

ТЕХНОЛОГИИ МАРКИРОВКИ ИЗОБРАЖЕНИЙ, СГЕНЕРИРОВАННЫХ С ПОМОЩЬЮ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ

Бурное развитие генеративных моделей (DALL·E, Midjourney, Gemini, Imagen и др.) привело к резкому росту объема сгенерированных изображений, которые зачастую трудно отличить от созданных человеком. Это создает риски для дезинформации и манипулирования общественным мнением. Данная проблема стимулировала появление технологий для маркировки такого рода изображений.

В докладе рассматриваются технологии C2PA и SynthID для маркировки и идентификации контента, а именно изображений, сгенерированных с использованием генеративных систем. Описываются принципы стеганографического встраивания водяного знака (метки) в изображение на пиксельном уровне, а также способы удаления данного водяного знака с изображения.

Ключевые слова: C2PA, SynthID, стеганография, невидимый цифровой водяной знак, метаданные, генеративные модели, диффузионные модели, маркировка изображений, удаление водяного знака.