

Д. Ланде, О. Пучков

OSINT

ІЗ ЗАСТОСУВАННЯМ ВЕЛИКИХ
ДАНИХ І ШТУЧНОГО ІНТЕЛЕКТУ



**Державна служба спеціального зв'язку та захисту інформації
Інститут спеціального зв'язку та захисту інформації
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»**

Д. Ланде, О. Пучков

**OSINT
ІЗ ЗАСТОСУВАННЯМ ВЕЛИКИХ ДАНИХ І
ШТУЧНОГО ІНТЕЛЕКТУ**

Монографія

Київ – 2026

*Рекомендовано до друку вченою радою Інституту спеціального зв'язку та захисту інформації Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
(протокол № 13 від 22 червня 2026 р.)*

Рецензенти:

Петров В.В. – академік НАН України, заслужений діяч науки і техніки України, лауреат Державної премії України в галузі науки і техніки, доктор технічних наук, професор.

Палагін О.В. – академік НАН України, заслужений винахідник України, лауреат Державної премії України в галузі науки і техніки, доктор технічних наук, професор.

Ланде Д.В., Пучков О.О.

Л22 OSINT із застосуванням великих даних і штучного інтелекту:
монографія. – Київ: ІСЗІ КПІ ім. Ігоря Сікорського, 2026. – 444 с.

ISBN 978-617-7903-31-3.

Монографію присвячено дослідженню інтеграції трьох ключових напрямків сучасного аналізу даних: Big Data (великих даних), OSINT (Open Source Intelligence, розвідки у відкритих джерелах) та AI (Artificial Intelligence, штучного інтелекту), з акцентом на методи створення і застосування семантичних мереж (Semantic Networks) та використання великих мовних моделей (Large Language Models, LLM).

Виклад матеріалу побудовано за принципом послідовного переходу від фундаментальних понять та технологій роботи з великими даними до методів розвідки у відкритих джерелах, інтелектуального аналізу інформації, застосування штучного інтелекту, графових моделей знань і систем підтримки прийняття рішень. Така логіка забезпечує цілісне розуміння сучасних інформаційно-аналітичних технологій, дозволяє простежити взаємозв'язок між окремими методами й інструментами та створює надійну методологічну основу для їх практичного застосування у сфері OSINT.

Видання розраховано на дослідників, OSINT-аналітиків, фахівців із кібербезпеки, розробників інтелектуальних систем, а також науково-педагогічних працівників і здобувачів освіти у галузях аналізу даних, інформаційної безпеки та прикладного штучного інтелекту.

УДК 004.5/.7.052:004.8](02)

ЗМІСТ

ЗМІСТ	3
ВСТУП.....	14
РОЗДІЛ 1. ОСНОВИ BIG DATA.....	28
1.1. Поняття та характеристики Big Data.....	28
1.2. Архітектурні підходи до обробки Big Data.....	30
1.3. Екосистема Apache Hadoop	37
1.4. Сучасні фреймворки обробки даних	41
1.5. NoSQL-системи керування базами даних.....	44
1.6. Пошукові та аналітичні платформи	47
1.7. Графові бази даних у контексті Big Data	50
1.8. Інструментальне забезпечення інформаційної аналітики.....	52
1.9. Проблеми та обмеження сучасних Big Data-технологій.....	55
Висновки до розділу 1	57
РОЗДІЛ 2. РОЗВІДКА У ВІДКРИТИХ ДЖЕРЕЛАХ	59
2.1. Поняття та еволюція OSINT	60
2.2. Основні джерела відкритої інформації	63
2.3. Методи збору даних у OSINT.....	65
2.4. Аналітичні методи в OSINT	67
2.5. Технологічна інфраструктура OSINT	69
2.6. Ключові виклики та обмеження OSINT.....	71
2.7. Перспективи розвитку OSINT	73
Висновки до розділу 2	74
РОЗДІЛ 3. ДЖЕРЕЛА ВІДКРИТИХ ДАНИХ: ТЕКСТ, МЕТАДАНИ, СТРУКТУРОВАНІ ПЛАТФОРМИ.....	76
3.1. Класифікація джерел відкритої інформації	77
3.1.1. Текстові джерела: новини, блоги, наукові публікації, законодавчі акти.....	77
3.1.2. Метадані.....	78
3.1.3. Структуровані платформи	79

3.2. Особливості аналізу тестових джерел.....	80
3.2.1. Видобування чистого тексту із документів	80
3.2.2. Новинні сайти та інформаційні агрегатори	82
3.2.3. Наукові бази.....	83
3.3. Метадані як джерело контекстуальної інформації.....	85
3.3.1. RSS/Atom-фіди	85
3.3.2. JSON-LD та мікророзмітка веб-сторінок	87
3.3.3. Фільтрація, часових рядів, геолокації	88
3.4. Соціальні мережі та месенджери.....	89
3.4.1. Telegram	89
3.4.2. X (Twitter), Facebook, YouTube.....	91
3.4.3. Форуми та мікроблогінг.....	92
3.5. Технічні та інфраструктурні платформи.....	94
3.5.1. WHOIS, DNS, Shodan, Censys: інформація про мережеві ресурси	95
3.5.2. GitHub, GitLab: код, issues, коментарі як джерело знань	96
3.5.3. Відкриті картографічні сервіси.....	98
3.6. Офіційні та правові джерела.....	99
3.6.1. Законодавчі бази.....	100
3.6.2. Судові реєстри, антикорупційні бази, реєстри юридичних осіб..	102
3.6.3. Парламентські стенограми, звіти, бюджетні документи.....	104
3.7. Енциклопедичні та довідкові ресурси.....	106
3.7.1. Wikipedia	106
3.7.2. Wikidata, DBpedia	108
3.7.3. Спеціалізовані довідники.....	111
3.8. Структуровані, напівструктуровані та неструктуровані формати відкритих даних	114
3.8.1. XML та RDF.....	115
3.8.2. JSON та CSV	117
3.8.3. Неструктуровані та напівструктуровані текстові дані	120
3.9. Методологічні та технологічні обмеження аналізу відкритих даних	122
3.9.1. Фрагментарність, дублювання, нестача структури	123

3.9.2. Часті зміни API, блокування скрейперів, капчі	124
3.9.3. Якість мови: сленг, помилки, багатомовність, машинний переклад	125
3.10. Інтеграція, нормалізація та підготовка відкритих даних до семантичного моделювання	126
3.10.1. Уніфікація форматів: перетворення XML → JSON → CSV	127
3.10.2. Ідентифікація та нормалізація сутностей.....	128
3.10.3. Підготовка даних до формування семантичних мереж.....	129
Висновки до розділу 3	130
РОЗДІЛ 4. ГЕОПРОСТОРОВІ ДЖЕРЕЛА В OSINT	132
4.1. Роль геопросторових даних у сучасному OSINT	132
4.1.1. Геолокація як контекст.....	133
4.1.2. Типи геопросторової інформації	133
4.1.3. Обмеження та переваги відкритих геоданих у порівнянні з комерційними джерелами.....	134
4.2. OpenStreetMap як ключове джерело OSINT	135
4.2.1. Структура геопросторових даних в OpenStreetMap	136
4.2.2. Модель даних OSM	137
4.2.3. Платформи доступу до OpenStreetMap	139
4.3. Методи екстрагування об'єктів із OSM	140
4.3.1. Запити Overpass QL	140
4.3.2. Автоматизоване завантаження через Python.....	142
4.3.3. Формати подання екстрагованих геопросторових даних.....	143
4.4. Інші відкриті геопросторові ресурси	145
4.4.1. Wikimapia, GeoNames, Natural Earth.....	145
4.4.2. Супутникові зображення.....	148
4.4.3. Картографічні сервіси з відкритим API	149
4.5. Побудова мереж на основі геопросторових даних.....	151
4.5.1. Від об'єктів до вузлів – ідентифікація сутностей.....	151
4.5.2. Формування зв'язків між просторовими об'єктами	153
4.5.3. Інтеграція геопросторових і текстових даних.....	154

