МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ І ПРИКЛАДНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

САПОЖНІКОВА ВЕРОНІКА ЄВГЕНІВНА

Допускається до захисту:
в.о. завідувача кафедри
інформаційних технологій
д-р. техн. наук, професор
_____ Наталія ВЕСЕЛОВСЬКА
«_____» _____ 2026 р.

МОДЕЛЮВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ МОНІТОРИНГУ ГЕНДЕРНО ЗУМОВЛЕНИХ СИТУАЦІЙ

Спеціальність 122 «Комп’ютерні науки»

Кваліфікаційна (бакалаврська) робота

Керівник:
Тетяна БАЦЕНКО, доцент кафедри
інформаційних технологій,
к. т. н., доцент
_____
(Підпис)

Оцінка _____/_____/_____
(бали/за шкалою ЄКТС/за національною шкалою)
Голова ЕК: _____
(Підпис)

Вінниця 2026

АНОТАЦІЯ

Сапожнікова В.Є. Моделювання інформаційної системи моніторингу гендерно зумовлених ситуацій. Спеціальність 122 «Комп'ютерні науки», освітня програма «Комп'ютерні науки». Донецький національний університет імені Василя Стуса, Вінниця 2026.

У кваліфікаційній (бакалаврській) роботі було розроблено модель інформаційно-аналітичної системи моніторингу гендерно зумовлених ситуацій (гендерно зумовленого насильства, ГЗН) у нотації IDEF0. Реалізовано аналітичний компонент з використанням Power BI для виявлення тенденцій, груп ризику та прихованої статистики звернень з метою підтримки управлінських рішень координаторів соціальних служб та неурядових організацій.

Ключові слова: гендерно зумовлені ситуації, ГЗН, IDEF0, Power BI, соціальний моніторинг, підтримка прийняття рішень.

65 с., 17 рис., 3 табл., 50 джерел.

ABSTRACT

Sapozhnikova V.E. Modeling an information system for monitoring gender-based violence cases. Specialty 122 «Computer Science», educational program «Computer Science». Vasyl Stus Donetsk National University, Vinnytsia 2026.

This work addresses the challenge of fragmented and analytically limited data on gender-based violence (GBV) by developing a model of an integrated information-analytical monitoring system using IDEF0 notation. The analytical component, implemented in Power BI, enables identification of regional risk patterns, vulnerable population groups, and gaps in official reporting to support evidence-based decision-making by social service coordinators and non-governmental organisations.

Keywords: gender-based violence, GBV, IDEF0, Power BI, social monitoring, decision support.

65 p., 17 fig., 3 tables, 50 sources.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ ТЕРМІНІВ

ГЗН - Гендерно зумовлені ситуації (гендерно зумовлене насильство);

НУО - Неурядова організація;

ЦСР - Цілі сталого розвитку;

BI - Business Intelligence (бізнес-аналітика);

BPMN - Business Process Model and Notation (нотація моделювання бізнес-процесів);

CEDAW - Convention on the Elimination of All Forms of Discrimination Against Women (Конвенція про ліквідацію всіх форм дискримінації щодо жінок);

CSV - Comma-Separated Values (значення, розділені комами);

DASH - Domestic Abuse, Stalking and Honour-Based Violence (контрольний список виявлення ризиків домашнього насильства);

DAX - Data Analysis Expressions (вирази аналізу даних);

DFD - Data Flow Diagram (діаграма потоків даних);

EIGE - European Institute for Gender Equality (Європейський інститут гендерної рівності);

ETL - Extract, Transform, Load (вилучення, перетворення, завантаження);

FRA - European Union Agency for Fundamental Rights (Агентство ЄС з основних прав);

GBVIMS - Gender-Based Violence Information Management System (інформаційна система управління даними про гендерно зумовлене насильство);

ICOM - Input, Control, Output, Mechanism (вхід, управління, вихід, механізм);

IDEF0 - Integration Definition for Function Modeling (методологія функціонального моделювання);

IPV - Intimate Partner Violence (насильство від інтимного партнера);

IRC - International Rescue Committee (Міжнародний комітет порятунку);

MARAC - Multi-Agency Risk Assessment Conference (міжвідомча конференція з оцінки ризику);

OCHA - Office for the Coordination of Humanitarian Affairs (Управління ООН з координації гуманітарних справ);

UNHCR - United Nations High Commissioner for Refugees (Верховний комісар ООН у справах біженців);

UNFPA - United Nations Population Fund (Фонд ООН у галузі народонаселення).

ВСТУП

Актуальність теми дослідження. Гендерно зумовлені ситуації, або гендерно зумовлене насильство (ГЗН), є системною соціальною проблемою, що зачіпає осіб незалежно від віку, гендерної ідентичності та соціального становища. Незважаючи на наявність міжнародних правових механізмів - зокрема Конвенції Ради Європи про запобігання насильству стосовно жінок і домашньому насильству та боротьбу з цими явищами (Стамбульська конвенція) та Конвенції про ліквідацію всіх форм дискримінації щодо жінок (CEDAW) - реальний масштаб явища залишається прихованим через розрізненість джерел даних та феномен прихованої статистики (недообліку) звернень. Дані національної поліції, медичних закладів та неурядових організацій (НУО) функціонують ізольовано, у несумісних форматах, без аналітичного компоненту, що унеможливлює виявлення тенденцій, оцінку груп ризику та формування обґрунтованих управлінських рішень. Існуючі інформаційні системи моніторингу ГЗН забезпечують лише реєстрацію справ у межах однієї інституції та не надають інтегрованого аналітичного інструментарію для підтримки прийняття рішень координаторами соціальних служб та НУО. Зазначене обумовлює доцільність моделювання інформаційно-аналітичної системи моніторингу ГЗН, аналітичний шар якої здатний продемонструвати як інтеграція даних з кількох інституційних джерел може слугувати основою для міжінституційної співпраці державних органів та неурядових організацій у сфері протидії ГЗН та підвищення ефективності розподілу ресурсів підтримки.

Мета дослідження. Розробити модель інформаційно-аналітичної системи моніторингу гендерно зумовлених ситуацій та реалізувати її аналітичний компонент для підтримки прийняття управлінських рішень.

Завдання дослідження.

1. Проаналізувати природу гендерно зумовлених ситуацій та існуючі інформаційні системи моніторингу для виявлення їх структурних обмежень.

2. Обґрунтувати вибір методології IDEF0 (Integration Definition for Function Modeling) для моделювання інформаційно-аналітичної системи моніторингу.
3. Запропонувати структуру інтегрованого датасету як модель вхідних даних заданої системи.
4. Побудувати модель інформаційної системи моніторингу в нотації IDEF0.
5. Розробити аналітичну модель у середовищі Power BI для підтримки прийняття управлінських рішень.

Об'єкт дослідження. Процес моніторингу гендерно зумовлених ситуацій на основі інтегрованих даних з різнорідних джерел.

Предмет дослідження. Модель інформаційно-аналітичної системи моніторингу гендерно зумовлених ситуацій та інструментарій обробки і аналізу інтегрованих масивів даних.

Методи дослідження. Системний аналіз, функціональне моделювання (IDEF0), аналіз даних засобами Business Intelligence (Power BI).

Практичне значення одержаних результатів. У роботі вперше запропоновано модель інформаційно-аналітичної системи моніторингу гендерно зумовлених ситуацій, що інтегрує дані з кількох інституційних джерел та застосовує розширену класифікацію інцидентів. Практичне значення полягає у тому, що розроблена модель забезпечує перехід від простої реєстрації інцидентів до управління на основі даних. Створений аналітичний компонент дозволяє виявляти регіональні закономірності, приховані групи ризику та розриви між джерелами звернень. Це надає координаторам соціальних служб та неурядових організацій інструментарій для підтримки прийняття обґрунтованих управлінських рішень - від цільового розподілу фінансових ресурсів до стратегічного планування відкриття кризових центрів у виявлених зонах ризику.

Апробація результатів дослідження. Результати кваліфікаційної (бакалаврської) роботи апробовано на трьох наукових конференціях:

1. X Міжнародна наукова конференція студентів, аспірантів та молодих вчених «Topical Issues of Humanities, Technical, and Natural Sciences», яка

відбулась 30 квітня 2026 року на базі Донецького національного університету імені Василя Стуса. Тези на тему: «Gender-Based Violence in Ukraine in the Context of Armed Conflict» (сертифікат учасника, сертифікат переможника).

2. Фаховий електронний журнал категорії Б «Наука і техніка сьогодні» № 5(59) 2026. Стаття на тему: «Аналіз систем моніторингу гендерно зумовленого насильства» (довідка про прийняття до публікації).
3. VI Всеукраїнська науково-практична конференція «Прикладні інформаційні технології» (ПІТ 2026). Тези на тему: «Вибір методу функціонального моделювання для інформаційно-аналітичних систем моніторингу».

Структура кваліфікаційної роботи. Робота складається зі вступу, трьох розділів, висновків та списку використаних джерел.

У першому розділі проаналізовано природу гендерно зумовлених ситуацій як соціального явища, досліджено існуючі інформаційні системи моніторингу та обґрунтовано вибір методу моделювання IDEF0.

У другому розділі розглянуто теоретичні засади побудови інформаційно-аналітичної системи: сформульовано вимоги до процесів обробки даних, визначено склад вихідного датасету, визначено аналітичний інструментарій системи (Power BI і мова DAX) та розкрито теоретичні засади його ключових компонентів.

У третьому розділі представлено практичну реалізацію системи: побудовано модель у нотації IDEF0 та базуючись на синтетичному датасеті побудовано аналітичний компонент у середовищі Power BI з інтерпретацією отриманих результатів.

Робота містить 65 сторінок, 17 рисунків, 3 таблиці та список літератури з 50 джерел.

РОЗДІЛ 1
ТЕОРЕТИЧНІ ЗАСАДИ ТА ОГЛЯД СИСТЕМ МОНІТОРИНГУ ГЕНДЕРНО ЗУМОВЛЕНИХ СИТУАЦІЙ

1.1 Концептуальні засади та класифікація форм гендерно зумовлених ситуацій

Гендерно зумовлені ситуації або гендерно зумовлене насильство (ГЗН), є однією з найпоширеніших форм порушення прав людини у світі. Всесвітня організація охорони здоров'я визначає насильство як навмисне застосування фізичної сили або влади, фактичне або у вигляді погрози, спрямоване проти себе, іншої особи, групи чи спільноти, що призводить або з великою ймовірністю може призвести до травми, смерті, психологічної шкоди, порушення розвитку або позбавлення [1]. У контексті гендерної нерівності таке насильство набуває специфічного характеру; воно вчиняється проти особи внаслідок її гендерної ідентичності або соціальної ролі, що відводиться їй суспільством.

Ключовим міжнародно-правовим документом у сфері протидії ГЗН є Конвенція Ради Європи про запобігання насильству стосовно жінок і домашньому насильству та боротьбу з цими явищами (Стамбульська конвенція), ухвалена у 2011 році та чинна з 2014 року [2]. Конвенція є першим юридично обов'язковим міжнародним інструментом, що забезпечує всебічну правову основу для протидії ГЗН. Принципово важливим є те, що Стамбульська конвенція запроваджує чітке розмежування між поняттями «стать» (sex) та «гендер» (gender): стать є біологічною характеристикою, тоді як гендер визначається як соціально сконструйовані ролі, поведінка, діяльність та атрибути, які суспільство вважає відповідними для жінок та чоловіків [2]. Це розмежування має принципове значення для систем моніторингу ГЗН. Врахування гендерної ідентичності жертви, не лише біологічної статі, дозволяє зафіксувати повний спектр потерпілих, включно з особами, насильство щодо яких залишається поза офіційною статистикою.

Конвенція визначає такі форми насильства, що підлягають криміналізації:

- психологічне насильство;
- переслідування (сталкінг);
- фізичне насильство;
- сексуальне насильство включно з зґвалтуванням;
- примусовий шлюб;
- каліцтво жіночих статевих органів;
- примусовий аборт та примусова стерилізація;
- сексуальні домагання [2].

Паралельно з цим, Комітет з ліквідації дискримінації щодо жінок (CEDAW) визнав ГЗН формою дискримінації за ознакою статі та зобов'язав держави вживати заходів для її усунення на законодавчому, інституційному та практичному рівнях [3].

Окремим чинником, що загострює прояви ГЗН та змінює ландшафт вразливості, є стан збройного конфлікту. В умовах воєнних дій руйнуються захисні інституційні механізми, зростає соціальна напруга та посилюється економічна залежність жертв, що разом суттєво збільшує ризики насильства та ще більше обмежує доступ до служб підтримки [4].

Масштаб ГЗН залишається критично великим попри десятиліття міжнародних зусиль. За даними Євроінституту гендерної рівності (EIGE) та Агентства ЄС з основних прав (FRA), отриманими в рамках загальноєвропейського опитування на основі 114 023 інтерв'ю, кожна третя жінка в ЄС стикалася з фізичним або сексуальним насильством починаючи з 15-річного віку [5]. При цьому лише чверть жертв повідомляє про пережите насильство до поліції, медичних або соціальних служб [5].

Феномен прихованої статистики звернень (недообліку) є ключовою перешкодою для ефективного моніторингу ГЗН. Офіційна статистика фіксує лише незначну частину реальних випадків через страх помсти, соціальну стигматизацію, недовіру до інституцій та відсутність доступу до служб підтримки [6]. Як зазначає EIGE, офіційні статистичні дані зазвичай базуються на зверненнях до поліції та судових справах і тому охоплюють лише частку від

реального рівня поширеності явища [6]. Це означає, що будь-яка система моніторингу, яка спирається виключно на одне джерело даних, систематично недооцінює масштаб проблеми. Саме тому інтеграція кількох інституційних джерел є не просто бажаною, а методологічно необхідною умовою для побудови комплексної інформаційно-аналітичної системи моніторингу ГЗН.

1.2 Аналіз існуючих інформаційних систем моніторингу гендерно зумовлених ситуацій

Існуючі інформаційні системи у сфері моніторингу ГЗН можна систематизувати за трьома інституційними категоріями, що фундаментально відрізняються за архітектурою, моделлю даних, можливостями інтеграції та аналітичним потенціалом: державні правоохоронні системи, транснаціональні статистичні платформи та інструменти неурядових організацій. Жодна з цих категорій не проєктувалась з урахуванням міждоменної інтеграції або аналітичного синтезу даних з різних джерел, що унеможливлює підтримку прийняття управлінських рішень на основі повної картини явища [7].

Державні правоохоронні системи належать до категорії операційних систем управління справами, що характеризуються транзакційною архітектурою, реляційною моделлю даних, орієнтованою на обробку транзакцій в реальному часі (OLTP), та суворим розмежуванням доступу [8]. Найбільш технічно задокументованою правоохоронною платформою ГЗН в Європі є іспанська система VioGén, що функціонує з 2007 року під управлінням Міністерства внутрішніх справ Іспанії [9]. Система координує дії поліції, органів юстиції, пенітенціарних установ та соціальних служб у межах єдиної централізованої платформи; станом на 2018 рік вона містила понад 510 000 справ та налічувала близько 30 000 авторизованих користувачів [10]. Ключовою функцією VioGén є оцінка ризику повторного насильства за допомогою двох стандартизованих опитувальників: VPR (Valoración Policial del Riesgo - поліцейська оцінка ризику) та VPER (Valoración Policial de Evolución del Riesgo - поліцейська оцінка еволюції ризику). Алгоритм VPR є статичною логістичною

регресією, навченою на історичних даних, без механізму автоматичного перенавчання [11]. Жорстка диадична схема "одна жертва - один кривдник" архітектурно унеможливлює фіксацію будь-яких форм насильства поза межами гетеросексуального насильства від інтимного партнера [10]. Система не має аналітичного шару для виявлення тенденцій та не інтегрує дані медичних закладів чи неурядових організацій [9].

Серед інших правоохоронних систем слід відзначити британський фреймворк DASH/MARAC (Domestic Abuse, Stalking and Honour-Based Violence), що застосовується в Англії та Уельсі. На відміну від централізованої архітектури VioGén, контрольний список DASH не інтегрований у жодну централізовану національну базу даних - він реалізується незалежно кожним регіональним поліцейським підрозділом [12]. Міжвідомча координація в рамках конференцій MARAC (Multi-Agency Risk Assessment Conference) здійснюється через захищений обмін електронною поштою, а не через синхронізацію баз даних у реальному часі [13]. Це забезпечує операційну гнучкість, але призводить до фрагментації даних між регіонами та унеможливлює загальнонаціональний аналіз.

Транснаціональні статистичні платформи представляють принципово інший клас інформаційних систем - сховища статистичних даних, оптимізовані для аналітичного зчитування, а не транзакційного оновлення [14]. База даних гендерної статистики Євроінституту гендерної рівності (EIGE) застосовує класичну зіркову схему: центральна таблиця фактів обчислених індикаторів пов'язана з вимірами географії, часового періоду та визначень показників [15]. Ця архітектура забезпечує ефективні аналітичні запити, але структурно унеможливлює аналіз на рівні індивідуальних записів та субнаціональну деталізацію. База даних UN Women Data Hub відтворює схожу архітектуру у глобальному масштабі, доповнюючи її байєсівською ієрархічною моделлю для оцінки індикатора ЦСР 5.2.1 (частка жінок, які зазнали насильства з боку інтимного партнера) у країнах без прямих даних опитувань [16]. Обидві платформи мають структурно вбудований часовий лаг від одного до двох років

між збором даних та їх публікацією, що унеможливлює оперативне реагування [7].

