Омбудсман України
Ombudsman of Ukraine

Фінансується Європейським Союзом

FIIAPP
COOPERACIÓN ESPAÑOLA

Ministry of Digital Transformation of Ukraine

aepd agencia española protección datos

# РЕКОМЕНДАЦІЇ З ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ В ОБРОБЦІ ПЕРСОНАЛЬНИХ ДАНИХ:

## НАЙКРАЩІ ПРАКТИКИ ЄС

В епоху, що відзначається надзвичайно глибокою інтеграцією штучного інтелекту (ШІ) до нашого повсякденного життя, починаючи від персоналізованих маркетингових стратегій і до складних систем прийняття рішень, ще ніколи захист персональних даних не відігравав такої критично важливої ролі. «Рекомендації з етики використання ШІ в обробці персональних даних: найкращі практики ЄС» — це важливий проєкт для підтримки крихкого балансу між прогресивними можливостями ШІ та нагальною потребою в захисті даних.

Цю публікацію підготовлено в рамках міжнародного проєкту EU4DigitalUA у співробітництві з Офісом Омбудсмена України та Міністерством цифрової трансформації України.

Ці рекомендації являють собою синтез компетенцій Іспанської агенції із захисту даних (AEPD) та втілюють у собі акумульований досвід Європейського Союзу у сфері захисту даних та регулювання у сфері ШІ. Тому цей документ є не лише набором рекомендацій, але й відображенням найкращих практик у одній із найбільш прогресивних регуляторних юридичних рамок для захисту даних у світі.

Цей документ, який було підготовлено з надзвичайною уважністю та ретельністю, і його метою є стати у пригоді в якості керівництва для організацій, а також для уповноважених та посадових осіб у галузі політики захисту даних. У цьому документі надається чітка й легка для прикладного застосування інформація про юридичні рамки задля гарантії того, що ШІ-системи розроблено, розгорнуто та скеровано в чіткій відповідності до етичних та правових стандартів обробки даних, особливо тих, що викладено в Загальному регламенті захисту даних (GDPR).

Керівні принципи включають основні принципи захисту даних, такі як прозорість, мінімізація даних та підзвітність, і налаштовані для вирішення унікальних проблем, пов'язаних із технологіями штучного інтелекту. Поєднуючи теоретичні основи захисту даних із практичними сценаріями застосування ШІ, цей документ прагне досягти тонкого балансу між технічним прогресом та невід'ємними правами на конфіденційність та захист даних.

Крім того, це керівництво виходить за рамки простого дотримання правил і втілює в собі зобов'язання щодо етичної відповідальності. Вона закликає організації активно вплітати конфіденційність у структуру систем штучного інтелекту, тим самим підвищуючи довіру та безпеку технологій штучного інтелекту та гарантуючи, що вони приносять користь громадськості, захищаючи індивідуальні свободи.

Зрештою, «Рекомендації з етики використання ШІ в обробці персональних даних: найкращі практики ЄС» символізують собою більше, ніж нормативний посібник; вони представляють собою зобов'язання щодо етичного управління персональними даними в епоху ШІ. Ці рекомендації є важливим кроком вперед у нашому спільному шляху до відповідального ШІ та виступають за майбутнє, де технологічні інновації та права на конфіденційність поєднуються, сприяючи прогресу, захищаючи гідність, повагу та конфіденційність людей.

Передмова - Андрій Ніколаєв,  юрист-експерт з етики штучного інтелекту, приватності та захисту даних.

Участь у підготовці публікації:

**Луіс де Сальвадор Карраско**
**Андрій Ніколаєв** – передмова, переклад та редагування україномовної версії та **Зоя Медюк**, підготовка україномовної версії.

Проєкт EU4DigitalUA являє собою частину діяльності Європейського Союзу, направленої на підтримку України. Погляди, думки та висновки, висловлені в тексті нижче, належать винятково авторові й не обов'язково відображають позицію проєкту, Європейського Союзу або Міжнародного та іберо-американського фонду адміністрації та державної політики FIIAPP F.S.P.

Омбудсман України
Ombudsman of Ukraine

Фінансується Європейським Союзом

FIIAPP
COOPERACIÓN ESPAÑOLA

Ministry of Digital Transformation of Ukraine

aepd agencia española protección datos

# РЕКОМЕНДАЦІЇ З ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ В ОБРОБЦІ ПЕРСОНАЛЬНИХ ДАНИХ:

## НАЙКРАЩІ ПРАКТИКИ ЄС

# INDEX

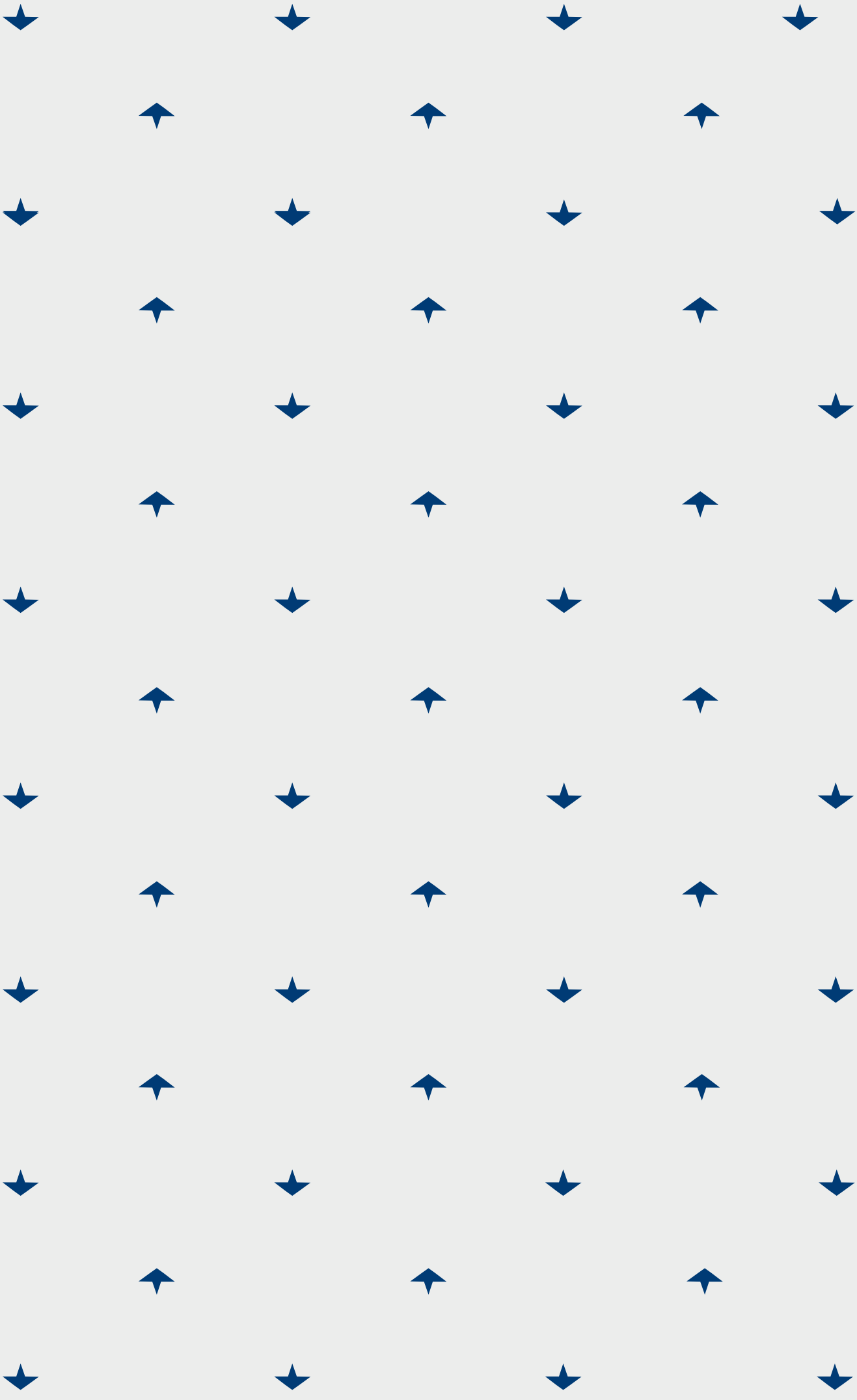
# СПИСОК СКОРОЧЕНЬ

| | |
|---|---|
| **AEPD** | Іспанський орган із захисту даних |
| **ШІ** | Штучний інтелект |
| **КРИМІНАЛЬНИЙ КОДЕКС** | Органічний закон 10/1995 від 23 листопада про Кримінальний кодекс |
| **DPA** | Орган із захисту даних |
| **EDPB** | Європейська рада із захисту даних |
| **EDPS** | Європейський інспектор із захисту даних |
| **ЄС** | Європейський Союз |
| **FIIAPP** | Міжнародний та іберо-американський фонд управління та державної політики |
| **GDPR** | РЕГЛАМЕНТ (ЄС) ЄВРОПЕЙСЬКОГО ПАРЛАМЕНТУ ТА РАДИ 2016/679 від 27 квітня 2016 року |
| **LOPDGDD** | Органічний закон 3/2018 від 5 грудня, «Захист персональних даних та гарантія цифрових прав» |
| **ML** | Машинне навчання |

**ОКРЕМІ ТЕРМІНИ:**

**Контролер даних, контролер:** Володілець персональних даних - в термінах законодавства України

**Процесор даних, процесор / оператор даних, оператор:** Розпорядник персональних даних - в термінах законодавства України



# 1. ВСТУП


Серед багатьох інших, одним із можливих визначень «системи штучного інтелекту» є:

> *«Система штучного інтелекту» (ШІ) означає систему, призначену для роботи з певним рівнем автономії, яка на основі вхідних даних, наданих машинами або людьми, робить висновок про те, як досягти набору заявлених цілей, використовуючи стратегії та знання на основі логіки або машинного навчання, а також генерує вихідну інформацію, таку як контент (генеративні системи штучного інтелекту), прогнози, рекомендації або рішення, що впливають на середовище, з яким вона взаємодіє.*

Наразі система штучного інтелекту - це парасолькове визначення, яке включає безліч різних автоматизованих систем обробки даних: системи машинного навчання, прогнозний аналіз, системи профілювання даних, будь-які розробки на основі наукових даних та загалом усі системи, що здійснюють автоматизоване прийняття рішень. Системи штучного інтелекту визначають спосіб розвитку систем обробки даних.

Необхідно враховувати, що ШІ-система буде складатися з алгоритму ШІ та інших елементів, які дозволять реалізувати та ефективно працювати алгоритму, і які можуть обумовлювати робочі параметри системи в багатьох аспектах.

З іншого боку, стаття 2, розділ 1 Регламенту (ЄС) 2016/679 Європейського Парламенту та Ради від 27 квітня 2016 року про захист фізичних осіб у зв'язку з обробкою персональних даних та вільним обігом цих даних, яка скасовує Директиву 95/46/ЄС (Загальний регламент про захист даних, або GDPR), встановлює матеріальну сферу в обробці персональних даних, а не в системах, технологіях чи методах інфраструктури:

> *1. Це Положення застосовується до повністю або частково автоматизованої обробки персональних даних, а також до неавтоматизованої обробки персональних даних, що містяться у файлі або призначені для включення до нього.*

Обробка визначатиметься, як зазначено в частині 1 статті 24 GDPR, за її характером, обсягом, контекстом та цілями:

> *Беручи до уваги характер, обсяг, контекст і цілі обробки, а також різної ймовірності та серйозності ризики для прав і свобод фізичних осіб, контролер застосовуватиме відповідні технічні та організаційні заходи для того, щоб гарантувати і мати можливість продемонструвати, що обробка відповідає цьому Регламенту. За необхідності ці заходи будуть переглядатися та оновлюватися.*

Характер діяльності з обробки передбачає спосіб, у який ця обробка ефективно здійснюється. Здійснення обробки можна розділити на різні операції обробки, як це встановлено в статті 4.2 GDPR:

> *«Обробка»: будь-яка операція або сукупність операцій, що виконуються з персональними даними або наборами персональних даних, незалежно від того, автоматизованими процедурами чи ні, такі як збирання, запис, організація, структурування, збереження, адаптація або модифікація, витяг, консультація, використання, поширення шляхом передачі, розповсюдження або будь-якої іншої форми доступу, порівняння або встановлення взаємозв'язків, обмеження, видалення або знищення.*

Операції обробки, в свою чергу, можуть виконуватися вручну або бути автоматизованими, повністю або частково. Автоматизоване розгортання може здійснюватися за допомогою різних систем, таких як мобільні системи, локальні системи зберігання, хмарні системи, системи шифрування, системи відеоспостереження та/або системи штучного інтелекту. Усі вони повинні передбачати достатнє документування та надання контролерам і операторам достатньої кількості інформації, щоб гарантувати виконання ними обов'язків щодо комплаєнсу.

# 2. ВИЗНАЧЕННЯ

- **Доступ** (Access): будь-яке використання даних відповідно до конкретних технічних, юридичних або організаційних вимог **без обов'язкової передачі або завантаження даних**.[1]
- **Автоматизовані рішення** (Automated decisions): рішення, прийняті на основі виключно автоматизованої обробки, які відображають здатність приймати рішення за допомогою технологічних засобів без участі людини[2]. Разом із тим пункт 71 Преамбули GDPR та стаття 22 GDPR обмежують застосування автоматизованих рішень та встановлюють право суб'єкта даних не підкорятися рішенням, заснованим виключно на автоматизованій обробці[3], які мають правові наслідки, що суттєво впливають на суб'єкта даних. Автоматичне профілювання віднесено до автоматизованих рішень. Робоча група за статтею 29 оцінила наслідки цього права в Керівних принципах WP29 щодо автоматизованого індивідуального прийняття рішень та профілювання для цілей Регламенту 2016/679. Однак слід враховувати, що автоматизовані рішення можуть включати або не включати профілювання, а профілювання може здійснюватися з використанням автоматизованих рішень або без них.
- **Обчислення до даних** (Compute-to-data): стратегія, за якою замість відправлення даних до обчислювальних ресурсів обчислювальні ресурси переносяться до місця знаходження даних. Таким чином, конфіденційність даних зберігається, а контролер (держатель даних) зберігає більший контроль над обробкою даних. Одним зі способів реалізації стратегії **Compute-to-data** є федеративне навчання, але це не єдиний спосіб.
- **Каталог даних** (Data catalogue): сукупність систематично організованих описів наборів даних, яка містить відкриту частину для користувача, де інформація про окремі параметри набору даних може бути доступна в електронному вигляді через онлайн-портал.[4]
- **Каталогізація даних** (Data cataloguing): обробка, яка здійснюється на даних або наборі даних, що дозволяє пов'язати з ними метадані, необхідні для подальшого використання даних. Як правило, це передбачає створення каталогів ресурсів (даних), які можуть бути доступні для кількох користувачів.[5]
- **Держатель даних** (Data Holder): юридична особа, включаючи органи державного сектору та міжнародні організації, або фізична особа, інша, ніж суб'єкт даних щодо відповідних конкретних даних, яка відповідно до чинного законодавства Союзу або національного законодавства має право надавати доступ до певних персональних чи неперсональних даних або поширювати їх.[6]

1 Стаття 2(13) DGA

2 Настанови щодо автоматизованого індивідуального прийняття рішень та профілювання для цілей Регламенту 2016/679

3 Для того, щоб вважати, що мало місце втручання людини, нагляд за рішенням повинен здійснюватися компетентною особою, уповноваженою змінювати рішення, і її втручання має бути значним, а не суто символічним.

4 Стаття 2(2)(ac) пропозиції EHDS

5 Адаптовано з публікації Data Office від 28.12.2020

6 Стаття 2(8) DGA

- **Якість даних** (Data quality): з точки зору Простору даних (Data Space)[7] якість даних є суб'єктивною характеристикою,[8] пов'язаною з корисністю набору даних для конкретної обробки.[9] Ця концепція відрізняється від принципу точності даних[10] у GDPR.
- **Обмін даними** (Data sharing): надання даних суб'єктом даних або держателем даних користувачеві безпосередньо або через Медіатора за добровільною угодою або відповідно до законодавства Союзу чи національного законодавства з метою спільного чи індивідуального використання таких даних, наприклад, на основі відкритих ліцензій або платних чи безкоштовних комерційних ліцензій.[11]
- **Суверенітет даних** (Data sovereignty): концепція, яка не визначена в європейському стандарті і загалом трактується як ідея про те, що місце, де збираються дані, визначає регулювання та управління, яке до нього застосовується, а також здатність урядів та компаній використовувати цифрові дані користувачів та компаній.
- **Простір даних** (Data Space): інфраструктура, заснована на спільному управлінні, організаційних, регуляторних та технічних механізмах, яка полегшує доступ до даних і, таким чином, розробку бізнес-моделей на основі дослідження та експлуатації даних.
- **Суб'єкт даних** (Data Subject): ідентифікована фізична особа, або фізична особа яку можна ідентифікувати.[12]
- **Користувач даних** (Data User): фізична або юридична особа, яка має законний доступ до певних персональних або неперсональних даних і право, зокрема надане GDPR у випадку персональних даних, використовувати ці дані в комерційних або некомерційних цілях.[13]
- **Фасилітатор** (Enabler): суб'єкт(и), що надає(-ють) послуги або засоби, які дозволяють поширювати або використовувати набори даних та впроваджувати заходи управління.[14]
- **Федеративне навчання** (Federated learning): метод машинного навчання, який навчає алгоритм через децентралізовану архітектуру пристроїв, що містять власні локальні та приватні дані. Створений Google у 2017 році, цей підхід відрізняється від методів, коли всі дані завантажуються централізовано на сервер. Це зберігає цілісність інформації, яка використовується для навчання, без шкоди для конфіденційності та безпеки.

7 Можна ознайомитися зі стандартами UNE 0079 та ISO 25012.

8 Найкращі практики використання даних в Інтернеті: словник якості даних (w3.org)

9 Стаття 2 (ad) запропонованого EHDS визначає корисність як ступінь придатності характеристик електронних даних про здоров'я для вторинного використання.

10 Стаття 5(1)(d) GDPR

11 Стаття 2(10) del DGA

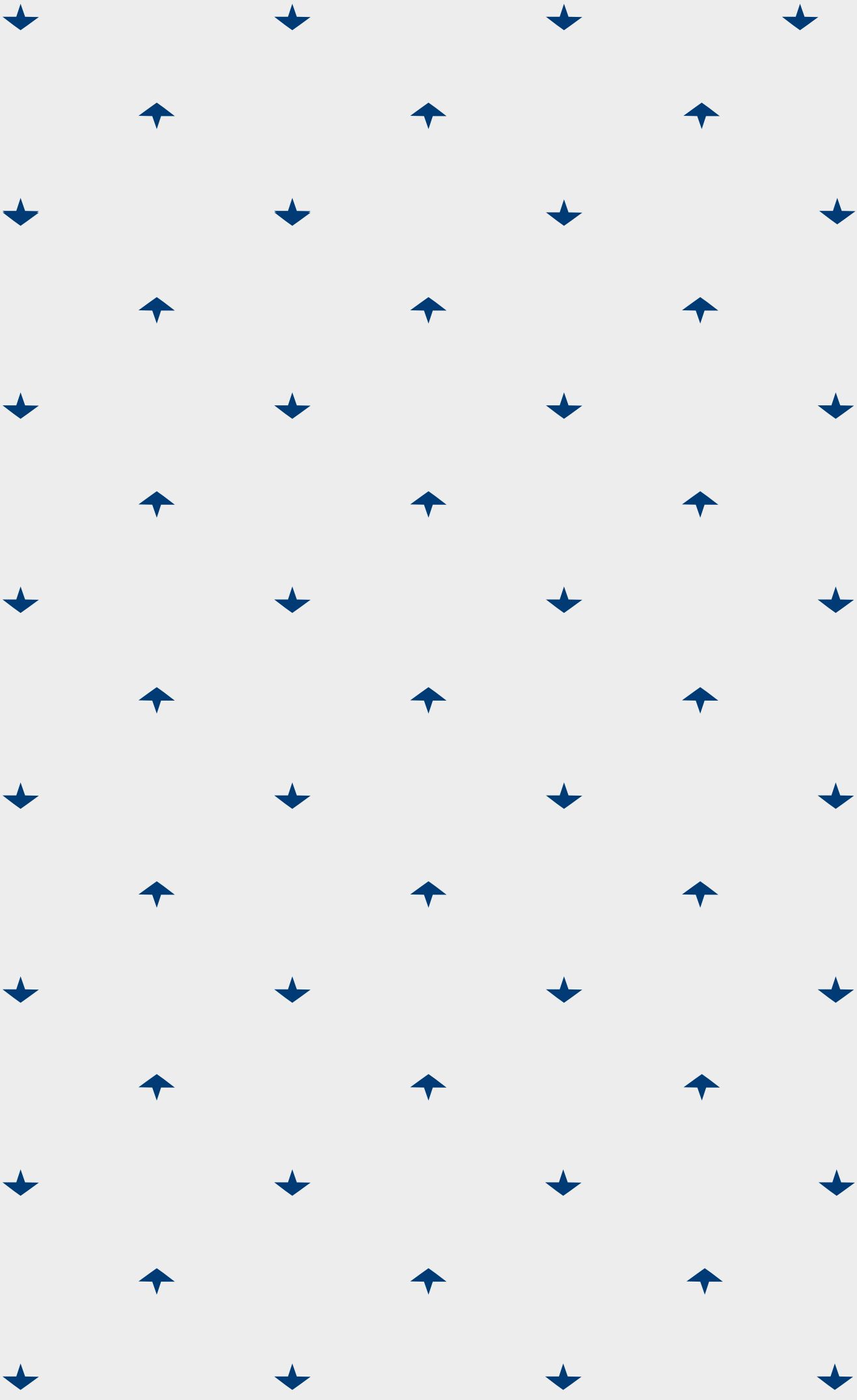
12 Стаття 4 GDPR

13 Стаття 2(9) DGA

14 Інструмент для розробки варіантів використання в просторах даних іспанського офісу даних [доступно лише іспанською мовою]

- **Гейткіпер / Регулятор доступу (Gatekeeper):** визначається в DMA як компанія, що надає послуги основної платформи, для цілей цього документа — послугу хмарних обчислень, — що має сильний вплив на внутрішній ринок і усталену та тривалу позицію.
- **Гіпермасштаб** (Hyperscale): це основна послуга платформи[15], виділена для надання послуг масового зберігання та обробки в хмарі, які можуть масштабувати розподілене обчислювальне середовище до тисяч серверів.
- **Медіатор, Медіатор даних** (Mediator, Data Mediator): суб'єкти, які встановлюють відносини між Суб'єктами даних та/або **Держателями** даних, з одного боку, та Користувачами даних, з іншого боку. У структурі DGA «компетентні органи»[16], «служби посередництва даних» (та їхній підтип «кооперативи даних») і «організації управління даними для альтруїстичних цілей» розглядаються як посередники. Згідно з пропозицією EHDS, **посередником** буде, серед іншого, центральна платформа для вторинного використання електронних даних про здоров'я. В інших областях їх називають «постачальником даних», «оператором простору даних» тощо.
- **Профілювання** (Profiling)**:** Стаття 4.4. GDPR визначає профілювання як будь-яку форму автоматизованої обробки персональних даних, що включає використання персональних даних для оцінки певних особистих аспектів, які стосуються фізичної особи, зокрема для аналізу або прогнозування аспектів, що стосуються цієї фізичної особи.[17]
  Будь-яка обробка, що включає профілювання, характеризується трьома елементами[18]:
  - Вона повинна полягати в автоматизованій обробці, включно з такими видами обробки, де частково задіяні люди.
  - Вона повинна виконуватися з використанням персональних даних.
  - Метою такого профілювання має бути оцінка особистих аспектів фізичної особи.
- **Псевдонімізація** (Pseudonymisation): обробка персональних даних таким чином, що персональні дані більше не можуть бути співвіднесені з конкретним суб'єктом даних без використання додаткової інформації, за умови, що така додаткова інформація зберігається окремо та підлягає технічним та організаційним заходам, що забезпечують, щоб персональні дані не були віднесені до фізичної особи, яка ідентифікована, або яку можна ідентифікувати.[19]
- **Безпечне середовище обробки** (Secure processing environment): фізичне або віртуальне середовище та організаційні засоби для забезпечення відповідності такому законодавству Європейського Союзу, як, наприклад, GDPR, зокрема щодо прав суб'єктів даних, прав інтелектуальної власності, комерційної та статистичної конфіденційності, цілісності та доступності, а також для забезпечення відповідності застосовному національному законодавству та надання оператору безпечного середовища обробки з можливістю визначати та контролювати всі дії з обробки даних, зокрема відображення, зберігання, завантаження та експорт даних, а також обчислення похідних даних за допомогою обчислювальних алгоритмів.[20]
- **Довірене середовище обробки** (Trusted execution environment): непорушне середовище обробки, розгорнуте на основному процесорі пристрою з апаратним і програмним забезпеченням, розробленим таким чином, щоб гарантувати цілісність і конфіденційність даних і обробки, що здійснюється на цьому процесорі, у разі будь-якого типу атак. Не плутати з Безпечним середовищем обробки (Secure Processing Environment), де, крім аспектів конфіденційності, цілісності та доступності даних, гарантуються юридичні зобов'язання, встановлені національним законодавством та законодавством ЄС[21].