4.6. Візуалізація та аналіз геомереж.....	156
4.6.1. Графове представлення географічних об'єктів	157
4.6.2. Аналіз у Gephi: кластеризація, центральність, модулярність	159
4.6.3. Візуалізація геопросторових мереж засобами CSV2Graph.....	160
4.7. Практичне використання відкритих геопросторових даних у системах OSINT	161
4.7.1. Моніторинг критичної інфраструктури під час конфліктів	162
4.7.2. Геопросторова верифікація повідомлень відкритих джерел	163
4.7.3. Аналіз змін просторового середовища на основі відкритих геопросторових даних.....	165
4.8. Етичні та технічні виклики.....	166
4.8.1. Просторова неоднорідність якості та повноти відкритих геопросторових даних.....	167
4.8.2. Верифікація відкритих геопросторових даних та джерела помилок.....	168
4.8.3. Етичні та правові аспекти використання відкритих геопросторових даних	169
Висновки до розділу 4	170
РОЗДІЛ 5. СЕМАНТИЧНИЙ НЕТВОРКІНГ ЯК ОСНОВА ПОБУДОВИ МЕРЕЖ ЗНАНЬ	173
5.1. Теоретичні основи семантичних мереж.....	174
5.1.1. Еволюція концепції семантичних мереж	175
5.1.2. Граф як модель знань	176
5.1.3. Класифікація семантичних мереж.....	177
5.2. Моделі семантичного нетворкінгу	178
5.2.1. Асоціативні мережі зі зв'язками на основі контекстної близькості .	179
5.2.2. Причинно-наслідкові мережі: моделювання логіки подій.....	181
5.2.3. Онтологічні мережі: формалізація предметної області.....	182
5.2.4. Динамічні та адаптивні мережі.....	184
5.3. Підходи до автоматизованої побудови семантичних мереж	186
5.3.1. Екстрагування сутностей та зв'язків за допомогою великих мовних моделей	187

5.3.2. Промпт-інженерія для видобутку знань.....	188
5.3.3. Концепція «рою віртуальних експертів»	189
5.3.4. Агрегування, оцінювання достовірності та верифікація семантичних зв'язків.....	191
5.4. Формальна модель семантичного нетворкінгу.....	193
5.4.1. Формальне представлення семантичної мережі знань	194
5.4.2. Оцінка достовірності зв'язків	195
5.4.3. Моделі трансформації мереж знань	197
5.5. Інтеграція з іншими парадигмами.....	199
5.5.1. Семантичний веб (RDF, OWL, SPARQL) та його зв'язок із LLM-мережами	201
5.5.2. Поєднання семантичного нетворкінгу з машинним навчанням ..	203
5.5.3. Взаємодія з графовими базами даних та системами візуалізації.	204
5.6. Етапи життєвого циклу семантичної мережі.....	206
5.6.1. Ініціалізація: вибір джерел, визначення мети аналізу	207
5.6.2. Побудова: ітеративне формування через LLM	209
5.6.3. Верифікація: оцінювання достовірності та узгодженості семантичної мережі.....	209
5.6.4. Модифікація: реконструкція знань, генерація сценаріїв, прогнозування	211
5.6.5. Експлуатація: використання семантичної мережі в інформаційно-аналітичних системах	212
5.6.6. Узагальнена модель життєвого циклу семантичної мережі.....	213
5.7. Обмеження та виклики сучасного семантичного нетворкінгу	214
5.7.1. Епістемологічні обмеження великих мовних моделей: галюцинації, упередженість та актуальність знань	216
5.7.2. Обчислювальні виклики: масштабованість, продуктивність і динамічне оновлення семантичних мереж	217
5.7.3. Організаційні та методологічні аспекти: експертна верифікація, пояснюваність і контроль якості знань.....	219
Висновки до розділу 5.....	221

РОЗДІЛ 6. ВЕЛИКІ МОВНІ МОДЕЛІ ЯК ІНСТРУМЕНТ АВТОМАТИЗОВАНОГО АНАЛІЗУ OSINT	223
6.1. Роль LLM у сучасному OSINT-аналізі.....	224
6.1.1. Від ручного моніторингу до генеративного інтелекту	225
6.1.2. Переваги LLM: масштабованість, багатомовність, контекстне розуміння	227
6.1.3. Обмеження: галюцинації, чорна скринька, затримка актуальності.....	228
6.2. Основні задачі OSINT, що вирішуються за допомогою LLM	230
6.2.1. Екстрагування іменованих сутностей	230
6.2.2. Класифікація документів за тематикою та рівнем загрози	232
6.2.3. Виявлення дезінформації, фейків	233
6.2.4. Генерація аналітичних звітів, резюме, рекомендацій	235
6.2.5. Аналіз тональності та емоційного забарвлення інформаційних повідомлень	236
6.3. Методологія автоматизованого аналізу текстів через LLM	238
6.3.1. Формування інформаційного масиву	239
6.3.2. Структуровані промпти: шаблони для екстракції пар понять, цитат, подій.....	240
6.3.3. Стандартизація виводу для подальшої обробки	241
6.4. Концепція «рою віртуальних експертів» у OSINT	244
6.4.1. Принцип декомпозиції проблеми через множину ролей	244
6.4.2. Паралельне виконання багатоагентного аналізу	246
6.4.3. Агрегація та верифікація результатів	248
6.5. Побудова семантичних мереж на основі LLM-відповідей	249
6.5.1. Від пар понять до графів: формування вузлів і ребер	250
6.5.2. Причинно-наслідкові та асоціативні зв'язки без прив'язки до хронології	251
6.5.3. Інтеграція з Neo4j та Gephi: імпорт, візуалізація, аналіз центральності	252
6.5.4. Розширення семантичних мереж.....	254
6.5.5. Кейс 1: Аналіз кібератаки на основі LLM та семантичної мережі..	256

