УДК 004.8

ШТУЧНИЙ ІНТЕЛЕКТ І МОРАЛЬ: ЧИ МОЖЛИВО ВИХОВАТИ «ЕТИЧНИЙ» AI?

Карабань Р. Є.
karabas736901.11@gmail.com
Черкаський державний фаховий бізнес-коледж
Люта М. В.
м. Черкаси, Україна

Технології швидко розвиваються, і штучний інтелект (AI) став важливою частиною життя, знаходячи застосування у медицині, транспорті, фінансах і повсякденних цифрових сервісах [1]. Однак разом із поширенням AI виникає питання його етичності, оскільки такі системи дедалі частіше приймають рішення, що впливають на життя людей. У зв'язку з цим проблема відповідальності, справедливості та безпечного використання штучного інтелекту набуває особливої актуальності.

У сучасному світі сформувалася окрема галузь досліджень – етика штучного інтелекту (AI Ethics), яка визначає принципи безпечного, прозорого та відповідального використання технологій [2]. Над її розвитком працюють фахівці різних галузей, що підкреслює міждисциплінарний характер проблеми.

Однією з ключових проблем є упередженість алгоритмів, що виникає через використання даних, які можуть містити помилки або соціальні стереотипи. Це може призводити до дискримінаційних рішень, зокрема у сфері працевлаштування або аналізу соціальних даних [3]. Подібні ризики спостерігаються і у фінансовій сфері, де системи кредитного скорингу можуть створювати нерівні умови доступу до ресурсів.

Важливим напрямом розвитку є використання explainable AI, що дозволяє пояснювати рішення систем і підвищує рівень довіри до них [5]. У критичних сферах, таких як транспорт і медицина, застосовується принцип контролю людини та мінімізації потенційної шкоди [4].

У фінансовій сфері AI використовується для кредитного скорингу. Проблема полягає в тому, що модель може враховувати неочевидні фактори,

наприклад район проживання або тип витрат, що призводить до нерівного доступу до кредитів. Тому впроваджуються explainable AI-системи, які пояснюють причини рішень (рівень доходу, кредитна історія, боргове навантаження).

Ще один важливий аспект – безпека використання. У безпілотних автомобілях системи постійно тестуються в різних умовах: дощ, туман, нічний час. У критичних ситуаціях алгоритми діють за заздалегідь визначеними принципами – мінімізація шкоди, дотримання правил і захист життя.

Складною проблемою є також питання відповідальності. Якщо система штучного інтелекту допускає помилку, складно визначити, хто саме винен: розробник, компанія чи користувач. Це особливо актуально у випадках медичних систем.

У медицині важливу роль відіграє якість даних. Якщо модель навчена переважно на даних однієї групи пацієнтів, вона може гірше працювати для інших. Тому системи проходять клінічну перевірку на різних вибірках. При цьому остаточне рішення завжди залишається за лікарем: AI лише допомагає, а не замінює спеціаліста.

У соціальних мережах рекомендаційні алгоритми можуть підсилювати дезінформацію або радикальний контент, оскільки орієнтуються на залученість користувачів. Для вирішення цього застосовують обмеження: зниження видимості недостовірного контенту, його маркування та модерацію.

У сфері безпеки системи розпізнавання облич можуть мати різну точність для різних груп людей, що призводить до помилкових ідентифікацій. Тому такі технології додатково тестуються, обмежуються або регулюються.

В освіті системи автоматичного оцінювання можуть бути упередженими до стилю відповідей або мови. Тому їх використовують лише як допоміжний інструмент, а не як остаточне джерело оцінки.

Для створення етичного штучного інтелекту застосовують кілька основних підходів:

- Програмування етичних правил – у системи закладають обмеження та правила поведінки. Наприклад, чат-боти не відповідають на небезпечні або шкідливі запити та уникають мови ненависті.
- Використання якісних і перевірених даних – моделі навчаються на достовірній і різноманітній інформації, щоб зменшити упередженість.
- Контроль і перевірка людиною – системи тестуються та коригуються перед впровадженням.

Проблема відповідальності за рішення AI залишається відкритою, оскільки складно визначити, хто несе відповідальність за помилки системи – розробник, організація чи користувач [2]. У соціальних мережах алгоритми можуть сприяти поширенню дезінформації, що потребує впровадження механізмів регулювання та контролю [3].

Для забезпечення етичності штучного інтелекту застосовуються такі підходи: використання якісних даних, програмування етичних обмежень, аудит алгоритмів, забезпечення прозорості рішень та участь людини у процесі прийняття рішень (human-in-the-loop) [4].

Штучний інтелект є потужним інструментом розвитку суспільства, однак його використання супроводжується низкою етичних викликів. Забезпечення справедливості, прозорості та безпеки AI-систем можливе лише за умови комплексного підходу, що поєднує технічні рішення та людський контроль. Водночас штучний інтелект не має власної моралі, тому відповідальність за його застосування повністю покладається на людину [1].

Список використаних джерел:

1. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson, 2021. 1136 p.
2. Floridi L. Ethics of Artificial Intelligence. Oxford: Oxford University Press, 2023. 320 p.
3. Analytics Steps. AI Ethics: Addressing Bias and Fairness in Machine Learning Models. URL: https://www.analyticssteps.com (дата звернення: 18.03.2026).

4. European Commission. Ethics Guidelines for Trustworthy AI. URL: https://digital-strategy.ec.europa.eu (дата звернення: 18.03.2026).
5. IBM. AI Ethics. URL: https://www.ibm.com/artificial-intelligence/ethics (дата звернення: 18.03.2026).

УДК 004.8:004.932

РОЗРОБКА ФОТОРЕАЛІСТИЧНОГО ЦИФРОВОГО АВАТАРА З ВИКОРИСТАННЯМ ГЕНЕРАТИВНИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ

Компанієць Ю. М.
yuliannakompaniets@gmail.com
Черкаський державний фаховий бізнес–коледж
Ночевнов Д. П.
м. Черкаси, Україна

Цифрові аватари сьогодні активно використовуються у відеоконференціях, онлайн-освіті, іграх, медичних тренажерах і середовищах розширеної реальності, а потреба у якісних і реалістичних аватарах постійно зростає. Якщо раніше їх створення ґрунтувалося переважно на трудомісткому ручному 3D-моделюванні, то поява генеративних моделей штучного інтелекту дала змогу автоматично синтезувати фотореалістичні обличчя та змінювати їх параметри на основі лише кількох вхідних фотографій.

GAN стали першою широко застосованою архітектурою для генерації реалістичних зображень, а StyleGAN – одним із найвідоміших рішень для синтезу облич із високою якістю та керованістю стилем [1]. Variational Autoencoder (VAE) реалізують інший підхід через латентний простір, поступаючись GAN у деталізації, але забезпечуючи стабільніше навчання та зручну інтерполяцію. Отже, GAN і VAE можна розглядати як недифузійні генеративні підходи, що становлять теоретичну основу для порівняння із сучасними дифузійними моделями. Дифузійні моделі виконують генерацію шляхом поетапного видалення шуму, а Latent Diffusion Model, на якій побудований Stable Diffusion, переносить цей процес у стиснутий латентний простір, що пришвидшує обчислення [2]. Додатково ControlNet надає таким