DOI 10.34185/1562–9945–5–166–2026–10
УДК: 004.056; 004.62

І.В. Клименко, Є.А. Лебідь

МЕТОДИ ЗАХИСТУ ТА ПСЕВДОНІМІЗАЦІЇ ПЕРСОНАЛЬНИХ ДАНИХ У СИСТЕМАХ ФОРМУВАННЯ КРОС–ФУНКЦІОНАЛЬНИХ КОМАНД ІТ–ПРОЕКТІВ

Анотація. В статті запропоновано комплексний аналіз безпекових вимог, методів і технологічних засобів захисту персональних даних у системах управління персоналом та формування крос–функціональних команд ІТ–проєктів. Визначено основні категорії даних, що використовуються під час формування профілів фахівців, включно з ідентифікаційними відомостями, професійним досвідом, «hard skills», «soft skills», ролями, результатами оцінювання та інформацією про участь у попередніх проєктах. Систематизовано основні ризики, пов'язані з несанкціонованим доступом, витоками даних, надмірними правами користувачів, повторною ідентифікацією суб'єктів та використанням персональної інформації поза межами визначеної мети оброблення. Розглянуто організаційно–процесні механізми захисту, технічні засоби автентифікації та авторизації, рольові й атрибутні моделі контролю доступу, принцип мінімальних привілеїв, аудит операцій, механізми запобігання витоку даних (Data Loss Prevention DLP) та засоби моніторингу операцій із персональними даними. Проведено порівняльний аналіз основних методів трансформації та знеособлення персональних ідентифікаційних даних (Personally Identifiable Information, PII), зокрема маскування, токенізації, псевдонімізації, анонімізації та диференційної приватності, з урахуванням рівня конфіденційності, можливості контрольованої повторної ідентифікації, збереження структури даних, придатності до подальшого аналізу, машинного навчання та data mining (DM), а також складності практичної реалізації. Показано, що для профілів потенційних кандидатів критичним є не лише зменшення ризику розкриття ідентичності користувачів, але й збереження аналітичної цінності професійних і поведінкових характеристик, необхідних для їх оцінювання та підтримки кадрових рішень. Обґрунтовано доцільність використання псевдонімізації як базового механізму розділення ідентифікаційного та аналітичного контурів оброблення даних. Запропоновано концептуальний підхід до формування та оцінки захищеного профілю фахівця, за якого прямі персональні ідентифікатори вилучаються з аналітичного контуру та замінюються контрольованими псевдонімами, тоді як професійні характеристики, навички, досвід і рольові ознаки залишаються доступними для алгоритмів аналізу та формування команд. Окрему увагу приділено автоматизованому виявленню PII із використанням правил, регулярних виразів, методів обробки природної мови (Natural

© Клименко І.В., Лебідь Є.А., 2026
Copyright for this paper by its authors. Use permitted under License CC BY 4.0.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Language Processing, NLP), машинного навчання та комбінованих механізмів ідентифікації чутливих атрибутів.

Ключові слова: комп'ютерні системи, інформаційні технології, аналіз даних, захист персональних даних, PII, псевдонімізація, контроль доступу, DLP, машинне навчання.

Постановка проблеми. На сьогоднішній день цифровізація процесів управління персоналом супроводжується стрімким зростанням обсягів даних, що використовуються для оцінювання професійного досвіду, компетентностей, ролей, результатів роботи та інших характеристик фахівців, а також активним впровадженням інтелектуальних засобів підтримки кадрових рішень [1–3]. Із стрімким поширенням систем штучного інтелекту, питання запобігання несанкціонованому розкриттю, повторній ідентифікації, дискримінаційним або непрозорим автоматизованим рішенням щодо людини та для забезпечення законності, підзвітності й довіри до оброблення даних є критично важливим [17]. За таких умов персональні дані дедалі частіше стають складовою частиною комплексних аналітичних профілів, що використовуються не лише для адміністративного обліку, але й для автоматизованого добору кандидатів, оцінювання їх відповідності вимогам проєктів. Водночас розширення складу таких профілів, інтеграція даних із різних джерел і використання алгоритмічних методів аналізу підвищують ризики несанкціонованого доступу, витоку інформації, надмірного використання персональних відомостей та повторної ідентифікації суб'єктів даних [2, 3]. Сучасні підходи до зниження зазначених ризиків передбачають комплексне застосування організаційно–процесних механізмів, засобів автентифікації й авторизації, рольових та атрибутних моделей контролю доступу, механізмів запобігання витоку даних (DLP), журналювання та моніторингу операцій із чутливими даними [4, 5]. Однак для систем, у яких персональні дані безпосередньо використовуються в аналітичних процедурах, лише обмеження доступу є недостатнім, оскільки воно не усуває необхідності мінімізувати обсяг ідентифікаційної інформації, доступної алгоритмам оброблення. Це зумовлює необхідність застосування методів трансформації та знеособлення даних, зокрема маскування, токенізації, анонімізації, псевдонімізації та диференційної приватності, ефективність яких визначається не лише рівнем конфіденційності, але й можливістю зберегти структуру, зв'язність та аналітичну корисність даних [6–9]. Додатковим напрямом розвитку відповідних систем є автоматизоване виявлення персональних ідентифікаційних даних (PII) із використанням правил, методів обробки природної мови (NLP), розпізнавання іменованих сутностей (Named Entity Recognition, NER) та машинного навчання, що дозволяє включати механізми ідентифікації чутливих атрибутів безпосередньо до контуру оброблення даних [10]. У сукупності це формує багатокомпонентне завдання, пов'язане з вибором і обґрунтуванням таких механізмів захисту персональних даних, які одночасно забезпечують належний рівень конфіденційності та мінімально впливають на придатність даних для подальшого аналітичного використання.