6.5.6. Кейс 2: Моніторинг інформаційної кампанії на основі LLM та каузальної семантичної мережі	257
6.5.7. Кейс 3. Юридична аналітика та виявлення недостовірних нормативних посилань	259
6.6. Інструменти та технології для LLM-аналізу OSINT	261
6.6.1. Платформи доступу: OpenAI API, Groq, Ollama, Hugging Face Inference	262
6.6.2. Локальні великі мовні моделі та їх використання в системах OSINT	263
6.6.3. Автоматизація через Python.....	265
6.7. Верифікація та оцінка якості LLM-результатів.....	266
6.7.1. Крос-перевірка через різні моделі та джерела	266
6.7.2. Метрики оцінювання достовірності семантичних зв'язків.....	267
6.7.3. Людський контроль: роль аналітика як ментора системи.....	269
Висновки до розділу 6	271
РОЗДІЛ 7. ГРАФОВІ БАЗИ ДАНИХ У OSINT: NEO4J, GEPHI, АНАЛІЗ ЗВ'ЯЗКІВ.....	273
7.1. Роль графових баз даних у сучасному OSINT.....	273
7.1.1. Від реляційних моделей до графового подання знань	274
7.1.2. Порівняння графових, реляційних та документних СКБД у контексті OSINT.....	276
7.2. Neo4j як основна платформа для зберігання семантичних мереж	280
7.2.1. Архітектура Neo4j: вузли, ребра, мітки, властивості	280
7.2.2. Інтеграція зовнішніх даних у Neo4j	284
7.3. Gephi як інструмент візуального аналізу мереж	286
7.3.1. Метрики центральності та виявлення спільнот у семантичних мережах.....	287
7.3.2. Експорт та інтеграція результатів мережевого аналізу	289
7.4. Підготовка даних для імпорту в графові системи	291
7.4.1. Формати подання графових даних для імпорту в Neo4j та Gephi..	291
7.4.2. Автоматизація підготовки графових даних засобами Python	294
7.5. Аналіз зв'язків у графах: ключові метрики та методи	296

7.5.1. Аналіз центральності вузлів у задачах OSINT.....	297
7.5.2. Аналіз спільнот та кластерів у семантичних мережах	299
7.5.3. Аналіз шляхів і зв'язності семантичних мереж.....	300
7.5.4. Аналіз часової еволюції семантичних мереж	301
7.6. Інтеграція LLM-екстрагованих даних у Neo4j	302
7.6.1. Інтеграція результатів SVE-аналізу до графової бази даних	303
7.6.2. Формування типізованих семантичних зв'язків	304
7.6.3. Онтологічне збагачення графової бази знань	306
7.7. Практичні кейси використання графових систем у OSINT	307
7.7.1. Аналіз кіберзагроз: від інструментів до акторів.....	308
7.7.2. Розслідування фінансових схем: від осіб до компаній через транзакції	309
7.7.3. Моніторинг інформаційних кампаній	310
7.8. Майбутнє графових технологій в OSINT.....	311
7.8.1. Гібридні інформаційно-аналітичні системи: графові моделі, векторні бази даних та великі мовні моделі	312
7.8.2. Автономне оновлення графів у реальному часі.....	313
7.8.3. Відповідальне використання графового аналізу та принципи White Hat AI.....	314
Висновки до розділу 7	316
РОЗДІЛ 8. ШТУЧНИЙ ІНТЕЛЕКТ У ПІДТРИМЦІ РІШЕНЬ: ВІД ВИЯВЛЕННЯ ЗАГРОЗ ДО ПРОГНОЗУВАННЯ ПОДІЙ	318
8.1. Роль ШІ в аналітичному циклі OSINT	319
8.1.1. Інтелектуальний конвеєр підтримки прийняття рішень	320
8.1.2. Розподіл когнітивних функцій між штучним інтелектом та експертом.....	322
8.1.3. Замкнений цикл когнітивного навчання систем підтримки прийняття рішень	323
8.2. Виявлення загроз на основі семантичних мереж	324
8.2.1. Виявлення структурних аномалій у семантичних мережах.....	325
8.2.2. Семантична класифікація загроз у графах знань.....	326
8.2.3. Інтегральне оцінювання рівня загрози в семантичній мережі	327

8.3. Методи раннього попередження	328
8.3.1. Виявлення слабких сигналів у потоках відкритої інформації	329
8.3.2. Виявлення інформаційних кампаній на основі аналізу семантичних наративів	330
8.3.3. Механізми автоматичного формування попереджень	332
8.4. Прогнозування подій через каузальні мережі	333
8.4.1. Семантична реконструкція причинно-наслідкових зв'язків	334
8.4.2. Семантичне моделювання альтернативних сценаріїв розвитку подій	335
8.4.3. Колективне оцінювання альтернативних сценаріїв методом рою віртуальних експертів	336
8.5. Підтримка стратегічного рішення	338
8.5.1. Автоматизована генерація аналітичних звітів із використанням великих мовних моделей	338
8.5.2. Візуалізація альтернатив	340
8.5.3. Рекомендації щодо контрзаходів	343
8.6. Інтеграція з інтелектуальними системами управління	346
8.6.1. Зв'язок із системами e-governance, цифровими двійниками	347
8.6.2. Програмні інтерфейси для передавання аналітичних результатів у режимі реального часу	349
8.6.3. Автоматизовані сценарії реагування	350
Висновки до розділу 8	352
РОЗДІЛ 9. ЕТИЧНІ, ПРАВОВІ ТА БЕЗПЕКОВІ АСПЕКТИ ВИКОРИСТАННЯ AI ТА OSINT	353
9.1. Етичні дилеми автоматизованої розвідки	354
9.1.1. Приватність і безпека: межі допустимого моніторингу	355
9.1.2. Розподіл відповідальності за помилкові висновки в системах AI та OSINT	357
9.1.3. Концепція White Hat AI: етика як основа проектування інтелектуальних систем	358
9.2. Правові рамки використання відкритих даних	361

9.2.1. GDPR, право на стирання та обробка персональних даних у системах OSINT	361
9.2.2. Правові аспекти автоматизованого збирання інформації із відкритих вебресурсів	363
9.2.3. Колізії у використанні відкритих даних: глобальний інформаційний простір і національне законодавство	365
9.3. Прозорість та підзвітність алгоритмів	367
9.3.1. Конфлікт FOIA (Freedom of Information Act) та «чорної скриньки» ІІІ	367
9.3.2. Пояснюваність автоматизованих висновків та прозорість виявлення загроз.....	369
9.3.3. Аудит семантичних мереж і відтворюваність інформаційно-аналітичного процесу.....	371
9.4. Юридичні ризики галюцинацій у офіційних документах.....	373
9.4.1. Юридичні галюцинації: вигадані нормативно-правові акти та помилкові посилання	373
9.4.2. Юридична відповідальність за використання недостовірних правових норм у документах	375
9.4.3. Механізми юридичної верифікації та інтеграція з офіційними правовими джерелами.....	377
9.5. Безпеківі ризики використання AI в OSINT	379
9.5.1. Ризики використання OSINT-технологій у системах Black Hat AI	379
9.5.2. Атаки на інтелектуальні інформаційно-аналітичні системи та порушення достовірності даних	381
9.5.3. Захист інформаційно-аналітичної інфраструктури інтелектуальних систем	383
9.6. Організаційні механізми контролю.....	385
9.6.1. Принцип «людина в центрі»: обов'язковий людський контроль над критичними рішеннями	385
9.6.2. Верифікація результатів великих мовних моделей	387
9.6.3. Управління та аудит інтелектуальних інформаційно-аналітичних систем	388

