http://ma123.ru **ЦИТИСЭ** № 2 (48) 2026 **ISSN** 2409-7616

© Н.Д. Берман

Научная статья
УДК 378.1

МОДЕЛЬ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ И АНТИФЕЙКОВОЙ ГРАМОТНОСТИ В УСЛОВИЯХ ТОТАЛЬНОГО ДИПФЕЙКА

Н.Д. Берман

Берман Нина Демидовна,
старший преподаватель, Тихоокеанский государственный университет, Хабаровск, Россия.
ORCID: 0000-0002-3573-048X
nina.berman@mail.ru

Аннотация. *В условиях стремительного развития технологий синтетических медиа (генеративно-состязательных сетей GAN, диффузионных моделей, технологий голосового клонирования) особую актуальность приобретает проблема «тотального дипфейка», высокореалистичных мультимодальных подделок, практически неотличимых от подлинного контента. Традиционные подходы к информационной безопасности и медиаграмотности, ориентированные на выявление текстовых фейков и фальсифицированных фотоизображений, становятся малоэффективными. Тотальный дипфейк затрагивает основы доверия к визуальным и аудиоданным, меняет механизмы распространения информации и требует пересмотра существующих подходов к верификации контента на институциональном и индивидуальном уровнях. Целью исследования является представление модели формирования компетенций в области информационной безопасности и антифейковой грамотности в условиях тотального дипфейка. В работе использованы результаты научных исследований по психологии восприятия и когнитивным искажениям, а также данные о современных технологиях дипфейков. Предложенная в работе модель описывает из каких компонентов складывается антифейковая грамотность в контексте тотального дипфейка и как происходит поэтапное формирование устойчивости от интуитивного уровня до рефлексивно-проактивного. Её внедрение будет способствовать повышению устойчивости обучающихся и общества в целом к угрозам, связанным с распространением синтетических медиаматериалов. Разработанные рекомендации могут быть интегрированы в образовательные программы, тренинги по цифровой грамотности и системы повышения квалификации.*

Ключевые слова: *дипфейк, синтетический контент, информационная безопасность, медиаграмотность, антифейковая грамотность, критическое мышление, генеративно-состязательные сети.*

Библиографическая ссылка: *Берман Н.Д. Модель формирования компетенций в области информационной безопасности и антифейковой грамотности в условиях тотального дипфейка // ЦИТИСЭ. 2026. № 2. С. 448-457.*

Research Full Article
UDC 378.1

MODEL FOR THE DEVELOPMENT OF COMPETENCIES IN THE FIELD OF INFORMATION SECURITY AND ANTI-FAKE LITERACY IN THE CONTEXT OF TOTAL DEEPFAKE

N.D. Berman

Nina D. Berman,
Senior Lecturer, Pacific National University,
Khabarovsk, Russian Federation.
ORCID: 0000-0002-3573-048X
nina.berman@mail.ru

Abstract. *With the rapid development of synthetic media technologies (generative adversarial networks (GANs), diffusion models, and voice cloning technologies), the problem of "total deepfakes"—highly realistic multimodal fakes virtually indistinguishable from authentic content—has become particularly pressing. Traditional approaches to information security and media literacy, focused on identifying text fakes and falsified photographs, are becoming ineffective. Total deepfakes undermine the foundations of trust in visual and audio data, change the mechanisms of information dissemination, and require a revision of existing approaches to content verification at the institutional and individual levels. The aim of this study is to present a model for developing competencies in information security and anti-fake literacy in the context of total deepfakes. This paper draws on research findings on the psychology of perception and cognitive biases, as well as data on modern deepfake technologies. The model proposed in this paper describes the components of anti-fake literacy in the context of pervasive deepfakes and how resilience develops gradually, from the intuitive to the reflexive-proactive level. Its implementation will contribute to increasing the resilience of students and society as a whole to the threats posed by the spread of synthetic media. The developed recommendations can be integrated into educational programs, digital literacy training, and professional development systems.*

Keywords: *deepfake, synthetic content, information security, media literacy, antifake literacy, critical thinking, generative adversarial networks.*

For citation: Berman N. D. (2026). Model for the development of competencies in the field of information security and anti-fake literacy in the context of total deepfake. *CITISE, 2*, 448-457. (In Russian).

