

**Рекомендації щодо відповідального використання  
генеративного штучного інтелекту в дослідженнях. Документ  
зацікавлених сторін Форуму ERA.**

Зміст

1. ВСТУП
2. КЕРІВНІ ПОЛОЖЕННЯ ЩОДО ВІДПОВІДАЛЬНОГО ВИКОРИСТАННЯ  
ГЕНЕРАТИВНОГО ШІ В ДОСЛІДЖЕННЯХ
  - 2.1. РЕКОМЕНДАЦІЇ ДЛЯ ДОСЛІДНИКІВ
  - 2.2. РЕКОМЕНДАЦІЇ ДЛЯ ДОСЛІДНИЦЬКИХ ОРГАНІЗАЦІЙ
  - 2.3. РЕКОМЕНДАЦІЇ ДЛЯ ОРГАНІЗАЦІЙ, ЩО ФІНАНСУЮТЬ ДОСЛІДЖЕННЯ
3. ПЕРЕДУМОВИ. ЦІЛІСНІСТЬ ДОСЛІДЖЕНЬ
4. НАДІЙНИЙ ШІ
5. ІНШІ ОСНОВИ ПРИНЦИПІВ
6. СИНТЕЗ ДЛЯ РОЗРОБКИ КЕРІВНИХ ПОЛОЖЕНЬ ЄС

## 1. ВСТУП

Штучний інтелект (ШІ) став невід'ємною частиною нашого повсякденного життя, змінюючи те, як ми живемо та працюємо. За останні кілька років спостерігається швидке та революційне прискорення розвитку можливостей ШІ, зумовлене значним прогресом у широкій доступності даних, обчислювальній потужності та методах машинного навчання. Зокрема, було досягнуто помітних успіхів у розробці базових моделей - моделей ШІ, навчених на великих обсягах немаркованих даних.

1. Ці досягнення призвели до появи того, що відомо як «ШІ загального призначення», здатного виконувати широкий спектр завдань. Це включає «генеративний ШІ», який може генерувати різні форми нового контенту (текст, код, дані, зображення, музику, голос, відео тощо), зазвичай на основі інструкцій (також відомих як підказки), наданих користувачем. Якість вихідних даних, що створюються цими моделями, така, що їх може бути важко відрізнити від контенту, створеного людиною.

Широке та швидке впровадження генеративного ШІ привернуло значну увагу та призвело до широких політичних та інституційних відповідей. ЄС бере на себе світову ініціативу із Законом про штучний інтелект, а також вживаються інші міжнародні зусилля з управління. Серед них Хіросімський процес<sup>2</sup>, очолюваний G7, Блетчлійська декларація<sup>3</sup>, підписана після першого Саміту з безпеки ШІ, Паризька заява, підписана після Саміту дій щодо ШІ 2025 року<sup>4</sup>, та Рамкова конвенція Ради Європи про ШІ<sup>5</sup>.

### **Генеративний ШІ та дослідження**

Генеративний ШІ надає багато можливостей для різних секторів. Однак він також містить ризики, такі як масштабне генерування дезінформації, проблеми інтелектуальної власності та захисту даних, а також інші неетичні способи використання зі значними суспільними та екологічними наслідками.

Дослідження є одним із секторів, які можуть найбільше постраждати від генеративного штучного інтелекту.

ШІ має великий потенціал для прискорення наукових відкриттів, що призведе до нових проривів у дослідженнях та значного підвищення продуктивності, а також покращить ефективність та темпи процесів досліджень та верифікації. Позитивні приклади використання цих інструментів дослідниками включають підтримку неносіям мови у створенні текстів кількома мовами, швидке створення текстових

резюме з різних джерел у надзвичайно великих корпусах, автоматичне отримання та контекстуалізація широкого обсягу знань та допомогу в написанні коду для аналізу даних<sup>6</sup>. Однак ця технологія також несе ризик зловживання. Деякі ризики пов'язані з технічними обмеженнями інструменту, а інші пов'язані з (навмисним чи ненавмисним) використанням інструменту таким чином, що це підриває надійну дослідницьку практику.

Інші ризики для досліджень у Європі можуть впливати з власницького характеру деяких інструментів (наприклад, відсутність відкритості, плата за доступ до сервісу, використання вхідних даних), концентрації власності або небажаної передачі критично важливих технологій та інтелектуальної власності<sup>7</sup>. Вплив генеративного штучного інтелекту на дослідження та різні аспекти наукового процесу вимагає рефлексії, наприклад, під час роботи з текстом (підсумовування статей, мозковий штурм або дослідження ідей, складання чи переклад). У багатьох відношеннях ці інструменти можуть зашкодити цілісності досліджень та поставити під сумнів здатність сучасних моделей боротися з оманливими науковими практиками та дезінформацією.

### **Чому ці рекомендації**

Різні установи, включаючи університети, дослідницькі організації, фінансові органи та видавці, видали рекомендації щодо належного використання цих інструментів, щоб забезпечити повне використання переваг цих інструментів. Поширення рекомендацій<sup>8</sup> та рекомендацій створило складний ландшафт, який ускладнює визначення того, яких рекомендацій слід дотримуватися в конкретному контексті.

З цієї причини Європейський форум дослідницького простору<sup>9</sup> (що складається з європейських країн та зацікавлених сторін у сфері досліджень та інновацій<sup>10</sup>) вирішив розробити рекомендації щодо використання генеративного ШІ в дослідженнях для: фінансових органів, дослідницьких організацій та дослідників, як у державних, так і в приватних дослідницьких екосистемах.