9.7. Відповідальне використання великих даних і штучного інтелекту в системах OSINT	390
Висновки до розділу 9	392
РОЗДІЛ 10. МАЙБУТНЄ OSINT	394
10.1. Від реактивного аналізу до прогностичної розвідки.....	395
10.2. Автономні аналітичні агенти.....	397
10.3. White Hat AI як стратегічна парадигма	400
10.4. Цифрові двійники в OSINT та управлінні	404
10.5. Інтеграція OSINT у системи e-governance	408
10.6. Технологічні тренди, що формують майбутнє OSINT	414
10.6.1. Мультимодальні моделі: інтегрований аналіз тексту, зображень, аудіо та відео	415
10.6.2. RAG та динамічна актуалізація знань	416
10.6.3. Гібридні графово-векторні архітектури семантичного пошуку	417
10.6.4. No-code і low-code як інтерфейс доступу до складної аналітики..	419
10.6.5. Поточкова аналітика, периферійні обчислення та подієво-орієнтовані системи	420
10.6.6. Локальні моделі та багатомодельна оркестрація.....	420
10.6.7. Походження даних, автентичність і довіра до цифрового контенту	421
10.6.8. Технології захисту приватності та контрольованого обміну даними.....	421
Висновки до розділу 10	423
ВИСНОВКИ ТА ПЕРСПЕКТИВИ ДОСЛІДЖЕНЬ	428
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	435

ВСТУП

Однією з визначальних особливостей сучасного етапу розвитку інформаційного суспільства є безперервне зростання обсягів цифрових даних, що генеруються людиною, інформаційними системами та різноманітними технічними пристроями. Щосекунди створюються значні масиви інформації у соціальних мережах, новинних ресурсах, мобільних додатках, сенсорних системах, відкритих державних реєстрах, корпоративних інформаційних системах та інших цифрових джерелах. Проте сам по собі великий обсяг даних не гарантує їхньої практичної цінності. Навпаки, швидке накопичення інформації нерідко супроводжується зростанням частки інформаційного шуму, дубльованого контенту, недостовірних повідомлень, дезінформації та матеріалів, автоматично згенерованих програмними засобами. У таких умовах першочергового значення набуває не стільки накопичення даних, скільки їх інтелектуальне опрацювання, інтерпретація та перетворення на знання, придатні для підтримки обґрунтованого прийняття рішень. Саме ця обставина зумовила появу й активний розвиток нових наукових і технологічних напрямів, серед яких особливе місце посідають обробка Big Data [1] (великих даних), OSINT [2] (Open Source Intelligence, розвідка у відкритих джерелах) та AI [3] (Artificial intelligence, штучний інтелект).

Ефективний аналіз сучасних інформаційних потоків ґрунтується на поєднанні трьох взаємодоповнювальних складових: технологій Big Data, методології OSINT та засобів штучного інтелекту. Кожна з них виконує власну функцію, проте лише їх комплексне застосування забезпечує можливість перетворення великих масивів розрізнених даних на знання, придатні для підтримки прийняття рішень.

Технології Big Data формують інфраструктурну основу такого аналізу. Вони забезпечують збирання, зберігання, індексацію та високопродуктивне опрацювання надвеликих обсягів структурованої й неструктурованої інформації, що надходить із численних джерел у режимі, наближеному до реального часу. Саме ці технології дозволяють працювати з інформаційними потоками, обсяг яких значно перевищує можливості традиційних систем керування базами даних. Для реалізації таких функцій широко застосовуються розподілені платформи та спеціалізовані системи оброблення даних [3], зокрема Hadoop, Elasticsearch, MongoDB, Apache Spark та інші.

Методологія OSINT визначає змістовний аспект аналітичного процесу. Вона охоплює принципи пошуку, відбору, оцінювання та інтеграції інформації з відкритих джерел, включаючи засоби масової інформації, соціальні мережі, офіційні реєстри, наукові публікації, спеціалізовані вебресурси та інші доступні інформаційні масиви. Важливими елементами OSINT є формування інформаційних потреб, побудова пошукових стратегій,

верифікація достовірності отриманих відомостей, виявлення взаємозв'язків між подіями та формулювання аналітичних гіпотез.

Штучний інтелект, і насамперед великі мовні моделі, виконує функцію інтелектуального опрацювання накопичених даних. На відміну від класичних інформаційно-пошукових систем, сучасні моделі здатні аналізувати семантичний зміст текстів, автоматично виділяти сутності та зв'язки між ними, будувати семантичні мережі, класифікувати документи, узагальнювати великі інформаційні масиви, генерувати пояснення, прогнозувати можливий розвиток подій і формувати рекомендації для особи, яка приймає рішення.

Таким чином, технології Big Data відповідають за масштабоване опрацювання інформації, методологія OSINT – за цілеспрямоване отримання та перевірку релевантних даних, а штучний інтелект – за їх інтерпретацію, узагальнення та трансформацію в нові знання. Саме синергія цих трьох складових визначає сучасний підхід до побудови інтелектуальних інформаційно-аналітичних систем.

Наведена триада на практиці дозволяє не лише виявляти в інформаційному просторі та реагувати на події, але й прогнозувати їх, створюючи основу для проактивного управління у сфері безпеки, політики, права та бізнесу.

Розвідка у відкритих джерелах сьогодні виходить далеко за межі традиційної розвідувальної діяльності. Вона стає універсальним інструментом глибокого розуміння складних систем – від кіберзагроз і соціальних настроїв до правових колізій і стратегічних ризиків. Основна мета OSINT – не просто зібрати дані, а виявити приховані зв'язки, реконструювати події, передбачити розвиток ситуації. Це особливо важливо в умовах, коли надходження інформації запізнюється, цензурується або навмисно спотворюється. Відкриті джерела, навпаки, пропонують живий, нефільтрований потік реальності, який, однак, вимагає складного інтелектуального опрацювання. Саме тому сучасна OSINT-аналітика переходить від пасивного моніторингу до активного семантичного моделювання, де кожен документ стає вузлом у мережі знань.

Класичні методи аналізу, від простого пошуку за ключовими словами до базових NER-систем (систем розпізнавання іменованих сутностей, Named Entity Recognition) [4], виявляються недостатніми перед обличчям сучасних викликів. Вони зосереджені на поверхневих ознаках: частоті слів, метаданих, формальних категоріях. Такі підходи не враховують глибоких зв'язків, причинно-наслідкових механізмів, контекстної багатозначності. Наприклад, дві події можуть бути пов'язані логічно, але не мати жодного хронологічного чи лексичного перетину, тобто традиційні системи їх не зв'яжуть.

Водночас навіть найсучасніші великі мовні моделі не позбавлені принципових обмежень. До них належать можливість генерування фактологічно недостовірних тверджень (так званих галюцинацій), обмежена

інтерпретованість процесу формування відповідей («ефект чорної скриньки»), залежність від обсягу контексту, що одночасно аналізується (контекстного вікна), а також обмежена актуальність знань, сформованих на основі навчальних даних. Унаслідок цього результати, отримані із застосуванням лише великих мовних моделей, без використання перевірених методів аналізу даних, процедур верифікації та предметно-орієнтованої методології, можуть виявитися неповними, недостатньо обґрунтованими або навіть помилковими. Саме тому сучасні інформаційно-аналітичні системи дедалі частіше будуються як комплексні рішення, у яких великі мовні моделі інтегруються з технологіями Big Data, методами OSINT, механізмами інтеграції з пошуковими системами (Retrieval-Augmented Generation, RAG) [5], базами знань, семантичними мережами та засобами пояснюваного штучного інтелекту (Explainable AI, XAI) [6].

