Міністерство освіти і науки України
Київський столичний університет імені Бориса Грінченка
Факультет інформаційних технологій та математики
Кафедра інформаційної та кібернетичної безпеки
імені професора Володимира Бурячка

«Допущено до захисту»
Завідувач кафедри інформаційної та
кібернетичної безпеки імені
професора Володимира Бурячка
кандидат технічних наук, доцент
Складанний П.М.

_____
(підпис)
«____»_____ 20__ р.

КВАЛІФІКАЦІЙНА РОБОТА
на здобуття другого (магістерського)
рівня вищої освіти

Спеціальність 125 Кібербезпека та захист інформації

Тема роботи:
КОГНІТИВНЕ МОДЕЛЮВАННЯ СЦЕНАРІЇВ В УПРАВЛІННІ РИЗИКАМИ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Виконала

студентка групи БІКСм-1-24-1.4д

Гаркушенко Аріна Максимівна

_____
(підпис)

Науковий керівник

Кандидат педагогічних наук, доцент

Шевченко С.М.

_____
(підпис)

Київ – 2025

Міністерство освіти і науки України
Київський столичний університет імені Бориса Грінченка
Факультет інформаційних технологій та математики
Кафедра інформаційної та кібернетичної безпеки
імені професора Володимира Бурячка

Освітньо-кваліфікаційний рівень – магістр
Спеціальність 125 Кібербезпека та захист інформації
Освітня програма 125.00.01 Безпека інформаційних та комунікаційних систем

«Затверджую»
Завідувач кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка
кандидат технічних наук, доцент
Складанний П.М.

_____
(підпис)
«____»_____ 20__ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТЦІ
Гаркушенко Аріні Максимівні

1. Тема роботи: «Когнітивне моделювання сценаріїв в управлінні ризиками інформаційної безпеки»
   Керівник: Шевченко Світлана Миколаївна, кандидат педагогічних наук, доцент
   Затверджено наказом ректора від _____
2. Термін подання роботи студентом: _____
3. Вихідні дані до роботи:
   3.1 науково-технічна та нормативна література з теми дослідження;
   3.2 методи: системно-структурний, порівняльний аналіз, методи когнітивного моделювання
   3.3 математичні моделі та методи: нечіткі когнітивні карти, сценарний підхід
4. Зміст текстової частини роботи (перелік питань, які потрібно розробити):
   4.1 Проаналізувати існуючі підходи до когнітивного моделювання, управління кіберризиками за допомогою сценарного аналізу
   4.2 Дослідити особливості застосування когнітивних карт, Байєсівських мереж і агентно-орієнтованих моделей у сфері кібербезпеки
   4.3 Розробити модель побудови когнітивних сценаріїв кіберінцидентів.
5. Перелік графічного матеріалу:
   5.1 Презентація доповіді, виконана в Microsoft PowerPoint.
6. Дата видачі завдання: _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів підготовки роботи	Термін виконання	Примітка
1.	Уточнення постановки завдання	14.11.2024	Виконано
2.	Аналіз літератури	20.12.2024	Виконано
3.	Обґрунтування вибору рішення	15.01.2025	Виконано
4.	Збір даних	25.02.2025	Виконано
5.	Виконання та оформлення розділу 1.	20.05.2025	Виконано
6.	Виконання та оформлення розділу 2.	01.09.2025	Виконано
7.	Виконання та оформлення розділу 3.	13.10.2025	Виконано
8.	Виконання та оформлення розділу 4.	20.10.2025	Виконано
9.	Вступ, висновки, реферат	24.10.2025	Виконано
10.	Апробація роботи на науково-методичному семінарі та/або науково-технічній конференції	26.10.2025	Виконано
11.	Оформлення та друк тестової частини роботи	10.11.2025	Виконано
12.	Оформлення презентацій	12.11.2025	Виконано
13.	Отримання рецензій	14.11.2025	Виконано
14.	Попередній захист роботи	17.11.2025	Виконано
15.	Захист в ЕК	17.12.2025	Виконано

Студентка _____ Гаркушенко Аріна Максимівна
(підпис)

Науковий керівник _____ Шевченко Світлана Миколаївна
(підпис)

РЕФЕРАТ

Кваліфікаційна робота присвячена технологіям використання когнітивного моделювання в системах управління ризиками кібербезпеки.

Робота складається зі вступу, чотирьох розділів, що містять 12 рисунків та 5 таблиць, висновків та списку використаних джерел, що містить 52 найменування. Загальний обсяг роботи становить 115 сторінок, з яких 7 сторінок займають перелік умовних скорочень та список використаних джерел.

Об'єктом дослідження в роботі є процес управління ризиками інформаційної безпеки в розподілених інформаційних системах.

Предметом дослідження є методи когнітивного моделювання сценаріїв.

Метою роботи є підвищення ефективності управління кіберризиками.

Для досягнення поставленої мети у роботі:

- проведено аналіз існуючих підходів до когнітивного моделювання, управління кіберризиками за допомогою сценарного аналізу;
- досліджено особливості застосування когнітивних карт, Байєсівських мереж і агентно-орієнтованих моделей у сфері кібербезпеки;
- обґрунтовано і розроблено модель побудови когнітивних сценаріїв кіберінцидентів.

Наукова новизна одержаних результатів полягає в тому, що запропоновано вдосконалений підхід до побудови когнітивних сценаріїв для управління ризиками, розроблено метод побудови когнітивних сценаріїв, що дозволяє моделювати причинно-наслідкові ланцюги для прийняття рішень.

Галузь застосування. Запропоновані підходи можуть бути використані для створення інтегрованих систем підтримки та прийняття рішень.

Ключові слова: КОГНІТИВНЕ МОДЕЛЮВАННЯ, КОГНІТИВНІ КАРТИ, СЦЕНАРНИЙ АНАЛІЗ, БАЙЄСІВСЬКІ МЕРЕЖІ, ШТУЧНИЙ ІНТЕЛЕКТ, ІНФОРМАЦІЙНА БЕЗПЕКА.

ЗМІСТ

СПИСОК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

КМ – когнітивне моделювання

КК – когнітивна карта

НКК – нечітка когнітивна карта

БМ – байєсівська мережа

АОМ – агентно-орієнтоване моделювання

ШІ – штучний інтелект

ML – Machine Learning – машинне навчання

SOC – Security Operations Center – центр операційної безпеки

CERT – Computer Emergency Response Team – команда реагування на надзвичайні ситуації

DL – Deep Learning – глибоке навчання

ВСТУП

Актуальність теми. Зі зростанням масштабів цифровізації та ускладнення кіберзагроз традиційні підходи до управління ризиками інформаційної безпеки демонструють недостатню ефективність. Сучасні атаки характеризуються динамічністю та адаптивністю, що робить їх складними для своєчасного виявлення, що призводить до того, що значна частина інцидентів залишається непоміченою доти, доки не спричиняє операційних втрат, порушення конфіденційності або повної зупинки бізнес-процесів. Це створює прогалини у прогнозуванні розвитку подій та знижує ефективність реагування. У цих умовах особливої актуальності набувають моделі, здатні враховувати причинно-наслідкові зв'язки, поведінкові закономірності, помилки користувачів, складність атаки та інші когнітивні і технічні фактори. Саме тому когнітивне моделювання та сценарний підхід сьогодні розглядаються як перспективні інструменти для підвищення якості прогнозування та прийняття рішень у сфері кібербезпеки. Використання нечітких когнітивних карт, агентно-орієнтованих моделей, байєсівських мереж та штучного інтелекту дозволяє проводити симуляції кіберінцидентів, прогнозувати можливу поведінку внутрішніх користувачів та атакуючих, а також оцінювати ефективність заходів захисту в різних умовах, що є особливо важливим там, де оперативність прийняття рішень та точність оцінки ризиків визначають успішність реагування на інциденти та мінімізацію наслідків атак.

Таким чином, актуальність роботи зумовлена необхідністю підвищення ефективності управління кіберризиками шляхом інтеграції когнітивного моделювання в процеси оцінки та прогнозування загроз, що дозволяє не лише виявляти уразливості, а й проактивно запобігати потенційним атакам.

Мета роботи полягає у підвищенні ефективності управління ризиками кібербезпеки шляхом розробки, застосування та аналізу когнітивних моделей і сценарного підходу для моделювання взаємодії факторів ризику, поведінки

атакуючих і користувачів та автоматизованого коригування ризиків. Для досягнення цієї мети в роботі необхідно вирішити такі **завдання**:

1) проаналізувати сучасні підходи до когнітивного моделювання та їх застосування у сфері інформаційної безпеки;

2) дослідити особливості сценарного підходу та визначити його роль в управлінні кіберризиками;

3) розробити та застосувати когнітивні моделі і сценарії для моделювання поведінки атакуючих та внутрішніх користувачів;

4) виконати симуляції та дослідити вплив різних концептів ризику на розвиток сценаріїв;

5) обґрунтувати можливості автоматизації моніторингу та коригування ризиків на основі когнітивних моделей;

6) визначити обмеження, виклики та перспективи розвитку когнітивного моделювання у сфері кібербезпеки.

Виходячи з цього, **об'єктом дослідження** є процес управління ризиками кібербезпеки в умовах невизначеності та динамічних загроз. **Предметом дослідження** є методи когнітивного моделювання та сценарного аналізу, їх застосування для оцінки, симуляції та коригування ризиків інформаційної безпеки.

Методи дослідження. У роботі застосовано методи когнітивного моделювання (нечіткі когнітивні карти, концептуальні графи), сценарний аналіз, моделювання поведінки агентів, методи симуляції включно зі сценарним проходженням ланцюгів атак, аналітичні та експертні підходи до оцінювання ризиків, а також елементи автоматизації логічних та причинно-наслідкових залежностей.

Наукова новизна одержаних результатів. Наукова новизна полягає у вдосконаленні підходів до управління кіберризиками шляхом комплексного поєднання когнітивного моделювання, сценарного аналізу та моделювання поведінки атакуючих і внутрішніх користувачів. У роботі сформовано інтегровану концептуальну модель сценаріїв кіберризиків, обґрунтовано

вплив взаємопов’язаних когнітивних факторів на розвиток інцидентів, а також запропоновано підхід до автоматизованого коригування ризиків на основі динамічних змін концептів.

Теоретичне та практичне значення полягає в обґрунтуванні необхідності та досліджені можливості застосування когнітивних моделей і сценаріїв для підвищення якості аналітичної роботи організацій, покращення прогнозування ризиків, моделювання поведінки порушників та оптимізації процесів прийняття рішень у системах кібербезпеки.

Галузь застосування. Результати роботи можуть бути використані у практиці управління ризиками кібербезпеки, у підрозділах моніторингу безпеки (SOC), у CERT-операціях, у побудові тренінгових симуляторів, а також у навчальному процесі при вивченні методів моделювання кіберзагроз та прийняття рішень.

Апробація результатів магістерської роботи. Основні положення роботи викладалися

1) на студентській науковій конференції «Безпека інформаційно-комунікаційних систем» (БІКС), Київський столичний університет імені Бориса Грінченка, Київ, Україна 26 жовтня 2025 року.

Тема доповіді «Когнітивне моделювання як інструмент інформаційної безпеки»;

2) на ІІІ міжнародній науково-практичній конференції «Сучасні аспекти діджиталізації та інформатизації в програмній та комп’ютерній інженерії», ДУІКТ, Україна, 4–6 грудня 2025 року.

Тема доповіді «Методи когнітивного моделювання сценаріїв кіберризику»;

3) Гаркушенко, А., Шевченко, С., Жданова, Ю., & (2025). Когнітивний підхід в інформаційній та кібербезпеці. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 1(29), 854–866. https://doi.org/10.28925/2663-4023.2025.29.945

Розділ 1 ТЕОРЕТИЧНІ ОСНОВИ КОГНІТИВНОГО МОДЕЛЮВАННЯ ТА УПРАВЛІННЯ РИЗИКАМИ В ІНФОРМАЦІЙНІЙ БЕЗПЕЦІ

1.1 Основи когнітивного моделювання: концепції, принципи, типи моделей.

Когнітивна наука є міждисциплінарною галуззю, яка знаходиться на перетині системного аналізу, комп'ютерних наук, математики, когнітивної психології та прикладної когнітивістики, а також швидко і динамічно розвивається.

Поштовхом до створення і розвитку когнітивної науки стало створення комп'ютерів та виникнення штучного інтелекту, що намагається імітувати людське мислення [1]. Сильний вплив на це мали ідеї математичної логіки, сформульовані науковцем-теоретиком Алан Тьюрінг, та інформаційної теорії, розробленої Клодом Шенноном. Ці науковці продемонстрували, що складні математичні ідеї та процеси можуть бути представлені формальними системами символів, дії з якими здійснюються за певними чіткими правилами.

Подальший перехід від формальної логіки до моделювання людського мислення, в свою чергу, було здійснено завдяки новаторській на той час роботі американських дослідників Герберта Саймона, Аллена Ньюелла та Кліффа Шоу. Їхні ранні когнітивні симуляції, зокрема представлені у програмах Logic Theorist (1956) та General Problem Solver (GPS) (1957) мали дуже великий вплив на формування інформаційно-процесингової психології. У цих роботах комп'ютер було використано не лише як інструмент для розрахунків, але і для того, щоб спробувати пояснити механізми людського вирішення проблем [2]. Ньюелл та Саймон заклали такі концептуальні основи у свої дослідження, які продовжують використовуватись для моделювання складних завдань у сучасній когнітивній психології. Окрім цього, ця філософія пізніше також лягла в основу розробки уніфікованих теорій пізнання, які є важливими для сучасних когнітивних структур, як, наприклад, ACT-R.

У сучасному розумінні, когнітивне моделювання – це характерна методологія дослідження когнітивної науки, результатом якої є теорії про когнітивні процеси, сформульовані у вигляді комп'ютерних програм [3]. Тобто це підхід у когнітивних науках, що формує теорії про мислення, пам'ять, навчання та прийняття рішень у вигляді комп'ютерних моделей, які, в свою чергу, дають можливість генерувати кількісні передбачення поведінки при різних задачах. Воно поєднує теорії, експериментальні дані та можливі реалізації у вигляді архітектур, що можуть бути перевірені на емпіричних даних.