Аналіз останніх досліджень і публікацій, присвячених захисту персональних даних у цифрових системах управління персоналом та аналітичної підтримки кадрових

рішень, вказує на необхідність комплексного поєднання організаційних та алгоритмічних механізмів забезпечення конфіденційності [1–10]. Окремий напрям досліджень пов'язаний із побудовою моделей розмежування доступу до даних різного рівня чутливості та застосуванням рольових і атрибутних механізмів контролю, які забезпечують диференційоване надання прав відповідно до функцій користувачів, характеристик даних та контексту їх оброблення [4]. Водночас у роботах, присвячених захисту від витоків інформації, підкреслюється необхідність доповнення механізмів контролю доступу засобами запобігання витоку даних (DLP), моніторингу операцій та виявлення внутрішніх загроз, оскільки санкціонований доступ до інформації сам по собі не виключає можливості її неправомірного використання або передачі [5]. Важливе місце займають дослідження методів трансформації та знеособлення даних, у яких аналізуються можливості анонімізації, псевдонімізації, маскування та інших підходів з урахуванням ризику повторної ідентифікації та втрати інформаційної корисності [6–9]. При цьому результати сучасних досліджень свідчать, що забезпечення максимально можливого рівня знеособлення не завжди є оптимальним для аналітичних систем, оскільки надмірна трансформація структури або зв'язків між записами може обмежувати можливості подальшого машинного навчання та підтримки рішень [7–9]. Перспективним напрямом є також автоматизація виявлення персональних PII із використанням правил, NLP–методів, NER та машинного навчання, що дозволяє інтегрувати процедури ідентифікації та трансформації чутливих атрибутів безпосередньо до контуру обробки інформації [10]. Сукупність розглянутих робіт засвідчує, що сучасні підходи здебільшого досліджують окремі складові захисту персональних даних (контроль доступу, DLP, анонімізацію, псевдонімізацію або автоматизоване виявлення PII), тоді як питання їх узгодженого застосування в інформаційних системах управління персоналом залишається опрацьованим недостатньо. Зокрема, невирішеною частиною загальної проблеми є формування такого підходу до захисту персональних даних кандидатів, який забезпечував би відокремлення ідентифікаційної інформації від аналітичного профілю, можливість контрольованої повторної ідентифікації за визначених умов та водночас збереження професійних, поведінкових і рольових характеристик, необхідних для подальшого аналізу та формування команд.

Таким чином, метою дослідження є розробка комплексного підходу до захисту персональних даних у системах обробки і управління профілями кандидатів, орієнтованого на автоматизоване виявлення персональних ідентифікаційних даних, обґрунтований вибір методів їх трансформації та псевдонімізації, а також збереження аналітичної цінності професійних і поведінкових характеристик фахівців.

Викладення основного матеріалу дослідження. Реалізація поставленої мети передбачає послідовне визначення безпекових вимог до обробки HR–даних, формування набору організаційних і технічних механізмів захисту, порівняльний аналіз методів знеособлення, побудову захищеного аналітичного профілю фахівця та оцінювання ефективності запропонованих рішень (рис. 1).

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Цифровізація процесів управління персоналом суттєво розширила склад і обсяг інформації, що використовується під час оцінювання фахівців та прийняття кадрових рішень. Сучасні HR–системи поєднують традиційні облікові дані працівників із результатами професійного оцінювання, відомостями про компетентності, досвід роботи, участь у проєктах, виконувані ролі та інші характеристики, які можуть використовуватися методами HR–аналітики й алгоритмічної підтримки прийняття рішень [1–3]. У системах управління персоналом така інформація набуває додаткової цінності, оскільки на її основі здійснюється зіставлення профілів фахівців із вимогами конкретного IT–проєкту, ранжування кандидатів та оцінювання можливості їх спільної роботи.

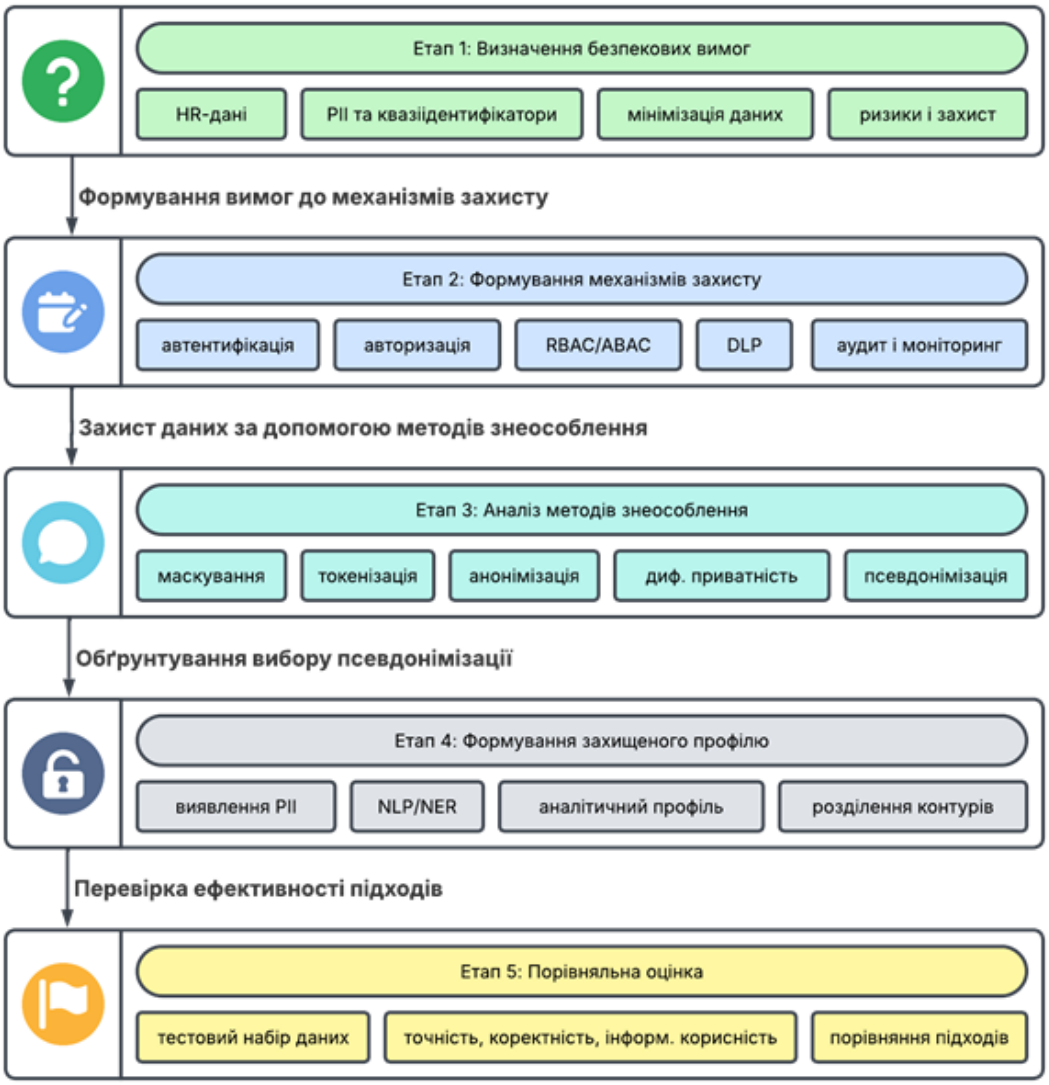


Рисунок 1 – Послідовність етапів захисту персональних даних у системах формування крос–функціональних команд

