

Научная статья
УДК 159.9
DOI: 10.18384/3033-6414-2026-2-92-104

ФОРМИРОВАНИЕ УМЕНИЯ РАСПОЗНАВАТЬ ФОТОГРАФИИ ЛИЦ, СГЕНЕРИРОВАННЫХ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

Васильева И. В.^{1,2*}, Асафова А. Н.³, Грищенко Д. С.⁴

¹ Тюменский государственный университет, г. Тюмень, Российская Федерация

² Тюменский институт повышения квалификации сотрудников МВД России, г. Тюмень, Российская Федерация

³ Независимый исследователь, г. Тюмень, Российская Федерация

⁴ Независимый исследователь, г. Тюмень, Российская Федерация

* Корреспондирующий автор, e-mail: i.v.vasileva@utmn.ru

Поступила в редакцию 10.02.2026

После доработки 30.03.2026

Принята к публикации 31.03.2026

Аннотация

Цель. Разработка стратегии обучения, позволяющей повысить умение распознавания реальных и сгенерированных ИИ фотографий.

Процедура и методы. В исследовании приняли участие 201 респондент в возрасте от 14 до 66 лет, студенты и специалисты различных профилей ($m=45$, $f=156$, $M_{\text{возраст}}=23,66$, $SD=8,65$). Исследование выполнено в формате формирующего эксперимента с начальным и итоговым замерами. Экспериментальная группа участников исследования принимала участие в обучении, основанном на синтезе стратегий обратной связи и прямого указания. Для обработки количественных данных используются U-критерий Манна–Уитни и T-критерий Уилкоксона в программе Statsoft STATISTICA 10.0.

Результаты. На начальном замере экспериментальная и контрольная группы были эквивалентны. Были обнаружены статистически значимые различия между начальным и итоговым замерами в экспериментальной группе. Между замерами в контрольной группе также существуют статистически значимые различия, но в меньшей степени, чем у экспериментальной группы.

Теоретическая и/или практическая значимость заключается в том, что стратегия обучения умению распознавать фотографии, сгенерированные искусственным интеллектом, основанная на объединении стратегий обратной связи и прямого указания, действительно повышает эффективность распознавания генераций участниками. Результаты могут быть полезны для дальнейших исследований в области восприятия и взаимодействия с ИИ.

Ключевые слова: восприятие изображений, искусственный интеллект (ИИ), лицо, опыт с ИИ, распознавание изображений

Благодарности и источники финансирования. Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации в рамках проекта «Фундаментальные проблемы методики разработки и связанного с ней правового и этического регулирования в сфере применения систем и моделей искусственного интеллекта» (FEWZ-2024-0052).

Для цитирования: Васильева И. В., Асафова А. Н., Грищенко Д. С. Формирование умения распознавать фотографии лиц, сгенерированных искусственным интеллектом // Психологические науки. 2026. № 2. С. 92–104. <https://doi.org/10.18384/3033-6414-2026-2-92-104>

Original research article

FORMING THE ABILITY OF RECOGNITION OF FACES IN AI-GENERATED PHOTOS

I. V. Vasileva^{1,2*}, A. N. Asafova³, D. S. Grischenko⁴

¹ University of Tyumen, Tyumen, Russian Federation

² Tyumen Institute for Advanced Training of Employees of the Ministry of Internal Affairs of Russia, Tyumen, Russian Federation

³ An Independent researcher, Tyumen, Russian Federation

⁴ An Independent researcher, Tyumen, Russian Federation

* Corresponding author, e-mail: i.v.vasileva@utmn.ru

Received by the editorial office 10.02.2026

Revised by the author 30.03.2026

Accepted for publication 31.03.2026

Abstract

Aim. To develop a training strategy that enhances the ability to distinguish between real and AI-generated photos.

Methodology. The study involved 201 respondents aged from 14 to 66 years, students and specialists of various profiles (m-45, f-156, Mage=23.66, SD=8.65). The study was conducted as a formative experiment with initial and final measurements. The experimental group participated in training based on a synthesis of feedback strategies and direct instruction. Quantitative data were processed using the Mann-Whitney U-test and Wilcoxon test in Statsoft STATISTICA 10.0.

Results. At the initial measurement, the experimental and control groups were equivalent. Statistically significant differences were found between the initial and final measurements in the experimental group. There were also statistically significant differences in the control group, though to a lesser extent than in the experimental group.

Research implications. The training strategy for recognizing AI-generated photos, based on combining feedback and direct instruction strategies, indeed improves participants' detection accuracy. The results may be useful for further research in the field of perception and interaction with AI.

Keywords: AI experience, artificial intelligence (AI), face, images perception, images recognition

Acknowledgments: This study was supported by the Ministry of Science and Higher Education of the Russian Federation within the frame-work of a State assignment (FEWZ-2024-0052).

For citation: Vasilyeva, I. V., Asafova, A. N. & Grishchenko, D. S. (2026). Forming the Ability of Recognition of Faces in AI-Generated Photos. In: *Psychological Sciences*, 2, 92–104. <https://doi.org/10.18384/3033-6414-2026-2-92-104>

ВВЕДЕНИЕ

В настоящее время стремительно развиваются технологии искусственного интеллекта (ИИ) и они доступны для широкого слоя населения¹. Это приводит не только к полезным эффектам, но и к распространению практик, которые применяются для совершения криминальных действий, в частности, к использованию дипфейков в мошеннических целях. Дипфейк – это изображения, видео- или аудиоматериалы, созданные при помощи искусственного интеллекта, которые могут изображать реальных или нереальных людей. Они могут использоваться, например, для манипуляции общественным мнением, создания компрометирующих материалов с дальнейшим шантажом и вымогательством, кражи личных и корпоративных данных, подделки документов. Исследования психических, эмоциональных реакций людей на дипфейки разнообразны и дают противоречивые результаты: от неприятия по отношению к сверхреалистичным дипфейкам до неразличимости дипфейков и нейтральной реакции [1; 2; 3].