15 Стаття 2(2) DMA
16 Стаття 7 DGA
17 Стаття 4.4 «профілювання»: будь-яка форма автоматизованої обробки персональних даних, що складається з використання персональних даних для оцінки певних особистих аспектів, які стосуються фізичної особи, зокрема для аналізу або прогнозування аспектів стосовно діяльності цієї фізичної особи на роботі, економічного становища, здоров'я, особистих уподобань, інтересів, надійності, поведінки, місцезнаходження або пересування;.
18 Настанови щодо автоматизованого індивідуального прийняття рішень та профілювання для цілей Регламенту 2016/679. Стаття 29 Робоча група
19 Стаття 4 GDPR
20 Стаття 2(20) DGA
21 Пункт 4.3 документа «ІНЖЕНЕРІЯ ЗАХИСТУ ДАНИХ. Від теорії до практики. Агентство Європейського Союзу з кібербезпеки (ENISA) [січень 2022 р.]»



# 3. ШІ В ОБРОБЦІ ПЕРСОНАЛЬНИХ ДАНИХ

## 3.1. АЛГОРИТМ І СИСТЕМА

Алгоритм — це те, що вважається основою системи штучного інтелекту, наприклад, вже навчена модель МН (машинного навчання). Саме про алгоритм іде мова, коли обговорюються всі ці моделі, і саме алгоритми можуть бути джерелом проблем, пов'язаних з надійністю, контекстом або упередженнями.

Однак слід враховувати, що алгоритми ШІ, як і будь-який алгоритм, розроблений у програмному забезпеченні, не є сутністю, яка працює у вакуумі. Точніше, багато проблем, які заважатимуть правильній роботі алгоритмів МН, може бути знайдено не в самому алгоритмі, а у всьому, що його оточує: у всій системі.

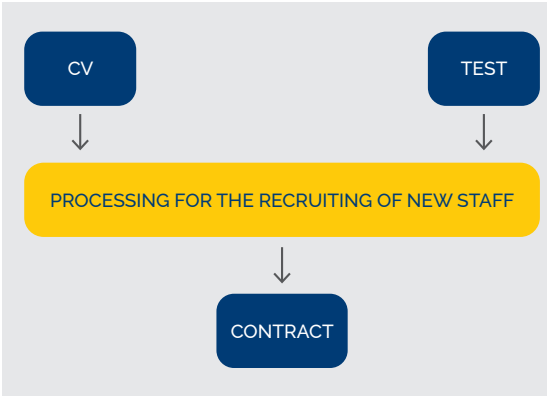
Кожен алгоритм повинен бути реалізований в коді, і на цьому етапі можна знайти більше слабких місць у фактичній продуктивності будь-якого алгоритму. У деяких випадках код, який ви використовуєте для розробки, буде використовуватися й на етапі виробничої експлуатації системи. У будь-якому випадку будь-який код не розробляється з нуля, для його розробки завжди буде підібрано набір бібліотек. Бібліотеки також є наборами коду, але вони реалізують основні функції, які в свою чергу працюють з використанням більш примітивних бібліотек, які залежать від операційної системи. Код, в якому реалізовано алгоритм, адаптується і виконується в системі, як-от персональний комп'ютер, мобільний телефон або сервер у хмарі. На цьому етапі в гру вступає більше змінних, таких як зв'язки, їхні метадані, якщо застосовно, характеристики використовуваних датчиків тощо.

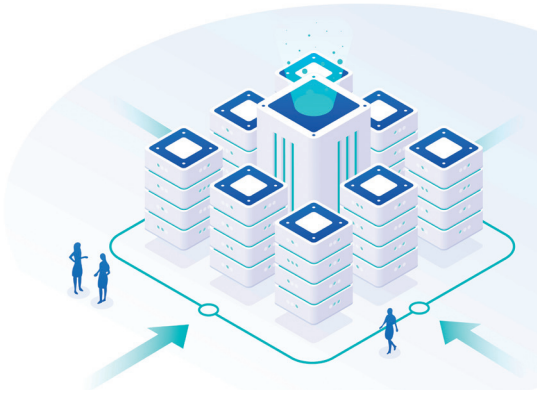
Коротше кажучи, при практичному застосуванні алгоритму МН найбільш доречним є говорити про систему МН, а не просто про алгоритм. Цей підхід не означає, що відповідна оцінка алгоритму не повинна проводитися, а скоріше, що, оцінюючи додаток МН, важливо знати, що ми говоримо про щось більш складне, з додатковими елементами, які можуть обумовити його продуктивність. Цей заклик до уваги не новий, оскільки в 80-х роках стала очевидною необхідність оцінювати алгоритми інтегровано з системами, в яких вони були реалізовані. В іншому випадку могли виникнути вкрай небажані, а в деяких випадках — і катастрофічні, ситуації, і вони таки виникали.

## 3.2. СИСТЕМА І ОБРОБКА ДАНИХ

Давайте почнемо цей новий розділ, зосередившись на штучному інтелекті і проаналізувавши приклад, щоб проілюструвати, що сама система штучного інтелекту є лише засобом для обробки, а не кінцевою метою.

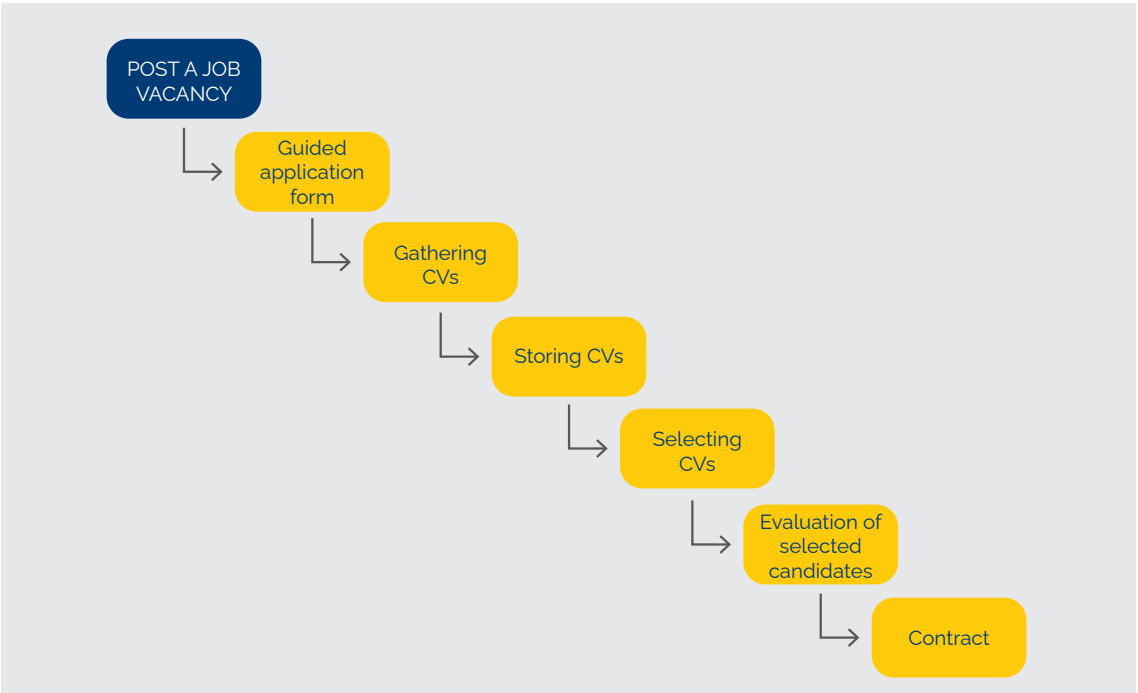
Розглянемо обробку, яку передбачає набір нового персоналу в компанії. Ця діяльність з обробки визначається її метою (рекрутинг), її сферою (тобто дані резюме, потенційні працівники, процес найму тощо), її контекстом (нормування праці, соціальна ситуація, статус такої роботи в суспільстві тощо) та її характером (або способом її реалізації).



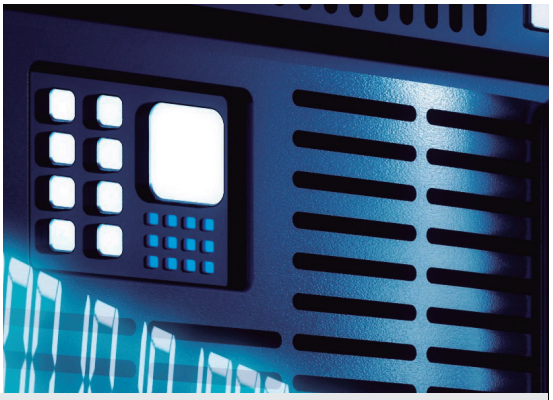



Після того, як контролер вирішує розпочати реалізацію такої обробки персональних даних, він повинен працювати над розробкою загального набору операцій обробки, які вимагаються або є необхідними для досягнення кінцевої мети обробки.

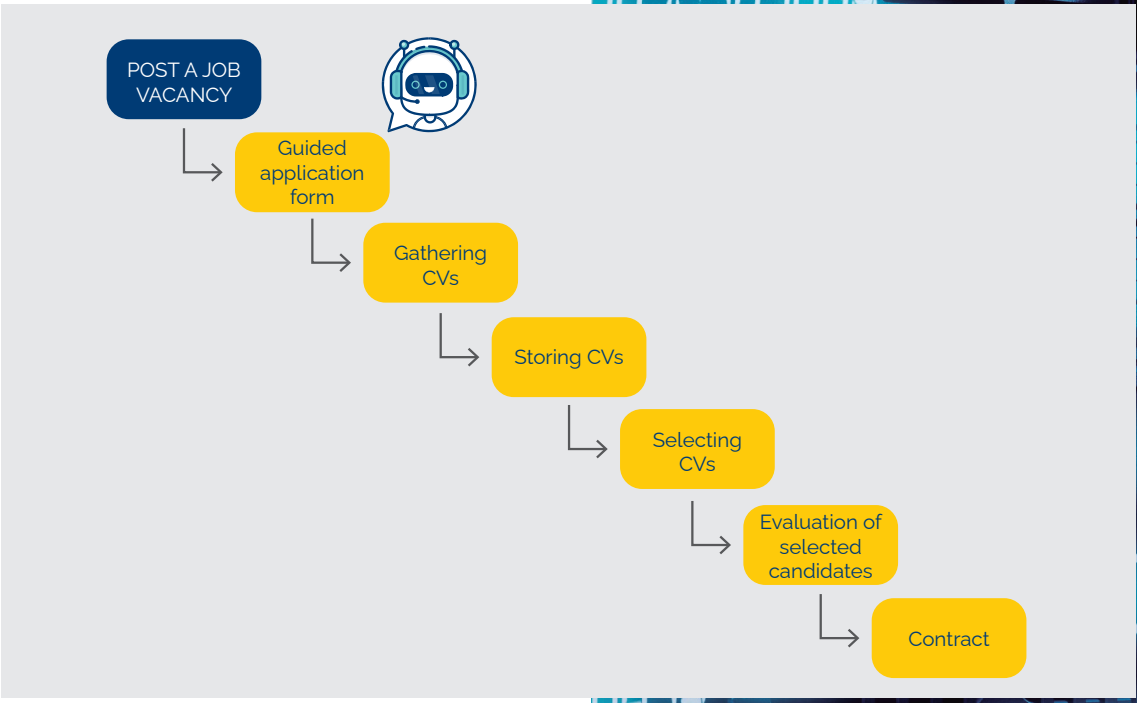
У цьому прикладі набір операцій може полягати в тому, щоб опублікувати пропозицію про роботу, мати процедуру управління пропозицією та збору резюме й заявок, фільтрувати та відбирати найцікавіші резюме, підготувати вступний тест, провести вступний тест, зробити остаточну оцінку, вибрати найбільш підходящого кандидата та укласти контракт з обраним кандидатом. Як бачите, деякі операції, такі як публікація пропозиції про роботу, швидше за все, не



передбачають обробки персональних даних, тоді як для реалізації інших така обробка необхідна.



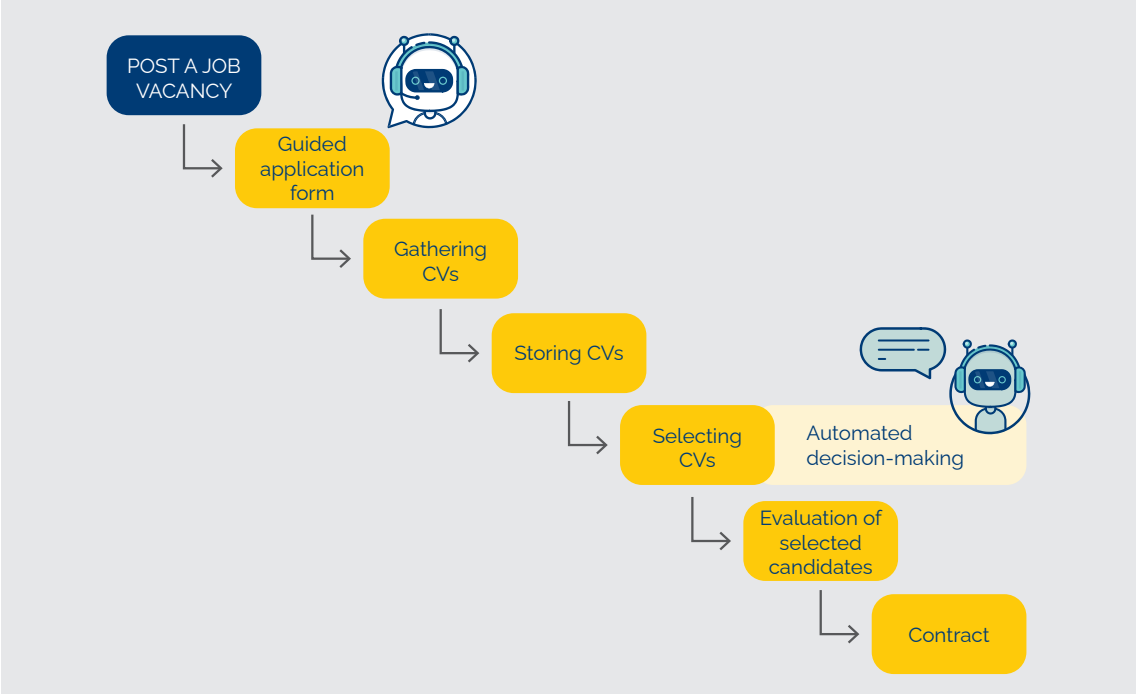
Засоби здійснення різних операцій є вибором контролера. Ці засоби можуть бути автоматизованими або ручними, з використанням власних потужностей або в хмарі, мобільними системами, шляхом аутсорсингу деяких операцій процесорам (розпорядникам) тощо. Наприклад,




процедура, яка супроводжуватиме кандидатів під час заповнення аплікаційної форми, яка включатиме їхні резюме, може бути реалізована чат-ботом, який є ШІ-системою.

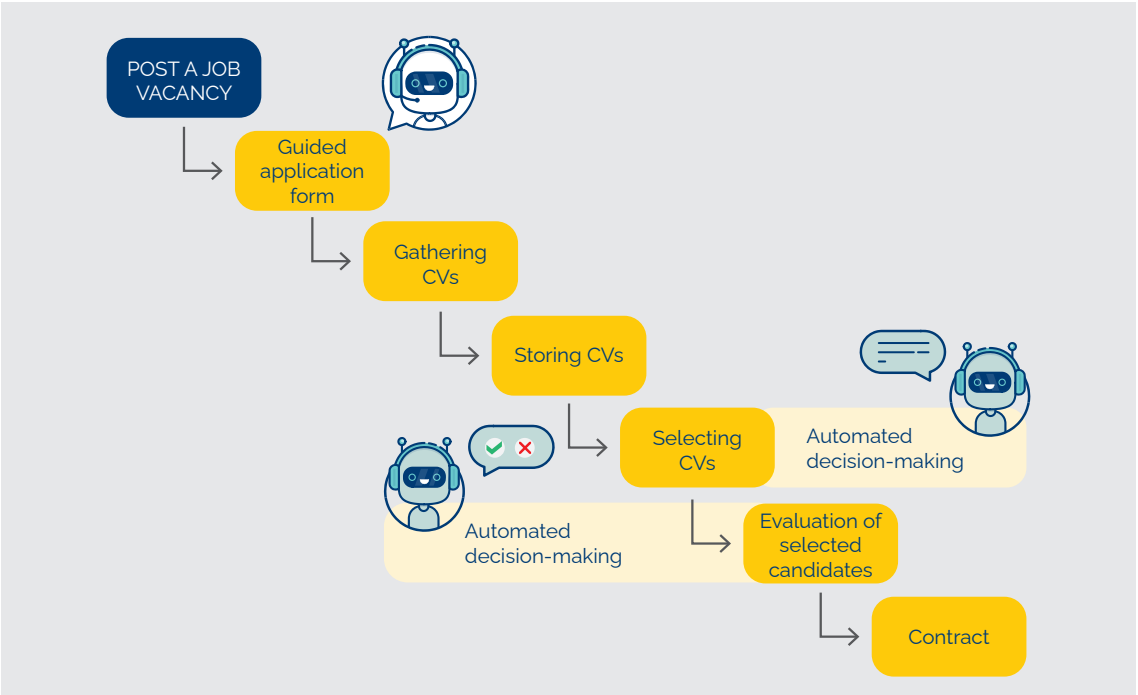
Більше того, кількість отриманих резюме може бути настільки величезною, що контролер може вирішити використовувати систему штучного інтелекту для автоматичного відбору найцікавіших резюме за певним критерієм, який також повинен встановити контролер. Контролер також приймає рішення чи буде такий відбір здійснюватися під наглядом людини і, як наслідок, чи буде це автоматизованим прийняттям рішень чи ні.

Контролер може піти далі та реалізувати оцінку відібраних кандидатів за допомогою іншої системи




штучного інтелекту, яка проводить та оцінює тести. Контролер також може вирішити, що остаточний відбір кандидатів, які будуть набрані, буде здійснюватися за результатами ШІ-системи або буде здійснюватися людський нагляд за таким відбором. Знову ж таки, контролер має повноваження застосовувати операції автоматизованого прийняття рішень чи не застосовувати.

Отже, в процесі обробки, подібної до цього прикладу, контролер може вирішити реалізувати її з використанням до трьох різних систем



штучного інтелекту. Операція фільтрації резюме за допомогою ШІ-системи сама по собі не може бути діяльністю з обробки персональних даних, оскільки така ізольована операція не може бути легітимною, якщо вона не включена в широку обробку з кінцевою та законною метою.

Підводячи підсумок: рішення щодо того, чи призведе використання ШІ-систем до автоматизованого прийняття рішень, чи ні приймає контролер. Тому факт автоматизованого прийняття рішень не є характеристикою самої ШІ-системи, а йдеться про те, чи включає контролер додаткову операцію людського нагляду за результатами, згенерованими ШІ-системою.

ШІ-системи — це засоби, які можуть бути обрані та реалізовані контролером як набір операцій у рамках конкретної діяльності з обробки персональних даних. ШІ-системи самі по собі не означають обробку персональних даних; вони є засобами для здійснення операцій з обробки даних. Діяльність з обробки може бути реалізована різними ШІ-системами одночасно. Ці ШІ-системи можуть бути реалізовані локально або в хмарі, вони можуть залучати процесорів (розпорядників) даних тощо. Вони є частиною діяльності з обробки даних, коли вони включені в деякі з необхідних визначених кроків для здійснення такої діяльності, і цей факт може призвести до виникнення конкретних ризиків для прав і свобод суб'єктів даних, які необхідно оцінювати та якими потрібно управляти.

## 3.3. ОБРОБКА ПЕРСОНАЛЬНИХ ДАНИХ ПОЗА СИСТЕМАМИ ШТУЧНОГО ІНТЕЛЕКТУ

Система штучного інтелекту може не бути частиною обробки персональних даних. ШІ-система може перебувати у стадії проєктування/розробки або еволюції, у стадії як її розповсюдження постачальником, так і функціонування в рамках обробки персональних даних.

Наприклад, система штучного інтелекту, що працює в середовищі промислового виробництва, такому як конвеєр, в якому персональні дані не використовувалися для її розробки і яка не має жодної взаємодії або рішення щодо людей, не буде залучена до обробки в матеріальній сфері GDPR.

Однак GDPR передбачає різні види діяльності з обробки, в яких може бути задіяно одну або кілька систем.

По-перше, ШІ-система може бути залучена до обробки, метою якої є проєктування та розробка автоматизованої системи. У випадку, якщо буде прийнято рішення про розробку такої автоматизованої системи з використанням технології ШІ, і така технологія ШІ потребує використання персональних даних для її розробки (наприклад, система штучного інтелекту машинного навчання або певні системи, керовані на основі правил), ми будемо припускати наявність обробки персональних даних. Слід зазначити, що при прийнятті рішення про те, як розробляти таку автоматизовану систему, можуть бути обрані інші рішення, крім ШІ, або ШІ, які потребують персональних даних.

По-друге, система штучного інтелекту може сама містити дані ідентифікованих осіб або осіб, яких можна ідентифікувати. Ця обставина не завжди зустрічається в ШІ-системі та не є унікальною для ШІ-систем. У такому випадку розповсюдження постачальником системи штучного інтелекту з такими характеристиками може включати передачу персональних даних, коли є дані осіб, яких можна ідентифікувати, та які можна витягти з ШІ-системи.

По-третє, обробка може складатися з однієї або кількох операцій з використанням однієї або

декількох систем штучного інтелекту для автоматичної обробки персональних даних, прийняття рішень щодо фізичних осіб або профілювання особи. Наприклад, коли в процесі набору персоналу ШІ-система використовується в операціях з відбору для впровадження попереднього фільтра резюме, що означатиме обробку персональних даних та автоматизоване прийняття рішень з юридичними наслідками або може суттєво вплинути на вас аналогічним чином. Це вже буде вибір відповідальної особи — залучати чи не залучати до такої діяльності з обробки даних кваліфікований людський нагляд.

Нарешті, четвертий вид обробки може відбуватися, коли деякі системи штучного інтелекту мають властивість розвиватися під час виконання обробки контролером/процесором (розпорядником) із використанням оброблюваних персональних даних. У зв'язку з можливістю еволюції ШІ-систем можуть статися такі випадки: обробка здійснюється третьою стороною для виконання власних цілей, або зазначеною третьою стороною для виконання цілей контролера/процесора (розпорядника), або контролером/процесором (розпорядником) для власних цілей.

Коротше кажучи, систему штучного інтелекту можна розглядати в рамках чотирьох груп обробки даних: проєктування/розробка, розповсюдження, експлуатація та еволюція. Кожна з них може бути під наглядом різних відповідальних сторін. Одна й та сама система штучного інтелекту може перебувати в рамках чотирьох процесів обробки, трьох, двох, одного або жодного.