Когнітивне моделювання стосується розробки обчислювальних моделей, які намагаються відтворити когнітивні процеси людини. Ця концепція є цінною в психології, штучному інтелекті (ШІ), когнітивних науках та у інших галузях, які передбачають взаємодію людини з комп'ютером [4].

Основними характеристиками когнітивного моделювання можна узагальнити так: орієнтація на глибинне пояснення психіки (глибинне розуміння), відтворення поведінкових процесів (реплікація), поєднання знань із різних наук, наявність прикладних та теоретичних цілей, а також постійний розвиток методів.

Розбираючи детальніше кожен з наведених пунктів, їх можна охарактеризувати наступним чином:

1. Глибинне розуміння: когнітивне моделювання намагається пояснити, як саме людина сприймає, запам'ятовує, приймає рішення чи реагує. Воно враховує не лише зовнішні прояви поведінки, а й внутрішні психологічні механізми, які не завжди усвідомлюються.

2. Відтворення процесів: моделі створюються таким чином, щоб відтворювати роботу людського мозку та дозволяти передбачати, як людина діятиме в тих чи інших умовах.

3. Міждисциплінарність: у цьому напрямі поєднуються підходи з психології, нейронаук, комп'ютерних наук, лінгвістики, штучного

інтелекту, що забезпечує широку та комплексну основу для дослідження мислення.

4. Цілі: моделі застосовуються як для теоретичного аналізу мислення, так і для практичних завдань, наприклад, покращення ефективності навчання, роботи інтерфейсів або створення інтелектуальних систем.
5. Динамічний розвиток: когнітивне моделювання не є сталою галуззю, воно змінюється під впливом нових наукових відкриттів про мозок і технічного прогресу. З часом моделі стають складнішими і точнішими, ще більше відображаючи глибші аспекти роботи свідомості та поведінки.

Принципами когнітивного моделювання є: принцип відповідності психологічним даним, принцип модульності, принцип процесуальності, адаптивності, принцип економії ресурсів та принцип інтеграції рівнів.

Принцип відповідності психологічним даним говорить про те, що побудована модель має узгоджуватись з практичними результатами поведінкової психології та нейронауки. Вона не може бути достовірною, якщо суперечить емпіричним фактам.

Принцип модульності передбачає, що когнітивні процеси мають ієрархічну структуру, в якій окремі функції виконуються спеціалізованими підсистемами.

Принцип процесуальності про те, що ключовим є не лише кінцевий результат, а і сам перебіг обробки інформації. Тобто хороша модель має пояснювати яким самим чином формується думка чи рішення.

В основі принципу адаптивності лежить те, що моделі мають передбачати механізми оновлення та розвитку, можливість перебудови чи зміни параметрів, адже, як правило, когнітивні системи змінюються під впливом часу, навчання та нового отриманого досвіду.

Принцип економії ресурсів зазначає, що моделі мають враховувати обмеженість когнітивної системи, тому що людина не може одночасно

обробляти неосяжні обсяги даних, тому використовує якісь спрощення чи шаблони.

Принцип інтеграції рівнів закладає в себе те, що модель має бути здатною пояснити не лише мислення в теорії, а і реальну поведінку суб'єкта.

Когнітивне моделювання є тісно пов'язаним, але концептуально відмінним від більш широких галузей штучного інтелекту (ШІ) та когнітивних обчислень. Детальніше кажучи, ШІ представляє собою найширшу категорію, яка охоплює технології, орієнтовані на автоматизацію. Когнітивні обчислення, у свою чергу, є галуззю ШІ, яка імітує людські розумові процеси для аналізу наданих даних та допомоги у прийнятті на їх основі рішень, часто посилюючи, а не замінюючи людські можливості. Когнітивний штучний інтелект (когнітивний ШІ) йде ще далі, імітуючи людські когнітивні функції, такі як навчання, міркування та вирішення проблем, динамічно розвиваючись та адаптуючись до нових ситуацій, використовуючи такі технології, як машинне навчання, обробка природної мови та глибоке навчання [5]. КМ у цьому виконує роль методологічної основи.

Процес когнітивного моделювання є циклічним і охоплює повний спектр від теоретичної побудови до чітко визначеного емпіричного аналізу. Це, в свою чергу, включає в себе розробку та аналіз когнітивної моделі складної системи.

Когнітивна модель – це формалізований опис механізмів, які пояснюють певний аспект людської когніції (наприклад, пам'ять, увага чи планування), реалізований у вигляді програмного або математичного представлення. Оскільки когнітивна модель часто реалізована саме у вигляді комп'ютерної програми, її можна розглядати як генеративну теорію даної когнітивної функції, оскільки вона не лише пояснює, але й фактично виконує цю функцію.

Розвиток когнітивного моделювання ґрунтується на декількох ключових філософських та концептуальних рамках, які дають визначення як слід будувати та оцінювати когнітивні теорії.

У 1980-х роках Девід Марр, британський нейробіолог та психолог, запропонував концептуальну основу, яка пізніше стала фундаментальною для когнітивної обчислювальної науки. У своїх роботах Марр стверджував, що повне розуміння будь-якої обчислювальної проблеми, зокрема когнітивного процесу, вимагає аналізу на трьох різних, але взаємопов'язаних рівнях. Цими рівнями є обчислювальний рівень, алгоритмічний рівень та рівень імплементації [6].

Обчислюваний рівень (Computational Level) є найбільш фундаментальним, і відповідає на питання «що?» і «чому?». Тобто, на цьому етапі ми ще не заглиблюємось в те, як це працює, але визначаємо мету обчислення, а саме, яку конкретно проблему має вирішити створена когнітивна система, та яка логіка її необхідності.

На алгоритмічному рівні (Algorithmic Level) ми вже фокусуємось на питанні «як?». Його основа це процеси та правила. Тут описується, яким саме чином досягається визначена раніше обчислювальна мета. Цей рівень вимагає специфікації репрезентацій (структур, в яких зберігається інформація) та процедур (алгоритмів), що використовуються для маніпуляції цими репрезентаціями. Простіше кажучи, на цьому рівні розробляються стратегії, моделі та плани для вирішення поставленої проблеми.

Рівень імплементації (Implementation Level) або ж рівень впровадження стосується вже фізичної реалізації процесів і репрезентацій, які були визначені раніше на алгоритмічному рівні. Він відповідає на питання про механізми, які здійснюють ці алгоритми. Дії на цьому етапі мають реалізовувати визначені раніше плани з урахуванням складнощів і викликів реального світу, які можуть виникнути у процесі.

Жоден з вище описаних рівнів не може існувати самостійно для якісного представлення та впровадження моделі. Марр неодноразово наголошував, що успішні когнітивні моделі мають не лише бути біологічно правдоподібними, але також демонструвати раціональність (обчислюваний рівень) та мати чітко визначені структури (алгоритмічний рівень). Ця концепція пізніше стала

підставою для еволюції моделей від чисто символьних до більш комплексних гібридних архітектур.

Ключовий теоретичний принцип, який лежить в основі когнітивного моделювання, був запропонований Гербертом Саймоном і відомий нам як концепція обмеженої раціональності. Цей принцип є альтернативною основною для математичного моделювання прийняття рішень, на відміну від більш класичних моделей, які розглядають прийняття рішення як пошук оптимального вибору (раціональність як оптимізація). У своїх дослідженнях Саймон стверджував, що раціональність людей у більшості випадків обмежена пізнавальними можливостями їхнього розуму. Тобто у такому разі приймачі рішення діють як «задовольняючі агенти», основною метою яких є знаходження задовільного рішення, а не оптимального варіанту. Для демонстрації власних слів, Саймон використав аналогію ножиць, де одне лезо – це когнітивні обмеження людей, а інше – структури навколишнього середовища, таким чином показуючи компенсацію обмежених ресурсів розуму завдяки використанню відомих структурних закономірностей в середовищі.

Цей принцип також має пряме відношення до кібербезпеки, оскільки якби користувачі завжди діяли оптимально, тобто використовували принципи класичної раціональності, то не потрапляли б, наприклад, на фішингові приманки. Моделювання обмеженої раціональності натомість дозволяє системам захисту робити прогнози за яких можливих умов (наприклад, емоційне виснаження, високе емоційне навантаження чи дефіцит часу) та в які моменти користувач має більшу ймовірність відхилитись від безпечної поведінки, шукаючи задовільний варіант, а не оптимально безпечний [7].

Аллен Ньюелл та Герберт Саймон також сформували гіпотезу фізичної символьної системи (PSSH) у 1976 році, яка говорить, що фізична система, яка здатна маніпулювати символами, може виявляти інтелектуальну поведінку [8].

Згідно PSSH, інтелект прирівнюється до здатності використовувати символи для обробки та представлення інформації. У цій гіпотезі мислення

розуміється як певний процес маніпулювання символами відповідно чітко встановлених правил. Саме завдяки цим концепціям було прокладено шлях до ранніх досліджень ШІ в яких припускалось, що когнітивні функції людини можуть бути змодельовані обчислювальними засобами, оскільки людський мозок функціонує подібно до комп'ютера, в якому когніція є наслідком символьних маніпуляцій. Зокрема, Джеррі Фодор створив гіпотезу мови думки (The language of thought hypothesis (LOTH)), яку можна назвати конкретною реалізацією PSSH у людській когнітивній системі [9].

Говорячи про взаємозв'язок і порівняння концепції обмеженої раціональності та гіпотези фізичної символьної системи, важливо зазначити, що PSSH не каже, що система завжди має знаходити оптимальні варіанти, вона говорить про те, яким типом системи є мозок.

Тобто, PSSH відповідає на питання «Що таке мислення як система?» і стверджує, що мислення може бути реалізоване у вигляді фізичної системи, яка виконує маніпуляції з символами за встановленими правилами. Вона про архітектуру/механізми системи. У той час як концепція обмеженої раціональності відповідає на питання «Яким саме чином люди приймають рішення в умовах обмеженості?» і стверджує, що у випадку обмеженості ресурсів, люди застосовують евристики і рішення з категорії «задовільно», аніж з категорії глобальної оптимізації. Це твердження про поведінку та стратегію. Буде помилкою вважати, що PSSH говорить, що система має завжди знаходити оптимальні рішення або мати необмежені ресурси, адже головна ідея заключається в тому, що система маніпулює символами і правилами. Прикладом може слугувати програма, що шукає шлях у графі. Вона може бути реалізована як символічна система і виконувати або повний (оптимально за хорошою евристикою) або обмежений пошук (наприклад, до певного часу або бюджету) і повернути перший знайдений прийнятний шлях. Кожен із запропонованих варіантів є символічним алгоритмом, що відрізняється лише стратегіями.

Когнітивні моделі класифікуються відповідно до фундаментальних припущень щодо того, як інформація представлена і як обробляється, відображаючи різні підходи до алгоритмічного рівня Марра. Можна виділити символьні моделі, конекціоністські моделі та гібридні [10].

Символьні моделі історично протягом довгого часу були домінуючими і найбільш використовуваними обчислювальними підходами до пізнання. Вони мають своє коріння від PSSH, і представляють інформацію у вигляді дискретних символів та явних правил, найчастіше виражених у формі «якщо-то» («if-then»). Ці моделі припускають, що когнітивні процеси керуються визначеними алгоритмами та полягають у послідовній маніпуляції символами. Обробка інформації в них здійснюється за принципом послідовної, дискретної логічної обробки, яка є аналогічною до програмування на мовах логічного типу. Це означає, що дана модель може пояснити логічні міркування, планування, вирішення конкретної задачі чи прийняття рішення в конкретних умовах крок за кроком.

Цей підхід не суперечить раціоналістській філософській школі, яка наголошує на важливості вродженого або набутого знання та міркування, заснованого на логіці. Саме тому від досі лишається успішним у моделюванні високорівневої когніції, логічного виведення та вирішення поставлених проблем.

Одними з найвідоміших прикладів реалізації символьного підходу є General Problem Solver (GPS) та SOAR.

General Problem Solver (GPS) – це комп’ютерна програма, основною ідеєю якої є виявлення загальних методів розв’язання проблем шляхом вираження задачі у вигляді пошуку у просторі, де існує початковий стан і цільовий стан, а також оператори, які переводять один стан в інший. Простіше кажучи, її мета розділити задачу на підзадачі, оцінити поточну різницю між ціллю і станом та застосувати оператор, який має зменшити цю різницю. Завдяки цьому процесу GPS має змогу знайти рішення математичних теорем,

текстових задач, логічних доказів та широко спектру інших проблем, що мають чітке визначення [11].

GPS має цінне історичне значення, як один із найперших фундаментальних проектів, які мали на меті змоделювати людське вирішення проблем за допомогою комп'ютерної програми. І хоча його універсальність була обмежена, саме він став поштовхом і стимулював розвиток інших когнітивних архітектур, а також підходів до моделювання мислення.

SOAR – архітектура загального призначення, розроблена Джоном Лердом, Алленом Ньюеллом та Полом Розенблумом в Університеті Карнегі-Меллона. Вона має на меті створення універсальних обчислювальних структурних модулів, які є необхідними для побудови загальних інтелектуальних агентів, тобто систем, здатних виконувати різноманітні завдання, опрацьовувати, застосовувати та засвоювати різні типи знань з метою відтворення повного набору людських когнітивних здібностей. SOAR реалізовує низку гіпотез щодо обчислювальних структур, які лежать в основі загального інтелекту. Багато з цих ідей поділяються і іншими когнітивними архітектурами, зокрема і ACT-R, проте відмінністю є те, що останні дослідження в межах SOAR були зосереджені переважно на розвитку загального штучного інтелекту (ШІ) і підвищенні його функціональності та ефективності, в той час як ACT-R традиційно спрямований на точне моделювання людських когнітивних процесів [12].