Неурядові організації формують третю категорію акторів у сфері моніторингу ГЗН і часто першими фіксують випадки насильства, зокрема ті, що не потрапляють до офіційної статистики і складають «темну цифру» (dark figure) неврахованої злочинності [17; 18]. Найбільш поширеною спеціалізованою платформою в гуманітарних контекстах є Gender-Based Violence Information Management System (GBVIMS), розроблена спільно OCHA, UNHCR та IRC і нині активна у понад 25 країнах [19]. Система підтримує офлайн-функціональність для польового збору даних без мережевого з'єднання, що є критично важливим у гуманітарних умовах [20]. Проте кожне розгортання GBVIMS функціонує як ізольована система - міжорганізаційна агрегація здійснюється через ручний процес звітування, а не через автоматизовану інтеграцію даних [19; 20].

Для збору первинних польових даних НУО широко використовують спеціалізовані інструменти цифрового збору - зокрема KoBoToolbox, розроблений Гарвардською гуманітарною ініціативою спільно з OCHA та IRC [21]. KoBoToolbox є безкоштовною відкритою платформою, що дозволяє створювати анкети, збирати дані в умовах обмеженого доступу до інтернету та експортувати результати у стандартизованих форматах, зокрема CSV та XLS [21]. Саме цей формат є стандартним вихідним форматом даних у гуманітарному секторі та слугує основою для подальшої аналітичної обробки. Однак локальні платформи НУО, як правило, функціонують ізольовано - дані зберігаються у несумісних форматах, не інтегруються з державними реєстрами та не мають спільного аналітичного шару [7].

Для систематичного порівняння всіх трьох категорій систем застосовано матрицю оцінювання за ключовими критеріями, що безпосередньо стосуються вимог до інтегрованого моніторингу ГЗН. Результати порівняльного аналізу наведено в табл. 1.1.

Таблиця 1.1 – Порівняльна матриця оцінювання систем моніторингу ГЗН за ключовими критеріями [7].

Система	Аналітична потужність	Широта класифікації ГЗН	Інтеграція кількох джерел	Виявлення прихованої статистики
VioGén	Відсутня	Вузька (лише IPV)	Відсутня	Відсутнє
DASH/MARAC	Відсутня	Часткова	Відсутня	Відсутнє
EIGE	Висока	Широка (агрег.)	Часткова (SDMX)	Відсутнє
UN Women	Висока	Широка (агрег.)	Часткова (SDMX)	Часткове
GBVIMS	Низька	Часткова	Відсутня	Відсутнє
Локальні НУО	Відсутня	Вузька	Відсутня	Відсутнє

При порівняльному аналізі систем моніторингу ГЗН варто розглядати такі критерії, що безпосередньо визначають здатність системи забезпечувати повну картину явища. Аналітична потужність визначає можливість виявлення тенденцій на агрегованому рівні. Широта класифікації визначає які групи жертв залишаються видимими для інституційного реагування. Інтеграція кількох джерел є необхідною умовою для подолання прихованої статистики. Виявлення прихованої статистики є кінцевою аналітичною метою без якої система відтворює викривлену картину явища [7].

Принципово важливим є те, що прогалини за цими критеріями не є незалежними - вони взаємно підсилюють одна одну. Система що має аналітичну потужність, але працює лише з агрегованими даними, не може виявити регіональні закономірності. Система, що має деталізацію на рівні справ, але охоплює лише один тип насильства, систематично виключає цілі групи жертв зі статистики. Система НУО, що фіксує найширший спектр звернень, але функціонує ізольовано, не дозволяє порівняти свої дані з поліцейськими - саме це порівняння є головним індикатором прихованої статистики [7]. Виявлені

прогалини слугують основою для формування функціональних вимог до моделі інформаційно-аналітичної системи моніторингу ГЗН, адже жодна з існуючих систем не подолала структурну взаємозалежність.

1.3 Обґрунтування вибору методу функціонального моделювання інформаційно-аналітичної системи моніторингу

Системний аналіз як методологічна основа проєктування складних інформаційних систем передбачає застосування методів декомпозиції та формалізації структури, що дозволяє послідовно переходити від загального опису системи до деталізації її окремих компонентів і створення аналітичного інструментарію підтримки прийняття рішень [22]. На етапі концептуального проєктування інформаційно-аналітичної системи моніторингу ГЗН вибір методу функціонального моделювання є принциповим рішенням, що визначає повноту опису системи та якість подальших проєктних рішень [23]. У науковій літературі виділяють кілька підходів до функціонального моделювання інформаційних систем, що відрізняються за рівнем абстракції, нотацією та областю застосування. Для порівняльного аналізу обрано три методи - DFD, BPMN та IDEF0, які охоплюють три принципово різних підходи: моделювання потоків даних, моделювання виконуваних процесів та функціональне моделювання систем.

Діаграма потоків даних (DFD, Data Flow Diagram) є інструментом структурного аналізу, що відображає рух інформації між процесами, сховищами та зовнішніми сутностями системи. Метод виник у контексті розробки програмного забезпечення наприкінці 1970-х років і здобув широке застосування завдяки наочності та простоті нотації. Зовнішні сутності - джерела та споживачі даних - позначаються прямокутниками з подвійним контуром, процеси відображаються еліпсами, а сховища даних - відкритими прямокутниками. DFD добре підходить для деталізації інформаційних потоків між компонентами програмної системи та є зрозумілим для широкого кола зацікавлених сторін. Метод підтримує ієрархічну декомпозицію та є незалежним від деталей

технічної реалізації, що робить його придатним для ранніх етапів проєктування. Разом з тим DFD не передбачає формального механізму для опису зовнішніх обмежень та регуляторних впливів на функції системи [24].

Нотація моделювання бізнес-процесів (BPMN, Business Process Model and Notation) є стандартом Object Management Group (OMG), орієнтованим на опис виконуваних бізнес-процесів [25]. Метод використовує розвинену нотацію: горизонтальні смуги для позначення учасників та їхніх зон відповідальності, шлюзи для розгалуження логіки виконання, події початку та завершення процесу, а також пунктирні стрілки для міжорганізаційних потоків повідомлень. BPMN є зрозумілим як для бізнес-аналітиків так і для розробників програмного забезпечення, і добре підходить для моделювання складної міжорганізаційної взаємодії з чіткими ролями виконавців. Він також дозволяє детально описати умови переходу між станами процесу та логіку прийняття рішень на рівні конкретних виконавців. Водночас BPMN орієнтований на опис послідовності виконання дій, тобто на рівень реалізації процесу, і не забезпечує ієрархічної декомпозиції функцій системи [25].

Методологія функціонального моделювання IDEF0 розроблена на основі методу SADT та набула статусу стандарту IEEE у 1998 році [26]. Метод призначений для представлення рішень, дій та діяльності існуючої або перспективної організації чи системи і може застосовуватись для моделювання широкого спектру систем що складаються з людей, машин, матеріалів, комп'ютерів та інформації різних видів. IDEF0 базується на нотації ICOM що чітко розмежовує чотири типи зв'язків кожного функціонального блоку: Input, Control, Output та Mechanism. Ключовою перевагою методу є підтримка ієрархічної декомпозиції: моделювання починається з контекстної діаграми A-0 що представляє систему як єдиний функціональний блок, і поступово деталізується через розкладання на дочірні блоки [27]. Такий підхід дозволяє послідовно переходити від загального опису до деталізації окремих функцій зберігаючи при цьому цілісність моделі та узгодженість між рівнями. IDEF0 є

незалежним від деталей технічної реалізації та забезпечує концептуальний рівень опису системи.

Для обґрунтованого вибору методу варто розглядати критерії що відображають вимоги до концептуального рівня проектування інформаційно-аналітичної системи моніторингу ГЗН. Концептуальне проектування передбачає опис системи незалежно від деталей реалізації - що є принциповим на ранніх етапах розробки. Ієрархічна декомпозиція дозволяє описати систему від загального контексту до конкретних функціональних блоків зберігаючи цілісність моделі. Незалежність від реалізації забезпечує стабільність моделі при зміні технічних рішень. Придатність для систем з нормативним регулюванням визначає можливість формалізувати зовнішні обмеження що впливають на функціонування системи - зокрема міжнародне законодавство та стандарти збору даних у сфері ГЗН. Порівняльний аналіз методів за зазначеними критеріями наведено в табл. 1.2.

Таблиця 1.2 – Порівняльний аналіз методів функціонального моделювання [23].

Критерій	DFD	BPMN	IDEF0
Концептуальне проектування	Ні	Ні	Так
Ієрархічна декомпозиція	Так	Ні	Так
Незалежність від реалізації	Так	Ні	Так
Моделювання ролей та учасників	Ні	Так	Ні
Придатність для міжінституційних систем	Ні	Частково	Так

З наведеної таблиці видно що кожен з розглянутих методів має власні переваги у певних контекстах. DFD є ефективним для деталізації потоків даних у програмних системах та підтримує декомпозицію, однак не забезпечує концептуального рівня опису та не дозволяє формалізувати регуляторні впливи. BPMN є потужним інструментом для опису міжорганізаційної взаємодії з чіткими ролями виконавців, однак орієнтований на виконуваний рівень деталізації і не підтримує ієрархічну декомпозицію функцій. IDEF0 є єдиним з розглянутих методів що одночасно забезпечує концептуальний рівень опису, ієрархічну декомпозицію та придатність для систем що функціонують в умовах

нормативного регулювання. Саме тому метод IDEF0 обрано для функціонального моделювання інформаційно-аналітичної системи моніторингу ГЗН.

Висновки до розділу 1

У першому розділі розкрито теоретичні засади моніторингу гендерно зумовлених ситуацій. Проаналізовано природу ГЗН як соціального явища на основі міжнародної нормативно-правової бази та встановлено принципове значення розмежування біологічної статі та гендерної ідентичності для систем моніторингу. Проведено аналіз існуючих інформаційних систем моніторингу ГЗН, виявлено їх спільну архітектурну проблему - нездатність консолідувати дані з різних інституційних джерел у єдиний аналітичний простір. Обґрунтовано вибір методу IDEF0 для функціонального моделювання системи. Сформовано теоретичну основу для подальшої розробки моделі інформаційно-аналітичної системи моніторингу ГЗН.

РОЗДІЛ 2
ПРОЕКТУВАННЯ ІНФОРМАЦІЙНО-АНАЛІТИЧНОЇ СИСТЕМИ МОНІТОРИНГУ ГЕНДЕРНО ЗУМОВЛЕНИХ СИТУАЦІЙ

2.1 Постановка задачі та визначення вимог до системи

Головною задачею проєктування інформаційно-аналітичної системи моніторингу ГЗН є формалізація процесу консолідації різнорідних інституційних даних та побудова аналітичного шару, здатного трансформувати первинні звернення у знання придатні для прийняття управлінських рішень. Ефективне вирішення цієї задачі вимагає чіткого визначення вхідних потоків даних, механізмів їх обробки, очікуваних аналітичних результатів та кінцевих користувачів системи. Саме ці компоненти формують основу для подальшого функціонального моделювання у нотації IDEF0.

Аналіз існуючих систем моніторингу ГЗН, проведений у підрозділі 1.2, виявив три структурні прогалини, що у сукупності унеможливлюють формування повної аналітичної картини явища: відсутність механізму інтеграції даних з кількох інституційних джерел, обмеженість аналітичного шару у системах обліку та відсутність індикатора прихованої статистики як міри розриву між офіційними й неурядовими джерелами [7].

На основі виявлених прогалин формулюється основна задача дослідження: побудувати модель інформаційно-аналітичної системи моніторингу гендерно зумовлених ситуацій у нотації IDEF0, що інтегрує дані кількох інституційних джерел, використовує уніфіковану структуру категоріальних ознак інцидентів та формує аналітичний шар для виявлення тенденцій, регіональних закономірностей, груп ризику і прихованої статистики звернень з метою підтримки управлінських рішень.

Цільова аудиторія проєктованої моделі охоплює два типи кінцевих користувачів, що розрізняються за інституційним контекстом та характером аналітичних потреб. Перший тип - аналітики та координатори державних соціальних служб і спеціалізованих органів у сфері протидії насильства, для яких

система моніторингу є інструментом обґрунтування управлінських рішень щодо розподілу ресурсів підтримки та планування мережі кризових центрів [28]. Другий тип - координатори неурядових організацій, що надають послуги постраждалим паралельно з державними структурами та в інституційно незалежний спосіб. На відміну від державних служб, що спираються на офіційно зареєстровані звернення, НУО фіксують значну частку випадків, які залишаються поза межами державної статистики через страх помсти, соціальну стигматизацію або недовіру до інституцій [6]. Для координаторів НУО аналітично важливим є не лише власна статистика, а її зіставлення з офіційними даними - розрив між цими показниками у конкретному регіоні слугує підставою для адвокаційної діяльності та обґрунтування фінансування [17]. Обидва типи користувачів є особами, що приймають управлінські рішення, а не технічними операторами, що визначає вимогу до аналітичного компонента моделі щодо інтерактивності та доступності без необхідності програмування.

Визначення вимог до моделі здійснюється у термінах нотації IDEF0, що передбачає чітке розмежування чотирьох типів зв'язків кожного функціонального блоку: вхідних даних (Input), управлінь (Control), результатів (Output) та механізмів реалізації (Mechanism) [26; 27]. Така структура є необхідною передумовою побудови функціональної моделі та забезпечує повноту і несуперечність опису системи.

Вхідними даними моделі є чотири потоки інформації від інституційних джерел різних рівнів: звернення до поліції, дані закладів охорони здоров'я, дані НУО, а також демографічні дані. Принциповою характеристикою вхідних даних є їхня структурна неоднорідність, що розглядається у підрозділі 2.2.

Управліннями моделі виступають три нормативні рамки, що визначають правила та обмеження обробки даних: міжнародні стандарти у сфері моніторингу ГЗН, що охоплюють як правові зобов'язання щодо класифікації та захисту даних, так і методологічні рамки збору статистики [2; 3]; державна політика у сфері протидії домашньому насильству та захисту персональних

даних; методологічні вимоги до класифікації форм насильства та категоризації інцидентів, визначені у підрозділі 1.1.

Механізмами реалізації моделі є два компоненти: програмні модулі обробки даних (модулі стандартизації, верифікації та об'єднання CSV файлів від різних інституційних джерел) та аналітичний інструментарій для візуалізації, теоретичні засади якого розглядаються у підрозділі 2.3.

Очікуваними результатами функціональної моделі є три типи аналітичних виходів. Аналітичні звіти з динамікою звернень дозволяють відстежувати зміни структури інцидентів за формою насильства та характером стосунків між сторонами, а також візуалізувати регіональний розподіл у вигляді теплових карт - інформацію, яку агреговані річні показники платформ на кшталт EIGE за своєю природою не здатні забезпечити, і яка слугує координаторам доказовою базою для виваження розміщення кризових центрів та обґрунтування розподілу ресурсів реагування [7; 28; 30]. Виявлені групи ризику є аналітичним виходом, що дозволяє ідентифікувати демографічні та соціальні характеристики постраждалих, що найчастіше звертаються до різних типів служб, та відповідно адаптувати профілі підтримки. Індикатор прихованої статистики відображає співвідношення між кількістю звернень, зафіксованих НУО, та кількістю офіційно зареєстрованих випадків у тому самому регіоні та часовому проміжку. Концептуально цей індикатор спирається на поняття «темної цифри» злочинності [18] та адаптоване до контексту ГЗН у методологічних розробках EIGE [6; 31]. Суттєве розходження між показниками НУО та офіційної статистики є аналітичним сигналом системного недообліку звернень у конкретному регіоні та підставою для перегляду інституційних механізмів реагування, що робить цей вихід ключовим для обох типів визначених користувачів.

Слід зазначити, що запропонована модель має ряд методологічних обмежень. Аналітичний компонент реалізується на синтетичному датасеті, що імітує структуру реальних даних моніторингу ГЗН, оскільки дані про інциденти ГЗН є конфіденційними та захищеними правилами збереження персональних

даних постраждалих осіб. Використання синтетичних наборів даних є поширеною практикою у сферах, де доступ до реальних індивідуалізованих записів обмежений вимогами приватності, зокрема в охороні здоров’я та соціальних дослідженнях [32; 33]. Такі датасети дають змогу відтворити структуру, типи змінних і логіку аналітичної обробки без розкриття чутливої інформації про реальних осіб. Попри обмеження, запропонована модель забезпечує відтворювану архітектурну основу для демонстрації аналітичного потенціалу інтеграції інституційних даних у сфері моніторингу ГЗН.