Ці рекомендації зосереджені на одному конкретному типі ШІ, що використовується в дослідницькому процесі: генеративному ШІ. Це важливий крок для запобігання зловживанням та забезпечення того, щоб генеративний ШІ відігравав позитивну роль у дослідницькій практиці. Одна з цілей цих рекомендацій полягає в тому, щоб наукова

спільнота використовувала цю технологію відповідально. Однак, розробка надійної системи для генеративного ШІ в наукових дослідженнях не може бути виключною відповідальністю політиків (на європейському та національному рівнях). Університети, дослідницькі організації, органи фінансування, дослідницькі бібліотеки, наукові товариства, видавці, керівники досліджень, допоміжний персонал досліджень та дослідники на всіх етапах їхньої кар'єри є важливими у формуванні дискусії щодо ШІ та того, як він може служити суспільним інтересам у дослідженнях. Усі вони повинні активно брати участь в обговореннях щодо відповідального та ефективного впровадження застосувань ШІ, підвищуючи обізнаність та культивуючи відповідальне використання ШІ як частини дослідницької культури, що базується на спільних цінностях. Правила та рекомендації повинні йти пліч-о-пліч із широким залученням тих, хто бере участь у державних та приватних дослідженнях, як організацій, так і окремих осіб, для розвитку культури використання генеративного ШІ в дослідженнях належним та ефективним чином.

Ці рекомендації мають на меті встановити спільні напрямки відповідального використання генеративного ШІ. Хоча вони не є обов'язковими, їх слід розглядати як допоміжний інструмент для дослідників, дослідницьких організацій та органів, що фінансують дослідження, включаючи тих, що подають заявку на Європейську рамкову програму досліджень та інновацій. Вони враховують ключові принципи доброчесності досліджень, а також вже існуючі рамки для використання ШІ загалом та в дослідженнях зокрема. Користувачам цих рекомендацій пропонується адаптувати їх до своїх конкретних контекстів та ситуацій, пам'ятаючи про пропорційність.

Ці керівні принципи доповнюють та розвивають політичну рамку ЄС щодо штучного інтелекту, включаючи Закон про штучний інтелект<sup>11</sup>. Вони доповнюють інші види політичної діяльності щодо впливу штучного інтелекту на науку. До них належать висновок Механізму наукових консультацій<sup>12</sup> (SAM) щодо штучного інтелекту та політичний документ<sup>13</sup>, опублікований Генеральним директором Європейської комісії з досліджень та інновацій, що визначає проблеми та можливості.

### **Ключові принципи**

Набір принципів, що лежать в основі цих керівних принципів, базується на вже існуючих відповідних рамкових принципах:

- Європейський кодекс поведінки щодо доброчесності досліджень<sup>14</sup>;
- робота та керівні принципи щодо надійного штучного інтелекту, розроблені Експертною групою високого рівня з питань штучного інтелекту<sup>15</sup>;

Спираючись на спільні риси керівних принципів, що розробляються на даний момент, від різних зацікавлених сторін, ключовими принципами, що лежать в основі цих керівних принципів щодо відповідального використання генеративного штучного інтелекту в дослідженнях, є:

- Надійність у забезпеченні якості досліджень, що відображається в дизайні, методології, аналізі та використанні ресурсів. Це включає аспекти, пов'язані з перевіркою та відтворенням інформації, отриманої за допомогою штучного інтелекту для досліджень. Це також передбачає усвідомлення можливих проблем рівності та недискримінації стосовно упередженості та неточностей.
- Чесність у розробці, проведенні, огляді, звітності та комунікації досліджень прозоро, справедливо, ретельно та неупереджено. Цей принцип включає розкриття інформації про використання генеративного ШІ.
- Повага до колег, учасників дослідження, суб'єктів дослідження, суспільства, екосистем, культурної спадщини та навколишнього середовища. Відповідальне використання генеративного ШІ повинно враховувати обмеження технології, її вплив на навколишнє середовище<sup>16</sup> та її суспільні наслідки (упередженість, різноманітність, недискримінація, справедливість та запобігання шкоді). Це включає належне управління інформацією, повагу до конфіденційності, конфіденційності та прав інтелектуальної власності, а також належне цитування.
- Відповідальність за дослідження від ідеї до публікації, за його управління та організацію, за навчання, нагляд та наставництво, а також за його ширший суспільний вплив. Це включає відповідальність за всі результати, які створює дослідник, що ґрунтується на понятті людської діяльності та нагляду.

## **Наступні кроки**

Генеративний штучний інтелект може створювати можливості та ризики, які важко передбачити та повністю зрозуміти сьогодні. Ці рекомендації базуються на поточному стані технологій та навколишньому політичному ландшафті. Враховуючи динамічний характер обох, цим рекомендаціям потрібно буде постійно адаптуватися та

розвиватися. Вони будуть регулярно оновлюватися, щоб забезпечити їхню корисність для дослідників та організацій. Ці рекомендації були розроблені та продовжуватимуть розроблятися спільно в рамках діяльності Форуму ERA. Користувачі цих рекомендацій можуть надавати відгуки та пропозиції щодо адаптації у формі зворотного зв'язку<sup>17</sup>.