Введение.

Развитие генеративного искусственного интеллекта привело к появлению новых форм цифрового мультимедийного контента, среди которых особое место занимают дипфейки (deepfake). Термин «дипфейк» происходит от «глубокого обучения» (deep learning), метода, используемого для генерации контента, и «подделки» (fake), указывающего на то, что сгенерированный контент не является подлинным. Технология дипфейков имеет множество полезных применений в образовании и развлечениях, но она также поднимает множество

этических, социальных проблем и вопросов безопасности, таких как кража личных данных, распространение ложной информации и нарушение конфиденциальности. Если на ранних этапах (2017–2019 гг.) дипфейки характеризовались заметными визуальными артефактами, позволявшими относительно легко отличить синтетическое видео от реального, то современное состояние технологий, так называемый «тотальный дипфейк», отличается высокой реалистичностью, при которой даже экспертный визуальный анализ часто не позволяет выявить подделку без специальных инструментов (исчезли типичные артефакты, такие как неестественное моргание, сбои в области зубов, некорректные отражения в глазах). Инструменты для создания дипфейков стали общедоступными, сегодня для генерации видео достаточно базовых навыков компьютерной грамотности и свободного ПО. Современные дипфейки позволяют синхронизировать подмену лица, голоса и текста (например, сгенерированная «речь» политического лидера в реальном времени), что создает эффект полного присутствия и усиливает доверие к фейковому контенту [1]. Дипфейки распространяются через социальные сети, мессенджеры, новостные ресурсы и блоги, достигая многомиллионной аудитории за считанные часы [2]. Параллельно происходит крах традиционной парадигмы «видео = доказательство». На протяжении десятилетий видеозапись считалась одним из наиболее надежных видов свидетельств в журналистике, судопроизводстве и социальных коммуникациях. Тотальный дипфейк подрывает этот фундамент, теперь любое видео, даже с естественной мимикой, потенциально может быть полностью синтетическим [3; 4].

Несмотря на растущие угрозы, уровень антифейковой грамотности общества остаётся критически низким и нуждаются в формировании устойчивости к дипфейковым манипуляциям [5]. Уже зафиксированы случаи мошенничества с голосовым клонированием (вымогание денег под видом родственника в беде), репутационных атак на публичных лиц и политических дезинформационных кампаний с использованием дипфейков. С другой стороны, отсутствует системная теоретическая модель, которая бы описывала, как именно должна формироваться информационная безопасность и антифейковая грамотность на уровне личности и социальных групп в новых условиях. Существующие подходы исследуют либо только технические, либо только поведенческие аспекты и не учитывают специфику именно тотального дипфейка как уникального когнитивного и технологического вызова. Проблемная ситуация заключается в несоответствии между нарастающей угрозой тотального дипфейка и отсутствием целостной концептуальной рамки для построения защитных компетенций у студентов и всего общества в целом.

Исследования показывают [6; 7], что пользователь, столкнувшийся с дипфейком в своем мессенджере или соцсети, в подавляющем большинстве случаев не предпринимает верификацию, даже если техническая возможность существует. Основной причиной являются сформированные когнитивные привычки, доверие к видео как к доказательству и автоматическая реакция репоста при сильном эмоциональном воздействии. Поэтому, без изменения этих привычек, любые технические барьеры остаются легко преодолимыми через социальную инженерию. В современном цифровом обществе назрела необходимость концептуального перехода от реактивной защиты к проактивной антифейковой грамотности. Реактивная защита подразумевает ответ на уже появившуюся угрозу (обновление базы детекторов, удаление разоблачённого дипфейка), а проактивная грамотность означает формирование у человека устойчивых когнитивных алгоритмов, действующих до контакта с дипфейком, привычка проверять источник. Анализ научной литературы показывает, что проблематика дипфейков разрабатывается в рамках нескольких дисциплин: информационная безопасность (основное внимание уделяется техническим и организационным мерам, предлагается защита системы, но не личности [8]), медиаграмотность даёт общие методики для противодействия фейковым новостям [9], но дипфейки подрывают доверие к

видеоформату, которое в классических моделях медиаграмотности не подвергается сомнению. Технические подходы к распознаванию синтетических медиа (решения достигли высокой точности распознавания до 95–99% [10]) без изменения человеческого фактора не способны решить проблему на уровне общества.