2.2 Структура інтегрованого масиву даних системи моніторингу

Для подальшого аналізу даних про гендерно зумовлені ситуації необхідно сформувати єдину структуровану основу, придатну для зіставлення записів із різних джерел. Як встановлено у підрозділі 2.1, вхідні дані надходять у форматі CSV від трьох інституційних джерел зі структурно неоднорідними схемами полів - поліції, закладів охорони здоров'я та НУО. Ця неоднорідність є структурним наслідком відсутності єдиного міжінституційного стандарту фіксації даних у сфері ГЗН, задокументованої у міжнародних методологічних настановах [29; 32]. Концепція інтегрованого масиву даних передбачає об'єднання адміністративних записів із кількох незалежних інституційних джерел в єдину структуру для аналітичної обробки - підхід, що є усталеним у соціальній політиці та моніторингових системах охорони здоров'я [34; 35]. У контексті цієї роботи інтегрований масив реалізується як плоска таблиця, в якій кожен рядок відповідає одному зверненню і містить усі необхідні атрибути для подальшого аналізу [19; 20].

Класифікаційну основу структури масиву формує двовимірна таксономія ГЗН. Перший вимір, форма насильства, охоплює: фізичне насильство (навмисне заподіяння фізичної шкоди), сексуальне насильство (будь-які сексуальні дії без згоди), психологічне насильство (залякування, маніпуляції, ізоляція), економічне насильство (контроль над фінансовими ресурсами, позбавлення економічної самостійності) та контролюючу поведінку (систематичний контроль дій,

пересувань та соціальних контактів жертви) [1; 2]. Виокремлення контролюючої поведінки як самостійної форми є принципово важливим - попри те що психологічне та економічне насильство від інтимного партнера є одними з найпоширеніших форм ГЗН, вони рідко фіксуються як окремі категорії в адміністративних реєстрах і найчастіше виявляються лише через аналіз текстових описів звернень до НУО [5]. Другий вимір - тип стосунків між жертвою та кривдником - базується на класифікації ВООЗ [1] та включає: насильство від інтимного партнера (IPV), домашнє насильство з боку членів родини, насильство з боку знайомих та незнайомців. Різні типи стосунків потребують якісно різних інституційних відповідей: IPV - кризових центрів і правового захисту, насильство від незнайомців - правоохоронного реагування. Двовимірна класифікація забезпечує повноту охоплення явища та дозволяє виявляти специфічні закономірності, що залишаються невидимими при одновимірному аналізі.

Концептуальна структура масиву охоплює чотири групи полів. Поля ідентифікації включають унікальний ідентифікатор запису, що забезпечує цілісність масиву при об'єднанні файлів від різних джерел, та код інституційного джерела - критично важливий для розрахунку індикатора прихованої статистики через зіставлення обсягів звернень між джерелами. Поля просторово-часового аналізу охоплюють дату події як основу для аналізу динаміки звернень у часі, назву населеного пункту для геовізуалізації та тип місцевості (місто/область), що дозволяє виявляти структурні відмінності у доступі до служб підтримки [28]. Демографічні поля характеризують суб'єктів інциденту: вікова категорія постраждалої особи фіксується відповідно до стандартизованих діапазонів UNFPA [28], стать постраждалої особи фіксується як базова демографічна характеристика відповідно до стандартних форм звітності інституційних джерел, поле повторності фіксує чи є звернення повторним для тієї самої особи - ознака, критична для виявлення хронічних ситуацій насильства та груп ризику. Поле опису інциденту містить узагальнений текстовий опис інциденту та використовується як додатковий пояснювальний атрибут запису. Основні

аналітичні категорії - форма насильства та тип стосунків між сторонами - у межах цієї роботи подаються як уже структуровані поля, що забезпечує придатність масиву до подальшої обробки.

Концептуальна структура полів інтегрованого масиву даних наведена у табл. 2.1.

Таблиця 2.1 – Концептуальна структура полів інтегрованого масиву даних системи моніторингу ГЗН

Назва поля	Тип	Опис
record_uid	числовий	Унікальний ідентифікатор запису
source_id	числовий	Код інституційного джерела (1 - поліція, 2 - медицина, 3 - НУО)
incident_date	дата	Дата події у стандартизованому форматі
location_name	текстовий	Назва населеного пункту або регіону
region_type	категоріальний	Тип місцевості: місто / село
victim_age	числовий	Вік постраждалої особи
victim_sex	категоріальний	Стать постраждалої особи
is_repeated_case	логічний	Ознака повторного звернення
injury_type	категоріальний	Тип зафіксованої травми
incident_description	текстовий	Узагальнений опис інциденту
violence_form	категоріальний	Форма насильства
relationship_type	категоріальний	Тип стосунків між сторонами

Слід зазначити, що через різний рівень деталізації записів у різних джерелах не всі поля масиву будуть заповнені для кожного звернення - це є очікуваною характеристикою інтегрованих адміністративних даних у сфері ГЗН [32].

Демографічні дані - зокрема показники чисельності населення регіонів - не входять до структури інтегрованого масиву інцидентів, а використовуються як довідкові дані на етапі аналітичної обробки. Їх застосування є методологічно необхідним для нормалізації показників поширеності ГЗН у розрізі регіонів - розрахунку кількості звернень на 100 000 населення - що є стандартною вимогою порівняльного аналізу в соціальній статистиці [28].

Принциповою проблемою при формуванні інтегрованого масиву є структурна неоднорідність вхідних даних від різних джерел. Аналіз міжнародних практик збору даних про ГЗН свідчить, що адміністративні джерела систематично відрізняються за складом полів, форматами значень та рівнем деталізації записів [29; 34]. Поліцейські звіти оперують юридичними категоріями злочинів і формалізованими полями складу правопорушення, при цьому опис інциденту, як правило, міститься в окремому полі вільного тексту [36]; медичні записи структуровані навколо кодів Міжнародної класифікації хвороб (МКХ) та клінічних описів травм, натомість інформація про характер стосунків між сторонами фіксується вкрай нерегулярно [37]; НУО-дані, особливо зібрані через платформи на кшталт KoBoToolbox та йому подібним, зберігають значну частку інформації у вигляді вільного тексту без стандартизованих категорій [21]. Показово, що тип стосунків між постраждалою особою та кривдником не завжди фіксується як окреме структуроване поле в адміністративних джерелах [38]. У поліцейських і медичних записах ця інформація найчастіше міститься у полі вільного тексту або не фіксується взагалі, тоді як у НУО-анкетах вона може бути присутня у структурованому вигляді. У реальній системі отримання цього поля потребувало б додаткового ETL-кроку або ручного введення на етапі стандартизації. Саме тому в запропонованій структурі інтегрованого масиву це поле визначено як обов’язковий категоріальний атрибут, необхідний для подальшого аналізу профілю інцидентів та виявлення груп ризику. Типовими проявами неоднорідності є також: різні формати дат, відмінні переліки допустимих

значень для категоріальних полів, а також відсутність окремих полів у частині джерел.

Вирішення проблеми неоднорідності у запропонованій моделі реалізується на етапі підготовки синтетичного масиву даних, що передує аналітичній обробці. Концептуально цей етап відповідає процесам вилучення, перетворення та завантаження даних (ETL) і може бути реалізований засобами Power Query - вбудованого модуля трансформації даних у середовищі Power BI Desktop [39]. До основних операцій такого етапу належать приведення форматів дат до єдиного стандарту, уніфікація категоріальних значень, нормалізація назв регіонів, перевірка типів даних та формування єдиного набору полів відповідно до структури, наведеної у табл. 2.1. Такий підхід дозволяє перетворити розрізнені записи з різних джерел на узгоджений датасет, придатний для побудови зв'язків із довідковими демографічними даними та подальшого аналітичного моделювання.

2.3 Теоретичні засади та обґрунтування застосування Power BI і DAX для аналітичного моделювання

Реалізація аналітичного компонента інформаційно-аналітичної системи моніторингу ГЗН потребує вибору інструментарію, здатного забезпечити повний цикл обробки даних - від стандартизації різнорідних вхідних файлів до формування інтерактивних звітів, придатних для підтримки управлінських рішень без необхідності програмування з боку кінцевого користувача. Для цього обрано платформу Microsoft Power BI, що поєднує три функціональні рівні в єдиному середовищі: рівень трансформації даних (Power Query), рівень семантичної моделі з аналітичними обчисленнями (модель даних і мова DAX) та рівень інтерактивної візуалізації [40]. Застосування BI-платформ у соціальному та медичному моніторингу підтверджується сучасними дослідженнями, які демонструють здатність зрілих BI-систем забезпечувати виявлення груп ризику, стратифікацію вразливих категорій населення та оцінку ефективності соціальних

політик - функції, структурно ідентичні аналітичним виходам проектованої моделі моніторингу ГЗН [41].

Першим функціональним рівнем є модуль Power Query - вбудований ETL-рушій платформи, призначений для підключення до джерел даних, їх очищення та структурування перед завантаженням у семантичну модель [39]. Абревіатура ETL (Extract, Transform, Load) описує трифазний процес: вилучення даних із джерел, їх перетворення відповідно до цільової схеми та завантаження в аналітичне середовище. У контексті моніторингу ГЗН, де вхідні дані надходять від структурно неоднорідних інституційних джерел - поліції, медичних закладів та НУО, - ETL-етап є методологічно необхідним: без попередньої стандартизації форматів, уніфікації категоріальних значень та нормалізації назв регіонів подальші аналітичні обчислення втрачають достовірність. Power Query фіксує кожну операцію трансформації як відтворюваний крок, що автоматично генерує код мовою M (Power Query Formula Language), забезпечуючи повну відтворюваність процесу підготовки даних [39]. Саме цей етап стандартизації та верифікації вхідних даних є першою функціональною ланкою системи моніторингу, без якої подальше аналітичне моделювання втрачає методологічну коректність.

Другим рівнем є семантична модель даних, що визначає склад таблиць, тип зв'язків між ними та логіку аналітичних обчислень. В основі її проектування лежить принцип вимірного моделювання (dimensional modeling): модель будується навколо таблиці фактів, що містить вимірювані події, та таблиць вимірів, що описують контекст цих подій [15]. У Power BI такий підхід реалізується через зіркову схему (star schema), де таблиця фактів пов'язана з таблицями вимірів відношенням «багато-до-одного», а фільтри з вимірів поширюються на таблицю фактів через механізм перехресної фільтрації [42]. Така структура є необхідною передумовою для коректного розрахунку нормалізованих показників, оскільки демографічні дані зберігаються окремо від таблиці інцидентів і стають доступними саме через зв'язок між таблицями [43].

У межах семантичної моделі Power BI розрізняє два принципово різних типи DAX-об'єктів: обчислювані стовпці та міри. Обчислюваний стовпець обчислюється під час завантаження даних у рядковому контексті - тобто окремо для кожного рядка таблиці - і зберігається як статичний атрибут запису. Він є доцільним для категоризації, що не залежить від поточного стану звіту: наприклад, визначення вікової групи на основі числового значення віку, формування україномовних міток для категоріальних полів або приєднання географічного коду з таблиці виміру для картографічної візуалізації. Міра, натомість, обчислюється динамічно у контексті фільтрації під час формування звіту і не займає місця в моделі даних [44]. Вона є доцільною для всіх агрегованих показників, значення яких мають змінюватись залежно від вибору користувача: підрахунку звернень у розрізі джерел, розрахунку показника поширеності, визначення частки повторних звернень або обчислення індексів, що порівнюють дані між інституційними джерелами.

Мова DAX (Data Analysis Expressions) є формульною мовою виразів, розробленою для аналітичних обчислень у середовищі Power BI [43]. Теоретичну основу DAX становить концепція контексту обчислення, що визначає набір даних, над яким виконується формула у конкретний момент [44; 45]. DAX розрізняє рядковий контекст, що виникає при обчисленні значень для кожного рядка таблиці, та контекст фільтрації, що формується активними фільтрами зрізів, осей діаграм і зв'язків між таблицями. Центральною функцією мови є CALCULATE, яка модифікує контекст фільтрації для обчислення виразу: приймає вираз як перший аргумент і один або кілька умов-фільтрів як наступні аргументи, формуючи новий контекст, у якому виконується обчислення [46]. Типовим прикладом є конструкція виду CALCULATE(COUNT(...), умова), що підраховує кількість записів, які відповідають заданій умові в межах поточного контексту фільтрації звіту. Саме цей механізм дозволяє будувати порівняльні показники між джерелами даних і є основою для розрахунку індексів співвідношення між офіційною та позаінституційною статистикою. Методологічна обґрунтованість таких ratio-based підходів підтверджується

дослідженнями, які фіксують систематичне заниження офіційних показників насильства порівняно з альтернативними джерелами даних [18]. Поряд із CALCULATE, у побудові аналітичних мір і обчислюваних стовпців застосовуються функції навігації між таблицями, перетворення та роботи з датами [47]. Функція RELATED забезпечує звернення до атрибуту з пов'язаної таблиці виміру в рядковому контексті таблиці фактів що є необхідним для обчислюваних стовпців, які потребують даних із довідникової таблиці. Функції YEAR, MAX і FORMAT забезпечують часовий вимір аналізу: виокремлення року з поля дати для групування динаміки, визначення крайньої дати у масиві та форматування значень для відображення у картках звіту. Функція SWITCH реалізує логіку умовного перетворення категоріальних значень - зокрема перекодування числових ідентифікаторів або англомовних категорій у локалізовані текстові мітки. Функція DIVIDE використовується замість арифметичного ділення для безпечного оброблення випадків ділення на нуль, що є важливим для показників, що обчислюються у розрізі одиниць аналізу з потенційно відсутніми значеннями у знаменнику [47].

Окремий клас DAX-функцій становлять ітераторні функції, що виконують покрядкове обчислення виразу для кожного рядка зазначеної таблиці з подальшою агрегацією результатів [48]. Прикладом такої функції є AVERAGEX, яка обчислює середнє арифметичне результатів виразу по рядках таблиці [49]. Типовим сценарієм її застосування є розрахунок середнього значення показника, що попередньо обчислюється окремо для кожної одиниці агрегації, що методологічно відрізняється від простого усереднення загального агрегату і дозволяє уникнути викривлення результату через нерівномірний розподіл обсягів між одиницями аналізу. У поєднанні з AVERAGEX використовується функція VALUES, яка повертає таблицю унікальних значень поля, наприклад, перелік регіонів, і виступає таблицею ітерації. Функція ISBLANK застосовується для коректної обробки порожніх значень у полях, де відсутність даних має бути явно ідентифікована як окрема категорія, а не трактуватися як нульове значення [48].

Таким чином, обраний аналітичний інструментарій утворює цілісний методологічний апарат, що забезпечує перехід від розрізнених інституційних записів до інтерактивних аналітичних показників, здатних відповідати на управлінські запити в реальному часі. Практична реалізація кожного з цих компонентів описана у підрозділі 3.2.

Висновки до розділу 2

У другому розділі сформовано концептуальну основу проектування інформаційно-аналітичної системи моніторингу ГЗН. Сформульовано головну задачу дослідження та визначено вимоги до моделі у термінах нотації IDEF0: чотири вхідні потоки даних від інституційних джерел, управлінь у вигляді міжнародних стандартів і державної політики, механізми реалізації та три типи аналітичних виходів. Запропоновано концептуальну структуру інтегрованого масиву даних із двовимірною класифікаційною схемою за формою насильства та типом стосунків між сторонами, що забезпечує повноту охоплення явища. Обґрунтовано використання синтетичного датасету як методологічно прийнятного підходу в умовах обмеженого доступу до реальних мікроданих. Визначено аналітичний інструментарій системи та розкрито теоретичні засади його ключових компонентів: Power Query для стандартизації різнорідних вхідних даних на ETL-етапі, семантичної моделі із зірковою схемою для структурування зв'язків між таблицями та мови DAX для реалізації аналітичних обчислень. Сформовано методологічну основу для практичної реалізації системи, представленої у третьому розділі.

РОЗДІЛ 3
РЕАЛІЗАЦІЯ МОДЕЛІ ІНФОРМАЦІЙНО-АНАЛІТИЧНОЇ СИСТЕМИ МОНІТОРИНГУ ГЕНДЕРНО ЗУМОВЛЕНИХ СИТУАЦІЙ

3.1 Функціональна модель системи моніторингу ГЗН в нотації IDEF0

Функціональна модель є формальним інструментом визначення меж системи, її взаємодії із зовнішнім середовищем та внутрішньої логіки перетворення даних. Для виконання завдання було обрано методологію IDEF0, що забезпечує ієрархічну декомпозицію системи від загального контексту до деталізованих функціональних блоків із збереженням узгодженості між рівнями [26; 27]. Практична реалізація моделі здійснена у середовищі AllFusion Process Modeler - спеціалізованому інструменті функціонального моделювання, що підтримує нотацію IDEF0 відповідно до стандарту IEEE та забезпечує графічну побудову діаграм з автоматичною перевіркою синтаксичної коректності зв'язків між блоками [50]. Середовище AllFusion Process Modeler реалізує повний набір елементів нотації ICOM: вхідні дані (Input) подаються стрілками з лівого боку функціонального блоку, управління (Control) - зверху, виходи (Output) - праворуч, механізми реалізації (Mechanism) - знизу. Кожен функціональний блок отримує унікальний ідентифікатор у форматі A-X, де X відповідає рівню та порядковому номеру блоку в ієрархії моделі.

