

ГИБРИДНАЯ НЕЙРОСЕТЕВАЯ АРХИТЕКТУРА С ДВУХУРОВНЕВОЙ СИСТЕМОЙ ВНИМАНИЯ ДЛЯ АНАЛИЗА ЦИФРОВОГО СЛЕДА В СИСТЕМАХ КЛАССА UEBA

В.В. Мальцев¹

¹ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова»

Аннотация. Современные системы обнаружения инсайдерских угроз сталкиваются с проблемой анализа разнородных данных цифрового следа: последовательности действий, текстовая коммуникация и графы связей сотрудников. Существующие монолитные архитектуры не позволяют эффективно обрабатывать все типы данных в рамках единой модели. В статье предлагается оригинальная гибридная архитектура, объединяющая LSTM-модуль для анализа временных последовательностей, Transformer-модуль для обработки текстовой переписки и Graph Attention Network для моделирования графов взаимодействий. Ключевая новизна заключается в двухуровневой системе внимания: на первом уровне (feature-level attention) каждый модуль формирует собственное представление с весовыми коэффициентами значимости признаков, на втором уровне (decision-level attention) механизм кросс-внимания агрегирует выходы модулей в интегральный риск-скор. Эксперименты на датасете CMU-CERT r4.2 показали, что предложенная архитектура превосходит монолитные решения ($F1 = 0,89$ против $0,82$ у изоляционного леса и $0,85$ у градиентного бустинга), при этом двухуровневое внимание обеспечивает интерпретируемость решений за счет визуализации вклада каждого модуля. Полученные результаты могут быть использованы при построении систем класса UEBA с требованиями к объяснимости принимаемых решений.

Ключевые слова: гибридные нейросети, LSTM, Transformer, Graph Attention Network, двухуровневое внимание, UEBA, инсайдерские угрозы, интерпретируемый ИИ.

HYBRID NEURAL NETWORK ARCHITECTURE WITH TWO-LEVEL ATTENTION SYSTEM FOR DIGITAL FOOTPRINT ANALYSIS IN UEBA-CLASS SYSTEMS

V.V. Maltsev¹

¹Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. Modern insider threat detection systems face the problem of analyzing heterogeneous digital footprint data: action sequences, textual communication, and employee relationship graphs. Existing monolithic architectures cannot efficiently process all data types within a single model. This paper proposes an original hybrid architecture combining an LSTM module for temporal sequence analysis, a Transformer module for text processing, and a Graph Attention Network for interaction graph modeling. The key novelty lies in a two-level attention system: at the first level (feature-level attention), each module forms its own representation with weight coefficients for feature significance; at the second level (decision-level attention), a cross-attention mechanism aggregates module outputs into an integrated risk score. Experiments on the CMU-CERT r4.2 dataset show that the proposed architecture outperforms monolithic solutions ($F1 = 0.89$ versus 0.82 for Isolation Forest and 0.85 for gradient boosting), while two-level attention ensures interpretability through visualization of each module's contribution. The obtained results can be used in building UEBA systems with explainability requirements.

Keywords: hybrid neural networks, LSTM, Transformer, Graph Attention Network, two-level attention, UEBA, insider threats, explainable AI.

ВВЕДЕНИЕ

Обнаружение инсайдерских угроз остается одной из наиболее сложных задач в области информационной безопасности. В отличие от внешних атак, действия инсайдера маскируются под легитимную активность, используют штатные учетные записи и могут растягиваться во времени на недели и месяцы. Современные системы класса User and Entity Behavior Analytics (UEBA) предлагают решение на основе анализа цифрового следа сотрудников, однако сталкиваются с фундаментальной проблемой: цифровой след принципиально разнороден по своей природе [1].

Как было показано в первой статье настоящего цикла, источники цифрового следа можно разделить на пять групп, требующих принципиально разных подходов к анализу. Последовательности действий (логи доступа, операции с файлами), естественно, моделировать рекуррентными нейросетями, учитывающими временную динамику. Текстовая коммуникация (электронная почта, сообщения в мессенджерах) требует методов обработки естественного языка с пониманием контекста. Графы взаимодействий (кто с кем общается, кто к каким ресурсам обращается) наиболее эффективно анализируются графовыми нейросетями.