Наведені особливості сучасного інформаційного середовища зумовлюють необхідність переходу до нової парадигми інформаційно-аналітичної діяльності. Її основу становить інтеграція технологій Big Data, методології OSINT та сучасних засобів штучного інтелекту, насамперед великих мовних моделей. Водночас ефективність такого підходу визначається не лише використанням окремих технологій, а й їх поєднанням у рамках єдиного керованого процесу, що забезпечує верифікацію інформації, семантичне збагачення даних, пояснюваність отриманих результатів та можливість їх практичного використання. Саме такий інтегрований інформаційно-аналітичний архітектурі присвячено цю книгу.

За останні десятиліття OSINT зазнав суттєвої еволюції. Якщо на початкових етапах його розвиток був пов'язаний переважно зі збиранням та систематизацією публічно доступних відомостей, то сьогодні OSINT сформувався як самостійний міждисциплінарний напрям інформаційної аналітики. Його методи застосовуються у сфері кібербезпеки, державного управління, правоохоронної діяльності, журналістських розслідувань, конкурентної розвідки, економічного аналізу, наукових досліджень та в багатьох інших галузях, де виникає потреба у своєчасному отриманні, перевірці та інтерпретації інформації.

Сучасний OSINT ґрунтується на комплексному аналізі великих обсягів відкритих даних, що надходять із різномірних інформаційних ресурсів. До таких джерел належать новинні вебресурси, соціальні мережі, блоги, форуми, відеохостинги, відкриті державні реєстри, законодавчі та судові бази даних, геопросторові інформаційні системи, наукові електронні бібліотеки, спеціалізовані технічні ресурси, бази даних кіберзагроз та численні інші інформаційні сервіси. Кожен із цих ресурсів характеризується власною структурою, форматом подання інформації, рівнем достовірності, періодичністю оновлення та специфікою предметної області.

У зв'язку з цим одним із центральних завдань сучасного інформаційного аналізу є інтеграція різномірних інформаційних потоків у єдиний

інформаційний простір. Така інтеграція передбачає узгодження текстових документів, метаданих, часових характеристик, геопросторової інформації, мережових атрибутів та інших типів даних, що дає змогу досліджувати не лише окремі повідомлення, а й закономірності, взаємозв'язки та тенденції розвитку складних інформаційних процесів.

Традиційні методи OSINT значною мірою спиралися на ручний пошук, експертний аналіз і безпосередню роботу аналітика з окремими джерелами інформації. Незважаючи на високу якість такого аналізу, його масштабованість залишалася обмеженою, а швидкість оброблення даних часто не відповідала темпам їх надходження. Розвиток технологій Big Data, машинного навчання та великих мовних моделей істотно змінив можливості інформаційно-аналітичних систем. Сучасні рішення забезпечують автоматизоване збирання, інтеграцію, класифікацію, семантичний аналіз, узагальнення та інтерпретацію інформації, що відкриває можливість переходу від аналізу окремих документів до комплексного моделювання інформаційного середовища, реконструкції взаємозв'язків між подіями та підтримки прийняття обґрунтованих рішень.

Застосування великих мовних моделей істотно розширило можливості інформаційно-аналітичних систем. Вони забезпечують не лише автоматизоване виділення фактів із текстових документів, а й формування структурованих моделей предметної області у вигляді семантичних мереж. У таких мережах вузли відповідають окремим сутностям (особам, організаціям, об'єктам, подіям, процесам тощо), а ребра відображають різноманітні типи взаємозв'язків між ними, зокрема причинно-наслідкові, функціональні, просторові, часові, ієрархічні або асоціативні. Подання знань у такому вигляді дає змогу перейти від аналізу ізольованих повідомлень до дослідження структури предметної області та закономірностей її розвитку.

Важливою перевагою семантичного моделювання є можливість реконструкції подій навіть за умов неповноти або фрагментарності вихідної інформації. Якщо окремі часові характеристики відсутні або є невизначеними, взаємозв'язки між сутностями можуть бути встановлені на основі логічних, причинно-наслідкових чи асоціативних залежностей. Наприклад, навіть за відсутності точної інформації про час здійснення кібератаки можна встановити зв'язок між суб'єктом атаки, використаним програмним інструментом та об'єктом впливу. Такі структуровані залежності становлять самостійну цінність і можуть бути використані під час аналізу ситуації, побудови сценаріїв розвитку подій та підтримки процесів прийняття рішень.

Таким чином, розвиток методів штучного інтелекту забезпечив перехід від накопичення та пошуку інформації до її семантичного моделювання, у межах якого основним об'єктом аналізу стають не окремі документи, а система взаємопов'язаних фактів, сутностей і відношень між ними.

У цьому контексті інформаційно-аналітичні системи, що використовують методологію OSINT, орієнтовані на розв'язання трьох взаємопов'язаних класів задач.

Перший клас становлять задачі верифікації інформації, спрямовані на оцінювання достовірності повідомлень, виявлення суперечностей та зменшення впливу дезінформації. Одним із перспективних підходів є застосування методу рою віртуальних експертів (Swarm of Virtual Experts, SVE), розглянутого в цій книзі. Його основна ідея полягає у незалежному аналізі однієї й тієї самої інформації кількома спеціалізованими агентами, що моделюють різні експертні ролі, після чого результати інтегруються для формування більш обґрунтованих висновків і зниження ризику помилок, характерних для окремих великих мовних моделей.

Другий клас задач пов'язаний із прогнозуванням можливого розвитку подій. Семантичні мережі дозволяють моделювати потенційні сценарії шляхом аналізу існуючих взаємозв'язків між сутностями, оцінювання можливих наслідків зміни окремих факторів та виявлення альтернативних траєкторій розвитку ситуації. Такий підхід є особливо важливим у задачах оцінювання кіберзагроз, інформаційної безпеки та підтримки прийняття управлінських рішень.

Третій клас охоплює задачі виявлення прихованих закономірностей і взаємозв'язків, які залишаються непомітними під час аналізу окремих документів. Поєднання інформації з різномірних джерел дає можливість встановлювати опосередковані зв'язки між особами, організаціями, подіями, фінансовими операціями або кіберінцидентами, що значно підвищує ефективність комплексного аналізу.

Поєднання зазначених функцій формує новий підхід до інформаційної аналітики, орієнтований не лише на опис уже відомих фактів, а й на їх комплексне осмислення, виявлення прихованих закономірностей та прогнозування подальшого розвитку досліджуваних процесів.

Реалізація такого підходу неможлива без використання технологій Big Data, які забезпечують технологічну основу функціонування сучасних інформаційно-аналітичних систем. Якщо методологія OSINT визначає джерела, принципи пошуку та верифікації інформації, а штучний інтелект здійснює її інтелектуальне опрацювання, то технології Big Data забезпечують масштабове збирання, зберігання, індексацію, інтеграцію та швидке оброблення інформаційних потоків.

Відкриті інформаційні ресурси характеризуються не лише надзвичайно великими обсягами даних, а й їх різномірністю, високою швидкістю надходження, неповнотою, неоднорідною структурою та різним рівнем достовірності. За таких умов традиційні централізовані реляційні бази даних часто виявляються недостатньо ефективними. Саме тому сучасні інформаційно-аналітичні системи базуються на розподілених масштабованих

архітектурах, які забезпечують не лише довготривале зберігання даних, а й їх швидку індексацію, повнотекстовий і семантичний пошук, інтеграцію різнорідних джерел, побудову аналітичних моделей та візуалізацію результатів у режимі, наближеному до реального часу.