# 4. ШІ ТА ПРИНЦИПИ ЗАХИСТУ ДАНИХ

## 4.1. ПРОЗОРІСТЬ

Термін «прозорість» в системах штучного інтелекту іноді використовується з іншим значенням, ніж принцип прозорості GDPR, і плутається із застосуванням принципу проактивної відповідальності або «підзвітності», як він розвивається в самому GDPR.

При розробці систем машинного навчання (МН-системи) повинна бути доступна об'єктивна та добре задокументована інформація про процеси проєктування, розробки, перевірки та валідації, щоб забезпечити правильне виконання таких завдань:

- Бути здатним дотримуватися вимог прозорості та інформування фізичних осіб, яких стосується обробка даних, що включає системи штучного інтелекту відповідно до статей 13 і 14 GDPR.
- Забезпечити ефективне здійснення прав зацікавлених сторін.
- Забезпечити виконання обов'язків контролера у зв'язку із застосуванням принципу проактивної відповідальності, як зазначалося раніше.
- Забезпечити здійснення своїх повноважень контролюючими органами відповідно до статті 58.1 GDPR, а також виконання функцій органів сертифікації та нагляду за дотриманням кодексів поведінки.

Розробник системи штучного інтелекту, який також може виконувати роль процесора (розпорядника даних) залежно від того, чи надає він розроблену ним ШІ-систему як послугу, повинен надавати учасникам кожного з чотирьох зазначених вище сценаріїв дуже різний тип та обсяг інформації. У цьому він нічим не відрізняється від будь-якого іншого постачальника технологій в технічній або науковій сфері.

Інформація, яка надається в чотирьох сценаріях, повинна бути спеціально орієнтована на навички та потреби кожного сценарію. Надання доступу до коду алгоритму фізичним особам, на яких впливає обробка даних на основі системи машинного навчання, навряд чи забезпечить інформацію, необхідну суб'єктам даних для дотримання статті 13 GDPR. І тим більше, якщо ці завдання з прозорості щодо суб'єктів даних обмежувалися би лише наданням доступу до коду.

### *4.1.1. ПРОЗОРІСТЬ GDPR*

Зокрема, прозорість для кінцевого користувача (користувачів) означає надання користувачам інформації, необхідної для прийняття ними раціонального рішення про те, як обробка впливає на них. Оскільки обробка має унікальні наслідки через включення систем штучного інтелекту в її виконання, інформація про прозорість щодо обробки повинна передавати вимір цих наслідків. Така інформація може включати сертифікати використовуваних систем машинного навчання, показники продуктивності для конкретного контексту використання та обмеження, категорію даних та важливість використовуваних даних, вплив на кінцевих користувачів та побічні ефекти, заборонене використання або невідповідні контексти, періоди зберігання, плани на випадок несправності, альтернативи в разі недоступності тощо.

Вищезазначені елементи відповідають пункту f частини 2 статті 13 GDPR стосовно зобов'язань щодо надання інформації зацікавленим сторонам. Крім того, слід пам'ятати, що дії щодо прозорості, як підтверджено Національним судом Іспанії, Адміністративно-процесуальна палата, Секція 1, у рішенні від 11 жовтня 2021 року, Rec. 1410/2019, виходять за межі того, що суворо встановлено у статтях 13 та 14, коли застосування суті принципів захисту даних ставиться під сумнів.

Застосування заходів прозорості, окрім елементів, перелічених у статтях 13 та 14 GDPR, саме по собі є заходом, що дозволяє керувати ризиком, який може виникнути для прав та свобод під час впровадження обробки, де використовуються системи.

### *4.1.2. ПРОЗОРІСТЬ І ЗРОЗУМІЛІСТЬ ДЛЯ ВОЛОДІЛЬЦЯ*

Інформацію, яка потрібна кінцевому користувачеві, не слід плутати з інформацією, яку контролер повинен знати для забезпечення виконання своїх зобов'язань щодо проактивної відповідальності під час обробки, в якій використовуються ШІ-системи як технологія, яка включена в реалізацію обробки, або як послуга, що надається третьою стороною, котра в такій мірі може бути розпорядником даних, якщо послуга передбачає обробку персональних даних. Інформацію, необхідну кінцевому користувачу, не слід плутати з інформацією, яку контролер повинен знати для виконання своїх зобов'язань з проактивної відповідальності щодо обробки з використанням системи штучного інтелекту як технології, що впроваджується у виконання обробки, або як послуги, наданої третьою стороною, котра в такій мірі може бути процесором даних, якщо це включає обробку персональних даних.

У цьому випадку відповідальній особі повинна бути надана інформація, яка дозволить їй прийняти об'єктивне рішення про те, чи існують гарантії відповідності нормативним вимогам, а також дозволити здійснювати належне управління ризиками для прав і свобод. Коротше кажучи, це також буде інформація, пов'язана із сертифікатами, показниками ефективності, конкретним контекстом використання та обмеженнями, категорією даних та вагою використовуваних даних, впливом на кінцевих користувачів та побічними ефектами, забороненим використанням або невідповідними контекстами, періодами зберігання, планами на випадок надзвичайних ситуацій, альтернативами тощо, як зазначено вище. Але контролер не є суб'єктом даних і має більше «відповідальностей» перед своїми користувачами/суб'єктами даних, тому обсяг, деталі, докази та мова, що використовуються, повинні сильно відрізнятися від обсягу, адресованого фізичній особі.

Зрештою, інформація, яка повинна надаватися контролюючим органам, органам сертифікації та органам нагляду за дотриманням кодексів поведінки, повинна бути такою, яка дозволяє їм виконувати всі свої функції інспекції та аудиту, та включає всю документацію системи машинного навчання стосовно планування, проєктування, розробки, перевірки та валідації системи. Зокрема, для того, щоб забезпечити відповідність статтям 24, 25, 32 та 35 GDPR.

### *4.1.3. ІНТЕЛЕКТУАЛЬНА ВЛАСНІСТЬ*

Виправдання захистом інтелектуальної власності запобігання здійсненню прозорості або підзвітності контролерами або діям контролюючих органів чи органів сертифікації є неприйнятним.

Права щодо інтелектуальної власності власників або розробників систем штучного інтелекту не відрізняються від прав розробників будь-якого іншого промислового або технологічного продукту, такого як компоненти безпеки та шифрування, мобільні системи, операційні системи, транспортні засоби, двигуни, літаки тощо. Усі вони однаково зобов'язані не тільки дотримуватися, але і демонструвати відповідність численним правилам. Вони повинні нести таку відповідальність як на рівні процесу розробки, так і щодо конкретного продукту, що вводиться в обіг на ринку.

Зокрема, цей обов'язок доводити дотримання GDPR встановлений в статті 24 GDPR. З цієї причини контролюючі органи та органи сертифікації, в свою чергу, пов'язані зобов'язаннями щодо дотримання таємниці.

## 4.2. МІНІМІЗАЦІЯ ДАНИХ

Реалізація принципів, прав та обов'язків GDPR, зокрема, принципу мінімізації даних вимагає від контролерів даних прийняття відповідних заходів для досягнення цієї мети[22]. Ці заходи можуть бути юридичними, організаційними, а також технічними інструментами. Що стосується останнього, GDPR стверджує у своїх початкових преамбулах, що технології повинні сприяти рішенням, які реалізують високий рівень захисту персональних даних[23]. У будь-якому випадку адекватність заходів буде встановлена відповідно до контексту, характеру, обсягу, цілей та ризиків для прав Суб'єктів даних, які беруть участь у обробці.

Принципи мінімізації та захисту даних за проєктом та за замовчуванням є суттєвими, якщо обробка пов'язана зі значними ризиками для фундаментальних прав фізичних осіб, зокрема, для реалізації принципу мінімізації даних. Беручи до уваги останні технічні розробки, всі сторони, залучені до поширення даних для цілей штучного інтелекту, повинні вжити технічних та організаційних заходів для захисту цих прав.

Такі заходи включають не тільки знеособлення, псевдонімізацію та шифрування. Ці методи не є єдиними, і в багатьох випадках можуть бути не найдоцільнішими. Все частіше використовуються й інші технології, що дозволяють вводити в дані алгоритми й отримувати цінну інформацію без необхідності передачі між сторонами або зайвого копіювання необроблених або структурованих даних[24]. Прикладами таких методів є диференціальна конфіденційність, узагальнення, приховування та рандомізація[25], використання синтетичних даних, федеративне навчання, безпечні середовища обробки та інші інструменти та технології, що підвищують конфіденційність (PET, privacy enhancing tools and technologies)[26]. Держави-члени повинні надавати підтримку органам державного сектору з метою оптимального використання цих методів і, таким чином, зробити якомога більше даних доступними для обміну[27].

### *4.2.1. ДОСТУП ДО ДАНИХ ТА ІНФОРМАЦІЇ*

Акт про управління даними (Data Governance Act, DGA) визначає «доступ» як будь-яке використання даних відповідно до конкретних технічних, юридичних або організаційних вимог без необхідності обов'язкової передачі **або завантаження даних**.[28] Іншими словами, в рамках даних доступу для розробки ШІ розрізняють дві концепції, які можуть здатися схожими, але дуже різними:

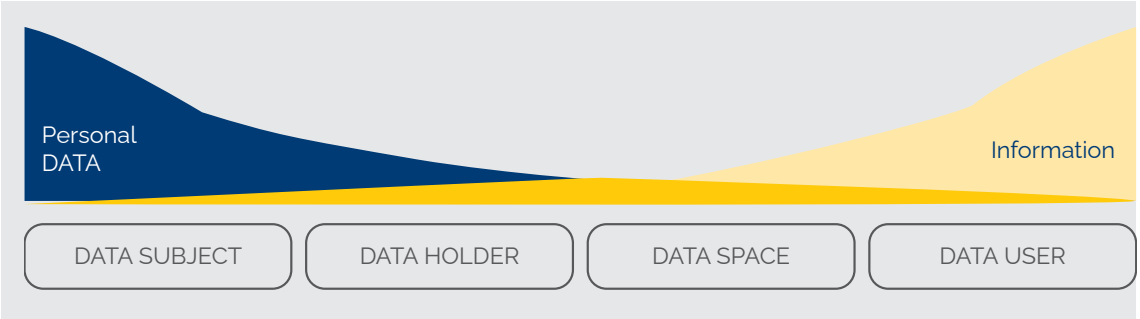
- Доступ до даних шляхом передачі або завантаження.
- Доступ до інформації, що генерується в результаті обробки даних, шляхом доступу, який не передбачає ні передачі, ні завантаження даних. Інформація – це те, що підвищує знання в даному контексті і є необхідною та доречною для досягнення цілей Користувача даних.

Дані можуть не містити необхідної інформації для даного контексту, і в будь-якому разі для отримання необхідної інформації буде необхідна попередня обробка даних. В рамках розробки ШІ для отримання необхідної інформації для кожної з операцій обробки не обов'язково повинні здійснюватися передача або поширення персональних даних. Таким чином, контролер може (і з точки зору захисту даних це є найбільш доцільним) надати доступ до персональних даних, але без поширення даних третім особам, тобто без обов'язкового передавання даних третім сторонам. Насправді, в інфраструктурі для обробки даних штучного інтелекту, як-от архітектура простору даних (Data Space), яка створена з урахуванням принципів захисту даних за проєктом, буде мінімізовано розголошення персональних даних (принцип мінімізації) та збережено можливість надавати інформацію, необхідну для досягнення цілей простору даних.

Залишаючи фактичний контроль над даними та цілями в руках Держателів даних, сприяє

22 Стаття 24 GDPR
23 Преамбула 6 GDPR
24 Преамбула 8 пропозиції ВМГО ДА
25 Преамбула 7 DGA
26 Технології, що підвищують конфіденційність
27 Преамбула 7 DGA

28 Стаття 2(13) DGA




**Малюнок 10:** Еволюція захисту даних у просторі даних, що дозволяє реалізувати мінімізацію в обробці даних для ШІ

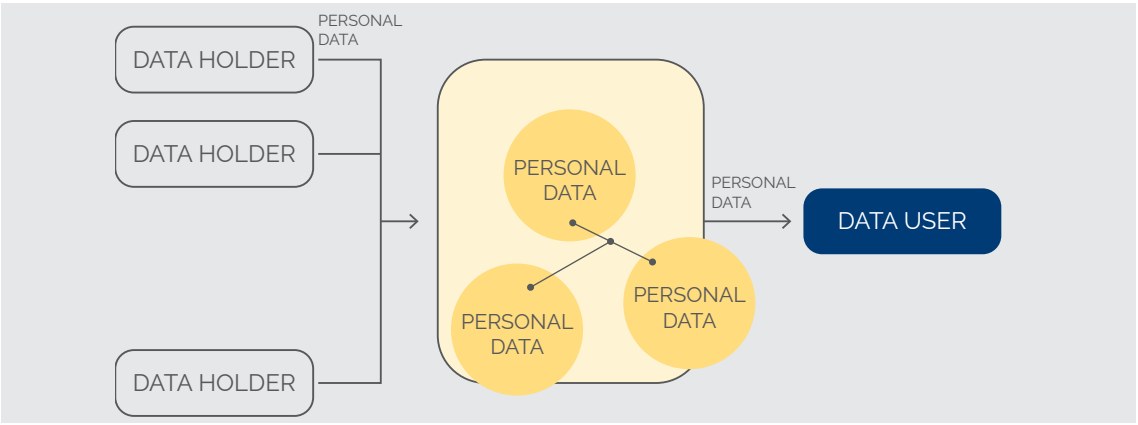
підвищенню довіру Держателів даних[29] до участі в просторі даних, а також сприятиме їхньому бажанню брати участь у розвитку цифрової економіки.

Як додаткова примітка, європейський розвиток систем, рішень та підтримка послуг, що впроваджують захист даних за проєктом, є рушієм цифрової економіки, що є загальною метою, яка обґрунтовує створення Простору даних. Крім того, якщо відсутність довіри між учасниками не дозволяє операціям обробки даних у рамках обробки даних для штучного інтелекту досягти зазначених цілей, такі операції обробки не відповідатимуть критеріям адекватності та необхідності.

### *4.2.2. АРХІТЕКТУРИ ТА ВАРІАНТИ ВИКОРИСТАННЯ ДЛЯ МАСОВОГО ДОСТУПУ ДО ДАНИХ У ШІ*

Основна та найбезпосередніша архітектура для масового доступу до даних в ШІ передбачає збір даних від декількох Держателів даних, концентрацію даних в одній точці та надання розробнику ШІ прямого доступу до даних.

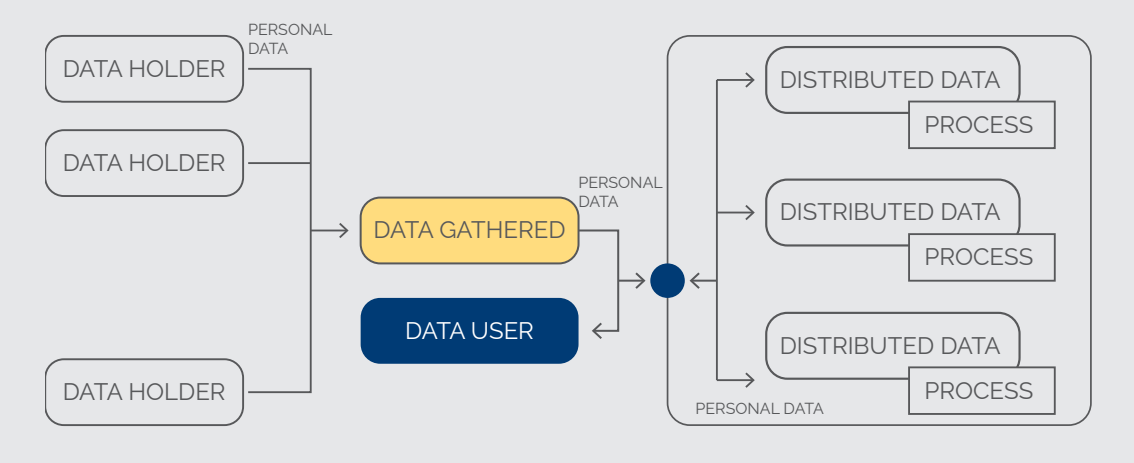
Оскільки реалізація обробки в точці концентрації даних передбачає необхідність доступу до потужних ресурсів зберігання та обробки, Користувачі даних можуть не мати таких ресурсів (наприклад, малі та середні підприємства, дослідницькі групи тощо), і у багатьох випадках не буде іншого вибору, окрім як скористатися сервісами хмарних обчислень.




**Малюнок 11:** Базова архітектурна схема доступу до даних

29 Необхідно враховувати небажання суб'єктів господарювання обмінюватися інформацією, яка може завдати шкоди їхнім комерційним інтересам або розкрити їхню бізнес-стратегію, окрім інших нормативно-правових актів щодо захисту інтелектуальної та промислової власності.

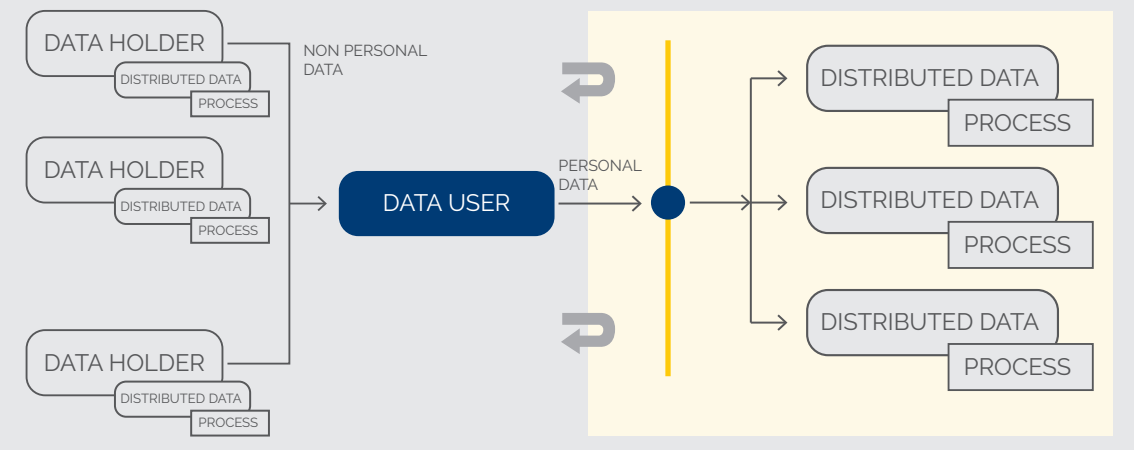
Ці архітектури передбачатимуть перекладання питань дотримання нормативних вимог, збереження принципів прав і свобод та інших супутніх наслідків на тих Користувачів даних, які




**Малюнок 12:** Схема базової архітектури масової обробки даних із використанням гіпермасштабування

мають менше ресурсів. Це може ускладнити досягнення цілей обробки через відсутність довіри Суб'єктів даних, небажання ділитися даними самими Держателями даних для захисту інтересів або комерційної таємниці тощо.

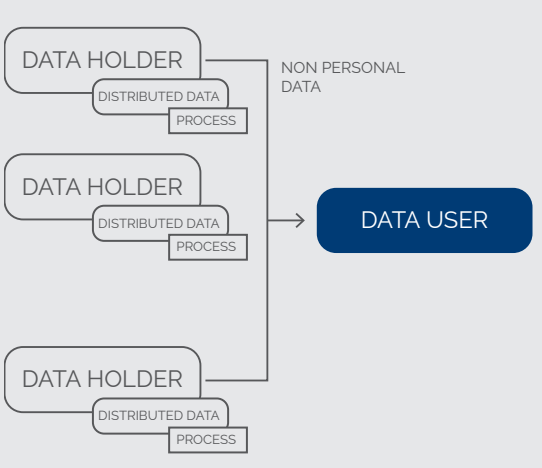
Існують різні підходи до Простору даних, які могли би вирішити ці проблеми заздалегідь. Одне з цих можливих рішень може бути досягнуто шляхом використання тих самих технологій, що використовуються в гіпермасштабах для розподіленої обробки, але в цьому випадку — для реалізації методів **Compute-to-data (обчислення до даних)**[30], тобто розподілена обробка даних буде виконуватися за місцем знаходження даних. Це дозволить уникнути передачі даних третім особам, обробка буде виконуватися за місцем знаходження даних, а розкриття персональних даних у мережах зв'язку і накопичення даних у великих сховищах буде зменшено.




**Малюнок 13:** Базова архітектурна схема з використанням стратегії обробки даних на місці (compute-to-data)

30 Як випливає з визначень, це означає, що обробка здійснюється за місцем знаходження даних, а не передається у «хмару», де здійснюється обробка. Це також пов'язано зі стратегіями Edge-Computing, які є однією з характеристик 5G, і які передбачають наближення обробки даних до власних систем кінцевих користувачів, з тією перевагою, що мережевий трафік розвантажений, а постачальникам послуг потрібно менше серверів, оскільки вони використовують обчислювальні ресурси терміналів користувачів.

Як і будь-яка стратегія захисту даних за проєктом, зокрема така, як знеособлення або диференціальна конфіденційність, це не буде єдиним рішенням для забезпечення відповідності нормативним вимогам та цілям доступу до великої кількості даних у випадку обробки персональних даних. Це також не є найкращим рішенням для всіх можливих випадків використання, а також



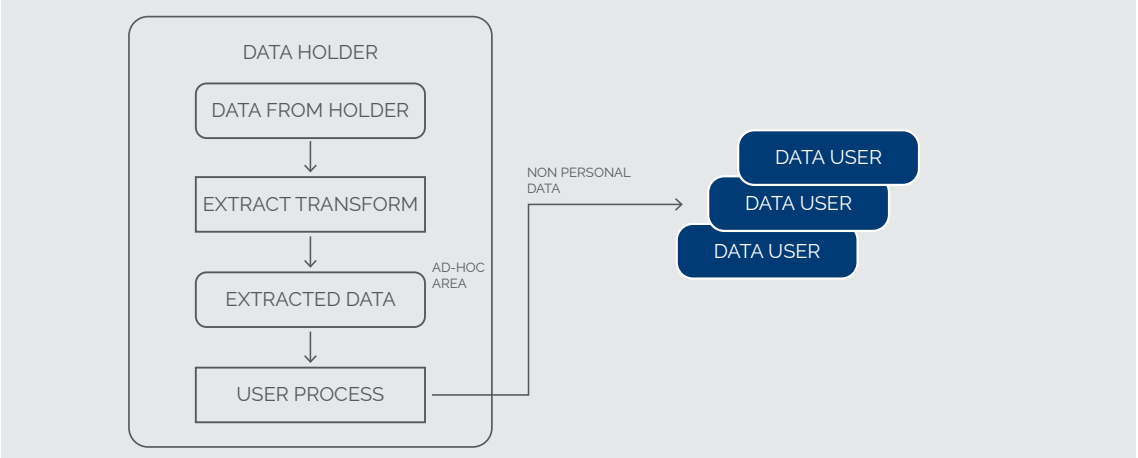

**Малюнок 14:** Діаграма архітектури з використанням стратегій обробки даних на місці (compute-to-data)

підходить не для всіх сценаріїв обробки даних. Завжди слід розглядати різні стратегії залежно від конкретного випадку використання, і дизайн стратегії доступу має передбачати їх реалізацію. Щоб це стало можливим, необхідно враховувати захист даних із самого початку при плануванні інфраструктур доступу до даних, таких як Data Spaces, а також при плануванні будь-якого проєкту масового доступу до даних, як, наприклад, проєкту на основі машинного навчання.

### *4.2.3. СТРАТЕГІЇ ОБРОБКИ ДАНИХ НА МІСЦІ (COMPUTE-TO-DATA) ТА ФЕДЕРАТИВНЕ НАВЧАННЯ*

Приклад того, що таке стратегії обробки даних на місці (compute-to-data), було наведено на початку цього розділу. Ці стратегії передбачають перенесення обробки даних до їхнього першоджерела, у цьому випадку до Держателя даних, із метою вилучення (вже не персональної) інформації від Держателя даних. Наприклад, стратегії обробки даних на місці можуть бути використані для реалізації федеративного навчання в тренуванні штучного інтелекту на основі машинного навчання.