Важно, чтобы люди использовали собственные интеллектуальные возможности для обеспечения своей безопасности, а технологии искусственного интеллекта рассматривали именно как инструмент [4]. Избыточная уверенность в способности распознать сгенерированный контент не способствует успеху в решении этой задачи [5]. В рамках настоящего исследования разрабатывается стратегия обучения умению распознавать сгенерированные и реальные фотографий. Эта стратегия может быть использована как в организации, так и индивидуально, для развития умений, обеспечивающих информационную безопасность.

В актуальный период продолжают развиваться уже существующие сервисы,

предоставляющие услуги искусственного интеллекта, а также разрабатываются новые. Среди них подавляющее большинство занимается генерацией именно фотоматериалов, которые уже являются довольно реалистичными. При этом сервисы не имеют никаких ограничений или они незначительны. Многие площадки являются бесплатными или предоставляют несколько бесплатных запросов. Технологии искусственного интеллекта могут упрощать жизнь, как обычным людям, так и мошенникам, поскольку для совершения киберпреступления не требуется большое количество времени или ресурсов.

Созданные с помощью технологий искусственного интеллекта изображения лиц можно разделить на три типа [6]:

1. Искусственные лица. Эти изображения создаются как интеграция большого количества фотографий настоящих лиц.

2. Изображения лиц с некоторыми модификациями, например, изменение пола, нескольких черт. Фактически происходит редактирование настоящих фотографий с помощью технологий искусственного интеллекта.

3. Смешанные лица. Эти изображения получены путём смешения изображений двух настоящих фотографий.

Активное внедрение технологий искусственного интеллекта в повседневную жизнь требует критичного отношения к информации, которая используется для принятия решения или обращается к эмоциональной реакции человека. Изображения людей, аналогичные фотографии, которые генерируют технологии искусственного интеллекта, несут больший риск, чем все остальные виды генераций, поскольку ранее не было необходимости проверять визуальную информацию, она априорно вызывала доверие, тогда как сейчас необходимо критически относиться не только к текстовому, но и к визуальному контенту [7]. Уровень развития технологий генеративного искусственного интеллекта таков,

¹ Туровец Ю. В., Вишневецкий К. О. Искусственный интеллект в России: кто, что и как внедряет. URL: <https://issek.hse.ru/news/862013645.html> (дата обращения: 12.10.2025).

что качество изображений очень высоко. Не только люди испытывают трудности в распознавании настоящих фотографий, но даже специальные программы, основанные на технологиях искусственного интеллекта, также не справляются идеально с этой задачей [8].

Новизна данного исследования заключается в проверке эффективности синтеза уже разработанных предыдущими исследователями стратегий обучения на русскоязычных участниках. Результаты исследования вносят свой вклад в область взаимодействия человека с искусственным интеллектом. Это может иметь значение для дальнейших разработок технологий и обучающих программ, направленных на оптимизацию взаимодействия человека с машинами.

Теоретический анализ

Теоретический обзор существующих исследований позволил оценить восприятие лиц человеком при постановке задачи различить реальные и сгенерированные ИИ-лица. Распознавание лица – это стремительный процесс, который может быть осложнён разными задачами: распознавания эмоции, отличия настоящей фотографии от сгенерированной и т. д. [9; 10]. Человек, решая задачу понять, какие фотографии сгенерированы искусственным интеллектом, а какие изображают реальных людей, чаще всего обращает внимание на специфичные детали. Э. Миллер [11] с коллегами опрашивала своих испытуемых, по каким именно деталям человек определил, что фотография перед ним является генерацией искусственного интеллекта или реальной. 84,62% опрошенных делали свой выбор на основе отдельных черт лица. Самая информативная черта лица, согласно результатам исследования, – это глаза (50%). Участники сообщали о «пустом взгляде» генераций искусственного интеллекта. Часть респондентов (52,88%) обратила внимание на неестественно сглаженную кожу на фотографиях ИИ, отсутствие

складок, морщин. На излишнюю симметричность сгенерированных фотографий обратили внимание 39% участников. В 32% случаев опрошенные заявили об идеальности сгенерированных фотографий. Несмотря на сложность поставленной задачи и неуспешность большинства испытуемых в её решении, по результатам данного исследования можно отметить, что люди в целом обращают внимание на мелкие детали и стараются анализировать предоставленные фотографии, прежде чем совершить выбор. При развитии умения распознавания фотографий можно обращать произвольное внимание испытуемых на детали, которые могут помочь правильно отличить генерацию искусственного интеллекта от фотографии реальных людей. В. А. Барабанщиков как одно из оснований оценки мимических паттернов как неестественных выделяет нетипичность демонстрируемого выражения лица для повседневного общения, например, чрезмерная выраженность или несоответствие нормам проявления эмоций [12]. Соответственно, для повышения эффективности обучения, которое будет предложено испытуемым, стоит подбирать фотографии со спокойной, не ярко выраженной эмоциональной экспрессией. Также была оценена эффективность тех стратегий обучения, которые уже были разработаны и опробованы на респондентах. Было обнаружено несколько способов, которые позволяют научить участников эксперимента распознавать сгенерированные и реальные фотографии [13]. Можно воспользоваться стратегиями обратной связи, прямого совета, помощи со стороны ИИ, повышения осведомлённости и коллаборации с другими участниками эксперимента. Из всех перечисленных наиболее эффективной оказалась стратегия обучения, основанная на обратной связи [14; 15; 16; 17; 18]. Тем не менее, у респондентов, которые прошли обучение с обратной связью, увеличивается тревога и снижается оценка собственной эффективности. Также

через некоторое время после обучения эффективность распознавания фотографий снова падает. Повышение умения распознавать сгенерированные ИИ фотографии также наблюдается при использовании стратегии прямого совета/указания [19; 20]. Однако советы, которые были использованы в рамках данных исследований, являются устаревшими. Современные технологии искусственного интеллекта позволяют создавать изображения, практически неотличимые от настоящих, а количество «артефактов» значительно снизилось. Необходимо разработать советы, актуальные в реалиях текущего технологического прогресса.

