}_n^{\widetilde{EM}}),$$

где три тензора (матрицы) трансформерного слоя представляют собой запрос (**Q**), ключ (**K**) и значение (**V**) соответственно (рис. 2). Такой механизм позволяет каждой модальности выборочно фокусировать внимание на признаках другого домена.

После рассмотрения N слоев эмоциональные и личностные признаки объединяются в виде:

$$\mathbf{H}_{fused} = \text{Concat}\Big(\mathbf{H}_N^{\widetilde{EM}}, \mathbf{H}_N^{\widetilde{PT}}\Big).$$

Затем признаки агрегируются с использованием контекстного усреднения — формулы (3) и (4). Объединенные признаки $\mathbf{H}_{fused}$ подаются в задачно-специфичные «головы», которые работают аналогично формулам (5) и (6). Для повышения устойчивости и сохранения одномодальной производительности усредняются предсказания задачно-специфичных и многозадачной кросс-доменной моделей:

$$\mathbf{y}^{EM} = \frac{1}{2}\Big(\mathbf{y}^{\widehat{EM}} + \mathbf{y}^{\widetilde{EM}}\Big),$$

$$\mathbf{y}^{PT} = \frac{1}{2}\Big(\mathbf{y}^{\widehat{PT}} + \mathbf{y}^{\widetilde{PT}}\Big),$$

где $\mathbf{y}^{\widehat{EM}}$ и $\mathbf{y}^{\widetilde{EM}}$ — предсказания эмоций из задачно-специфичных и многозадачной моделей; $\mathbf{y}^{\widehat{PT}}$ и $\mathbf{y}^{\widetilde{PT}}$ — предсказания личностных характеристик тех же моделей.

Полуконтролируемое обучение. Для реализации разработанного метода обучение нейросетевых моделей проводится на двух задачно-специфичных корпусах, каждый из которых посвящен единственной задаче — распознаванию эмоций или оцениванию личностных характеристик. Метки для несоответствующей задачи в обоих случаях заполняются значением *NaN*. На каждом шаге обучающая выборка формируется путем случайного выбора примеров текстовых высказываний из обоих корпусов; тогда n_1 и n_2 обозначают количество выбранных примеров, а размер обучающей выборки $B = n_1 + n_2$.

Для обучения нейросетевых моделей применяется гибридная функция потерь, объединяющая контролируемую и полуконтролируемую составляющие. Контролируемая потеря L_s вычисляется только на аннотированных данных:

$$L_s = W_s^{EM} L_s^{EM} + W_s^{PT} L_s^{PT},$$

где L_s^{EM} — кросс-энтропийная потеря для распознавания эмоций; L_s^{PT} — средняя абсолютная ошибка для оценивания личностных характеристик. При этом предсказания, полученные на неаннотированных данных, исключаются из расчета функции потерь. Динамически адаптируемые веса потерь $\{W_s^t\}_{t \in T}$ устанавливаются равными 1 для задачи распознавания эмоций и 0,2 — для задачи оценивания личностных характеристик.

Неаннотированные примеры текстовых высказываний помечаются с использованием псевдомаркировки и задачно-зависимых порогов уверенности. Для распознавания эмоций псевдометки принимаются, когда максимальная вероятность предсказания превышает порог τ^{PT}. Для распознавания личностных характеристик выходы преобразуются в бинарные метки с использованием порога полярности, равного 0,5 только в том случае, если они выходят за пределы зоны неопределенности — т. е. выше τ^{PT} или ниже 1 — τ^{PT}. Полуконтролируемая функция потерь L_{ss} вычисляется по формуле:

$$L_{ss} = W_{ss}^{EM} L_{ss}^{EM} + W_{ss}^{PT} L_{ss}^{PT},$$

где L_{ss}^{EM} — кросс-энтропийная потеря для распознавания эмоций; L_{ss}^{PT} — бинарная кросс-энтропийная потеря для оценивания личностных характеристик. Гибридная функция потерь вычисляется:

$$L = L_s + L_{ss}.$$

Полуконтролируемые веса потерь W_s^t и W_{ss}^t устанавливаются равными 1.

Экспериментальные исследования метода

Экспериментальная установка. Процесс обучения нейросетевых моделей включает многоэтапную настройку. Сначала обучаются задачно-специфичные, а затем — кросс-доменные модели. В ходе обучения посредством поиска по сетке подбираются оптимальные значения для следующих параметров модели: тип контекстного слоя (трансформер или Mamba), размер

скрытых состояний (128, 256, 512 или 1024), размерность признаков (128, 256, 512 или 1024), число слоев (1, 2, 3, 4) и число «голов» внимания (2, 4, 8, 16). Аналогичным образом подбираются параметры для обеспечения полуконтролируемого машинного обучения: вклад контролируемых и полуконтролируемых функций потерь (0, 1 с шагом 0,1) и пороги полярности для генерации псевдометок (0,5, 0,6, 0,7). Параметры машинного обучения, такие как размер обучающей выборки (32), скорость обучения 10^{-4}, оптимизатор Adam и вероятность отключения нейронов (0,15) — фиксированы для всех нейросетевых моделей.

Показатели эффективности. Для оценки производительности разрабатываемых нейросетевых моделей используются различные показатели, специфичные для каждой задачи. Для распознавания эмоций применяются показатели, такие как средняя взвешенная F1-мера (mMF1) и средняя невзвешенная точность mWACC, предложенные в работе [17]. Для оценивания личностных характеристик используются коэффициент корреляции конкордации Лина (mean Concordance Correlation Coefficient, mCCC) и точность, определяемая как 1 — средняя абсолютная ошибка [18]. Оба показателя вычисляются для каждой характеристики, после чего усредняются.