Узагальнений профіль фахівця в межах такого підходу доцільно розглядати як сукупність декількох взаємопов'язаних груп даних. До першої групи належать безпосередні ідентифікаційні відомості: ім'я та прізвище, адреса електронної пошти, номер телефону, ідентифікатори працівника, облікові записи та інші атрибути, що безпосередньо дозволяють встановити особу. Друга група охоплює професійні характеристики

– освіту, досвід роботи, технічні компетентності (Hard Skills; HSK), поведінкові та комунікативні характеристики (Soft Skills; SSK), поточну та попередні ролі, участь у проєктах, результати професійного оцінювання та інші ознаки, що характеризують придатність фахівця до виконання певних завдань. Окрему групу становлять службові та контекстні відомості, зокрема організаційна приналежність, доступність, історія переміщень між проєктами та взаємодії з іншими членами команд, результати тестів на визначення психотипу кандидата тощо. Сукупність таких даних дає можливість сформувати достатньо детальний аналітичний профіль, але одночасно підвищує ризики порушення конфіденційності [2, 3].

З позиції інформаційної безпеки зазначені дані доцільно розділяти на три категорії:

- прямі ідентифікатори;
- квазіідентифікатори;
- неідентифікаційні професійні характеристики.

Прямі ідентифікатори однозначно або з високою ймовірністю дозволяють встановити особу без використання додаткової інформації. Квазіідентифікатори окремо можуть не забезпечувати однозначної ідентифікації, проте їх комбінація здатна суттєво звузити множину можливих суб'єктів і створити умови для повторної ідентифікації [6, 7]. До таких атрибутів у HR–домені можуть належати, наприклад, рідкісна спеціалізація, унікальна комбінація технологій, досвід роботи над конкретним проєктом, організаційний підрозділ або поєднання посади та професійного стажу. Неідентифікаційні професійні характеристики безпосередньо не встановлюють особу, однак мають ключове значення для процедур аналізу та формування команд. Тому їх повне вилучення або надмірна трансформація може суттєво знизити інформаційну корисність набору даних.

Основні загрози в процесі обробки персональних даних пов'язані не лише з несанкціонованим отриманням доступу до інформації, але й із надмірними правами легітимних користувачів, неконтрольованим копіюванням та передачею відомостей, використанням інформації поза початково визначеною метою, дискримінаційним або непрозорим рішенням щодо кандидата, а також можливістю повторної ідентифікації осіб на основі взаємопов'язаних атрибутів [5, 6]. Особливістю систем формування команд є те, що значна частина професійних даних повинна залишатися доступною для аналітичних алгоритмів навіть тоді, коли безпосередня ідентифікація кандидата на відповідному етапі не обов'язкова. Внаслідок цього традиційна модель, за якої захист забезпечується переважно через обмеження доступу до всього запису працівника, є недостатньо гнучкою.

Формування безпекових вимог до вищезазначених систем доцільно здійснювати на основі наступних принципів [4, 11]:

- мінімізація даних;
- обмеження мети оброблення (Purpose Limitation; PL);
- мінімально необхідні привілеї (Least Privileges; LP);

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

- функціональне розмежування доступу.

Принципи мінімізації даних, обмеження мети та мінімально необхідних привілеїв передбачають використання лише тих атрибутів і прав доступу, які є необхідними для виконання конкретної функції, з урахуванням ролі користувача, типу даних і рівня їх чутливості [4]. Для HR–систем це безпосередньо пов'язано з доцільністю розділення ідентифікаційного та аналітичного контурів: перший забезпечує контрольований доступ до даних, необхідних для встановлення особи, тоді як другий оперує переважно професійними, поведінковими та рольовими характеристиками, необхідними для ранжування кандидатів і формування команди.

Такий підхід зменшує обсяг ідентифікаційної інформації, доступної аналітичним алгоритмам, і створює передумови для застосування псевдонімізації, за якої прямі ідентифікатори замінюються контрольованими псевдонімами зі збереженням можливості авторизованого відновлення відповідності між профілем і конкретною особою [8, 9]. Відповідно, до базових функціональних вимог належать: класифікація PII, розмежування доступу, мінімізація використання прямих ідентифікаторів, контроль витоків, трансформація персональних атрибутів та збереження аналітичної корисності професійних характеристик. Реалізація цих вимог потребує комплексного застосування організаційних, технічних та алгоритмічних механізмів, що розглядаються на наступному етапі дослідження.

Забезпечення конфіденційності персональних даних у системах управління персоналом потребує поєднання організаційних, технічних та програмних механізмів, оскільки окреме використання будь–якого одного засобу не дозволяє повною мірою контролювати всі етапи життєвого циклу чутливої інформації. У сучасних HR–системах дані про працівників і кандидатів можуть надходити з різних джерел, зберігатися у декількох інформаційних контурах, передаватися між прикладними сервісами та використовуватися як безпосередньо користувачами, так і автоматизованими аналітичними модулями. За таких умов механізми захисту повинні охоплювати не лише початкове надання доступу до даних, але й контроль подальших операцій із ними, аудит, запобігання витокам та обмеження використання інформації поза межами визначених функцій [4, 11].

Базовий рівень захисту формують організаційно–процесні механізми, які визначають правила роботи з персональними даними, відповідальність учасників та допустимі сценарії їх оброблення. До таких механізмів належать політика доступу, класифікація інформації за рівнями чутливості, розподіл відповідальності між адміністраторами, HR–фахівцями, керівниками проєктів та операторами аналітичних систем, а також регламентація процедур отримання, використання, оновлення, архівування і видалення даних. Важливим елементом є управління життєвим циклом інформації, оскільки ризики порушення конфіденційності виникають не лише під час активного використання профілю, але й у процесах резервного копіювання, експорту, передачі між системами та зберігання історичних записів [5, 11].