Цель: разработка стратегии обучения, позволяющей повысить умение распознавать реальные и сгенерированные искусственным интеллектом фотографии лиц.

Гипотезы.

Основная: обучение умению распознавать сгенерированные и реальные фотографии лиц, основанному на объединении стратегий обратной связи и прямого указания, значительно повысит эффективность участников в умении идентифицировать фотографии лиц, сгенерированные искусственным интеллектом.

Дополнительная: результаты контрольной группы, которая не проходила обучение, не изменятся или изменятся незначительно за счёт насмотренности.

Организация исследования

В исследовании принял участие 201 человек в возрасте от 14 до 66 лет, студенты и специалисты различных областей. Из них 45 мужчин, 156 женщин. $M_{\text{возраст}} = 23,66$, $SD = 8,65$. Из них 114 человек составили контрольную группу, 87 были участниками экспериментальной группы. Контрольная группа состояла из 89 женщин и 25 мужчин, экспериментальная группа состояла из 67 женщин и 20 мужчин. Участники исследования были представителями разных профессиональных групп и направлений подготовки: психолого-педагогических, социо-

гуманитарных, инженерно-технических, естественно-научных. Контрольная и экспериментальная группы по этому признаку были уравновешены, состав участников был гетерогенным.

Исследование состоит из трёх этапов. На первом этапе проводится начальный замер, который позволит оценить начальный уровень эффективности распознавания фотографий лиц у участников. Тестирование состоит из восьми проб: первые четыре пробы состоят из трёх фотографий реальных лиц и одного сгенерированного лица. Следующие четыре пробы, наоборот, состоят из трёх фотографий сгенерированных лиц и одного реального лица. В первом случае участникам исследования предлагается выбрать лицо, которое, по их мнению, является сгенерированным искусственным интеллектом. Во втором случае, соответственно, необходимо выбрать лицо, которое является реальным. Пробы из четырёх фотографий берутся для того, чтобы снизить фактор случайности, который может повлиять на исследование с двумя фотографиями [21]. Такой формат замера эффективности уже проводился нами ранее при исследовании связи коммуникативных характеристик и эффективности распознавания лиц на фотографиях реальных людей и на изображениях людей, сгенерированных с помощью искусственного интеллекта [22]. Как показало это исследование, тестирование в таком формате является сложным для испытуемых, поскольку переменная «эффективность распознавания генераций» отклоняется от нормального распределения ($K-S > 0,20$; Lilliefors test $< 0,01$; $S-W = 0,02$). Средний показатель среди испытуемых: 9,5 из 32 возможных. Высокая сложность поставленной задачи поможет оценить изменение эффективности.

На втором этапе экспериментальная выборка проходила обучение. Обучение основано на синтезе двух стратегий, которые ранее использовались другими исследователями в этой области. Наилучшим

образом себя показали стратегия обратной связи и стратегия прямого указания/совета [13]. Стратегия обратной связи подразумевает такое же тестирование, как на первом этапе, но с сообщением участникам того, какой был правильный ответ. Стратегия прямого указания предполагает описание возможных деталей, которые помогут распознать лицо, сгенерированное ИИ. Советы будут составляться на основе опыта предыдущих исследователей и на основе просмотра сформированной базы фотографий для исследования.

Итоговый замер проводился в том же формате, что и начальный, но с использованием другого стимульного материала (других фотографий лиц). Подбор стимульного материала производился вручную. Фотографии реальных лиц были взяты с фотостока Unsplash, который позволяет свободно использовать опубликованные фотографии без водяных знаков. Фотографии отбирались согласно следующим критериям: лицо не должно выражать яркую эмоциональную экспрессию, изображение должно быть цветным, на фотографии не должны присутствовать резкие тени или оригинальные световые схемы. Таким образом, было отобрано 24 фотографии мужчин и 24 фотографии женщин.

Сгенерированные лица проходили проверку согласно таким же критериям. Также было необходимо, чтобы на сгенерированных ИИ фотографиях лиц не присутствовали слишком видимые артефакты. Артефакт на генерации искусственного интеллекта – это любое искажение реальности [23]. Фотографии генерировались при помощи сервиса This Person Does Not Exist, который работает на модели GAN. Вторым источником сгенерированных лиц выступала база данных, которая состояла из фотографий, сгенерированных моделями Stable Diffusion 1.5, 2.1; SDXL 1.0 checkpoint.

Для стимульного материала, который был использован в начальном и итоговом

замерах, был рассчитан индекс трудности задания. Для каждой отдельной фотографии из 64 индекс колеблется от 0,25 до 0,75, при этом для большей части фотографий этот показатель составляет от 0,41 до 0,75, что позволяет судить о гомогенности задания по уровню трудности. Общий индекс трудности задания для всех фотографий составляет 0,55.