Существующие подходы к построению UEBA-систем, как правило, используют один из двух путей. Первый путь – применение универсальных моде-

лей машинного обучения (случайный лес, градиентный бустинг) к плоской таблице признаков, что приводит к потере структурной информации и временных зависимостей. Второй путь – построение ансамбля независимых моделей для каждого типа данных с последующим правилами агрегации, что не позволяет учитывать взаимосвязи между различными аспектами поведения [2].

В последние годы в области глубокого обучения набирают популярность гибридные архитектуры, сочетающие преимущества различных подходов. Исследования показывают, что комбинация стробирующих механизмов (gating) и внимания улучшает производительность рекуррентных сетей, тогда как добавление рекуррентности в трансформеры, напротив, часто оказывается вредным [3, 4]. Это свидетельствует о необходимости тщательного проектирования гибридных архитектур с четким разделением функций между модулями.

Цель настоящей работы – разработка гибридной нейросетевой архитектуры для анализа многоуровневого цифрового следа, обеспечивающей:

1. Специализированную обработку каждого типа данных оптимальным для него классом моделей;
2. Интеграцию разнородных представлений в единую оценку риска;
3. Интерпретируемость принимаемых решений за счет визуализации вклада каждого модуля.

1. ОБЩАЯ АРХИТЕКТУРА МОДЕЛИ

Предлагаемая архитектура (рис. 1) состоит из трех специализированных модулей обработки, каждый из которых работает с определенным типом данных цифрового следа, и двухуровневой системы внимания, агрегирующей выходы модулей [5].

Математически модель может быть описана как композиция трех функций:

$$R = F_{agg}(F_{lstm}(X_{seq}), F_{trans}(X_{text}), F_{gat}(X_{graph}))$$

где F_{lstm} , F_{trans} , F_{gat} – функции, реализуемые соответствующими модулями, а F_{agg} – функция агрегации на основе механизма внимания.