Исследовательские корпуса. Поскольку на данный момент не существует корпусов, аннотированных для обеих задач, в настоящей работе использованы два специализированных корпуса, а именно, CMU-MOSEI [17] для многомодального распознавания эмоций и FIv2 [18] для оценивания личностных характеристик.

CMU-MOSEI — крупнейший многомодальный корпус для анализа сентимента и распознавания эмоций. Он содержит более 23 500 видеовыражений, соответствующих отдельным предложениям, от более чем 1000 демонстраторов, размещенных на популярном видеохостинге. Каждое видео размечено по 6 базовым эмоциям: злость (Anger, An), отвращение (Disgust, Di), страх (Fear, Fe), радость (Happiness, Ha), печаль (Sadness, Sa), удивление (Surprise, Su) и нейтральное состояние (Neutral, Ne). Длина видеозаписей варьируется, средняя продолжительность составляет 4,5 с.

Корпус FIv2 состоит из 10 000 коротких видеоклипов длительностью по 15 с, снятых более чем на 3000 каналах видеоблогов, размещенных на популярном видеохостинге, аннотированных по модели «Большой пятерки» личностных характеристик: открытость опыту (Openness, O), добросовестность (Conscientiousness, C), доброжелательность (Agreeableness, A), экстраверсия (Extraversion, E) и эмоциональная стабильность (Non-Neuroticism, N).

Корпусы содержат спонтанную речь, что обеспечивает применимость разрабатываемого метода в реальных условиях. Оба корпуса имеют фиксированные подмножества для обучения, валидации и тестирования. Распределение эмоциональных классов и оценок личностных характеристик представлено на рис. 3. Распределения по обоим задачам являются несбалансированными, что может негативно сказаться на эффективности модели. Текстовые транскрипции для корпусов извлекаются нейросетевой моделью Whisper Turbo [45], а количество слов в высказываниях варьируется. Для корпуса CMU-MOSEI среднее количество слов составило 20, диапазон значений равен [2; 289], для FIv2 среднее количество слов — 42 при диапазоне [2; 82]. Такая вариативность может улучшать обобщающую способность моделей на новых данных.

Экспериментальные результаты. Результаты анализа моделей представлены в табл. 1. Рейтинг эффективности моделей рассчитан с использованием теста Фридмана [46]. Результаты показали сравнительную эффективность различных экстракторов признаков

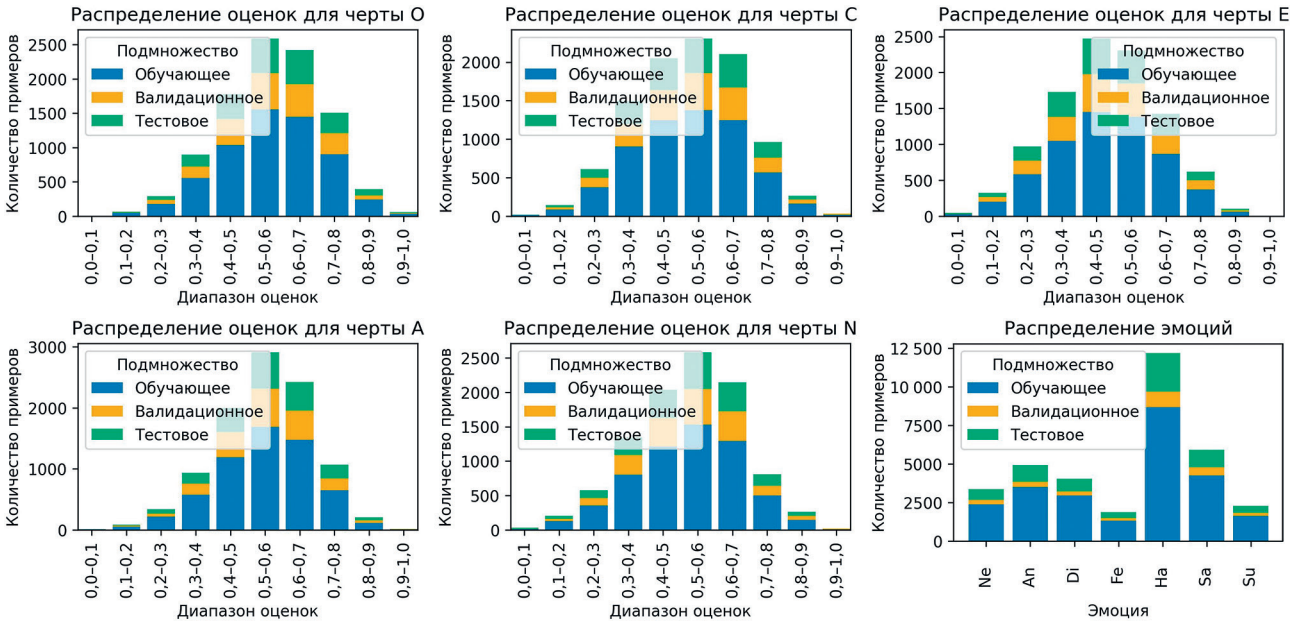

Рис. 3. Диаграммы распределения эмоциональных классов и оценок характеристик личности

Fig. 3. Distribution diagrams of emotional classes and personality trait scores

Таблица 1. Сравнение эффективности архитектур нейросетевых моделей
Table 1. Comparison of neural network architectures performance