Основний цикл обробки SOAR полягає у взаємодії між процедурною пам'яттю (знанням, яким чином виконувати дії) та робочою пам'яттю (представлення поточної ситуації) для підтримки вибору та застосування операторів. Поведінка агента SOAR керується вже згаданим вище принципом раціональності, згідно якого агент вибирає оператор у тому випадку, якщо він знає, що його застосування призведе до цілі.

Окрім того, ключовим механізмом навчання в SOAR є чанкінг (chunking), тобто у випадку, якщо агент стикається з проблемою або заходить у тупик, то він переходить до підзадач. Результати обробки у підзадачах

компілюються у нові правила. І у майбутньому, у схожих ситуаціях, ці правила будуть вже спрацьовувати автоматично, таким чином перетворюючи складне міркування на реактивну обробку. В додачу до цього, SOAR підтримує навчання з підкріпленням для точного налаштування числових переваг при оцінці операторів.

Конекціоністські моделі**,** які також відомі як моделі паралельної розподіленої обробки (DPD) або штучні нейронні мережі (ANN), мають за основу принципи нейронних обчислень. Вони часто представлені у вигляді нейронних мереж і імітують обробку інформації на рівні, більш близькому до біологічного функціонування мозку. Ці моделі мають в основі те, що обчислення виконуються простими, взаємопов’язаними одиницями, що функціонують паралельно, тобто фокусуються на паралельній розподіленій обробці. Знання представлені тут не у вигляді явних правил, а скоріше як розподілена активація та ваги зв’язків між численними взаємопов’язаними вузлами [13]. Навчання в даному випадку відбувається шляхом коригування сили цих зв’язків.

Ця парадигма в свою чергу співвідноситься з емпіристською філософською школою, роблячи наголос на важливості навчання асоціацій з великих даних [14].

Конекціоністські моделі забезпечують гнучкість, здатність ефективно навчатись з великих масивів даних, високу біологічну правдоподібність та стійкість до шуму. Вони успішно використовуються для моделювання таких процесів, як сприйняття, пам’ять, візуальне розпізнавання та формування понять. Головним недоліком даних моделей є так звана «проблема чорного ящика», простіше кажучи, їхня непрозорість або ж проблема інтерпретованості, оскільки часто важко або взагалі неможливо визначити як саме було прийнято рішення в мережі, і це створює перешкоди для їх впровадження у критично важливі системи безпеки.

Гібридні моделі на даний час є найбільш перспективними та широко використовуються для моделювання складного пізнання. Вони виникли як

спроба подолати обмеження чисто символьних або ж чисто конекціоністських підходів шляхом їхньої інтеграції. Саме тому вони поєднують в собі сильні сторони символьних (прозорість міркування) та конекціоністських (адаптація та навчання) підходів. Таким чином гібридний підхід дозволяє використовувати конекціонізм для низькорівневої обробки даних (наприклад, розпізнавання шаблонів) та символізм для високорівневого планування та логіки, що надає змогу моделювати вже складну когнітивну гнучкість. Завдяки цьому, гібридні архітектури дозволяють моделювати, як швидкі, автоматичні процеси, так і повільні, свідомі, забезпечуючи комплексне розуміння. Прикладом гібридної моделі може слугувати ACT-R.

ACT-R (Adaptive Control of Thought – Rational) – це когнітивна архітектура, яку розробив Джон Р. Андерсон, і яка спрямована на моделювання широкого спектру когнітивних процесів, у який входить пам’ять, увага, навчання, вирішення задач, а також взаємодія людини та комп’ютера [15]. Архітектура має символічні та субсимволічні елементи, і саме через це вона вважається однією з найповніших платформ для моделювання людського мислення. Символьна структура представлена системою продукцій, субсимвольна складається з математичних рівнянь, які працюють паралельно, і контролюють швидкість та вибір символьних процесів. Ця архітектура складається з певних модулів, які представляють собою когнітивні функції. Ці модулі включають декларативну пам’ять, процедурну пам’ять, а також механізми активації, витіснення та затримки [16].

Говорячи детальніше, декларативна пам’ять містить фактичну інформацію у вигляді окремих фрагментів, які є схематичними структурами з атрибутами (простіше кажучи, факти, наприклад «2+2=4»). Доступ до цих даних здійснюється через механізм, що ґрунтується на субсимвольній активації. Процедурна пам’ять зберігає продукційні правила (пари «умова-дія»), які визначають виконання завдань. Ці правила змагаються між собою за вибір на основі показника корисності, що формується з попереднього досвіду успішності. Буфери в свою чергу являють собою тимчасові сховища даних, які

забезпечують обмін інформацією між різними модулями. Наприклад, буфер пошуку для декларативної пам’яті). І стек цілей містить активні цілі та підцілі, що дає змогу організовувати процес розв’язання задач у вигляді ієрархічної структури.

ACT-R успішно застосовується для моделювання і пояснення явищ у різних галузях: від спискової пам’яті, розуміння тексту, діагностичного мислення та ймовірнісного навчання до водіння, польотів та використання у навчальних середовищах [17].

Перевагами цієї архітектури є інтеграція символічного та субсимволічного підходів, а також можливість отримувати кількісні передбачення (наприклад, час реакції чи ймовірність помилки), велика база досліджень.

Обмеженнями можна вважати те, що дані моделі можуть бути складними для побудови, присутня обмеженість у моделюванні дуже динамічних або складних нелінійних середовищ, є велика залежність від параметрів. Залежність ACT-R від чітко встановлених фіксованих правил обмежує її адаптивність та масштабність в порівнянні з деякими іншими архітектурами [18].

Нижче наведена порівняльна таблиця моделей:

Таблиця 1.1

Порівняння когнітивних моделей

Тип моделі	Представлення знань	Механізм обробки	Основна перевага	Ключове обмеження
Символічний	Явні символи та правила	Послідовне міркування та маніпуляція правилами «Якщо-То»	Висока прозорість та точна логіка	Жорсткість, низька адаптація
Конекціоністський	Розподілені ваги зв’язків	Паралельна розподілена обробка	Гнучкість, біологічна правдоподібність	Непрозорість
Гібридний	Модулі (символічні та субсимволічні)	Комбінована, модульна обробка	Синтез міркування та навчання,	Висока обчислювальна складність,

			адаптивність, комплексність	інтеграційні виклики

Від вибору форми залежить спосіб функціонування моделі та її архітектура.

1.2 Особливості кібербезпеки як сфери застосування когнітивного моделювання

Для того, щоб зрозуміти як користувачі думають, реагують та приймають рішення, когнітивне моделювання використовують у таких сферах як освіта, охорона здоров'я, реклама, оборона, розробка програмного забезпечення.

У контексті кібербезпеки когнітивне моделювання важливе для розв'язання задач, у яких фігурує людська поведінка, таких як, наприклад, розпізнавання фішингу, прийняття рішень під тиском часу, соціальна інженерія чи внутрішні загрози.

Важливо зазначити, що в умовах, коли кіберзагрози все дедалі більше можуть залежати від людського фактору, інтеграція психологічного компоненту в систему захисту є критично необхідною. Зі зростанням кількості психологічно-орієнтованих атак, когнітивне моделювання надає інструментарій для кількісного опису людської поведінки. КМ дозволяє формалізувати яким саме чином суб'єкт (користувач або зловмисник) обробляє, зберігає та використовує інформацію. Це, в свою чергу, дає змогу перевести оцінку ризику, який пов'язаний з людським фактором, з якісної площини у вимірювану кількісну модель, яка є необхідною умовою для побудови і впровадження адаптивних систем кіберзахисту. Огляд досліджень показує, що моделювання нападників, захисників та користувачів дає змогу створювати реалістичні симуляції та тести безпеки.

Конкретно у контексті кібербезпеки застосування когнітивного моделювання базується на кількох архітектурних підходах, які мають свої

власні переваги та обмеження. Це розглянуте вище символічне моделювання, моделювання з використанням нейронних мереж і ймовірнісне та агентне моделювання.

В даному випадку мережеве моделювання (на основі нейронних мереж) передбачає використання різних методів AI/ML. Серед них особлива увага приділяється ансамблевому машинному навчанню, яке показує ефективність у підвищенні стабільності та точності прогнозів. Ефективність ансамблів зумовлена агрегуванням результатів кількох класифікаторів, що, в свою чергу, знижує ризик випадкових помилок окремих моделей. Тобто замість використання однієї моделі (наприклад, одна нейронна мережа чи один алгоритм класифікації) використовується кілька моделей одночасно і їх результати поєднуються. Це має важливе значення для центрів операцій безпеки (SOC), де аналіз здійснюється на основі гетерогенних джерел даних (мережеві журнали, поведінкові патерни, показники аномалій), а загрози характеризуються високою варіативністю і динамікою [19].

Говорячи про ймовірнісне (наприклад, баєсівські моделі, які розглянуто детальніше далі) та агентне моделювання, можна сказати, що це дозволяє ефективно аналізувати складну динаміку загроз і систем безпеки, особливо в умовах, коли є потреба в моделюванні багаторівневих взаємодій між різними елементами.

Фундаментальним елементом когнітивної кібербезпеки є принцип human-in-the-loop (HITL), який визначається як спільний підхід, заснований на інтеграції людського внеску та експертизи в безперервний життєвий цикл машинного навчання. У цьому циклі людина активно бере участь у тренуванні, оцінці та операційній діяльності моделей, керуючи, а також даючи активний зворотній зв’язок та рекомендації [20].

Необхідність HITL випливає з того, що навіть найдосконаліші на перший погляд моделі мають проблеми з обробкою неоднозначності, упереджень чи граничних випадків, які відхиляються від їх тренувальних даних [21]. Інтеграція людини у цей процес, робить результат більш надійним,

забезпечує підвищення точності, а також слідкує за етичністю прийняття рішень і прозорістю.

Окрім того, що людина є хорошим помічником, вона також є джерелом потенційних уразливостей і початковим вектором можливої атаки. Наприклад, коли переходить за шкідливими посиланнями, розкриває конфіденційні дані чи банально використовує слабкі паролі. Соціальна інженерія використовує людські якості, стани чи вразливості, як от довіра, страх, перевантаженість чи поспіх. Згідно статистичних даних, більшість кібератак ініціюється саме через помилки, допущені людьми [22]. Важливо зазначити, що у цей спектр входять як навмисні дії, так і несвідомі помилки користувачів. Саме тому, зрозуміти поведінку людини у сфері інформаційної та кібербезпеки – означає зрозуміти механізм атаки та можливі точки її зупинення.

Окрім цього, в умовах швидкого розвитку технологій, включаючи штучний інтелект, який може ускладнювати атаки, необхідність у формалізованих, передбачуваних моделях людської когніції стає ще більш важливою та потрібною. В даному випадку когнітивне моделювання пропонує методологічну основу, яка забезпечує архітекторам безпеки нові, ширші інструменти для управління ризиками. Правильне застосування ефективних технік та моделей, а також свідоме ставлення до інформаційної безпеки може суттєво зменшити ризик успішних кібератак.

Когнітивне моделювання виступає інструментарієм, спеціально розробленим для роботи з динамічними складними системами, де є ступінь невизначеності. Його цінність полягає в тому, щоб формалізувати знання, які або є нечіткими, або просто мають суб’єктивний характер, завдяки чому вираховується як якісні, так і кількісні фактори впливу на систему.

В такому випадку, одна із очевидних переваг використання КМ, зокрема когнітивних карт, розглянутих далі, є прогнозування не чисельних показників, а тенденцій розвитку процесів у системі, яке досягається шляхом моделювання динаміки імпульсного процесу системи при гіпотетичних змінах вхідних факторів або ж їх комбінацій. І оскільки цей підхід дозволяє приймати рішення

на основі розуміння можливих векторів еволюцій загроз та станів захищеності, а не лише на ретроспективному аналізі подій, то він може мати вирішальне значення для стратегічного управління кіберконфліктом [23].

Окрім того, хочеться ще окремо сказати про центр операцій з безпеки (Security Operations Center, SOC). SOC – це централізований підрозділ або команда фахівців з кібербезпеки, що відповідає за моніторинг, виявлення, аналіз та реагування на загрози інформаційної безпеки. Тобто виконує функцію захисту організації, забезпечуючи цілодобовий моніторинг мережі та розслідування інцидентів безпеки, зібраних аналітиками SOC.

Важливо зауважити, що це середовище вимагає постійної когнітивної пильності. Операційне підвищене навантаження SOC аналітика може підпадати під визначення обмежень людської когніції, зокрема під описану вище концепцію обмеженої раціональності, адже навіть експерти все одно залишаються людьми, можуть мати лімітовану пам’ять чи перевантаження від постійного поглинання великої кількості інформації. В деяких випадках це може призводити до когнітивних упереджень у їх судженнях, що є критично небезпечним при оцінці загроз.

Одним із прикладів когнітивних обмежень у даному концепті є феномен втоми безпеки. Цей стан викликається безперервною необхідністю бути зосередженим, підтримувати рівень пильності та реагувати на нескінченний потік попереджень та вимог безпеки. Втома має безпосередній зв’язок зі зниженням якості та ефективності прийнятих рішень та зі збільшенням кількості пропущених загроз.

Можливість мінімізації впливу цих описаних вище когнітивних обмежень може бути досягнути завдяки застосування КМ. Оскільки зменшити обсяг даних є неможливою задачею (це суперечитиме вимогам моніторингу), то ефективним шляхом є оптимізація взаємодії людини з цими даними. У даному випадку когнітивне моделювання, моделюючи процеси прийняття рішень, дозволяє автоматично пріоритизувати алерти чи хибні спрацювання. Така дія безпосередньо впливає на зниження «втоми безпеки», адже зменшує

кількість навантаження і зайвого шуму, який у іншому випадку вимагав би постійного ручного висококонцентрованого втручання.

1.3 Сценарний підхід в управлінні ризиками інформаційної безпеки

Сценарний підхід є фундаментальною методологією, яка забезпечує перехід від статичних методів, заснованих на контрольних списках, до проактивного системного передбачення, та формує цілісний причинно-наслідковий наратив про потенційне порушення безпеки.