Впровадження цих стратегій передбачає, що стратегія обміну даними визначає зобов'язання щодо управління та менеджменту інформації[31] в розподіленому середовищі. У деяких випадках




**Малюнок 15:** Схема спеціальних просторів в Держателя даних для забезпечення інфраструктури обробки даних на місці (compute-to-data)

31 Примітка редактора: В оригінальному тексті використовуються терміни «Governance» та «Management». Governance означає структуру правил, практик та процесів, за допомогою яких організація керується і контролюється. Воно зосереджене на забезпеченні нагляду, стратегічного напряму і підзвітності. Management, навпаки, охоплює повсякденну діяльність і операції, необхідні для досягнення цілей організації, та зосереджено на реалізації стратегій і політик, встановлених у межах governance.

процеси, які Користувач даних має намір запускати на об'єктах Держателів даних, повинні бути перевірені або сертифіковані, перш ніж бути розповсюджені останніми.

У цих випадках Держателі даних повинні мати спеціальні простори у своїх інфраструктурах з повним відокремленням від своїх операційних систем (подібно до зони DMZ[32]).

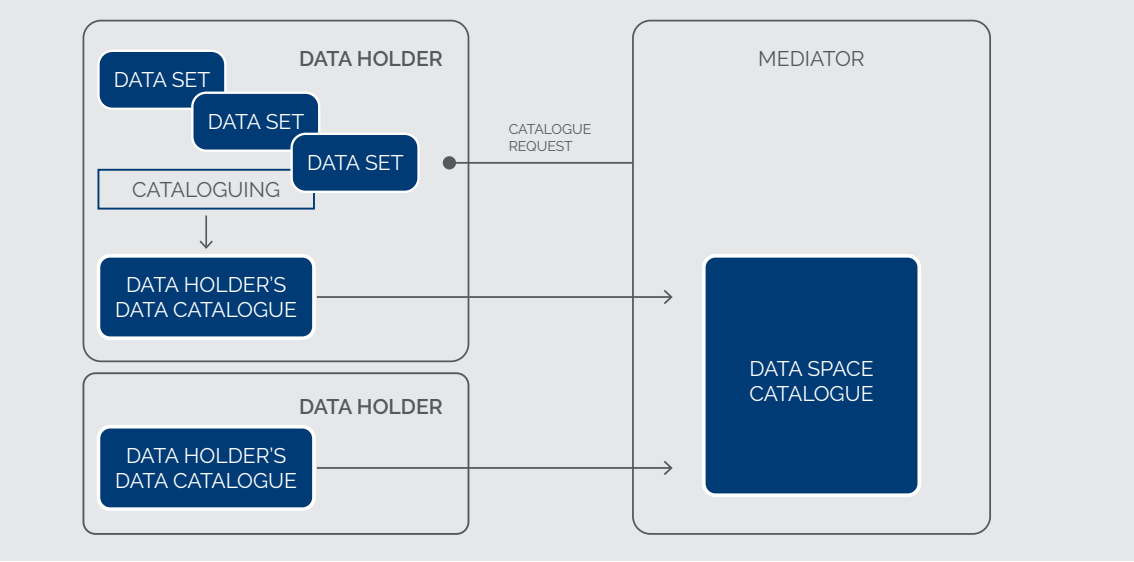
Цей варіант використання може бути розроблений у багатьох варіантах залежно від конкретної обробки.

Забезпечення та демонстрації відповідності законодавству про захист даних у цьому випадку використання може бути реалізовано через юридичні, організаційні та технічні заходи. Наприклад, юридичні заходи можуть включати зобов'язання щодо попереднього аудиту або сертифікації процесів Користувачів даних. Організаційні заходи та політики захисту даних можуть включати, серед іншого, нагляд людини за завантаженням та виконанням процесів у Держателя даних, невиконання процесів Користувача в даних та операційних системах Держателя даних, витягання робочого набору даних, який не містить усього набору даних, оцінку того, що процес Користувача не генерує та не передає персональні дані, або оцінку отриманих результатів.

Технічні заходи можуть включати, наприклад, створення спеціальних областей для виконання процесів із фізичною ізоляцією від систем Держателя даних, спеціальних областей, що утворюють безпечні середовища обробки тощо.

### *4.2.4. ВИПАДОК ОБРОБКИ НА МІСЦІ (COMPUTE-TO-DATA) : КАТАЛОГІЗАЦІЯ*

Каталогізація — це обробка даних або набору даних, яка дозволяє асоціювати з даними метадані, необхідні для подальшого використання. Це життєво важлива діяльність, коли ми маємо справу із середовищами машинного навчання. Метадані принаймні включатимуть описи типів даних і місця розташування даних, але можуть поширюватися на оцінку якості даних, що передбачає ретельну обробку даних або навіть визначення того, чи існують персональні дані в наборі даних. Таким чином, можуть бути створені каталоги ресурсів (даних), які можуть бути доступні декільком учасникам віртуальним, проміжним або реальним способом.




**Малюнок 16:** Схема архітектури каталогізації

32 DMZ стосується демілітаризованої зони або безпечної зони, не пов'язаної з операційними системами суб'єкта господарювання.

Каталогізація наборів даних — перше завдання, яке потрібно виконати в рамках машинного навчання. Каталогізація може здійснюватися різними способами: шляхом доступу Медіаторів та/або Користувачів до систем Держателів, шляхом передачі даних Медіаторам та/або Користувачам, або процесом у самих Держателів як окремий випадок обробки на місці (compute-to-data).

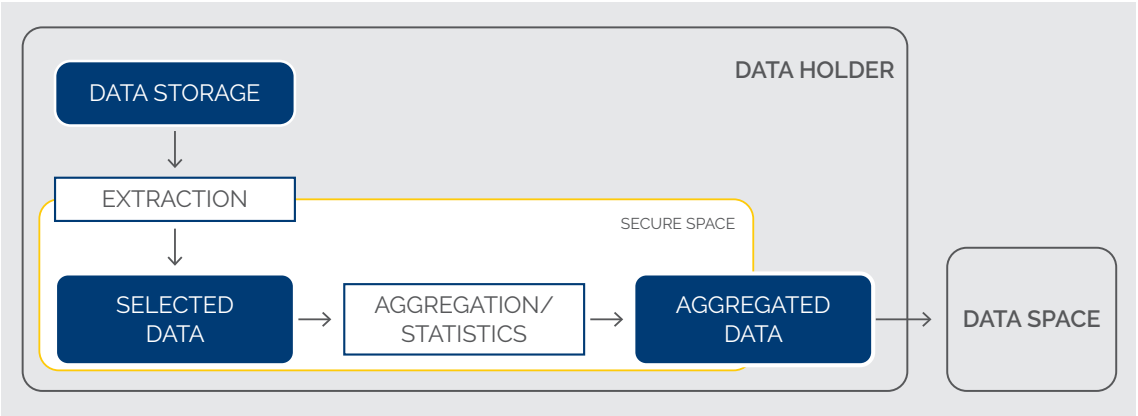
В останньому випадку каталогізація не потребує передачі наборів даних, що знаходяться у Держателів даних, третім сторонам, а лише передача метаданих, отриманих в результаті аналізу.

### *4.2.5. АНОНІМІЗАЦІЯ: ОБРОБКА, ЯКА ВИМАГАЄ ЗНЕОСОБЛЕННЯ АГРЕГОВАНИХ ДАНИХ ВЛАСНИКІВ ДАНИХ З ДИСОЦІАЦІЄЮ ДАНИХ ВІД РІЗНИХ ВЛАСНИКІВ ДАНИХ*

Заходи, які можуть бути застосовані для забезпечення та демонстрації відповідності законодавству про захист даних, будуть аналогічними до стратегій обробки даних на місці (compute-to-data). Процес каталогування та генерації метаданих також може бути частиною впровадження технік захисту даних за проєктом, таких як використання тегування та ієрархій тегів для управління привілеями доступу. знеособлення: Обробка, яка потребує знеособлених агрегованих даних держателів даних з дисоціацією даних від різних держателів даних.

У цьому випадку Медіатори або Користувачі даних за власною ініціативою або у відповідь на запит Користувача вимагають знеособлену або неперсональну інформацію від Держателів даних. При цьому важливо, щоб витягнута інформація не потребувала об'єднання даних про одного і того ж суб'єкта, які зберігаються у різних Держателів даних.

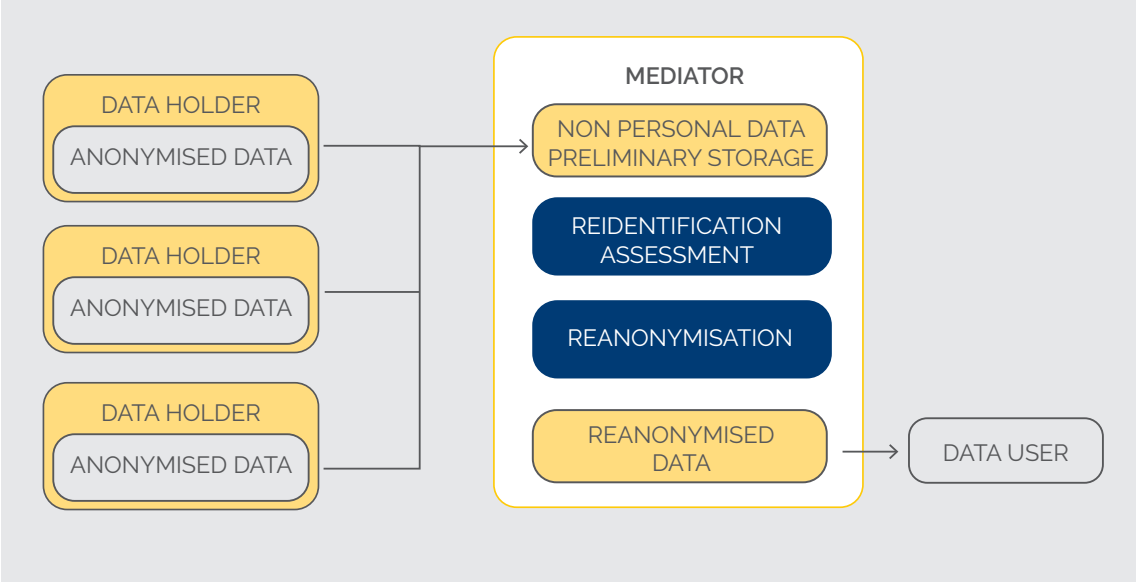
Прикладом є дослідження мобільності на основі даних геолокації від операторів телекомунікацій.




**Малюнок 17:** Схема архітектури для випадку використання анонімних даних без зв'язку між даними від різних Держателів даних

Зазвичай користувач пов'язаний з одним оператором телекомунікацій, і випадки, коли профіль мобільності користувача залежить від інформації від двох операторів, є рідкісними та незначними для досягнення мети обробки.

У цьому випадку може бути розроблена така архітектура:




**Малюнок 18:** Схема архітектури для випадку консолідації знеособлених даних від різних Держателів даних

Для забезпечення та демонстрації відповідності вимогам законодавства про захист даних у цьому випадку використання можуть бути впроваджені, наприклад, юридичні заходи, такі як проведення аналізу ризиків для перевірки неможливості повторної ідентифікації.

Можливі організаційні заходи та політики захисту даних можуть включати, зокрема, такі кроки:

- Виконання процесу агрегації або інших видів аналізу даних не на основних робочих даних, а на вже вилучених даних, дотримуючись принципу мінімізації.
- Проведення цих процесів у спеціально призначених середовищах.

Серед технічних заходів, які можуть бути застосовані, може бути оцінка неможливості повторної ідентифікації в отриманих даних.

### *4.2.6. ЗНЕОСОБЛЕННЯ: ОБРОБКА, ЯКА ПЕРЕДБАЧАЄ КОНСОЛІДАЦІЮ АНОНІМНИХ ДАНИХ ВІД РІЗНИХ ДЕРЖАТЕЛІВ ДАНИХ*

Цей випадок використання може відбуватися в поєднанні з іншими випадками використання, показаними в цьому посібнику, зокрема з випадком використання обробки неперсональних даних.

У просторі даних має бути враховано, що чим більший обсяг отриманих даних, то вищі шанси на ідентифікацію, незважаючи на те, що дані є неперсональними даними, особливо коли дані отримані з різних джерел. Тому серед функцій Медіаторів Простору даних особливо важливо провести первинну перевірку якості знеособлення набору даних, перш ніж зробити його доступним для Користувача даних.

У випадках, коли виявлена можливість повторної ідентифікації, механізми знеособлення доведеться знову застосовувати в Просторі даних у безпечному середовищі, де будуть видалені будь-які ключі повторної ідентифікації, а також всі дані з цього першого огляду.

Для забезпечення та демонстрації відповідності правилам захисту даних у цьому випадку використання можуть бути застосовані, наприклад, юридичні заходи, зокрема запобіжні заходи для обмеження розповсюдження знеособлених даних, доки вони не будуть належним чином оцінені на ефективність знеособлення або повторно знеособлені, а також обмеження поширення або зберігання знеособлених даних за допомогою юридичних угод, що, зокрема, виходять за рамки положень GDPR.

Організаційні заходи та політики захисту даних можуть включати, серед іншого, видалення наборів даних, які призводять до повторної ідентифікації, та інформування Держателів даних про цю можливу вразливість.

Технічні заходи можуть включати впровадження контрольованого та безпечного середовища для тимчасового зберігання знеособлених даних із різних джерел, які будуть об'єднані та згодом знеособлені.

### *4.2.7. ЗНЕОСОБЛЕННЯ: ГЕНЕРУВАННЯ ТА ВИКОРИСТАННЯ СИНТЕТИЧНИХ ДАНИХ*

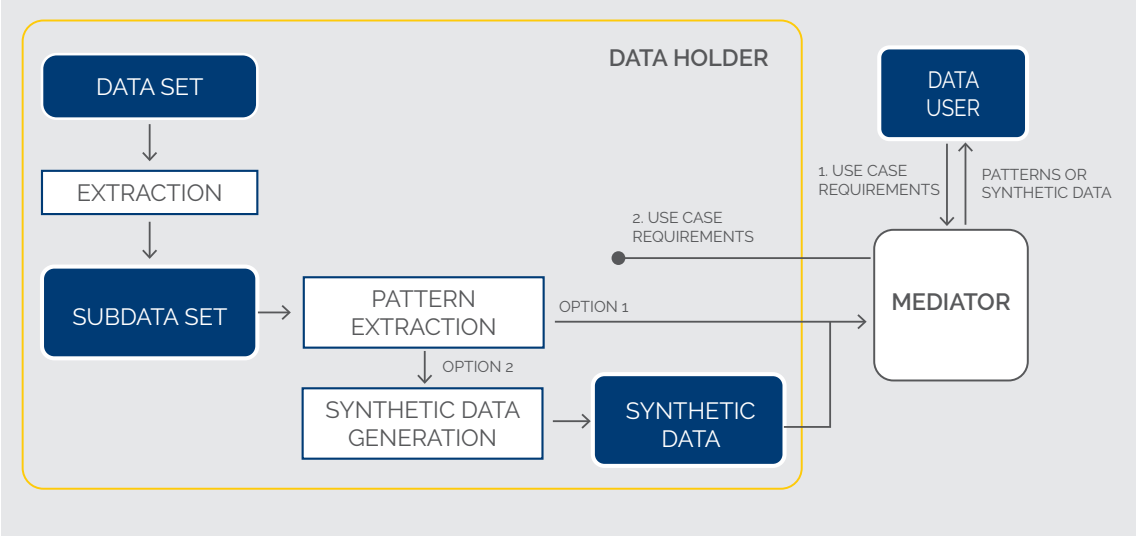
Іншою стратегією мінімізації даних є використання синтетичних даних. Синтетичні дані — це не випадкові дані, а такі, що відповідають тим самим вимогам, що і реальні дані для конкретної мети. Вимоги будуть залежати від конкретного варіанту використання: певний статистичний розподіл, відповідність певному типу шаблонів тощо. Ці шаблони повинні бути вилучені з персональних даних шляхом обробки таких персональних даних та створення неперсональної інформації[33]. Щойно починають використовуватися персональні дані, процес генерування синтетичних даних є або буде частиною операції з обробки, що підлягає відповідності GDPR.

Держатель даних може обробляти персональні дані у спеціальному середовищі для аналізу даних, потім генерувати шаблони, отримані з його збережених даних, і, таким чином, це дозволяє самому Держателю даних створювати синтетичні дані або просто випускати шаблони для екосистеми Простору даних (можливо, самого

33 Прикладом такого варіанту використання є пілотний проєкт для обміну даними, який здійснюється на європейському рівні з даними центральних банків кожної країни, де перед тим як зробити дані доступними, генерується синтетична база даних, яка має ті ж характеристики, що і оригінал. Це проєкт, очолюваний Генеральним директоратом Європейської Комісії з питань фінансової стабільності, фінансових послуг та ринків капіталу в рамках проєкту зі створення Центру даних у Платформі цифрових фінансів ЄС.

Користувача даних або Фасилітатора (Enabler)) для створення синтетичних наборів даних.

Для цього випадку пропонується така архітектура:




**Малюнок 19:** Схема архітектури для випадку використання синтетичних даних

Для забезпечення та демонстрації відповідності правилам захисту даних у цьому випадку можуть бути впроваджені юридичні, організаційні та технічні заходи. Наприклад, юридичні заходи можуть включати зобов'язання створити синтетичний набір даних у Держателя даних або вимогу проведення аудитів чи сертифікації інструментів генерування синтетичних даних.

Організаційні заходи та заходи політики захисту даних можуть включати процес аналізу даних у безпечному середовищі, зокрема вилучення підмножини даних з операційних систем.

Технічні заходи можуть включати, наприклад, оцінку анонімності синтетичних результатів.

### *4.2.8. ЗНЕОСОБЛЕННЯ: БЕЗПЕЧНЕ БАГАТОСТОРОННЄ ОБЧИСЛЕННЯ*

Безпечне багатостороннє обчислення[34] (Secure Multiparty Computation або SMPC) — це криптографічний протокол, який за допомогою технології адитивного секретного розподілу (Additive Secret Sharing) дозволяє сегментувати секретні дані на різні частини так, що при обміні інформацією вихідні дані не можуть бути розкриті жодним із джерел. У протоколі бажаний результат отримується без необхідності розкривати будь-які конфіденційні дані, а отриманий результат не зазнає жодних відхилень.

Ця стратегія корисна в певних сценаріях і для реалізації потребує технологічної підтримки.

### *4.2.9. ЗНЕОСОБЛЕННЯ: ДИФЕРЕНЦІЙОВАНА КОНФІДЕНЦІЙНІСТЬ*

Диференціальна конфіденційність[35] гарантує, шляхом додавання випадкового шуму до вихідної інформації, що в результаті процесу аналізу даних, до яких застосовано цей метод, не буде втрати корисності отриманих результатів. Цей процес заснований на законі великих чисел

34 Стаття блогу AEPD під назвою «Конфіденційність за задумом: безпечні багатокомпонентні обчислення: адитивний обмін секретами | AEPD [травень 2022]»

35 Стаття блогу AEPD під назвою «Анонімізація та псевдонімізація (II): Диференційована конфіденційність | AEPD [жовтень 2021]»

— статистичному принципі, який говорить, що при зростанні розміру вибірки середні значення, отримані з нього, наближаються до реального середнього значення інформації. Таким чином, додавання випадкового шуму до всіх даних компенсує ці ефекти та створює «фактично еквівалентне» значення.

Один із прикладів використання можна знайти в Бюро перепису населення США[36], яке застосовує диференційовану конфіденційність для забезпечення точності своєї статистики та запобігання розкриттю особистої інформації навіть через статистику і таким чином підвищує довіру громадян та безпеку даних, які вони надають.

### *4.2.10. ЗНЕОСОБЛЕННЯ: ДОКУМЕНТИ СПРЯМОВАНІ НА ЗНЕОСОБЛЕННЯ*

У пункті 9 Преамбули Регламенту про управління даними (DGA) зазначається, що у випадку повторного використання даних необхідно забезпечити, щоб знеособлення було вбудоване в саму концепцію даних, а формати даних дозволяли ефективне знеособлення «за задумом». Зазначено: «Для полегшення захисту персональних і конфіденційних даних та прискорення процесу надання таких даних для повторного використання відповідно до цього Регламенту держави-члени повинні заохочувати органи державного сектору створювати та надавати дані відповідно до принципу «відкритості за задумом і за замовчуванням», як зазначено в статті 5(2) Директиви (ЄС) 2019/1024, і сприяти створенню та використанню даних у форматах і структурах, які полегшують знеособлення.»

### *4.2.11. ІНШІ МЕТОДИ ЗАХИСТУ ДАНИХ*

Не прагнучи бути вичерпними, скажемо лише, що існують й інші методи, що використовуються для захисту даних під час обміну даними. Наприклад, гомоморфне шифрування, відновлення приватної інформації або федеративні методи навчання в машинному навчанні. Нижче наведено короткий огляд кожної з цих методик.

Гомоморфне шифрування[37] — це технологія конфіденційності за замовчуванням, яка підходить для випадків, коли контролер передає частину діяльності процесору на аутсорсинг і хоче технічно гарантувати, що процесор не матиме доступу до даних.

У традиційній схемі контролер даних для захисту конфіденційності під час передачі передає інформацію процесору в зашифрованому вигляді. Щойно процесор отримує інформацію, вона розшифровується та обробляється. Однак ця схема створює як юридичні, так і технічні ризики, тому в ідеалі, щоб мінімізувати ризики, процесор не повинен мати можливості розшифрувати інформацію, і вся обробка повинна здійснюватися на даних, зашифрованих контролером даних. Це завадить недобропорядному процесору або третій стороні отримати доступ до даних і використовувати їх для різних цілей. Одним із способів досягнення цього захисту є так зване гомоморфне шифрування.

Гомоморфне шифрування, таким чином, дозволяє виконувати операції з зашифрованими даними і отримувати результати, також зашифровані, еквівалентні операціям, що виконуються безпосередньо над вихідною інформацією.

З іншого боку, відновлення приватної інформації (Private Information Retrieval, PIR) [38] — це криптографічний метод, який дозволяє користувачу отримати запис із бази даних, не розкриваючи зберігачу даних (data custodian) запис, який було отримано, і не пов'язуючи інформацію, з якої можна було б зробити висновок про те, хто виконує доступ.[39]

Це можна застосувати, наприклад, у компанії, яка хоче, щоб її клієнти мали доступ до бази даних. У середовищі за замовчуванням кожного разу, коли клієнт отримує доступ до бази даних, зберігач даних (data custodian) знає, до якого

36 Диференційована конфіденційність та перепис населення 2020 року (census.gov)

37 Стаття блогу AEPD під назвою «Шифрування та конфіденційність III: гомоморфне шифрування [червень 2020 р.]»

38 ІНЖЕНЕРІЯ ЗАХИСТУ ДАНИХ, Від теорії до практики. Агентство Європейського Союзу з кібербезпеки (ENISA) [січень 2022]

39 Наприклад, у разі медичного, фінансового або поліцейського розслідування Утримувач даних або Медіатор простору даних не буде проінформований про те, що дані певної особи перевіряються.