## **2. РЕКОМЕНДАЦІЇ ЩОДО ВІДПОВІДАЛЬНОГО ВИКОРИСТАННЯ ГЕНЕРАТИВНОГО ШІ У ДОСЛІДЖЕННЯХ**

Розгляньте використання генеративного ШІ третіми сторонами, які можуть взаємодіяти з вами/вашою організацією.

- Дослідники та організації пам'ятають, що треті сторони можуть використовувати інструменти генеративного ШІ. Наприклад, у контексті зустрічей, ведення нотаток, підбиття підсумків обговорень або створення оглядів документів (які можуть бути конфіденційними). Тому вони можуть зіткнутися з ризиками, подібними до тих, що описані в цьому керівництві, а саме щодо конфіденційності, захисту даних або прав інтелектуальної власності, пов'язаних із захищеною або оригінальною інформацією.
- Інформуйте треті сторони про використання таких інструментів та потенційні ризики, наприклад, поширюючи ці рекомендації.
- У своїх контактах із третіми сторонами повідомляйте їм, чи використовуєте ви ці інструменти ШІ і як ви захищаєте зібрану інформацію.

### **РЕКОМЕНДАЦІЇ ДЛЯ ДОСЛІДНИКІВ**

Для відповідального використання генеративного ШІ дослідники повинні:

#### *1. Нести повну відповідальність за наукові результати.*

- Дослідники несуть відповідальність за цілісність контенту<sup>18</sup>, який вони створюють, згенерованого за допомогою інструментів штучного інтелекту або за їхньої підтримки.
- Дослідники дотримуються критичного підходу до використання результатів, створених генеративним штучним інтелектом, і усвідомлюють обмеження інструментів, такі як упередженість<sup>19</sup>, галюцинації<sup>20</sup> та неточності.
- Системи штучного інтелекту не є ні авторами, ні співавторами. Авторство передбачає свободу дій та відповідальність, тому воно лежить на дослідниках-людях. Моделі генеративного ШІ можуть мати обмеження, які впливають як на їхні відповіді, так і на зручність використання. Найпоширенішими типами є:

- Упередженість навчальних даних: Упередженість у даних, що використовуються для навчання моделей генеративного ШІ, може призвести до перекошених відповідей, що відображає неточності або системні упередження у вихідному матеріалі.
- Упередженість підказки<sup>16</sup>: Моделі можуть узгоджувати свої відповіді зі сприйнятими переконаннями або уподобаннями користувача, явище, відоме як підлабузницька поведінка, що потенційно призводить до оманливо прийнятних або упереджених результатів.
- Вигадані цитати та неправильні резюме: Моделі генеративного ШІ можуть генерувати правдоподібні, але неправильні цитати, які можуть ввести в оману користувачів, які покладаються на ці джерела для дослідження або перевірки. Навіть коли цитати правильні (стаття існує, має правильну назву), згенерований ШІ резюме цих статей може бути неправильним. Тому дослідники несуть відповідальність за перевірку всіх посилань та змісту цитацій та резюме.
- Інтерпретованість: генеративні моделі ШІ функціонують як «чорні скриньки», що ускладнює розуміння того, як генеруються конкретні відповіді. Ця непрозорість підкреслює важливість перехресної перевірки, особливо в автоматизованому аналізі даних, де відповіді моделі можуть суттєво впливати на висновки.

## *2. Використовуйте генеративний ШІ прозоро.*

- Дослідники, щоб бути прозорими, детально описують, які інструменти генеративного ШІ були використані суттєво<sup>21</sup> у їхніх дослідницьких процесах. Коли генеративний ШІ змістовно формує результати, дослідники прозоро зазначають його використання відповідно до рекомендацій свого журналу або стандартів у своїй дисципліні в розділі методів (або еквіваленті), відповідально оцінюючи ступінь внеску<sup>22</sup>. Цей принцип прозорості також застосовується до використання ШІ для виявлення або скринінгу поганих практик (тобто виявлення галюцинацій або плагіату), зберігаючи при цьому людський нагляд.
- Дослідники враховують стохастичну (випадкову) природу інструментів генеративного ШІ, яка полягає в тенденції створювати різні результати з тих самих вхідних даних. Дослідники прагнуть відтворюваності та надійності своїх результатів і висновків. Вони розкривають або обговорюють обмеження, що використовуються

інструментами генеративного ШІ, включаючи можливі упередження у згенерованому контенті, а також можливі заходи щодо їх пом'якшення.

*3. Звертайте особливу увагу на питання, пов'язані з конфіденційністю, недоторканністю та правами інтелектуальної власності, під час обміну конфіденційною або захищеною інформацією за допомогою інструментів штучного інтелекту.*

- Дослідники пам'ятають, що згенеровані або завантажені вхідні дані (текст, дані, підказки, зображення тощо) можуть бути використані для інших цілей, таких як навчання моделей штучного інтелекту.

Тому вони захищають неопубліковані або конфіденційні роботи (такі як власні або чужі неопубліковані роботи), намагаючись не завантажувати їх у зовнішню систему штучного інтелекту, якщо немає гарантій, що дані не будуть використані повторно, наприклад, для навчання майбутніх мовних моделей або для повторного використання даних, яке неможливо відстежити та перевірити.

- Дослідники намагаються не надавати персональні дані третіх сторін зовнішнім генеративним системам штучного інтелекту, якщо суб'єкт даних (фізична особа) не дав їм свою згоду, і дослідники не мають чіткої мети, для якої персональні дані будуть використані, щоб забезпечити дотримання правил ЄС щодо захисту даних<sup>2324</sup>.