Энкодер	Контекстная модель	Тип обучения	Домен обучения	CMU-MOSEI		FIv2		Рейтинг
				mWACC, %	mMF1, %	mACC, %	mCCC, %	
Jina-v3	Mamba	КО	ОДО	62,20	60,53	**88,82**	27,72	5,00
	Mamba	ПКО	КДО	**62,52**	**61,03**	88,80	25,44	**2,00**
	Трансформер	КО	ОДО	62,20	60,88	88,75	**29,93**	5,33
	Трансформер	ПКО	КДО	62,16	60,40	88,75	27,87	6,67
BGE-en	Mamba	КО	ОДО	62,40	60,69	88,80	27,09	4,67
	Mamba	ПКО	КДО	62,43	60,98	88,75	27,02	3,67
	Трансформер	КО	ОДО	62,48	60,72	**88,82**	29,16	2,67
	Трансформер	ПКО	КДО	62,31	60,79	**88,82**	28,71	3,33
Glove + CNNs [33]		КО	ОДО	—	—	88,33	—	9,00
BERT + BiGRU [26]		КО	ОДО	62,00	57,13	—	—	9,00

Примечание: КО — контролируемое обучение; ПКО — полуконтролируемое обучение; ОДО — однодоменное обучение; КДО — кросс-доменное обучение; mWACC — средняя взвешенная точность классификации; mMF1 — средняя взвешенная F1-мера; mACC — средняя точность; mCCC — средний коэффициент корреляции конкордации Лина. Наилучшие значения метрик выделены полужирным шрифтом.

(Jina-v3 и BGE-en) в сочетании с контекстными моделями (Mamba и трансформер) в условиях контролируемого однодоменного и полуконтролируемого кросс-доменного машинного обучения. Полученные данные демонстрируют, что энкодер Jina-v3 обеспечивает более высокие результаты по сравнению с BGE-en. Это свидетельствует о способности энкодера Jina-v3 обобщать знания на другие языки.

При сравнении контекстных моделей Mamba демонстрирует значительно более высокую производительность по сравнению с моделями типа трансформер в задаче распознавания эмоций. В то же время модели на архитектуре трансформер демонстрируют более высокую эффективность при оценивании личностных характеристик. Различие в производительности объясняется структурными особенностями корпусов. FIv2 содержит высказывания фиксированной длительности (15 с), тогда как CMU-MOSEI включает высказывания переменной длины. В табл. 1 также приведено сравнение наиболее производительной модели с конфигурацией Jina-v3 + трансформер (рейтинг равен 2,00) с современными методами. Дополнительно для оценки статистической значимости достигнутых улучшений рассчитаны доверительные интервалы модели используя метод бутстрап-перевыборки [47]. Показатель эффективности mWACC = 62,00, полученный методом [26], лежит в доверительном интервале представленного метода ([61,70; 63,24]), но значительно ближе к его нижней границе (разница 0,30 против 1,24 до верхней), что свидетельствует о меньшей эффективности метода [26]. При этом значение mACC = 88,33, полученное методом [33], не превышает нижней границы доверительного интервала предложенного метода ([88,53; 89,08]). Таким образом, результаты работы метода демонстрируют статистически значимые улучшения по сравнению с современными решениями [26, 33].

На рис. 4 представлена тепловая карта внимания модели к словам (токенам), наиболее значимым для предсказания эмоций и личностных характеристик. Визуализация выполнена с использованием метода Gradient-weighted Class Activation Mapping [48] для предложения на русском языке, сгенерированного крупной языковой моделью Qwen3-4B-Instruct [49] в ответ на запрос: «Сгенерируй предложения на русском языке, отражающие эмоцию печали и раскрывающее личностные характеристики человека, не превышающее 30 слов». Выбор русскоязычного примера обусловлен необходимостью продемонстрировать обобщающую способность предложенного подхода на языках, отличных от английского.

Несмотря на то, что нейросетевая модель обучалась только на английском языке, использование энкодера, предварительно обученного на 30 языках (включая русский), обеспечивает ее применимость к другим языкам. Модель корректно предсказывает отрицательную валентность эмоции, присваивая наибольшие вероятности таким классам, как отвращение (Di — 0,49), злость (An — 0,27), печаль (Sa — 0,13) и удивление (Su — 0,08). Наиболее значимыми словами для первых двух эмоций являются «мои чувства». Подобные негативные высказывания приводят к низким предсказанным значениям по всем личностным характеристикам, кроме открытости опыту (O — 0,59). При оценивании личностных характеристик наибольшую значимость имеют слова «просто бросил меня».

В целом результаты показывают, что нейросетевая модель выделяет ключевые семантические маркеры, соответствующие содержанию текста, и способна одновременно распознавать эмоции и оценивать личностные характеристики, демонстрируя применимость не только к англоязычным, но и к многоязычным данным.