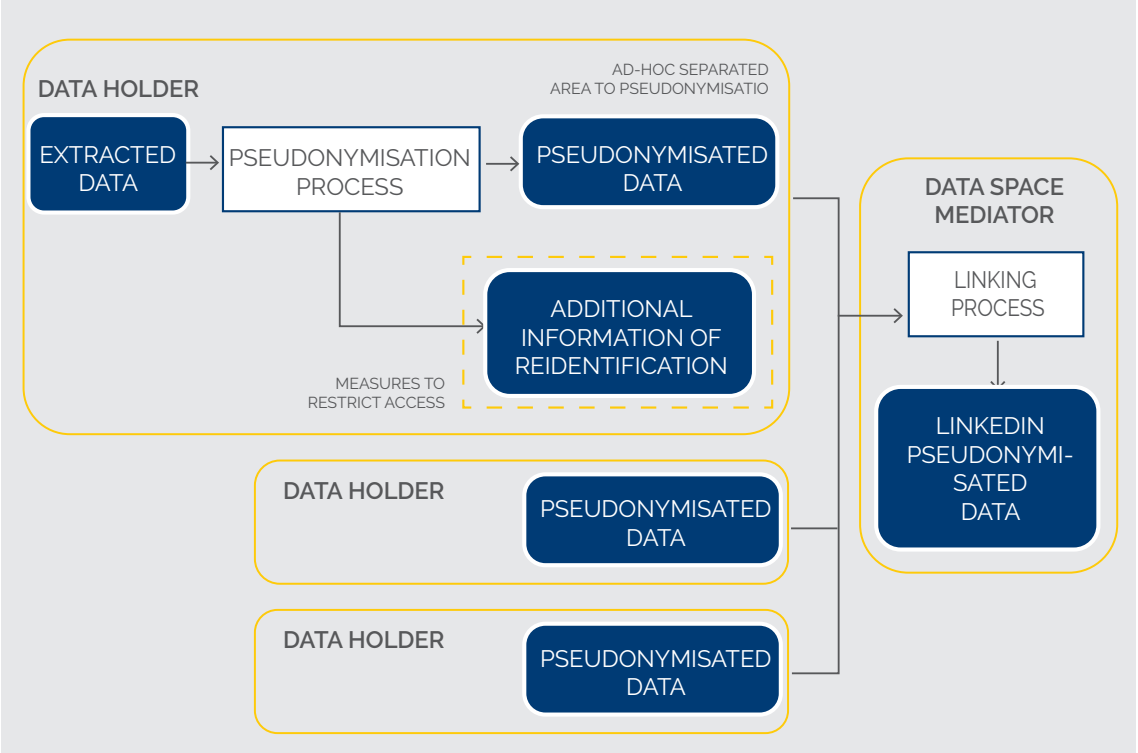
запису було отримано доступ. Згодом контролер зможе знати, які записи в базі даних цікавлять клієнтів.

Врешті, ми також можемо виділити федеративні методи[40] навчання, як горизонтальні, так і вертикальні, для додатків штучного інтелекту на основі машинного навчання. Федеративні методи навчання є різновидом технологій покращення конфіденційності (PET / Privacy-Enhancing Technology), які дозволяють розробляти системи машинного навчання без необхідності передавати персональні дані між учасниками. Ці методи можуть бути як горизонтальними, так і вертикальними, і є ключовими в нових сценаріях поліпшення та розвитку суспільства, таких як простори даних.

Федеративне навчання дозволяє створювати моделі машинного навчання, де замість централізації даних у великому сховищі для аналізу, моделі надсилаються до місця, де знаходяться дані. Ця стратегія типу «compute-to-data», дозволяє здійснювати локальну обробку даних, а згодом агрегувати результати локальних моделей і консолідувати інформацію, отриману з навчання, у повну модель. Таким чином, це дозволяє створювати федеративні простори даних, в яких кожен учасник зберігає контроль, суверенітет і захист даних, обираючи в кожен момент, хто може використовувати дані і для якої конкретної мети.

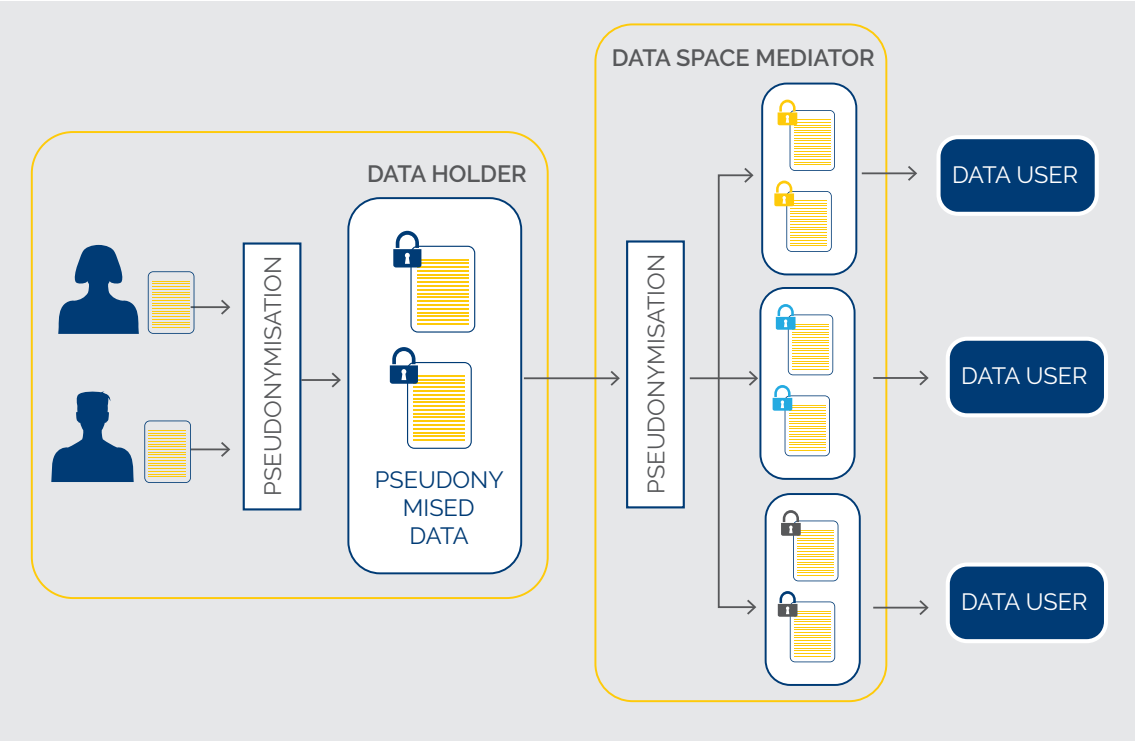
### 4.2.12. ПСЕВДОНІМІЗАЦІЯ ДАНИХ

Псевдонімізація здійснюється через набір операцій у межах обробки даних (у деяких дуже специфічних випадках це може бути окрема операція обробки) і призначена як захід безпеки, коли неможливо досягти цілей обробки через знеособлення. Однією з таких цілей може бути необхідність пов'язати дані одного і того ж суб'єкта між різними Держателями даних, інша




**Малюнок 20:** Схема архітектури для випадку використання псевдонімізації

40 Стаття блогу AEPD під назвою «Федеративне навчання: штучний інтелект без шкоди для конфіденційності | AEPD [квітень 2023]»




**Малюнок 21:** Схема архітектури псевдонімізації одного набору даних для різних користувачів даних

– коли дані отримуються не пакетним режимом, а на постійній основі (наприклад, отримуються з мобільних пристроїв або IoT), і ще одна – коли суб'єкта даних потрібно інформувати конкретно про результат обробки їхніх даних (наприклад, клінічні дослідження), і, таким чином, необхідна періодична та вибіркова повторна ідентифікація для забезпечення життєво важливих інтересів суб'єктів даних.

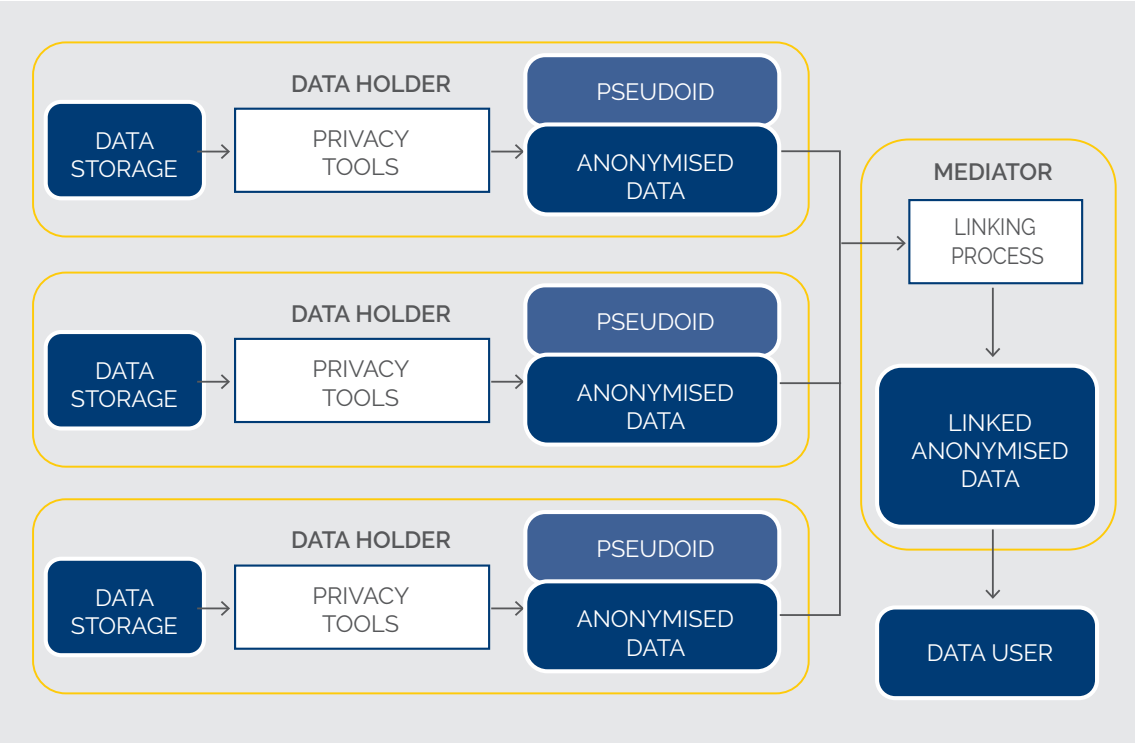
Крім того, в певних секторах, таких як клінічні дослідження, існують конкретні вимоги та правила[41], тому те, що розробляється тут, буде в руслі цих правил. У деяких секторах регулюється роль Фасилітатора як довіреної особи, яка здійснює процес псевдонімізації та відповідає за зберігання додаткової інформації про повторну ідентифікацію, наприклад монітор у випадку клінічних досліджень.

В процесі псевдонімізації також можуть використовуватися технології поліморфного шифрування та псевдонімізації (Polymorphic Encryption and Pseudonymisation techniques - PEP)[42]. Кожній особі присвоюються різні псевдоніми для кожного Користувача даних, який запитує доступ до даних Суб'єкта даних, таким чином уникаючи пов'язування псевдонімів між кількома третіми сторонами.

У конкретному випадку просторів даних про здоров'я (Health Data Spaces) кожен пацієнт може мати унікальний ідентифікатор. Медіатор (Mediator) може трансформувати цей ідентифікатор у різні псевдоніми залежно від одержувача та контексту або мети обміну даними. Кожен псевдонім повідомляється кожному одержувачу разом із поліморфними зашифрованими даними. Оскільки для кожного одержувача генерується новий псевдонім, псевдоніми, що використовуються для одного і того ж пацієнта, не можуть бути пов'язані і тому вважаються непов'язаними та зберігають конфіденційність даних пацієнта.

41 Додаткове положення 17 LOPDGDD, або Кодекс поведінки, що регулює обробку персональних даних у сфері клінічних випробувань та інших клінічних досліджень і фармаконагляду.

42 Розділ 2.2.1 документа «ІНЖЕНЕРНИЙ ОБМІН ПЕРСОНАЛЬНИМИ ДАНИМИ, нові варіанти використання та технології. Агентство Європейського Союзу з кібербезпеки (ENISA). [січень 2023]».




**Малюнок 22:** Схема архітектури для випадку використання анонімних даних із зв'язком між даними різних власників даних

### *4.2.13. ОБРОБКА, ЩО ВИМАГАЄ ЗНЕОСОБЛЕНИХ ДАНИХ, ДЕ ВАЖЛИВО ПОВ'ЯЗАТИ ПЕРСОНАЛЬНУ ІНФОРМАЦІЮ, ОБРОБЛЕНУ РІЗНИМИ ДЕРЖАТЕЛЯМИ ДАНИХ*

Це випадок використання, який може виникнути, коли такі стратегії, як безпечні багатосторонні обчислення або диференціальна конфіденційність, не можуть бути застосовані. Приклад може виникнути, коли ви хочете проаналізувати продукти або послуги, які один і той же Суб'єкт даних використовує у різних Держателів даних, для чого необхідно спочатку пов'язати їх, але які в кінцевому підсумку будуть відображатися анонімно.

У цьому випадку можна провести попередній процес заміни ідентифікаторів та псевдоідентифікаторів новими псевдоідентифікаторами, не пов'язаними з персональними даними. Це повинно бути зроблено за допомогою механізму, попередньо узгодженого всіма Держателями даних в рамках управління простором даних, щоб при передачі записів у простір даних можна було пов'язати ті, що відповідають одному і тому ж користувачеві.

Після отримання в просторі даних буде створена консолідація анонімних даних, а інформація, яка використовується для пов'язування записів буде відкинута.

Цей підхід може вимагати використання спеціально створених просторів для впровадження процесів вилучення та знеособлення у Держателів даних, а також аналізу повторної ідентифікації у Просторі даних консолідованого набору даних.

### *4.2.14. ОБРОБКА ЗА УМОВИ НЕМОЖЛИВОСТІ АНОНІМІЗАЦІЇ ДАНИХ*

Засоби захисту, які можуть бути застосовані, можуть бути похідними від тих, що вже згадані для обробки персональних, псевдонімізованих та знеособлених даних. Обробка за умови неможливості знеособлення даних

Деякі операції обробки за своєю природою не можуть досягти адекватного рівня якості із знеособленими даними. Цю обставину слід належним чином оцінити в DPIA (Оцінка впливу на захист даних). Контролери даних повинні провести аналіз відповідності, необхідності та суворої пропорційності даних.

В рамках оцінки ризиків, зокрема оцінки пропорційності обробки, існує три варіанти, розташовані в порядку найменшого та найвищого ризику для здійснення обробки:

1. Передача бажаної обробки від Користувача даних до Держателів даних (наприклад, використовуючи методи федеративного навчання).
2. Передача або надання доступу до персональних даних Медіатору та переміщення обробки в Безпечне середовище обробки, надане Медіатором (застосовуючи, де це необхідно, псевдонімізацію або анонімізацію та інші стратегії).
3. Передача або надання доступу до персональних даних Медіатору та подальша передача від Медіатора до Користувача даних.

Цей останній випадок буде останнім, який має бути розглянуто після відкидання всіх попередніх на основі аналізу суворої необхідності. Крім того, саме він вимагатиме більш обмежувальної оцінки вже згаданої необхідності та пропорційності обробки, тобто передбачаючи більш суворі гарантії захисту даних.

До будь-якої обробки, що передбачає передачу персональних даних від Держателя даних, повинен застосовуватися попередній аналіз мінімізації даних. Зокрема, повинні бути застосовані один або більше з наступних методів:

- Видалення непотрібних полів і метаданих (наприклад, у випадку зображень).
- Зменшення рівня деталізації переданих даних.[43]
- Зменшення частоти зібраних подій.[44]
- Додавання шуму зі статистичними характеристиками, який не погіршує необхідну якість.
- Заплутування.
- Кластеризація (наприклад, від 40 до 45 років).
- Перемішування даних.
- Токенізація.
- Застосування методів шифрування.

Серед методів шифрування, які будуть використовуватися, необхідно розглянути використання сучасних стратегій, таких як гомоморфне шифрування, як описано вище, а також шифрування на основі атрибутів, проксі-перешифрування (proxy re-encryption), поліморфне шифрування[45] та інші.

Крім того, в рамках масового доступу до даних слід зазначити, що персональні дані, зібрані Держателем даних, можуть підлягати таким додатковим операціям обробки, які повинні бути записані та включені до Оцінки впливу на захист даних (DPIA):

- Обробка безпосередньо у Держателя даних для досягнення цілей простору даних або Користувачів даних.
- Передача персональних даних до простору даних.
- Передача персональних даних із Простору даних до Користувача даних.

### *4.2.15. БЕЗПЕЧНІ СЕРЕДОВИЩА ОБРОБКИ*

Як згадувалося вище, може виникнути ситуація, коли для реалізації певних операцій обробки необхідно надати доступ до незнеособлених персональних даних Держателів даних у Просторі даних, оскільки інакше цілі обробки не

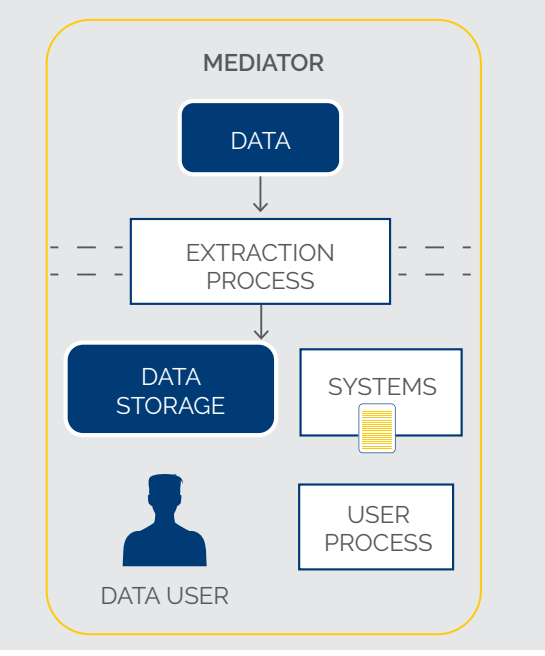
43 *Примітка редактора: Наприклад, округлення дат до місяця замість точного дня, зменшення точності геолокації до міста, замість точних координат*

44 *Примітка редактора: Зменшення частоти, з якою фіксуються та зберігаються певні події або дані. Наприклад, замість того, щоб записувати дані кожну секунду, дані можуть записуватися кожну хвилину або годину. Це допомагає зменшити обсяг зібраної інформації та знизити ризик розкриття персональних даних.*

45 ІНЖЕНЕРНИЙ ОБМІН ПЕРСОНАЛЬНИМИ ДАНИМИ, нові варіанти використання та технології. Агентство Європейського Союзу з кібербезпеки (ENISA). [січень 2023].

можуть бути досягнуті. Деякі з попередніх архітектур стосувалися використання спеціально створеної області для зберігання, вилучення та попередньої обробки. Безпечне середовище обробки пов'язане з такими спеціальними областями і може бути визначене як області або послуги, що надаються тим, хто фізично зберігає дані, і які дозволяють раніше уповноваженому персоналу безпосередньо отримувати доступ до даних та аналізувати їх, вільний доступ до яких становитиме неприйнятний ризик, навіть з урахуванням правових гарантій[46]. У разі доступу Користувачів даних до персональних даних, що зберігаються у Медіаторів, вони являють собою організаційний та технічний захід для мінімізації обробки та зберігання даних (майже доводячи їх до нуля) в руках Користувачів даних.

Доступ і повторне використання даних у безпечному середовищі обробки не можна вважати альтернативою правовим підставам обробки, вичерпно перерахованим у статті 6 GDPR.[47]




**Малюнок 23:** Схема безпечного простору у Медіатора

У випадку захищених даних, що зберігаються органами державного сектору, пункт 15 Преамбули Регламенту про управління даними (DGA) роз'яснює, що повторне використання локально або віддалено в безпечному середовищі обробки може бути дозволено за умови виконання вимог щодо проведення DPIA (Оцінка впливу на захист даних) та консультацій з наглядовим органом відповідно до статей 35 та 36 GDPR, і якщо ризики для прав та інтересів суб'єктів даних визнано мінімальними. Органи державного сектору встановлюють умови, які зберігають цілісність функціонування технічних систем безпечного середовища обробки, що використовується, і залишають за собою право перевіряти процес, засоби та результати обробки даних, що здійснюється повторним користувачем для збереження цілісності захисту даних, а також права забороняти використання результатів, які містять інформацію, що ставить під загрозу права та інтереси третіх осіб[48].

З іншого боку, у статті 50 Пропозицій щодо Регламенту Європейського простору даних з охорони здоров'я (Proposal for a Regulation on the European Health Data Space) EHDS (Європейська система даних про здоров'я) зазначає: «(1) Органи контролю використання та доступу до медичних даних повинні надавати доступ до електронних медичних даних лише через безпечне середовище обробки, з технічними та організаційними заходами та вимогами щодо безпеки та інтероперабельності» і «(2) Органи контролю використання та доступу до медичних даних повинні забезпечити, щоб електронні медичні дані могли бути завантажені Держателями даних і доступні Користувачам даних у безпечному середовищі обробки. Користувачі даних можуть завантажувати з безпечного середовища обробки лише неперсональні електронні медичні дані».

У тій же статті пропозиції EHDS перераховані заходи безпеки, які повинні бути реалізовані в таких середовищах, і які можуть служити орієнтиром для інших інфраструктур масового доступу до даних, таких як Простори даних:

46 SOMA_D2.1.pdf Оцінка рішення «Безпечний простір», включаючи налаштування управління та обробки даних (disinfobservatory.org)

47 Параграф 81 документу «Спільний висновок EDPB-EDPS 3/2021 щодо пропозиції щодо регламенту Європейського Парламенту та Ради з управління даними (Закон про управління даними) [10 травня 2021 р.]».

48 Стаття 5(4) DGA

- Обмеження доступу до безпечного середовища обробки лише уповноваженим особам, зазначеним у відповідному дозволі на дані.
- Мінімізація ризику несанкціонованого читання, копіювання, модифікації або видалення електронних медичних даних, розміщених у безпечному середовищі обробки, за допомогою сучасних технологічних засобів.
- Обмеження введення електронних медичних даних та інспекції, модифікації або видалення електронних медичних даних, розміщених у безпечному середовищі обробки, для обмеженої кількості уповноважених осіб, які можуть бути ідентифіковані.
- Забезпечення, щоб Користувачі даних мали доступ лише до тих електронних медичних даних, які охоплюються їхнім дозволом на дані, за допомогою індивідуальних та унікальних ідентифікаторів користувачів та конфіденційних режимів доступу.
- Ведення ідентифікованих журналів доступу до безпечного середовища обробки протягом необхідного періоду часу для перевірки та аудиту всіх операцій обробки в цьому середовищі.
- Забезпечення дотримання та моніторинг заходів безпеки, зазначених у цій статті, для зменшення потенційних загроз безпеці.

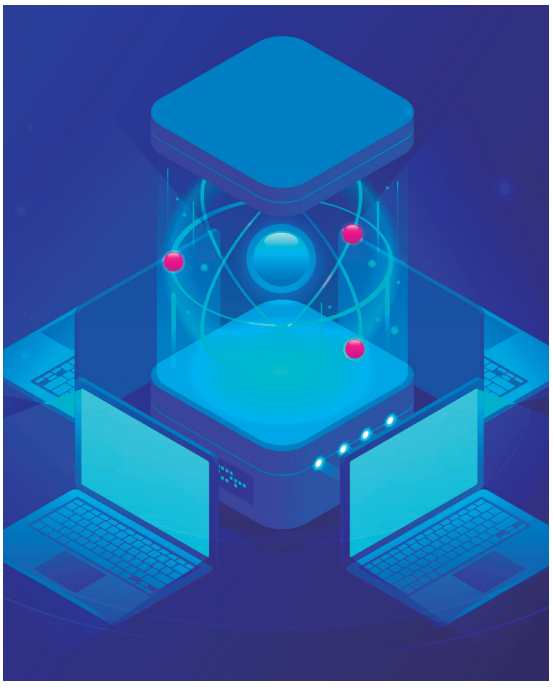
Безпечні середовища обробки можуть бути визначені на двох рівнях:

- Безпечне локальне/фізичне середовище обробки, яке передбачає, що обробка даних і оператор фізично переміщуються на територію, де знаходяться дані, і там вони підлягають контролю доступу та обробки. Потім отриману інформацію можна витягти з безпечного простору.
- Віддалене/віртуальне безпечне середовище обробки, де, хоча обробка виконується на території, де зберігаються дані, оператор може дистанційно керувати процесом обробки так, що оператор не має доступу до даних, але має доступ до отриманої інформації. Керування здійснюється через безпечні віртуальні мережі або навіть фізичні приватні мережі.

В обох випадках вихідні дані не залишають фізичне місце зберігання інформації.