Політики обробки персональних даних повинні бути безпосередньо пов'язані з принципами мінімізації та обмеження мети їх використання. Для кожної категорії ко-

ристувачів або програмних компонентів доцільно визначати не лише перелік доступних атрибутів, але й допустимі операції над ними та контекст, у якому такий доступ дозволяється. Наприклад, аналітичному модулю формування команд може бути необхідна інформація про досвід, технологічний стек, компетентності та попередні ролі кандидата, але не його ім'я, контактні дані або інші прямі ідентифікатори. Такий підхід дозволяє перейти від традиційної моделі доступу до цілісного профілю працівника до більш гранульованого керування окремими групами атрибутів. Для HR–систем особливо важливим є розмежування доступу до різних категорій даних, оскільки право перегляду загальної професійної інформації не повинно автоматично означати право доступу до контактних, фінансових, оціночних або інших чутливих відомостей.

Одним із найбільш поширених підходів до такого розмежування є рольове керування доступом (Role–Based Access Control; RBAC), за якого права призначаються відповідно до організаційної ролі користувача. У межах HR–домену такими ролями можуть бути HR–менеджер, рекрутер, керівник проєкту, системний адміністратор або оператор аналітичної підсистеми. Перевагою RBAC є відносна простота адміністрування та зрозумілість політик, однак для складних інформаційних систем рольової моделі може бути недостатньо, якщо рішення про доступ залежить від додаткових умов [4].

Більш гнучким є атрибутне керування доступом (Attribute–Based Access Control, ABAC), у якому рішення формується на основі сукупності характеристик суб'єкта, ресурсу, операції та контексту виконання [4]. Такий підхід дозволяє враховувати, наприклад, організаційний підрозділ користувача, тип HR–даних, рівень їх чутливості, належність кандидата до конкретного проєкту, час або місце виконання операції. Для систем формування крос–функціональних команд ABAC є особливо перспективним, оскільки дозволяє формувати політики, за яких керівнику проєкту доступний аналітичний профіль кандидата, але приховані ідентифікаційні атрибути, тоді як HR–фахівець за наявності відповідних повноважень може отримувати повний запис.

Незалежно від моделі розмежування доступу, важливим залишається принцип мінімально необхідних привілеїв, який обмежує доступ до критичних атрибутів лише тими користувачами й програмними компонентами, для яких він є функціонально необхідним [4, 11]. Доповненням до цього виступають аудит і журналювання операцій із персональними даними, що забезпечують фіксацію доступу, змін, експорту та інших дій і дозволяють виявляти відхилення від стандартних сценаріїв роботи. Оскільки санкціонований доступ не виключає можливості неправомірного використання інформації, важливим складником захисту є механізм запобігання витоку даних, орієнтований на контроль передачі та виявлення чутливих відомостей у файлах, повідомленнях і потоках даних [5]. Для HR–систем їх доцільно поєднувати з класифікацією атрибутів за рівнем чутливості та моніторингом нетипової активності, зокрема масового завантаження профілів, надмірної кількості звернень до чутливих полів або різкого збільшення обсягів експорту [5]. Разом із тим RBAC/ABAC, DLP, аудит і моніторинг переважно визначають, хто і за яких умов отримує доступ до даних, але не змінюють сам зміст доступних записів. Тому навіть коректно авторизований аналітичний модуль може отри-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

мувати надлишкові ідентифікаційні відомості, не потрібні для формування команди. Це зумовлює необхідність застосування додаткових механізмів, орієнтованих уже на трансформацію самих персональних даних.

Таким чином, забезпечення конфіденційності персональних даних у системах управління персоналом не може бути зведене виключно до контролю доступу, шифрування або запобігання витокам. Для аналітичних систем необхідним є додатковий рівень захисту, орієнтований на трансформацію самого набору даних і мінімізацію кількості ідентифікаційних атрибутів, що потрапляють до аналітичного контуру. Такий рівень може реалізовуватися шляхом маскування, токенізації, псевдонімізації, анонімізації та застосування методів диференційної приватності. Порівняння зазначених підходів із позицій рівня захисту, оборотності перетворення та збереження інформаційної корисності становить наступний етап дослідження.

Застосування контролю доступу, DLP, аудиту та моніторингу знижує ризики несанкціонованого використання персональних даних, однак не усуває необхідності захищати самі записи, що надходять до аналітичного контуру. У HR–системах це зумовлює пошук компромісу між конфіденційністю та збереженням інформаційної корисності даних [6–9]. До базових методів захисної трансформації належать маскування, токенізація, псевдонімізація та анонімізація. Маскування передбачає повне або часткове приховування значень і характеризується простотою реалізації, однак може бути недостатнім за наявності квазіідентифікаторів та ускладнювати подальше зіставлення записів [6, 7]. Токенізація замінює чутливі значення штучними токенами та дозволяє підтримувати сталі ідентифікатори, але потребує захищеного сервісу. Псевдонімізація передбачає відокремлення ідентифікаційної інформації від аналітичного профілю зі збереженням можливості контрольованого відновлення відповідності [8, 11]. Для систем управління персоналом це дозволяє використовувати компетентності, досвід, ролі та інші професійні характеристики без передачі прямих ідентифікаторів до аналітичної підсистеми, а після завершення ранжування – встановити конкретного кандидата авторизованими засобами [8]. Анонімізація орієнтована на практично незворотне усунення можливості ідентифікації суб’єкта. Вона може бути ефективною для статистичного аналізу й агрегованих HR–даних, однак посилення анонімізації зменшує інформаційну корисність і не завжди виключає ризик повторної ідентифікації за квазіідентифікаторами [6, 7, 9]. Оскільки незворотна втрата зв’язку між кандидатом і його профілем недопустима, то застосування анонімізації не може бути базовим механізмом для розглядуваної системи [6, 9].

Окремим класом методів є диференційна приватність, що спрямована не стільки на трансформацію окремого запису, скільки на обмеження інформації, яку можна отримати про конкретного суб’єкта. Концептуально механізм диференційної приватності забезпечує близькість розподілів результатів для наборів даних, що відрізняються одним записом. Для сусідніх наборів D і D' механізм M задовольняє ε–диференційну приватність, якщо для будь–якої множини можливих результатів S:

$$\Pr[M[D] \in S] \leq e^{\varepsilon} \Pr[M[D'] \in S], \qquad (1)$$

де параметр ε визначає допустимий рівень розкриття інформації: зменшення його значення посилює гарантії приватності, але, як правило, збільшує спотворення результату та знижує аналітичну точність [13, 14]. Водночас для задачі безпосереднього ранжування кандидатів роль диференційної приватності є більш обмеженою. Додавання шуму до індивідуальних професійних характеристик здатне змінювати порядок кандидатів або призводити до помилкової оцінки відповідності вимогам проєкту. Сучасні дослідження диференційної приватності також підкреслюють фундаментальний компроміс між силою захисту та корисністю результатів аналізу [13, 14]. Тому в досліджуваній системі цей підхід доцільно розглядати передусім як додатковий механізм для агрегованої HR–аналітики, статистичної звітності або окремих процедур машинного навчання, а не як основний засіб формування робочих профілів.