У контексті OSINT поняття Big Data набуває специфічного змісту, що виходить за межі класичних трьох «V» (Volume – обсяги, Velocity – швидкість, Variety – різноманітність), доповнюючись критичними вимірами Veracity (достовірність) та Volatility (мінливість). Саме ці два параметри стають визначальними для розвідувальної практики: величезний обсяг даних з соціальних мереж, новинних агрегаторів чи геопросторових платформ часто містить дезінформацію, маніпулятивні нарративи або навмисно сфабриковані артефакти, що робить верифікацію не технічною процедурою, а стратегічним елементом аналізу.

Однією з характерних особливостей сучасного інформаційного середовища є надзвичайно висока швидкість появи, оновлення та зникнення цифрового контенту. Значна частина повідомлень постійно редагується, доповнюється, видаляється або замінюється новими версіями, унаслідок чого інформаційний простір набуває вираженої часової динаміки. За таких умов достовірність і зміст інформації дедалі більше залежать від моменту її фіксації та контексту, у якому вона була отримана.

У зв'язку з цим технології Big Data вже не можна розглядати виключно як засіб зберігання та високопродуктивного опрацювання великих обсягів даних. Не менш важливими стають задачі оцінювання достовірності інформації, виявлення маніпулятивного контенту, аналізу часової еволюції інформаційних потоків, реконструкції причинно-наслідкових залежностей між подіями та виявлення прихованих взаємозв'язків між інформаційними об'єктами. Особливого значення набуває аналіз семантичних структур, динаміки поширення інформаційних повідомлень, трансформації нарративів, мемів та інших інформаційних конструкцій, що визначають розвиток суспільних інформаційних процесів.

Таким чином, у системах OSINT технології Big Data виконують значно ширшу функцію, ніж забезпечення масштабованого зберігання та оброблення інформації. Вони створюють технологічну основу для безперервного моніторингу інформаційного простору, інтеграції різнорідних джерел, оцінювання якості даних та формування інформаційної бази, придатної для подальшого семантичного аналізу й підтримки прийняття рішень. За таких умов визначальним фактором стає не стільки обсяг накопиченої інформації, скільки її релевантність, достовірність, повнота та можливість перевірки.

Наприклад, окремий інформаційний ресурс або канал у соціальних мережах може щоденно генерувати сотні чи навіть тисячі повідомлень, проте лише незначна їх частина містить інформацію, що має аналітичну цінність. Завдання інфраструктури Big Data полягає не лише у накопиченні таких

потоків, а й у їх автоматизованому індексуванні, семантичному аналізі, виявленні релевантних повідомлень, інтеграції з іншими джерелами та формуванні структурованого інформаційного представлення предметної області.

Робота з великими обсягами різномірної інформації потребує використання спеціалізованих програмно-технологічних платформ, здатних забезпечити масштабоване збирання, зберігання, індексацію, пошук та аналіз даних. На відміну від традиційних реляційних баз даних, сучасні інформаційно-аналітичні системи дедалі частіше базуються на гнучких архітектурах, орієнтованих на ефективне опрацювання структурованих, напівструктурованих і неструктурованих даних. До таких технологій належать документоорієнтовані та інші NoSQL-системи, розподілені пошукові платформи, графові бази даних, засоби потокового оброблення інформації та масштабовані програмні комплекси для розподілених обчислень. У наступних розділах книги буде детально розглянуто принципи побудови та практичне застосування таких систем, як Elasticsearch, MongoDB, Neo4j, Apache Spark, Hadoop та інших компонентів сучасної екосистеми Big Data.

Практичні інформаційно-аналітичні системи, як правило, поєднують декілька взаємодоповнювальних технологій, кожна з яких виконує власну функцію. Одні компоненти забезпечують автоматизоване збирання інформації з відкритих джерел, інші – її довготривале зберігання, повнотекстову та семантичну індексацію, інтеграцію різномірних даних, побудову графових моделей або підтримку аналітичної обробки та візуалізації. Саме така багатокomпонентна архітектура забезпечує ефективну роботу з інформаційними потоками практично необмеженого обсягу та складності.

Однією з найважливіших переваг технологій Big Data є можливість інтеграції різномірних інформаційних ресурсів у єдиному інформаційному просторі. У процесі аналізу можуть одночасно використовуватися текстові документи, повідомлення із засобів масової інформації та соціальних мереж, метадані, часові характеристики, геопросторова інформація, мережеві атрибути, мультимедійний контент, дані відкритих державних реєстрів, інформація про кіберзагрози, результати роботи великих мовних моделей та інші інформаційні об'єкти. Їх комплексне опрацювання створює передумови для побудови інтегрованих інформаційних моделей, встановлення взаємозв'язків між об'єктами різної природи та подальшого застосування методів семантичного аналізу й штучного інтелекту.

Далі увагу буде зосереджено не лише на теоретичних засадах технологій Big Data, а й на їх практичному використанні в системах OSINT та інтелектуальної аналітики. На прикладі сучасних програмних платформ буде показано підходи до побудови масштабованих інформаційно-аналітичних систем, інтеграції різномірних джерел даних, створення семантичних мереж,

застосування великих мовних моделей та підтримки процесів прийняття рішень.

Одним із фундаментальних принципів технологій Big Data є інтеграція різномірних інформаційних ресурсів у межах єдиного інформаційного простору. Сучасні інформаційно-аналітичні платформи забезпечують не лише накопичення окремих документів, а й формування комплексних інформаційних об'єктів, які об'єднують усі доступні відомості про певну подію, об'єкт або процес незалежно від їхнього походження та формату.

Такий інтегрований інформаційний об'єкт може містити первинний текст документа, його метадані, результати автоматичного виділення сутностей і взаємозв'язків, часові характеристики, геопросторову інформацію, мультимедійні матеріали, посилання на зовнішні інформаційні ресурси, структуровані записи з відкритих баз даних, а також результати подальшого семантичного аналізу. Усі ці компоненти утворюють єдиний багатовимірний контекст, який значно перевищує інформаційну цінність окремого документа.

Саме інтегроване подання даних створює основу для застосування сучасних методів штучного інтелекту. Великі мовні моделі отримують можливість аналізувати не ізольовані тексти, а комплекс взаємопов'язаних відомостей, встановлювати приховані закономірності, будувати семантичні мережі, реконструювати причинно-наслідкові зв'язки та формувати обґрунтовані аналітичні висновки. Таким чином, технології Big Data забезпечують не лише масштабове зберігання інформації, а й створюють інформаційну основу для її інтелектуального опрацювання.

У наступних розділах книги буде показано, як такі інтегровані інформаційні моделі реалізуються на практиці із застосуванням сучасних платформ зберігання, індексації та семантичного аналізу даних, а також великих мовних моделей і графових структур знань.

Таким чином, великі дані виступають не як самодостатній інструмент, а як невидимий каркас, що зв'язує джерела, аналітику та візуалізацію в єдиний цілісний процес – від шуму до знань.