Проектована модель охоплює три рівні ієрархічної декомпозиції. Глибина декомпозиції на кожному рівні визначається складністю відповідного процесу: для блоків, що реалізують багатоетапні перетворення з відмінними управліннями, одного рівня опису недостатньо для повноти функціональної специфікації. Перший рівень, контекстна діаграма A-0, представляє систему як єдиний функціональний блок у взаємодії із зовнішнім середовищем і визначає межі системи та її зовнішні зв'язки. Другий рівень, діаграма декомпозиції A0, розкриває внутрішню функціональну архітектуру через чотири послідовні блоки обробки даних. Третій рівень охоплює дві додаткові декомпозиції: діаграму A1, що деталізує ETL-процес збору та стандартизації даних від трьох різнорідних

інституційних джерел, та діаграму A3, що розкриває багатоетапну структуру аналітичної обробки від збагачення семантичної моделі до розрахунку порівняльних індикаторів звітності. Такий підхід відповідає рекомендованій практиці IDEF0-моделювання, де кожен рівень декомпозиції поглиблює деталізацію без втрати узгодженості із загальною архітектурою [26].

Контекстна діаграма A-0 відображає систему моніторингу ГЗН на найвищому рівні абстракції - як єдиний функціональний блок «Моніторинг гендерно зумовлених ситуацій», що взаємодіє із зовнішнім середовищем через чотири типи зв'язків нотації ICOM. Побудовану діаграму наведено на рис. 3.1.

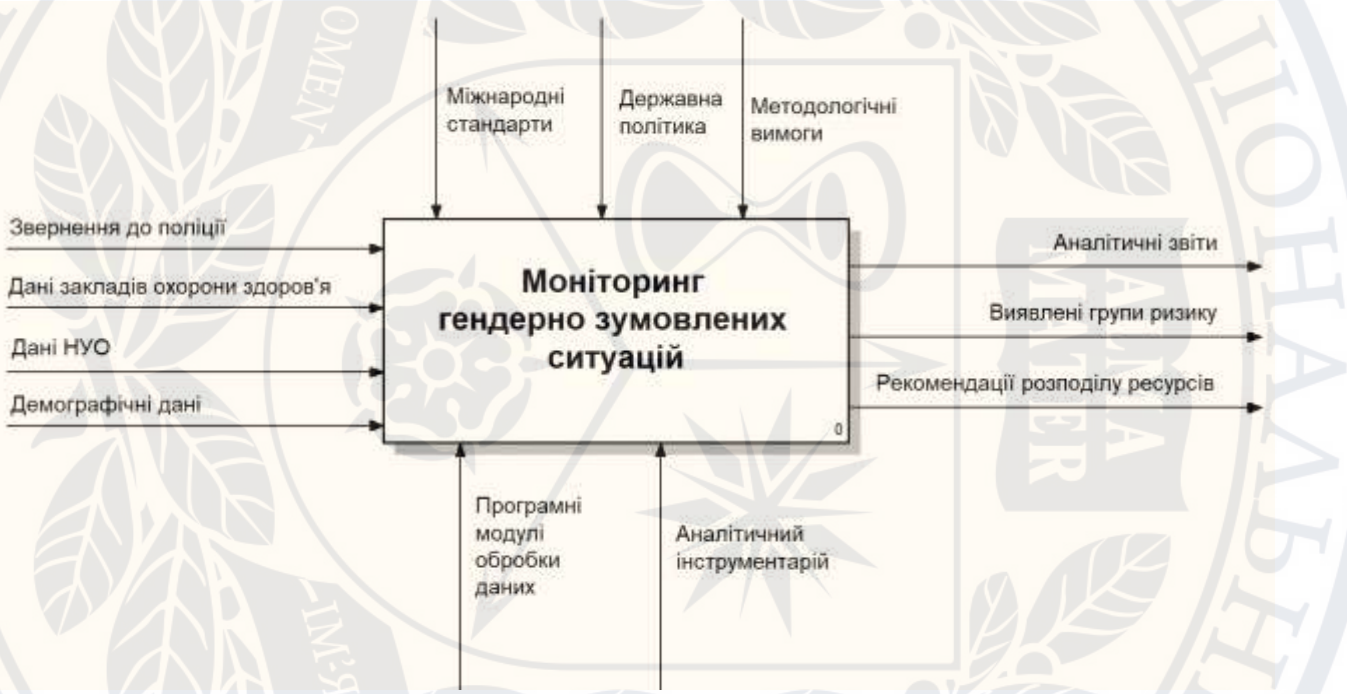

Рис. 3.1 – Контекстна діаграма A-0 моделі системи моніторингу ГЗН

На діаграмі реалізовано ряд архітектурних рішень, що відображають специфіку системи моніторингу ГЗН. Три інституційні вхідні потоки - звернення до поліції, дані закладів охорони здоров'я та дані НУО - зображені як окремі стрілки, а не об'єднані в один вхід, що підкреслює їхню структурну неоднорідність і необхідність індивідуальної обробки на етапі стандартизації [36; 37]. Демографічні дані представлено окремим вхідним потоком, що фіксує їхню участь у функціонуванні системи поряд з інституційними джерелами, однак, на відміну від них, цей потік задіється не на етапі збору та обробки, а

безпосередньо на аналітичному рівні, що стане очевидним на діаграмі декомпозиції A0.

Три управління, зображені стрілками зверху функціонального блоку, відображають різні рівні нормативного регулювання системи. Міжнародні стандарти охоплюють зовнішні методологічні рамки збору та представлення статистики у сфері ГЗН, зокрема настанови UNFPA щодо виробництва статистики насильства стосовно жінок, методологічні керівництва EIGE щодо якості адміністративних джерел даних, а також вимоги Стамбульської конвенції та CEDAW щодо обов'язкових категорій обліку і захисту персональних даних постраждалих осіб [2; 3; 28; 31]. Державна політика включає національне законодавство та політику захисту персональних даних, що регулює умови збору, зберігання та обміну інформацією між інституційними джерелами. Методологічні вимоги є внутрішнім управлінням, що конкретизує стандарти категоризації інцидентів, розроблені в межах цієї роботи на основі зазначених зовнішніх рамок, зокрема двовимірну класифікаційну схему за формою насильства та типом стосунків між сторонами, описану у підрозділі 2.2, що забезпечує однорідність аналітичних категорій незалежно від інституційного джерела даних.

Два механізми реалізації, зображені стрілками знизу функціонального блоку, відповідають двом функціонально відмінним типам операцій у системі. Програмні модулі обробки даних забезпечують структурне перетворення вхідних потоків, приведення форматів, уніфікацію категоріальних значень і формування єдиного масиву; у цій роботі на роль такого модуля запропоновано Power Query - вбудований ETL-рушій платформи Power BI, теоретичні засади якого розглянуто у підрозділі 2.3 [39]. Аналітичний інструментарій забезпечує обчислення показників і формування інтерактивних звітів - на цю роль запропоновано платформу Microsoft Power BI з аналітичним компонентом на основі мови DAX [40; 44].

Три виходи системи, аналітичні звіти, виявлені групи ризику та рекомендації розподілу ресурсів, відображають три рівні аналітичної цінності

результатів: описовий, діагностичний і прескриптивний. Такий поділ є свідомим проектним рішенням, що відповідає потребам двох визначених типів кінцевих користувачів (аналітиків, соціальних служб і координаторів НУО) та забезпечує адресність кожного аналітичного виходу [28; 30].

Якщо контекстна діаграма A-0 визначає межі системи та її взаємодію із зовнішнім середовищем, то діаграма декомпозиції A0 розкриває внутрішню функціональну архітектуру, замінюючи єдиний блок чотирма взаємопов'язаними функціональними блоками:

- A1 «Збір та стандартизація даних»;
- A2 «Обробка та верифікація даних»;
- A3 «Аналітична обробка»;
- A4 «Формування звітів та візуалізацій».

Вихід кожного блоку є вхідним потоком наступного, утворюючи послідовний ланцюг перетворення даних від первинних інституційних записів до кінцевих аналітичних результатів. Управління і механізми контекстної діаграми розподіляються між блоками відповідно до їхньої функціональної ролі - методологічні вимоги регулюють переважно ранні блоки обробки, тоді як міжнародні стандарти і державна політика поширюються на аналітичний і звітний рівні. Побудовану діаграму декомпозиції A0 наведено на рис. 3.2.

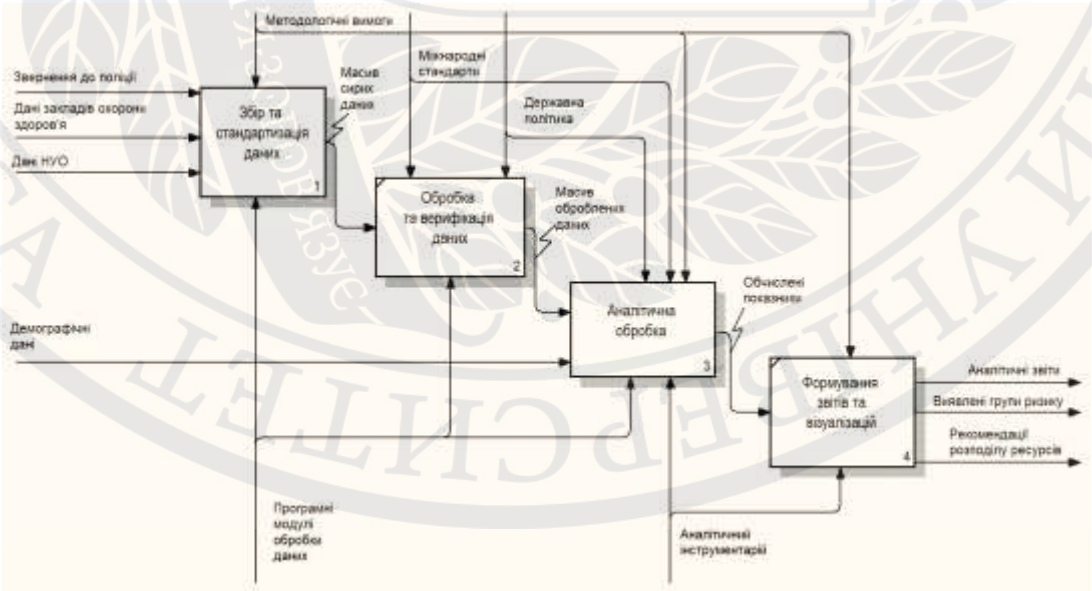

Рис. 3.2 – Діаграма декомпозиції A0 функціональної моделі системи моніторингу ГЗН

Блок А1, «Збір та стандартизація даних», виконує першу функціональну ланку системи - отримання та первинну уніфікацію вхідних потоків від трьох інституційних джерел. Вхідними даними є звернення до поліції, дані закладів охорони здоров'я та дані НУО. Управлінням є методологічні вимоги, що визначають цільову структуру інтегрованого масиву - склад обов'язкових полів, допустимі категоріальні значення та формат дат. Механізмом реалізації виступають програмні модулі обробки даних. Виходом блоку є масив сирих даних - структурно узгоджений, але ще не верифікований набір записів, що передається до наступного блоку.

Наступний крок, блок А2, «Обробка та верифікація даних», забезпечує якість і цілісність інтегрованого масиву перед аналітичною обробкою. Вхідними даними є масив сирих даних з попереднього блоку, А1. Управліннями є міжнародні стандарти та державна політика, що визначають вимоги до повноти, достовірності та захисту персональних даних. Механізмом реалізації є програмні модулі, що виконують перевірку типів даних, контроль заповненості ключових ідентифікаційних полів, зокрема унікального ідентифікатора запису та ідентифікатора інституційного джерела, відсутність яких унеможливлює коректну агрегацію та розрахунок порівняльних індикаторів, а також виявлення і обробку пропущених значень у критичних полях та валідацію дат. Окремо контролюється логічна узгодженість записів: числові значення вікового поля перевіряються на відповідність допустимому діапазону, а категоріальні поля - на відповідність класифікаційній схемі. Виходом блоку є масив оброблених даних - верифікований і придатний до аналітичної обробки набір записів, що передається до блоку А3.

Блок А3, «Аналітична обробка», реалізує обчислення аналітичних показників на основі верифікованих даних. Вхідними даними є масив оброблених даних з блоку А2, а також демографічні дані, що надходять безпосередньо із зовнішнього середовища, оминаючи попередні блоки обробки, оскільки це довідникові ресурси, які не потребують стандартизації в рамках системи. Управліннями є методологічні вимоги, міжнародні стандарти, що

визначають рекомендовані індикатори моніторингу ГЗН, та державна політика, що визначають стандарти розрахунку показників соціальної статистики. Механізмами реалізації виступають програмні модулі обробки даних, для побудови зв'язків між таблицями і збагачення моделі похідними атрибутами, та аналітичний інструментарій, що реалізує обчислення агрегованих показників і порівняльних індикаторів звітності. Виходом блоку є обчислені показники, що передаються до блоку А4.

Заключний блок А4, «Формування звітів та візуалізацій», завершує ланцюг перетворення даних, перетворюючи обчислені показники на аналітичні продукти, адресовані кінцевим користувачам системи. Вхідними даними є обчислені показники з блоку А3. Управліннями є міжнародні стандарти та державна політика, що визначають вимоги до змісту і форми аналітичної звітності у сфері ГЗН. Механізмом реалізації є аналітичний інструментарій, що забезпечує формування інтерактивних звітів з можливістю фільтрації за часовим проміжком, регіоном та іншими параметрами без необхідності програмування з боку кінцевого користувача. Виходами блоку є всі три фінальні результати системи - аналітичні звіти, виявлені групи ризику та рекомендації розподілу ресурсів, що відповідають виходам контекстної діаграми А-0.

Для поглибленої деталізації двох більш методологічно складних блоків побудовано діаграми третього рівня декомпозиції. Блок А1 обрано для декомпозиції з огляду на те, що збір та стандартизація даних від трьох різнорідних інституційних джерел є нетривіальним ETL-процесом, що охоплює кілька якісно відмінних підфункцій - підключення до джерел, уніфікацію структури та об'єднання таблиць. Наочна реалізація цього процесу в межах цієї роботи була б методологічно ненадійною через відсутність доступу до реальних структур даних кожного джерела - адміністративні дані поліції та медичних закладів є конфіденційними і не перебувають у відкритому доступі [29; 34], - а у випадку НУО - через індивідуальність систем ведення обліку кожної організації [7; 19]. Тому декомпозиція А1 є концептуальною специфікацією ETL-процесу, що визначає його логічну структуру і може слугувати орієнтиром при інтеграції

системи з реальними джерелами даних. Блок А3 деталізовано через складність аналітичної обробки, що поєднує побудову семантичної моделі, розрахунок нормалізованих показників і обчислення порівняльних індикаторів звітності - підфункції, що задіюють різні механізми і мають відмінні управління.

Побудову декомпозиції блоку А1 наведено на рис. 3.3.

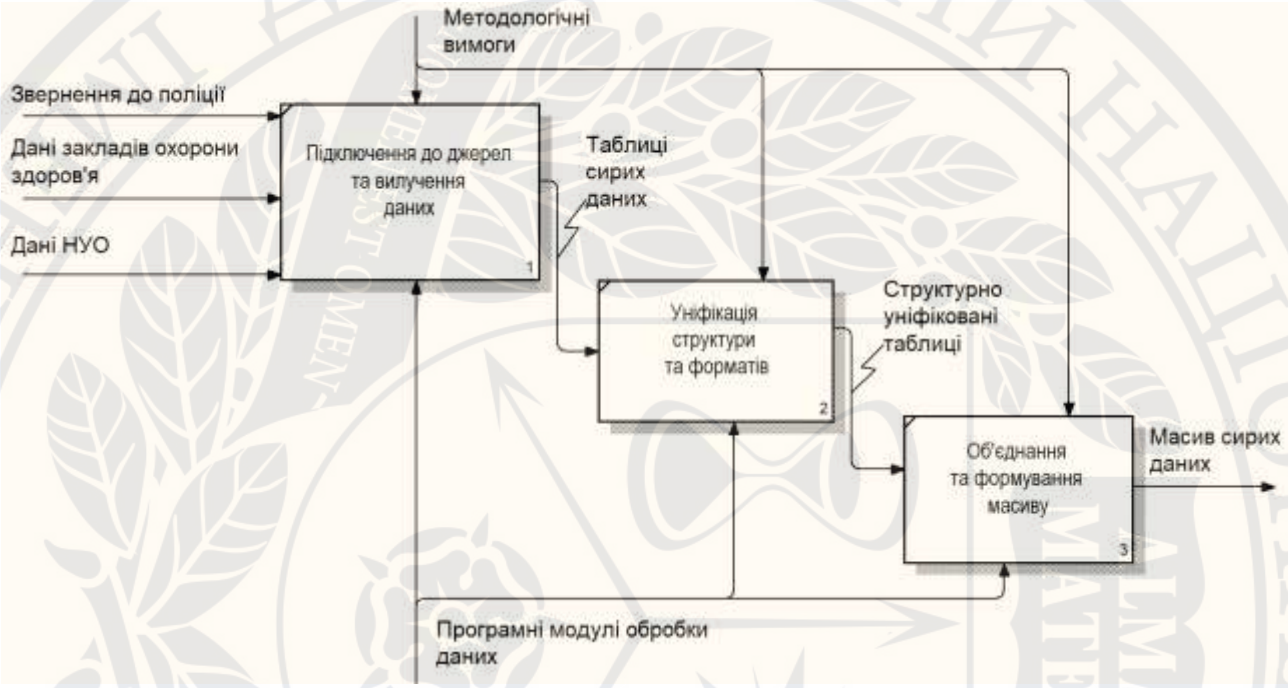

Рис. 3.3 – Діаграма декомпозиції А1 «Збір та стандартизація даних»

Діаграма декомпозиції А1 розкриває процес збору та стандартизації даних через три послідовні підфункції, що реалізують повний ETL-цикл від підключення до джерел до формування єдиного масиву.