Целью исследования является представление модели формирования компетенций в области информационной безопасности и антифейковой грамотности в условиях тотального дипфейка.

Материалы и методы исследования.

В статье были использованы результаты теоретических разработок и практических исследований в сфере психологии восприятия, когнитивных искажениях, технологий дипфейков. Проанализированы современные инструменты создания и детекции синтетического контента, а также изучены механизмы восприятия медиаматериалов аудиторией.

Результаты исследования и их обсуждение.

Достижения в области искусственного интеллекта (ИИ) позволили разработать алгоритмы, создающие видеоконтент, практически неотличимый от реального мира. Под тотальным дипфейком будем понимать синтетический мультимедийный контент (видео, аудио, фото, либо их комбинация с синхронизированной генерацией текста или лицевой мимики), созданный с использованием технологий искусственного интеллекта и машинного обучения, в котором полностью отсутствует подлинная основа (контент не является модификацией реального материала, а генерируется «с нуля»), достигается максимальная степень реалистичности, имитируется не только внешность, но и поведение (мимика, жесты, интонации, речевые особенности, стиль общения). И если для обнаружения классического дипфейка возможен визуальный анализ, то для тотального дипфейка уже требуется специализированные программы.

Антифейковая грамотность применительно к дипфейкам является интегральной компетенцией, включающей:

– знание типов дипфейков и их артефактов, природы синтетических медиа;

– умение оперативно верифицировать медиаконтент, комбинируя интуитивные эвристики (поиск несоответствий, проверка источника) с цифровыми инструментами анализа;

– привычку (автоматизированное поведение) подвергать сомнению визуально-аудиальный контент, особенно вызывающий сильные эмоции;

– этическую позицию на нераспространение непроверенного или заведомо синтетического контента, информирование окружающих о потенциальных рисках такого контента.

Отличие от традиционной медиаграмотности заключается в смещении фокуса критического анализа, где помимо семантики и источника информации, подвергается проверке сам формат представления данных (видео может не отражать реальность, а имитировать её).

Информационную безопасность (в контексте дипфейков) будем рассматривать в двух аспектах: объективный (защита человека от финансового, репутационного или психологического ущерба вследствие воздействия дипфейков) и субъективный (формирование у человека устойчивых паттернов поведения, минимизирующих вероятность стать жертвой или невольным распространителем дипфейка).

Методологическая основа модели формируется за счёт интеграции трёх групп теорий:

– когнитивно-психологических (анализируются когнитивные искажения, механизмы восприятия и обработки информации, обуславливающие уязвимость к фейкам) [11; 12];

– педагогических (определяются стратегии и методы обучения, способствующие формированию антифейковой грамотности и информационной устойчивости) [13];

– системно-информационных (задаются принципы объединения различных уровней и компонентов в единую функциональную систему).

Даниэль Канеман в книге «Думай быстро и медленно» описывает два различных режима когнитивной обработки информации: системы мышления 1 и 2 [14]. Система 1 (быстрая, автоматическая и интуитивная, работающая практически без усилий) позволяет принимать решения и выносить суждения на основе закономерностей и опыта, отвечает за мгновенное доверие к реалистичному видео, особенно если оно вызывает сильный аффект (гнев, страх, восторг). В отличие от нее, Система 2 – медленная, обдуманная и осознанная, требующая целенаправленных усилий. Этот тип мышления используется для решения сложных задач, где необходимо больше размышлений и обдуманного подхода, и именно она должна активироваться для верификации информации, но по умолчанию находится в «спящем режиме».