Сценарний підхід поєднує в собі математичні моделі, програмні рішення, логічні схеми та організаційні механізми, які дозволяють визначити оптимальну послідовність застосування окремих методів, встановити взаємозв’язки між ними та сформувати системний процес прогнозування [24]. Іншими словами, сценарний аналіз забезпечує можливість опису, порівняння та оцінювання альтернативних сценаріїв зміни у системі з урахуванням різних факторів невизначеності, ризиків і поведінкових аспектів. Завдяки цьому підходу можна не лише передбачити потенційні наслідки певних рішень чи загроз, але й виявити критичні точки впливу, визначити найбільш і найменш ймовірнісні траєкторії розвитку подій, а також сформувати стратегії реагування та управління ризиками, які ґрунтуються на фактах. Тобто сценарний аналіз також забезпечує менеджеру інформаційної безпеки повний інструментарій для ефективної протидії ще до фактичної реалізації, а не лише прогнозує можливі інциденти.

Основними принципами сценарного підходу є прогнозування, гнучкість, багатоваріантність, а також урахування поведінкових характеристик учасників інформаційної взаємодії.

Принцип прогнозування передбачає моделювання можливих станів системи ІБ в майбутньому на основі вже наявних даних або тенденцій, які прослідковуються. Прогнозування дозволяє не лише реагувати на загрози

постфактум, а здійснювати проактивне управління ризиками, таким чином створюючи основу для випереджального прийняття рішень.

Гнучкість може проявлятись у можливості адаптації вже побудованих сценаріїв, якщо змінюються зовнішні або внутрішні умови, що є важливим враховуючи, що кіберзагрози постійно розвиваються та вдосконалюються, а статичні жорстко фіксовані моделі швидко втрачають свою актуальність. Завдяки цьому принципу сценарний аналіз дає змогу оперативно переглядати оцінку ризиків, оновлювати сценарії та змінювати пріоритети в режимі реального часу.

Третій принцип про багатоваріантність оцінки закладає в себе те, що розвиток ситуації ніколи не є однозначним. Змінність середовища, поява нових технологій, поведінкові особливості користувачів та інші подібні фактори формують множину можливих сценаріїв. Тому замість побудови одного загального прогнозу, сценарний підхід передбачає створення різних альтернативних моделей розвитку подій, які коливаються від найменш імовірних до критичних.

Принцип про врахування поведінкових характеристик учасників інформаційної взаємодії говорить про те, що у сфері кібербезпеки взаємодіють люди (наприклад, користувачі, адміністратори чи аналітики), технічні системи, а також потенційні зловмисники, і кожен з них має власні цілі, обмеження чи стратегії за якими вони діють. Тому сценарний підхід включає у себе аналіз поведінкових патернів різних учасників взаємодії у системі, а також можливих реакцій, що дозволяє передбачити не лише технічні, а й соціально-психологічні аспекти безпеки.

Ефективна розробка сценаріїв ІБ вимагає ґрунтовної методологічної основи, яка починається зі стратегічного контексту і закінчується детальним моделюванням зловмисника. Цей процес зазвичай є системним та послідовним. Детальніше про етапи створення і розробки сценаріїв розглянуто у розділ 3.

Коли ми говоримо про сценарний аналіз, то можна також згадати такі нормативні документи як ISO/IEC 27005 та NIST SP 800-30.

ISO/IEC 27005 – це міжнародний стандарт, що надає методологічні вказівки з управління ризиками інформаційної безпеки і є частинкою стандартів ISO 27000. Цей стандарт описує цикл ризик-менеджменту (наприклад, встановлення контексту, оцінка, обробка, прийняття ризиків, комунікація та моніторинг), а також містить рекомендації щодо ідентифікації активів, загроз, властивостей і оцінки ймовірності і впливу [25]. Він застосовується до організацій будь-якого типу та розміру, які мають на меті керувати ризиками, що можуть компрометувати інформаційну безпеку.

Стандарт описує етапи управління ризиками і говорить про опис потенційних подій/інцидентів як сценаріїв для того, щоб оцінити наслідки та ймовірність їх реалізації. Таким чином, сценарій слугує необхідним мостом між технічним аналізом та бізнес-оцінкою, визначаючи наслідки через попередню вартість активів. В такому випадку, робиться акцент на визначенні контексту і чинників (сюди входять характеристики системи, бізнес-процеси, зацікавлені сторони), які впливають на вибір сценарію і його оцінку. Більш того, ISO 27005 окремо охоплює безпеку людських ресурсів, включаючи користувачів, працівників, підрядників.

NIST SP 800-30 Rev.1 – це нормативний документ, розроблений Національним інститутом стандартів і технологій США, який описує систематичний підхід до оцінювання ризиків ІБ. Він не обмежений державними установами, хоч і початково розроблявся саме для них, натомість навпаки активно використовується на сьогоднішній день у приватних компаніях, банках та інших установах, оскільки його структура добре підходить для практичної реалізації ризик-менеджменту, що робить його цінним інструментом.

NIST SP 800-30 визначає, як системно оцінити рівень ризику, що може виникнути внаслідок кіберзагроз, технічних вразливостей, людського чинника чи впливу інцидентів.

Метою документу є надання логіки, яка дозволяє виявити можливі сценарії загроз, визначити ймовірність їх реалізації, оцінити можливі наслідки, класифікувати ризики за критичністю та визначити або способи їх зниження, або прийняття. Це все відбувається у чотири прописані етапи: підготовка, проведення оцінки, документування та підтримка і оновлення [26].

Говорячи детальніше, на першому етапі, підготовка до оцінки ризику (prepare), визначається мета, система, її межі, контекст, зацікавлені сторони, збирається необхідна документація. В результаті чого отримується чіткий контекст і перелік активів, які треба оцінити.

На другому етапі, проведення оцінки ризики (conduct), відбувається ідентифікація загроз, вразливостей, можливих сценаріїв, а також оцінка ймовірності та впливу. У результаті отримується перелік ризиків із кількісними та якісними показниками.

На третьому етапі, документування результатів (communicate), відбувається формування звіту і його передача відповідальним особам. Організація документує виявлені ризики, їх рівень та рекомендації з їх усунення, щоб на основі цього прийняти обґрунтоване рішення. Тобто, в результаті отримується сформований звіт для прийняття рішень щодо захисту.

На фінальному четвертому етапі, підтримка та оновлення (maintain), відбувається моніторинг змін системи та оновлення оцінки. В результаті підтримується актуальність ризик-менеджменту в часі. Цей етап про те, що оцінка ризиків – це не одноразова дія, а безперервний процес, який вимагає регулярного перегляду та оновлення оцінки.

NIST наголошує, що ризик має розглядатись не абстрактно, а через конкретні сценарії («загроза – використовує вразливість – впливає на актив – спричиняє наслідки»). NIST також радить документувати припущення та підґрунтя для оцінок ймовірності/впливу.

Підсумовуючи, обидва документи вимагають системний та документований підхід, що дає можливість зіставити сценарії та визначити пріоритети обробки ризиків. Проте вони також мають і відмінності, а саме те,

що ISO/IEC 27005 надає загальні вказівки з управління ризиками в межах стандарту і служить рамковою основою для управління ризиками ІБ, в той час як NIST SP 800-30 описує конкретні етапи процесу управління ризиками і фокусується на конкретних методиках. Таким чином перший стандарт забезпечує гнучкість та варіативність побудови сценаріїв, а другий їх структурованість і формальність.

Також важливо зазначити, що сценарний підхід набуває реальної цінності лише тоді, коли в його основі реальні актуальні дані, а також структурована інформація про загрози.

1.4 Взаємозв'язок когнітивного моделювання і сценарного аналізу

Когнітивне моделювання та сценарний аналіз є взаємодоповнювальними методами аналізу в кібербезпеці, особливо коли ми говоримо про прогнозування поведінки учасників інформаційного процесу та оцінку наслідків, якщо загроза буде реалізована. Когнітивне моделювання забезпечує механізм пояснення і відтворення процесу того, як користувачі, аналітики або зловмисники приймають рішення, в той час як сценарний аналіз розглядає різні варіанти розвитку подій у системі, враховуючи різні фактори і чинники. Іншими словами, сценарний аналіз відповідає на питання «що може статись і як це вплине на систему?», а когнітивне моделювання відповідає на питання «як саме учасник системи прийме рішення?», наприклад, когнітивна модель може описати ймовірність, що користувач перейде за шкідливим посиланням. Таким чином, взаємна інтеграція цих підходів дозволяє робити оцінку ризиків повнішою, більш реалістичною та прогнозованою.

Як вже було написано раніше, когнітивне моделювання має здатність об'єднувати два критично важливих види моделювання, а саме структурно-системне та імітаційне. І саме це є ключовим для інтеграції з сценарним аналізом. Когнітивне моделювання в даному випадку надає необхідну

структуровану основу (наприклад, когнітивні карти) для подальшого формування і динамічного розігрування гіпотетичних сценаріїв [27].

Зі стратегічної точки зору, використання сценарного планування має вирішальне значення для того, щоб прогнозувати можливі зміни та формувати альтернативі дієві стратегії управління. А когнітивне моделювання забезпечує технологічну базу, яке перетворює концептуальне сценарне планування вже на формалізований прогностичний інструмент.

Сценарний підхід, що застосовується в рамках когнітивного моделювання, часто можуть називати «динамічним імітаційним моделюванням». Таким чином, він займається «розігруванням» різних варіантів розвитку подій, що залежать від управлінської моделі, яку вибрали, чи від поведінки непередбачуваних зовнішніх чинників. Головна функція когнітивної моделі у даному випадку полягає у тому, щоб сприяти повному і адекватному врахуванні повного комплексу ефектів для їхніх факторів, в тому числі включаючи їх динаміку та взаємозв'язки у різних умовах [28]. Цей підхід дозволяє не лише виявити різні вектори розвитку подій, але і розробити прогнози щодо наслідків до яких призведе та чи інша вибрана стратегія. Кінцевою метою є висунення пропозицій з приводу оптимальної стратегії, яка буде спрямована на реалізацію вибраного сценарію.

Якщо порівнювати традиційний сценарний аналіз та сценарний аналіз на основі когнітивного моделювання, особливо з використанням нечітких когнітивних карт, про які детальніше розказано у розділі 2, то можна побачити, що другий являє собою технологічно вдосконалену версію класичного СА. В таблиці 2 наведена порівняльна характеристика класичного СА та СА на основі когнітивного моделювання з використанням нечітких когнітивних карт.

Таблиця 1.2

Порівняння класичного сценарного аналізу та сценарного аналізу на основі когнітивного моделювання на основі НКК

Критерій порівняння	Традиційний сценарний аналіз (СА)	Сценарне моделювання на основі НКК
Формалізація структури	Опис ключових чинників, які часто базується на лінійних прозорих зв'язках	Формалізована структура (функціональний граф) зі здатністю відображати нелінійні та взаємозалежні відносини
Обробка інформації	Переважно експертні оцінки, результати залежіть від досвіду, тенденцій, інтуїції та суб'єктивних оцінок	Інтеграція якісних та кількісних даних, можливість роботи в умовах дефіциту кількісних даних завдяки нечіткій логіці. Впливи між факторами обчислюються за матрицею зв'язків, можливе використання симуляцій та ітераційних обчислень. Значно нижчий рівень суб'єктивності
Динаміка і зміни	Моделювання розвитку подій здебільшого статичне, після побудови сценаріїв їх структура майже не змінюється. Перехід між станами зазвичай описується дискретно	Моделювання динамічне та імітаційне. Зміна одного фактору викликає ланцюжкову реакцію. Є можливість моделювати часову еволюцію, а також адаптація моделі під нові дані
Облік невизначеності	Невизначеність зазвичай вираховується шляхом створення декількох гіпотетичних сценаріїв (песимістичний, оптимістичний, базовий). Однак ступінь впевненості слабо формалізована	Врахування невизначеності вбудовано у структуру моделі. Можлива зміна ваг зв'язків, впливів чи ймовірностей, використання нечіткої логіки або байєсівських підходів. Невизначеність представлена кількісно

Тобто традиційний аналіз більш сфокусований на описовому передбаченні можливих траєкторій розвитку подій. Експертна оцінка є його основним інструментом, роблячи цей підхід менш чутливим до динамічних змін у системі чи прихованих причинно-наслідкових залежностей. Він підходить для стратегічного планування, але набагато менш ефективний у випадках ситуацій зі складною, багатофакторною та швидкозмінною взаємодією, яка в тому числі характерна для кібербезпеки.

В той час як сценарне моделювання на основі тих самих нечітких когнітивних карт, навпаки, дає змогу формалізувати складні взаємозв'язки, в тому числі нечіткі, опосередковані чи зворотні впливи між подіями. Завдяки

цьому стає можливим прогнозування динамічних поведінкових процесів, коригування моделі у режимі реального часу та врахування невизначеності. Це робить цей підхід ефективним для моделювання поведінки порушників, ескалації загроз, прийняття рішень під тиском, в тому числі тиску часу, а також для аналізу ризиків у розподілених інформаційних системах.

Висновки до розділу 1

Когнітивне моделювання є важливим інструментом кібербезпеки, адже забезпечує аналітичні інструменти, здатні функціонувати в умовах невизначеності та обмеженості даних, що є притаманним для інформаційної інфраструктури. Воно також є необхідним, адже дуже часто саме людський фактор стає джерелом вразливостей.

З одного боку, когнітивне моделювання захищає власні когнітивні ресурси, пом’якшуючи дії таких обмежень, як наприклад, втома безпеки чи когнітивні упередження, таким чином знижуючи пропущені загрози. З іншого боку, КМ також використовується в наступальному сенсі, дозволяючи прогнозувати та експлуатувати індивідуальні упередження зловмисників, досягаючи більшого прогнозування їх рішень.

Ця інтеграція когнітивної науки забезпечує архітекторам безпеки більш надійний та довготривалий підхід до управління ризиками, оскільки, хоч вразливості мережі/програмного забезпечення еволюціонують з часом, принцип людської когніції та прийняття рішень залишається більш сталим. Тому когнітивні моделі є стійкішими до швидкої застарілості, на відміну від технологічних рішень, які фокусуються лише на конкретні мережеві вразливості.