Полученный стимульный материал был оформлен в виде чат-бота в социальной сети Telegram (https://t.me/generated_photosBot). Возможности данного приложения позволяют обеспечить случайное распределение участников между экспериментальной и контрольной выборками с вероятностью 1:2 и переносить ответы в онлайн-таблицу для дальнейших статистических подсчетов. Конфиденциальность участников исследования обеспечивается при помощи ключа безопасности. Чат-бот создавался при помощи онлайн-конструктора Salebot. После составления сценариев работы бота были проведены тестирования его функционала непосредственно в Telegram. Исследование проводилось весной 2025 г.

Собранные данные были проверены на соответствие закону нормального распределения. Затем были проведены расчёты, которые позволяют говорить о наличии статистически значимых различий между первым и повторным тестированием и между экспериментальной и контрольной группами. Были использованы методы непараметрической статистики: критерий U Манна–Уитни и T-критерий Уилкоксона, для полученных результатов был рассчитан размер эффекта методом Вендта как показателя рангово-бисериальной корреляции. Данные обрабатывались в программе StatSoft STATISTICA 10.

Результаты исследования

Результаты проверки данных на нормальное распределение по всем измененным параметрам показали незначительные отклонения. Для переменной

«эффективность распознавания фотографий на начальном замере» в контрольной группе ($n=114$) $K-S < 0,01$, $Lilliefors < 0,01$, что указывает на формальное отклонение от нормальности; $S-W=0,92$, – этот критерий не отвергает нормальность. Сходная ситуация образовалась для переменной «эффективность распознавания фотографий на итоговом замере» в экспериментальной группе ($K-S < 0,05$, $Lilliefors < 0,01$, $S-W=0,94$). В экспериментальной группе ($n=87$) переменная «эффективность распознавания фотографий на начальном замере» показывает отклонения от нормального распределения по всем критериям ($K-S < 0,01$, $Lilliefors < 0,01$, $S-W = 0,88$). Переменная «эффективность распознавания фотографий на итоговом замере» формально отклоняется от нормального распределения ($K-S < 0,01$, $Lilliefors < 0,01$), хотя $S-W=0,95$ не отклоняет гипотезу о нормальности распределения данных. Для согласованности с общим подходом были применены методы непараметрической статистики для

оценки различий между переменными и группами.

Для наглядности представим графический результат сравнения замеров эффективности распознавания фотографий на начальном и итоговом этапах исследования (рис. 1).

Рисунок 1 построен на основании средних данных для каждой выборки:

1. Сравнение показателей эффективности распознавания фотографий лиц участниками контрольной и экспериментальной выборок на начальном замере.

U-критерий Манна–Уитни выявил статистически значимые различия между контрольной ($n=114$) и экспериментальной ($n=87$) группами ($U=4847$ при $p>0,05$) по исходному уровню показателя эффективности. Это указывает на отсутствие значимых различий между выборками в начале эксперимента. Можно утверждать, что распределение участников по группам не внесло систематического смещения, последующие различия связаны с эффектом вмешательства.

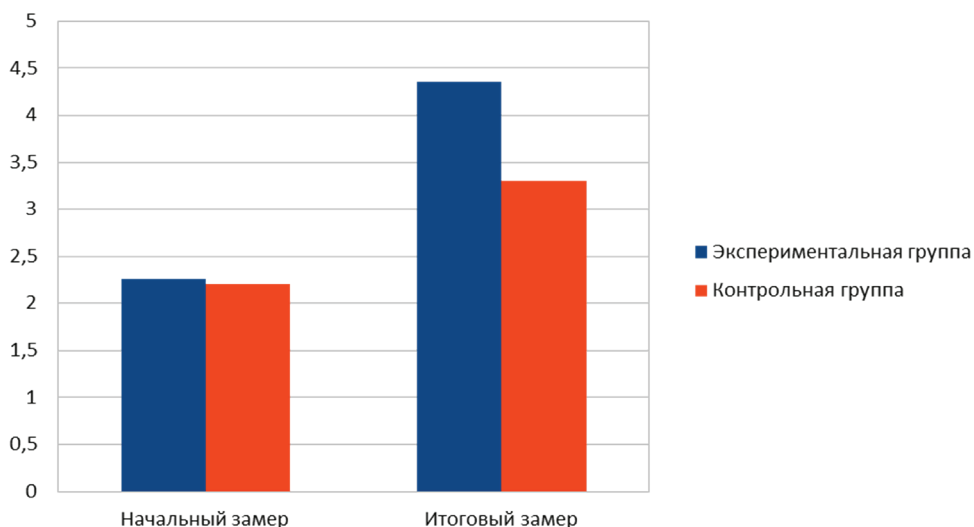


Рис. 1 / Fig. 1. Графическое представление сравнения показателей эффективности распознавания фотографий лиц участниками контрольной и экспериментальной выборок на начальном и итоговом замерах / Graphical representation of the comparison that shows the performance indicators of facial photo recognition by participants in the control and experimental samples at the initial and final measurements

Источник: данные авторов.

2. Сравнение показателей эффективности распознавания фотографий лиц участниками контрольной и экспериментальной выборок на итоговом замере.

U-критерий Манна–Уитни выявил статистически значимые различия между контрольной ($n=114$) и экспериментальной ($n=87$) группами ($U=3502$, $p < 0,001$) по уровню показателя эффективности при повторном тестировании. Значения в экспериментальной группе выше, чем в контрольной на 35,3%. Средние ранги показали, что значения переменной «эффективность распознавания фотографий» после повторного тестирования в экспериментальной группе ($\text{Mean Rank}=117,75$) были достоверно выше, чем в контрольной ($\text{Mean Rank}=88,22$), с небольшим размером эффекта ($r=0,25$).

3. Сравнение показателей эффективности распознавания фотографий лиц участниками экспериментальной группы на начальном и итоговом замерах.