Блок А11 «Підключення до джерел та вилучення даних» реалізує перший крок - отримання вхідних записів з трьох джерел і завантаження сирих записів у робоче середовище. Вхідними даними є три інституційні потоки - звернення до поліції, дані закладів охорони здоров'я та дані НУО. Управлінням є методологічні вимоги, що визначають очікуваний склад полів для кожного джерела. Механізмом є програмні модулі обробки даних. Виходом блоку є три окремі таблиці сирих даних, що передаються до А12.

Блок А12 «Уніфікація структури та форматів» виконує приведення трьох таблиць до єдиної схеми полів, визначеної у табл. 2.1. Вхідними даними є таблиці з А11. Управлінням є методологічні вимоги до цільової структури інтегрованого

масиву. Механізмом є програмні модулі, що здійснюють перейменування колонок відповідно до стандартних назв полів, приведення типів даних, уніфікацію допустимих значень категоріальних полів та нормалізацію назв регіонів для забезпечення коректного зв'язку з демографічною довідниковою таблицею. Виходом блоку є три структурно уніфіковані таблиці, що передаються до A13.

Блок A13 «Об'єднання та формування масиву» реалізує злиття трьох уніфікованих таблиць в єдиний інтегрований масив. Вхідними даними є три структурно уніфіковані таблиці з A12. Управлінням є ті самі методологічні вимоги. Механізмом є програмні модулі, що забезпечують об'єднання трьох таблиць в єдиний масив із збереженням повного обсягу записів з кожного джерела. Принципово важливим на цьому кроці є збереження ідентифікатора інституційного джерела для кожного запису - поля, що є необхідним для подальшого розрахунку порівняльних індикаторів звітності. Виходом блоку є масив сирих даних - єдина таблиця з повним набором уніфікованих полів, що є вихідним потоком блоку A1.

Якщо декомпозиція A1 деталізує процес підготовки даних для подальшої обробки, то декомпозиція A3 розкриває якісно інший за характером процес - перетворення верифікованого масиву на аналітичні результати. Три підфункції блоку утворюють послідовний ланцюг: від побудови моделі та збагачення даних похідними атрибутами до розрахунку агрегованих показників і, нарешті, обчислення порівняльних індикаторів звітності, що є ключовим аналітичним виходом системи. Побудовану діаграму наведено на рис. 3.4.

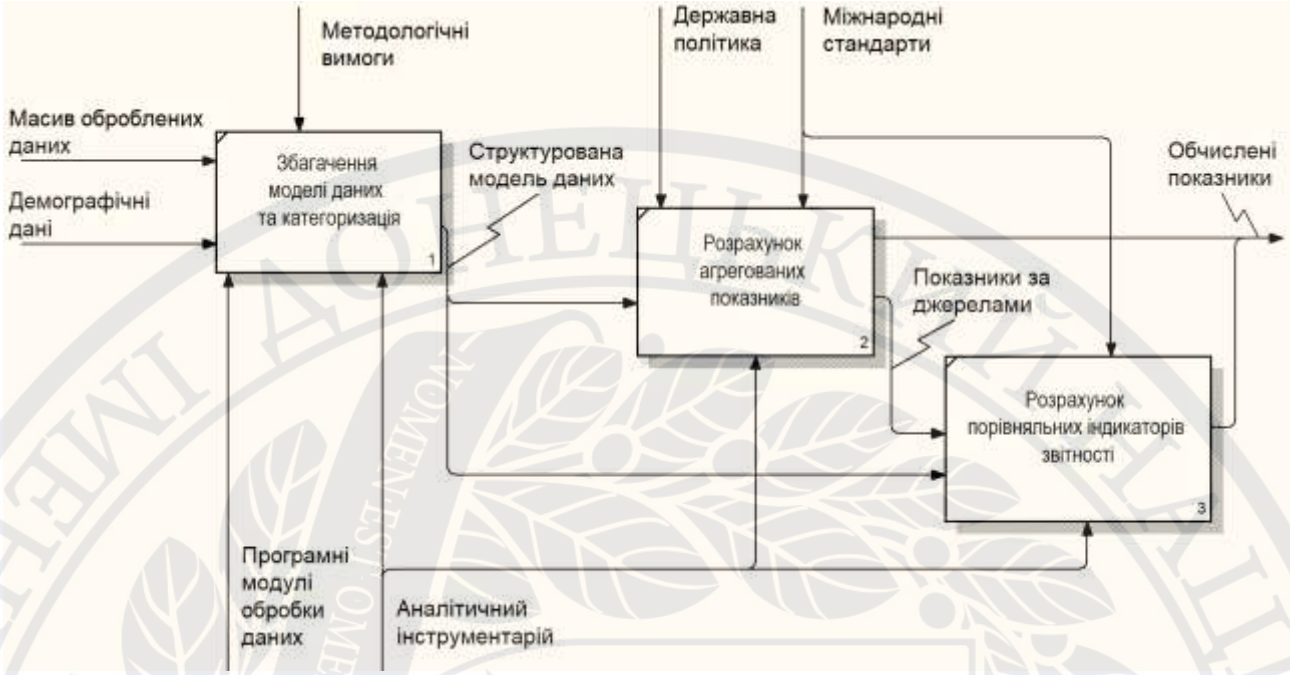

Рис. 3.4 – Діаграма декомпозиції А3 «Аналітична обробка»

Блок А31 «Збагачення моделі даних та категоризація» реалізує побудову структурованої моделі даних на основі верифікованого масиву. Вхідними даними є масив оброблених даних з блоку А2 та демографічні дані із зовнішнього середовища, присутність яких на цьому рівні зумовлена необхідністю побудови зв'язку між таблицею інцидентів і демографічною довідниковою таблицею для подальшої нормалізації показників. Управлінням є методологічні вимоги до класифікації. Механізмами є обидва компоненти системи - програмні модулі обробки даних забезпечують підключення довідникових таблиць і формування зв'язків між ними, наприклад, демографічної таблиці регіонів та таблиці інституційних джерел, тоді як аналітичний інструментарій відповідає за формування похідних атрибутів: вікових груп на основі числового значення віку, україномовних міток категоріальних полів, текстових назв інституційних джерел, географічних кодів для картографічної візуалізації, часових атрибутів для динамічного аналізу тощо. Виходом блоку є структурована модель даних, що слугує спільною вхідною основою для блоків А32 і А33.

Блок А32 «Розрахунок агрегованих показників» обчислює кількісні аналітичні міри на основі сформованої моделі даних. Вхідними даними є структурована модель з А31. Управліннями є державна політика та міжнародні

стандарти, що визначають стандарти вимірювання показників соціальної статистики [28; 31]. Механізмом є аналітичний інструментарій, що реалізує обчислення показників, по типу кількості звернень у розрізі інституційних джерел, нормалізованого показника поширеності, частки повторних звернень або часових індикаторів актуальності даних. Виходами блоку є два потоки: показники за джерелами, що передаються до блоку А33 як основа для розрахунку порівняльних індикаторів, та обчислені показники, що виходять безпосередньо до блоку А4.

Блок А33 «Розрахунок порівняльних індикаторів звітності» обчислює індикатори, що характеризують розрив між інституційними джерелами даних. Вхідними даними є структурована модель даних з А31, що забезпечує доступ до повного набору атрибутів записів, необхідних для регіональної деталізації індикаторів, та показники за джерелами з А32, що слугують базовими лічильниками для розрахунку відношення між джерелами. Управлінням є міжнародні стандарти, що встановлюють вимоги до порівняльного аналізу між інституційними джерелами даних як інструменту оцінки повноти офіційної звітності [28; 31]. Механізмом є аналітичний інструментарій, що реалізує обчислення порівняльних індикаторів звітності - наприклад, регіонального індексу недообліку [18].

Виходом блоку є обчислені показники, що разом з виходом А32 формують повний набір обчислених показників, що передається до блоку А4.

Побудована трирівнева IDEF0-модель формує концептуальну основу інформаційно-аналітичної системи моніторингу ГЗН - вона фіксує що має робити система, якими даними оперувати і в яких нормативних межах функціонувати. Усі три виходи контекстної діаграми А-0 простежуються через кожен рівень декомпозиції до конкретних функціональних блоків, що підтверджує узгодженість моделі між рівнями ієрархії. У такому вигляді модель може слугувати орієнтиром для подальшого розвитку системи - як у частині розширення кола інституційних джерел, так і в частині доповнення переліку аналітичних виходів відповідно до потреб користувачів.

3.2 Реалізація моделі даних та аналітичного компоненту

Реалізація моделі аналітичного компонента інформаційної системи моніторингу ГЗН базується на двох взаємопов'язаних складових: синтетичному масиві даних, що відповідає етапу виходу блоку А2 функціональної моделі, та аналітичній моделі у середовищі Power BI, що реалізує обчислювальний і візуалізаційний шар системи. Масив є стандартизованим і верифікованим, тобто придатним до безпосередньої аналітичної обробки. Разом вони утворюють повний ланцюг від вхідних даних до аналітичних виходів, визначених функціональною моделлю у підрозділі 3.1.

Для цілей цієї роботи країною моделювання обрано Україну, що зумовило відповідне наповнення синтетичного масиву та географічну прив'язку аналітичних виходів. Запропонована модель є універсальною і може бути адаптована до іншої країни за умови заміни відповідного нормативного контексту та регіональних довідникових даних.

Як зазначено у підрозділі 2.1, використання синтетичних даних зумовлено об'єктивними обмеженнями доступу до реальних даних у сфері ГЗН: адміністративні записи поліції та медичних закладів є конфіденційними, а дані НУО - фрагментованими та індивідуальними за форматом кожної організації [29; 34]. Синтетичний підхід дозволяє зберегти методологічну коректність дослідження, відтворюючи структурні властивості та реалістичні розподіли без ризику порушення конфіденційності постраждалих осіб [32; 33]. Таким чином, сформований масив є методологічно репрезентативною основою для демонстрації повного аналітичного циклу системи.

Основний масив містить 1 353 записи про інциденти ГЗН і відповідає структурі інтегрованого масиву, визначеній у табл. 2.1. Записи охоплюють інциденти від трьох інституційних джерел (поліції, медичних закладів та НУО) і містять поля числового, текстового, категоріального та логічного типів. Наповнення масиву відтворює розподіли та пропорції, що відповідали б реальному інтегрованому масиву даних моніторингу ГЗН, сформованому згідно

з попередньо визначеними методологічними вимогами. Фрагмент масиву у середовищі Power Query наведено на рис. 3.5.

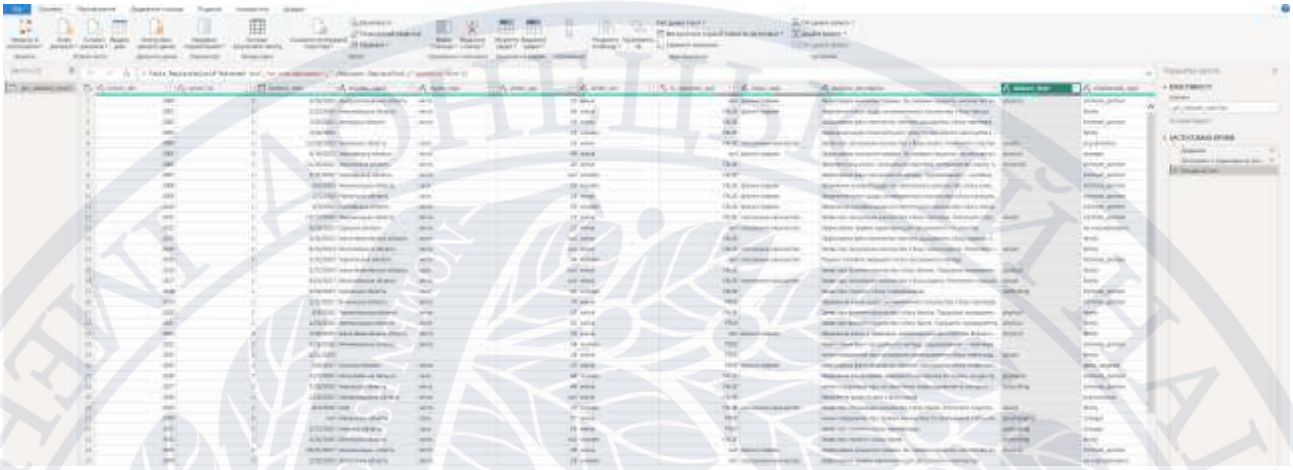

Рис. 3.5 – Фрагмент масиву даних у середовищі Power Query

Відповідно до моделі системи, демографічні дані є окремим вхідним потоком системи, що на поточному етапі використовуються для збагачення робочого масиву. Для забезпечення нормалізації показників поширеності цей вхід реалізовано у вигляді довідникового файлу region_population, що містить демографічні показники 25 регіонів України - без урахування Автономної Республіки Крим та міста Севастополь. Зв'язок між таблицями встановлено за ключовим полем location_name у середовищі Power BI відношенням «один-до-багатьох» кожному регіону в таблиці region_population відповідає множина записів у таблиці фактів. Слід зазначити, що демографічні дані є наближеними: точні актуальні показники чисельності населення в умовах воєнного стану є обмежено доступними, тому вони використовуються виключно для демонстрації аналітичних принципів. Розширення демографічної довідникової таблиці актуальними офіційними даними є одним із напрямів подальшого розвитку системи.

Встановлення зв'язку між таблицями та результуючу модель даних наведено на рис. 3.6.

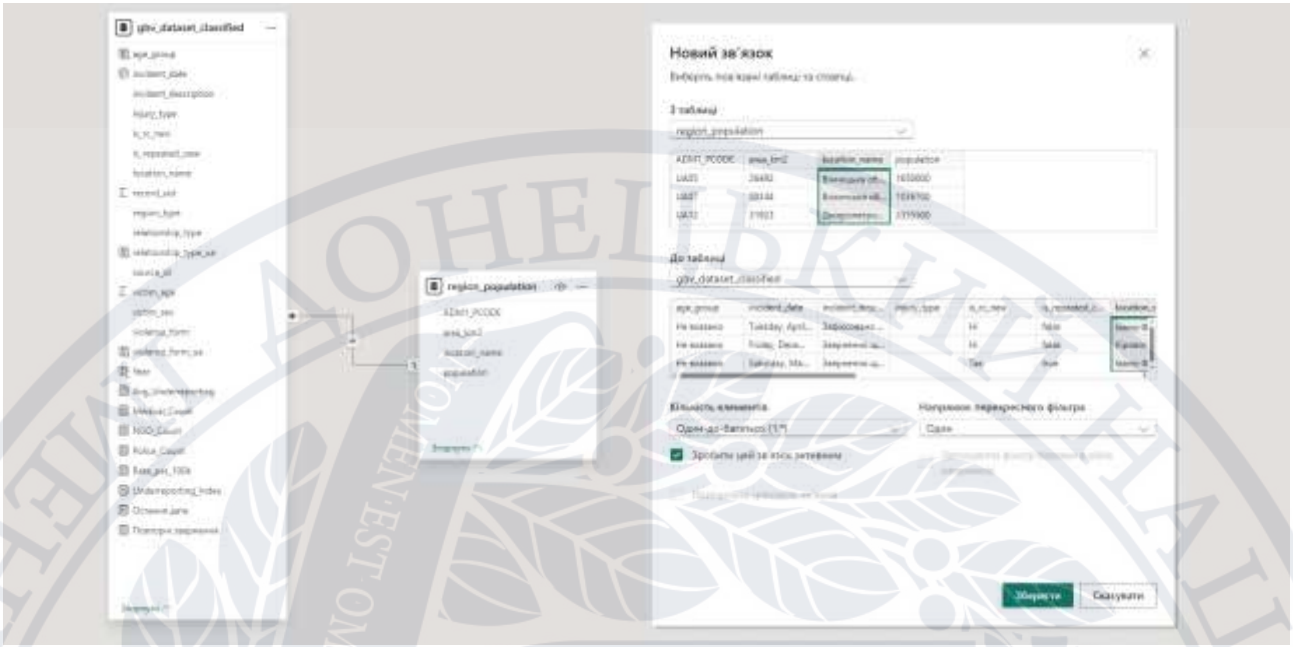

Рис. 3.6 – Модель даних у середовищі Power BI: зв'язок між таблицями

Сформована модель даних є основою аналітичного компонента системи. На її базі побудовано три сторінки інтерактивного звіту, кожна з яких відповідає окремому аналітичному виходу функціональної моделі IDEF0.