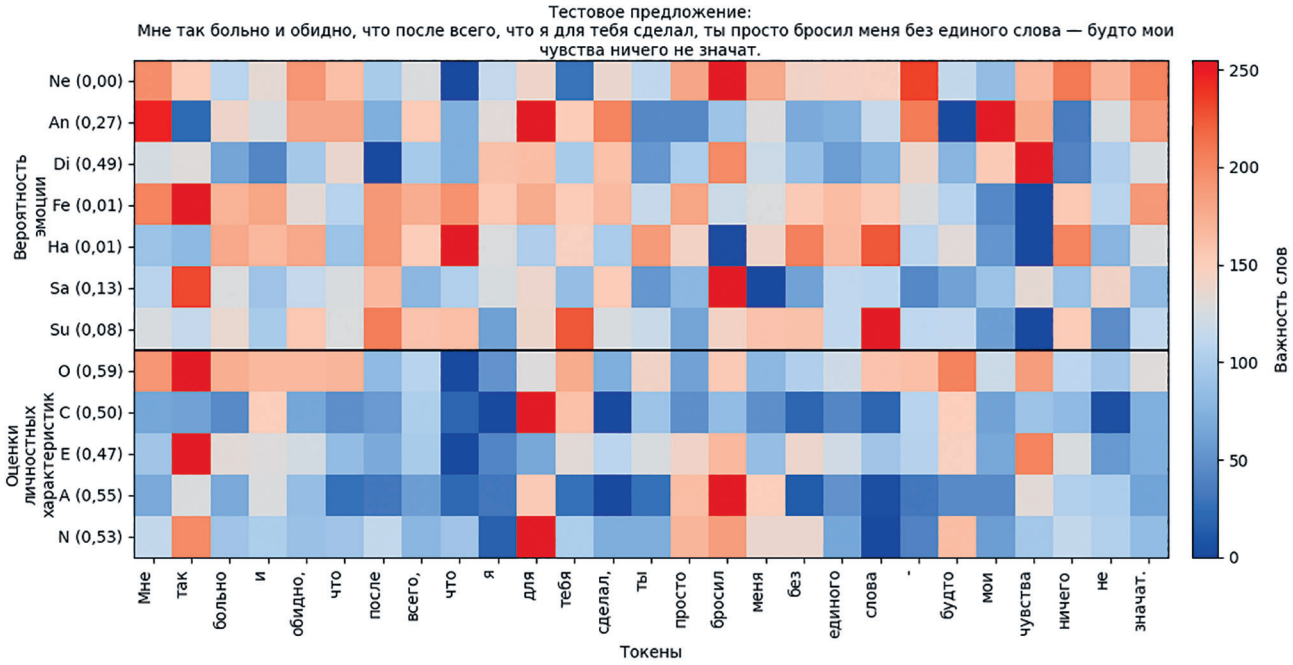

Рис. 4. Визуализация тепловой карты внимания модели к токенам.

Цветовая шкала отражает значимость каждого слова (токена) для соответствующего прогноза: синий — низкая значимость, красный — высокая

Fig. 4. Heatmap visualization of the model attention to tokens

Заключение

Предложен новый метод многозадачного анализа психологического портрета человека по тексту на основе полуконтролируемого кросс-доменного обучения, который позволяет одновременно распознавать эмоции и оценивать личностные характеристики без необходимости совместной разметки данных. Метод реализован в виде двухэтапной архитектуры: на первом этапе обучаются независимые однозадачные модели, на втором — совместная модель с перекрестным вниманием. Финальное предсказание формируется путем усреднения их выходов.

Выполненные эксперименты на корпусах CMU-MOSEI и FIv2 показали, что лучшие результаты достигаются при использовании энкодера Jina-v3 и контекстной модели Mamba: средняя взвешенная точность классификации (mWACC) составила 62,52 %, а средняя взвешенная F1-мера (mMF1) — 61,03 % на корпусе CMU-MOSEI; средняя точность (mACC) составила 88,80 %, а средний коэффициент корреляции конкордации Лина (mCCC) — 25,44 % на FIv2. Предложенная нейросетевая модель демонстрирует устойчивую передачу знаний между задачами и превосходит современные аналоги. Визуализация внимания методом Gradient-weighted Class Activation Mapping подтверждает интерпретируемость ее прогнозов.

В дальнейшем планируется исследование многомодальных нейросетевых архитектур и расширение обучающих данных за счет включения неанглоязычных корпусов. Предложенный метод предполагается интегрировать в многомодальные системы психологического профилирования для повышения точности и надежности распознавания эмоций и оценивания личностных характеристик в реальных сценариях использования, включая анализ пользовательского контента в цифровых средах, поддержку принятия решений в системах при подборе персонала и психотерапевтической практике.

Литература

1. Karanatsiou D., Sermpezis P., Gruda D., Kafetsios K., Dimitriadis I., Vakali A. My tweets bring all the traits to the yard: Predicting personality and relational traits in Online Social Networks // ACM Transactions on the Web (TWEB). 2022. V. 16. N 2. P. 1–26. https://doi.org/10.1145/3523749
2. Двойникова А.А, Маркитантов М.В., Рюмина Е.В., Уздяев М.Ю., Величко А.Н., Рюмин Д.А., Лякso Е.Е., Карпов А.А. Анализ информационного и математического обеспечения для распознавания аффективных состояний человека // Информатика и автоматизация. 2022. Т. 21. № 6. С. 1097–1144. https://doi.org/10.15622/ia.21.6.2

References

1. Karanatsiou D., Sermpezis P., Gruda D., Kafetsios K., Dimitriadis I., Vakali A. My tweets bring all the traits to the yard: Predicting personality and relational traits in Online Social Networks. *ACM Transactions on the Web (TWEB)*, 2022, vol. 16, no. 2, pp. 1–26. https://doi.org/10.1145/3523749
2. Dvoynikova A.A., Markitantov M.V., Ryumina E.V., Uzdiaev M.Y., Velichko A.N., Ryumin D.A., et al. Analysis of infoware and software for human affective states recognition. *Informatics and Automation*, 2022, vol. 21, no. 6. pp. 1097–1144. (in Russian). https://doi.org/10.15622/ia.21.6.2