В экспериментальной группе ($n=73$) при помощи парного T-критерия Уилкоксона были выявлены статистически значимые различия между переменными «эффективность распознавания фотографий на начальном замере» и «эффективность распознавания фотографий на итоговом замере» ($T=232$ при $p < 0,001$). Значения переменной «эффективность распознавания фотографий на итоговом замере» были достоверно выше, чем значения переменной «эффективность распознавания фотографий на начальном замере» с очень большим размером эффекта ($r=0,72$). В 83,9% случаев значения после вмешательства были выше, чем до него.

4. Сравнение показателей эффективности распознавания контрольной группы на начальном и итоговом замерах.

В контрольной группе ($n=91$) парный T-критерий Уилкоксона показал значимые различия между переменными «эффективность распознавания фотографий на начальном замере» и «эффективность распознавания фотографий на итоговом

замере» ($T=896$ при $p < 0,001$), но с малым размером эффекта ($r=0,16$). В экспериментальной группе ($n=73$) различия были более выраженными ($r=0,72$), что свидетельствует о влиянии вмешательства. В 79,8% случаев значения после вмешательства были выше, чем до него.

Обсуждение результатов

Согласно подсчётам, проведённым для нашего исследования, на начальном замере экспериментальная группа справилась с предложенным заданием на 21,4%, к итоговому замеру это показатель увеличился до 41,1% (значение p для z-критерия $=0,0051$). Контрольная группа на начальном замере показала эффективность в 27,6%, а на итоговом – в 41,4% (значение p для z-критерия $=0,0284$). Значение p z-критерия для сравнения результатов начального замера у контрольной и экспериментальной группы равно 0,3060, для сравнения результатов итогового замера – 0,9659.

Данные результаты подтверждают предположение о том, что обучение, основанное на синтезе стратегий обратной связи и прямого указания, способно повысить эффективность участников, поскольку между итоговыми замерами экспериментальной и контрольной групп существует различие в пользу экспериментальной группы. Говорить о значительном повышении эффективности было бы некорректно, поскольку размер эффекта небольшой, однако, разница между группами испытуемых всё же есть. Если сравнивать начальный и итоговый замеры только у экспериментальной группы, также можно говорить о подтверждении предположения о повышении эффективности, поскольку размер эффекта является большим. Предположение о том, что результаты контрольной группы не изменятся или изменятся незначительно, не подтвердилось, поскольку в 79,8% случаев результаты контрольной группы выше при итоговом замере. Тем не менее, размер эффекта для контрольной

ной группы значительно меньше, чем для экспериментальной группы.

Такие результаты могут быть объяснены следующими причинами. Во-первых, согласно результатам нашего предыдущего исследования [22], группа, имеющая опыт взаимодействия с искусственным интеллектом, справляется с задачей различения сгенерированных ИИ фотографий лиц лучше, чем группа без такого опыта. В рамках данного исследования опыт взаимодействия с искусственным интеллектом был повышен искусственным образом как при прохождении первого тестирования, так и при прохождении обучения. Можно предположить влияние роли «насмотренности» на полученные результаты. Этим же явлением можно объяснить повышение эффективности в контрольной группе, поскольку в процессе естественного обучения они также приобрели опыт взаимодействия с искусственным интеллектом и «насмотренность», однако в меньшем объёме, чем экспериментальная группа.

Во-вторых, роль может сыграть эффект повторного тестирования. Испытуемые из контрольной группы могли уловить отдельные паттерны или стимулы и использовать их при контрольном замере неосознанно. Исследования, посвящённые имплицитному научению, утверждают, что при повторном тестировании результаты испытуемых улучшаются и использование тестирований при обучении намного эффективнее, чем их отсутствие [23]. Контроль над прохождением исследования не осуществлялся, респонденты могли самостоятельно анализировать вопросы тестирования по своему желанию.

Рост точности в экспериментальной группе можно объяснить структурированной обратной связью. Указание на правильные ответы помогло обратить внимание респондентов на ключевые артефакты ИИ-генераций. Конкретные советы снижали когнитивную нагрузку, поскольку испытуемые заранее были подготовлены к распознаванию перечис-

ленных признаков. Также на результаты исследования могла повлиять асимметрия проб. В первых четырёх случаях респонденты учатся искать аномалии при сравнении фотографий с ИИ и реальными людьми, последние четыре пробы учили отделять «дипфейки».

На практике результаты данного исследования могут быть применены для составления программ тренинговых занятий и упражнений, которые нацелены на противостояние мошенничеству с применением технологий искусственного интеллекта [24]. Психологи могут проводить тренинги для различных категорий людей в общественно значимых сферах. Важно обучать журналистов, сотрудников силовых структур, юристов распознаванию «дипфейков» для противодействия преступным действиям, опознанию «фейковой» информации. Возможно проводить обучение распознаванию фальшивых резюме (в частности, генерации вместо фотографии кандидата) для менеджеров по подбору персонала и рекрутеров, особенно в ситуации удалённого найма. Также можно проводить обучение в любых других корпоративных системах для предотвращения мошенничества с целью получения корпоративных данных. Такие тренинги могут быть полезны для общего повышения уровня информационной безопасности среди населения в формате онлайн-курса, учебной дисциплины в школе, элективного курса в университете.

ЗАКЛЮЧЕНИЕ

Перед проведением эмпирического исследования были поставлены следующие гипотезы:

1. Обучение умению распознавать сгенерированные и реальные фотографии, основанное на объединении стратегий обратной связи и прямого указания, значительно повысит эффективность участников.

2. Результаты контрольной группы, которая не проходила обучение, не из-

меняться или изменяться незначительно за счёт эффекта «насмотренности».