Другий випадок, безпечне середовище обробки з віртуальним доступом, показав свою вразливість на практиці, навіть коли доступ дозволяється тільки через фізичні приватні мережі, що призвело до витоків даних з великим соціальним впливом. Тому його використання повинно доповнюватися іншими заходами захисту.

Крім того, такі безпечні середовища обробки повинні бути реалізовані через довірені середовища виконання.[49] Хоча такі середовища повинні бути реалізовані для всієї обробки персональних даних, саме в безпечних просторах це більш критично. Довірене середовище виконання (trusted execution environment, TEE), як визначено ENISA, є непорушним середовищем обробки на головному процесорі пристрою. Працюючи паралельно з операційною системою і використовуючи як апаратне, так і програмне забезпечення, TEE розроблені таким чином, щоб бути більш безпечними, ніж традиційні середовища обробки. Їх називають також повнофункціональним середовищем виконання операційної системи (Rich Execution Environment, REE), в якому працює операційна система пристрою та застосунки.



49 Розділ 4.3 документа «ІНЖЕНЕРІЯ ЗАХИСТУ ДАНИХ, від теорії до практики. Агентство Європейського Союзу з кібербезпеки (ENISA) [січень 2022 р.]»

## 4.3. ТОЧНІСТЬ

Термін «точність» визначається в статті 5.1 GDPR як один із принципів захисту персональних даних. У разі обробки, яка включає рішення на основі штучного інтелекту, наявність упереджень у моделях висновків (inference models[50]) тісно пов'язана з точністю або якістю даних[51].

Відповідно до другого абзацу пункту 71 Преамбули GDPR, дані, пов'язані з суб'єктами даних[52], повинні бути точними, незалежно від того, зібрані вони безпосередньо чи отримані на основі зроблених висновків. Точніше, визначено, що контролер даних повинен *«використовувати відповідні математичні або статистичні процедури для профілювання»*, які гарантують, що дані, пов'язані з суб'єктом даних, є точними. Тобто, обов'язковим є доведення та документування, що процедури, які використовуються для отримання інформації про суб'єкта даних, є точними і, отже, стабільними та передбачуваними.

Коли контролер вирішує використовувати компонент ШІ для обробки даних, він повинен гарантувати, що дані, які обробляються, генеруються та пов'язані із суб'єктом даних, відповідають вищезазначеним вимогам.

### *4.3.1. ФАКТОРИ, ЩО ВПЛИВАЮТЬ НА ТОЧНІСТЬ*

Що стосується необхідності точності даних про суб'єктів даних, отриманих в наслідок ШІ обробки[53], існує три фактори, які можуть вплинути на таку точність:

- Саме впровадження системи штучного інтелекту. Існують ШІ-системи, такі як експертні системи на основі правил, де сама концепція системи може вносити помилки, які можуть призвести до помилкових висновків, або системи на основі машинного навчання, які не здатні моделювати бажану обробку. Як зазначалося раніше, помилки можуть виникати, оскільки елементи поза штучним інтелектом, такі як біометричні зчитувачі, можуть вносити помилки у вхідні дані. З іншого боку, помилки програмування або проєктування можуть призвести до неправильного впровадження моделі на практиці. У таких випадках можна сказати, що упередження «вбудоване» у рішення, прийняті при побудові моделі аналізу (упередження оцінки та агрегації).
- Навчальний або перевірочний набір даних пошкоджений помилками, навмисно помилковою інформацією[54] або упередженнями, які унеможливлюють точність висновків. Такі упередження можуть бути зумовлені низькою якістю даних, відсутністю даних або вибірковим відбором. Вони також можуть виникати через помилки представлення та вимірювання, пов'язані з тим, як сформований набір даних.
- Упереджена еволюція моделі ШІ. Для штучного інтелекту, що використовує адаптивні техніки, важливо враховувати, що якщо штучний інтелект використовується переважно групою суб'єктів даних із певними характеристиками, це може призвести до виникнення нових упереджень у моделі через зворотний зв'язок.

Для забезпечення необхідної точності навчальних даних необхідно впровадити метрики та методи очищення і відстеження, що гарантують достовірність і цілісність наборів даних. Хоча це не стосується всіх рішень на основі ШІ, навчальні та операційні дані в деяких моделях можуть поділятися на «жорсткі» дані та «м'які» дані. «Жорсткі» дані — це об'єктивні дані, які можна виміряти, наприклад, оцінки, відсоток відвідуваності, аналітичні показники, результати тестів тощо. «М'які» дані — це якісні дані, що містять суб'єктивний компонент або компонент невизначеності. Прикладами «м'яких» даних є дані, отримані шляхом обробки природних мов, думок, особистих оцінок, опитувань тощо. Хоча «жорсткі» дані не позбавлені

50 *Примітка редактора: Моделі інференції (inference models) - це моделі штучного інтелекту, які використовуються для виведення висновків або прогнозів на основі вхідних даних. У контексті захисту даних ці моделі повинні бути точними і передбачуваними, щоб забезпечити відповідність вимогам GDPR.*

51 Термін «якість даних» міститься в статті 47.2.d GDPR.

52 Усі суб'єкти даних мають однакові права щодо захисту даних, а не лише групи, класифіковані як «групи ризику», і «позитивна дискримінація» неможлива.

53 *Примітка редактора: Мова йде про «inferred data», що отримані в наслідок обробки за допомогою так званих «inference models» штучного інтелекту.*

54 Помилкова інформація також може з'явитися в результаті атаки, здійсненої над набором даних, або в результаті включення бекдорів через свідоме маніпулювання даними.

помилок або упереджень, контролер повинен ретельно оцінити проблеми точності, які можуть виникнути внаслідок використання «м'яких» даних або надання більшої ваги «м'яким» даним як джерелу інформації.

Упередження не є виключною проблемою систем штучного інтелекту і може виникати в будь-якій автоматизованій/неавтоматизованій системі обробки, яка приймає рішення або виконує профілювання. Існують такі методи, як оцінка впливу алгоритму (Algorithm Impact Assessment, AIA),[55] які спрямовані на вивчення та виявлення можливих упереджень в рішеннях штучного інтелекту та на забезпечення справедливості у впровадженні моделі. Такі методи повинні аналізувати логіку, яка впроваджується, щоб уникнути закладених на етапі проєктування неточностей і використовувати зрілі тестові моделі та перевірки впровадження для виявлення помилок проєктування[56].

### 4.3.2. ПРОФІЛЮВАННЯ ТА РІШЕННЯ

Виходячи з положень пункту 71 Преамбули GDPR ми можемо стверджувати, що будь-яке рішення або профілювання — це дані, пов'язані з ідентифікованим суб'єктом даних. Тому рішення та профілі є персональними даними. Це означає, що на них поширюються всі обов'язки, передбачені в GDPR, зокрема, дотримання принципів мінімізації та точності.

Рішення та профіль, пов'язані із суб'єктом даних, повинні бути точними. Неважливо, як було отримано таке рішення або профіль: за рішенням людини або за допомогою автоматизованого рішення чи профілювання, на основі персональних даних або будь-яких інших даних, з використанням машинного навчання або без нього. Якщо система штучного інтелекту на основі машинного навчання не забезпечує точне профілювання або рішення щодо ідентифікованого суб'єкта даних, така система не працює належним чином, і контролер повинен забезпечити, щоб такі системи не використовувалися для обробки персональних даних.

### 4.3.3. КОМБІНАЦІЯ ПРОФІЛЮВАННЯ

Іноді компонент ШІ може створювати профілі або приймати рішення щодо одного і того ж суб'єкта даних, але за різних обставин і для різних цілей. Наприклад, при обробці даних в рамках кримінального провадження використовуються рішення на основі ШІ, які дозволяють проаналізувати профіль суб'єкта даних на предмет невиконання судової повістки та одночасно зробити висновок про ризик повторного вчинення тим самим суб'єктом даних кримінального правопорушення. У такому випадку, якщо це не підтверджено документацією, бажано, щоб такі профілі створювалися незалежно один від одного. Профіль, створений в рамках дослідження одного питання, не повинен враховуватися або використовуватися в процесі генерації висновків з іншого питання.

### 4.3.4. БІОМЕТРИЧНА ІНФОРМАЦІЯ

Точність особливо важлива, коли мова йде про обробку, що базується на біометричній інформації, наприклад ШІ-рішення для розпізнавання обличчя,

55 Алгоритмічні оцінки впливу — це інструменти, які дозволяють проводити оцінку зацікавленими сторонами, щоб переконатися, що компоненти ШІ відповідають певним параметрам якості та чи підходять вони для виконання завдання. Посилання на нього можна знайти, наприклад, за адресою https://ainowinstitute.org/aiareport2018.pdf

56 У недавніх статтях висловлено стурбованість з приводу ризику ненавмисного упередження в цих моделях, що призведе до негативного впливу на певних осіб в декількох групах. Хоча було запропоновано багато показників упереджень та визначень власного капіталу, не існує консенсусу щодо визначень та методик, які необхідно використовувати на практиці для цілей оцінки та аудиту таких систем. Незважаючи на це, було розроблено кілька методик для оцінки алгоритмічної дискримінації, таких як Aequitas, набір інструментів аудиту упередженості з відкритим кодом, розроблений Центром науки про дані та публічної політики.

відбитків пальців, голосу тощо. У такому випадку необхідно враховувати технічні фактори (помилкові спрацьовування, помилкові негативи та інші), а також вплив на збір персональних даних фізичних особливостей або інвалідності. Особливо, коли рішення на основі штучного інтелекту не може правильно ідентифікувати суб'єкта даних[57], помилкове профілювання може виникнути ще до початку самої обробки[58].

Контролер повинен враховувати, що навіть якщо такі користувачі можуть становити меншість, необхідно запровадити альтернативні механізми, щоб уникнути дискримінації суб'єктів на тій підставі, що рішення на основі штучного інтелекту не може зчитувати їх біометричні характеристики[59].

### *4.3.5. ОЦІНКА ТОЧНОСТІ ЯК БЕЗПЕРЕРВНИЙ ПРОЦЕС*

У випадку, коли рішення на основі штучного інтелекту еволюціонують, зокрема коли вони отримують зворотний зв'язок як від взаємодії із суб'єктом даних, так і від взаємодії з іншими суб'єктами даних, необхідно проводити повторну оцінку моделі. Зміщення у точності профілювання, виконаного на основі штучного інтелекту, через «ефект фільтраційної бульбашки» або інші обставини, може підсилювати упередження, які користувач має щодо себе або інших людей .

57 У дослідженні «Дискримінація, штучний інтелект та алгоритмічне прийняття рішень» («Discrimination, artificial intelligence, and algorithmic decision-making»), опублікованому Європейською Радою, наведено такі приклади. Програмне забезпечення для відстеження обличчя від Hewlett Packard не розпізнавало обличчя темного кольору як обличчя. А додаток Google Фото позначив фотографію афроамериканської пари як «горил». Фотокамера Nikon постійно запитувала людей з азіатського походження: «Хтось моргнув?» Чоловіку азіатського походження автоматично відхилили фотографію паспорта, тому що «очі суб'єкта закриті», але очі були відкриті. Буоламвіні і Гебру виявили, що «темношкірі жінки є найбільш неправильно класифікованою групою (з рівнем помилок до 34,7%). Максимальний рівень помилок для чоловіків зі світлою шкірою становить 0,8%».

58 Існують випадки людей з нечіткими відбитками пальців, які при використанні системи доступу через ідентифікацію відбитків пальців постійно стикаються з проблемами доступу.

59 Таким чином, щоб уникнути дискримінації за те, що вони не є «біометрично придатними».

## 4.4. ЗБЕРІГАННЯ ДАНИХ

У випадку, коли рішення на основі штучного інтелекту еволюціонують, зокрема коли вони отримують зворотний зв'язок як від взаємодії із суб'єктом даних, так і від взаємодії з іншими суб'єктами даних, необхідно проводити повторну оцінку моделі. Зміщення у точності профілювання, виконаного на основі штучного інтелекту, через «ефект фільтраційної бульбашки»[60] або інші обставини, може підсилювати упередження, які користувач має щодо себе або інших людей[61].

Одним із принципів захисту даних є обмеження періоду зберігання даних, який встановлений у статті 5.1.e. GDPR. Цей принцип передбачає, що дані повинні зберігатися таким чином, щоб ідентифікація суб'єктів даних дозволялася не довше, ніж це необхідно для цілей обробки персональних даних.

Що стосується зберігання даних, які використовуються для розробки та перевірки системи штучного інтелекту, стверджується, що для цілей аудиту та для можливості оцінки їх певних якісних характеристик (наприклад, відсутність упереджень), необхідно зберігати ці дані після закінчення періоду розробки. Іншими словами, вони мають бути доступні протягом невизначеного періоду, щоб мати можливість проводити процеси аудиту та сертифікації постфактум.

Ця ідея ґрунтується на двох припущеннях. Перше — у розробці системи машинного навчання не було впроваджено принципу підзвітності або захисту даних за задумом (data protection by design). Тобто набір даних має бути оцінений заздалегідь, щоб його можна було використовувати для розробки та тестування, а також задокументувати цю оцінку щодо того, чи були дані дійсно відповідними. Все це має бути зроблено

60 Бульбашковий фільтр або Echo Chamber є дуже поширеною ситуацією в пошукачах вмісту, які вивчають смаки користувача і пропонують лише той вміст, який, на їхню думку, подобається користувачеві, і ігнорують всю іншу інформацію таким чином, що користувач замкнений у «бульбашці», яка заважає йому отримати доступ до всіх можливих варіантів.

61 Наприклад, штучний інтелект, який оцінює психологічний статус суб'єкта, може мати зворотний зв'язок із баченням, яке людина має про себе, тим самим направляючи їх до дрейфу власного сприйняття.

заздалегідь, перед розробкою та впровадженням системи штучного інтелекту в експлуатацію.

Друге припущення полягає в тому, що системи штучного інтелекту є або можуть бути лише «поганими системами штучного інтелекту». Це означає, що немає іншого варіанту, окрім проведення аудиту чорної скриньки постфактум. Аудит, який повинен піти ще далі і виконати зворотне проєктування самої системи. Останнє саме по собі є суперечністю, оскільки, якщо процес розробки був проведений неправильно, буде важко зберегти навчальні та тестові дані, що ускладнює визначення, які дані використовувалися, а які ні, і для якої мети та на якому етапі процесу розробки системи штучного інтелекту.

Отже, вимога, яка повинна бути висунута до систем штучного інтелекту, полягає в застосуванні захисту даних ще на стадії проєктування в рамках зрілої методології розробки, в якій заздалегідь проводиться перевірка, а докази зберігаються у вигляді навчальних і верифікаційних наборів.

## 4.5. ЗВІТНОСТІ

Одним із найважливіших нововведень GDPR у порівнянні з попередніми регламентами захисту даних є встановлення нового принципу — проактивної відповідальності або «підзвітності», який встановлений у статті 5.2. Цей принцип настільки унікальний, що він не включений до початкового набору принципів, перелічених у розділі 1 вищезгаданої статті 5, а йому присвячено окремий розділ у стандарті.

Термін «підзвітність» є добре відомим в англосаксонському світі і тому визначення його в GDPR немає. Підзвітність означає розумний, свідомий та об'єктивний спосіб виконання покладених обов'язків, пояснюваність та простежуваність кожної дії та рішення, компетентність у виконанні зобов'язань, прийняття зобов'язань, демонстративність та очікування підзвітності.

### *4.5.1. ПІДЗВІТНІСТЬ ЗАСОБІВ*

Принцип підзвітності повинен застосовуватися відповідальними за дотримання вимог GDPR у своїх процесах обробки даних[62]. Однак слід враховувати, що принцип підзвітності важко виконати в обробці, якщо засоби, що використовуються для її реалізації (які визначають природу обробки), самі по собі не є «підзвітними». У цьому випадку, якщо системи машинного навчання, на яких базується одна або кілька операцій обробки, не нададуть контролеру необхідну документацію, докази, дані перевірок та іншу інформацію, останній не зможе забезпечити необхідну підзвітність.

Тому контролер буде зобов'язаний вимагати необхідну інформацію та співпрацю від обробників даних, які надають або впроваджують операції, що базуються на системах машинного навчання. Крім того, контролеру знадобиться така інформація від технічних постачальників, які не мають статусу обробників, але послуги та інструменти яких можуть впливати на обробку персональних даних. Особливо це стосується випадків, коли їхні послуги включають додаткові операції обробки. Такі постачальники повинні надати контролеру можливість бути належним чином обізнаним і надати докази своєї якості, такі як аудиторські звіти або отримані сертифікати, що мають значення у контексті захисту даних.

### *4.5.2. МОДЕЛЬ РОЗВИТКУ ЗРІЛОСТІ*

Якщо система штучного інтелекту буде включена в процес обробки даних, незалежно від того, чи є вона компонентом, послугою або комерційним продуктом, вона повинна бути розроблена та впроваджена відповідно до моделі зрілого процесу розробки, яка гарантує її якісні параметри, продуктивність, документацію та надання інформації контролеру/обробнику в обсязі, достатньому для оцінки відповідності вимогам GDPR.

Той факт, що системи штучного інтелекту, наприклад система машинного навчання (ML), дотримуються стратегії розвитку, заснованої на налаштуванні шляхом навчання алгоритму, замість розробки алгоритму безпосередньо

62 Примітка редактора: GDPR застосовує принцип підзвітності (accountability) саме до контролерів даних.

програмістом, може свідчити про те, що знання програмування або комп'ютерної інженерії не потрібні. Але це тільки на етапі прототипування. Загальні алгоритми машинного навчання можуть бути отримані від третіх сторін, а потім навчені. Це дозволяє швидко створювати перші прототипи і означає, що на ранніх стадіях розробки або перевірки концепції проєкту ШІ дозволяє експериментувати з даними для перевірки концепцій та отримання результатів, як і інші інструменти, пов'язані з наукою про дані, такі як статистичні інструменти.

Як уже зазначалося, реалізація моделі чи ідеї в реальній системі має багато наслідків за межами моделі ШІ: бібліотеки, побічні канали, проблеми безпеки, проблеми з датчиками тощо. Розробка системи машинного навчання повинна здійснюватися професійною командою з використанням зрілої методології розробки в рамках проєкту на основі науки про дані. Це означає, принаймні, що залучення спеціаліста з Data science необхідне для очищення даних, сертифікації бібліотек та основних інструментів розробки, перевірки та аудиту кінцевих продуктів і, звичайно, консультацій та моніторингу делегата захисту даних (стаття 39.1.c GDPR), як тільки обробляються персональні дані.

Зрілість процесу розробки — це застосування принципу проактивної відповідальності в разі використання персональних даних при розробці. Це також спосіб виконання зобов'язань щодо захисту даних за задумом (на стадії проєктування), які встановлені в статті 25 GDPR.

### 4.5.3. ЧОРНА СКРИНЬКА

Неможливо вважати, що чорна скринька є частиною природи системи штучного інтелекту. Концепція чорної скриньки не є частиною природи будь-якого програмного забезпечення, алгоритму або технологічного компонента як такого. Чорна скринька — це стан, який може мати будь-який з цих елементів через те, що документація системи відсутня, або через те, що протягом усього процесу розробки ігнорувалися найелементарніші процедури зрілої розробки продукту чи послуги. Системи штучного інтелекту, як і будь-який промисловий або технологічний продукт, можуть бути розроблені зріло та відповідально, що призведе до задокументованого та простежуваного процесу, який включає верифікацію та валідацію[63] системи, що забезпечують об'єктивні докази якості продукту. Система штучного інтелекту, яка є чорною скринькою через те, що її не було розроблено належним чином, є погано розробленою системою штучного інтелекту.

### 4.5.4. ВЕРИФІКАЦІЯ ТА ВАЛІДАЦІЯ

Тестування та/або верифікація компонента ШІ є ключовим етапом у процесі його розробки, а також важливим критерієм для контролера при виборі конкретного компонента серед інших варіантів від різних розробників. Однак контролер повинен пам'ятати, що включення верифікованого компонента ШІ в процес обробки даних не гарантує, що сам процес буде валідований, і не підтверджує придатність компонента ШІ для конкретної обробки. Тестування компонента ШІ гарантує, що результати його проєктування та розробки відповідають вимогам цього компонента. Валідація процесу обробки охоплює значно ширший обсяг. Вона повинна забезпечувати, що кінцеві продукти та послуги відповідають вимогам для конкретного застосування або передбаченого використання[64]. Валідація гарантує, що процес обробки, а також ШІ-рішення, на якому він базується, досягає запланованих результатів для певного продукту або послуги.

Іншими словами, валідація процесу обробки, що включає компонент ШІ, повинна здійснюватися в умовах, які відображають реальний контекст, де очікується розгортання цього процесу[65]. Крім того, процес валідації потребує періодичного перегляду, з огляду на те, що такий контекст

63 *Примітка редактора:* ***Верифікація*** *(verification): процес перевірки того, що система відповідає своїм специфікаціям та вимогам.* ***Валідація*** *(validation): процес перевірки того, що система виконує свої призначені функції в реальних умовах.*

64 ISO 9001:2015 Системи менеджменту якості. Вимоги

65 «Практичний посібник зі штучного інтелекту в рамках охорони здоров'я», згаданий у бібліографії, описує випадок дитячої системи діагностики пневмонії, яка отримала точність 97%. Однак, коли такий алгоритм був застосований до населення Мадрида, точність знизилася до 64%. Аналіз даних тренінгу показав, що населення було віком від 0 до 5 років, тоді як сфера застосування педіатричного лікування в Мадриді включала дітей віком до 14 років.

або сама обробка можуть змінюватися та розвиватися.

### *4.5.5. ПЕРЕВІРКА КОДУ*

Одним із аспектів, який викликає суперечки щодо ШІ, є питання перевірки коду як інструменту оцінки системи штучного інтелекту.

Розглянемо приклад алгоритму ШІ, реалізованого на нейронній мережі шляхом налаштування набору параметрів. Якщо ми зосередимося на аудиті поведінки алгоритму щодо налаштування цих параметрів, то повинні визнати, що перегляд коду такої нейронної мережі надасть аудитору мало інформації. Зрештою, нейронна мережа, включаючи ті, що використовують «глибинне навчання», є не більше ніж набором однакових функцій із наборами входів і виходів, що виконують однаковий простий процес і працюють ітеративно. Виконання тієї чи іншої функції залежить від точного налаштування параметрів або ваг, пов'язаних із кожним із входів.

Хоча це правильне наближення з теоретичної точки зору, необхідно враховувати, що відстань між теорією і практикою може бути дуже великою. Також слід розрізняти перевірку лише функціональності алгоритму та перевірку всієї системи штучного інтелекту. Реалізація алгоритму штучного інтелекту або будь-якої системи штучного інтелекту є комп'ютерним програмним кодом і тому має ті ж вразливості та проблеми реалізації або потенційного втручання, що й будь-який інший комп'ютерний код.

По-перше, навчання автоматичної системи навчання має виконуватися на коді, придбаному або розробленому, в якому втілено алгоритм загального призначення, що потребує налаштування параметрів. Цей алгоритм може містити помилки та погані практики кодування (від глобальних змінних до проблем набору тексту), які можуть призвести до проблем під час експлуатації, які не були виявлені під час перевірки.

Також не поодинокими є випадки, коли під впливом стресу, пов'язаного з процесом розробки, розробники намагалися вирішити деякі проблеми налаштування алгоритму в останню хвилину, зробивши невелику маніпуляцію з кодом. Такі проблеми поширені незалежно від методології, що використовується для розробки додатків. Для досвідченого ока, яке проводить перевірку коду, що є частиною аудиту якості систем за методом білого ящика, таку обставину легко виявити. По-друге, при роботі з комерційними або промисловими системами ШІ не використовується той самий код, який був отриманий під час розробки, для кінцевої реалізації в продуктах. Як вже зазначалося, один і той самий алгоритм може бути реалізований на різних операційних системах з використанням різних бібліотек та пристроїв. Високорівневий код може бути однаковим, але низькорівневий код відрізнятиметься. Невеликі відмінності в реалізації можуть мати значний вплив на продуктивність і надійність системи ШІ та впливати на обробку даних. Крім того, завжди може виникнути проблема зловмисної маніпуляції кодом будь-якою зі сторін, залучених від фази розробки до етапу обслуговування системи машинного навчання. Будь-який хакер, незадоволений співробітник або недобросовісна організація можуть внести невеликі зміни в код, які вплинуть на результати, що отримуються системою ШІ, на основі певної вхідної змінної або проміжної умови.