Слід зазначити, що у більшості візуалізацій звіту застосовано фільтрацію порожніх значень - записи з незаповненими ключовими категоріальними полями виключаються з розрахунків відповідної візуалізації. Це є відомим методологічним обмеженням, зазначеним у підрозділі 2.2: неповна заповненість окремих полів є очікуваною характеристикою інтегрованих адміністративних даних.

Перша сторінка інтерактивного звіту реалізує перший аналітичний вихід функціональної моделі - «Аналітичні звіти» - і забезпечує загальну картину динаміки та структури звернень у розрізі ключових категоріальних вимірів. Загальний вигляд сторінки наведено на рис. 3.7.

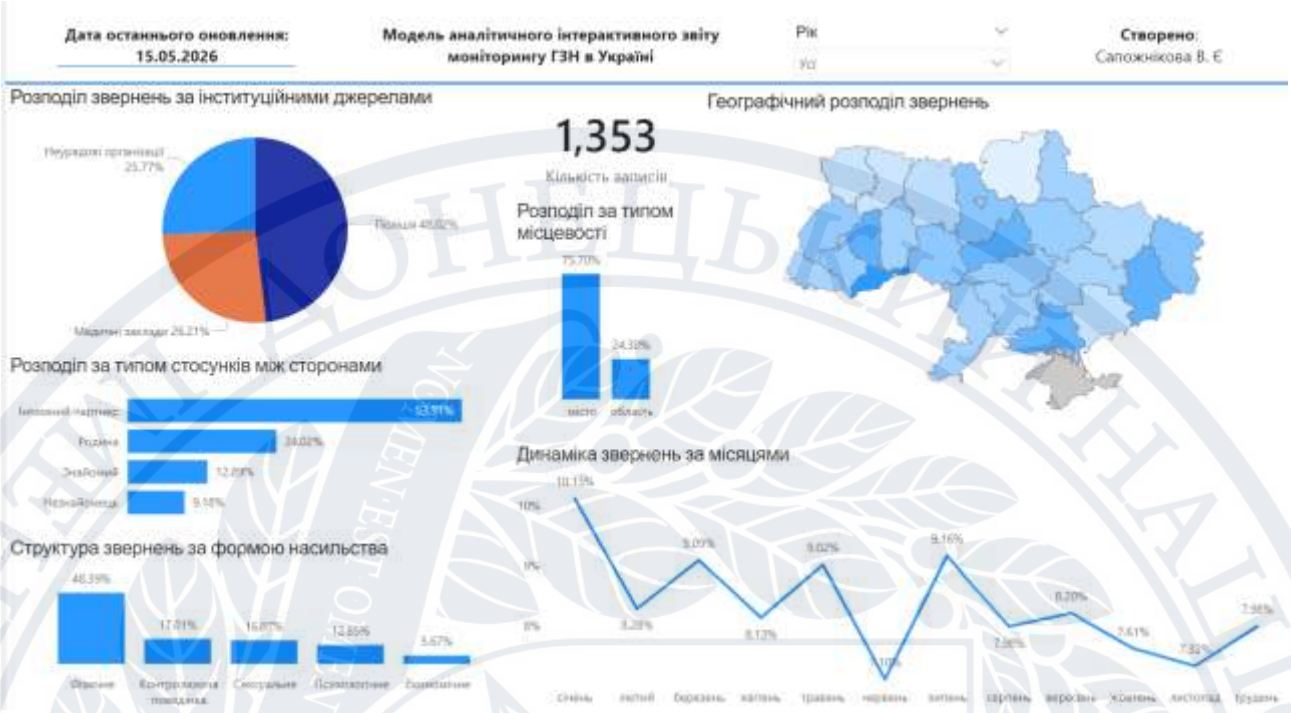

Рис. 3.7 – Сторінка аналітичних звітів інтерактивної моделі моніторингу ГЗН

Центральним елементом сторінки є картка із загальною кількістю записів у поточному контексті фільтрації - 1,353. Значення розраховується динамічно і змінюється відповідно до активних фільтрів зрізів, відображаючи актуальний обсяг вибірки в реальному часі.

Карта із заливкою «Географічний розподіл звернень» візуалізує регіональний розподіл звернень по 25 областях та місту Київ через інтенсивність кольорового відтінку - темніший відтінок відповідає вищій концентрації звернень у регіоні. Саме цей тип візуалізації дозволяє миттєво виявляти регіони з підвищеним рівнем звернень і дозволяє пріоритизувати регіони з вищою концентрацією звернень для розгортання додаткових кризових центрів та мобільних бригад. Автономна Республіка Крим та місто Севастополь відображаються сірим кольором як регіони, що не охоплені масивом даних. Лінійний графік «Динаміка звернень за місяцями» відображає розподіл звернень протягом року у відсотках від загальної кількості. Фіксується пік у січні (10,13%) зі спадом у червні (7,10%) та відносною стабілізацією в другій половині року. Графік демонструє здатність системи виявляти часові закономірності, що є підставою для посилення кампаній інформування та збільшення штату гарячих ліній у передпіковий період.

Стовпчикова діаграма «Розподіл за типом місцевості» показує що 75,70% звернень надходять з міських населених пунктів проти 24,30% із сільської місцевості. Цей вимір є важливим для аналітичного виходу «Рекомендації розподілу ресурсів», оскільки дозволяє виявляти диспропорції між міськими та сільськими територіями і обґрунтовувати спрямування ресурсів на розвиток сільської інфраструктури підтримки - мобільних бригад, телемедицини та дистанційних консультацій.

Кругова діаграма «Розподіл звернень за інституційними джерелами» відображає структуру масиву за полем source_id: поліція формує найбільшу частку - 48,02%, медичні заклади - 26,21%, НУО - 25,77%. Розподіл між джерелами відображає типову для систем моніторингу ГЗН асиметрію між офіційною та позаінституційною статистикою, що є методологічною основою для розрахунку індикаторів прихованої статистики на третій сторінці звіту. Крім того, регіони з нижчою за середню часткою НУО-звернень можуть потребувати додаткового фінансування місцевих організацій

Горизонтальна стовпчикова діаграма «Розподіл за типом стосунків між сторонами» відображає структуру масиву за полем relationship_type. Оскільки вхідні дані реєструються англійською мовою відповідно до міжнародного стандарту категоризації, для забезпечення україномовної звітності створено обчислюваний стовпець, що перетворює англомовні значення даного поля на локалізовані мітки. DAX-формулу обчислюваного стовпця наведено на рис. 3.8.

```
relationship_type_ua =
SWITCH(gbv_dataset_classified[relationship_type],
    "intimate_partner", "Інтимний партнер",
    "family", "Родина",
    "lgbtq_targeted", "ЛГБТК+",
    "acquaintance", "Знайомий",
    "stranger", "Незнайомець",
)
```

Рис. 3.8 – DAX-формула обчислюваного стовпця relationship_type_ua

За результатами, більше половини звернень (53,91%) стосуються насильства з боку інтимного партнера, родинне насильство складає 24,02%,

звернення щодо знайомих - 12,89%, незнайомців - 9,18%. Така структура ілюструє можливості системи щодо виявлення домінуючих типів стосунків у розрізі інцидентів, що є підставою для створення спеціалізованих програм для жертв насильства з боку інтимного партнера та програм для кривдників як додаткового заходу захисту.

Стовпчикова діаграма «Структура звернень за формою насильства» відображає розподіл інцидентів за категоріями класифікаційної схеми, визначеної у підрозділі 2.2. Аналогічно до попередньої візуалізації, для відображення застосовано обчислюваний стовпець з локалізованими україномовними мітками, сформований на основі англомовного поля violence_form. Фізичне насильство складає найбільшу частку - 48,39%, що майже втричі перевищує наступні категорії. Контролююча поведінка (17,01%), сексуальне насильство (16,07%) та психологічне насильство (12,85%) демонструють близькі значення, що вказує на відносно рівномірне охоплення цих форм у масиві. Економічне насильство представлено найменше - 5,67%. Розподіл демонструє здатність системи фіксувати повний спектр форм насильства відповідно до міжнародної класифікаційної схеми, що є принциповою перевагою порівняно з системами, які обмежуються лише правовими категоріями [7]. Зокрема, висока частка контролюючої поведінки є підставою для впровадження навчання для поліції та суддів щодо її виявлення - вона часто залишається поза офіційним обліком.

Лінійний графік «Динаміка звернень за місяцями» відображає розподіл звернень протягом року у відсотках від загальної кількості. Фіксується виражений пік у січні (10,13%) з подальшим спадом до мінімального значення у червні (7,10%). Друга половина року демонструє відносну стабілізацію у діапазоні 7,32-9,16%, з локальним піком у липні (9,16%) та поступовим зниженням до грудня (7,98%). Така хвилеподібна динаміка є характерною для часових рядів звернень і підтверджує доцільність включення місячного виміру до аналітичної моделі - він дозволяє виявляти сезонні закономірності, що є

аналітично значущим для планування інституційного реагування та розподілу ресурсів підтримки.

Якщо перша сторінка забезпечує загальну картину динаміки та структури звернень, то друга зосереджується на демографічному вимірі, адже відповідає аналітичному виходу «Виявлені групи ризику» і дозволяє перейти від агрегованих показників до профілю конкретних категорій постраждалих. Загальний вигляд сторінки наведено на рис. 3.9.

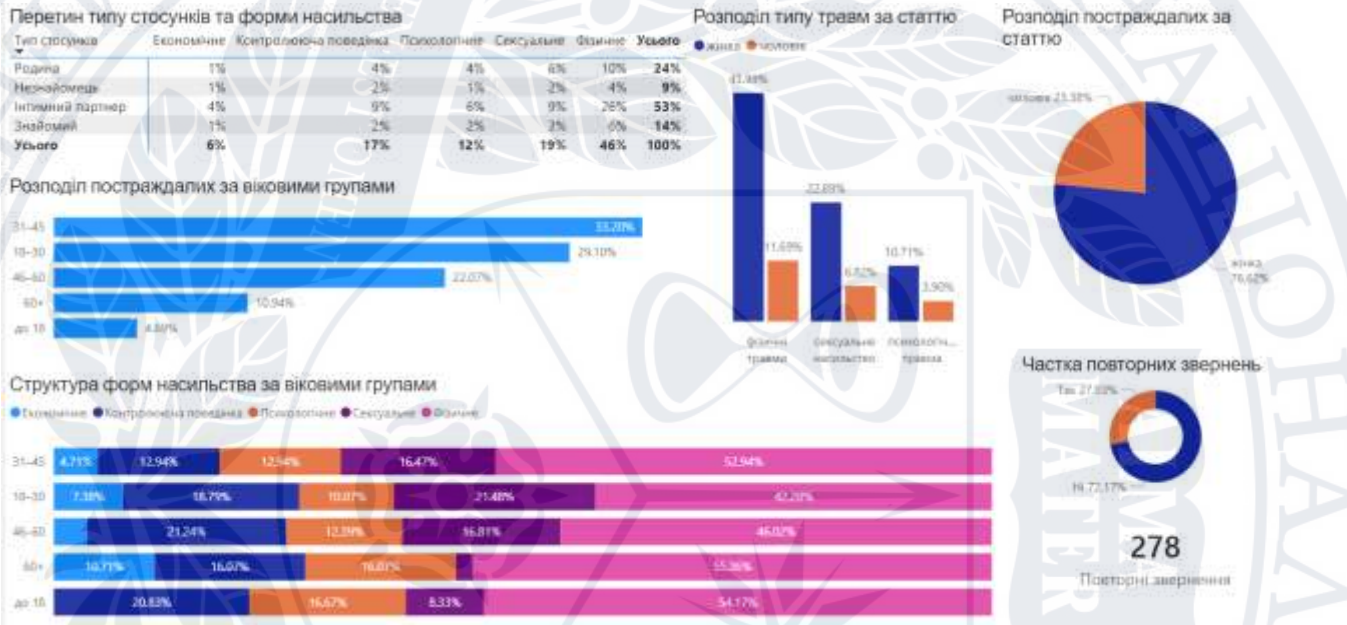

Рис. 3.9 – Сторінка профілю постраждалих інтерактивної моделі моніторингу ГЗН

Кругова діаграма «Розподіл постраждалих за статтю» показує що 76,62% постраждалих є жінками, 23,38% - чоловіками. Стовпчаста діаграма «Розподіл типу травм за статтю» деталізує цей розподіл у розрізі зафіксованих медичних наслідків: фізичні травми складають найбільшу частку як серед жінок (43,99%), так і серед чоловіків (11,69%), сексуальне насильство - 22,89% і 6,82% відповідно, психологічна травма - 10,71% і 3,90%. Частка чоловіків серед постраждалих є підставою для створення окремих програм підтримки, оскільки наявні послуги часто орієнтовані переважно на жінок. Ці дві візуалізації разом ілюструють можливості системи щодо формування гендерного профілю постраждалих - у наведеному прикладі вони демонструють структурну

нерівномірність у характері наслідків між статями, яка у реальному масиві даних могла б слугувати підставою для диференційованих програм підтримки.

Зведена матриця «Перетин типу стосунків та форми насильства» є аналітично центральним елементом сторінки - вона відображає розподіл інцидентів за двома класифікаційними вимірами одночасно і дозволяє виявляти стійкі комбінації, що є маркерами підвищеного ризику. Зокрема, інтимний партнер формує 53% усіх звернень, причому 26% з них - це фізичне насильство, а 9% - контролююча поведінка та сексуальне насильство відповідно. Родинне насильство (24%) також переважно має фізичний характер (10%). Саме цей тип візуалізації дозволяє перейти від одновимірних розподілів до структурного аналізу взаємозв'язків між категоріями, що є основою для адресного формування профілів підтримки - розробки спеціалізованого протоколу оцінки ризику для випадків насильства з боку інтимного партнера з фізичним характером, що є найбільш поширеною комбінацією..

Гістограма «Розподіл постраждалих за віковими групами» відображає вікову структуру масиву. Оскільки вихідне поле victim_age містить числові значення віку, для змістовної аналітичної візуалізації доцільно згрупувати їх у вікові діапазони - саме для цього було створено обчислюваний стовпець age_group, який за допомогою функції SWITCH присвоює кожному запису відповідну вікову категорію. DAX-формулу обчислюваного стовпця наведено на рис. 3.10.

```
1 age_group =
2 SWITCH(TRUE(),
3     gbv_dataset_classified[victim_age] <= 17, "до 18",
4     gbv_dataset_classified[victim_age] <= 30, "18-30",
5     gbv_dataset_classified[victim_age] <= 45, "31-45",
6     gbv_dataset_classified[victim_age] <= 60, "46-60",
7     "60+"
8 )
```

Рис. 3.10 – DAX-формула обчислюваного стовпця age_group

Найчисельнішою є група 31-45 років (33,20%), за нею - 18-30 років (29,10%) та 46-60 років (22,07%). Особи старшого віку (60+) складають 10,94%,

неповнолітні - 4,69%. Така вікова деталізація є аналітично значущою для адаптації комунікаційних кампаній та форм підтримки під вікові групи - онлайн-консультування для молоді та спрощені формати для осіб старшого віку.

Нормована гістограма з накопиченнями «Структура форм насильства за віковими групами» є логічним продовженням попередньої діаграми - якщо та показує кількісний розподіл постраждалих за віком, то ця розкриває якісний вимір: яким чином форми насильства розподіляються всередині кожної вікової групи. У наведеному прикладі фізичне насильство домінує в усіх групах, проте його частка зростає для осіб старшого віку (60+: 55,36%), тоді як контролююча поведінка є відносно вищою для груп 18-30 (18,79%) та 46-60 (21,24%). Для найменшої групи - до 18 років - помітна вища частка економічного насильства (20,83%), що може вказувати на специфіку залежності неповнолітніх. Поєднання вікового і формального вимірів в одній візуалізації дозволяє формувати диференційовані профілі ризику для кожної вікової категорії окремо.

Кільцева діаграма «Частка повторних звернень» та картка з абсолютним значенням ілюструють можливості системи щодо відстеження повторних звернень. У наведеному прикладі 27,83% записів (278 з 1 353) позначені як повторні звернення. Цей показник реалізовано через міру Repeat_Rate, що розраховується як відношення кількості записів з ознакою is_repeated_case = TRUE до загальної кількості записів у поточному контексті фільтрації. Висока частка повторних звернень є підставою для впровадження обов'язкового повторного контакту з постраждалими через певний час після першого звернення.

Третя сторінка реалізує аналітичний вихід «Рекомендації розподілу ресурсів» і зосереджується на оцінці повноти інституційної звітності через міжджерельний аналіз та індикатори прихованої статистики. Загальний вигляд сторінки наведено на рис. 3.11.