По результатам исследования можно говорить об обнаружении доказательств в пользу первой гипотезы. Вторая гипотеза была отклонена.

Таким образом, можно утверждать о подтверждении эффективности разработанного обучения, которое в дальнейшем можно применять в различных сферах жизнедеятельности. Дальнейшие перспективы исследования могут включать в себя проверку переноса сформированного умения на новые поколения генеративных моделей; оптимизацию длительности и формата обучения под разные цели и сроки; а также учёт особенностей такого обучения для представителей разных возрастных групп [25].

Ограничения. В качестве возможного ограничения может выступать уровень технической осведомлённости пользова-

теля. Редко встречающийся формат исследования с элементами геймификации мог помешать людям с низким уровнем компетентности в работе с мессенджерами и чат-ботами быть эффективными в распознавании изображений. Помимо этого, существуют технические ограничения со стороны самого бота: при большом потоке людей он замедлял свою работу. Также среди неучтённых переменных можно перечислить низкую скорость интернета, использование различных устройств при прохождении исследования (мобильное устройство, персональный компьютер), разная цветопередача на разных устройствах, отсутствие контроля над условиями прохождения. Стоит отметить, что полученные результаты актуальны именно на данный момент с учётом текущих возможностей нейросетей. Неизвестно, останется ли обучение таким же эффективным для новых типов генераций.

ЛИТЕРАТУРА

1. Katsyri J., de Gelder B., Takala T. Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images // *Perception*. 2019. № 48 (10). P. 968–991. DOI: 10.1177/0301006619869134.
2. Natale D. A., Simonetti M. E., La Rocca S. Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research // *Computers in Human Behavior Reports*. 2023. № 10 (3). URL: <https://www.sciencedirect.com> (дата обращения: 10.10.2025). DOI: 10.1016/j.chbr.2023.100288.
3. Miller E. J., Foo Y. Z., Mewton P. et al. How do people respond to computer-generated versus human faces? A systematic review and meta-analyses // *Computers in Human Behavior Reports*. 2023. № 10. URL: <https://www.sciencedirect.com> (дата обращения: 10.10.2025). DOI: 10.1016/j.chbr.2023.100283.
4. Korteling J. E., van de Boer-Visschedijk G. C. et al. Human- versus Artificial Intelligence // *Frontiers in Artificial Intelligence*. 2021. № 4. URL: <https://www.frontiersin.org> (дата обращения: 10.10.2025). DOI: 10.3389/frai.2021.622364.
5. Köbis N. C., Doležalová B., Soraperra I. Fooled twice: People cannot detect deepfakes but think they can // *iScience*. 2021. № 24 (11). URL: <https://www.cell.com/iscience> (дата обращения: 10.10.2025).
6. Rathgeb C., Tolosana R., Vera-Rodriguez R. et al. Handbook of Digital Face Manipulation and Detection. From Deepfakes to Morphing attacks. Boston: Springer, 2022. 481 p. DOI: 10.1007/978-3-030-87664-7.
7. Connolly B. Digital Trust: Social Media Strategies to Increase Trust and Engage Customers. London: Bloomsbury Business, 2020. 256 p.
8. Fletcher J. Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance // *Theatre Journal*. 2018. Vol. 70. № 4. P. 455–471. DOI: 10.1353/tj.2018.0097.
9. Willis J., Todorov A. First Impression: Making Up Your Mind After a 100-Ms Exposure to a Face // *Psychological Science*. 2006. № 17 (7). P. 592–598. DOI: 10.1111/j.1467-9280.2006.01750.x.
10. Nemrodov D., Niemeier M., Mok J. N. Y. et al. The time course of individual face recognition: A pattern analysis of ERP signals // *NeuroImage*. 2016. № 132. P. 469–476. DOI: 10.1016/j.neuroimage.2016.03.006.