Нарешті, код алгоритму ШІ є лише частиною системного коду машинного навчання. Крім коду алгоритму необхідно враховувати наявність бібліотек для введення даних і процесів, які виконуються на виході алгоритму ШІ. Всі вони можуть містити програмні помилки, бути підданими навмисній маніпуляції або впливу з боку третіх осіб.

Перевірка коду є одним із етапів аудиту будь-якої системи, і вона стає тим більш важливою, чим більш критичною є робота цієї системи, незалежно від того, як вона розроблена.

### *4.5.6. УПРАВЛІННЯ РИЗИКАМИ*

У пропозиції Європейського регламенту щодо штучного інтелекту та інших етичних класифікацій ШІ-систем є перелік областей ШІ-систем високого ризику — концепція, яку слід кваліфікувати стосовно визначення високого ризику в GDPR.

Щодо високого ризику, стаття 35 GDPR «Оцінка впливу на захист даних» (DPIA) встановлює, що відповідальна особа повинна визначити, чи є обробка високоризиковою для прав і свобод фізичних осіб, враховуючи природу, обсяг, контекст та цілі обробки, зокрема, якщо використовуються нові технології. Існування високого ризику в обробці визначається у невичерпний спосіб у частині 3 статті 35 та переліку у частині 4 статті 35 GDPR, у пунктах частини 2 статті 32 або Преамбули 75 GDPR, у частині 2 статті 28 Органічного закону 3/2018 від 5 грудня про захист персональних даних і гарантії цифрових прав (LOPDGDD), у прикладах з Керівництва WP248 Групи статті 29 (тепер Європейський комітет із захисту даних - EDPB), у спеціальних регламентах, які вимагають DPIA, у конкретних випадках і умовах, описаних у керівництвах, опублікованих EDPB для конкретних обробок, у конкретних випадках і умовах, описаних у кодексах поведінки відповідно до статті 40 та у механізмах сертифікації відповідно до статті 42 GDPR. Отже, класифікація ризиків систем штучного інтелекту пропонує, але не замінює оцінку ризиків обробки персональних даних, встановлену в GDPR. Класифікація ризиків систем штучного інтелекту доповнює оцінку ризиків GDPR таким чином, що вона підтримує оцінку високого ризику тих операцій з обробки, які включають такі типи систем. Управління ризиками в різних обробках, де використовується система ШІ, повинно здійснюватися з урахуванням не лише природи будь-якої з її операцій (однієї або більше систем ШІ), але й ризиків для прав і свобод, що виникають від спільного використання всіх систем, які реалізують обробку, обсягу чи розширення такої обробки, контексту, у якому вона здійснюється, цілей, а також невизначеності, що може бути внесена використанням нових технологій.

У прикладі, описаному вище про обробку в рекрутингу, цілком можливо, що єдиною включеною ШІ-системою є чат-бот з низьким ризиком. Але вся обробка може мати високий ризик, якщо, наприклад, потрібні якісь дані, які можуть означати високий ризик для суб'єкта даних. Такі дані не обов'язково повинні відноситися до чутливих даних, це може бути, наприклад, адреса жертви гендерного насильства.

## 4.6. БЕЗПЕКА

У статті 32 GDPR встановлено, що і контролер, і процесор повинні застосовувати відповідні технічні та організаційні заходи для гарантування належного рівня захисту прав і свобод суб'єктів даних. Такі заходи повинні бути адаптовані з урахуванням витрат на впровадження, характеру, обсягу, контексту та цілей обробки, а також змінних ризиків ймовірності та тяжкості. Не існує стандартного рішення для всіх процесів обробки, і тим більше для тих, що включають систему штучного інтелекту. Рішення необхідно оцінювати за допомогою аналізу ризиків, який повинен бути пов'язаний з ризиками для прав і свобод суб'єктів даних з точки зору захисту даних.

### *4.6.1. СПЕЦИФІЧНІ ЗАГРОЗИ В КОМПОНЕНТАХ ШІ*

Крім аналізу заходів безпеки, які є загальними для будь-якої системи, існують конкретні гарантії діяльності з обробки, яка включає системи штучного інтелекту. Ці гарантії повинні враховувати специфічні загрози, що виникають з факту розробки компонентів ШІ третіми сторонами або з розкриття даних третім сторонам.

Існують типології атак і захисту щодо систем штучного інтелекту, які були проаналізовані[66]. Серед різних заходів безпеки доцільно приділити особливу увагу управлінню такими типами загроз:

- Доступ та маніпулювання навчальним набором даних, наприклад шляхом зараження шкідливими шаблонами.

66 Опитування щодо загроз безпеці та захисних методів машини, Qiang Liu et al., IEEE Access, ISSN:2169-3536, лютий 2018

- Включення троянських програм[67] і бекдорів[68] під час розробки ШІ — в сам код або в інструменти розробки[69].
- Маніпулювання API користувача, що дозволяє отримати доступ до моделі, на рівні як чорної скриньки, так і білої скриньки, для маніпулювання параметрами моделі, витоку моделі третім особам, атак на цілісність або доступність висновків (інференцій)[70].
- Атаки за допомогою «ворожого машинного навчання»[71], що вимагає проведення аналізу надійності[72] та контролю над надходженням даних до моделі.
- Атаки через імітацію шаблонів, які система визнає допустимими[73].
- Повторна ідентифікація персональних даних, включених у модель (приналежність, висновок або[74] інверсія моделі[75]), внутрішніми та зовнішніми [76]користувачами.
- Шахрайство або введення в оману штучного інтелекту з боку суб'єктів даних, особливо в тих випадках, коли це може спричинити збитки для інших суб'єктів даних,[77] що передбачає необхідність проведення аналізу надійності у світлі таких дій та проведення аудитів.
- Витік третім особам результатів профілювання або рішень, визначених штучним інтелектом (також пов'язаних з API користувача).
- Витік або доступ до журналів, отриманих в результаті висновків, створених під час взаємодії з суб'єктами даних.

### *4.6.2. ЖУРНАЛИ АБО ЗАПИСИ АКТИВНОСТІ*

- Існування файлів журналів або записів активності, виконання аудитів (автоматизованих або ручних) та сертифікація процесу є невід'ємною частиною стратегій «підзвітності» або стратегій проактивної відповідальності, але вони також виникають із законодавчих вимог, спеціально встановлених у секторальному регулюванні.
- Файли журналів необхідні для підтримки процесів аудиту та механізмів безпеки, які стосуються захисту даних. Ці файли журналів повинні надавати докази для таких завдань:
- Встановити, хто і за яких обставин отримує доступ до персональних даних, які можуть бути включені в модель.

Забезпечити можливість відстеження щодо оновлення моделей висновків (інференції), зв'язку API користувача з моделлю та виявлення зловживань або спроб вторгнення.

Забезпечити можливість відстеження, щоб забезпечити управління розкриттям даних між усіма сторонами, що втручаються в рішення на основі штучного інтелекту відповідно до зобов'язань, що випливають з пункту 66 Преамбули GDPR.

67 Трояни в IA, IARPA https://www.iarpa.gov/index.php?option=com_content&view=article&id=1150&Itemid=448

68 Прикладом маніпулювання алгоритмом розпізнавання зображень https://www.bleepingcomputer.com/news/security/ai-training-algorithms-susceptible-to-backdoors-manipulation/ також вбудовування бекдору в моделі згорткових нейронних мереж через невидимі збурення https://arxiv.org/pdf/1808.10307.pdf

69 Уразливість в бібліотеці для розробки моделей МН: https://cyware.com/news/critical-vulnerability-in-numpy-could-allow-attackers-to-perform-remote-code-execution-33117832

70 З метою збільшення частоти хибнопозитивних результатів і частоти помилкових негативів.

71 Техніка атаки, що полягає в подачі ШІ прикладом даних, які можуть не відрізнятися від звичайних даних з точки зору людського сприйняття, але можуть включати невеликі порушення, які змушують ШІ робити помилкові висновки.

72 Досліджуючи ландшафт просторової стійкості, Логан Енгстрем MIT https://arxiv.org/pdf/1712.02779.pdf

73 Орієнтований на такі програми, як розпізнавання обличь або виявлення вторгнень і часто пов'язаний з методами «змагального машинного навчання».

74 Коли можна встановити, чи є певна особа, її дані, частиною моделі навчання.

75 Атака шляхом інверсії моделі МН відбувається, коли зловмисник має доступ до певних персональних даних користувача, включених до моделі ШІ, і може зробити висновок про додаткову особисту інформацію таких осіб, аналізуючи входи та виходи моделі.

76 Точніше, коли моделі ШІ купуються третіми особами у розробника.

77 Типовим прикладом шахрайства в аналізі резюме за допомогою штучного інтелекту є запис неіснуючих переваг тим же кольором, що і фон документа, який неможливо прочитати людині, але не машині, таким чином, щоб система відбору кандидатів могла бути обдурена на шкоду чесним кандидатам. Ця система вже використовувалася для обману пошуковика Google, щоб той індексував сторінки за ключовими словами, які не були видимі для користувача і не мали ніякого відношення до реального вмісту сторінки.


Забезпечити контроль за якісними параметрами інференції, коли ШІ використовується для прийняття рішень або в процесах допомоги в прийнятті рішень.

Законний інтерес контролера як законна підстава для обробки персональних даних у файлах записів про діяльність для цілей безпеки пояснюється в пункті 49 Преамбули [78]GDPR «в тій мірі, в якій це суворо необхідно і пропорційно для цілей забезпечення мережевої та інформаційної безпеки». В інших випадках галузеве регулювання встановлює зобов'язання щодо збереження та обробки записів активності, наприклад як Закон 10/2010 про запобігання відмиванню грошей та боротьбу з фінансуванням тероризму[79], і тому в такому випадку правовою основою для виконання обробки буде дотримання застосовного юридичного зобов'язання контролера даних. Аналогічно інші правові підстави можуть бути використані для легітимізації обробки даних у реєстрах журналів.

У будь-якому випадку контролер повинен знати про зобов'язання та обмеження, встановлені галузевим регулюванням, оскільки такі правові підстави не дозволяють обробляти персональні дані, що містяться у файлі журналу, для інших цілей, таких як оцінка продуктивності або еволюція системи штучного інтелекту. Таким чином, контролер повинен забезпечити впровадження гарантій, щоб уникнути доступу до такого запису та його використання для цілей, для яких немає законних підстав.

Розробники систем штучного інтелекту, коли вони використовують рішення на основі машинного навчання, повинні впроваджувати інші реєстри з метою документування та дотримання принципу підзвітності, щоб забезпечити простежуваність походження навчальних даних та валідацію таких даних, а також записи аналізів щодо достовірності таких даних та відповідних результатів.

78 Преамбула 49 — Опрацювання персональних даних в обсязі, суворо необхідному та пропорційному для цілей забезпечення мережевої та інформаційної безпеки, вважається законним інтересом контролера даних, тобто здатністю мережі або інформаційної системи протистояти, при певному рівні впевненості, випадковим подіям або незаконним чи зловмисним діям, які ставлять під загрозу доступність, автентичність, цілісність і конфіденційність збережених або переданих персональних даних, а також безпека пов'язаних послуг, що пропонуються або доступні через ці мережі та системи державними органами, групами реагування на комп'ютерні надзвичайні ситуації (CERT), групами реагування на інциденти комп'ютерної безпеки (CSIRT), постачальниками електронних комунікаційних мереж і послуг та постачальниками технологій і послуг безпеки, становить законний інтерес відповідного контролера даних. Це може включати, наприклад, запобігання несанкціонованому доступу до мереж електронних комунікацій та розповсюдження шкідливого коду, а також припинення атак «відмови в обслуговуванні» та пошкодження комп'ютерних та електронних комунікаційних систем.

79 Стаття 25. Збереження документів.
1. Зобов'язані суб'єкти зберігають документацію протягом десяти років з моменту виконання зобов'язань, встановлених цією Угодою, і після цього вони підлягають стиранню. Через п'ять років після припинення ділових відносин або здійснення нерегулярної операції доступ до документації, що зберігається, отримують лише органи внутрішнього контролю зобов'язаного суб'єкта, включаючи технічні підрозділи запобігання та, залежно від обставин, особи, відповідальні за їхній юридичний захист.
Точніше, зобов'язані суб'єкти зберігають їх для використання в кожному дослідженні або розслідуванні, з точки зору можливого відмивання грошей або фінансування тероризму Виконавчою службою комісії або будь-яким іншим юридично уповноваженим органом.
a) Копія документів, які можуть бути запитані для застосування заходів належної обачності протягом десяти років після припинення ділових відносин або завершення операції.
b) Оригінал документа або копію з підтверджуючою силою документів або записів, які належним чином підтверджують операції, сторони таких операцій та ділові відносини протягом десяти років після завершення операції або припинення ділових відносин.
2. Зобов'язані суб'єкти, за винятками, встановленими нормативно-правовими актами, зберігають копії ідентифікаційних документів, зазначених у статті 3.2, в оптичному, магнітному або електронному форматі таким чином, щоб забезпечувалася цілісність даних, точне зчитування даних, неможливість маніпулювання та належне зберігання і локалізація таких даних.
У будь-якому випадку, картотека зобов'язаних суб'єктів даних повинна забезпечувати належне управління та доступність документації як для цілей внутрішнього контролю, так і для належної уваги до часу та способу виконання вимог органами влади.

## 4.7. АУДИТ СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ ТА ЗВОРОТНЕ ПРОЄКТУВАННЯ

У зв'язку з усім вищезазначеним, аудит системи ШІ може мати підхід білого ящика або чорного ящика. Це не нове, але є звичним для аудиту будь-якої технологічної системи. Перший підхід передбачає, що аудитор матиме доступ до всієї документації з дизайну, впровадження, верифікації та валідації системи, що підлягає аудиту, для висновку про те, що вона відповідає стандартам якості. Однак аудит чорного ящика припускає, що аудитор не має доступу до зазначеної інформації і буде намагатися визначити якість системи, в даному випадку ШІ, через поведінковий аналіз. Це передбачає проведення максимальної кількості тестів, щоб дослідити всі можливі комбінації або категорії комбінацій, які дозволяють перевірити поведінку системи.

Такий тип аудиту чорного ящика підходить, коли неможливо дізнатися реалізацію системи, або тому, що вона не була задокументована, або документація була втрачена, або процес розробки не був зрілим, або тому, що не бажають надавати документацію системи, або вона належить третій стороні, яку не хочуть попереджати, наприклад, у випадку безпекових аудитів типу пентестингу.

Коротше кажучи, аудит чорного ящика зазвичай проводиться, коли є перешкода для проведення аудиту білого ящика, який може досягти щонайменше того ж, що й аудит чорного ящика, але з набагато більшими гарантіями того, що поведінка системи та якість процесу розробки були ретельно перевірені.

Аудити білого ящика дозволяють визначити, крім якості продукту, якість процесу розробки, досліджуючи, наприклад, у випадку систем ШІ, причини прийняття дизайнерських рішень, вибір моделі нейронної мережі, набір даних для розробки та тестування, тип реалізації системи, результати тестів верифікації та валідації тощо. Коротше кажучи, всі кроки, які вживають розробники, щоб забезпечити кінцеву якість системи ШІ, наприклад, відсутність упереджень. Якщо так, то це дає гарантії хорошої (або принаймні контрольованої) поведінки продукту за межами контексту валідаційних тестів, яким він піддавався.

У цьому відношенні аудит чорного ящика не може стати процесом зворотного проєктування, тому що відсутність зрілого та якісного процесу розробки призводить до неможливості проведення аудиту білого ящика. Зворотне проєктування - це процес або метод, за допомогою якого намагаються через дедуктивне мислення зрозуміти, як працює пристрій, процес, система або алгоритм, маючи мало або взагалі не маючи інформації про те, як він функціонує. Усі процеси зворотного проєктування складаються з трьох основних кроків: вилучення інформації, моделювання та огляд. Зазвичай його використовують як частину аналізу для отримання дизайнерських особливостей продуктів із малою або взагалі відсутньою додатковою інформацією про процедури, які брали участь в їх оригінальному виробництві. У деяких випадках метою процесу зворотного проєктування може бути створення неіснуючої документації системи або допомога в покращенні розуміння базового алгоритму.

Розгляд розробки системи ШІ як непрозорого процесу без гарантій якості, який буде підданий процесу аудиту чорного ящика у кожному окремому випадку, щоб допомогти зворотному проєктуванню того, як вона дійсно працює та які її реальні результати, суперечить принципу захисту даних за дизайном, а також безпеці за дизайном та зрілості розробки технологічних систем.

# 5. ПРАВОВІ ПІДСТАВИ ДЛЯ ОБРОБКИ

Обробка персональних даних у межах процесів, що включають системи штучного інтелекту, потребує спеціальної правової підстави, визначеної у статті 6 GDPR. Для повторного використання даних, що зберігаються державними органами, законодавство Союзу або держав-членів повинно забезпечити відповідну правову підставу відповідно до GDPR, і державні органи повинні чітко її визначити.

Правова підстава для обробки персональних даних у межах систем ШІ може базуватися на будь-якій з підстав, зазначених у статті 6 GDPR, включаючи дотримання юридичних зобов'язань або законний інтерес, якщо обробка не здійснюється державними адміністраціями у виконанні їхніх функцій. Важливо, щоб правова підстава була чітко та правильно сформульована у документації процесу обробки.

Тісно пов'язаним із законністю обробки є принцип обмеження мети. Межі того, що становить законну обробку та подальшу сумісну обробку даних, повинні бути дуже чіткими для всіх зацікавлених сторін. У випадку сумісної обробки обробка в контексті, що включає ШІ, має відповідати вимогам статті 5(1)(b) GDPR (обмеження мети) та статті 6(4) GDPR (тест на сумісність). Якщо має здійснюватися подальша обробка, контролер повинен спочатку забезпечити, що така обробка сумісна з початковою метою, і відповідно спроєктувати її. Сумісність або несумісність нової мети повинна оцінюватися відповідно до критеріїв, зазначених у статті 6(4).

Крім того, якщо мета полягає в архівуванні в суспільних інтересах, наукових та історичних дослідженнях або статистичних цілях, це повинно відповідати статті 89(1) GDPR (гарантії та винятки щодо обробки для наукових цілей), з урахуванням статті 50 GDPR. Думка 3/2013 Робочої групи статті 29 надає корисні рекомендації щодо реалізації принципу обмеження мети, а також щодо належного використання різних правових підстав для обробки персональних даних і залишається значущою також під GDPR.

Якщо законність обробки базувалася на згоді, подальша обробка на основі тесту на сумісність відповідно до статті 6(4) GDPR може суперечити принципу вимог до згоди. Тому, якщо обробка базувалася на згоді, подальша обробка може здійснюватися лише в тому випадку, якщо контролер запитує конкретну згоду на цю іншу окрему мету або якщо контролер може довести, що він спирається на законодавство Союзу або держави-члена для захисту цілей, зазначених у статті 23 GDPR.

Стаття 5(1)(b) GDPR передбачає, що подальша обробка персональних даних для архівування в суспільних інтересах, наукових та історичних дослідженнях або статистичних цілей не вважається несумісною з початковими цілями. Це не означає, що ці цілі завжди вважаються сумісними, а скоріше, що початковою точкою аналізу є можливість сумісності. Для визначення, чи сумісна мета подальшої обробки з метою, для якої персональні дані були спочатку зібрані, контролер, після виконання всіх вимог щодо законності початкової обробки, повинен враховувати, зокрема: будь-який зв'язок між цими цілями та цілями передбачуваної подальшої обробки; контекст, у якому були зібрані персональні дані,

зокрема розумні очікування суб'єктів даних щодо подальшого використання, природу персональних даних; наслідки передбачуваної подальшої обробки для суб'єктів даних; і наявність відповідних гарантій як у початкових, так і в передбачуваних подальших операціях обробки. Контролери повинні також бути обережними, щоб не розширювати межі «сумісних цілей» статті 6(4) і пам'ятати, що обробка буде в межах розумних очікувань суб'єктів даних.

Існування правової підстави не звільняє від виконання всіх принципів, прав та обов'язків, встановлених у GDPR. Зокрема, модель відповідності на основі проактивної підзвітності, визначена у GDPR, вимагає більше, ніж просто вибір правової підстави для обробки відповідно до категорій у статті 6:

- По-перше, у випадку обробки спеціальних категорій даних необхідно продемонструвати, що виконані умови для зняття заборони на таку обробку, викладені у статті 9(2) GDPR.
- Зняття заборони на обробку спеціальних категорій даних, зазначених у статті 9(2)(g) (суттєвий суспільний інтерес), (h) (цілі профілактичної або трудової медицини, оцінка здатності працівника виконувати роботу, медична діагностика, надання медичної чи соціальної допомоги або лікування, або управління системами та службами охорони здоров'я та соціального забезпечення) та (i) (суспільний інтерес у галузі громадського здоров'я) GDPR, повинно бути покрито регламентом, який має силу закону, і повинні бути встановлені відповідні гарантії.
- GDPR вимагає явної оцінки необхідності, яка також включає аналіз доцільності обробки для тих, хто має право відповідно до статті 6(1)(b) до (f), і для зняття заборон на основі статті 9(2)(b), (c), (f) та (g).
- GDPR вимагає оцінки пропорційності обробки для тих, хто має правову підставу (стаття 6(1)(c) GDPR), суспільний інтерес або виконання офіційних повноважень (стаття 6(1)(e) GDPR), для зняття заборон на обробку спеціальних категорій даних відповідно до статей 9(2)(g) (суттєвий суспільний інтерес) та 9(2)(j) (цілі архівування в суспільних інтересах, наукові або історичні дослідження або статистичні цілі).
- Для будь-якої обробки з високим ризиком необхідно провести оцінку впливу на захист даних (DPIA) для управління цим ризиком і пройти оцінку доцільності, необхідності та суворої пропорційності.
- Якщо обробка включає прийняття рішень виключно на основі автоматизованої обробки, включаючи профілювання, що має юридичні наслідки або значно впливає на суб'єкта даних у подібний спосіб, умови, що дозволяють таку обробку відповідно до статті 22 GDPR, повинні бути виконані.

Щодо анонімізації, слід зазначити, що це є обробка персональних даних, і, як і всяка обробка, вона повинна відповідати тим самим вимогам, які викладені вище.

Нарешті, існують інші обмеження щодо обробки персональних даних, які не випливають з GDPR. Наприклад, у DGA (Data Governance Act) послуга посередництва даних, щоб вважатися такою, не може використовувати дані щодо тих, хто надає свої послуги, для інших цілей, окрім їх надання користувачам даних, і повинна надавати послуги посередництва через юридичну особу, незалежну від інших видів діяльності постачальника таких послуг. Такі послуги також не можуть виконувати конверсії форматів персональних даних, якщо не виконані певні умови, і суб'єктам даних надається можливість виключення. DGA також обмежує можливості обробки даних організаціями альтруїзму даних, які добровільно вирішили подати заявку на реєстрацію в відповідному національному реєстрі, у сенсі, що вони не можуть використовувати дані для цілей, відмінних від тих, що є суспільно корисними, для яких суб'єкт даних або власник даних дозволяє обробку.

# 6. ПРАВА СУБ'ЄКТІВ ДАНИХ

Будь-які контролери даних, які використовують рішення на основі ШІ для обробки персональних даних, проведення профілювання або прийняття автоматизованих рішень, повинні усвідомлювати, що суб'єкти даних мають права, пов'язані із захистом їх персональних даних, які необхідно враховувати.

Таким чином, за замовчуванням контролери даних повинні враховувати необхідність включення відповідних механізмів і процедур для обробки будь-яких отриманих претензій, і ці процедури повинні відповідати масштабу обробки, яку вони проводять.