Якщо Big Data забезпечує інфраструктурну основу, а OSINT – аналітичну цільову функцію, то штучний інтелект, зокрема LLM [7], виступає як інтелектуальний трансформатор, що перетворює неструктуровані масиви текстів на структуровані знання. У контексті OSINT ШІ більше не є лише допоміжним інструментом – він стає центральним механізмом семантичного осмислення. Однак ефективне застосування LLM вимагає не просто запуску промпта, а глибокого розуміння їхніх можливостей, обмежень і архітектурних особливостей. Лише при наявності методологічного каркасу – такого як «рій віртуальних експертів» та причинно-наслідкове моделювання ШІ стає надійним партнером у аналітичному процесі.

Сучасні LLM (ChatGPT, Gemini, DeepSeek тощо) демонструють деяку здатність до контекстного розуміння, багатомовності, адаптації до домену (предметної області) та генерації структурованого виводу [8]. Вони можуть видобувати сутності, класифікувати тексти, пояснювати зв'язки, при чому робити це в масштабі, недосяжному для людини.

Проте фахівець має чітко усвідомлювати чотири ключові обмеження [9]:

- галюцинації – генерація правдоподібної, але фактично хибної інформації. Це особливо небезпечно в юридичній, кібербезпечній або медійній аналітиці, де помилка може мати серйозні наслідки;
- чорна скринька – відсутність прозорості у механізмах прийняття рішень. Навіть при однаковому промпті різні запуски можуть давати різні результати, що ускладнює верифікацію;
- лінь (laziness) – схильність моделі давати поверхневі, узагальнені відповіді замість глибокого аналізу, особливо коли промпт не містить чітких інструкцій;
- застарілість даних (knowledge cutoff) – обмежена актуальність інформації через фіксовану дату навчання моделі. LLM не «бачать» події, що відбулися після цієї дати, і не мають доступу до реального часу без інтеграції з пошуковими інструментами. Для OSINT це критично: аналіз подій минулого тижня, відстеження поточних інформаційних операцій або верифікація свіжих фейків вимагають оперативних даних, яких у «голій» моделі немає. Навіть найпотужніший LLM, навчений на даних до 2024 року, буде безпорадним перед подіями 2025–2026 років, що робить його допоміжним інструментом, а не джерелом первинної інформації.

Ці обмеження вимагають системного підходу до верифікації: використання декількох моделей, крос-перевірка через зовнішні джерела, агрегація через «рій віртуальних експертів».

Традиційні NER-системи (spaCy [10], Stanza [11]) обмежені фіксованими класами сутностей (PERSON, ORG, LOC). LLM дозволяють переходити до семантичного рівня: видобування пар понять у форматі (суб'єкт, предикат, об'єкт), наприклад: «група APT28», «використала», «інструмент Zloader») («законопроект № X», «порушує», «статтю Y Конституції»).

Такий підхід дозволяє будувати семантичні мережі без попередньої онтології, тобто модель сама виявляє типи зв'язків. Ключовим є точне формулювання промпта, який визначає:

- які типи сутностей шукати;
- які типи зв'язків вважати релевантними;
- у якому форматі подавати результат (CSV, JSON, GEXF).

Для забезпечення точності промпт, наприклад, може включати інструкцію: «Видобий усі трійки у форматі [суб'єкт | дія | об'єкт]. Якщо немає зв'язку – поверни порожній список. Не вигадуй.»

Однією з важливих тенденцій розвитку сучасних систем штучного інтелекту є перехід від використання окремих промптів до побудови формалізованих інформаційно-аналітичних процесів. Для цього застосовуються спеціалізовані фреймворки, які забезпечують опис послідовності дій мовної моделі у вигляді стандартизованих логічних конструкцій. Такий підхід наближає взаємодію з великими мовними моделями до принципів програмування, дозволяючи формалізувати складні сценарії аналізу без необхідності розроблення традиційного програмного коду.

Основою таких фреймворків є представлення аналітичного процесу як послідовності взаємопов'язаних операцій, що включають перевірку умов, повторювану обробку даних, виклик окремих функціональних модулів, розгалуження логіки виконання та агрегацію отриманих результатів. Формальний опис подібних процесів забезпечує їх відтворюваність, повторне використання, масштабованість та можливість незалежної перевірки окремих етапів аналізу. Математичні засади такого підходу, а також запропонований авторський фреймворк будуть детально розглянуті в наступних розділах книги.

Важливою перевагою формалізації промптів є можливість їх інтеграції у загальний інформаційно-аналітичний конвеєр. Результати роботи великих мовних моделей можуть автоматично перетворюватися на структуровані подання даних, придатні для подальшого аналізу, накопичення у сховищах даних, побудови семантичних мереж, графів знань, візуалізації та використання в системах підтримки прийняття рішень. Це дозволяє розглядати великі мовні моделі не як ізольований інструмент генерації тексту, а як один із компонентів комплексної інформаційно-аналітичної архітектури.

Особливе місце в такій архітектурі посідає багатоагентний підхід, за якого один і той самий інформаційний об'єкт аналізується кількома спеціалізованими агентами або віртуальними експертами, кожен із яких виконує окрему аналітичну функцію відповідно до своєї предметної спеціалізації. Подальша інтеграція отриманих результатів забезпечує більш повне представлення досліджуваної предметної області, підвищує достовірність висновків, сприяє виявленню альтернативних інтерпретацій та створює умови для їх подальшої експертної верифікації.

Таким чином, стандартизація промптів і формалізація процесів взаємодії з великими мовними моделями перетворюють їх із засобу виконання окремих запитів на компонент відтворюваних, масштабованих і контрольованих інформаційно-аналітичних технологій. Саме такий підхід є одним із методологічних принципів побудови сучасних систем аналізу відкритих джерел інформації, що поєднують технології Big Data, OSINT та штучного інтелекту.

У системах OSINT основним об'єктом дослідження стають не лише окремі факти, а й взаємозв'язки між ними, які формують цілісне уявлення про події, процеси, об'єкти та їх взаємодію. Однією з найбільш придатних моделей такого подання є семантична мережа – графова структура, у якій вузли відповідають сутностям предметної області, а ребра відображають різноманітні типи семантичних відношень між ними. Подання знань у вигляді графів забезпечує можливість одночасного аналізу структури, контексту та взаємозалежностей інформаційних об'єктів, що робить семантичні мережі важливим інструментом сучасної інформаційної аналітики.

Класичні методи інформаційного пошуку [12] та аналізу текстів переважно ґрунтуються на лексичних характеристиках документів, таких як ключові слова, n-грами, статистичні ваги термінів або їх частоти. Незважаючи на високу ефективність у задачах пошуку, такі підходи обмежено відображають змістові зв'язки між поняттями та не дозволяють безпосередньо аналізувати структуру знань. Розвиток великих мовних моделей відкрив можливість переходу до семантичного індексування, за якого з текстів автоматично виділяються сутності, їхні властивості та взаємозв'язки, що надалі формують графи знань або семантичні мережі. Це створює передумови для побудови моделей предметної області без необхідності попереднього жорсткого визначення їх структури та забезпечує можливість масштабованого опрацювання великих інформаційних масивів.