Для формалізованого зіставлення розглянутих підходів доцільно використовувати сукупність критеріїв, які відображають не лише безпекові властивості, але й функціональні вимоги системи формування команд. До основних критеріїв належать рівень забезпечення конфіденційності, збереження інформаційної корисності, керована оборотність трансформації, придатність до інтеграції з процедурами машинного навчання та глибинний аналіз даних (DM), а також складність і ресурсні витрати реалізації. Узагальнену оцінку методу m можна подати як

$$Q_m = w_1 P_m + w_2 U_m + w_3 R_m + w_4 I_m - w_5 C_m, \qquad (2)$$

де P_m визначає рівень забезпечення конфіденційності, U_m – збереження корисності даних, R_m – керована оборотність, I_m – придатність до інтеграції з аналітичними процедурами, C_m – складність і витрати реалізації, а набір $\{w_1, \dots, w_5\}$ – вагові коефіцієнти критеріїв.

Для забезпечення інтерпретованості інтегральної оцінки доцільно нормувати часткові критерії в єдиному діапазоні $P_m, U_m, R_m, I_m, C_m \in [0,1]$, а вагові коефіцієнти задавати з умовою

$$\sum_{j=1}^{n} w_j = 1, w_j \geq 0, j = 1..5. \qquad (3)$$

Від’ємний знак перед C_m відображає те, що зростання складності та витрат знижує загальну доцільність застосування методу, тоді як збільшення інших чотирьох показників є бажаним. При цьому інтегральний показник Q_m не слід трактувати як універсальну характеристику методу знеособлення: його значення визначається конкретним сценарієм використання та системою ваг. Для задачі формування команд відносно високі ваги мають отримувати U_m, R_m та I_m, оскільки захищений набір повинен залишатися придатним для аналізу, підтримувати зв’язність профілю в різних процедурах та забезпечувати контрольований перехід від рекомендованого профілю до конкретного фахівця.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

За такої постановки максимізація конфіденційності не є єдиною метою вибору методу. Надмірне зменшення інформаційної деталізації може підвищувати P_m, але одночасно знижувати U_m та I_m, тоді як повна необоротність трансформації фактично мінімізує R_m. Саме тому для аналітичних HR–систем важливим є не максимальне усунення інформації про суб'єкта, а контрольоване відокремлення його ідентичності від аналітичного профілю. У цьому контексті псевдонімізація забезпечує найбільш збалансоване поєднання конфіденційності, збереження структури та зв'язності даних, можливості їх повторного аналітичного використання і керованого відновлення відповідності. Це дає підстави розглядати її як базовий механізм формування захищеного профілю фахівця.

Проведений аналіз показав, що для систем управління персоналом доцільним є поєднання автоматизованого виявлення персональних ідентифікаційних даних із їх подальшою псевдонімізацією. Такий підхід дозволяє вилучити прямі ідентифікатори з аналітичного контуру, зберігши професійні та поведінкові характеристики, необхідні для ранжування кандидатів і формування команд. Псевдонімізація при цьому забезпечує збереження зв'язності записів і можливість контрольованого встановлення відповідності між аналітичним профілем та конкретною особою [8, 11]. Автоматизоване виявлення PII доцільно реалізовувати як комбіновану процедуру. Для структурованих і формалізованих атрибутів, як то номерів телефонів, адрес електронної пошти, ідентифікаційних номерів, а також інших даних із визначеним синтаксисом, ефективними є правила, регулярні вислови та словникові механізми. Для неструктурованих текстових полів, зокрема резюме, описів досвіду, характеристик і результатів оцінювання, своєю чергою, доцільно застосовувати методи оброблення природної мови, розпізнавання іменованих сутностей та машинного навчання. Сучасні дослідження підтверджують ефективність гібридних схем, у яких механізми на основі правил (Rule–Based; RB) поєднуються з NLP, NER та ML для контекстно залежного виявлення персональної інформації [10].

Якість виявлення PII може оцінюватися стандартними показниками класифікації, точністю (Precision), повнотою (Recall) та F1–мірою:

$$\begin{cases} Precision = \dfrac{TP}{TP + FP} \\ Recall = \dfrac{TP}{TP + FN} \\ F_1 = 2 \cdot \dfrac{Precision \cdot Recall}{\text{Precision} + Recall} \end{cases}, \qquad (4)$$

де величина TP визначає кількість правильно виявлених PII–елементів; FP – кількість помилково ідентифікованих елементів; FN – кількість пропущених PII. Використання сукупності цих показників дозволяє врахувати як помилкове віднесення звичайних атрибутів до чутливих, так і пропуск персональної інформації.

Для формалізації подальшої обробки вихідний профіль кандидата C_i можна представити у вигляді $C_i = (ID_i, PII_i, H_i, S_i, E_i, R_i)$, де ID_i – внутрішній ідентифікатор, PII_i – сукупність персональних ідентифікаційних даних, H_i та S_i – множини «hard skills» і «soft skills», E_i – характеристики професійного досвіду, R_i – рольові характеристики. Після застосування функції псевдонімізації $\varphi(\cdot)$ можна визначити $PID_i = \varphi(ID_i, PII_i)$ і сформувати аналітичний профіль $D_i^A = (PID_i, H_i, S_i, E_i, R_i)$, у якому PID_i є псевдонімом кандидата, що підтримує однозначне зв'язування його записів у межах аналітичних процедур без безпосереднього розкриття особи. Додаткова інформація, необхідна для відновлення відповідності, зберігається окремо від аналітичного профілю, що відповідає загальному принципу псевдонімізації персональних даних [8, 11].

Запропонований підхід передбачає логічне розмежування ідентифікаційного та аналітичного контурів. В ідентифікаційному контурі зберігаються PII та засоби встановлення відповідності між псевдонімом і конкретною особою, тоді як до аналітичного контуру передаються лише псевдонімізовані профілі D_i^A. Алгоритми ранжування та формування команд працюють із професійними, поведінковими й рольовими характеристиками кандидатів без необхідності доступу до їхніх прямих ідентифікаторів. Реідентифікація виконується лише після завершення аналітичної процедури для відібраних кандидатів та за наявності відповідних прав доступу.