Теория нулевого доверия (Zero Trust) гласит «никогда не доверяй, всегда проверяй». Этот принцип для контента должен стать поведенческой установкой, согласно которой ни один медиаобъект (видео, аудио, изображение) не принимается как достоверный по умолчанию, независимо от формата (видео не даёт дополнительных преимуществ перед текстом), источника распространения (даже официальный канал или знакомый человек не могут являться гарантией правдивой информации), эмоционального воздействия и визуального правдоподобия (современные дипфейки сложно отличить от реальности без специализированных программ). Практическим следствием применения данной концепции в образовательном процессе становится внедрение алгоритмизированных протоколов верификации, обучающийся усваивает не абстрактный принцип сомнения, а конкретную последовательность шагов: сделать паузу перед любым репостом, задействовать технические инструменты анализа, оценить контекст появления видео и только после этого выносить суждение о его достоверности. Теория нулевого доверия Zero Trust трансформируется в поведенческую модель, формирующую устойчивость к информационным манипуляциям.

Дайана Халперн в своей книге определяет критическое мышление как «использование когнитивных техник или стратегий, которые увеличивают вероятность получения желаемого конечного результата» [15] (в данном случае не быть обманутым), включают «метакомпоненты, которые используются для планирования, оценки и наблюдения» (перед репостом проверь источник информации), «компоненты усвоения знания, которые заключают в себе способность обработать информацию» (осознание собственной уязвимости к дипфейкам), «исполняющие компоненты, представляющие из себя навыки мышления, которые используются» («пауза» перед действием).

Более ранние теории информационной безопасности относили человека к категории «слабого звена». Современные исследователи пересмотрели эту позицию, фокусируясь на активном формировании защитных поведенческих реакций. Это достигается через осознанное принятие рисков и создание устойчивых привычек [16]. Достичь этого можно, развивая в обучающемся способность к распознаванию рисков и доводя до автоматизма определённые алгоритмы безопасного поведения. Наиболее действенным инструментом здесь выступает методика «прививки». Практическим примером такого подхода служит разбор учебных дипфейков, позволяя выработать навыки распознавания подделок и повышая устойчивость к подобным угрозам в реальной жизни.

Системный подход требует рассматривать формирование антифейковой грамотности не как линейную совокупность знаний и умений, а как эмерджентное свойство взаимодействия трех подсистем: технической (инструменты), когнитивной (эвристики) и поведенческой (гигиена).

Достижение устойчивости личности и социальной группы к дипфейковым манипуляциям в предлагаемой модели определяется совокупностью сформированных

компетенций: способности распознавать синтетический контент, выстраивать адекватную реакцию на его обнаружение и проактивно снижать информационную уязвимость (изменение информационных привычек). Объектами формирования (элементы личности и социальной группы, подвергающиеся трансформации в рамках модели) являются:

– когнитивные установки (снижение уровня доверия к видео как безусловному доказательству и формирование критического отношения к непроверенным визуальным материалам);

– знания (расширение знаний о феномене дипфейков, включая понимание их природы, классификацию по типам, а также представление о технологиях и инструментах их создания);

– навыки верификации (освоение конкретных алгоритмов верификации информации, позволяющих выявлять признаки синтетического контента);

– модели поведения (внедрение устойчивых поведенческих привычек в цифровой коммуникации: пауза перед распространением контента, использование инструментов фактчекинга для предварительной проверки).

Защита от дипфейков формируется только при одновременном развитии в трех направлениях:

–техническое – знание основных типов дипфейков (замена лица, ресинхронизация губ, генерация голоса, полная генерация сцены), понимание артефактов (в ранних генеративно-состязательных нейронных сетях (GAN) обнаруживались проблемы с зубами, ушами, отражениями, а в современных моделях можно заметить аномалии в частоте моргания, микротики век, несоответствие освещения направлению источника света), умение пользоваться инструментами детекции (онлайн-сервисы, боты в Telegram, расширения браузера), владение базовыми навыками форензики (проверка метаданных, обратный поиск кадра);

–когнитивное (критическое мышление) – изменение установок восприятия от безусловного доверия к медиаконтенту к установке «нулевого доверия» к видео/аудио по умолчанию, применение эвристики «паузы сомнения» (искусственное замедление работы интуитивной Системы 1 (по Канеману), когда контент вызывает сильные эмоции, чтобы включить аналитическое мышление), осознание собственной уязвимости перед реалистичным синтетическим контентом, умение задавать вопросы: откуда взялся этот контент, кто его автор и распространитель, зачем это опубликовано и какую реакцию хотят вызвать, соответствует ли реальным событиям?