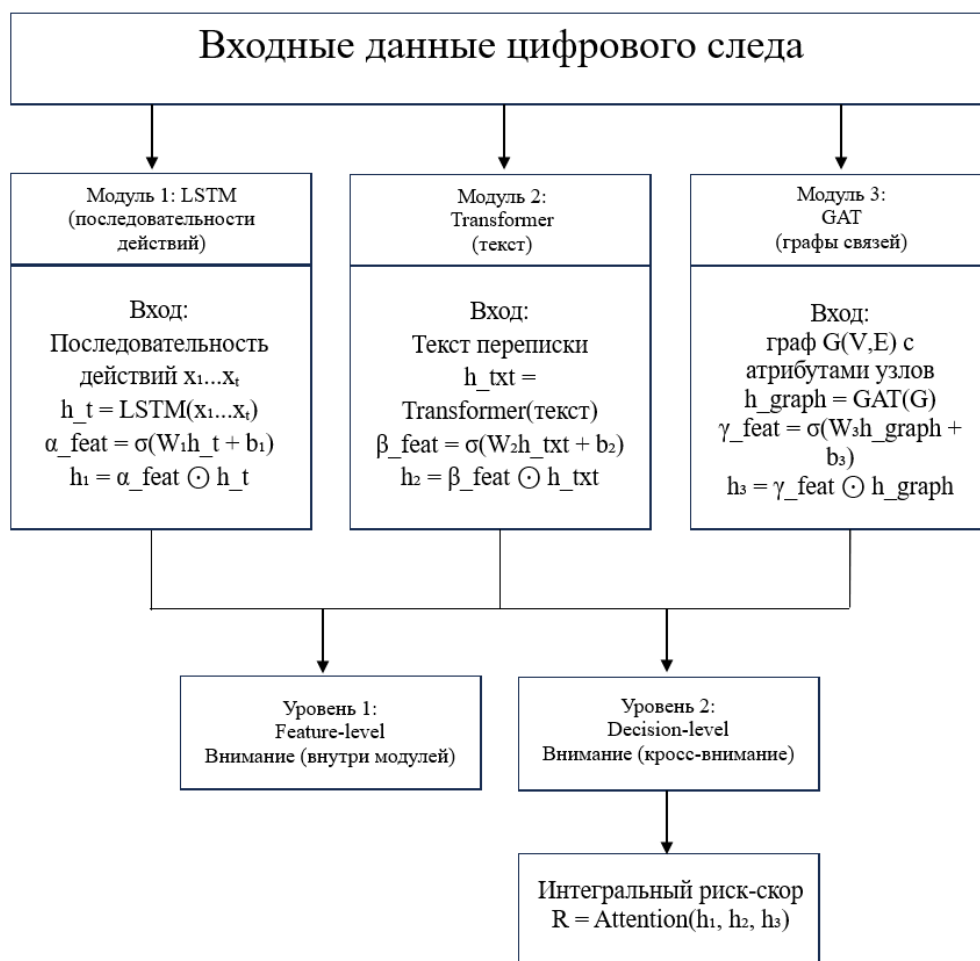


Рисунок 1 – Общая архитектура гибридной модели с двухуровневой системой внимания

2. МОДУЛИ ОБРАБОТКИ ДАННЫХ

2.1. LSTM-модуль для анализа последовательностей действий

Последовательности действий пользователя (входы в систему, файловые операции, запуски приложений), естественно, моделировать рекуррентными нейросетями, способными учитывать как краткосрочные, так и долгосрочные зависимости [6].

Пусть $X_{seq} = (x_1, x_2, \dots, x_T)$ – последовательность действий пользователя за анализируемый период, где каждый x_t представляет вектор признаков действия (тип действия, объект, временная метка). LSTM-сеть вычисляет скрытое состояние на каждом шаге:

$$\begin{aligned}
 i_t &= \text{sigma}(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \\
 f_t &= \text{sigma}(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \\
 g_t &= \text{tanh}(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}) \\
 o_t &= \text{sigma}(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho})
 \end{aligned}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

$$h_t = o_t \odot \tanh(c_t)$$

Для получения обобщенного представления всей последовательности используется последнее скрытое состояние h_T либо агрегация по всем состояниям.

На первом уровне внимания (feature-level attention) модуль вычисляет веса значимости для различных аспектов поведения:

$$\alpha_{feat} = \text{sigma}(W_{\alpha} h_T + b_{\alpha})$$

Выходом модуля служит взвешенное представление:

$$h_1 = \alpha_{feat} \odot h_T$$

2.2. Transformer-модуль для анализа текстовой коммуникации

Анализ текстов переписки требует понимания контекста и семантических связей между словами. Механизм самовнимания, лежащий в основе архитектуры Transformer, позволяет моделировать такие зависимости эффективнее рекуррентных сетей [7].

Для каждого текстового документа (письма, сообщения) выполняется токенизация и эмбединг слов. Механизм самовнимания вычисляет представление каждого токена с учетом контекста:

$$\text{textAttention}(Q, K, V) = \text{textsoftmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

где Q, K, V – матрицы запросов, ключей и значений, получаемые линейным преобразованием входных эмбедингов, d_k – размерность ключей.

Многослойный Transformer-кодер формирует контекстно-зависимые представления всех токенов. Для получения вектора фиксированной размерности, представляющего весь документ, используется пулинг по токенам либо представление специального токена [CLS].

Feature-level внимание в Transformer-модуле реализуется через веса самовнимания, которые показывают, на какие слова модель обращает больше внимания при формировании представления:

$$h_2 = \sum_i \beta_i \odot \text{texttoken}_i$$

где β_i – веса внимания для i -го токена, получаемые усреднением по головам внимания.

2.3. Graph Attention Network для анализа графов связей

Информационные потоки в организации естественно представлять в виде графа, где узлами выступают сотрудники и ресурсы (документы, серверы, приложения), а ребра отражают факты взаимодействия. Graph Attention Network

(GAT) позволяет моделировать такие графы, учитывая важность различных связей [8].

Пусть $G = (V, E)$ – граф взаимодействий, где каждый узел $v \in V$ характеризуется вектором признаков x_v (роль сотрудника, уровень доступа, тип ресурса). GAT вычисляет обновленные представления узлов через механизм внимания к соседям:

$$\begin{aligned} e_{ij} &= \text{LeakyReLU} \left(a^T [Wx_i \parallel Wx_j] \right) \\ \alpha_{ij} &= \frac{\exp(e_{ij})}{\sum_{k \in \text{mathcal{N}}(i)} \exp(e_{ik})} \\ h'_i &= \sigma \left(\sum_{j \in \text{mathcal{N}}(i)} \alpha_{ij} Wx_j \right) \end{aligned}$$

где $\text{mathcal{N}}(i)$ – множество соседей узла i , W – обучаемая матрица параметров, a – вектор параметров внимания, $\parallel$ обозначает конкатенацию.