Узагальнена послідовність обробки профілів кандидатів при формуванні команди зображена на рис. 2.



Рисунок 2 – Узагальнена послідовність обробки профілів кандидатів

Практична реалізація такого контуру може бути виконана із застосуванням «Google Cloud Sensitive Data Protection», що підтримує автоматизоване виявлення чутливої інформації за вбудованими та користувацькими детекторами, включаючи словники, регулярні вислови та контекстні правила, а також операції редагування, маскування й криптографічної псевдонімізації [15]. Для зворотної псевдонімізації сервіс підтримує детерміноване криптографічне перетворення з використанням AES–SIV, яке дозволяє отримувати стабільні токени та виконувати контрольовану реідентифікацію за наявності відповідного ключа [16]. Така реалізація відповідає запропонованому розділенню контурів: аналітичні компоненти оперують псевдонімізованими значеннями, а криптографічні ключі та процедури відновлення ідентичності залишаються доступними лише захищеному компоненту системи.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Таким чином, запропонований підхід поєднує автоматизоване виявлення PII, їх контрольовану псевдонімізацію та функціональне розмежування ідентифікаційної й аналітичної обробки. Це дозволяє мінімізувати поширення прямої персональної інформації в системі без втрати зв'язності професійних профілів.

Оцінка того, наскільки запропонований підхід дозволяє мінімізувати доступ аналітичної підсистеми до персональних ідентифікаційних даних без істотного зниження інформаційної корисності профілів має охоплювати два взаємопов'язані аспекти: ефективність виявлення та усунення PII з аналітичного контуру і збереження характеристик, необхідних для подальшого ранжування кандидатів та формування команд [7–10]. Для проведення порівняльного аналізу доцільно сформувати синтетичний або попередньо знеособлений набір профілів кандидатів. Як зазначалося вище, до складу профілю мають входити прямі ідентифікатори, контактні відомості, посада, професійний досвід, технологічний стек, «hard skills», «soft skills», рольові характеристики та історія участі у проєктах (рис. 3).

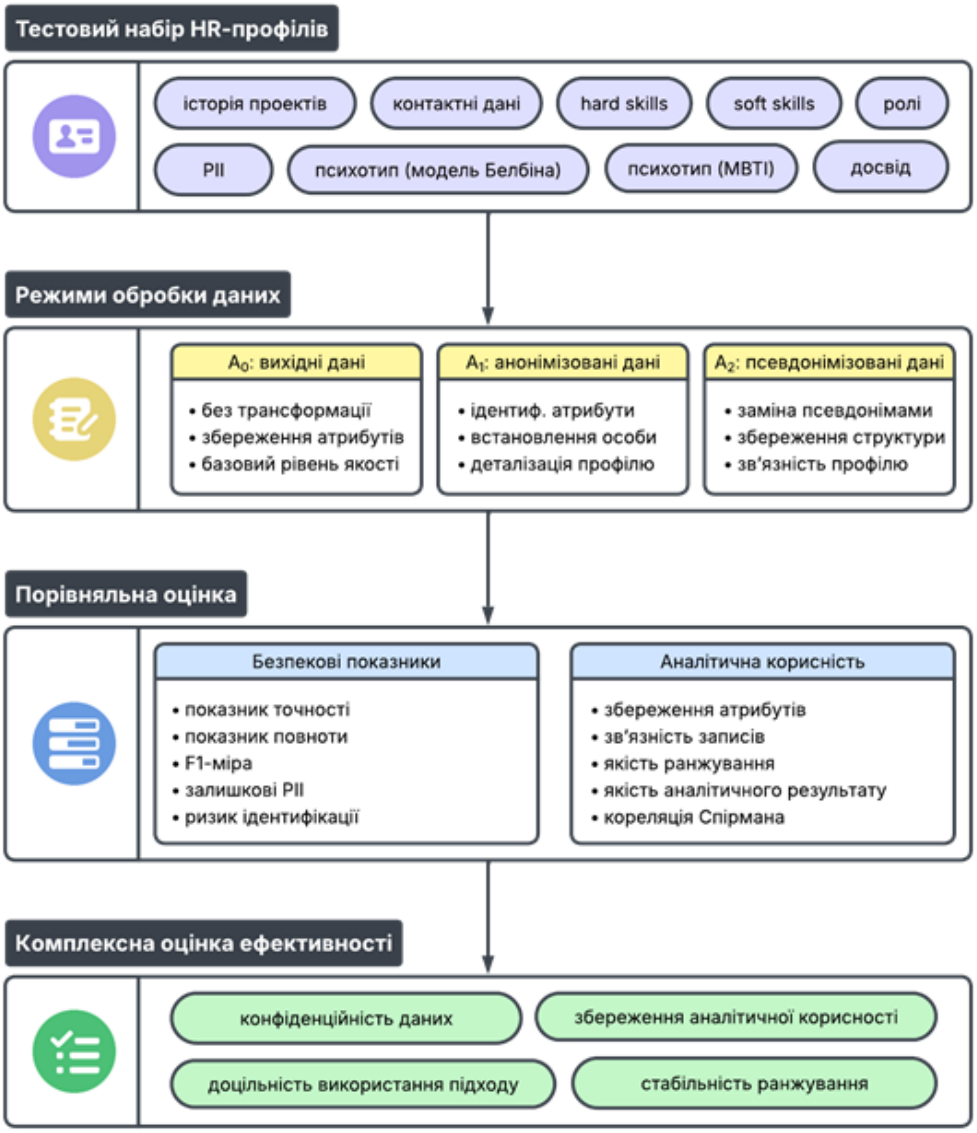


Рисунок 3 – Процес оцінки ефективності псевдонімізації профілів кандидатів

Використання синтетичного набору на початковому етапі дозволяє повністю контролювати склад PII та квазіідентифікаторів, формувати еталонну розмітку для перевірки алгоритмів виявлення чутливих даних і водночас уникати залучення реальних персональних відомостей. Передбачається формування тестового набору HR–профілів із подальшим порівнянням трьох режимів оброблення даних: вихідного, анонімізованого та псевдонімізованого. Оцінювання виконується за двома групами критеріїв – безпековими показниками та характеристиками аналітичної корисності, що дозволяє одночасно врахувати рівень захисту PII, збереження структури й атрибутів профілю та стабільність результатів ранжування кандидатів. Сукупність цих показників формує основу для комплексного оцінювання ефективності та визначення доцільності застосування режиму у системах формування крос–функціональних команд.