3. Sorin V., Brin D., Barash Y., Konen E., Charney A., Nadkarni G., Klang E. Large language models and empathy: systematic review // Journal of Medical Internet Research. 2024. V. 26. P. e52597. https://doi.org/10.2196/52597
4. Rajesh S.G., Madangarli S.V., Pisharady G.S., Subrahmanyam R. Enhancement of Virtual Assistants through MultiModal AI for Emotion Recognition // IEEE Access. 2025. V. 13. P. 102159–102179. https://doi.org/10.1109/ACCESS.2025.3577664
5. Kovacevic N., Holz C., Gross M., Wampfler R. On multimodal emotion recognition for human-chatbot interaction in the wild // Proc. of the International Conference on Multimodal Interaction. 2024. P. 12–21. https://doi.org/10.1145/3678957.3685759
6. Bao Y., Wang Y., Qi Y., Yang Q., Liu R., Feng L. Emotion-Assisted multi-modal Personality Recognition using adversarial Contrastive learning // Knowledge-Based Systems. 2025. V. 317. P. 113504. https://doi.org/10.1016/j.knosys.2025.113504
7. Mohammadi G., Vuilleumier P. A multi-componential approach to emotion recognition and the effect of personality // IEEE Transactions on Affective Computing. 2020. V. 13. N 3. P. 1127–1139. https://doi.org/10.1109/TAFFC.2020.3028109
8. Li Y., Bell P., Lai C. Transfer Learning for Personality Perception via Speech Emotion Recognition // Proc. of the Annual Conference of the International Speech Communication Association Interspeech. 2023. P. 5197–5201. https://doi.org/10.21437/Interspeech.2023-2061
9. Chandraumakantham O., Gowtham N., Zakariah M., Almazyad A. Multimodal emotion recognition using feature fusion: an LLM-based approach // IEEE Access. 2024. V. 12. P. 108052–108071. https://doi.org/10.1109/ACCESS.2024.3425953
10. Gan P., Sowmya A., Mohammadi G. CLIP-based model for effective and explainable apparent personality perception // Proc. of the 1st International Workshop on Multimodal and Responsible Affective Computing. 2023. P. 29–37. https://doi.org/10.1145/3607865.3613178
11. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // Proc. of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019. V. 1. P. 4171–4186. https://doi.org/10.18653/v1/N19-1423
12. Boitel E., Mohasseb A., Haig E. MIST: Multimodal emotion recognition using DeBERTa for text, Semi-CNN for speech, ResNet-50 for facial, and 3D-CNN for motion analysis // Expert Systems with Applications. 2025. V. 270. P. 126236. https://doi.org/10.1016/j.eswa.2024.126236
13. Li Y., Wang Y., Cui Z. Decoupled multimodal distilling for emotion recognition // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023. P. 6631–6640. https://doi.org/10.1109/CVPR52729.2023.00641
14. Agrawal T., Balazia M., Müller P., Brémond F. Multimodal vision transformers with forced attention for behavior analysis // Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2023. P. 3381–3391. https://doi.org/10.1109/WACV56688.2023.00339
15. Ryumina E., Markitantov M., Ryumin D., Karpov A. Gated Siamese Fusion Network based on multimodal deep and hand-crafted features for personality traits assessment // Pattern Recognition Letters. 2024. V. 185. P. 45–51. https://doi.org/10.1016/j.patrec.2024.07.004
16. Peng C., Chen K., Shou L., Chen G. CARAT: Contrastive feature reconstruction and aggregation for multi-modal multi-label emotion recognition // Proceedings of the AAAI Conference on Artificial Intelligence. 2024. V. 38. N 13. P. 14581–14589. https://doi.org/10.1609/aaai.v38i13.29374
17. Bagher Zadeh A., Liang P.P., Poria S., Cambria E., Morency L.P. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph // Proc. of the 56th Annual Meeting of the Association for Computational Linguistics. 2018. V. 1. P. 2236–2246. https://doi.org/10.18653/v1/P18-1208
18. Escalante H.J., Kaya H., Salah A., Escalera S., Gucluturk Y., Guclu U., et al. Modeling, Recognizing, and Explaining Apparent Personality from Videos // IEEE Transactions on Affective Computing. 2020. V. 13. N 2. P. 894–911. https://doi.org/10.1109/TAFFC.2020.2973984
19. Ouali Y., Hudelot C., Tami M. Semi-supervised semantic segmentation with cross-consistency training // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. P. 12671–12681. https://doi.org/10.1109/CVPR42600.2020.01269

3. Sorin V., Brin D., Barash Y., Konen E., Charney A., Nadkarni G., Klang E. Large language models and empathy: systematic review. *Journal of Medical Internet Research*, 2024, vol. 26, pp. e52597. https://doi.org/10.2196/52597
4. Rajesh S.G., Madangarli S.V., Pisharady G.S., Subrahmanyam R. Enhancement of Virtual Assistants through MultiModal AI for Emotion Recognition. *IEEE Access*, 2025, vol. 13, pp. 102159–102179. https://doi.org/10.1109/ACCESS.2025.3577664
5. Kovacevic N., Holz C., Gross M., Wampfler R. On multimodal emotion recognition for human-chatbot interaction in the wild. *Proc. of the International Conference on Multimodal Interaction*, 2024, pp. 12–21. https://doi.org/10.1145/3678957.3685759
6. Bao Y., Wang Y., Qi Y., Yang Q., Liu R., Feng L. Emotion-Assisted multi-modal Personality Recognition using adversarial Contrastive learning. *Knowledge-Based Systems*, 2025, vol. 317, pp. 113504. https://doi.org/10.1016/j.knosys.2025.11350,
7. Mohammadi G., Vuilleumier P. A multi-componential approach to emotion recognition and the effect of personality. *IEEE Transactions on Affective Computing*, 2020, vol. 13, no. 3, pp. 1127–1139. https://doi.org/10.1109/TAFFC.2020.3028109
8. Li Y., Bell P., Lai C. Transfer Learning for Personality Perception via Speech Emotion Recognition. *Proc. of the Annual Conference of the International Speech Communication Association Interspeech*, 2023, pp. 5197–5201. https://doi.org/10.21437/Interspeech.2023-2061
9. Chandraumakantham O., Gowtham N., Zakariah M., Almazyad A. Multimodal emotion recognition using feature fusion: an LLM-based approach. *IEEE Access*, 2024, vol. 12, pp. 108052–108071. https://doi.org/10.1109/ACCESS.2024.3425953
10. Gan P., Sowmya A., Mohammadi G. CLIP-based model for effective and explainable apparent personality perception. *Proc. of the 1st International Workshop on Multimodal and Responsible Affective Computing*, 2023, pp. 29–37. https://doi.org/10.1145/3607865.3613178
11. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, vol. 1, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423
12. Boitel E., Mohasseb A., Haig E. MIST: Multimodal emotion recognition using DeBERTa for text, Semi-CNN for speech, ResNet-50 for facial, and 3D-CNN for motion analysis. *Expert Systems with Applications*, 2025, vol. 270, pp. 126236. https://doi.org/10.1016/j.eswa.2024.126236
13. Li Y., Wang Y., Cui Z. Decoupled multimodal distilling for emotion recognition. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 6631–6640. https://doi.org/10.1109/CVPR52729.2023.00641
14. Agrawal T., Balazia M., Müller P., Brémond F. Multimodal vision transformers with forced attention for behavior analysis. *Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023, pp. 3381–3391. https://doi.org/10.1109/WACV56688.2023.00339
15. Ryumina E., Markitantov M., Ryumin D., Karpov A. Gated Siamese Fusion Network based on multimodal deep and hand-crafted features for personality traits assessment. *Pattern Recognition Letters*, 2024, vol. 185, pp. 45–51. https://doi.org/10.1016/j.patrec.2024.07.004
16. Peng C., Chen K., Shou L., Chen G. CARAT: Contrastive feature reconstruction and aggregation for multi-modal multi-label emotion recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, no. 13, pp. 14581–14589. https://doi.org/10.1609/aaai.v38i13.29374
17. Bagher Zadeh A., Liang P.P., Poria S., Cambria E., Morency L.P. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018, vol. 1, pp. 2236–2246. https://doi.org/10.18653/v1/P18-1208
18. Escalante H.J., Kaya H., Salah A., Escalera S., Gucluturk Y., Guclu U., et al. Modeling, Recognizing, and Explaining Apparent Personality from Videos. *IEEE Transactions on Affective Computing*, 2020, vol. 13, no. 2, pp. 894–911. https://doi.org/10.1109/TAFFC.2020.2973984
19. Ouali Y., Hudelot C., Tami M. Semi-supervised semantic segmentation with cross-consistency training. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12671–12681. https://doi.org/10.1109/CVPR42600.2020.01269