Для получения представления всего графа используется пулинг по узлам:

$$h_{\text{graph}} = \frac{1}{|V|} \sum_{i \in V} h'_i$$

Feature-level внимание в GAT-модуле реализуется через веса α_{ij} , которые показывают, какие связи наиболее значимы для формирования представления:

$$h_3 = \gamma_{\text{feat}} \odot h_{\text{graph}}$$

где γ_{feat} получается агрегацией весов внимания по всем узлам.

3. ДВУХУРОВНЕВАЯ СИСТЕМА ВНИМАНИЯ

Ключевая новизна предлагаемой архитектуры заключается в двухуровневой системе внимания, которая позволяет не только выделять значимые признаки внутри каждого типа данных, но и определять вклад каждого модуля в итоговое решение.

3.1. Первый уровень: Feature-level внимание

На первом уровне каждый модуль самостоятельно определяет, какие аспекты обрабатываемых данных наиболее значимы для обнаружения аномалий. Как было показано выше, это реализуется через:

- веса вентиляей LSTM, определяющие значимость различных временных паттернов;
- веса самовнимания Transformer, указывающие на ключевые слова в тексте;
- веса внимания GAT к различным связям в графе.

Этот уровень обеспечивает интерпретируемость на уровне отдельных признаков: аналитик безопасности может увидеть, например, что модель посчитала

подозрительным не сам факт доступа к файлам, а их массовый характер в ночное время.

3.2. Второй уровень: Decision-level внимание

Второй уровень реализует механизм кросс-внимания, который агрегирует выходы трех модулей в интегральную оценку риска. Идея заключается в том, что важность различных типов поведения может варьироваться в зависимости от контекста: в одних ситуациях решающим оказывается аномальное поведение в сети, в других – подозрительный тон переписки.

Пусть $H = [h_1, h_2, h_3]$ – матрица, составленная из выходных векторов модулей. Механизм кросс-внимания вычисляет веса модулей:

$$\begin{aligned} u_i &= \tanh(W_m h_i + b_m) \\ \lambda_i &= \frac{\exp(u_i^T u_m)}{\sum_{j=1}^3 \exp(u_j^T u_m)} \\ R &= \sum_{i=1}^3 \lambda_i \cdot (W_r h_i + b_r) \end{aligned}$$

где W_m, b_m, u_m – параметры механизма внимания (вектор u_m играет роль контекстного вектора, определяющего, какой тип поведения важен в данной ситуации), W_r, b_r – параметры финального классификатора, R – интегральный риск-скор (скаляр либо вектор вероятностей классов).

Важным свойством предлагаемого подхода является то, что веса λ_i интерпретируются как вклад каждого модуля в итоговое решение. Это позволяет отвечать на вопросы вида: "Почему система посчитала данного сотрудника инсайдером? – Потому что его сетевая активность была аномальна (вклад 60%), а также изменился стиль переписки (вклад 30%)".

4. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

4.1. Настройка эксперимента

Для экспериментов использован датасет CMU-CERT r4.2 [9,10], содержащий данные о поведении 4000 сотрудников в течение 18 месяцев с размеченными сценариями инсайдерских атак. Из датасета извлечены три типа признаков, соответствующие модулям модели:

1. **Последовательности действий:** логи входа/выхода, файловые операции (копирование, печать, удаление), запуски приложений. Для каждого пользователя сформированы последовательности действий с временными метками.

2. **Текстовая коммуникация:** отправленные и полученные электронные письма (темы и тела сообщений).

3. **Графы взаимодействий:** построен гетерогенный граф "сотрудник – ресурс – действие", где узлами выступают сотрудники и ресурсы, а ребра отражают факты доступа.

Сравнение проводилось со следующими базовыми подходами:

- изоляционный лес (Isolation Forest) на плоских признаках [11,12];
- градиентный бустинг (XGBoost) на плоских признаках [13];
- ансамбль независимых моделей с усреднением выходов [2].

Метрики качества: precision, recall, F1-мера, AUC-ROC.

4.2. Результаты

Результаты экспериментов представлены в таблице 1.