У межах поставленої задачі доцільно порівнювати щонайменше три режими оброблення даних:

- A_0: вихідні дані;
- A_1: анонімізовані дані;
- A_2: псевдонімізовані дані.

Режим A_0 використовується як базовий і відображає результати аналітичної обробки вихідного набору без захисної трансформації. Режим A_1 характеризує більш жорстке знеособлення зі зменшенням або втратою частини зв'язків між атрибутами, тоді як A_2 відповідає запропонованому підходу з відокремленням ідентифікаційних даних та збереженням псевдонімізованого аналітичного профілю. За потреби може бути введено додатковий режим маскування або токенізації, однак для базового порівняння трьох зазначених сценаріїв достатньо.

Оцінювання результатів доцільно здійснювати за двома групами метрик. Перша група характеризує ефективність захисту персональних даних. До неї належать точність, повнота та F1–міра автоматизованого виявлення PII, частка персональних ідентифікаційних атрибутів, що залишилися доступними після трансформації, а також можливість контрольованого відновлення відповідності між псевдонімізованим профілем і конкретною особою [10]. Для анонімізованих даних додатково може враховуватися оцінка залишкового ризику повторної ідентифікації [6, 7, 9]. Другу групу становлять показники збереження аналітичної корисності даних. До них доцільно віднести частку збережених професійних атрибутів, можливість підтримання зв'язків між записами, стабільність ранжування кандидатів, зміну якості рекомендацій щодо складу команди та додаткові часові або обчислювальні витрати, пов'язані із застосуванням захисної трансформації. Такий поділ дозволяє уникнути ситуації, коли метод оцінюється лише за рівнем конфіденційності без урахування його практичної придатності для подальшої аналітичної обробки.

Зміну якості аналітичного результату після застосування методу захисту можна визначити як $\Delta Q_m = Q_m^{prot} - Q^{orig}$, де значення Q_m^{prot} є показником якості ранжуван-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

ня або рекомендації після застосування методу m, а Q^{orig}, своєю чергою, відповідний показник для вихідних даних. Для зручності порівняння різних наборів або метрик також може використовуватися відносна зміна $\delta Q_m = \left(\left(Q_m^{prot} - Q^{orig}\right)/Q^{orig}\right) \cdot 100\%$. Значення ΔQ_m, близьке до нуля, свідчитиме про збереження аналітичної якості після захисної трансформації, тоді як значне від'ємне відхилення вказуватиме на втрату інформаційної корисності. Для псевдонімізованого режиму A_2 бажаним результатом є поєднання мінімальної доступності прямих ідентифікаторів у аналітичному контурі з $|\Delta Q_2| \to 0$. У такому випадку подальша оцінка може підтвердити, що захист досягається без істотного впливу на функціональну придатність профілів.

Для оцінювання збереження структури ранжування доцільно додатково застосувати коефіцієнт рангової кореляції Спірмена:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{n(n^2 - 1)}, \qquad (5)$$

де d_i – різниця між позиціями i–го кандидата у ранжуваннях, отриманих для вихідного та трансформованого наборів, а n – кількість кандидатів. Значення ρ, близьке до 1, означатиме, що захисна трансформація практично не змінила порядок кандидатів. Такий показник є особливо доречним для розглядуваної задачі, оскільки дозволяє оцінювати не лише загальну точність рекомендацій, а й стабільність самого результату ранжування.

Таким чином, вищезазначений підхід передбачає послідовне порівняння вихідних, анонімізованих і псевдонімізованих профілів кандидатів за безпековими та аналітичними критеріями і дозволяє використовувати його як основу для подальшої реалізації безпекових вимог у системах обробки персональних даних, в тому числі у системах формування крос–функціональних команд IT–проєктів.

Висновки. У результаті проведеного дослідження було сформовано комплексний підхід до захисту персональних даних у системах управління персоналом, орієнтований на поєднання безпекових механізмів із збереженням аналітичної корисності професійних і поведінкових характеристик фахівців. Основні результати дослідження включають в себе:

1. Систематизовано склад і безпекові характеристики HR–даних, що використовуються під час формування профілів кандидатів, та визначено прямі ідентифікатори, квазіідентифікатори і професійні атрибути, необхідні для подальшого аналітичного оброблення.

2. Визначено основні вимоги до захисту персональних даних, що включають мінімізацію оброблюваної інформації, обмеження мети її використання, принцип мінімально необхідних привілеїв, розмежування доступу, аудит операцій, DLP–механізми та моніторинг нетипової активності.

3. Проведено порівняльний аналіз методів трансформації та знеособлення PII, зокрема маскування, токенізації, псевдонімізації, анонімізації та диференційної приват-

ності, з урахуванням рівня конфіденційності, збереження інформаційної корисності, керованої оборотності, придатності до аналітичної обробки та складності реалізації.

4. Обґрунтовано доцільність використання псевдонімізації як базового механізму для систем формування команд, оскільки вона забезпечує контрольоване відокремлення ідентичності кандидата від його аналітичного профілю зі збереженням зв'язності даних і можливості авторизованої зворотної ідентифікації.

5. Запропоновано підхід до автоматизованого виявлення та псевдонімізації PII, що поєднує правила і регулярні вислови для формалізованих атрибутів із методами NLP, NER та машинного навчання для аналізу неструктурованих даних.

6. Сформовано концептуальний підхід, що передбачає порівняння вихідних, анонімізованих і псевдонімізованих HR–даних за безпековими та аналітичними показниками, включаючи Precision, Recall, F1–міру, залишкові PII, збереження атрибутів, стабільність ранжування та коефіцієнт рангової кореляції Спірмена.

Запропонований підхід створює методичну основу для побудови захищеної підсистеми аналізу профілів кандидатів, у якій зменшення доступності персональної ідентифікаційної інформації поєднується зі збереженням характеристик, необхідних для формування крос–функціональних команд. Практична реалізація підходу та результати порівняльного аналізу режимів анонімізації й псевдонімізації буде розглянуто в подальших дослідженнях.

ЛІТЕРАТУРА / REFERENCES

1. Álvarez–Gutiérrez, F. J., Stone, D. L., Castaño, A. M., & García–Izquierdo, A. L. (2022). Human resources analytics: A systematic review from a sustainable management approach. *Journal of Work and Organizational Psychology, 38*(3), 129–147. https://doi.org/10.5093/jwop2022a18.