- Дослідники розуміють технічні, етичні та безпекові наслідки щодо конфіденційності, недоторканності та прав інтелектуальної власності. Вони перевіряють, наприклад, свої інституційні правила, параметри конфіденційності інструментів, хто керує інструментом (державні чи приватні установи, компанії тощо), де працює інструмент та наслідки для будь-якої завантаженої інформації. Це може варіюватися від закритих середовищ, розміщення на сторонній інфраструктурі з гарантованою конфіденційністю, до відкритих платформ, доступних через Інтернет.

*4. Дотримуйтесь чинного національного, європейського та міжнародного законодавства. Зокрема, результати генеративного ШІ можуть бути особливо чутливими щодо захисту прав інтелектуальної власності та персональних даних.*

- Дослідники звертають увагу на потенційну можливість плагіату (текст, код, зображення тощо) під час використання результатів генеративного ШІ. Дослідники поважають авторство інших та цитують їхню роботу, де це доречно. Результат

генеративного ШІ (наприклад, велика мовна модель) може базуватися на результатах інших осіб і вимагати належного розпізнавання та цитування<sup>25</sup>.

- Результат генеративного ШІ може містити персональні дані. Якщо це стає очевидним, дослідники несуть відповідальність за належне поводження з будь-якими персональними даними, а також повинні дотримуватися правил ЄС щодо захисту даних.

*5. Постійно навчайтеся правильно використовувати інструменти генеративного ШІ, щоб максимізувати їхні переваги, зокрема шляхом проходження навчання.*

- Інструменти генеративного ШІ швидко розвиваються, і регулярно відкриваються нові способи їх використання<sup>26</sup>. Дослідники слідкують за передовим досвідом та діляться ним з колегами та іншими зацікавленими сторонами.

- Дослідники прагнуть мінімізувати вплив генеративного ШІ на навколишнє середовище<sup>27</sup> шляхом оцінки того, чи найкраще підходить інструмент ШІ для поставленого завдання, і який саме, та використовуючи найефективніші методи підказування.

*6. Утримуйтеся від використання інструментів генеративного ШІ суттєво<sup>28</sup> у делікатній діяльності, яка може вплинути на інших дослідників або організації (наприклад, експертна оцінка<sup>29</sup>, оцінка дослідницьких пропозицій тощо).*

- Уникнення використання інструментів генеративного ШІ усуває потенційні ризики несправедливого ставлення або оцінки, які можуть виникнути через обмеження цих інструментів (такі як галюцинації та упередження).

- Крім того, це захистить оригінальну неопубліковану роботу колеґ-дослідників від потенційного розкриття або включення до моделі ШІ (за умов, детально описаних вище у рекомендації для дослідників №3).

## **2.2. РЕКОМЕНДАЦІЇ ДЛЯ ДОСЛІДНИХ ОРГАНІЗАЦІЙ**

Для відповідального використання генеративного ШІ дослідницькі організації повинні:

1. Сприяти, спрямовувати та підтримувати відповідальне використання генеративного ШІ в дослідницькій діяльності.

- Дослідницькі організації забезпечують та/або сприяють навчанню для всіх рівнів кар'єри та дисциплін, включаючи керівників досліджень та допоміжний персонал

досліджень, з використання генеративного ШІ, особливо (але не виключно) з питань перевірки результатів, збереження конфіденційності, усунення упереджень та захисту прав інтелектуальної власності та конфіденційних знань.

- Дослідницькі організації сприяють атмосфері довіри, де дослідників заохочують прозоро розкривати інформацію про використання генеративного ШІ без побоювань щодо негативних наслідків.<sup>30</sup>

- Дослідницькі організації надають підтримку та рекомендації для забезпечення дотримання етичних та правових вимог (правила ЄС щодо захисту даних, захист прав інтелектуальної власності тощо).

2. Активно відстежувати розвиток та використання систем генеративного ШІ у своїх організаціях.

- Дослідницькі організації повинні пам'ятати про дослідницьку діяльність та процеси, для яких вони використовують генеративний ШІ, щоб краще підтримувати його майбутнє використання<sup>26</sup>. Ці знання можуть:

- бути використані для надання подальших рекомендацій щодо використання генеративного ШІ, допомогти визначити потреби в навчанні та зрозуміти, яка підтримка може бути найбільш корисною;

- допомогти передбачити та запобігти можливому неправильному використанню та зловживанню інструментами ШІ;

- бути опублікованими та поширеними серед наукової спільноти.

- Дослідницькі організації аналізують обмеження технологій та інструментів і надають відгуки та рекомендації своїм дослідникам.

- Дослідницькі організації відстежують вплив генеративного ШІ на навколишнє середовище в межах своїх організацій та сприяють ініціативам з підвищення обізнаності, щоб допомогти своїм співробітникам вибрати найбільш сталий варіант.

3. Посилатися на ці рекомендації щодо генеративного ШІ або інтегрувати їх у свої загальні дослідницькі керівні принципи щодо належної дослідницької практики та етики.

- Використовуючи ці рекомендації як основу для обговорення, дослідницькі організації відкрито консультуються зі своїми дослідницькими співробітниками та

зацікавленими сторонами щодо використання генеративного ШІ та відповідної політики.