20. Lian Z., Liu B., Tao J. SMIN: Semi-supervised multi-modal interaction network for conversational emotion recognition // IEEE Transactions on Affective Computing. 2023. V. 14. N 3. P. 2415–2429. https://doi.org/10.1109/TAFFC.2022.3141237
21. Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., et al. Unsupervised cross-lingual representation learning at scale // Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747
22. Hosseini S.S., Yamaghani M.R., Arabani S.P. Multimodal modelling of human emotion using sound, image and text fusion // Signal, Image and Video Processing. 2023. V. 18. N 1. P. 71–79. https://doi.org/10.1007/s11760-023-02707-8
23. Deng L., Liu B., Li Z. Multimodal sentiment analysis based on a cross-modal multihead attention mechanism // Computers, Materials and Continua. 2024. V. 78. N 1. P. 1157–1170. https://doi.org/10.32604/cmc.2023.042150
24. Goncalves L., Leem S.-G., Lin W.-C., Sisman B., Busso C. Versatile audio-visual learning for emotion recognition // IEEE Transactions on Affective Computing. 2023. V. 16. N 1. P. 306–318. https://doi.org/10.1109/TAFFC.2024.3433386
25. Cui Z., Li Y., Wang Y. Incomplete multimodality-diffused emotion recognition // Proc. of the Advances in Neural Information Processing System. 2023. P. 17117–17128. https://doi.org/10.52202/075280-0748
26. Arumugam L., Arumugam S., Chidambaram P., Govindasamy K. A multi-modal deep learning approach for human emotion recognition // Cognitive Neurodynamics. 2025. V. 19. N 1. P. 123. https://doi.org/10.1007/s11571-025-10304-3
27. Li D., Xing B., Liu X., Xia B., Wen B., Kälviäinen H. DEEMO: De-identity multimodal emotion recognition and reasoning // arXiv. 2025. arXiv:2504.19549. https://doi.org/10.48550/arXiv.2504.19549
28. Zhang D., Ju X., Li J., Li S., Zhu Q., Zhou G. Multi-modal multi-label emotion detection with modality and label dependence // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020. P. 3584–3593. https://doi.org/10.18653/v1/2020.emnlp-main.291
29. Zhang Y., Chen M., Shen J., Wang C. TAILOR versatile multi-modal learning for multi-label emotion recognition // Proceedings of the AAAI Conference on Artificial Intelligence. 2022. V. 36. N 8. P. 9100–9108. https://doi.org/10.1609/aaai.v36i8.20895
30. Ryumina E., Ryumin D., Axyonov A., Ivanko D., Karpov A. Multi-corpus emotion recognition method based on cross-modal gated attention fusion // Pattern Recognition Letters. 2025. V. 190. P. 192–200. https://doi.org/10.1016/j.patrec.2025.02.024
31. Naz A., Khan H.U., Bukhari A., Alshemaimri B., Daud A., Ramzan M. Machine and deep learning for personality traits detection: a comprehensive survey and open research challenges // Artificial Intelligence Review. 2025. V. 58. N 8. P. 239. https://doi.org/10.1007/s10462-025-11245-3
32. Soto C.J., Jackson J.J. Five-factor model of personality // Journal of Research in Personality. 2013. V. 42. P. 1285–1302.
33. Ouarka A., Baha T.A., Es-Saady Y., El Hajji M. A deep multimodal fusion method for personality traits prediction // Multimedia Tools and Applications. 2024. V. 84. N 25. P. 29665–29687. https://doi.org/10.1007/s11042-024-20356-y
34. Liu W., Sun Z., Wei S., Zhang S., Zhu G., Chen L. PS-GCN: psycholinguistic graph and sentiment semantic fused graph convolutional networks for personality detection // Connection Science. 2024. V. 36. N 1. P. 2295820. https://doi.org/10.1080/09540091.2023.2295820
35. Akber M.A., Ferdousi T., Ahmed R., Asfara R., Rab R., Zakia U. Personality and emotion — a comprehensive analysis using contextual text embeddings // Natural Language Processing Journal. 2024. V. 9. P. 100105. https://doi.org/10.1016/j.nlp.2024.100105
36. Motlagh S.M.H., Rezvani M.H., Khounsiavash M. AI methods for personality traits recognition: a systematic review // Neurocomputing. 2025. V. 640. P. 130301. https://doi.org/10.1016/j.neucom.2025.130301
37. Zhang Y., Yang Q. A Survey on multi-task learning // IEEE Transactions on Knowledge and Data Engineering. 2021. V. 34. N 12. P. 5586–5609. https://doi.org/10.1109/TKDE.2021.3070203
38. Li Y., Kazemeini A., Mehta Y., Cambria E. Multitask learning for emotion and personality traits detection // Neurocomputing. 2022. V. 493. P. 340–350. https://doi.org/10.1016/j.neucom.2022.04.049
39. Talaat F.M., El-Gendy E.M., Saafan M.M., Gamel S.A. Utilizing social media and machine learning for personality and emotion

