

ЗАЩИТА ИЗОБРАЖЕНИЙ ОТ НЕСАНКЦИОНИРОВАННОГО РЕДАКТИРОВАНИЯ ИИ-МОДЕЛЯМИ

С. П. Никитенкова, А. С. Пряничникова

Национальный исследовательский Нижегородский государственный университет им. Н. И. Лобачевского, г. Нижний Новгород

Ключевые слова: искусственный интеллект, защита изображений, спектральный анализ.

Защита личной информации и конфиденциальных данных играет ключевую роль в современном мире, поэтому одним из наиболее актуальных вопросов в эпоху ИИ является защита личных фотографий от нежелательного редактирования.

ИИ-модели позволяют легко создавать большое количество персонализированных фотографий, используя ограниченное количество изображений, удостоверяющих личность [1]. Фотография может быть сделана в общественных местах или взята из социальных сетей. И если ранее для серьезной корректировки сцены или добавления объектов требовалось программное обеспечение для редактирования фотографий и соответствующие знания, но теперь с помощью ИИ-моделей text-to-image это можно сделать с помощью простой фразы (промта) на естественном языке. Порог создания новых изображений низок, что открывает двери для злоумышленников, которые могут легко редактировать изображения или создавать дипфейки. В связи с этим особую актуальность приобретает разработка простых и доступных методов защиты персональных фотографий.

Все методы защиты изображений от несанкционированного редактирования принято делить на активные и пассивные. Первые направлены на предотвращение изменений, вторые – на анализ изображений с целью сделать вывод о наличии или отсутствии постороннего вмешательства.

Самый известный из активных методов – это так называемая «иммунизация» изображений. Идея этого метода заключается во внесении в исходное изображение незначительных изменений пикселей, невидимых человеческому глазу, но обнаруживаемые ИИ-моделями. Возмущения критически нарушают работу модели, вынуждая ее генерировать нереалистичные или очень искаженные результаты. Примером такого подхода является программа PhotoGuard, разработанная в Массачусетском технологическом институте [2]. Ключевым недостатком данного метода является его нестабильность – любое последующее изменение, например, кадрирование или поворот, ведет к снижению его эффективности.

Еще одним методом активной защиты является использование

цифровых водяных знаков. В этом случае в изображение внедряются метки, которые должны сохраняться даже после редактирования. Устойчивость водяных знаков к модификациям критически важна для обеспечения подлинности фотографии. Однако результаты исследований показали, что водяные знаки неустойчивы к различным шумовым воздействиям и геометрической трансформации [3].

В предлагаемой работе для защиты изображения предлагается использовать спектральный подход и трактовать изображение как 2D-сигнал. Анализ частотной области выявил, что ИИ-инструменты генерации изображений сильнее всего искажают средние и высокие частоты. На основе этого был разработан нелинейный метод искажения частот исходного изображения. Предложенный алгоритм имеет простую реализацию и устойчив к jpeg-сжатию, а также к «print/scan» атаке. Метод был протестирован на модели Nano Banana.

Список использованных источников

1. Huang Y. et al. Diffusion model-based image editing: A survey/ IEEE Transactions on Pattern Analysis and Machine Intelligence. 2025.
2. Salman H. et al. Raising the cost of malicious ai-powered image editing /arXiv preprint arXiv:2302.06588. 2023.
3. Cao J. et al. Secure and Robust Watermarking for AI-generated Images: A Comprehensive Survey /arXiv preprint arXiv:2510.02384. 2025.

Никитенкова Светлана Павловна, к.т.н., доцент, доцент каф. БИС, nikitenkova@rf.unn.ru.
Пряничникова Алина Сергеевна, студ., pryanichnikova04@mail.ru.

УДК 004.056

К ВОПРОСУ ОБ ОБЕСПЕЧЕНИИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ АВТОМАТИЗИРОВАННЫХ СИСТЕМ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ

П.А. Летюк, Р.М. Яппаров

Уфимский университет науки и технологий, г. Уфа

Ключевые слова: информационная безопасность, критическая информационная инфраструктура, АСУ ТП, модель угроз, защита информации.

Современный этап развития информационных технологий характеризуется широким внедрением автоматизированных систем управления технологическими процессами в различных отраслях экономики. Такие системы применяются в энергетике, промышленности, коммунальном хозяйстве и других сферах, обеспечивающих жизнедеятельность общества. Их функционирование непосредственно