УДК 004.89

**О. Б. Мокін**[1]
**Л. Р. Кулик**[1]

# ПРОБЛЕМА ЗАБЕЗПЕЧЕННЯ КОНСИСТЕНТНОСТІ ГЕНЕРАЦІЇ ДИФУЗІЙНИХ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

[1] Вінницький національний технічний університет

***Анотація***
*Показано актуальність проблеми забезпечення консистентності генерування зображень за допомогою дифузійних моделей глибокого навчання та наведено приклади існуючих підходів забезпечення консистентності, їх переваги та недоліки.*
**Ключові слова:** глибоке навчання, дифузійні генеративні моделі, консистентність генерації.

***Abstract***
*There has been shown the relevance of the problem of ensuring the consistency of image generation using deep learning diffusion models and provided examples of existing approaches with their pros and cons.*
**Keywords**: deep learning, diffusion generative models, generation consistency.

За останнє десятиріччя глибоке навчання стало одним з найбільш активно зростаючих напрямків не лише в межах розвитку штучного інтелекту, але й у світі сучасних технологій загалом. А буквально за останній рік на перший план вийшли так звані генеративні дифузійні моделі глибокого навчання, які демонструють вражаючу ефективність генерації у багатьох областях [1, 2]. Однак, при роботі з такими моделями виникають проблеми, пов'язані з консистентністю генерації, тобто здатністю забезпечити стабільні, змістовні та очікувані результати генерування, наприклад:

1) створене за допомогою дифузійної моделі зображення може бути розмитим або мати недостатню деталізацію (як всього зображення, так і окремих його частин), що може бути спричинено багатьма факторами, у тому числі параметрами шуму, використаного для ініціалізації моделі, якістю навчальних даних тощо;

2) моделі можуть генерувати недостатню кількість деталей та об’єктів (або не всі бажані чи вказані у підказці/prompt’і) на результуючих зображеннях або демонструвати значну чутливість навіть до відносно малих змін вхідних даних;

3) актуальна також проблема стабільності тренування генеративних моделей, коли навчання моделі може бути невідтворюваним при повторному запуску, що також може впливати на розвиток та покращення такої моделі в рамках створення системи генерування консистентних зображень;

4) складність або неможливість забезпечення консистентності об’єктів на різних згенерованих зображеннях із взаємопов’язаними сценами.

Серед існуючих методів контролю над стилем та змістом згенерованих зображень, варто згадати перспективні моделі та підходи, такі як DreamBooth [3] та Textual Inversion [4]. Окрему увагу заслуговує популярна система ControlNet [5], яка дозволяє враховувати пози людей, глибину об’єктів та сцен, сегментаційні маски та інше. Однак, всі ці підходи також мають свої недоліки, зокрема недостатній рівень автоматизації, оскільки, незважаючи на додаткові можливості, вони перекладають побудову структури зображення чи окремих об’єктів з дифузійної моделі на користувача, тобто створення сцени виконується в ручному режимі.

З метою забезпечення консистентності можна застосувати корегування моделі під час процесу генерування з метою наближення результатів роботи моделі до бажаних. Корегування може здійснюватися, наприклад, шляхом «ін’єкції» корегувальних ваг у спеціальну дотреновану LLM (Large Language Model) [6], яка відповідає за формування сцени. Декомпозиція задачі генерації на формування сцени та її наповнення дозволяє зберегти зв’язок між різними елементами вже створеної сцени.

## Висновки

Показано, що проблема забезпечення консистентності генерування зображень за допомогою дифузійних моделей глибокого навчання є вкрай актуальною для галузі штучного інтелекту (та глибокого навчання зокрема). Наведено приклади найбільш ефективних та перспективних підходів забезпечення консистентності, їх переваги та недоліки.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ


1. Jonathan Ho, Ajay Jain, Pieter Abbeel “Denoising Diffusion Probabilistic Models” arXiv:2006.11239 [cs.LG], Jun. 2020.
2. Flavio Schneider “ArchiSound: Audio Generation with Diffusion” arXiv:2301.13267 [cs.SD], Jan. 2023.
3. Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, Kfir Aberman “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation” arXiv:2208.12242 [cs.CV], Aug. 2022.
4. Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, Daniel Cohen-Or “An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion” arXiv:2208.01618 [cs.CV], Aug. 2022.
5. Lvmin Zhang, Maneesh Agrawala “Adding Conditional Control to Text-to-Image Diffusion Models” arXiv:2302.05543 [cs.CV], Feb. 2023.
6. Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, Luke Zettlemoyer “QLoRA: Efficient Finetuning of Quantized LLMs” arXiv:2305.14314 [cs.LG], May. 2023.



***Мокін Олександр Борисович*** – доктор технічних наук, професор, професор кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: abmokin@gmail.com

***Кулик Леонід Русланович*** – аспірант факультету інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: leonidkulik2707@gmail.com

***Mokin Oleksandr*** – Dr. Sc. (Eng.), Professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: abmokin@gmail.com

***Kulyk Leonid*** – a graduate student, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, email : leonidkulik2707@gmail.com