Ще кращим і набагато ефективнішим є поєднання когнітивного моделювання та сценарного аналізу. Вони разом формують інтегральний методологічний комплекс, що критично важливий для управління складними, багаторівневими системами в умовах невизначеності. В цьому випадку

когнітивне моделювання забезпечує інструментальну основу, а саме формалізує та структурує якісні та експертні знання, перетворюючи їх у функціональний граф, що візуалізує системну логіку та впорядковує інформацію про цілі та дії. Сценарний аналіз, реалізований на цій основі, у свою чергу виконує роль динамічного імітаційного моделювання, прокручуючи різні варіанти розвитку подій з урахуванням динаміки усіх факторів.

Інтеграція когнітивного моделювання та сценарного аналізу має такі переваги, як подолання невизначеності, підвищення точності оцінки ризиків, врахування людського фактору у сценаріях інцидентів, обґрунтованість рішень та можливість виявлення слабких місць у процесах реагування.

Розділ 2 МЕТОДИ КОГНІТИВНОГО МОДЕЛЮВАННЯ СЦЕНАРІЇВ КІБЕРРИЗИКУ

2.1 Когнітивні карти та їх використання для моделювання загроз

У сучасних інформаційних системах, зокрема у кібербезпеці, надзвичайно важливо не лише фіксувати наявні загрози, а й розуміти механізм їх виникнення та розвитку, в чому дуже допомагають моделі, що мають здатність відображати багатофакторність, невизначеність, зворотні зв'язки та динаміку змін у системі. Одним із найбільш ефективних інструментом у цьому є когнітивні карти.

Вперше термін когнітивних карт був запропонований в психології американським психологом Е. Толменом у його роботі «Когнітивні карти у щурів і людини». Толмен давав визначення когнітивним картам як внутрішнє ментальне представлення навколишнього середовища, що формує основу для поведінки та прийняття рішень. Саме це заклало базу для майбутнього моделювання, роблячи акцент і підкреслюючи, що в основі складних систем лежать інтуїтивне розуміння та ментальні моделі.

Для того, щоб перетворити цю концепцію на формалізований інструмент системного аналізу був необхідний математичний апарат. В той час було адаптовано теорію графів когнітивних карт для вирішення соціально-економічних завдань. Запроваджено поняття зваженої когнітивної карти та імпульсного процесу, де ребра (тобто зв'язки) почали позначатись не лише знаками, що вказували на позитивний чи негативний вплив, а і числовими значеннями (вагами), що відображали суть цього впливу.

Ключова формалізація, яка була початком методу, який ми бачимо і активно використовуємо зараз, почалась тоді, коли у 1986 році Б. Коско запропонував нечіткі когнітивні карти (НКК, Fuzzy Cognitive Maps, FCMs) [29]. Нечіткі когнітивні карти поєднали в собі математичний формалізм теорії графів та теорію нечітких множин Л. Заде. В результаті чого було створено

метод, який дозволяє одночасно враховувати експертні думки та кількісні параметри системи. І це добре ілюструє фундаментальну здатність НКК, а саме структуроване захоплення та формалізація суб'єктивних експертних знань. І можливість перетворювати інтуїтивну думку на динамічну модель є критично важливою у контексті моделювання загроз, особливо нових або гібридних, де емпіричні дані поки обмежені.

Якщо давати більш чітке визначення, то нечіткі когнітивні карти (НКК) – це орієнтований граф, вузли якого представляють собою поняття, події або фактори, а ребра відображають вплив одного фактору на інший. Кожен зв'язок має величину (вагу), яка показує силу впливу, його напрям (позитивний чи негативний) та ступінь невизначеності взаємодії. Ваги зазвичай знаходяться в діапазоні [-1,1]. Позитивний знак (+) означає, що зростання в концепті А веде до зростання в концепті В, негативний знак (-) – що зростання в концепті А спричиняє зменшення в концепті В.

НКК схожі на нейронні мережі у деяких аспектах, але призначенні для представлення причинно-наслідкових зв'язків. Ось покроковий огляд того, як вони працюють:

Крок 1: Визначення змінних

На цьому кроці визначаються всі відповідні змінні в системі, тобто всі ключові компоненти, які впливають один на одного.

Крок 2: Визначення зв'язків

Після того як змінні було визначено, наступним кроком стає визначення як кожна змінна впливає на інші, тобто призначаються ваги ребрам (зв'язкам). І саме тут вже прослідковується нечітка логіка. Ці ваги коливаються в уже зазначеному діапазоні, і додатні значення (наприклад, +0,7) вказують на те, що зі збільшенням однієї змінної вона позитивно впливає на іншу змінну, від'ємні значення (наприклад, -0,2) вказують на зворотній зв'язок, а 0, що впливу немає.

Крок 3: Побудова нечіткої когнітивної карти

Тепер, на основі визначених зв’язків, створюється графічна карта, де вузли з’єднані ребрами (стрілками) і кожне з них позначене вагою. Побудована карта представляє собою візуально петлі зворотного зв’язку та взаємодії.

Крок 4: Виконання моделювання

Оскільки НКК є динамічними, то після побудови є можливість моделювання, щоб відслідкувати як з часом змінюється система при змінні певного параметру або декількох. Можна спостерігати за поведінкою системи коригуючи значення вхідних змінних (на основі гіпотетичних сценаріїв чи реальних даних).

Крок 5: Валідація та аналіз чутливості

На цьому етапі відбувається перехресна перевірка та порівняння з емпіричними даними незалежними експертами. Аналіз чутливості допомагає виявити ключові фактори ризику.

Приклад методу НКК з його компонентами зображено на рисунку 2.1.

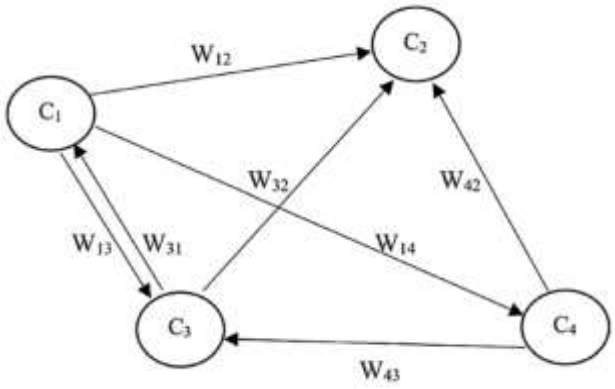


Рис. 2.1 Зразок НКК

де C_i виражає вузли або поняття, пов'язані з зваженими дугами. Кожен зв'язок між поняттями C_i та C_j має вагу W_{ij} (вага між вузлом і та вузлом j), яка може бути позитивною, негативною або нейтральною (вказує на те, що два поняття, що розглядаються, не мають зв'язку).

Тобто, вузли C_i, фактори системи, можуть бути, наприклад, рівень загрози, завантаженість операторів чи наявність контрзаходів. Ваги зв'язків W_{ij} – міра впливу одного фактору на інших в інтервалі від -1 до 1, а вектор станів показує поточні значення факторів у певний момент часу.

На відміну від звичайних когнітивних карт, ваги та значення НКК можуть бути нечіткими, тобто описуватись у певних діапазонах або ж взагалі лінгвістично (наприклад, «низький», «середній» чи «критичний рівень загрози»).

Окрім того, важливою перевагою нечітких когнітивних карт є те, що вони мають здатність моделювати зворотні зв'язки (feedback loops) та цикли [30]. На відміну від моделей, що мають більші обмеження, наприклад, орієнтовані ациклічні графи, які не можуть містити замкнених циклів, НКК можуть імітувати більш реальну динаміку складних систем [31]. Наприклад, у моделі загроз зростання атаки (концепт 1) може призвести до збою системи (концепт 2), але цей збій може викликати негативний зворотній зв'язок, впливаючи на підвищення рівню впровадження заходів безпеки (концепт 3), що, в свою чергу, зменшує початкову уразливість.

Таким чином, НКК дозволяє проводити динамічне моделювання. Динаміка НКК моделюється ітераційно, використовуючи процес, відомий як імпульсний процес або пряме виведення. Стан концепту C у наступний момент часу t+1, розраховується на основі його початкового стану, а також сумарного зваженого впливу від усіх інших його концептів, та зовнішніх входів (u). Розрахунок стану вузла C на момент часу t+1, розраховується за ітераційною формулою 2.1 [32]:

$$C(t+1) = f(C(t) + \sum_{j \neq i} C_j(t) W_{ji} + u_i) \qquad (2.1)$$

де f – нелінійна функція активації, яка використовується для нормалізації значення стану вузлів у діапазоні [-1,1], тим самим забезпечуючи стабільність моделювання.

Це дозволяє симулювати розвиток інцидентів у часі, перевіряти різні сценарії і у підсумку оцінювати ефективність контрзаходів.

Таким чином, ітераційний процес робить можливим для моделі прогноз патернів розвитку системи. Після кількох ітерацій система може збігатись до фіксованого стану (стан рівноваги) або увійти в граничний цикл, що дає змогу прогнозувати стійкі патерни розвитку системи чи траєкторії атаки.

Ключова математична перевага НКК, транзитивність причинного впливу, має вирішальне значення для динамічного аналізу, адже на відміну від ймовірнісних моделей в яких причинність не завжди транзитивна, НКК дозволяє відстежувати та сумувати сукупний вплив від початкових причин (наприклад, впровадження чогось нового) до кінцевого наслідку (наприклад, зниження ризику).

Нечіткі когнітивні карти мають здатність програвати широкий діапазон сценаріїв, включаючи зміни у політиці безпеки чи зміни у зовнішньому середовищі. Проте надійність НКК прямопропорційно залежить від якості знань, які надають експерти. Тобто, у нечітких когнітивних картах вузли та зв’язки будуються на основі того, які фактори експерти вважають важливими, того як експерти бачать причинно-наслідкові зв’язки, як оцінюють силу і напрям впливів ваг, а також від інтерпретації ступеня невизначеності цих зв’язків. Тобто НКК можуть відображати не об’єктивну реальність, а модель реальності в голові експертів. Для того, щоб набувати знання, необхідно використовувати структуровані методи, наприклад, напівструктуровані інтерв’ю [33]. Важливо поєднувати експертні оцінки з емпіричними даними, наприклад, використовуючи лог файли чи статистики інцидентів. Також для зниження суб’єктивності використовується метод злиття карт. Цей підхід передбачає злиття карт, створених різними експертами, шляхом зважування та додавання їхніх базових матриць суміжності. Об’єднана карта, зазвичай, є кращою репрезентацією загального доменного знання, особливо якщо вибірка експертів більш наближена до репрезентативної.

Сучасні розробки, в додачу до цього, передбачають інтеграцію НКК з алгоритмами машинного навчання. Важливо зазначити, що машинне навчання не створює карти повністю, а лише допомагає уточнювати ваги або додавати нові зв’язки на основі даних. На практиці це працює таким чином: спочатку експерт створює базову НКК, враховуючи концепти і попередні зв’язки, далі ML алгоритми отримують дані (наприклад, лог файли, інциденти, статистика), опісля ML коригує вагові коефіцієнти зв’язків, визначає сильні і слабкі впливи, в результаті чого модель стає більш наближеною до реальних процесів і містить менше суб’єктивної думки експерта. Перевагами такої інтеграції є зменшення суб’єктивності, те, що модель стає більш адаптивною і може ефективніше оновлюватись зі змінами середовища, а також з’являється можливість працювати з великими потоками даних. Обмеженнями у свою чергу є те, що ML не розуміє змісту причинності (може виявити кореляцію, проте не завжди гарантує логічність зв’язку), часто потрібні великі масиви даних, і те, що експерт все одно виконує верифікацію отриманих даних.

Як було вже сказано раніше, одне із основних обмежень НКК це суб’єктивність, проте також треба враховувати те, що вони демонструють велику чутливість до значень. Навіть невеликі зміни у вагах або початкових станах концептів можуть призводити до суттєво різних варіантів симуляції. Це вимагає ретельного налаштування та верифікації моделі. Окрім того, варто сказати про проблеми масштабованості, оскільки із зростанням складності систем та кількості змінних, управління та аналіз НКК стають все більш складними. Проте їхня здатність моделювати складні взаємозалежності та зворотні зв’язки робить нечіткі когнітивні карти цінним інструментом. Їх подальший розвиток, що має вирішити вищеописані проблеми, орієнтований, на нашу думку, на гібридизацію з методами машинного навчання.

2.2 Байєсівські мережі та їх застосування для оцінки невизначеності ризиків

Ще одним важливим інструментом у кількісному аналізі для моделювання та управління невизначеністю є байєсівські мережі (БМ), які дозволяють моделювати причинно-наслідкові зв’язки та оновлювати оцінки ризиків на основі цих даних.

Використання графічний структур для опису причинно-наслідкових залежностей почало формуватися ще на початку XX столітті. Одним із перших дослідників, який застосував орієнтовані графи для представлення причинних зв’язків між змінними був генетик Севолл Райт. У 1921 році він запропонував шляхові діаграми та обґрунтував метод аналізу шляхів, що використовувався для оцінки впливу одних факторів на інші [34-35]. Пізніше, аж у 1970-х роках, ці ідеї вже стали основою для методології моделювання структурних рівнянь, яка широко застосовувалась в соціальних та поведінкових науках. Саме ця графічна логіка згодом стала фундаментом для побудови орієнтованих ациклічних графів.

Справжній прорив у розвитку підходів до роботи з невизначеністю та причинністю стався завдяки Джуда Перлу, який сформував математичний апарат байєсівських мереж, поєднавши теорію ймовірностей із структурною моделлю залежностей між змінними [36]. Це дало змогу не просто фіксувати кореляції, а будувати моделі, здатні роботи обґрунтовані висновки, зокрема в умовах неповної інформації.