- Дослідницькі організації застосовують ці рекомендації, коли це можливо. За потреби, їх можна доповнити конкретними додатковими рекомендаціями та/або винятками, які слід опублікувати для забезпечення прозорості.

4. Коли це можливо та необхідно, впроваджувати локально розміщені або хмарні генеративні інструменти штучного інтелекту, якими вони керують самостійно<sup>31</sup>. Це дозволяє їхнім співробітникам вносити свої наукові дані в інструмент, який забезпечує захист даних та конфіденційність.

- Організації забезпечують належний рівень кібербезпеки цих систем, особливо тих, що підключені до Інтернету.

### **2.3. РЕКОМЕНДАЦІЇ ДЛЯ ОРГАНІЗАЦІЙ, ЩО ФІНАНСУЮТЬ ДОСЛІДЖЕННЯ**

Організації, що фінансують дослідження, працюють у різних контекстах та дотримуються різних мандатів та правил, які можуть не узгоджуватися з єдиним набором інструкцій. Наведені нижче рекомендації окреслюють набір заходів та належних практик для організацій, які слід впроваджувати таким чином, що найкраще відповідає їхнім конкретним обставинам та цілям.

Щоб генеративний ШІ використовувався відповідально, організації, що фінансують дослідження, повинні:

1. Сприяти та підтримувати відповідальне використання генеративного ШІ в дослідженнях.

- Організації, що фінансують дослідження, розробляють інструменти фінансування, які є відкритими, сприйнятливими та підтримують відповідальне та етичне використання технологій генеративного ШІ в дослідницькій діяльності.

- Організації, що фінансують дослідження, вимагають, щоб фінансовані дослідження та грантоотримувачі відповідали чинному національному, європейському та міжнародному законодавству (де це застосовується) та належним практикам використання генеративного ШІ.

- Організації, що фінансують дослідження, заохочують дослідників, керівників досліджень, персонал допоміжного дослідження та дослідницькі організації використовувати генеративний ШІ етично та відповідально, включаючи дотримання правових вимог та вимог дослідницьких стандартів.

2. Переглянути використання генеративного ШІ у своїх внутрішніх процесах. Вони будуть лідерами, забезпечуючи його прозоре та відповідальне використання.

- Організації, що фінансують дослідження, залишаються повністю відповідальними за використання генеративного ШІ у своїй діяльності відповідно до принципу підзвітності, який наголошує на відповідальності та людському контролю.

- Організації, що фінансують дослідження, використовують генеративний ШІ прозоро, зокрема для діяльності, пов'язаної з управлінням оцінюванням та оцінюванням, без пошкодження їхньої справедливості та конфіденційності контенту.<sup>32</sup> Генеративний ШІ може бути використаний для покращення внутрішніх процесів для спонсорів, але він не може брати участь в оцінюванні наукового змісту проектів.

- Вибираючи та використовуючи інструменти генеративного ШІ, організації, що фінансують дослідження, повинні ретельно розглянути дотримання інструменту стандартами та проблеми його використання,

під час обміну конфіденційною або захищеною інформацією з інструментами ШІ. Організації враховують елементи, що стосуються якості, прозорості, цілісності, захисту даних, конфіденційності та поваги до прав інтелектуальної власності, а також його впливу на навколишнє середовище.

3. Вимагати від заявників прозорості щодо використання ними генеративного штучного інтелекту, що полегшує способи звітування про це.

- Заявники заявляють, чи суттєво вони<sup>33</sup> використовували інструмент(и) генеративного штучного інтелекту для підготовки своєї заявки.

- Заявники надають інформацію про роль генеративного штучного інтелекту в дослідницькій діяльності, що пропонується та виконується.

- Організації, що фінансують дослідження, сприяють створенню атмосфери, в якій дослідники можуть прозоро та без побоювань бути у невідповідному становищі<sup>24</sup>.

4. Моніторинг та активна участь у швидкозмінному ландшафті генеративного ШІ.

- Організації, що фінансують дослідження, сприяють та фінансують навчальні та освітні програми для етичного та відповідального використання ШІ в наукових дослідженнях.

5. Використовуючи генеративний ШІ як частину внутрішніх процесів, враховуйте ризики, пов'язані з прихованими підказками<sup>34</sup> у контенті, отриманому третіми сторонами, зокрема, у заявках та пропозиціях.

- Організації, що фінансують дослідження, усвідомлюють, що контент може містити приховані підказки або невидимі інструкції, призначені для впливу на інструменти генеративного ШІ.

- Де це доречно, вони встановлюють чіткі правила, що забороняють такі маніпуляції, та запроваджують відповідні запобіжні заходи, зокрема в ІТ-системах, що беруть участь в обробці цієї інформації.

## **ПЕРЕДУМОВИ**

### **ДОБРОСЛІДНІСТЬ ДОСЛІДЖЕНЬ**

Європейський кодекс поведінки щодо доброчесності досліджень, розроблений усіма європейськими академіями (ALLEA), встановлює низку принципів для проведення якісних досліджень, включаючи етичні аспекти. Ці принципи включають:

- надійність у забезпеченні якості досліджень, що відображається в дизайні, методології, аналізі та використанні ресурсів;
- чесність у розробці, проведенні, огляді, звітності та комунікації досліджень прозоро, справедливо, ретельно та неупереджено;
- повага до колег, учасників досліджень, суб'єктів досліджень, суспільства, екосистем, культурної спадщини та навколишнього середовища;
- відповідальність за дослідження від ідеї до публікації, за його управління та організацію, за навчання, нагляд та наставництво, а також за його ширший суспільний вплив.