Особливістю семантичного моделювання є те, що взаємозв'язки між об'єктами визначаються насамперед їхнім логічним змістом, а не лише часовою послідовністю виникнення подій. У багатьох прикладних задачах причинно-наслідкові, функціональні або асоціативні залежності виявляються більш інформативними, ніж їх хронологічне розташування. Це дає змогу реконструювати структуру досліджуваних процесів навіть за умов неповноти, фрагментарності або навмисного спотворення часових характеристик. Семантична мережа у такому випадку виступає формальною моделлю предметної області, що відображає систему взаємозв'язків між її елементами незалежно від повноти часової інформації.

Практичне використання семантичних мереж передбачає застосування спеціалізованих програмних засобів, призначених для їх зберігання, аналізу та візуалізації. Для побудови масштабованих графових сховищ широко використовуються графові системи керування базами даних, зокрема Neo4j [13], які забезпечують ефективне зберігання графових структур, виконання складних запитів і підтримку аналітичних операцій. Візуальний аналіз семантичних мереж реалізується за допомогою таких інструментів, як Gephi [14], що надають широкий набір алгоритмів компонування графів, кластеризації, ранжування вузлів та дослідження їх структурних характеристик. Для автоматизації побудови інтерактивних графових моделей застосовуються також спеціалізовані програмні засоби, наприклад

CSV2Graph [15], які забезпечують перетворення структурованих даних у графове подання та інтеграцію його з інформаційно-пошуковими системами. Принципи побудови таких систем, їх архітектурні особливості та практичне застосування будуть детально розглянуті в наступних розділах книги.

Сучасні програмні засоби аналізу семантичних мереж формують єдине інформаційно-аналітичне середовище, у якому знання не лише накопичуються, а й стають доступними для структурованого аналізу, візуалізації, верифікації та подальшого використання в процесах підтримки прийняття рішень. Поєднання графових моделей, засобів семантичного аналізу та сучасних платформ оброблення даних забезпечує перехід від аналізу окремих документів до дослідження складних систем взаємопов'язаних інформаційних об'єктів.

Водночас широке використання великих мовних моделей в інформаційно-аналітичних системах супроводжується низкою методологічних викликів, пов'язаних із можливістю виникнення фактологічних помилок, неоднозначністю інтерпретацій, обмеженою пояснюваністю отриманих результатів та необхідністю їх незалежної перевірки. Одним із перспективних напрямів подолання цих обмежень є багатоагентний підхід, відповідно до якого складне аналітичне завдання декомпонується на низку взаємопов'язаних підзадач, що виконуються спеціалізованими віртуальними експертами. Така організація аналізу забезпечує можливість зіставлення альтернативних інтерпретацій, підвищення достовірності отриманих результатів та їх подальшої експертної верифікації. Запропонована в книзі методологія рою віртуальних експертів (Swarm of Virtual Experts, SVE) [16] детально розглядається у відповідних розділах як один із механізмів побудови відтворюваних інформаційно-аналітичних процесів.

Не менш важливим напрямом розвитку сучасних інформаційно-аналітичних систем є інтеграція великих мовних моделей із зовнішніми джерелами знань. Архітектура Retrieval-Augmented Generation поєднує генеративні можливості штучного інтелекту з автоматизованим пошуком актуальної інформації, що забезпечує використання під час аналізу не лише знань, сформованих у процесі навчання моделі, а й поточних відомостей із відкритих інформаційних ресурсів. Це особливо важливо для задач OSINT, у яких достовірність і актуальність інформації безпосередньо впливають на якість аналітичних висновків.

Матеріал монографії побудовано за принципом послідовного переходу від концептуальних засад до практичної реалізації сучасних інформаційно-аналітичних технологій. У ній розглядаються теоретичні основи технологій Big Data, методології OSINT, семантичного моделювання, великих мовних моделей, багатоагентних систем, архітектури RAG, графів знань і засобів підтримки прийняття рішень. Окрему увагу приділено програмно-технологічним аспектам реалізації таких систем, методам інтеграції

різнорідних джерел інформації, побудові семантичних мереж, графовому аналізу та застосуванню сучасних програмних платформ, зокрема Elasticsearch, MongoDB, Neo4j, Gephi та інших інструментів екосистеми Big Data.

Усі розділи книги підпорядковано єдиній методологічній логіці, що відображає повний цикл сучасної інформаційно-аналітичної діяльності: від збирання та інтеграції даних до побудови моделей знань, їх аналізу, прогнозування та формування обґрунтованих висновків. Саме така інтеграція технологій Big Data, методології OSINT і засобів штучного інтелекту становить концептуальну основу викладеного в монографії підходу.

Викладений у книзі матеріал побудовано за принципом послідовного переходу від загальних теоретичних положень до методів, алгоритмів, програмно-технологічних рішень і прикладів їх практичного застосування. Такий підхід забезпечує цілісне розуміння сучасних технологій інформаційної аналітики, сприяє відтворюваності запропонованих методів і створює передумови для їх адаптації до різних предметних областей, у яких виникає потреба в аналізі великих масивів відкритих даних.

У перших розділах розглядаються теоретичні засади технологій Big Data, методології OSINT та сучасного штучного інтелекту. Наступні розділи присвячено методам семантичного аналізу, побудові графів знань і застосуванню великих мовних моделей для автоматизованого видобування та верифікації знань. Значну увагу приділено програмно-технологічним аспектам реалізації інформаційно-аналітичних систем, інтеграції різнорідних джерел даних, використанню графових баз даних, засобів візуалізації та сучасних платформ оброблення великих даних. Завершальні розділи містять приклади практичного застосування запропонованих підходів, розглядають питання підтримки прийняття рішень, а також перспективи розвитку інтелектуальних інформаційно-аналітичних систем.

Книга адресована широкому колу фахівців, професійна діяльність яких пов'язана зі збиранням, інтеграцією, аналізом та інтерпретацією інформації з відкритих джерел. Вона може бути корисною аналітикам у сфері кібербезпеки, інформаційної та конкурентної розвідки, журналістських розслідувань, фахівцям із розроблення систем штучного інтелекту та аналізу даних, дослідникам у галузі інформаційних технологій, а також спеціалістам із права, державного управління, міжнародних відносин і політичної аналітики, які застосовують сучасні методи опрацювання інформації для дослідження складних соціотехнічних процесів.

Метою книги є систематизація теоретичних засад, методів і технологій побудови сучасних інформаційно-аналітичних систем, що інтегрують можливості Big Data, методології OSINT, семантичного моделювання та штучного інтелекту. Основну увагу приділено методам інтеграції різнорідних джерел інформації, побудови семантичних моделей і графів знань,

застосуванню великих мовних моделей, а також використанню сучасних інформаційних технологій для підтримки аналізу, прогнозування та прийняття рішень.

Розглянуті у книзі концепції, моделі та програмно-технологічні рішення можуть бути використані як методологічна основа для розроблення інтелектуальних інформаційно-аналітичних систем, проведення наукових досліджень, підготовки фахівців у галузі аналізу відкритих джерел інформації та впровадження сучасних засобів семантичного аналізу й штучного інтелекту в практику інформаційно-аналітичної діяльності. Таким чином, структура книги відображає послідовний перехід від фундаментальних теоретичних положень до методів, алгоритмів, програмно-технологічних рішень і практичних застосувань сучасних інформаційно-аналітичних систем, побудованих на інтеграції технологій Big Data, методології OSINT та штучного інтелекту.