–этическое (поведенческое) –формирование внутренней позиции, когда правила безопасности в сети становятся частью самого человека и он не распространяет то, что не проверил, использует паузу перед любым репостом, не генерирует правдоподобный дипфейк ради смеха или розыгрыша, осознает ответственность за то, что публикует и пересылает.

Формирование устойчивости к дипфейкам представляет собой циклический процесс. Исходным моментом цикла выступает столкновение с дипфейком (или его учебной моделью) как достоверным медиаконтентом. Обучающийся видит синтетическое видео, которое на первый взгляд выглядит совершенно реальным. Дальше следует фаза когнитивного диссонанса, возникающая в результате последующего разоблачения (либо идет проверка контента через внешний источник верификации, либо с помощью специальных программ детекции). Обучающийся фиксирует противоречие между первоначальным доверием к видео и установленным фактом его синтетической природы, что создаёт внутреннее противоречие «Я только что был уверен, что это правда, а оказывается, меня обманули!». И это психологическое напряжение запускает следующий шаг, рефлексию. Ключевым условием эффективности данного этапа является смещение фокуса с самообвинения («почему моё доверие к этому видео оказалось ошибочным») на выявление ключевых признаков дипфейка (аномалии мимики, неестественная синхронизация аудио и видео, артефакты синтеза и т.д.),

каналов восприятия, оказавшихся наиболее уязвимыми (визуальный, аудиальный, контекстный), определение, что повлияло на первоначальное доверие (эффект первичности, склонность верить визуальным данным, эффект авторитета и пр.). Рефлексия завершается формированием вывода о необходимом корректирующем действии. Четвёртый этап (приобретение нового когнитивного или инструментального ресурса) представляет собой процесс научения, в ходе которого обучающийся изучает методы верификации медиаконтента, использует сервисы детекции дипфейков, осваивает эвристические правила оценки информации (например, «правило трёх источников»). Далее полученные знания и навыки закрепляются посредством практической отработки на специально подобранных учебных кейсах, где заведомо известно, какие из видео дипфейки. Обучающиеся учатся применять освоенные инструменты для определения различных дипфейков. Многократное успешное применение создаёт предпосылки для автоматизации, навык начинает переходить в привычку, и пауза сомнения включается без сознательного усилия, как только перед глазами оказывается эмоционально заряженное видео. На заключительном этапе цикла происходит расширение контекста. Обучающийся не останавливается на одном типе дипфейков, если сначала он научился распознавать подмену лица в видео, то потом начинает применять те же самые приёмы к голосовым дипфейкам, к синхронизации губ, к полностью сгенерированным сценам.

Цикл повторяется на более сложных уровнях, каждый раз повышая уровень защиты, потому что технологии дипфейков не стоят на месте, и каждый следующий цикл готовит к угрозам, которых раньше не существовало. Ключевой элементом всего процесса должна стать когнитивная пауза (задержка между получением контента и действием доверия/репоста). Она должна прерывать автоматическую реакцию «увидел – поверил – поделился» и давать время для осознанного анализа (активации Системы 2). Обучение контекстно-критическому анализу без освоения инструментального уровня (технические средства детекции) и определения базовых визуальных или акустических признаков дипфейка будет недостаточным, у обучающегося не сформируется «когнитивный каркас» для глубокого анализа. Необходимо оценивать в первую очередь надёжность источника (проверенный новостной канал, официальный аккаунт), чем отсутствие визуальных артефактов у данного видео. Модель учит определять прежде всего канал доставки, а не контент (хотя и контент проверяется). Никогда не полагаться на единственный признак, если видео прошло один детектор, это еще не гарантия, требуется как минимум два разных способа верификации (инструментального и контекстного или двух разных инструментов). Эффективность модели повышается, если обучающийся не только проверяет, но и анализирует свои ошибки (почему поверил, какой канал пропустил). Без рефлексии даже инструментальная проверка не переходит в когнитивную привычку. Нужно научить избегать крайности, чтобы предотвратить превращения обучающегося в параноика, отвергающего реальные видео. Это достигается через обучение вероятностной оценке (степень уверенности) и сравнению с другими источниками, а не через абсолютное «всё ложь».

