**UDC 004.8:004.056:342.7**

***Diana Ovcharenko***

*Assistant of the Department of Information Technologies Security,*
*Kharkiv National University of Radio Electronics*

# THE PHENOMENON OF ACCIDENTAL VISUAL CONVERGENCE IN AI-GENERATED RECOGNIZABLE REAL-PERSON LIKENESSES: RISKS AND PROTECTIVE MEASURES

**Abstract.** This paper explores the issue of the accidental reproduction of the facial features of real individuals during the generation of portraits by artificial intelligence algorithms. The mathematical, combinatorial, and psychological foundations of this phenomenon are examined. Furthermore, the legal, reputational, and technical challenges arising from this phenomenon are outlined and a conceptual multi-level risk assessment model is proposed.

**Keywords:** generative artificial intelligence, synthetic images, accidental visual convergence, recognizable appearance, personal data, privacy, memorization of training data, cyber risks, biometric similarity, commercial use of images.

The rapid development and widespread accessibility of generative artificial intelligence models in recent years, particularly Generative Adversarial Networks (GANs) and diffusion models (such as DALL-E and Midjourney) have expanded the capabilities for creating realistic synthetic portraits. At the same time, the proliferation of these technologies has given rise to new legal, ethical, and information security risks.

When creating ostensibly entirely synthetic portraits based on user text prompts, situations increasingly arise where an AI-generated face turns out to be nearly identical or highly similar to the likeness of a real person who never consented to the use of their personal data. This undesirable phenomenon of accidental visual convergence creates a significant risk of false accusations regarding the unauthorized use of private photographs, infringement of image rights, and the use of such images for commercial purposes.

In this regard, there is an urgent scientific and practical need to conceptualize this issue - specifically by providing a mathematical and

psychological rationale - and to identify the reasons why it remains insufficiently researched.

One of the insufficiently explored aspects of the use of generative artificial intelligence is the unintentional creation of images that bear a significant visual resemblance to the appearance of real people who did not consent to the use of their personal likenesses.

Unlike the intentional creation of deepfakes—where reproducing a specific individual's appearance is the user's deliberate goal—in this scenario, such a resemblance can emerge during the execution of a completely neutral text prompt that contains no names, photographs, biometric data, or other identifiers of a specific person.

Therefore, an accidental resemblance between a synthetic portrait and a real person is not necessarily the result of copying files from the training dataset, as generative models primarily encode learned patterns in their parameters rather than directly storing original reference images. Nevertheless, the memorization and reproduction of specific training examples cannot be ruled out.

This issue stems from several fundamental factors:

1. Statistical modeling of anthropometric characteristics. Generative models can "memorize" specific patterns in the distribution of human facial features—such as proportions, the shapes of the eyes, nose, and lips, facial contours, and so on. If a prompt lacks sufficient specificity, the model may combine features that, to some extent, resemble the appearance of a real person. However, it cannot be said that all generated faces necessarily gravitate toward a single, averaged-out image.

2. Combinatorial and probabilistic patterns. The greater the number of generated images, the higher the probability that visually similar combinations of features will appear. However, the likelihood of accidental visual convergence depends on the distribution of facial characteristics, the specific criterion for similarity, and the number of images being compared. Without an appropriate

mathematical model and empirical data, an exact match with a real person cannot be considered mathematically inevitable.

3. Memorization of training data. [1, 2] A generative model may also reproduce specific characteristics or fragments of images that were present in the training dataset. Unlike a random coincidence, this can be attributed to the model overfitting to specific training examples. Therefore, it is important to distinguish between statistical synthesis and the potential reproduction of training data in order to determine the nature of the similarity.

4. Psychological mechanisms [3] of perception. Due to evolutionary development, the human cognitive system is optimized for recognizing faces and identifying connections. Even when subtle differences exist between a generated image and a real person, the human cognitive system may emphasize perceived similarities while overlooking certain discrepancies.

It can be said that the unintentional generation of a recognizable appearance creates two primary types of risk:

The first aspect concerns the protection of the individual whose appearance has been reproduced. The use of a generated portrait can lead to privacy violations, reputational damage, misidentification, and the use of the image in a misleading context—for instance, on social media, in advertising, in the media, or within computer games. Under certain circumstances, the personal and biometric data of real individuals may also be put at risk.