Вже подальший розвиток цих ідей стосувався вирішення проблеми причинності, яка тривалий час залишалась дискусійною. У 1990-х і 2000-х роках Перл, разом з іншими дослідниками, такими як Спітерс, Глаймор, Шейнс та інші [37], сформував формальну теорію причинних графів. Важливим кроком стало введення оператора do (·), який описує інтервенцію в систему, можливість не просто спостерігати, як змінні поводяться природнім чином, а моделювати наслідки цілеспрямованого втручання. Наприклад, не «що відбувається, коли змінюється X?», а «що станеться, якщо примусово встановити X на визначене значення?».

Ця концепція виявилась критично важливою для сфер, які вимагають проектування рішень, адже там недостатньо просто передбачати можливі сценарії, а потрібно ще й мати інструмент, який дозволяє обґрунтовано впливати на ситуацію, оцінюючи ефекти конкретних дій.

Таким чином, шлях від простих візуальних діаграм Райта до формальної теорії причинних графів Перла позначив перехід від описового аналізу до повноцінних причинно-орієнтованих обчислювальних моделей, здатних підтримувати прийняття рішень у реальних складних системах.

Байєсівська мережа формально представляє собою пару (G,B), де G це граф, що визначає якісну або структурну частину моделі, а B – таблиця умовних ймовірностей.

Таким чином, за своєю структурою БМ – це орієнтований ациклічний граф, вершини якого відповідають змінним домену (наприклад, тип загрози або стан системи), а ребра (спрямовані зв’язки) описують причинно-наслідкові залежності між ними.

На рисунку 2.2 можна побачити приклад графу.

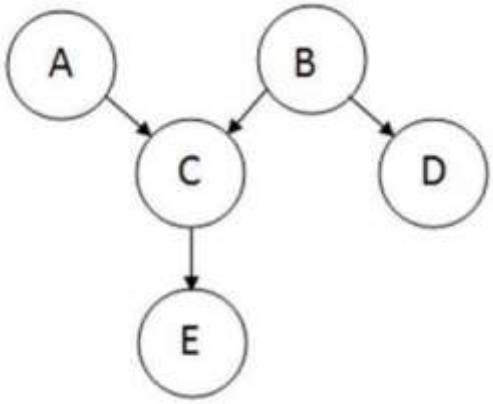


Рис.2.2 спрямований ациклічний граф

Як можна побачити на зображені, вершини графа (вузли), що представляють змінні, зображені колами, що містять у собі назву змінної. Зв’язки між ними називають дугами або ребрами, і малюються стрілками між вузлами, ілюструючи залежність між змінними. Окрім того, для будь-якої пари вузлів є батьківський вузол. Вузол де починається дуга називається батьківським, а той, де закінчується, дочірнім. Наприклад, на зображені 2

вузол B є батьківським, а вузол C – дочірнім. Окрім того, якщо вузли можуть бути досягнуті з інших вузлів, їх також називають нащадками, тобто для вже показаного вузла B, вузли C і E також можна назвати нащадками, і навпаки вузол B є предком для C і E.

Кожен вузол має таблицю умовної ймовірності (CPT), яка визначає ймовірнісний розподіл величини залежно від її батьківських вузлів. Якщо ж вузол не має батьків, то його ймовірнісний розподіл є маргінальним, тобто визначається безумовно. Простіше кажучи, якщо у вузла немає батьків, то на нього нічого не впливає в моделі і його значення не залежить від інших змінних. Тобто він описується сам по собі, незалежно, і тоді ймовірність події є теж сама по собі, без урахування інших факторів.

Основною математичного апарату БМ є теорія умовної ймовірності та теорема Байєса.

Умовна ймовірність P(A|B) визначає, наскільки є ймовірним настання події A, якщо подія B вже відбулась. Цей тип ймовірності є фундаментальним у Байєсівському аналізі, де головним завданням є оновлення впевненості щодо певних гіпотез після отримання нових даних.

Теорема Байєса формалізує процес перегляду початкових (апріорних) ймовірностей гіпотез у світлі зібраних спостережень. Якщо задані початкові ймовірності гіпотез $P(H_k)$ та ймовірність появи спостережуваного результату A за умови істинності кожної гіпотези $P(\mathrm{A}|H_k)$, то оновлені ймовірності визначаються виразом:

$$P(H_k|A) = \frac{P(A|H_k) \cdot P(H_k)}{P(A)} \qquad (2.2)$$

де P(A) – загальна ймовірність появи спостереження A, яка розраховується як сума ймовірностей цього спостереження за всіма можливими гіпотезами:

$$P(A) = \sum_k P(A|H_k)P(H_k) \qquad (2.3)$$

Суть байєсівського підходу полягає в тому, що він дозволяє не просто робити прогноз на основі даних, а систематично оновлювати рівень впевненості щодо причинних або прихованих змінних у міру надходження нової інформації. Іншими словами, ми переходимо від «передбачення на основі припущень» до «передбачення на основі реальних доказів, що накопичуються».

Банальним прикладом, що це ілюструє можна уявити як випадок, коли ми маємо гіпотезу «сьогодні буде дощ». Наша початкова впевненість (апріорна ймовірність) в цьому 30%. Опісля ми бачимо «свідоцтво»: сильні темні хмари. Байєсівський підхід допомагає математично визначити як змінюється впевненість у початковій гіпотезі, коли з’являються нові дані. Після оновлення, наприклад, можна сказати про ймовірність 80%, і це вже апостеріорна ймовірність.

Окрім цього, однією з найважливіших властивостей байєсівських мереж є можливість представлення складних систем залежностей між змінними в компактній формі. Це стає можливим завдяки принципу локальної умовної незалежності, який вбудований у структуру орієнтованого ациклічного графа (DAG). Згідно з цим принципом, кожна змінна в мережі вважається залежною лише від своїх безпосередніх батьківських вузлів.

Тому спільний розподіл ймовірностей для набору випадкових величин $X = \{X_1, X_2, \dots X_n\}$ можна розкласти у вигляді:

$$P(X_1, X_2, \dots X_n) = \prod_{i=1}^{n} P(X_i | Parents(X_i)) \qquad (2.4)$$

На відміну від загального підходу, де для повного опису спільного розподілу необхідно вираховувати всі залежності між змінними (що може швидко стати обчислювально непідйомним), БМ дозволяє оперувати лише тими залежностями, які є структурно обґрунтованими. Це означає, що замість зберігання величезної кількості параметрів, модель працює з відносно малими

таблицями умовних ймовірностей, де кожен вузол розглядається у контексті лише своїх найближчих породжуючих факторів. У результаті зменшується:

- обсяг пам’яті, потрібний для зберігання моделі;
- час, потрібний на обчислення ймовірностей;
- складність моделювання нових сценарії.

Фактично, декомпозиція спільного розподілу є тим механізмом, який робить БМ практичними для реальних систем, зокрема в аналізі ризиків, кібербезпеці, фінансах і медичних діагностичних системах.

Проте важливо зауважити, що для змінних з дискретними станами обсяг CTP зростає експоненційно зі збільшенням кількості батьків або кількості можливих станів у цих вузлах, і це явище відоме як «вибух CTP». Також це є одним із основних обмежень масштабування байєсівських мереж для складних систем, адже через це часто при побудові моделей аналітикам може доводитись штучно обмежувати число батьківських вузлів або зменшувати деталізацію наборів станів для того, щоб модель залишалась обчислювальною. Якщо уявити кожен вузол БМ як питання з варіантами відповідей, то щоб визначити відповідь потрібно знати стан інших питань від яких воно залежить. І коли одна змінна залежить від двох чи трьох це є цілком нормальним та легко обчислювальним, але коли вона починає залежати від десятка та більше і кожна має по 3-4 можливі значення – кількість можливих комбінація суттєво зростає і стає вже більш складною для обчислень. Наприклад, якщо кожна змінна має 4 стани, і в неї є 5 батьківських вузлів, CPT матиме 1024 можливих комбінацій, які треба заповнити ймовірностями. Ось чому CPT «вибухає». Тому на практиці або зменшують кількість зв’язків між вузлами, або агрегують дані, або застосовують гібридні методи (наприклад, машинне навчання (ML), що допомагає з автоматичними обчисленнями).

Байєсівські мережі не обмежуються роботою лише з дискретними змінними. Поряд також можуть використовуватись неперервні та гібридні моделі. У такому випадку традиційні таблиці умовних ймовірностей для

неперервних вузлів замінюються параметричними умовними розподілами. Зазвичай, вони описуються через математичні характеристики, такі як середнє значення, дисперсія або коефіцієнти регресійних залежностей, що визначаються окремо для кожної комбінації значень дискретних батьківських вузлів. Завдяки цьому підходу стає можливо безпосередньо моделювати безперервні величини, наприклад, фінансові індикатори (курси або відсоткові ставки), технічні параметри системи (температуру, навантаження мережі чи час реакції) або поведінкові метрики. Це дозволяє уникнути дискретизації, тобто штучного розподілу неперервних значень на категорії, яке зазвичай призводить до втрати точності чи спотворення вихідних даних. Таким чином, застосування неперервних або гібридних БМ забезпечує більш гнучке та реалістичне моделювання складних систем, в яких важливо спрогнозувати як ймовірнісні залежності, так і числові закономірності процесів, шляхом того, що мережа може моделювати ці дані прямо, без округлення чи поділу на категорії.

Формально умовні залежності та незалежності в БМ визначаються за допомогою концепції d-separation [38]. Вона встановлює, коли набір змінних здатний «блокувати» або «розблоковувати» передачу ймовірнісної інформації між іншими змінними.

d-separation визначає умовну незалежність, аналізуючи шляхи між змінними і тип структури вздовж цих шляхів:

1. Ланцюг $X \rightarrow N \rightarrow Y$: Якщо вузол N відомий, то X і Y стають умовно незалежними, оскільки N вже пояснює зв’язок.

Наприклад, X – кількість часу, який школяр навчався, N – результат вступного іспиту, Y – ймовірність вступу в університет. Поки ми не знаємо результатів тесту (N), інформація X впливає на наше припущення щодо вступу (Y), але коли результат стає відомим, то більше немає сенсу дивитись на час навчання. Отже X і Y стають умовно незалежними, якщо відоме N.

2. Спільна причина X←N→Y: Спостереження вузла N «відрізає» кореляцію між його наслідками, роблячи X та Y незалежними.

Наприклад, хмарність (X) ← дощ (N) → мокрий асфальт (Y). Оскільки ми вже знаємо, що йшов дощ, то хмарність та мокрий асфальт більше не накладаються одне на одне.

3. Спільний ефект X→N←Y: Навпаки якщо N або його нащадки спостерігаються, то X та Y стають залежними.

Наприклад, наполегливість (X)→успіх (N)←талант (Y). Якщо відомо, що людина досягла успіху, але не має таланту, то це підвищує ймовірність, що успіх пояснюється наполегливістю, або навпаки. Це явище також називають ефектом пояснення, коли спостереження ефекту робить причини взаємозалежними.

Таблиця підсумку правила d-separation наведена нижче:

Таблиця 2.1

Правила d-separation (умовне блокування шляхів)

Структура	Коли незалежні	Коли залежні	Пояснення
Ланцюг X→N→Y	Якщо N відоме	Якщо N невідоме	Інформація про проміжний вузол N пояснює вплив X на Y, блокує шлях поширення ймовірності
Спільна причина X←N→Y	Якщо N відоме	Якщо N невідоме	Відомий стан спільної причини N виключає спостережувану кореляцію між її наслідками
Спільний ефект X→N←Y	Якщо N невідоме	Якщо N відоме	Спостереження за спільним ефектом N створює умовну залежність між причинами X та Y, «ефект пояснення»

На практиці d-separation дає:

1. Визначення умовної незалежності: менше обчислень, простіші моделі;
2. Правильне тлумачення причинності: надійніші рішення, менше помилкових висновків;
3. Контроль поширення спостереження: швидке оновлення ймовірностей;
4. Модульність та масштабованість: можливість працювати з великими системами;
5. Оптимізація збору даних: економія ресурсів і підвищення точності.

Таким чином, d-separation є формальним, графовим критерієм, який визначає, коли множина спостережених змінних перешкоджає поширенню ймовірнісної інформації між парами змінних у байєсівській мережі. Воно лежить в основі розуміння як спостереження змінює (або не змінює) взаємозв'язки між змінними і є ключовим інструментом для коректного побудування та інференції в DAG-моделях.

Важливо зауважити, що традиційні (статичні) байєсівські мережі описують систему у вигляді фіксованої структури залежностей між змінними в один момент часу. Однак у багатьох реальних задачах стан системи змінюється, і ці зміни мають часовий характер. У таких випадках необхідно також враховувати динаміку переходів між станами, а не лише статичний розподіл.

Динамічні байєсівські мережі (ДБМ) розширюють класичні БМ, моделюючи схему як послідовність часових зрізів [39]. У цій структурі залежності існують як всередині одного моменту часу, так і між змінними в моменті часу t та t+1, що дозволяє описувати еволюцію системи у вигляді стохастичних процесів.

Модель загалом будується на припущенні марковості першого порядку:

$$P(X^{t+1}|X^t, X^{t-1}, \dots) = P(X^{t+1}|X^t) \quad (2.5)$$

Тобто майбутній стан залежить лише від поточного. Це суттєво спрощує модель, оскільки зменшує кількість необхідних зв'язків між змінними в різні моменти часу. На практиці більшість ДБМ використовують саме перший порядок маркова, оскільки він забезпечує баланс між точністю моделювання та обчислювальною ефективністю.

На основі цього принципу встановлюється стандартна архітектура ДБМ, відома як 2-Slice Temporal Bayesian Network (2-TBN) [40]. У такій моделі весь часовий процес представляється у вигляді повторюваних часових зрізів однакової структури. Фактично, ДБМ, визначена на часовому проміжку від 1 до T, є послідовністю однакових двослайсових мереж, з'єднаних між собою ребрами, що описують причинно-наслідкові залежності.

2-TBN складається з двох компонентів:

1)модель початкового стану M_1, яка визначає розподіл змінних у момент часу t=1;

2)перехідна модель $M_{\rightarrow}$, яка описує ймовірнісний зв'язок між станами в сусідніх моментах часу.

У більшості застосувань також робиться припущення про часову стаціонарність. Це означає, що перехідна модель не змінюється протягом усього інтервалу часу.