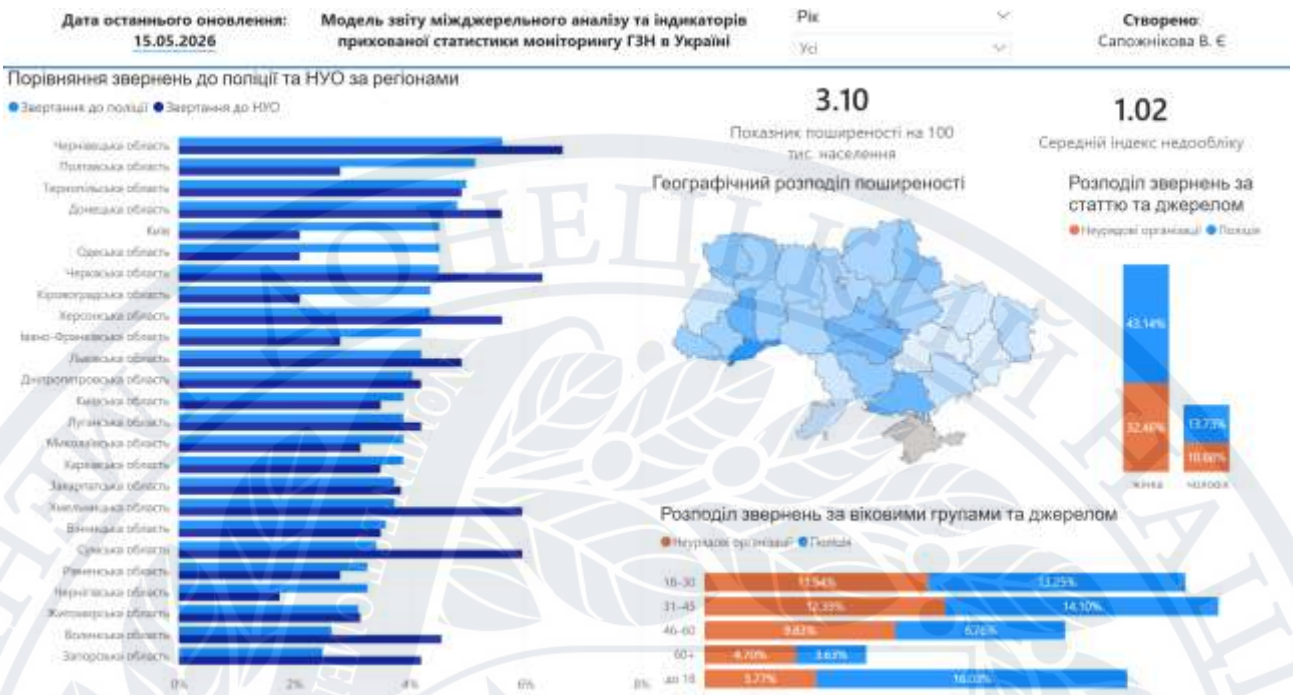

Рис. 3.11 – Сторінка міжджерельного аналізу та індикаторів прихованої статистики

Гістограма «Порівняння звернень до поліції та НУО за регіонами» відображає для кожного з 25 регіонів співвідношення між часткою поліцейських звернень і часткою звернень до НУО від загального обсягу. Ця візуалізація є відправною точкою для ідентифікації регіонів з найбільшою асиметрією між офіційною та позаінституційною статистикою. У наведеному прикладі помітна варіативність між регіонами - в окремих областях частка звернень до НУО суттєво перевищує частку поліцейських, що в реальному масиві даних слугувало б сигналом для перегляду інституційних механізмів реагування.

Ключовим аналітичним елементом сторінки є індекс недообліку - показник, що характеризує розрив між позаінституційними та офіційними зверненнями у конкретному регіоні. Концепцію прихованої статистики вперше систематизовано у публікації «On exploring the "dark figure" of crime» [18], де відношення між незареєстрованими та зареєстрованими випадками описується як:

$$D = \frac{N_{\text{незареєстровані}}}{N_{\text{зареєстровані}}} \text{ [18]} \qquad (3.1)$$

Оскільки у контексті моніторингу ГЗН прямий облік незареєстрованих випадків є неможливим, звернення до НУО та медичних закладів слугують

проксі-індикатором позаінституційного шару статистики. Відповідно, у цій роботі формулу адаптовано до наявної структури джерел даних, де чисельник представлено сукупністю позаінституційних звернень, а знаменником - офіційно зареєстрованими поліцейськими зверненнями:

$$I_u = \frac{N_{\text{НУО}} + N_{\text{мед}}}{N_{\text{поліція}}} \quad (3.2)$$

де I_u - індекс недообліку для відповідного регіону, $N_{\text{НУО}}$ - кількість звернень до НУО, $N_{\text{мед}}$ - кількість звернень до медичних закладів, $N_{\text{поліція}}$ - кількість поліцейських звернень у відповідному регіоні. Звернення до НУО та медичних закладів обрано як чисельник, оскільки саме ці джерела фіксують випадки, що з більшою ймовірністю залишаються поза офіційною поліцейською статистикою, відповідно до методологічних рекомендацій EIGE [29; 31]. У середовищі Power BI індекс представлено через міру Underreporting_Index, реалізацію якої наведено на рис. 3.12.

```
1 Underreporting_Index =
2 DIVIDE([NGO_Count] + [Medical_Count], [Police_Count])
```

Рис. 3.12 – DAX-формула міри Underreporting_Index

Узагальнене значення по всіх регіонах відображається через картку «Середній індекс недообліку», розрахований через міру Avg_Underreporting з використанням ітераторної функції AVERAGEX - вона обчислює індекс окремо для кожного регіону і лише потім усереднює результати, що є методологічно коректнішим підходом порівняно з усередненням загальних агрегатів, яке викривляло б результат через нерівномірний розподіл обсягів між регіонами. Узагальнений показник дозволяє спрямовувати адвокаційні зусилля на впровадження стандартизованих протоколів фіксації між інституційними джерелами. DAX-формулу міри Avg_Underreporting наведено на рис. 3.13.

```
1 Avg_Underreporting =
2 AVERAGEX(
3     VALUES(gbv_dataset_classified[location_name]),
4     [Underreporting_Index]
5 )
```

Рис. 3.13 – DAX-формула міри Avg_Underreporting

Значення індексу інтерпретується відносно порогового значення 1: якщо $I_u > 1$, позаінституційних звернень зафіксовано більше ніж поліцейських, що є сигналом потенційного недообліку в офіційній статистиці відповідного регіону; якщо $I_u \approx 1$ або $I_u < 1$- джерела демонструють відносну рівновагу. Регіони з систематично високим індексом є підставою для ініціювання навчання поліцейських щодо прийому заяв про ГЗН.

Картка «Показник поширеності на 100 тис. населення» є нормалізованим індикатором порівнянності між регіонами різного розміру [28; 31]. Математично показник визначається як:

$$R = \frac{N}{P} \times 100\,000 \text{ [28]} \tag{3.3}$$

де R - показник поширеності, N - кількість звернень у регіоні, P - чисельність населення регіону. У DAX цю формулу реалізовано через міру Rate_per_100k, що наведено на рис. 3.14.

```
Rate_per_100k =
DIVIDE(
    COUNT(gbv_dataset_classified[record_uid]),
    CALCULATE(
        SUM(region_population[population]),
        TREATAS(
            VALUES(gbv_dataset_classified[location_name]),
            region_population[location_name]
        )
    )
) * 100000
```

Рис. 3.14 – DAX-формула міри Rate_per_100k

Карта із заливкою «Показник поширеності за регіонами» візуалізує цей показник у географічному розрізі - темніший відтінок відповідає вищій концентрації звернень відносно чисельності населення регіону, що дозволяє виявляти регіони з непропорційно високим рівнем зафіксованих інцидентів.

У сукупності карта та картка дозволяють порівнювати регіони не за абсолютними числами а за нормалізованим показником - регіони з низьким

абсолютним числом але високим показником поширеності потребують не менш пильної уваги.

Стовпчаста діаграма «Розподіл звернень за статтю та джерелом» та діаграма «Розподіл звернень за віковими групами та джерелом» розкривають демографічний профіль звернень у розрізі інституційних джерел. У наведеному прикладі помітна відносна рівномірність гендерного розподілу між джерелами, тоді як вікова структура демонструє певні відмінності. У реальному масиві такі відмінності могли б свідчити про потребу у створенні молодіжних консультаційних просторів або цифрових каналів звернення для вікових груп що частіше звертаються до НУО ніж до поліції..

Кожна з трьох моделей інтерактивних звітів містить уніфіковану верхню панель, що виконує функцію навігаційного та інформаційного заголовку. Панель включає назву конкретної сторінки звіту, картку дати останнього оновлення та роздільник за роками, що забезпечує єдиний інтерфейс фільтрації для всього звіту незалежно від активної сторінки.

Поле дати останнього оновлення реалізовано як текстове поле з динамічним значенням - статичний підпис «Дата останнього оновлення:» міститься безпосередньо у текстовому полі, тоді як числове значення дати підтягується через міру Остання дата (рис. 3.15). Міра визначає максимальне значення поля incident_date у поточному контексті фільтрації та форматує його у зручний для відображення вигляд. Ключовою перевагою цього підходу є автоматичне оновлення значення при розширенні масиву новими записами без будь-якого ручного втручання.

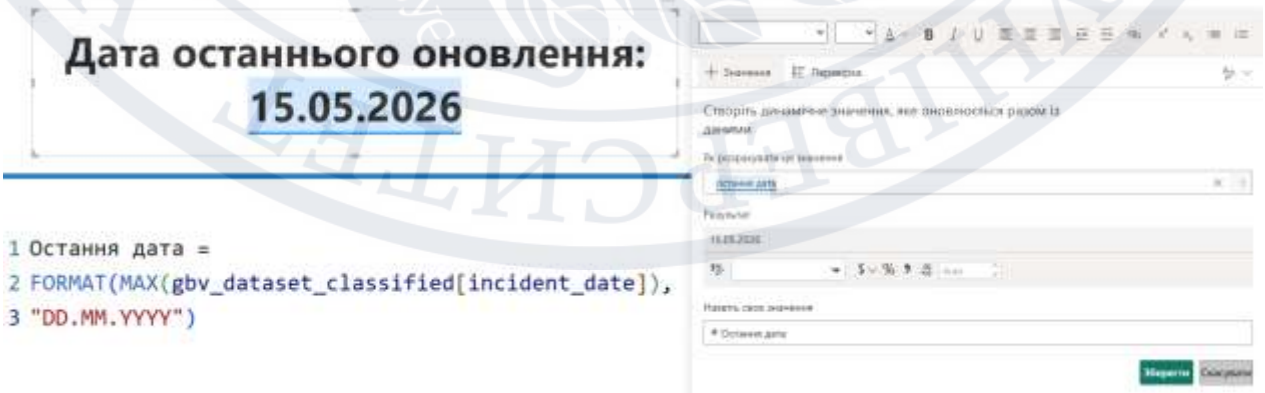

Рис. 3.15 – Реалізація динамічного поля дати останнього оновлення

Роздільник за роками реалізовано через обчислюваний стовпець Year, що виокремлює рік з поля incident_date (рис. 3.16).

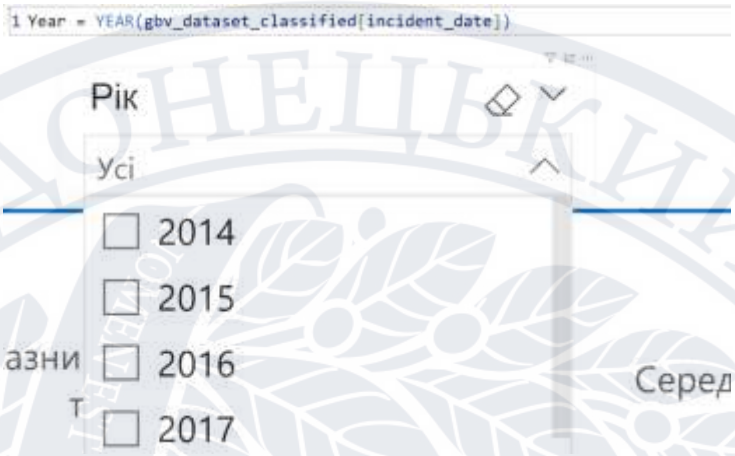

Рис. 3.16 – Реалізація роздільника за роками

Його присутність на кожній сторінці забезпечує можливість порівняльного аналізу в часовому розрізі - обираючи конкретний рік, користувач отримує автоматичний перерахунок усіх візуалізацій звіту відповідно до обраного періоду, що дозволяє відстежувати динаміку показників між роками.

Усі візуалізації звіту інтерактивні - вибір будь-якого елементу на одній діаграмі автоматично фільтрує решту візуалізацій поточної сторінки, що дозволяє отримати деталізований профіль обраної підвибірки без додаткових налаштувань. Для ілюстрації цієї властивості на рис. 3.17 наведено приклад сторінки 1 з активованим регіональним фільтром.

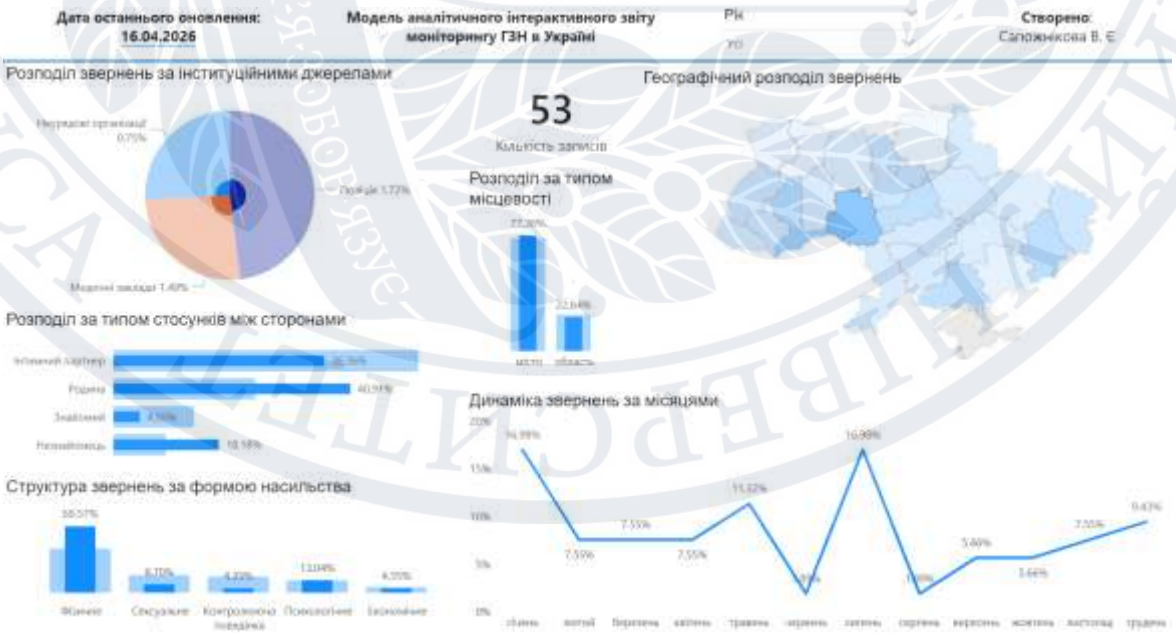

Рис. 3.17 – Демонстрація інтерактивної фільтрації

Реалізована аналітична модель не обмежується статичним відображенням - платформа Power BI забезпечує гнучкі можливості поширення звіту відповідно до потреб користувачів. Звіт може бути опублікований у Power BI Service для онлайн-доступу через браузер без встановлення додаткового програмного забезпечення, вбудований у вебсторінку або портал через iframe, а також експортований у форматах PDF або PowerPoint для офлайн-представлення та включення до офіційних звітних документів [40].

Розроблена в роботі інформаційно-аналітична модель моніторингу гендерно зумовлених ситуацій має безпосередню практичну цінність для органів державної влади, місцевого самоврядування, соціальних служб та неурядових організацій. Вона дозволяє перейти від фрагментарного обліку випадків ГЗН до управління на основі даних (data-driven decision making), забезпечуючи:

1. Виявлення регіонів із критичним рівнем недообліку звернень (індекс недообліку >1) для цільового перегляду інституційних механізмів реагування.
2. Ідентифікацію груп ризику за статтю, віком, типом насильства та стосунків для розробки адресних програм підтримки.
3. Обґрунтування розподілу фінансових і матеріальних ресурсів між регіонами та між міськими й сільськими територіями.
4. Планування відкриття кризових центрів у зонах із високою концентрацією звернень та низьким доступом до служб.
5. Моніторинг ефективності міжвідомчої координації через зіставлення даних поліції, медицини та НУО.

Таким чином, модель є дієвим інструментом підтримки прийняття рішень для координаторів соціальних служб та НУО, а також для адвокаційної діяльності щодо посилення протидії ГЗН.