Для тих персональних даних, які розподіляються між декількома контролерами, наприклад, якщо система ШІ включає персональні дані, призначені для тренування системи або її розвитку, необхідно, відповідно до принципу підзвітності, включити ефективну модель управління інформацією, яка дозволяє відстежувати інформацію з метою ідентифікації відповідного контролера та надання можливості суб'єктам даних здійснювати свої права. Ця модель управління інформацією також повинна бути надана, коли обробка включає одного або більше контролерів, і включати у відповідну угоду ті завдання, які призначені контролеру у зв'язку із здійсненням прав.

Нарешті, слід враховувати, що стаття 11 GDPR встановлює, що при здійсненні прав на доступ, видалення або обмеження обробки, якщо неможливо ідентифікувати суб'єкта даних, контролер не зобов'язаний зберігати, отримувати або підтримувати додаткову інформацію для ідентифікації суб'єкта даних з єдиною метою виконання положень GDPR. Проте відповідний суб'єкт даних має право надати додаткову інформацію, яка забезпечить його ідентифікацію та, таким чином, дозволить йому здійснювати свої права.

## A. ПРАВО НА ДОСТУП

Право на доступ повинно бути забезпечене контролером у будь-якій діяльності з обробки, яка включає системи ШІ. Це включає будь-які дані для навчання, які можуть бути включені в системи ШІ і які можуть бути вилучені контролером, що експлуатує систему ШІ.

## B. ПРАВО НА ВИДАЛЕННЯ

Право на видалення передбачає проактивний підхід з боку контролера даних для гарантування, як це встановлено у пункті 39[80] Преамбули GDPR, що дані видаляються, коли вони більше не потрібні для цілей обробки, і, зокрема,

80 Преамбула 39 (...) гарантує, що період, протягом якого зберігаються персональні дані, обмежений суворим мінімумом. Персональні дані повинні оброблятися тільки в тому випадку, якщо мета обробки не може бути розумно досягнута іншими способами. (...) Часові рамки повинні бути встановлені контролером для стирання або періодичного перегляду.

включення процедур для періодичного перегляду відповідних наборів даних та умов їх видалення.

Стаття 17.1 встановлює обов'язок видалити дані без невиправданої затримки у таких випадках:

- Персональні дані більше не потрібні для цілей, для яких вони були зібрані або іншим чином оброблені;
- Суб'єкт даних відкликає свою згоду, і така згода є єдиною підставою для такої обробки;
- Суб'єкт даних заперечує проти обробки і немає інших законних підстав, що переважають;
- Суб'єкт даних заперечує проти обробки своїх даних для цілей прямого маркетингу;
- Персональні дані оброблялися незаконно;
- Персональні дані повинні бути видалені для виконання певного юридичного зобов'язання;
- Персональні дані були отримані у зв'язку з наданням будь-якої послуги інформаційного суспільства.

Дані, зібрані для етапу навчання, відповідно до аспектів, викладених у статті 11 GDPR, та принципу мінімізації даних, повинні бути очищені або позбавлені всієї інформації, яка не є строго необхідною для навчання моделі.

Після завершення етапу навчання системи ШІ організація повинна виконати її видалення, якщо дані не використовуються для інших цілей, сумісних із початковими цілями відповідно до положень статті 6.4[81] GDPR. Якщо суб'єкти даних подають запити на видалення, контролер даних повинен приймати рішення на основі конкретних випадків, враховуючи можливі обмеження цього права, передбачені відповідними нормативними актами.

Зокрема, коли система ШІ надається контролерам та фізичним особам, якщо вона включає дані від суб'єктів даних:

- Дані потрібно або видалити, або оцінити, що це частково або повністю неможливо, оскільки це зашкодить системі;
- Повинні бути встановлені відповідні правові підстави для передачі даних третій стороні, особливо коли включаються дані спеціальних категорій;
- Суб'єкти даних повинні бути поінформовані (як зазначено вище);
- Повинні бути надані докази того, що були впроваджені відповідні заходи захисту за замовчуванням і за проєктом (зокрема, мінімізація даних);
- Залежно від ризику для суб'єктів даних та обсягу або категорій даних, провести оцінку впливу на конфіденційність.

Якщо контролер даних зберігає дані суб'єкта даних для цілей персоналізації послуги, що надається системою ШІ, після завершення відносин щодо надання цієї послуги відповідні дані повинні бути видалені.

### *6.1.1. ОБМЕЖЕННЯ НА СТИРАННЯ*

Стаття 17.3 GDPR встановлює певні обмеження на видалення даних. Крім того, стаття 32 LOPDGDD[82] встановлює обов'язок контролера даних блокувати дані, які мають бути виправлені або видалені.

81 Стаття 6.4. Якщо опрацювання з метою, відмінною від тієї, для якої було зібрано персональні дані, не ґрунтується на згоді суб'єкта даних або на законодавстві Союзу чи держави-члена, що є необхідним та пропорційним заходом у демократичному суспільстві для забезпечення цілей, зазначених у статті 23(1), Контролер повинен з'ясувати, чи є опрацювання для іншої мети сумісним з метою, для якої первинно збираються персональні дані, враховуючи, серед іншого (...)

82 LOPDGDD (абревіатура від іспанського «Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales») — це «Закон про захист персональних даних і забезпечення цифрових прав». Цей закон було ухвалено в Іспанії для адаптації Загального регламенту захисту даних (GDPR) до національного законодавства. LOPDGDD забезпечує додаткові положення та гарантії, які стосуються як захисту персональних даних, так і цифрових прав громадян в Іспанії. LOPDGDD не лише стосується питань, пов'язаних із захистом персональних даних, але також включає положення щодо цифрових прав, що робить його комплексним законом для захисту даних і цифрових прав в Іспанії.

## C. БЛОКУВАННЯ ДАНИХ

Блокування є обов'язком контролера даних, єдиною метою якого є забезпечення можливості відповідати на будь-які можливі зобов'язання, що виникають у зв'язку з обробкою даних, для отримання доказів можливого невиконання вимог і виключно протягом строку їхньої дії.

Стаття 32 LOPDGDD визначає заблоковані дані як стан даних, що зберігаються поза межами обробки, застосовуючи будь-які технічні або організаційні заходи, які запобігають будь-якому виду обробки, включаючи перегляд, за винятком випадків, коли такі дані повинні бути надані відповідним судам або суддям, державному прокурору чи компетентним органам державної влади, зокрема органам захисту даних.

Таким чином, необхідність включення зазначених заходів для блокування будь-яких даних, пов'язаних із процесом висновку (або принаймні вхідних даних і результатів), які можуть знадобитися для відповіді на запит або скаргу відповідного суб'єкта даних, повинна розглядатися як вимога при проєктуванні обробки. Ці механізми також стосуються журналів файлів[83], які будуть розглянуті нижче.

## D. ПРАВО НА ВИПРАВЛЕННЯ

Контролер даних зобов'язаний реагувати на здійснення суб'єктами даних їхнього права на виправлення, особливо коли це право виникає з висновків та профілів, створених відповідним рішенням на основі ШІ.

З іншого боку, якщо модель включає неточні навчальні дані, які не впливають на користувача компонента ШІ, і такі неточні дані не можуть бути пов'язані з жодним суб'єктом даних при розповсюдженні рішення на основі ШІ, подання неточних навчальних даних може бути доцільним як частина стратегій абстракції та обфускації, спрямованих на гарантування ефективного застосування[84] принципу мінімізації даних. Якщо такі стратегії запобігають повторній ідентифікації індивіда, посилаючись на вищезгадану статтю 11 GDPR, право на виправлення не буде застосовуватися.

Однак, якщо сама модель включає неточні персональні дані третіх суб'єктів даних, які можуть бути повторно ідентифіковані, і таким чином пов'язані з ними неправильною інформацією, право на виправлення повинно бути дотримане.

83 Файли, що записують всі події, що виконуються в системі.

84 Наприклад, застосування диференційованих методів конфіденційності.

# 7. АВТОМАТИЗОВАНЕ ПРИЙНЯТТЯ РІШЕНЬ ТА ДЕРЖАВНІ ОРГАНИ

Використання систем штучного інтелекту в процесах, які здійснюються державними органами, застосовується вже багато років, і існує кілька прикладів цього:

- Використання систем автоматичного перекладу, які широко застосовуються та доступні в різних установах, навіть у Єврокомісії, або онлайн-сервісах.
- Використання для пошуку інформації пошукових систем в Інтернеті, які підтримуються системами штучного інтелекту.
- Автоматичні транскрипції тексту, мовні коректори тощо.

Важливо розрізняти автоматизацію або робототехніку від використання систем штучного інтелекту. Автоматизація або робототехніка використовується для прискорення публічних процесів, що стосуються формальних дій. Наприклад, коли необхідно вжити дії при настанні закінчення терміну в публічному процесі, іншими словами, коли існує певний період для подання клопотань, і цей період добігає кінця, публічний процес має перейти на наступний етап.

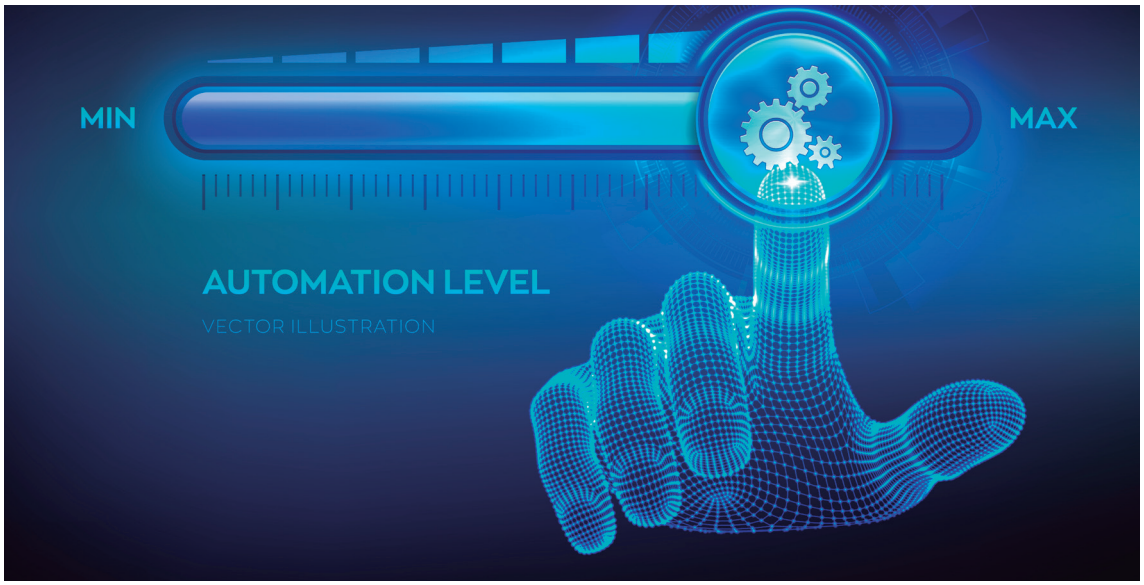
Розглянемо систему штучного інтелекту, яка в рамках процесу (обробки персональних даних) підготує пропозицію або остаточне рішення щодо скарг і апеляцій від громадян. У такому випадку система штучного інтелекту замінює людину, державного службовця, який би виконував деякі операції в процесі обробки.

## 7.1. МІНІМАЛЬНІ ВИМОГИ

Ми можемо замислитися, що вимагаємо від людини, відповідальної за будь-яку операцію в діяльності з обробки даних. Коротко ми б вимагали:

- Щоб людина була кваліфікована для виконання конкретного завдання: перед тим, як приймати такі рішення, вона пройшла акредитоване навчання, а після навчання отримала сертифікацію від незалежної третьої сторони та/або пройшла період стажування під наглядом.
  - Аналогічно, система штучного інтелекту повинна бути перевірена на відповідність конкретному завданню та сертифікована щодо мінімального рівня якості результатів.
  - Початкова робота повинна здійснюватися з дотриманням принципу обережності та під наглядом людини, доки не буде гарантовано правильну роботу.
- Щоб людина не приймала рішень щодо людей або не використовувала дані без легітимності, компетентності або обґрунтування.
  - Документація проєктування системи штучного інтелекту гарантує, що жодні зайві дані, окрім тих, що повинні бути використані в рамках процесу, не беруть участі у процесі прийняття рішень.
  - Принцип мінімізації даних застосовується в дизайні.

- Щоб людина була неупередженою та об'єктивною у своїх рішеннях.
  - Оцінка упереджень у навчальних даних повинна бути доступна в документації проєктування.
  - Тестування упереджень повинне бути доступне в документації проєктування.
  - Аудити в разі еволюції системи штучного інтелекту повинні проводитися періодично.
- Щоб людські рішення були послідовними, передбачуваними та стабільними.
  - Тестування послідовності, передбачуваності та стабільності повинне бути доступне в документації проєктування.
  - Аудит системи штучного інтелекту повинен проводитися періодично.
- Щоб громадяни могли знати та виправляти помилкову інформацію, яку використовує система штучного інтелекту.
  - Повинна бути прозорість щодо даних громадян, що вводяться в систему штучного інтелекту.
  - Повинна бути можливість реалізувати свої права згідно з GDPR.
- Щоб громадянин мав можливість відмовитися від того, щоб така система штучного інтелекту обробляла його дані.
  - Громадянин повинен мати можливість відмовитися від використання принаймні такої системи на основі штучного інтелекту.
- Щоб кожне рішення, як це відбувається в багатьох випадках, було затверджене вищим керівником (процес візування).
  - У процесі прийняття рішень повинна бути людська інтервенція.
- Щоб він був зобов'язаний дотримуватися таємниці щодо даних, які він обробляє, і не використовував їх для інших цілей.
  - Система штучного інтелекту не розкриває дані третім сторонам, або дані використовуються для інших цілей без згоди суб'єкта даних.
- Щоб після завершення роботи з процесом, він не зберігав зайву інформацію або не мав необмеженого доступу до файлів.
  - Система штучного інтелекту не повинна зберігати жодних даних про особу, що може бути ідентифікована або ідентифікується.
- Щоб процес прийняття особистих рішень був інформований і прозорий.
  - Необхідна прозорість реалізована в логіці системи штучного інтелекту.
- Щоб впровадження цього процесу прийняття рішень у загальну процедуру установи або служби (у широкому сенсі, поза індивідуальним процесом) керувалося аналізом ризиків.
  - Впровадження системи штучного інтелекту в рамках процесу має бути оцінено в загальному контексті всієї обробки даних за допомогою DPIA.
- Щоб у процесі збору та зберігання даних була точність.
  - Система штучного інтелекту (поза алгоритмом ШІ) була оцінена на точність вхідних даних.
- Щоб офіцер був достатньо гнучким, коли він стикається з ситуаціями, які виходять за межі стандартних.
  - Аудити в реальному часі та попередження повинні здійснюватися кваліфікованими людськими операторами.
- Щоб рішення могли бути оскаржені.
  - Автоматичні рішення повинні гарантувати процедуру оскарження до того, як вони набудуть незворотного ефекту.
- Щоб існував процес «звітності» про прийняті рішення.




- Повинні бути журнали роботи системи штучного інтелекту.
- Повинні бути журнали інцидентів системи штучного інтелекту.
- Повинні бути як автоматичні, так і ручні, а також регулярні та виняткові аудити.

- Щоб була чітко визначена відповідальність державного службовця та державного органу щодо впливу на громадян щодо результатів їхніх рішень.

  - Розробник або постачальник системи штучного інтелекту повинен за контрактом надати всі документи та інформацію, що дозволяє отримати докази про перевірку та валідацію системи та її експлуатацію.
  - Орган, який вибрав систему штучного інтелекту для впровадження, повинен дотримуватися належної обачності у процесі відбору такої системи, це повинно бути підзвітно та ґрунтуватися на задокументованих спостереженнях.
  - Орган, який впровадив систему штучного інтелекту в процеси (обробка персональних даних), повинен виконувати свої обов'язки як контролер або обробник.
  - Ніхто не може відмовлятися від своєї відповідальності та відповідальності у разі негативного впливу на громадян. Система штучного інтелекту, як і будь-яка інша система, не є відповідальною.

## 7.2. АВТОМАТИЗОВАНІ РІШЕННЯ

Програми, що пропонують або підтримують сервіс на основі рішень ШІ, можуть приймати певні рішення, які впливають на людей, їхнє приватне життя, фізичну безпеку, соціальний статус і взаємодію з іншими особами.

GDPR гарантує право не піддаватися автоматизованим рішенням, включаючи профілювання[85], коли:

- Відсутнє людське втручання. Щоб вважати, що існує людське втручання, відповідне рішення повинно бути проконтрольоване компетентною особою, уповноваженою змінити таке рішення шляхом значущої дії, а не лише символічної.
- З цих рішень виникають юридичні наслідки.
- Або суб'єкт даних зазнає подібного та значного впливу[86].

85 Автоматизовані рішення можуть прийматися з профілюванням або без нього; Профілювання може відбуватися без прийняття автоматизованих рішень. Однак профілювання та автоматизоване прийняття рішень не обов'язково є окремими видами діяльності. Щось, що починається як простий автоматизований процес прийняття рішень, може стати таким, що базується на профілюванні, залежно від того, як використовуються дані. (WP 251)

86 Залежно від відповідного випадку, можна вважати, що такі аспекти мають значний вплив: моніторинг осіб на різних веб-сайтах, пристроях та послугах, зміна очікувань та побажань залучених осіб, зміна способу подання реклами або використання їхніх знань про вразливості суб'єктів даних. (WP 251)

Виняток з цього може бути зроблено, коли обробка:

- Ґрунтується на явній згоді та впроваджені гарантії для захисту прав і свобод.
- Необхідна для укладення або виконання договору, не зачіпає спеціальні категорії даних і включає гарантії для захисту прав і свобод.
- Ґрунтується на правовій базі Європейського Союзу або держави-члена і не включає спеціальні категорії персональних даних.
- Ґрунтується на правовій базі Європейського Союзу або держави-члена і необхідна для захисту фундаментального суспільного інтересу.

Вищезазначений розділ «інформація», а особливо підрозділ «Відповідна інформація про реалізовану логіку», а також розділ «Прозорість» охоплюють вимоги до інформації щодо таких операцій обробки.

Коли правовою підставою обробки є явна згода, контролер повинен спроєктувати обробку таким чином, щоб вона захищала вільний вибір користувачів. Це робиться, по-перше, наданням життєздатних і еквівалентних альтернатив автоматизованим рішенням у момент, коли вимагається їх згода. Крім того, гарантується, що якщо суб'єкт даних вирішує не піддаватися автоматизованим рішенням, то рішення щодо такого суб'єкта даних не буде упередженим і не суперечить інтересам суб'єкта даних. Якщо вищезазначені умови не виконані, згода не може вважатися вільно наданою[87]. Ці альтернативи повинні бути впроваджені на етапі проєктування обробки.

Як кращі практики, і поза будь-якими вимогами, що випливають із захисту даних, людський нагляд може бути обраним варіантом в рамках обробки даних на основі ШІ, і в цілому щодо автоматизованих рішень. Підхід «вимикач мертвої людини»[88] слід уникати в проєктуванні системи: користувачі повинні мати можливість ігнорувати алгоритм у певний час у всіх випадках і документувати ситуації, в яких цей курс привілейований. З цієї причини рекомендується документувати будь-які інциденти або виклики автоматизованих рішень з боку відповідного суб'єкта даних, щоб аналіз дозволяв виявляти ситуації, які вимагають людського втручання, оскільки обробка не працює так, як очікувалося.

Щодо обробки даних, що стосуються неповнолітніх, у Роз'ясненні 71 зазначено, що рішення, засновані виключно на автоматизованій обробці, включаючи профілювання[89], з юридичними або подібними значущими наслідками, не повинні застосовуватися до неповнолітніх. Однак, оскільки ця заборона не відображена у відповідних статтях, вона не вважається абсолютною, і винятки можуть бути розглянуті, коли вони неминучі для захисту добробуту відповідного неповнолітнього і впроваджені відповідні та специфічні для дітей гарантії.

Щодо спеціальних категорій даних, слід зазначити, що автоматичні рішення, побудовані на обробці біометричних даних для ідентифікації і навіть для автентифікації, можуть ґрунтуватися лише на згоді або при застосуванні значного суспільного інтересу і впровадженні відповідних заходів для захисту прав і свобод суб'єкта даних і законних інтересів.

87 «Звіт автоматизованого суспільства 2019» описує автоматичні методи відбору кандидатів на основі згоди, які вимагають доступу суб'єкта даних до своєї адреси електронної пошти, щоб алгоритм оцінював їхній поштовий трафік і, таким чином, отримував профіль як майбутнього працівника. Якщо альтернативи немає, або просте подання резюме тягне за собою штраф у процесі відбору, згода не є дійсною.

88 Перемикач мерця / Dead man switch

89 Висновок WP29 02/2013 щодо додатків на розумних пристроях (WP 202), прийнятий 27 лютого 2013 року, щодо розділу 3.10 конкретно для дітей, вказує на сторінці 26, що «Зокрема, контролери даних не повинні обробляти дані дітей для цілей поведінкової реклами, ні прямо, ні опосередковано, оскільки це вийде за рамки розуміння дитини і, отже, перевищить межі законної обробки».

## 7.3. ЛЮДСЬКЕ ВТРУЧАННЯ

GDPR встановлює в статті 22 право не піддаватися рішенню, заснованому виключно на автоматизованій обробці, включаючи профілювання, яке спричиняє юридичні наслідки або значно впливає на суб'єкта даних. У цій статті наведені винятки, що супроводжуються певними гарантіями, які ми не будемо аналізувати в цьому тексті.

Як уже зазначалося, той факт, що система ШІ включає автоматизоване рішення без людського втручання, не є характеристикою самої системи ШІ, а скоріше характеристикою обробки, в яку буде включено одну або кілька систем машинного навчання (ML). Іншими словами, це рішення відповідальної особи за обробку при проєктуванні цієї обробки, яким чином буде прийматися рішення на основі результату однієї з її операцій, у цьому випадку реалізованої за допомогою автоматизованої обробки.

Щоб рішення було засноване виключно на автоматизованій обробці, людське втручання має бути усунуто з процесу прийняття рішення. Коли йдеться про визначення людського втручання, необхідно уточнити, що залучена особа повинна мати повноваження для зміни цього рішення, здатність до дії і повноваження для нагляду за цим рішенням.

Наприклад, система ML у контексті кримінального провадження, яка оцінює вирок ув'язненого, підлягатиме людському втручанню, коли результат системи ML буде оцінений суддею з необхідними часом і ресурсами для аналізу висновку, і суддя матиме мандат для цього в рамках процесу. З іншого боку, якщо результат системи ML передається для виконання безпосередньо виконавцю, це не можна вважати людським втручанням у процес прийняття рішення. Виконавець виконує, але не матиме необхідних знань для оскарження рішення і повноважень для цього. У цьому випадку виконавець буде ще однією частиною механізму виконання рішення.

У цьому контексті слід враховувати одне з найважливіших упереджень, яке може виникнути у системі машинного навчання (ML), що надає підтримку в прийнятті рішень або профілюванні осіб. Це упередження не обов'язково впливає на саму систему ML, і його складно усунути за допомогою проєктування. Воно впливає на людей, які використовують та інтерпретують результати, надані системою ML. Когнітивні упередження виникають через схильність людей обирати легші шляхи в процесі мислення, щоб не витрачати зайві зусилля на прийняття рішень. Одне з найнебезпечніших упереджень у відношенні до «інтелектуальних» систем - це упередження авторитету: схильність дотримуватися правил і підкорятися тим, кого вважають авторитетними в групі. Мало що так важко оскаржувати, як авторитет машини або системи, позначеної як «інтелектуальна».

Це упередження складно контролювати, оскільки воно стосується людей. Щоб уникнути цього, необхідно навчати операторів, стандартизувати процес прийняття рішень і забезпечувати нагляд за людьми, залученими у різні фази обробки даних.


Омбудсман України
Ombudsman of Ukraine


Фінансується
Європейським Союзом


FIIAPP
COOPERACIÓN ESPAÑOLA


Ministry
of Digital Transformation
of Ukraine


aepd
agencia
española
protección
datos