У динамічних байєсівських мережах структура 2-TBN визначає, яким саме чином стан системи в момент часу t залежить від стану в попередній момент t-1. У межах 2-TBN розглядають два типи зв'язків між змінними:

1. Міжзрізові ребра: це орієнтовані ребра, що з'єднують змінні попереднього часового зрізу X_{t-1} зі змінними у поточному зрізі X_t. Вони описують динаміку переходів станів і реалізують часову причинність моделі. Саме виявлення цих міжчасових залежностей є ключовим завданням структурного навчання ДБМ, оскільки воно дозволяє визначити, які саме фактори мають вплив на розвиток системи у часі.

2. Внутрішньозрізові ребра: ці ребра встановлюють залежності між змінними всередині одного й того ж часового зрізу X_t. Існують реалізації ДБМ, у яких внутрішньозрізові зв'язки навмисно не включаються. Таке обмеження спрощує структурного навчання, оскільки модель фокусується виключно на часових залежностях. Проте, відсутність внутрішньозрізових ребер має наслідок: модель не може відобразити миттєві причинно-наслідкові взаємодії, які виникають без часової затримки (наприклад, одночасний вплив фізіологічного параметра на інший у межах того ж моменту часу) [40].

Таким чином, структурне навчання в ДБМ відбувається подібно до навчання статичних БМ, але доповнюється етапом виявлення темпоральних залежностей між часовими зрізами.

ДБМ особливо корисні у ситуаціях, де часові залежності ускладнені прихованими (латентними) станами, які не спостерігаються напряму. У таких випадках у модель вводять латентні змінні, що дозволяє виявляти приховані режими поведінки системи, описувати перехід між цими режимами та виконувати прогнозування з урахуванням невизначеності.

Приховані (латентні) стани це характеристики системи, які впливають на її поведінку, але ми не можемо їх побачити прямо. Наприклад, стрес у людини (ми можемо спостерігати його видимі ознаки як сонливість, зниження продуктивності, підвищений пульс, але не сам стрес як число) чи рівень зношення технічного обладнання (видимими є лише індикації, як зміна шуму чи температури). Це все непрямо спостережувані причини, але вони визначають, як система змінюється з часом. Тому додають приховану змінну H^t, яка представляє стан системи в моменту часу t і визначає поведінку спостережуваних величин.

Включення часової компоненти розв'язує ключове обмеження статичних підходів у ризик-аналізі. Статична модель дозволяє визначити P(ризик), тобто ймовірність ризику взагалі. Натомість динамічна модель дає можливість оцінити P(ризик у t+1|поточний стан у t), тобто як зміниться ризик

на наступному кроці з урахуванням поточного контексту. Це дає змогу прогнозувати коли ризик стане критичним, моделювати сценарії розвитку подій, своєчасно виявляти аномалії та зміну поведінкових режимів та підвищувати точність рішень у системах підтримки прийняття рішень.

2.3 Агентно-орієнтовані моделі поведінки користувачів та зловмисників

Агентно-орієнтоване моделювання (АОМ) є сучасною методологією для дослідження складних динамічних систем, зокрема у сфері кібербезпеки. Основною особливість цього підходу є орієнтація на індивідуальні компоненти системи – агенти – та їхню взаємодію.

Іншими словами, агентно-орієнтоване моделювання – це підхід, який описує складні системи як сукупність автономних агентів, які взаємодіють за простими правилами. У контексті кібербезпеки АОМ використовується для відтворення поведінки двох класів акторів: легітимних користувачів (наприклад, персонал, клієнти, автоматизовані сервери) та зловмисників (хакери, інсайдери, автоматизовані боти). Його завдання зрозуміти, як мікроповедінка агентів породжує макропатерни загроз, вразливостей та інцидентів, а також оцінити ефективність заходів захисту у реалістичних сценаріях.

Однією з ключових причин застосування агентно-орієнтованого моделювання в кібербезпеці є його здатність відтворювати складні системи, де різні учасники мають власні цілі, рівень знань, стратегії та реакції. Це робить АОМ особливо актуальним у випадках, коли поведінка системи не може бути описана лише на основі технічних параметрів. Підсумовуючи, основними перевагами є:

1. Гетерогенність середовища та поведінки: у реальних системах користувачі та зловмисники суттєво відрізняються за рівнем обізнаності, намірами, мотивацією та технічною підготовкою. АОМ

дозволяє моделювати на рівнях абстракції, де модель може бути охарактеризована від низького рівня, що охоплює конкретні вразливості та індивідуальну поведінку зловмисника, і до високого рівня, який включає наміри найвищого рівня зловмисника. Таким чином, це також дає змогу охопити весь життєвий цикл кібератаки.

2. Сценарне тестування та симуляції інцидентів: AOM дозволяє створювати і програвати альтернативні сценарії атак та захисту без ризику для реальної інфраструктури у так званих «пісочницях». Це можуть бути наприклад, відтворення атак між координованою діяльністю ботнетів та роботою компонентів захисної інфраструктури, включно з системами виявлення вторгнень (IDS), засобами фільтрації трафіку та політиками доступу. Це цінно для аналізу ризиків, підготовки персоналу, розробки стратегій реагування, адже в результаті дослідники отримують не лише статичні характеристики загроз, але й поведінкову динаміку протиборства, оцінюючи, як захисні механізми реагуватимуть на зміни темпу, масштабу чи структури атаки.
3. Нелінійні процеси та емерджентні ефекти: навіть дуже прості локальні правила поведінки можуть призвести до непередбачуваних наслідків на рівні системи. Наприклад, одиничний успішний фішинг може спричинити ланцюгову компрометацію цілої мережі, що інколи важко спрогнозувати класичними моделями.
4. Підтримка розробки політик безпеки: модель може використовуватись для тестування ефективності нових процедур, інструкцій або автоматизованих механізмів реагування до їх впровадження.

Особливо важливим також є те, що AOM дозволяє моделювати не лише окремі технічні дії зловмисника, а й його стратегічні наміри, що має вирішальне значення для аналізу складних, довготривалих та цілеспрямованих атак типу Advanced Persistent Threat (APT), де дії атакуючого визначаються не випадковістю, а чіткою метою, наприклад, отриманням доступу до

конфіденційної інформації або поступовим порушенням роботи системи. Включення мотиваційного рівня у модель дозволяє відтворювати логіку прийняття рішень зловмисника у процесі атаки, а не лише її технічну реалізацію. Це, у свою чергу, надає можливість передбачати майбутні кроки атакуючого агента та оцінювати ризики ескалації, що суттєво перевищує можливості традиційних реактивних підходів, орієнтованих лише на фіксацію фактів проникнення.

Основними елементами моделі у свою чергу є:

1) агенти, кожен тип якого має власну логіку дій та інформаційні можливості;
2) середовище, яким може бути корпоративна мережа з визначеною топологією, набір серверів та сервісів, політики доступу, інструменти автентифікації чи системи журналювання.
3) правила поведінки, які окреслюють, як агент реагує на події, взаємодії або зміни умов; правила можуть бути стохастичними, евристичними або ґрунтуватись на алгоритмах машинного навчання.
4) комунікаційні механізми, тобто канали взаємодії між агентами, що включають як технічні протоколи (TCP/IP, HTTP,SSH), так і соціальні (електронні листи, корпоративні чати), таким чином дозволяючи моделювати як соціотехнічні атаки, так і суто технічні.
5) симуляційна платформа, в якій часовий механізм симуляції може бути дискретним (по кроках) або подіє-орієнтованим, і передбачає фіксацію подій для подальшого аналізу.

Говорячи детальніше про агента у даному випадку – це автономний об’єкт з атрибутами (стан, роль, ресурси) і набором правил поведінки (рішення, реакції, навчання). Він може навчатись, змінювати поведінку залежно від досвіду і адаптуватись до змін у середовищі. Агентами можуть виступати як люди (наприклад, корпоративний користувач, адміністратор мережі, аналітик, зловмисник), так і технічні елементи (сервер, мережевий

вузол, IoT-пристрій). Кожен агент функціонує у певному оточенні (інфраструктура, політики доступу, протоколи взаємодії) і впливає на інші елементи системи через комунікацію та дії.

Ключовими атрибутами для інтелектуального агента є:

- автономність: здатність діяти без зовнішнього втручання, самостійно;
- реактивність: здатність сприймати зміни в зовнішньому середовищі та реагувати на них;
- соціальність: здатність взаємодіяти та вести комунікацію з іншими агентами;
- проактивність: здатність бути орієнтованим на досягнення власних цілей.

Взаємодія між окремими учасниками системи є ключовою характеристикою мультиагентних підходів. Саме наявність комунікації та впливу агентів один на одного відрізняє мультиагентні системи від звичайних ізольованих моделей. Агент у такій системі не діє сам по собі, а формує зв’язки з іншими агентами, обмінюється інформацією, визначає стратегії поведінки та адаптується у разі потреби. Такі зв’язки можуть бути як парними, так і груповими, стабільними або динамічними, що надає системі здатність до самоорганізації. Завдяки цим процесам, взаємодія виконує не тільки роль комунікаційного механізму, але й стає чинником еволюційних перетворень, адже поведінка одного агента може змінюватись під впливом дій інших. Інакше кажучи, на рівні мультиагентних систем (MAC) можуть виникати нові властивості та ефекти, які не можна передбачити, аналізуючи окремих агентів ізольовано.

У моделюванні поведінки агентів правила визначають, як агент приймає рішення та реагує на зміни в середовищі. Рівень формалізації правил сильно впливає на реалістичність симуляції і на її потреби в даних. Основними підходами є ймовірнісні правила, евристики та агенти, які навчаються.

Правила на основі ймовірностей та порогів є найпростішим підходом, де подія X відбувається з ймовірністю p, яка може залежати від атрибуту агента і стану середовища. Це дозволяє задати стохастичну (часто біноміальну) поведінку без складних логік. Приклад, математично може бути таким, де нехай подія X – дія агента (умовно, клікнути на підозрілий лист), тоді

$$P(X|\text{контекст}) = f(a) \quad (2.6)$$

де a – вектор характеристик (роль агента, рівень навчання, наявність фільтрів, час доби, тощо), а f – скалярна функція.

Умовно, користувач з базовим навчанням відкриває фішинговий лист з ймовірністю 0,25, а просунутий – 0,05. Атака сканування портів успішна з ймовірністю 0,6 за відсутності IDS, та 0,1 при активному.

Перевагами такого правила є простота, невелика потреба в даних та очевидна інтерпретованість. Недоліками є обмежена адаптивність, статичність ймовірностей та складність відображення складних послідовних стратегій.

Роль евристик та правил прийняття рішень, корисні на етапі швидкої побудови моделі, при дефіциті даних або коли потрібно зберегти інтерпретованість моделі. Вони можуть бути застосовані як стартові політики в навчанні агентів.

Прикладами можуть бути, якщо лист має вкладення і відправник не у списку контактів ймовірність відкриття нижча, ніж за інших факторів, умовно 0,2.

Перевагами у такому випадку є швидке впровадження, зрозумілі правила та полегшення перевірки. Мінусами є ризик надмірної простоти, те, що тоді вони можуть не відображати складну адаптивну поведінку, а також давати помилкові сигнали при зміні контексту.

Також замість фіксованих правил агент використовує алгоритми навчання (часто – підкріплювальне навчання, RL), щоб самостійно відшукувати стратегії, що максимізують певну мету (наприклад, ефективність атаки, зменшення ризику для захисника).

Прикладами такого застосування зі сторони зловмисника може бути, коли RL-агент вчиться підбирати вектори атак, якій найімовірніше призведуть до ескалації прав доступу та мінімального виявлення. Або фішинговий тренінг, коли симульований користувач з адаптивною поведінкою змінює шанси «клікнути» після навчання.

Варіантами алгоритмів у такому випадку можуть бути табличні методи (Q-learning), для простих дискретних просторів; метод політики, для складних або безперервних дій; мультиагенти RL, для випадків, де одночасно навчаються кілька агентів з різними цілями.

Переваги тут у тому, що агенти адаптуються до динаміки середовища, можуть виявляти складні політики та мають еволюцію тактик. Виклики полягають у великій потребі в даних/симуляціях, проблемі стабільності під час навчання, ризику небажаних «вигаданих» стратегій, потреба у коректно визначених винагородах.

Зараз існують вже також і ШІ-агенти, які здатні без втручання людини виконувати завдання. Для того, щоб виконувати поставлені задачі, вони мають бути оснащені набором функціональних модулів, а саме:

1. Модуль пам’яті, який забезпечує збереження та відновлення інформації про попередні події та накопичений досвід агента. Його зазвичай розбивають на короткочасну (про поточний контекст взаємодії: останні події, поточні цілі, тимчасові зміни; необхідний для негайних рішень) та довготривалу пам’ять (містить сталу інформацію, наприклад, про профіль дій, історію успішних та невдалих сценаріїв, правила, вивчені політики). Пам’ять дозволяє моделювати такі явища, як, наприклад, повторна експлуатація однакової вразливості, накопичення індикаторів компрометації або навчання на попередніх невдалих атаках.

2. Модуль профілю, який описує постійні або напівпостійні атрибути агента, такі як роль (наприклад, користувач, адмін, ботнет-вузол), права й ресурси (доступи, привілеї, обчислювальні можливості), мотивації та

цілі (наприклад, фінансова вигода, саботаж, збори даних) та поведінкові установки (ризикова схильність, обережність, прагнення до прихованості). Цей профіль формує ідентичність агента і слугує базою для вибору стратегій та моделей прийняття рішень.