Для формирования компетенций в области информационной безопасности и антифейковой грамотности необходимо знание базовой цифровой грамотности как предпосылки. Для человека, не умеющего пользоваться поиском, браузером или мессенджером, инструментальный уровень (применение специального ПО) недоступен. Модель в таком случае ограничивается социальными механизмами (информационная гигиена, правило «3 секунды паузы перед репостом», помощь других людей). Модель не должна быть статичной, необходимо обновление с появлением новых типов дипфейков. То, что было эффективным признаком в 2024 году (артефакты в контексте генеративно-состязательных сетей GAN), может не работать в 2026 (диффузионные модели без артефактов), на каждом этапе обучения должны включаться актуальные кейсы.

Формирование навыка детекции дипфейков должно следовать от явных, легко наблюдаемых признаков к скрытым, требующим инструментов или аналитики. Попытка сразу обучать инструментальной детекции или контекстному анализу ведёт к когнитивной перегрузке [17, 18]. Наиболее эффективный способ понять уязвимости технологии – это воспроизвести её в упрощённом, контролируемом виде. Если обучаемый сам создаст примитивный дипфейк (например, с помощью мобильного приложения или онлайн-сервиса),то он увидит, какие именно артефакты возникают при автоматической генерации и начнёт их замечать у других, поймёт, как мало усилий нужно для создания обманчивого видео и осознает масштаб угрозы, снизит моральный порог для критического анализа («я сам могу такое сделать, значит, и другие смогут»). Знание активно конструируется обучаемым через действие, а не пассивно усваивается. Создание артефакта (дипфейка) даёт более глубокое понимание, которое является недостижимым при простом просмотре примеров.

Заключение.

В исследовании была предпринята попытка ответить на вызов, который тотальный дипфейк бросает современной системе информационной безопасности и медиаобразования. Проведённый анализ показал, что традиционные подходы (классическая информационная безопасность, рассматривающая человека как «слабое звено»; общая медиаграмотность, не специфицированная под дипфейки; технические детекторы, игнорирующие человеческий фактор) по отдельности не способны обеспечить устойчивость личности к тотальным дипфейкам. Предложенная в работе модель описывает из каких компонентов складывается антифейковая грамотность в контексте тотального дипфейка и как происходит поэтапное формирование устойчивости от интуитивного уровня до рефлексивно-проактивного.

Тотальный дипфейк является не рядовым обновлением технологий фальсификации, а результатом развития генеративно-состязательных сетей (GAN), трансформеров, алгоритмов синтеза речи и обработки естественного языка [19]. В совокупности они позволяют создавать контент, неотличимый от реального, без существенных временных и ресурсных затрат. Тотальный дипфейк затрагивает основы доверия к визуальным и аудиоданным, меняет механизмы распространения информации и требует пересмотра существующих подходов к верификации контента на институциональном и индивидуальном уровнях. В цифровом мире, где дипфейки становятся нормой, информационная безопасность и антифейковая грамотность становится не просто полезной компетенцией, а необходимостью, чтобы уметь защищаться от манипуляций, обмана и конфликтов.