2. Edwards, M. R., Charlwood, A., Guenole, N., & Marler, J. (2024). HR analytics: An emerging field finding its place in the world alongside simmering ethical challenges. *Human Resource Management Journal, 34*(2), 326–336. https://doi.org/10.1111/1748–8583.12435.

3. Chatterjee, S., Chaudhuri, R., Vrontis, D., & Siachou, E. (2022). Examining the dark side of human resource analytics: An empirical investigation using the privacy calculus approach. *International Journal of Manpower, 43*(1), 52–74. https://doi.org/10.1108/IJM–02–2021–0087.

4. Tall, A. M., & Zou, C. C. (2023). A framework for attribute–based access control in processing big data with multiple sensitivities. *Applied Sciences, 13*(2), 1183. https://doi.org/10.3390/app13021183.

5. Herrera Montano, I., García Aranda, J. J., Ramos Diaz, J., Molina Cardín, S., de la Torre Díez, I., & Rodrigues, J. J. P. C. (2022). Survey of techniques on data leakage protection and methods to address the insider threat. *Cluster Computing, 25*, 4289–4302. https://doi.org/10.1007/s10586–022–03668–2.

6. Weitzenboeck, E. M., Lison, P., Cyndecka, M., & Langford, M. (2022). The GDPR and unstructured data: Is anonymization possible? International Data Privacy Law, 12(3), 184–206. https://doi.org/10.1093/idpl/ipac008.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

7. Tomás, J., Rasteiro, D., & Bernardino, J. (2022). Data anonymization: An experimental evaluation using open–source tools. Future Internet, 14(6), 167. https://doi.org/10.3390/fi14060167.
8. Abu Attieh, H., Müller, A., Wirth, F. N., & Prasser, F. (2025). Pseudonymization tools for medical research: A systematic review. BMC Medical Informatics and Decision Making, 25, 128. https://doi.org/10.1186/s12911–025–02958–0.
9. Schneider, M. J., Bailie, J., & Iacobucci, D. (2025). Why data anonymization has not taken off. Customer Needs and Solutions, 12, 9. https://doi.org/10.1007/s40547–025–00158–5.
10. Mishra, K., Pagare, H., & Sharma, K. (2025). A hybrid rule–based NLP and machine learning approach for PII detection and anonymization in financial documents. *Scientific Reports*, *15*, 22729. https://doi.org/10.1038/s41598–025–04971–9.
11. European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L 119*, 1–88.
12. National Institute of Standards and Technology. (2020). *Security and Privacy Controls for Information Systems and Organizations* (NIST Special Publication 800–53, Revision 5). U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800–53r5.
13. Huang, W., Zhuo, M., Zhu, T., Zhou, S., & Liao, Y. (2023). Differential privacy: Review of improving utility through cryptography–based technologies. *Concurrency and Computation: Practice and Experience, 35*(5), e7565. https://doi.org/10.1002/cpe.7565.
14. Slavković, A., & Seeman, J. (2023). Statistical data privacy: A song of privacy and utility. *Annual Review of Statistics and Its Application, 10*, 189–218. https://doi.org/10.1146/annurev–statistics–033121–112921.
15. Google Cloud. (2026). *Sensitive Data Protection documentation*. Google Cloud Documentation.
16. Google Cloud. (2026). *De–identify and re–identify sensitive data*. Google Cloud Documentation.
17. Sartor G., Lagioia F. (2020). *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. European Parliamentary Research Service, Scientific Foresight Unit (STOA). PE 641.530. https://doi.org/10.2861/293. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf

Received 11.06.2026.
Accepted 13.08.2026.
Published 30.10.2026

Methods for protecting and pseudonymizing personal data in systems for forming cross–functional IT project teams

Abstract. The article presents a comprehensive analysis of the security requirements, methods, and technological tools for protecting personal data in human resource management systems and systems for forming cross–functional teams for IT projects. The main categories of data used to create professional profiles are identified, including identification data,

professional experience, hard skills, soft skills, roles, assessment results, and information about participation in previous projects. The principal risks associated with unauthorized access, data breaches, excessive user privileges, re–identification of data subjects, and the use of personal information beyond the specified purpose of processing are systematized. Organizational and procedural protection mechanisms, technical means of authentication and authorization, role–based and attribute–based access control models, the principle of least privilege, activity auditing, Data Loss Prevention (DLP) mechanisms, and tools for monitoring operations involving personal data are examined. A comparative analysis of the main methods for transforming and de–identifying Personally Identifiable Information (PII), including masking, tokenization, pseudonymization, anonymization, and differential privacy, is conducted with regard to the level of confidentiality, the possibility of controlled re–identification, preservation of data structure, suitability for further analysis, machine learning, and data mining (DM), as well as the complexity of practical implementation. The study demonstrates that, for potential candidate profiles, it is essential not only to reduce the risk of disclosing users' identities but also to preserve the analytical value of professional and behavioral characteristics required for candidate assessment and support of personnel–related decisions. The feasibility of using pseudonymization as a basic mechanism for separating the identification and analytical data–processing layers is substantiated. A conceptual approach to the creation and evaluation of a protected professional profile is proposed, in which direct personal identifiers are removed from the analytical layer and replaced with controlled pseudonyms, while professional characteristics, skills, experience, and role–related attributes remain available to algorithms for analysis and team formation. Particular attention is paid to the automated detection of PII using rules, regular expressions, Natural Language Processing (NLP) methods, machine learning, and combined mechanisms for identifying sensitive attributes.

Keywords: computer systems, information technologies, data analysis, personal data protection, PII, pseudonymization, access control, DLP, machine learning.

Клименко Іван Вікторович – доцент, к.е.н., доцент кафедри комп'ютерних і інформаційних технологій Українського державного університету науки і технологій.
ORCID: https://orcid.org/0000–0001–5149–3974
Лебідь Євген Андрійович – аспірант кафедри комп'ютерних і інформаційних технологій Українського державного університету науки і технологій.
ORCID: https://orcid.org/0009–0007–4277–2083

Klymenko Ivan – Associate Professor, Candidate of Economics Sciences, Associate Professor of the Department of Computer Information Technologies, Ukrainian State University of Science and Technology.
ORCID: https://orcid.org/0000–0001–5149–3974
Lebid Yevhen – PhD Student of the Department of Computer Information Technologies, Ukrainian State University of Science and Technology.
ORCID: https://orcid.org/0009–0007–4277–2083

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)