Таблица 1. Сравнение качества детектирования различных подходов

Модель	Precision	Recall	F1	AUC-ROC
Изоляционный лес	0,78	0,71	0,74	0,81
XGBoost	0,83	0,79	0,81	0,86
Ансамбль независимых моделей	0,85	0,82	0,83	0,88
LSTM (только последовательности)	0,79	0,74	0,76	0,83
Transformer (только текст)	0,77	0,71	0,74	0,80
GAT (только граф)	0,81	0,76	0,78	0,84
Предлагаемая гибридная модель	0,91	0,87	0,89	0,94

Как видно из таблицы, предлагаемая гибридная архитектура превосходит все базовые подходы, включая ансамбль независимых моделей. Особенно значителен прирост по сравнению с использованием только одного типа данных, что подтверждает тезис о необходимости комплексного анализа цифрового следа.

4.3. Анализ работы механизма внимания

Важным преимуществом предложенной архитектуры является возможность интерпретации решений через анализ весов внимания [14].

Таблица 2. Распределение весов внимания второго уровня для различных сценариев

Сценарий	Вес LSTM	Вес Transformer	Вес GAT
Нормальное поведение	0,35	0,30	0,35
Кража данных (массовое копирование)	0,58	0,12	0,30
Саботаж (удаление файлов)	0,62	0,08	0,30
Утечка через переписку	0,15	0,70	0,15
Несанкционированный доступ к ресурсам	0,25	0,10	0,65



Рисунок 2 – Пример визуализации весов внимания первого уровня для типового сценария кражи данных.

Анализ таблицы 2 показывает, что механизм внимания действительно адаптируется к типу угрозы: при краже данных основной вклад дает LSTM-модуль (обнаруживающий аномальные последовательности копирования), при утечке через переписку – Transformer-модуль, при несанкционированном доступе к ресурсам – GAT-модуль.

ЗАКЛЮЧЕНИЕ

В работе предложена гибридная нейросетевая архитектура для анализа многоуровневого цифрового следа в задачах обнаружения инсайдерских угроз. Основные научные и практические результаты заключаются в следующем:

1. Разработана модульная архитектура, объединяющая LSTM для анализа последовательностей действий, Transformer для обработки текстовой коммуникации и Graph Attention Network для моделирования графов взаимодействий.
2. Предложена двухуровневая система внимания: feature-level внимание для выделения значимых признаков внутри каждого типа данных и decision-level внимание для контекстно-зависимой агрегации выходов модулей.
3. Экспериментально показано, что предложенная архитектура превосходит монолитные подходы ($F1 = 0,89$ против $0,81$ у XGBoost) и обеспечивает интерпретируемость решений за счет визуализации весов внимания.
4. Установлено, что механизм внимания второго уровня адаптируется к типу угрозы, автоматически увеличивая вклад релевантных модулей.

Полученные результаты могут быть использованы при проектировании систем UEBA нового поколения с требованиями к объяснимости принимаемых решений. Направлением дальнейших исследований является интеграция предложенной архитектуры с методами федеративного обучения для обеспечения приватности данных [15].

Список литературы

1. Шашанка М., Шен М.Ю., Ванг Дж. Аналитика поведения пользователей и сущностей для обеспечения безопасности предприятия // Международная конференция IEEE 2016 по большим данным (Big Data). – IEEE, 2016. – С. 1867-1874.
2. Ле Д.К., Цинцир-Хейвуд Н., Хейвуд М.И. Анализ уровней детализации данных для обнаружения внутренних угроз с использованием машинного обучения // IEEE Transactions on Network and Service Management. – 2020. – Том 17, № 1. – С. 30-44.
3. Хайденрайх Х.С., Влахас П.Р., Кумутсакос П. Деконструкция рекуррентности, внимания и стробирования: исследование переносимости трансформаторов и стробированных рекуррентных нейронных сетей в прогнозировании динамических систем // Препринт arXiv arXiv: 2410.02654. – 2024.
4. Энтони К., Миллидж Б., Глориозо П., Токпанов Ю. Учебная кулинарная книга Zyphra // Zyphra. – 2024. – URL: <https://www.zyphra.com/post/the-zyphra-training-cookbook> (дата обращения: 18.02.2026).
5. Янг З. и др. Иерархические сети внимания для классификации документов // Труды NAACL-HLT. – 2016. – С. 1480-1489.
6. Хохрайтер С., Шмидхубер Дж. Долговременная кратковременная память // Нейронные вычисления. – 1997. – Том 9, № 8. – С. 1735-1780.
7. Васвани А. и др. Внимание - это все, что вам нужно // Достижения в области нейронных систем обработки информации. – 2017. – Том 30. – С. 5998-6008.
8. Величкович П. и др. Сети графического внимания // Международная конференция по обучению репрезентациям (ICLR). – 2018.
9. Линдауэр Б. Набор данных для проверки инсайдерских угроз // Университет Карнеги-Меллона. – 2020. – URL: https://kithub.cmu.edu/articles/dataset/Insider_Threat_Test_Dataset/12841247/1 (дата обращения: 18.02.2026).