### **НАДІЙНИЙ ШІ**

Етичні рекомендації щодо надійного ШІ, розроблені Експертною групою високого рівня ЄС з питань ШІ, зосереджені на розробці відповідальних систем ШІ (а не на їх використанні). Ця робота над надійним ШІ містить низку етичних принципів, які були перетворені на операційні вимоги до

розробки систем ШІ. Ці вимоги можуть бути корисним натхненням для розробки рекомендацій щодо використання генеративного штучного інтелекту.

Чотири етичні принципи для систем штучного інтелекту:

1. повага до людської автономії;
2. запобігання шкоді;
3. справедливість;
4. пояснимість.

Ці етичні принципи були використані для розробки цих семи ключових операційних вимог:

1. людська діяльність та нагляд<sup>35</sup>;
2. технічна надійність та безпека;
3. конфіденційність та управління даними;
4. прозорість;
5. різноманітність, недискримінація та справедливість;
6. екологічне та суспільне благополуччя;
7. підзвітність.

## **ІНШІ ОСНОВИ ПРИНЦИПІВ**

Принципи KU Leuven були використані в аналогічних вправах та для розробки цих керівних принципів:

- прозорість використання генеративного ШІ, залежно від типу використання;
- перевірка правильності згенерованого результату з увагою до правильного вказання джерела;
- повага до персональних даних та конфіденційної інформації шляхом невведення їх на платформи, які не керуються на власних серверах;
- відповідальність за правильне використання генеративного ШІ (перш за все допомога та підтримка) та опубліковані результати.

Фахівці зі ШІ, генеративного ШІ, інформатики, психології та соціології, що працюють в Інституті перспективних досліджень Амстердамського університету, визначили принципи підзвітності, прозорості та незалежного нагляду (стосовно розробників генеративного ШІ). На основі цих принципів вони запропонували практичні

рекомендації щодо відповідального використання генеративного ШІЗ6. Ці принципи можна доповнити принципами з інших джерел. До них належать принципи, висунуті Асоціацією комп'ютерної техніки щодо генеративних технологій штучного інтелекту, які мають широку сферу застосування для тих, хто розробляє, впроваджує та використовує технології. Іншими джерелами є Структура цінностей для організації досліджень, розроблена Science Europe, або принципи STM37 для етичного та надійного штучного інтелекту.

## **СИНТЕЗ ДЛЯ РОЗРОБКИ КЕРІВНИХ ПРАВ ЄС**

Принципи ALLEA в Європейському кодексі поведінки щодо доброчесності досліджень служать керівними принципами для етичних та відповідальних досліджень. Така ж система цінностей застосовується під час використання генеративного штучного інтелекту в дослідженнях. Тому принципи ALLEA можуть бути використані як основа для цих керівних принципів та збагачені частинами з інших рамок, описаних у попередніх розділах.

### *Силки:*

<sup>1</sup> <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=COM:2024:28:FIN>

<sup>2</sup> <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-aisystems>

<sup>3</sup> <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchleydeclaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>

<sup>4</sup>

Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet

<sup>5</sup> The Framework Convention on Artificial Intelligence - Artificial Intelligence

<sup>6</sup> Results of a survey by Nature, <https://doi.org/10.1038/d41586-023-02980-0> and follow up survey <https://doi.org/10.1038/d41586-024-03940-y>

<sup>7</sup> Council Recommendation of 23 May 2024 on enhancing research security (ST/9097/2024/INIT)

<sup>8</sup> Different guidelines that have been identified are listed in the References section of this document.