The second aspect concerns the protection of the good-faith user. A creator of AI-generated content—who had no intention of replicating a specific person's appearance—may face claims regarding the unauthorized use of another individual's likeness, encounter content takedowns, and suffer both financial and reputational losses. This issue becomes particularly significant when such images are used for commercial purposes.

Another challenge lies in the uncertainty regarding the allocation of responsibility among the developer of a generative model, the provider of an AI

service, and the end user. The applicability of the GDPR, the EU AI Act, and other legal regulations depends on the specific circumstances of data processing, the nature of the image, the manner of its use, and the relevant jurisdiction. Visual similarity alone does not constitute automatic proof of a legal violation, which complicates the process of establishing whether an individual's appearance has been used unlawfully. Nevertheless, these risks can be minimized by establishing a comprehensive protection system that encompasses technical, organizational, and legal dimensions.

1. Technical Level. It is necessary to explore the implementation of mechanisms for verifying generated outputs ("Output Guardrails"), specifically algorithms for assessing biometric similarity ("Biometric Similarity Filtering"). Such mechanisms could proactively identify potentially risky images and flag them for additional review before they are delivered to the client. Implementing these checks requires accounting for recognition error rates, the limitations of reference datasets, and the additional processing of biometric data.

Other potential areas of focus include the controlled sampling of generated facial features, the use of "negative prompts," and the regulation of modifications made to generated portraits. The effectiveness of these proposed methods requires experimental validation.

2. Organizational Level. Generation parameters, text prompts, model identifiers, configuration settings, and the random seed values should be documented. Such records can assist in investigating the circumstances surrounding an incident, although they do not, in themselves, prove a violation or establish a lack of involvement.

It is advisable to adopt the ISO/IEC 42001 standard for AI management systems, as well as technologies that verify the provenance of digital content, such as C2PA and SynthID. C2PA ensures the preservation of data regarding the origin and processing history of content, while SynthID handles digital watermarking. These technologies possess distinct characteristics and limitations;

they cannot, on their own, prove that a generated image bears no resemblance to a real person. Furthermore, before publishing generated portraits for commercial purposes, it is advisable to check for visual matches using reverse image search—and, if necessary, specialized facial comparison tools.

3. Legal level. It is necessary to further examine the criteria for allocating responsibility among the participants in the content generation process. Particular attention should be paid to defining the conditions for protecting good-faith users who did not intentionally reproduce the appearance of a specific individual. A promising avenue involves developing legal mechanisms that take into account the degree of control each participant exerts over image generation, the foreseeability of the risk, and the appropriateness of the precautionary measures taken.

The unintentional creation by artificial intelligence of an appearance recognizable as that of a real person is a pressing issue situated at the intersection of several fields: generative technologies, information security, privacy, and legal regulation. Accidental visual convergence may arise from the statistical and combinatorial characteristics of the generation process, whereas the memorization of training data constitutes a distinct mechanism that warrants further analysis.

Protecting both data subjects and good-faith users requires a comprehensive, multi-layered approach.

Future research should focus on the quantitative assessment of the probability of accidental visual resemblance, the experimental comparison of generative models, and the evaluation of the effectiveness of technical measures designed to prevent problematic coincidences during the commercial use of AI-generated content.

As a promising approach to addressing this problem, a conceptual multi-level model for assessing the risk of unintentionally generating a recognizable human appearance (MRA-UFL) is proposed. The model entails a comprehensive

analysis of four components: the origin of the visual resemblance (Provenance), the degree of biometric similarity (Similarity), the context in which the generated image is used (Context), and the availability of verifiable evidence regarding the generation process (Evidence). The model proposes separating the aggregate assessment of potential harm from a separate measure of uncertainty. Future research will focus on the mathematical implementation of the model, the determination of weighting coefficients and threshold values, and the experimental validation of its effectiveness.

## References


1. Carlini N., Hayes J., Nasr M. et al. Extracting Training Data from Diffusion Models. *32nd USENIX Security Symposium (USENIX Security 23).* 2023. P. 5253–5270. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/carlini

2. Somepalli G., Singla V., Goldblum M., Geiping J., Goldstein T. Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2023. P. 6048–6058. DOI: https://doi.org/10.1109/CVPR52729.2023.00586

3. White D., Burton A. M. Individual differences and the multidimensional nature of face perception. Nature Reviews Psychology. 2022. Vol. 1. P. 287–300. DOI: https://doi.org/10.1038/s44159-022-00041-3