20. Lian Z., Liu B., Tao J. SMIN: Semi-supervised multi-modal interaction network for conversational emotion recognition. *IEEE Transactions on Affective Computing*, 2023, vol. 14, no. 3, pp. 2415–2429. https://doi.org/10.1109/TAFFC.2022.3141237
21. Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., et al. Unsupervised cross-lingual representation learning at scale. *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747
22. Hosseini S.S., Yamaghani M.R., Arabani S.P. Multimodal modelling of human emotion using sound, image and text fusion. *Signal, Image and Video Processing*, 2023, vol. 18, no. 1, pp. 71–79. https://doi.org/10.1007/s11760-023-02707-8
23. Deng L., Liu B., Li Z. Multimodal sentiment analysis based on a cross-modal multihead attention mechanism. *Computers, Materials and Continua*, 2024, vol. 78, no. 1, pp. 1157–1170. https://doi.org/10.32604/cmc.2023.042150
24. Goncalves L., Leem S.-G., Lin W.-C., Sisman B., Busso C. Versatile audio-visual learning for emotion recognition. *IEEE Transactions on Affective Computing*, 2023, vol. 16, no. 1, pp. 306–318. https://doi.org/10.1109/TAFFC.2024.3433386
25. Cui Z., Li Y., Wang Y. Incomplete multimodality-diffused emotion recognition. *Proc. of the Advances in Neural Information Processing System*, 2023, pp. 17117–17128. https://doi.org/10.52202/075280-0748
26. Arumugam L., Arumugam S., Chidambaram P., Govindasamy K. A multi-modal deep learning approach for human emotion recognition. *Cognitive Neurodynamics*, 2025, vol. 19, no. 1, pp. 123. https://doi.org/10.1007/s11571-025-10304-3
27. Li D., Xing B., Liu X., Xia B., Wen B., Kälviäinen H. DEEMO: De-identity multimodal emotion recognition and reasoning. *arXiv*, 2025. arXiv:2504.19549. https://doi.org/10.48550/arXiv.2504.19549
28. Zhang D., Ju X., Li J., Li S., Zhu Q., Zhou G. Multi-modal multi-label emotion detection with modality and label dependence. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 3584–3593. https://doi.org/10.18653/v1/2020.emnlp-main.291
29. Zhang Y., Chen M., Shen J., Wang C. TAILOR versatile multi-modal learning for multi-label emotion recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, no. 8, pp. 9100–9108. https://doi.org/10.1609/aaai.v36i8.20895
30. Ryumina E., Ryumin D., Axyonov A., Ivanko D., Karpov A. Multi-corpus emotion recognition method based on cross-modal gated attention fusion. *Pattern Recognition Letters*, 2025, vol. 190, pp. 192–200. https://doi.org/10.1016/j.patrec.2025.02.024
31. Naz A., Khan H.U., Bukhari A., Alshemaimri B., Daud A., Ramzan M. Machine and deep learning for personality traits detection: a comprehensive survey and open research challenges. *Artificial Intelligence Review*, 2025, vol. 58, no. 8, pp. 239. https://doi.org/10.1007/s10462-025-11245-3
32. Soto C.J., Jackson J.J. Five-factor model of personality. *Journal of Research in Personality*, 2013, vol. 42, pp. 1285–1302.
33. Ouarka A., Baha T.A., Es-Saady Y., El Hajji M. A deep multimodal fusion method for personality traits prediction. *Multimedia Tools and Applications*, 2024, vol. 84, no. 25, pp. 29665–29687. https://doi.org/10.1007/s11042-024-20356-y
34. Liu W., Sun Z., Wei S., Zhang S., Zhu G., Chen L. PS-GCN: psycholinguistic graph and sentiment semantic fused graph convolutional networks for personality detection. *Connection Science*, 2024, vol. 36, no. 1, pp. 2295820. https://doi.org/10.1080/09540091.2023.2295820
35. Akber M.A., Ferdousi T., Ahmed R., Asfara R., Rab R., Zakia U. Personality and emotion — a comprehensive analysis using contextual text embeddings. *Natural Language Processing Journal*, 2024, vol. 9, pp. 100105. https://doi.org/10.1016/j.nlp.2024.100105
36. Motlagh S.M.H., Rezvani M.H., Khounsiavash M. AI methods for personality traits recognition: a systematic review. *Neurocomputing*, 2025, vol. 640, pp. 130301. https://doi.org/10.1016/j.neucom.2025.130301
37. Zhang Y., Yang Q. A Survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021, vol. 34, no. 12, pp. 5586–5609. https://doi.org/10.1109/TKDE.2021.3070203
38. Li Y., Kazemeini A., Mehta Y., Cambria E. Multitask learning for emotion and personality traits detection. *Neurocomputing*, 2022, vol. 493, pp. 340–350. https://doi.org/10.1016/j.neucom.2022.04.049