<sup>9</sup> <https://european-research-area.ec.europa.eu/era-forum>

<sup>10</sup> <https://european-research-area.ec.europa.eu/era-history>

- 11 <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- 12 <https://scientificadvice.eu/advice/artificial-intelligence-in-science>
- 13 [https://research-and-innovation.ec.europa.eu/document/download/1e2a4c9c-d3f1-43e9-9488-c8152aabf25f\\_en](https://research-and-innovation.ec.europa.eu/document/download/1e2a4c9c-d3f1-43e9-9488-c8152aabf25f_en)
- 14 European Code of Conduct for Research Integrity Revised Edition 2023
- 15 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 16 When assessing the environmental impact of AI, it is important to examine and monitor both its software and hardware life cycles. See *Artificial Intelligence (AI) end-to-end: The environmental impact of the full AI lifecycle needs to be comprehensively assessed – Issue note*.
- 17 Feedback form for the guidelines: [https://ec.europa.eu/eusurvey/runner/feedback\\_GenAIResearch](https://ec.europa.eu/eusurvey/runner/feedback_GenAIResearch)
- 18 Content should not be fabricated, falsified or plagiarised. For example, researchers do not use generative AI to falsify, alter or manipulate original research data. More details in Section 3.1 of *The European Code of Conduct for Research Integrity*, ALLEA 2023.
- 19 Ethical prompting should be applied. See the Ethical Prompting for Generative AI guidelines of the ETH Zurich for more information. <https://www.research-collection.ethz.ch/handle/20.500.11850/672207>, ETH Zurich 2024. On sycophantic behaviour see for example <https://doi.org/10.48550/arXiv.2310.13548>
- 20 "hallucination" refers to the generation of false, nonsensical, or inaccurate information by large language models (LLMs) or other generative AI systems.
- 21 In this case, for example, using generative AI as a basic editorial support tool is not a substantial use. However, using it as a method or source, interpreting data analysis, carrying out a literature review, identifying research gaps, formulating research aims, developing hypotheses, etc. could have a substantial impact.
- 22 References to the tool could include the name, version, date, etc. and how it was used and affected the research process. If relevant, researchers make the input (prompts) and output available, in line with open science principles. Specifically, wherever possible, final output shall follow the FAIR principles (Findable, Accessible, Interoperable, Reusable) and be itself machine actionable.
- 23 [https://commission.europa.eu/law/law-topic/data-protection/eu-data-protection-rules\\_en](https://commission.europa.eu/law/law-topic/data-protection/eu-data-protection-rules_en). See also the *First Orientations for ensuring data protection compliance when using Generative AI systems* by the European Data Protection Supervisor.
- 24 Apart from personal consent there could be other lawful basis that could justify the processing of personal data.
- 25 The technology is evolving, and generative AI tools will be more capable of providing correct citations. Tools providing citations could be more practical and could be favoured, but the final responsibility on the citation and its correctness remains with the researcher.
- 26 See as an example of considerations to follow Nogueira *et al*, 2025 <https://doi.org/10.1177/03400352241304121>
- 27 This can be seen as part of the broader initiatives to reduce the carbon footprint of research, following the principles for sustainable coding and sustainable research (see for example <https://doi.org/10.1371/journal.pcbi.1009324> and <https://doi.org/10.1371/journal.pcbi.1008148>). Preliminary data on the environmental cost of generative AI can be found in the AI Index Report 2023.
- 28 In this case, for example, using generative AI to search background info for a review using it to better understand punctually some ideas or concepts is not a substantial use, while delegating the evaluation or the assessment of a paper, drafting peer review report, helping to understand the main elements of a paper, or evaluating a funding proposal is a substantial use.
- 29 *ChatGPT is transforming peer review — how can we use it responsibly?*
- 30 Disclosure of the use of generative AI for assistance when writing may lead to a lower quality score during assessment, see for example Li *et al*, 2024, DOI: 10.18653/v1/2024.emnlp-main.279 or Suchikova, Y., & Tsybuliak, N. (2025). The Purity Myth: Why Stigmatizing GAI in Academic Writing Is Harmful. *Science Communication*, 47(5), 753-761. <https://doi.org/10.1177/10755470241313233>. Research organisations should try to avoid situations like these ones.
- 31 Or governed by trustworthy third parties e.g. partner research organizations, the EU or trusted countries.
- 32 In activities related with the assessment and evaluation of the scientific content in itself, generative AI should be avoided as described in the recommendation for researchers number 6.
- 33 As mentioned in the recommendations for researchers, using generative AI as a basic support tool for authors is not a substantial use. However, interpreting data analysis, carrying out a literature review, identifying research gaps, formulating research aims, developing hypotheses, etc. could have a substantial impact.
- 34 A hidden prompt is an instruction secretly embedded in a document or input for an AI system, without being visible to the human reader. Its purpose is to influence how the AI responds, for example by steering it toward a more favourable assessment. In practice, this can be done through invisible text, formatting tricks, or metadata that a person may not notice but an AI tool may still process.
- 35 Human agency and oversight includes human-in-the-loop (HITL), human-on-the-loop (HOTL) and human-in-command (HIC) approaches. HITL involves human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL involves human intervention during the design cycle of the system and monitoring how the system operates. HIC involves overseeing the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation (Ethics Guidelines for Trustworthy AI).
- 36 <https://www.nature.com/articles/d41586-023-03266-1>
- 37 The International Association of Scientific, Technical and Medical Publishers

### **Список літератури, статті і документи:**

Birhane, A., Kasirzadeh, A., Leslie, D. & Wachter, S., 'Science in the Age of Large Language Models', *Nature Reviews Physics*, Vol. 5, 2023, pp. 277-280.

<https://doi.org/10.1038/s42254-023-00581-4>

Conroy, G., 'Scientific Sleuths Spot Dishonest ChatGPT Use in Papers', *Nature*, 2023.

<https://doi.org/10.1038/d41586-023-02477-w>

Conroy, G., 'How ChatGPT and Other AI Tools Could Disrupt Scientific Publishing', *Nature*, Vol.

622, 2023, pp. 234-236. <https://doi.org/10.1038/d41586-023-03144-w>

Ganjavi, C., Eppler, M., Pekcan, A., Biedermann, B., Abreu, A., Collins, G., Gill, I. Cacciamani, G., 'Bibliometric Analysis of Publisher and Journal Instructions to Authors on Generative-AI in Academic and Scientific Publishing', *arXiv*, 2023. <https://doi.org/10.48550/arXiv.2307.11918>

Lenharo, M., 'ChatGPT turns two: how the AI chatbot has changed scientists' lives', *Nature*, Vol. 636, pp. 281-282 (2024). <https://doi.org/10.1038/d41586-024-03940-y>

Li, Z., Liang, C., Peng, J., Yin, M., 'How Does the Disclosure of AI Assistance Affect the Perceptions of Writing?' *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4849-4868 (2024). [10.18653/v1/2024.emnlp-main.279](https://doi.org/10.18653/v1/2024.emnlp-main.279)

Lorenz, P., Perset, K. & Berryhill, J., 'Initial Policy Considerations for Generative Artificial Intelligence', *OECD Artificial Intelligence Papers*, No. 1, OECD Publishing, Paris, 2023.