Висновки до розділу 3

У третьому розділі реалізовано практичну складову інформаційно-аналітичної системи моніторингу ГЗН відповідно до концептуальної основи, сформованої у попередніх розділах. Побудовано трирівневу функціональну модель системи в нотації IDEF0, що охоплює контекстну діаграму A-0, діаграму

декомпозиції A0 та дві додаткові декомпозиції блоків A1 і A3. Встановлено повну узгодженість між рівнями ієрархії моделі - всі три аналітичні виходи контекстної діаграми простежуються до конкретних функціональних блоків декомпозиції. Сформовано синтетичний масив даних обсягом 1 353 записів, що відтворює структуру верифікованого інтегрованого масиву відповідно до схеми, визначеної у підрозділі 2.2. Реалізовано аналітичний компонент системи у середовищі Power BI: побудовано семантичну модель даних, розроблено набір DAX-мір і обчислюваних стовпців та створено три сторінки інтерактивного звіту, що відповідають аналітичним виходам IDEF0-моделі - «Аналітичні звіти», «Виявлені групи ризику» та «Рекомендації розподілу ресурсів». Для ключового аналітичного індикатора системи - індексу недообліку - здійснено адаптацію методологічного підходу до оцінки прихованої статистики через зіставлення даних різних інституційних джерел до структури наявного масиву.

ВИСНОВКИ

У кваліфікаційній (бакалаврській) роботі вирішено задачу моделювання інформаційно-аналітичної системи моніторингу гендерно зумовлених ситуацій та реалізовано її аналітичний компонент для підтримки прийняття управлінських рішень. За результатами дослідження отримано такі висновки.

Проаналізовано природу гендерно зумовлених ситуацій як соціального явища та систематизовано форми ГЗН відповідно до міжнародної нормативно-правової бази. Проведено порівняльний аналіз шести існуючих інформаційних систем моніторингу ГЗН і виявлено їхню спільну структурну проблему - нездатність консолідувати дані з різних інституційних джерел в єдиний аналітичний простір, що унеможливлює оцінку прихованої статистики звернень.

Обґрунтовано вибір методології IDEF0 для функціонального моделювання системи шляхом порівняльного аналізу з альтернативними нотаціями DFD та BPMN. Встановлено, що IDEF0 є найбільш придатною нотацією для моделювання інформаційно-аналітичних систем з кількома інституційними джерелами даних, оскільки єдина серед розглянутих методологій підтримує концептуальне проєктування, ієрархічну декомпозицію, незалежність від реалізації та придатність для міжінституційних систем одночасно.

Запропоновано структуру інтегрованого масиву даних із двовимірною класифікаційною схемою за формою насильства та типом стосунків між сторонами, що охоплює дванадцять полів і забезпечує повноту охоплення явища незалежно від інституційного джерела. Обґрунтовано використання синтетичного датасету як методологічно прийнятного підходу в умовах обмеженого доступу до реальних даних. Запропоновано аналітичний інструментарій системи та обґрунтовано теоретичні засади його ключових компонентів: Power Query для стандартизації різнорідних вхідних даних на ETL-етапі, семантичної моделі для структурування зв'язків між таблицями та мови DAX для реалізації аналітичних обчислень.

Побудовано трирівневу функціональну модель системи в нотації IDEF0, що охоплює контекстну діаграму A-0, діаграму декомпозиції A0 та дві додаткові

декомпозиції блоків A1 і A3. Встановлено повну узгодженість між рівнями ієрархії - всі три аналітичні виходи контекстної діаграми простежуються до конкретних функціональних блоків декомпозиції. Побудована модель може слугувати архітектурним орієнтиром при адаптації системи до реальних інституційних джерел.

Розроблено аналітичну модель у середовищі Power BI: побудовано семантичну модель даних, розроблено набір DAX-мір і обчислюваних стовпців та створено три сторінки інтерактивного звіту, що відповідають аналітичним виходам IDEF0-моделі - «Аналітичні звіти», «Виявлені групи ризику» та «Рекомендації розподілу ресурсів». Для ключового індикатора системи - індексу недообліку - здійснено адаптацію методологічного підходу до оцінки прихованої статистики через зіставлення даних різних інституційних джерел. Реалізований звіт забезпечує інтерактивну фільтрацію та може бути опублікований у Power BI Service або експортований у форматах PDF і PowerPoint.

Розроблена модель має безпосередню практичну цінність для органів державної влади, соціальних служб та неурядових організацій - вона дозволяє перейти від фрагментарного обліку випадків ГЗН до управління на основі даних, забезпечуючи виявлення груп ризику, обґрунтування розподілу ресурсів між регіонами і моніторинг ефективності міжвідомчої координації через зіставлення даних різних інституційних джерел.

Перспективами подальшого розвитку системи є інтеграція з реальними інституційними джерелами даних, розширення демографічної довідникової таблиці актуальними офіційними даними - деталізацією до рівня громади, додаванням поля часу між інцидентом та зверненням і ознаки повторного кривдника, а також застосування методів обробки природної мови для автоматичної категоризації текстових описів інцидентів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Krug E.G., Dahlberg L.L., Mercy J.A., Zwi A.B., Lozano R. World Report on Violence and Health. Geneva, 2002. URL: https://www.who.int/publications/i/item/9241545615 (дата звернення: 04.05.2026).
2. Council of Europe. Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention). Strasbourg, 2011. URL: https://www.coe.int/en/web/istanbul-convention/about-the-convention (дата звернення: 04.05.2026).
3. CEDAW Committee. General Recommendation No. 19 on Violence against Women. New York, 1992. URL: https://www.refworld.org/legal/resolution/cedaw/1992/96542 (дата звернення: 04.05.2026).
4. Sapozhnikova V. Gender-Based Violence in Ukraine in the Context of Armed Conflict. Topical Issues of Humanities, Technical, and Natural Sciences : тези X Міжнар. наук. конф. студентів, аспірантів та молодих вчених (30 квітня 2026 р.). Вінниця : ДонНУ імені Василя Стуса, 2026. (сертифікат учасника, сертифікат переможника)
5. Eurostat, FRA, EIGE. EU Gender-based Violence Survey: Evidence for Policy and Practice. Vilnius, 2026. URL: https://eige.europa.eu/publications-resources/publications/eu-gender-based-violence-survey-evidence-policy-and-practice (дата звернення: 05.05.2026).
6. EIGE. Data Collection on Violence Against Women. Vilnius, 2025. URL: https://eige.europa.eu/gender-based-violence/data-collection (дата звернення: 05.05.2026).
7. Баценко Т. В., Сапожнікова В. Є. Аналіз систем моніторингу гендерно зумовленого насильства. *Наука і техніка сьогодні*. № 5(59), 2026.

8. Sheth A.P., Larson J.A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys. 1990. Vol. 22, No. 3. P. 183–236. URL: https://dl.acm.org/doi/10.1145/96602.96604 (дата звернення: 09.05.2026).
9. Ministerio del Interior de España. Sistema VioGén: Manual de usuario. Madrid, 2023. URL: https://sistemaviogen.ses.mir.es/publico/viogen/publicaciones.html (дата звернення: 09.05.2026).
10. González-Álvarez J.L., López-Ossorio J.J., Urruela C., Rodríguez-Díaz M. Integral Monitoring System in Cases of Gender Violence. VioGén System. Behavior & Law Journal. 2018. Vol. 4. No. 1. P. 29–40. URL: https://www.researchgate.net/publication/325579709 (дата звернення: 09.05.2026).
11. Echeburúa E., Amor P.J., Loinaz I., de Corral P. Escala de Predicción del Riesgo de Violencia Grave contra la Pareja. Psicothema. 2010. Vol. 22. No. 4. P. 1054–1060. URL: https://www.psicothema.com/pdf/3840.pdf (дата звернення: 09.05.2026).
12. Robinson A.L. Domestic Violence MARACs for Very High-Risk Victims in Cardiff, Wales: A Process and Outcome Evaluation. Cardiff, 2010. URL: https://www.researchgate.net/publication/237442284_Domestic_Violence_MARACs_Multi-Agency_Risk_Assessment_Conferences_for_Very_High-Risk_Victims_in_Cardiff_Wales_A_Process_and_Outcome_Evaluation (дата звернення: 09.05.2026).
13. SafeLives. Dash Risk Identification Checklist: Practice Guidance. Bristol, 2022. URL: https://safelives.org.uk/resources-library/dash-risk-checklist-for-idvas/ (дата звернення: 09.05.2026).
14. Inmon W.H. Building the Data Warehouse. New York, 1992. URL: https://books.google.com.ua/books?id=duRQAAAAMAAJ&redir_esc=y (дата звернення: 09.05.2026).
15. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Indianapolis, 2013. URL:

https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-toolkit/ (дата звернення: 09.05.2026).

16. UN Women. Progress on the Sustainable Development Goals: The Gender Snapshot 2023. New York, 2023. URL: https://www.unwomen.org/en/digital-library/publications/2023/09/progress-on-the-sustainable-development-goals-the-gender-snapshot-2023 (дата звернення: 09.05.2026).
17. Walby S., Towers J., Balderston S. et al. The Concept and Measurement of Violence Against Women and Men. Bristol, 2017. URL: https://policy.bristoluniversitypress.co.uk/the-concept-and-measurement-of-violence-against-women-and-men (дата звернення: 09.05.2026).
18. Biderman A.D., Reiss A.J. On exploring the "dark figure" of crime. The Annals of the American Academy of Political and Social Science. 1967. Vol. 374. P. 1–15. URL: https://ideas.repec.org/a/sae/anname/v374y1967i1p1-15.html (дата звернення: 09.05.2026).
19. GBVIMS Steering Committee. Gender-Based Violence Information Management System: Overview and Implementation Guide. New York, 2017. URL: https://www.gbvims.com/gbvims-tools/user-guide/ (дата звернення: 09.05.2026).
20. IRC/UNHCR. GBV Information Management System: User Guide. Geneva, 2017. URL: https://www.gbvims.com/wp/wp-content/uploads/User-Guide_ToPrintColor.pdf (дата звернення: 09.05.2026).
21. KoBoToolbox. About the Organization. Cambridge, 2024. URL: https://www.kobotoolbox.org/about-us/the-organization (дата звернення: 09.05.2026).
22. Нескородєва Т. В., Федоров Є. Є., Січко Т. В., Нескородєва А. Р. Експертні та рекомендаційні системи: навч. посіб. для здобувачів вищої освіти спеціальностей 122 «Комп'ютерні науки», 125 «Кібербезпека», 113 «Прикладна математика». Вінниця: ДонНУ імені Василя Стуса, 2023. 224 с.

23. Баценко Т. В., Сапожнікова В. Є. Вибір методу функціонального моделювання для інформаційно-аналітичних систем моніторингу // Прикладні інформаційні технології: тези доп. Учасників VII Всеукр. наук.-практ. конф. здобувачів вищої освіти та молодих вчених (ПІТ 2026). Вінниця : ДонНУ імені Василя Стуса, 2026.
24. Aleryani A. Analyzing Data Flow: A Comparison between Data Flow Diagrams and User Case Diagrams in Information Systems Development. 2024. URL: https://www.researchgate.net/publication/379476935 (дата звернення: 09.05.2026).
25. Chinosi M., Trombetta A. BPMN: An introduction to the standard // Computer Standards & Interfaces. 2012. Vol. 34. P. 124–134. URL: https://www.researchgate.net/publication/220619834_BPMN_An_introduction_to_the_standard (дата звернення: 09.05.2026).
26. IEEE. Standard for Functional Modeling Language - Syntax and Semantics for IDEF0. IEEE Std 1320.1-1998. New York : IEEE, 1998. URL: https://ieeexplore.ieee.org/document/749110 (дата звернення: 09.05.2026).
27. Li Q., Chen Y.L. IDEF0 Function Modeling. In: Modeling and Analysis of Enterprise and Information Systems. Berlin, Heidelberg: Springer, 2009. URL: https://link.springer.com/chapter/10.1007/978-3-540-89556-5_5 (дата звернення: 09.05.2026).
28. UNFPA. Guidelines for Producing Statistics on Violence Against Women. New York, 2014. URL: https://unstats.un.org/unsd/publication/seriesf/seriesF_110e.pdf (дата звернення: 10.05.2026).
29. EIGE. Quality Assessment of Administrative Data Sources on Gender-Based Violence. Vilnius : EIGE, 2019. URL: https://eige.europa.eu/publications/quality-assessment-administrative-data-sources-gender-based-violence (дата звернення: 10.05.2026).
30. Pratt T. C., Cullen F. T. Assessing Macro-Level Predictors and Theories of Crime: A Meta-Analysis of Meta-Analyses. Crime and Justice. 2005. Vol. 32.

P. 373–450. URL: https://www.researchgate.net/publication/288149655_Assessing_Macro-Level_Predictors_and_Theories_of_Crime_A_Meta-Analysis (дата звернення: 10.05.2026).

31. EIGE. Methodological Guide for Developing Gender Statistics. Vilnius, 2023. URL: https://eige.europa.eu/gender-statistics/dgs/methodology (дата звернення: 10.05.2026).
32. Gonzales A., Guruswamy G., Smith S.R. Synthetic data in health care: A narrative review. PLOS Digital Health. 2023. Vol. 2, No. 1. e0000082. URL: https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000082 (дата звернення: 18.05.2026).
33. Kokosi T., Harron K. Synthetic data in medical research. BMJ Medicine. 2022. Vol. 1, No. 1. e000167. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9951365/ (дата звернення: 18.05.2026).
34. Rothbard A. Quality Issues in the Use of Administrative Data Records. Philadelphia : University of Pennsylvania, 2008. URL: https://aisp.upenn.edu/wp-content/uploads/2015/06/Data-Quality-Paper_Final.pdf (дата звернення: 12.05.2026).
35. Fantuzzo J., Culhane D. Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government. New York, 2016. URL: https://aisp.upenn.edu/wp-content/uploads/2017/09/The-IDS-Approach_Fantuzzo-et-al.-2017_Final.pdf (дата звернення: 12.05.2026).
36. Halford E., Dixon A., Farrell G. et al. Improving police recorded crime data for domestic violence and abuse through natural language processing. Frontiers in Sociology. 2025. URL: https://www.frontiersin.org/journals/sociology/articles/10.3389/fsoc.2025.1686632/full (дата звернення: 12.05.2026).
37. Notko M., Holma J., Husso M. et al. The Documentation and Characteristics of Hospitalized IPV Patients Using Electronic Medical Records Data. Journal of

Family Violence. 2019. URL: https://link.springer.com/article/10.1007/s10896-019-00081-z (дата звернення: 12.05.2026).

38. EIGE. Mapping the Current Status and Potential of Administrative Data Sources on Gender-Based Violence in the EU. Vilnius, 2014. URL: https://eige.europa.eu/gender-based-violence/data-collection/mapping-current-status-and-potential-administrative-data-sources-gender-based-violence-eu (дата звернення: 12.05.2026).
39. Microsoft Corporation. What is Power Query? Microsoft Learn, 2024–2026. URL: https://learn.microsoft.com/en-us/power-query/power-query-what-is-power-query (дата звернення: 18.05.2026).
40. Microsoft Corporation. Power BI documentation. Microsoft Learn, 2024–2026. URL: https://learn.microsoft.com/en-us/power-bi/ (дата звернення: 18.05.2026).
41. Roorda E., Bruijnzeels M., Struijs J., Spruit M. Business intelligence systems for population health management: a scoping review. JAMIA Open. 2024. Vol. 7, Issue 4. Article ooae122. URL: https://academic.oup.com/jamiaopen/article/7/4/ooae122/7909841 (дата звернення: 18.05.2026).
42. Microsoft Corporation. Understand star schema and the importance for Power BI. Microsoft Learn, 2024. URL: https://learn.microsoft.com/en-us/power-bi/guidance/star-schema (дата звернення: 18.05.2026).
43. Microsoft Corporation. Model relationships in Power BI Desktop. Microsoft Learn, 2024. URL: https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-relationships-understand (дата звернення: 18.05.2026).
44. Microsoft Corporation. DAX overview. Microsoft Learn, 2024. URL: https://learn.microsoft.com/en-us/dax/dax-overview (дата звернення: 18.05.2026).
45. Microsoft Corporation. Context in DAX Formulas. Microsoft Support, 2024. URL: https://support.microsoft.com/en-us/office/context-in-dax-formulas-2728fae0-8309-45b6-9d32-1d600440a7ad (дата звернення: 18.05.2026).

46. Russo M., Ferrari A. The Definitive Guide to DAX: Business Intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel. 2nd ed. Redmond : Microsoft Press, 2020. 752 p.
47. Microsoft Corporation. CALCULATE function (DAX). Microsoft Learn, 2024. URL: https://learn.microsoft.com/en-us/dax/calculate-function-dax (дата звернення: 18.05.2026).
48. Microsoft Corporation. DAX function reference. Microsoft Learn, 2024. URL: https://learn.microsoft.com/en-us/dax/dax-function-reference (дата звернення: 18.05.2026).
49. Microsoft Corporation. AVERAGEX function (DAX). Microsoft Learn, 2024. URL: https://learn.microsoft.com/en-us/dax/averagex-function-dax (дата звернення: 18.05.2026).
50. Січко Т.В. Методи моделювання бізнес-процесів підприємства засобами системного аналізу. *Галицький економічний вісник*. 2016. № 2. С.190-201.