11. Miller E. J., Steward B. A., Witkower Z. et al. AI hyperrealism: Why AI faces are perceived as more real than human ones // *Psychological Science*. 2023. № 34 (12). P. 1390–1403. DOI: 10.1177/09567976231207095.
12. Барабанщиков В. А. Восприятие выражений лица. М.: Институт психологии РАН, 2009. 448 с.
13. Diel A., Lalgı T., Schroeter I. C. et al. As good as chance: A systematic review and meta-analysis of human deepfake detection performance based on 56 papers // *Expert Systems with Applications*. 2024. № 252. URL: <https://www.sciencedirect.com> (дата обращения: 10.10.2025). DOI: 10.31219/osf.io/cxv4r.
14. Diel A., Teufel M., Bäuerle A. Inability to detect deepfakes: Deepfake detection training improves detection accuracy, but increases emotional distress and reduces selfefficacy // *OSF Preprints*. 2024. URL: <https://osf.io/preprints/osf> (дата обращения: 10.10.2025). DOI: 10.2139/ssrn.5002585.
15. Mader B., Banks M. S., Farid H. Identifying computer-generated portraits: The importance of training and incentives // *Perception*. 2017. № 46 (9). P. 1062–1076. DOI: 10.1177/0301006617713633.
16. Hulzebosch N., Ibrahim S., Worring M. Detecting CNN-generated facial images in real-world scenarios // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020. URL: <https://arxiv.org/abs/2005.05632> (дата обращения: 10.10.2025). DOI: 10.1109/CVPRW50498.2020.00329.
17. Nightingale S. J., Farid H. AI-synthesized faces are indistinguishable from real faces and more trustworthy // *Proceedings of the National Academy of Sciences*. 2022. № 119 (8). URL: <https://www.pnas.org>. (дата обращения: 10.10.2025). DOI: 10.1073/pnas.2120481119.
18. Mohamed N. B., Bogdanel G., Moreno H. G. Is Training Useful to Detect Deepfakes? A Preliminary Study // *Royal Society Open Science*. 2025. № 12. URL: <https://royalsocietypublishing.org>. (дата обращения: 10.10.2025). DOI: 10.23919/CISTI58278.2023.10211622.
19. Bray S. D., Johnson S. D., Kleinberg B. Testing human ability to detect ‘deepfake’ images of human faces // *Journal of Cybersecurity*. 2023. № 9 (1). URL: <https://academic.oup.com/cybersecurity> (дата обращения: 10.10.2025). DOI: 10.1093/cybsec/tyad011.
20. Somoray K., Miller D. J. Providing detection strategies to improve human detection of deepfakes: An experimental study // *Computers in Human Behavior*. 2023. № 149 (4). URL: <https://www.sciencedirect.com> (дата обращения: 10.10.2025). DOI: 10.1016/j.chb.2023.107917.
21. Yegemberdiyeva G., Amirgaliyev B. Study of AI Generated and Real Face Perception // *Proceedings of the 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*. 2021. URL: <https://ieeexplore.ieee.org> (дата обращения: 10.10.2025). DOI: 10.1109/SIST50301.2021.9465908.
22. Васильева И. В., Асафова А. Н., Грищенко Д. С. Коммуникативные характеристики личности и оценка фотографий, сгенерированных искусственным интеллектом // *Теоретическая и экспериментальная психология*. 2025. Т. 18 (1). С. 27–44. DOI: 10.11621/ТЕР-25-02.
23. Devi S. P., Mishra A. K. The testing effect: Looking through implicit theories’ perspectives // *Psychological Studies*. 2023. № 68 (1). P. 92–100. DOI: 10.1007/s12646-022-00710-6
24. Ng Y.–L. An error management approach to perceived fakeness of deepfakes: The moderating role of perceived deepfake targeted politicians’ personality characteristics // *Current Psychology*. 2023. № 42 (1). P. 25658–25669. DOI: 10.1007/s12144-022-03621-x.
25. Schetinger V., Oliveira M. M., da Silva R. et al. Humans Are Easily Fooled by Digital Images // *Computers & Graphics*. 2015. № 68. URL: <https://arxiv.org/abs/1509.05301> (дата обращения: 10.10.2025). DOI: 10.48550/arXiv.1509.05301.

REFERENCES

1. Katsyri, J., de Gelder, B. & Takala, T. (2019). Virtual Faces Evoke Only a Weak Uncanny Valley Effect: An Empirical Investigation With Controlled Virtual Face Images. In: *Perception*, 48 (10), 968–991. DOI: 10.1177/0301006619869134.
2. Natale, D. A., Simonetti, M. E. & La Rocca, S. (2023). Uncanny Valley Effect: A Qualitative Synthesis of Empirical Research to Assess the Suitability of Using Virtual Faces in Psychological Research. In: *Computers in Human Behavior Reports*, 10 (3). URL: <https://www.sciencedirect.com> (accessed: 10.10.2025). DOI: 10.1016/j.chbr.2023.100288.
3. Miller, E. J., Foo, Y. Z. & Mewton, P. et al. (2023). How Do People Respond to Computer-generated Versus Human Faces? A Systematic Review and Meta-analyses. In: *Computers in Human Behavior*