Список источников:

1. Неренц Д.В. Дипфейк как одна из главных информационных угроз XXI века // Филология: научные исследования. 2025. № 9. С. 96-111. URL: https://www.elibrary.ru/tqdnwc
2. Кондратская В.Л., Басалай С.С. Дипфейки в тренде: возможности использования технологии в коммуникациях // МедиаВектор. 2025. № 17. С. 37-41. URL: https://www.elibrary.ru/hzszvo
3. Бодров Н.Ф., Лебедева А.К. Дипфейк: технологии, правовое регулирование и судебно-экспертное исследование. - Москва: Общество с ограниченной ответственностью "Проспект", 2026. - 224 с. URL: https://www.elibrary.ru/dqbtne
4. Красножон Е.А. Дипфейки: новая реальность в криминалистике // Юстиция. 2025. № 2. С. 91-97. URL: https://www.elibrary.ru/qrgcpf
5. Маглинова Т.Г., Горских Д.А., Паршин Д.С., Дрогин Е.В. Социальная инженерия: мощное оружие киберпреступников // Естественно-гуманитарные исследования. 2025. № 4(60). С. 330-333. URL: https://www.elibrary.ru/kxgkzj

6. Самойлов И.Н., Камышанский Д.Ю. Потенциальные и реальные угрозы технологии deepfake (дипфейк) // Интеллектуальная собственность. Авторское право и смежные права. 2024. № 4. С. 17-22. URL: https://www.elibrary.ru/cdbuas

7. Kaur A., Noori Hoshyar A., Saikrishna V., Firmin S., Xia F. Deepfake video detection: challenges and opportunities // Artif Intell Rev. 2024. Vol. 57, 159. DOI: https://doi.org/10.1007/s10462-024-10810-6

8. Андриянов А.М., Баюк О.В. Дипфейк как угроза информационной безопасности России и инструмент манипуляции и обмана // Научно-технический вестник Поволжья. 2025. № 1. С. 169-172. URL: https://www.elibrary.ru/iaemqz

9. Поздняков Е.И. Актуальные методы противодействия фейковым новостям // Гражданин. Выборы. Власть. 2023. № 3(29). С. 173-180. URL: https://www.elibrary.ru/tghuom

10. Sharma V.K., Garg R., Caudron Q. A systematic literature review on deepfake detection echniques // Multimed Tools Appl. 2025. Vol. 84. P. 22187–22229. DOI: https://doi.org/10.1007/s11042-024-19906-1

11. Кулиева А.К., Тихонов Р.В., Иванчей И.И. Теории и практики измерения осознания в экспериментальных когнитивных исследованиях // Экспериментальная психология. 2021. Т. 14, № 4. С. 164-181. DOI: https://doi.org/10.17759/exppsy.2021140409

12. Nistor N., Artmann B., Isik D., Neziri N., Stanciu D. Knowledge types and cognitive processing modes associated with online news credibility assessment: An interview study // Computers in Human Behavior. 2025. Vol. 171. DOI: https://doi.org/10.1016/j.chb.2025.108721

13. Берман Н.Д. Цифровые навыки и цифровая грамотность студентов // Russian Journal of Education and Psychology. 2021. Т. 12, № 2-2. С. 26-34. URL: https://www.elibrary.ru/rzypeq

14. Kahneman Thinking, fast and slow. - Farrar, Straus and Giroux. 2011.

15. Халперн Д. Психология критического мышления. - СПб.: Питер, 2000. - 512 с.

16. Мантуленко В. В. Информационная безопасность личности в контексте российского образования на современном этапе // Непрерывное образование: XXI век. 2024. № 2(46). С. 124-138. URL: https://www.elibrary.ru/rdquuz

17. Евплов Н.А. Оценка методов обнаружения дипфейков в видеоконтенте // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2025. № 10-2. С. 17-22. URL: https://www.elibrary.ru/mwqajb

18. Рыбаков А.А., Фрейлехман С.А. Автоматическое детектирование аудиодефектов с применением параллельных вычислений // Программные продукты и системы. 2022. № 3. С. 428-437. URL: https://www.elibrary.ru/ywtpfn

19. Sharma V.K., Garg R., Caudron Q. A systematic literature review on deepfake detection techniques // Multimed Tools Appl. 2025. Vol. 84(20). P. 22187–22229. DOI: https://doi.org/10.1007/s11042-024-19906-1

References:

1. Nerents D. V. (2025). Deepfake as one of the main information threats of the 21st century. *Philology: scientific research, 9,* 96–111. (In Russian).: https://www.elibrary.ru/tqdnwc

