

Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>

Ткачишин Андрій

аспірант Вінницького державного педагогічного університету імені Михайла Коцюбинського

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ТА ІМЕРСИВНИХ ТЕХНОЛОГІЙ У ПРОТИДІЇ КІБЕРБУЛІНГУ

Розвиток інформаційних технологій значно розширив можливості спілкування та доступу до інформації. Разом з тим зросла популярність соціальних мереж, онлайн-ігор та месенджерів, що сприяло поширенню нових форм агресивної поведінки в цифровому середовищі. Однією з найбільш актуальних проблем сучасного суспільства є кібербулінг [3; 4].

Особливо вразливими до кібербулінгу є діти та підлітки, для яких цифровий простір став невід'ємною частиною повсякденного життя. Тому пошук ефективних способів попередження та протидії кібербулінгу є важливим завданням для освітян [5; 6; 7].

Штучний інтелект є одним із найдієвіших інструментів боротьби з кібербулінгом. Сучасні алгоритми машинного навчання здатні аналізувати великі обсяги текстових повідомлень, коментарів та публікацій у соціальних мережах з метою виявлення образливої або агресивної лексики [1].

Системи на основі штучного інтелекту можуть:

- автоматично розпізнавати образливі висловлювання;
- визначати рівень агресії у повідомленнях;
- виявляти повторні випадки переслідування користувачів;
- повідомляти модераторів про потенційні випадки кібербулінгу та блокувати небезпечний контент ще до його поширення [5].

Використання технологій обробки природної мови (Natural Language Processing, NLP) дозволяє враховувати не лише окремі слова, а й контекст повідомлень, що підвищує точність виявлення кібербулінгу [1].

Окрім штучного інтелекту використовують також імерсивні технології, які охоплюють віртуальну реальність, доповнену реальність та змішану реальність.

Вони створюють ефект занурення користувача у цифрове середовище та забезпечують новий рівень взаємодії з інформацією [2].

У процесі протидії кібербулінгу такі технології використовуються для моделювання ситуацій онлайн-цькування, розвитку емпатії до постраждалих, формування навичок безпечної поведінки в інтернеті, навчання алгоритмам реагування на випадки кібербулінгу та проведення інтерактивних тренінгів для учнів і студентів [2].

Завдяки віртуальній реальності користувач може побачити ситуацію очима жертви кібербулінгу, що сприяє кращому усвідомленню психологічних наслідків агресивної поведінки в мережі[2].

Найбільш дієвим методом протидії кібербулінгу є поєднання штучного інтелекту з імерсивними технологіями. Такі системи здатні не лише виявляти випадки кібербулінгу, але й здійснювати профілактичну роботу[1; 2].

Доприкладу, штучний інтелект може аналізувати поведінку користувачів у віртуальному середовищі та пропонувати індивідуальні навчальні сценарії. У свою чергу, імерсивні технології забезпечують практичне відпрацювання навичок безпечного спілкування та реагування на прояви агресії.

В освітньому середовищі такі рішення можуть використовуватися для проведення тренінгів з цифрової безпеки та формування культури відповідальної поведінки в Інтернеті.

Таким чином можна зробити висновок що ШІ - це технології, що дозволяють комп'ютерам імітувати людське мислення та приймати рішення.

Імерсивні технології створюють ефект занурення користувача у віртуальне середовище. Основна мета яких, аналіз даних, навчання, прогнозування та автоматизація процесів.

Дані технології мають різне призначення, але можуть ефективно доповнювати одне одного. ШІ забезпечує аналіз та обробку інформації, тоді як імерсивні технології створюють умови для практичного навчання та отримання досвіду. У протидії кібербулінгу їх поєднання дозволяє одночасно виявляти небезпечний контент і навчати користувачів безпечній поведінці в цифровому середовищі.

Список використаних джерел

1. Биков В. Ю., Спірін О. М., Пінчук О. П. Сучасні завдання цифрової трансформації освіти. Інформаційні технології і засоби навчання. 2020. Т. 77, № 3. С. 1–15.
2. Литвинова С. Г. Імерсивні технології в освіті: сучасний стан та перспективи використання. Освітологічний дискурс. 2022. № 1. С. 45–58.

3. Livingstone S., Smith P. K. Annual Research Review: Harms experienced by child users of online and mobile technologies. *Journal of Child Psychology and Psychiatry*. 2014. Vol. 55(6). P. 635–654.

4. Kowalski R. M., Limber S. P., Agatston P. W. *Cyberbullying: Bullying in the Digital Age*. 3rd ed. Hoboken : Wiley-Blackwell, 2018. 272 p.

5. Hinduja S., Patchin J. W. *Cyberbullying: Identification, Prevention, and Response*. Cyberbullying Research Center. 2023. 42 p.

6. UNESCO. *Behind the Numbers: Ending School Violence and Bullying*. Paris : UNESCO Publishing, 2019. 70 p.

7. UNICEF. *Cyberbullying: What Is It and How to Stop It*. New York : UNICEF, 2023.

Ушкальов Володимир

кандидат економічних наук, доцент кафедри економіки підприємства
та організації бізнесу

Харківського національного економічного університету
імені Семена Кузнеця

ШТУЧНИЙ ІНТЕЛЕКТ ТА «ТРАГЕДІЯ ДОБРОЧЕСНОСТІ»

У 1968 році професор Каліфорнійського університету Гарретт Джеймс Хардін опублікував статтю з екологічної проблематики під назвою «The Tragedy of the Commons» [1]. В ній на прикладі використання мешканцями громади спільного пасовища було продемонстровано, що спільне користування ресурсом призводить до стрімкого падіння раціональності та ефективності його використання. На думку вченого за наявності спільної власності прагнення кожного окремого користувача ресурсу максимізувати індивідуальну корисність спонукає людей експлуатувати ресурс нерационально. Через це ресурс швидко знищується, а загальний економічний потенціал спільноти та суспільства – зменшуються. І хоча дещо пізніше Елінор Остром своїми дослідженнями довела, що «трагедія спільного» не є невідворотною, не завжди призводить до катастрофи та може бути попереджена [2], ситуація з поширенням технологій штучного інтелекту (ШІ) знову змушує констатувати актуальність сформульованої майже 70 років тому проблеми.

Регламент (ЄС) 2024/1689, що відомий як «AI Act», і який набув чинності у 2026 році, визначає ШІ як інструмент для людей «з кінцевою метою підвищення добробуту людства» [3]. Метою запровадження регламенту (як в ньому й