- Reports*, 10. URL: <https://www.sciencedirect.com> (accessed: 10.10.2025). DOI: 10.1016/j.chbr.2023.100283.
4. Korteling, J. E., van de Boer-Visschedijk, G. C. et al. (2021). Human- Versus Artificial Intelligence. In: *Frontiers in Artificial Intelligence*, 4. URL: <https://www.frontiersin.org> (accessed: 10.10.2025). DOI: 10.3389/frai.2021.622364.
 5. Köbis, N. C., Doležalová, B. & Soraperra, I. (2021). Fooled Twice: People Cannot Detect Deepfakes but Think They Can. In: *iScience*, 24 (11). URL: <https://www.cell.com/iscience> (accessed: 10.10.2025).
 6. Rathgeb, C., Tolosana, R. & Vera-Rodriguez, R. et al. (2022). *Handbook of Digital Face Manipulation and Detection. From Deepfakes to Morphing attacks*. Boston: Springer publ. DOI: 10.1007/978-3-030-87664-7.
 7. Connolly, B. (2020). *Digital Trust: Social Media Strategies to Increase Trust and Engage Customers*. London: Bloomsbury Business publ.
 8. Fletcher, J. (2018). Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance. In: *Theater Journal*, 70, 4, 455–471. DOI: 10.1353/tj.2018.0097.
 9. Willis, J. & Todorov, A. (2006). First Impression: Making Up Your Mind After a 100-Ms Exposure to a Face. In: *Psychological Science*, 17 (7), 592–598. DOI: 10.1111/j.1467-9280.2006.01750.x.
 10. Nemrodov, D., Niemeier, M. & Mok, J. N. Y. et al. (2016). The Time Course of Individual Face Recognition: A Pattern Analysis of ERP Signals. In: *NeuroImage*, 132, 469–476. DOI: 10.1016/j.neuroimage.2016.03.006.
 11. Miller, E. J., Steward, B. A. & Witkower, Z. et al. (2023). AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones. In: *Psychological Science*, 34 (12), 1390–1403. DOI: 10.1177/09567976231207095.
 12. Barabanshchikov, V. A. (2009). *Perception of Facial Expressions*. Moscow: Institute of Psychology of the RAS publ. (in Russ.).
 13. Diel, A., Lalgı, T. & Schroeter, I. C. et al. (2024). As Good as Chance: A Systematic Review and Meta-analysis of Human Deepfake Detection Performance Based on 56 Papers. In: *Expert Systems with Applications*, 252. URL: <https://www.sciencedirect.com> (accessed: 10.10.2025). DOI: 10.31219/osf.io/cxv4r.
 14. Diel, A., Teufel, M. & Bäuerle, A. (2024). Inability to Detect Deepfakes: Deepfake Detection Training Improves Detection Accuracy, but Increases Emotional Distress and Reduces Selfefficacy. In: *OSF Preprints*. URL: <https://osf.io/preprints/osf> (accessed: 10.10.2025). DOI: 10.2139/ssrn.5002585.
 15. Mader, B., Banks, M. S. & Farid, H. (2017). Identifying Computer-generated Portraits: The Importance of Training and Incentives. In: *Perception*, 46 (9), 1062–1076. DOI: 10.1177/0301006617713633.
 16. Hulzebosch, N., Ibrahimı, S. & Worrıng, M. (2020). Detecting CNN-generated Facial Images in Real-world Scenarios. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. URL: <https://arxiv.org/abs/2005.05632> (accessed: 10.10.2025). DOI: 10.1109/CVPRW50498.2020.00329.
 17. Nightingale, S. J. & Farid, H. (2022). Ai-synthesized Faces Are Indistinguishable From Real Faces and More Trustworthy. In: *Proceedings of the National Academy of Sciences*, 119 (8). URL: <https://www.pnas.org> (accessed: 10.10.2025). DOI: 10.1073/pnas.2120481119.
 18. Mohamed, N. B., Bogdanel, G. & Moreno, H. G. (2025). Is Training Useful to Detect Deepfakes? A Preliminary Study. In: *Royal Society Open Science*, 12. URL: <https://royalsocietypublishing.org> (accessed: 10.10.2025). DOI: 10.23919/CISTI58278.2023.10211622.
 19. Bray, S. D., Johnson, S. D. & Kleinberg, B. (2023). Testing Human Ability to Detect ‘deepfake’ Images of Human Faces. In: *Journal of Cybersecurity*, 9 (1). URL: <https://academic.oup.com/cybersecurity> (accessed: 10.10.2025). DOI: 10.1093/cybsec/tyad011.
 20. Somoray, K. & Miller, D. J. (2023). Providing Detection Strategies to Improve Human Detection of Deepfakes: An Experimental Study. In: *Computers in Human Behavior*, 149 (4). URL: <https://www.sciencedirect.com> (accessed: 10.10.2025). DOI: 10.1016/j.chb.2023.107917.
 21. Yegemberdiyeva, G. & Amırgalıyev, B. (2021). Study Of AI Generated and Real Face Perception. In: *Proceedings of the 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*. URL: <https://ieeexplore.ieee.org> (accessed: 10.10.2025). DOI: 10.1109/SIST50301.2021.9465908.
 22. Vasilyeva, I. V., Asafova, A. N. & Grishchenko, D. S. (2025). Communicative Characteristics of Personality and the Evaluation of Photographs Generated by Artificial Intelligence. In: *Theoretical and Experimental Psychology*, 18 (1), 27–44. DOI: 10.11621/TEP-25-02 (in Russ.).

23. Devi, S. P. & Mishra, A. K. (2023). The Testing Effect: Looking Through Implicit Theories' Perspectives. In: *Psychological Studies*, 68 (1), 92–100. DOI: 10.1007/s12646-022-00710-6
24. Ng, Y.-L. (2023). An Error Management Approach to Perceived Fakeness of Deepfakes: The Moderating Role of Perceived Deepfake Targeted Politicians' Personality Characteristics. In: *Current Psychology*, 42 (1), 25658–25669. DOI: 10.1007/s12144-022-03621-x.
25. Schetinger, V., Oliveira, M. M. & da Silva, R. et al. (2015). Humans Are Easily Fooled by Digital Images. In: *Computers & Graphics*, 68. URL: <https://arxiv.org/abs/1509.05301> (accessed: 10.10.2025). DOI: 10.48550/arXiv.1509.05301.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Васильева Инна Витальевна (г. Тюмень) – доктор психологических наук, директор департамента, департамент психологии и дефектологии «Школа образования» Тюменского государственного университета; профессор кафедры философии, иностранных языков и гуманитарной подготовки сотрудников ОВД Тюменского института повышения квалификации сотрудников МВД России; ORCID: 0000-0003-0740-7260; e-mail: i.v.vasileva@utmn.ru

Асафова Анастасия Николаевна (г. Тюмень) – независимый исследователь; ORCID: 0009-0000-8296-4919; e-mail: asafova03@mail.ru

Грищенко Дарья Станиславовна (г. Тюмень) – независимый исследователь; ORCID: 0009-0008-7036-0155; e-mail: green4misty@gmail.com

INFORMATION ABOUT THE AUTHORS

Inna V. Vasileva (Tyumen) – Dr. Sci. (Psychology), Assoc. Prof., Head of the Department, Department of Psychology and Defectology, School of Education, University of Tyumen; Prof., Department of Philosophy, Foreign Languages, Humanitarian Training, Tyumen Institute for Advanced Training of Employees of the Ministry of Internal Affairs of Russia; ORCID: 0000-0003-0740-7260; e-mail: i.v.vasileva@utmn.ru

Anastasia N. Asafova (Tyumen) – Independent researcher; ORCID: 0009-0000-8296-4919; e-mail: asafova03@mail.ru

Darya S. Grischenko (Tyumen) – Independent researcher; ORCID: 0009-0008-7036-0155; e-mail: green4misty@gmail.com