2. Kondratskaya V. L., & Basalai S. S. (2025). Deepfakes in trend: possibilities of using technology in communications. *MediaVector*, 17, 37–41. (In Russian). https://www.elibrary.ru/hzszvo

3. Bodrov N. F., & Lebedeva A. K. (2026). *Deepfake: technologies, legal regulation and forensic research.* Moscow: Prospect Limited Liability Company, 224 p. (In Russian). https://www.elibrary.ru/dqbtne

4. Krasnozhon E. A. (2025). Deepfakes: A New Reality in Forensic Science. *Justice, 2,* 91–97. (In Russian). https://www.elibrary.ru/qrgcpf

5. Maglinova T. G., Gorskikh D. A., Parshin D. S., & Drogin E. V. (2025). Social Engineering: A Powerful Weapon of Cybercriminals. *Research in Natural Sciences and Humanities*, *4*(60), 330–333. (In Russian). https://www.elibrary.ru/kxgkzj

6. Samoilov I. N., & Kamyshansky D. Yu. (2024). Potential and Real Threats of Deepfake Technology. *Intellectual Property. Copyright and Related Rights*, *4*, 17–22. (In Russian). URL: https://www.elibrary.ru/cdbuas

7. Kaur A., Noori Hoshyar A., Saikrishna V., Firmin S., & Xia F. (2024). Deepfake video detection: challenges and opportunities. *Artif Intell Rev*, 57, 159. https://doi.org/10.1007/s10462-024-10810-6

8. Andrianov A. M., & Bayuk O. V. (2025). Deepfake as a threat to Russia's information security and a tool for manipulation and deception. *Scientific and Technical Bulletin of the Volga Region*, *1*, 169–172. (In Russian). https://www.elibrary.ru/iaemqz

9. Pozdnyakov E. I. (2023). Current methods of countering fake news. *Citizen. Elections. Power*, *3*(29), 173–180. (In Russian). https://www.elibrary.ru/tghuom

10. Sharma V.K., Garg R., & Caudron Q. (2025). A systematic literature review on deepfake detection echniques. *Multimed Tools Appl*, *84*, 22187–22229. https://doi.org/10.1007/s11042-024-19906-1

11. Kulieva A.K., Tikhonov R.V., & Ivanchey I.I. (2021). Theories and practices of measuring awareness in experimental cognitive studies. *Experimental Psychology*, *14*(4), 164–181. (In Russian). https://doi.org/10.17759/exppsy.2021140409

12. Nistor N., Artmann B., Isik D., Neziri N., Stanciu D. (2025). Knowledge types and cognitive processing modes associated with online news credibility assessment: An interview study. *Computers in Human Behavior*, *171*. https://doi.org/10.1016/j.chb.2025.108721

13. Berman N. D. (2021). Digital skills and digital literacy of students. *Russian Journal of Education and Psychology*, *12*(2-2), 26–34. (In Russian). https://www.elibrary.ru/rzypeq

14. Kahneman D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

15. Halpern D. *Psychology of critical thinking.* St. Petersburg: Piter, 2000.

16. Mantulenko V. V. (2024). Information security of the individual in the context of Russian education at the present stage. *Continuous education: 21st century*, *2*(46), 124–138. (In Russian). https://www.elibrary.ru/rdquuz

17. Evplov N. A. (2025). Evaluation of deepfake detection methods in video content. *Modern science: current problems of theory and practice. Series: Natural and technical sciences*, 10-*2*, 17–22. (In Russian). https://www.elibrary.ru/mwqajb

18. Rybakov A. A., & Freilekhman S. A. (2022). Automatic detection of audio defects using parallel computing. *Software products and systems*, *3*, 428–437. (In Russian). URL: https://www.elibrary.ru/ywtpfn

19. Sharma V.K., Garg R., & Caudron Q. (2025). A systematic literature review on deepfake detection techniques. *Multimed Tools Appl*, *84*(20), 22187–22229. https://doi.org/10.1007/s11042-024-19906-1

Submitted: 17 April 2026 Accepted: 17 May 2026 Published: 18 May 2026