<https://doi.org/10.1787/fae2d1e6-en>

Lund, B.D., Wang, T., Mannuru, N.R., Nie, B., Shimray, S. & Wang, Z., 'ChatGPT and a New Academic Reality: Artificial Intelligence-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing', *Journal of the Association for Information Science and Technology*, Vol. 74, No. 5, 2023, pp. 570-581. [http://dx.doi.org/10.1002/asi.24750](https://doi.org/10.1002/asi.24750)

Nogueira, L. A., Moltubakk, S. T., Fagervik, A., & Langfeldt, I. B. 'Cutting through the noise: Assessing tools that employ artificial intelligence.', *IFLA Journal*, 0(0), 2025

<https://doi.org/10.1177/03400352241304121>

Parrilla, J.M., 'ChatGPT Use Shows that the Grant-Application System is Broken', *Nature*, 2023.

<https://doi.org/10.1038/d41586-023-03238-5>

International Association of Scientific, Technical and Medical Publishers (STM), 'AI Ethics in Scholarly Communication: STM Best Practice Principles for Ethical, Trustworthy and Humancentric AI', 2021.

<https://coilink.org/20.500.12592/hqbzq79>

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., ... & Perez, E. (2023). 'Towards understanding sycophancy in language models.' *arXiv*

<https://doi.org/10.48550/arXiv.2310.13548>

Van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL., 'ChatGPT: Five Priorities for Research', *Nature*, Vol. 614, pp. 224-226 (2023) <https://doi.org/10.1038/d41586-023-00288-7>

16  
Van Noorden R. & Perkel J., 'AI and Science: What 1,600 Researchers Think', *Nature*, Vol. 621, pp. 672-675 (2023) <https://doi.org/10.1038/d41586-023-02980-0>

United Nations Environment Programme, 'Artificial Intelligence (AI) end-to-end: The Environmental Impact of the Full AI Lifecycle Needs to be Comprehensively Assessed - Issue Note' (2024) <https://wedocs.unep.org/handle/20.500.11822/46288>

Zou, J., 'ChatGPT is transforming peer review — how can we use it responsibly?', *Nature* Vol. 635, p.10 (2024) <https://doi.org/10.1038/d41586-024-03588-8>

### **Проаналізовано рекомендації:**

Aalto University:

<https://www.aalto.fi/en/services/responsible-use-of-artificial-intelligence-in-the-research-process>

All European Academies (ALLEA): <https://allea.org/wp-content/uploads/2023/06/EuropeanCode-of-Conduct-Revised-Edition-2023.pdf> (DOI: 10.26356/ECOC)

Arizona State University:

<https://researchintegrity.asu.edu/export-controls-and-security/artificial-intelligence>

Association for Computing Machinery (ACM):

<https://www.acm.org/binaries/content/assets/public-policy/ustpc-approved-generative-aiprinciples>

Australian Government Tertiary Education Quality and Standards Agency (TEQSA):

<https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificialintelligence>

Bockting CL, van Dis EAM, van Rooij R, Zuidema W, Bollen J. Living guidelines for generative AI - why scientists must oversee its use. *Nature*. 2023 Oct;622(7984):693-696.

<https://doi.org/10.1038/d41586-023-03266-1>

Committee on Publication Ethics (COPE):

<https://publicationethics.org/cope-position-statements/ai-author>

Deakin University: <https://deakin.libguides.com/generative-AI/ethics-evaluation>

Deutsche Forschungsgemeinschaft (DFG):

[https://www.dfg.de/download/pdf/dfg\\_im\\_profil/geschaeftsstelle/publikationen/stellungnahmen\\_papiere/2023/230921\\_statement\\_executive\\_committee\\_ki\\_ai.pdf](https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/stellungnahmen_papiere/2023/230921_statement_executive_committee_ki_ai.pdf)

Elsevier: <https://www.elsevier.com/about/policies/publishing-ethics>

ETH Zurich: <https://www.research-collection.ethz.ch/handle/20.500.11850/672207>

Harvard University:

<https://provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard>

International Association of Scientific, Technical and Medical Publishers (STM):

[Global Principles for AI](https://www.stm.org/global-principles-for-ai)

17

International Association of Scientific, Technical and Medical Publishers (STM):

[Ethical and practical guidelines for the use of generative AI in the publishing process](https://www.stm.org/ethical-and-practical-guidelines-for-the-use-of-generative-ai-in-the-publishing-process)

International Society for Computational Biology (ISCB):

<https://www.iscb.org/iscb-policy-statements/iscb-policy-for-acceptable-use-of-large-languagemodels>

KU Leuven: <https://research.kuleuven.be/en/integrity-ethics/integrity/practices/genai/genai>

Maastricht University: <https://www.maastrichtuniversity.nl/news/guidelines-use-generative-ai>

Science Europe:

<https://www.scienceeurope.org/our-priorities/research-culture/research-values-framework>

United Nations Educational, Scientific and Cultural Organization (UNESCO):

<https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>

University of Ljubljana: <https://www.uni-lj.si/mma/uien/2023092013475160/?m=1695210471>

University of Toronto:

<https://www.sgs.utoronto.ca/about/guidance-on-the-use-of-generative-artificial-intelligence/>