

Додаток 1

до Положення Про відповідальне впровадження та використання технологій штучного інтелекту в Національному університеті «Одеська політехніка»

АЛГОРИТМ ОЦІНКИ РИЗИКІВ ПРИ ВИБОРІ ТА ВИКОРИСТАННІ СИСТЕМ ШІ

Цей алгоритм призначений для попередньої оцінки придатності систем ШІ для використання в Одеській політехніці з огляду на ризики порушення прав людини, інтелектуальної власності та захисту персональних даних.

I. Оцінка розробника та платформи (юридичний аудит)

Експертна рада з питань штучного інтелекту (AI Expert Council) повинна перевірити систему за певними критеріями.

1. Юрисдикція та походження:

- «червоний прапорець» (неприйнятний ризик): країною реєстрації розробника, місцем знаходження серверів або кінцевим бенефіціаром є російська федерація, Іран чи КНДР.
- високий ризик: країни з високим індексом порушення прав людини або наявність підсанкційних осіб серед власників;
- середній ризик: країни поза межами України та ЄС (крім зон «червоних прапорців»).

2. Права інтелектуальної власності:

- «червоний прапорець» (високий ризик): якщо розробник не гарантує, що згенерований контент не порушує прав третіх осіб, або забороняє використання результатів в освітній чи науковій діяльності.
- варто звернути увагу: чи залишає розробник за собою права на ваші промпти або бази даних.

3. Приватність та захист даних:

- «червоний прапорець» (неприйнятний ризик): відсутність можливості видалити персональні дані з промптів або несанкціонована передача промптів третім особам без належного захисту.
- високий ризик: система не гарантує конфіденційності інформації, внесеної користувачем.

II. Оцінка згенерованого контенту (операційний аудит)

Перед оприлюдненням результатів ШІ необхідно провести перевірку за такими пунктами.

1. Контроль з боку людини та достовірність:

- «червоний прапорець»: контент не пройшов перевірку людиною на відповідність етичним та юридичним нормам;
- варто звернути увагу: чи є факти в контенті перевіреними? чи не вводить використання цього контенту в оману (наприклад, у чутливих темах)?

2. Безпека та конфіденційність результату:

- «червоний прапорець»: у згенерованому контенті виявлено персональні дані реальних осіб (імена, обличчя, голоси) або комерційну таємницю Університету.

3. Етика та інклюзивність:

- високий ризик: контент містить ознаки дискримінації, дезінформації, розпалювання ворожнечі або пропаганду насильства;
- середній ризик: відсутність маркування про те, що контент створений за допомогою ШІ.

III. Шкала оцінки та прийняття рішень

1. Неприйнятний ризик (наявність хоча б одного «червоного прапорця»): використання системи ШІ або згенерованого нею контенту в Університеті категорично заборонено.
2. Високий ризик: використання можливе лише за умови критичної необхідності та впровадження додаткових засобів контролю, оскільки існує висока ймовірність репутаційних втрат.
3. Середній/низький ризик: використання дозволено за умови дотримання загальних правил атрибуції та етики, визначених цим Положенням.

Низькі ризики системи автоматично не гарантують безпеку контенту. Кожен результат роботи ШІ повинен проходити оцінювання людиною на предмет ризиків його подальшого використання.

✓