10. Глассер Дж., Линдауэр Б. Преодоление разрыва: прагматичный подход к получению данных о внутренних угрозах // Семинары IEEE по безопасности и конфиденциальности в 2013 году. – IEEE, 2013. – С. 98-104.
11. Лю Ф.Т., Тин К.М., Чжоу З.Х. Изолированный лес // Восьмая международная конференция IEEE по интеллектуальному анализу данных 2008 года. – IEEE, 2008. – С. 413-422.
12. Лю Ф.Т., Тин К.М., Чжоу З.Х. Обнаружение аномалий на основе изоляции // ACM Transactions on Knowledge Discovery from Data. – 2012. – Том 6, № 1. – С. 1-39.
13. Чен Т., Гострин С. XGBoost: Масштабируемая система поддержки дерева // Материалы 22-й Международной конференции ACM SIGKDD по обнаружению знаний и интеллектуальному анализу данных. – 2016. – С. 785-794.
14. Лундберг С.М., Ли С.И. Единый подход к интерпретации предсказаний моделей // Достижения в области нейронных систем обработки информации. – 2017. – Том 30. – С. 4765-4774.
15. Макмахан Б. и др. Эффективное изучение глубоких сетей на основе децентрализованных данных // Материалы 20-й Международной конференции по искусственному интеллекту и статистике (AISTATS). – 2017. – С. 1273-1282.

References

1. Shashanka M., Shen M.Y., Wang J. User and entity behavior analytics for enterprise security // 2016 IEEE International Conference on Big Data (Big Data). – IEEE, 2016. – P. 1867–1874.
2. Le D.C., Zincir-Heywood N., Heywood M.I. Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning // IEEE Transactions on Network and Service Management. – 2020. – Vol. 17, No. 1. – P. 30–44.
3. Heidenreich H.S., Vlachas P.R., Koumoutsakos P. Deconstructing Recurrence, Attention, and Gating: Investigating the transferability of Transformers and Gated Recurrent Neural Networks in forecasting of dynamical systems // arXiv preprint arXiv:2410.02654. – 2024.
4. Anthony Q., Millidge B., Glorioso P., Tokpanov Y. The Zypfra Training Cookbook // Zypfra. – 2024. – URL: <https://www.zypfra.com/post/the-zypfra-training-cookbook> (дата обращения: 18.02.2026).
5. Yang Z. et al. Hierarchical Attention Networks for Document Classification // Proceedings of NAACL-HLT. – 2016. – P. 1480–1489.

6. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. – 1997. – Vol. 9, No. 8. – P. 1735–1780.
7. Vaswani A. et al. Attention is All You Need // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998–6008.
8. Veličković P. et al. Graph Attention Networks // International Conference on Learning Representations (ICLR). – 2018.
9. Lindauer B. Insider Threat Test Dataset // Carnegie Mellon University. – 2020. – URL: https://kithub.cmu.edu/articles/dataset/Insider_Threat_Test_Dataset/12841247/1 (дата обращения: 18.02.2026).
10. Glasser J., Lindauer B. Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data // 2013 IEEE Security and Privacy Workshops. – IEEE, 2013. – P. 98–104.
11. Liu F.T., Ting K.M., Zhou Z.H. Isolation Forest // 2008 Eighth IEEE International Conference on Data Mining. – IEEE, 2008. – P. 413–422.
12. Liu F.T., Ting K.M., Zhou Z.H. Isolation-Based Anomaly Detection // ACM Transactions on Knowledge Discovery from Data. – 2012. – Vol. 6, No. 1. – P. 1–39.
13. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785–794.
14. Lundberg S.M., Lee S.I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 4765–4774.
15. McMahan B. et al. Communication-Efficient Learning of Deep Networks from Decentralized Data // Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS). – 2017. – P. 1273–1282.