recognition using PERS // Neural Computing and Applications. 2023. V. 35. P. 23927–23941. https://doi.org/10.1007/s00521-023-08962-7
40. Sturua S., Mohr I., Akram M.K., Günther M., Wang B., Krimmel M., et al. Jina Embeddings V3: multilingual text encoder with low-rank adaptations // Lecture Notes in Computer Science. 2025. V. 15576. P. 123–129. https://doi.org/10.1007/978-3-031-88720-8_21
41. Choure A.A., Adhao R.B., Pachghare V.K. NER in Hindi language using transformer model: XLM-RoBERTa // Proc. of the IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS). 2022. P. 1–5. https://doi.org/10.1109/icbds53701.2022.9935841
42. Xiao S., Liu Z., Zhang P., Muennighoff N., Lian D., Nie J.-Y. C-Pack: Packed resources for general chinese embeddings // Proc. of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024. P. 641–649. https://doi.org/10.1145/3626772.3657878
43. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., et al. Attention is all you need // Proc. of the 31st Conference on Neural Information Processing Systems. 2017. P. 1–11.
44. Gu A., Dao T. Mamba: linear-time sequence modeling with selective state spaces // arXiv. 2023. arXiv:2312.00752. https://doi.org/10.48550/arXiv.2312.00752
45. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision // Proc. of the 40th International Conference on Machine Learning. 2023. P. 28492–28518.
46. Demšar J. Statistical comparisons of classifiers over multiple data sets // Journal of Machine Learning Research. 2006. V. 7. P. 1–30.
47. Efron B., Tibshirani R. An Introduction to Bootstrap. Chapman and Hall/CRC, 1994. 456 p.
48. Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization // Proc. of the IEEE International Conference on Computer Vision (ICCV). 2017. P. 618–626. https://doi.org/10.1109/iccv.2017.74
49. Yang A., Li A., Yang B., Zhang B., Hui B., Zheng B., et al. Qwen3 technical report // arXiv. 2025. arXiv:2505.09388. https://doi.org/10.48550/arXiv.2505.09388

39. Talaat F.M., El-Gendy E.M., Saafan M.M., Gamel S.A. Utilizing social media and machine learning for personality and emotion recognition using PERS. *Neural Computing and Applications*, 2023, vol. 35, pp. 23927–23941. https://doi.org/10.1007/s00521-023-08962-7
40. Sturua S., Mohr I., Akram M.K., Günther M., Wang B., Krimmel M., et al. Jina Embeddings V3: multilingual text encoder with low-rank adaptations. *Lecture Notes in Computer Science*, 2025, vol. 15576, pp. 123–129. https://doi.org/10.1007/978-3-031-88720-8_21
41. Choure A.A., Adhao R.B., Pachghare V.K. NER in Hindi language using transformer model: XLM-RoBERTa. *Proc. of the IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*, 2022, pp. 1–5. https://doi.org/10.1109/icbds53701.2022.9935841
42. Xiao S., Liu Z., Zhang P., Muennighoff N., Lian D., Nie J.-Y. C-Pack: Packed resources for general chinese embeddings. *Proc. of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024, pp. 641–649. https://doi.org/10.1145/3626772.3657878
43. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., et al. Attention is all you need. *Proc. of the 31st Conference on Neural Information Processing Systems*, 2017, pp. 1–11.
44. Gu A., Dao T. Mamba: linear-time sequence modeling with selective state spaces. *arXiv*, 2023. arXiv:2312.00752. https://doi.org/10.48550/arXiv.2312.00752
45. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision. *Proc. of the 40th International Conference on Machine Learning*, 2023, pp. 28492–28518.
46. Demšar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 2006, vol. 7, pp. 1–30.
47. Efron B., Tibshirani R. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994, 456 p.
48. Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. https://doi.org/10.1109/iccv.2017.74
49. Yang A., Li A., Yang B., Zhang B., Hui B., Zheng B., et al. Qwen3 technical report. *arXiv*, 2025. arXiv:2505.09388. https://doi.org/10.48550/arXiv.2505.09388

Авторы

Коряковская Дарья Олеговна — стажер-исследователь, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, https://orcid.org/0009-0009-6207-8413, da.koryakovskaya@gmail.com
Аксёнов Александр Александрович — кандидат технических наук, старший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, SC 57203963345, https://orcid.org/0000-0002-7479-2851, a.aksenov95@mail.ru
Рюмина Елена Витальевна — кандидат технических наук, младший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, SC 57220572427, https://orcid.org/0000-0002-4135-6949, ryumina_ev@mail.ru
Рюмин Дмитрий Александрович — кандидат технических наук, старший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, SC 57191960214, https://orcid.org/0000-0002-7935-0569, dl_03.03.1991@mail.ru

Authors

Darya O. Koryakovskaya — Intern Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, https://orcid.org/0009-0009-6207-8413, da.koryakovskaya@gmail.com
Alexandr A. Axyonov — PhD, Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, SC 57203963345, https://orcid.org/0000-0002-7479-2851, a.aksenov95@mail.ru

Elena V. Ryumina — PhD, Junior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, SC 57220572427, https://orcid.org/0000-0002-4135-6949, ryumina_ev@mail.ru

Dmitry A. Ryumin — PhD, Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, SC 57191960214, https://orcid.org/0000-0002-7935-0569, dl_03.03.1991@mail.ru

Статья поступила в редакцию 09.11.2025
Одобрена после рецензирования 01.02.2026
Принята к печати 23.03.2026

Received 09.11.2025
Approved after reviewing 01.02.2026
Accepted 23.03.2026