3. Модуль планування є центральним компонентом, який формує стратегії поведінки. Сучасні архітектури можуть використовувати техніки «chain-of-thought» або ReAct (розуміння → дія) для генерації внутрішніх логічних міркувань, що дозволяє агенту будувати послідовні аргументовані плани.
4. Модуль виконання дій відповідає за приведення вже визначених планів у дію, наприклад, відправлення мережевих запитів у симуляторі, створення фішингового листа, виконання привілеєвих операцій або запуск захисних процедур (оновлення конфігурації, блокування IP). Цей модуль з'являється з API симуляційного середовища і реєструє результати виконання (успіх, помилка, час, побічні ефекти).
5. Модуль оркестрації, бо у складних мультиагентних системах часто присутній спеціальний компонент, що координує взаємодію між автономними мікроагентами. Оркестратор у свою чергу визначає розподіл ролей і черговість виконання підзадач, забезпечує контроль над складними, багатоетапними операціями та агрегує результати від підлеглих агентів і приймає рішення про наступні кроки. Оркестрація зменшує складність реалізації комплексних стратегій, дозволяючи декомпозувати великі цілі на керовані підзадачі.

Слід також дотримуватися наступних практичних рекомендацій.

1. Подумати про рівень деталізації, адже проектуючи агента, потрібно знайти компроміс між реалізмом та придатністю до обчислень (надто детальна модель ускладнює валідацію й пришвидшує симуляцію, занадто абстрактна у свою чергу втрачає інформативність);

2. Зауважити на прозорість та валідацію. Важливо зберігати пояснюваність рішень агента (особливо в захисних сценаріях), рекомендується поєднувати символічні правила з ML, щоб забезпечити інтерпретованість;
3. Враховувати етичні та правові аспекти, оскільки моделювання поведінки людей потребує уважного ставлення до приватності, збереження анонімності деяких джерел даних. Якщо модель використовується для тестування соціоінженерних атак, то необхідно встановлювати чіткі етичні рамки та дозволи;
4. А також варто поєднувати координацію з автономністю локальних агентів, оскільки агент-оркестратор корисний для відтворення складних, людино-орієнтованих кампаній, але надмірна централізація може зменшити реалістичність моделі.

Набір модулів, описаний вище, в додачу з зауваженнями та рекомендаціями, створює повноцінного ШІ-агента, здатного моделювати широкий спектр поведінки в кіберпросторі – від звичайного необізнаного користувача до цілеспрямованого та адаптивного зловмисника.

Агентно-орієнтоване моделювання забезпечує необхідну ієрархію абстракції для моделювання поведінки, що критично важливо для вивчення складних загроз.

На операційному (низькому) рівні: моделі фокусуються на точних, повторюваних діях агентів. Це може включати моделювання взаємодії мережевих пакетів, використання конкретних уразливостей або опис функціональності ботів за допомогою кінцевих автоматів (FSM). Наприклад, FSM може описувати послідовність кроків, які виконує ботнет-агент для ініціації DDoS-атаки.

На тактичному (середньому) рівні здійснюється взаємодія, координація та реакція. Агенти захисту (IDS, брандмауери) тут моделюються як системи, навчені виконувати певні перевірки на різних рівнях мережевої моделі

(системи, мережі, глобальної мережі) та звітувати про виявлення шкідливої активності.

На стратегічному (високому) рівні концентруються наміри та стратегії. Він вимагає онтології верхнього рівня кібер-атакуючого. Дослідження показують, що більшість атак засновані на намірах, що вимагає моделювання відповідної причини для атаки. Моделювання на цьому рівні дозволяє зрозуміти, чому зловмисник обрав саме цю ціль та послідовність дій, що є ключовим для прогнозування поведінки складних APT-груп [41].

Можливість «спускатись» від стратегічних мотивів атакуючого (на високому рівні) до тактичних рішень, і, нарешті, до операційних кроків (на низькому рівні) – експлуатації конкретної вразливості – дозволяє будувати цілісні симуляції, які можуть обробляти безліч різних типів сценаріїв, включно з тими, які або неможливо або небажано реплікувати в реальному середовищі.

В таблиці 2.2 наведено підсумок ієрархії рівнів моделювання поведінки агентів у кіберпросторі:

Таблиця 2.2

Ієрархія рівнів моделювання поведінки агентів

Рівень моделювання	Об’єкт моделі	Типова формалізація	Приклад застосування
Стратегічний (високий)	Наміри та цілі, онтологія атакування	Когнітивні моделі (BDI)	Моделювання APT-кампаній, стратегії кібершпигунства
Тактичний (середній)	Взаємодія, координація, реакція захисту	Мультиагентні протоколи, LLM-планування	Симуляція DDoS-атак ботнетів, IDS/брандмауерів
Операційний (низький)	Індивідуальні дії, уразливості, пакети	Кінцеві автомати (FSM), евристики	Реакція на конкретну уразливість, моделювання ботів

Якість моделі залежить від достовірності параметрів – ймовірностей, порогів, ваг, винагород.

Джерелами даних для налаштування моделі можуть слугувати:

1. Журнали SIEM/IDS/Firewall: логи доступу, виявлення атак, блокування з IP/таймстепами. Використання для оцінки частотності подій, часових шаблонів, кореляцій;
2. Результати фішинг-тестів та навчань: дані про відсоток успішних клікнувших до/після навчання, час реакції. Використання для встановлення початкових ймовірностей ризик-поведінки користувачів;
3. Пентести та threat intelligence: тактики, методи, індикатори компрометації. Використання для побудови сценаріїв атак, визначення можливих векторів;
4. Опитування та інтерв'ю співробітників: дані про поведінку, рівень обізнаності. Використання для параметризації профілю користувача;
5. Дані про інциденти, що вже стались: реальні кейси з докладними кроками інциденту. Використовується для валідації сценаріїв та навчання динамічних агентів.

Як методи отримання знань та оцінки параметрів використовуються експертне опитування, статистичне навчання з логів, гібридні підходи та вивчення через симуляції.

Експертні опитування широко використовуються, коли емпіричних даних недостатньо або вони не відображають потрібних контекстів. До найпоширеніших підходів належать структуровані інтерв'ю, метод Дельфі та ієрархічний аналіз. У структурованих інтерв'ю експертів просять описати ймовірності подій, причинно-наслідкові залежності або пріоритети загроз у чітко заданій формі. Метод Дельфі додає кілька раундів анонімної оцінки з подальшим узагальненням і поверненням до агрегованого результату до повторної оцінки. Ієрархічний аналіз дозволяє декомпозувати проблеми на ієрархію критеріїв і провести парні порівняння для отримання відносних ваг.

Головною перевагою експертних методів є можливість кодування якісних доменних знань там, де даних немає, або швидке отримання початкових оцінок для прототипних моделей. Водночас їхні слабкі місця це

суб’єктивність оцінок, схильність до когнітивних упереджень, необхідність ретельної агрегації та квантифікації невизначеності (наприклад, через інтервали довіри або розподіли). Тому при застосуванні експертних оцінок важливо документувати процедуру опитування, кількість експертів, критерії відбору та підходи до агрегування (зважені середі, методи комбінування розподілів), а також здійснювати перевірку сумісності оцінок.

Статистичне навчання з логів та інших даних передбачає, що якщо доступні кількісні дані (журнали SIEM/IDS, мережеві логи, результати пентестів, історія інцидентів), то параметри моделей можна оцінювати статистично. Стандартні підходи включають частотні оцінки, метод максимальної правдоподібності та байєсівську оцінку. Практична робота починається з підготовки даних, куди входить очищення, кореляція подій, часове вирівнювання записів, анотація, а також усунення артефактів (даних, які не відображають справжніх подій системи, а являються побічними продуктами роботи інструменту моніторингу або специфічних налаштувань). Статистичні методи дають більш об’єктивні, відтворювані оцінки та дозволяють витягти часові залежності або переходи станів. Однак, такі підходи вимагають великої кількості якісних даних, якщо логи неповні або марковані неточно, то оцінки можуть бути упередженими.

Гібридні стратегії є поєднанням використання експертних знань та навчання з даних, і використовується найчастіше. В такому випадку спочатку експерти задають початкові припущення та апріорні розподіли, а потім ці розподіли уточнюються на підставі наявних даних. Такий порядок робить модель більш стійкою, бо експерти прагматично компенсують дефіцит даних, а статистика запобігає надмірній суб’єктивності. На практиці це може виглядати як спочатку задання розподілів для ваг НКК або CPT у байєсівській мережі на основі опитування, потім використання логів для обчислення правдоподібності і у підсумку отримання апостеріорних оцінок параметрів. Перевагою гібридних підходів є краща інтерпретованість і здатність перетворювати розпливчасті або неточні знання на вимірювані числа.

Недоліками у свою чергу може бути підвищена складність реалізації та потреба в контролі узгодженості між даними й експертними судженнями.

Симуляційне навчання та генерація даних (для RL та «віртуального експерименту») може використовуватись, коли потрібно навчити адаптивних агентів або перевірити наслідки різних параметрів, за допомогою використання симуляторів. Тоді створюють віртуальне середовище з завідомо контрольованою структурою, де агент навчається або генерує траєкторії поведінки. Результати таких тренувань стають даними для подальшого калібрування емпіричних моделей. Важливо зауважити, що чим ближче симулятор до реальних умов, тим корисніші згенеровані дані. Окрім того, бувають випадки коли поведінка, яка ефективна в симуляторі, може виявитись неефективною в реальному середовищі, що називається проблемою переносу. Це може статись, наприклад, через неповному моделі середовища, неправильну статистику вхідних даних, чутливість політик або через невраховані взаємодії з людьми. Для того, щоб зменшити цей розрив можуть використовувати техніки типу domain randomization чи adversarial training.

Перша полягає в тому, щоб замість побудови одного «точного» симулятора, при навчанні політики створюються багато варіантів середовища, випадково варіюючи параметри симуляції (наприклад, затримки, рівень шуму, ймовірність помилок, конфігурації мережі, характеристики користувачів). Таким чином агент навчається в множині різнорідних симуляторів і шукає стратегії, які у подальшому будуть стабільно працювати в широкому діапазоні умов. Наприклад, при моделюванні фішингових кампаній можна змінювати такі параметри як частота розсилок, форма листа, рівень технічної грамотності користувача; при тренуванні агента – змінювати шаблони трафіку, затримки, частоту логів. Проте це все одно не може гарантувати покриття всіх реальних умов і може вимагати дуже великої кількості симуляцій та обчислювальних ресурсів, а також важливо правильно обирати інтервали варіацій, бо якщо вони занадто широкі, то агент не навчиться нічому корисному, якщо занадто вузькі – слабкий ефект.

Ідея адверсаріального тренінгу полягає в тому, щоб під час навчання політики мають протистояти найгіршим або найшкідливішим варіантам середовища. Іноді це реалізується як гра двох агентів, наприклад, захисник навчається в присутності адаптивного атакуючого агента, що спеціально шукає слабкі місця. В такому випадку атакуючий агент змінює умови та вхідні дані так, щоб максимізувати помилки захисника, а той у свою чергу має адаптуватись. У результаті знаходяться та усуваються вразливі місця політик, вона стає стійкішою до цілеспрямованих атак або несподіваних сценаріїв. Недоліками такого тренування може бути те, що це може бути складним з точки зору стабільності навчання, а також те, що вони вимагають налаштування та балансу сил між агентами. Існує ризик, що атака, оптимізована проти конкретного захисника в симуляторі, не відображатиме реальні креативні стратегії супротивника.

Також валідація та верифікація є обов'язковими для забезпечення достовірності, гнучкості та надійності АОМ.

Верифікація спрямована на перевірку того, чи правильно реалізований задум моделі, тобто чи відповідає програмна або формальна реалізація початковій концепції. На цьому етапі зосереджуються не на тому чи збігається модель з реальністю, а на тому, чи коректно вона побудована з точки зору внутрішньої логіки. Це включає в себе:

1) перевірку повноти поведінкових сценаріїв, адже модель повинна охоплювати всі можливі рішення та стани агентів, інакше частина поведінки виявиться непромодельованою;
2) виявлення надмірних суперечностей, що можуть спотворювати результати симуляції;
3) аналіз властивостей безпеки та стабільності, щоб бачити чи залишаються агенти в допустимих станах при непередбачуваних подіях;
4) експертну оцінку знань, закладених у модель.

Методики, на кшталт, Usage Case Maps (UCM) або поведінкових діаграм допомагають візуалізувати взаємодію агентів, що в свою чергу полегшує виявлення логічних помилок, дублювань або відсутніх переходів.

UCM – це метод формального опису поведінки системи та взаємодії її компонентів у вигляді сценаріїв, які відображають, як саме відбувається процес виконання дій у часі. Оскільки в АОМ важливо не лише визначити, що може робити агент, але й як його дії вписані в загальну поведінку системи, то UCM також показують як взаємодія між агентами організована в часі, які події запускають реакції, які альтернативні сценарії можуть виникати (наприклад, атака вдається чи не вдається) і де знаходяться критичні точки (вразливості, місця контролю). Це в свою чергу, як і зазначено вище, допомагає виявляти логічні помилки ще до програмної реалізації моделі.

Валідація перевіряє, наскільки модель достовірно відображає реальний світ. Якщо верифікація відповідає на питання «чи правильно побудована модель?», то валідація відповідає на «чи відображає модель реальну систему». На цьому етапі модель порівнюють з реальними статистичними даними, поведінковими патернами користувачів, відомими сценаріями атак та результатами експериментів і пентестів.

Калібрування – це регулювання параметрів моделі (наприклад, імовірностей, сил впливу, вагових коефіцієнтів), щоб модель поводилась максимально наближено до спостережуваної реальності. Цей етап особливо актуальний для систем, де поведінка агентів не повністю детермінована, тобто містить елементи випадковості та адаптації. Соціальні та кібернетичні системи часто демонструють стохастичні та нелінійні властивості, тому одразу підібрати точні параметри майже неможливо і їх доводиться уточнювати на основі емпіричних даних.

АОМ добре інтегрується в інші підходи, такі як когнітивні моделі, байєсівські та динамічні мережі, а також НКК. Наприклад, комбінування АОМ із когнітивними моделями дозволяє значно підвищити правдоподібність поведінки агентів. Завдяки можливості моделювати процеси прийняття

рішень, увагу, пам’ять і когнітивні упередження, агенти стають ближчими до реальних користувачів або зловмисників. БМ та ДБМ можуть виступати як модулі, що оцінюють ймовірність виникнення певних подій залежно від контексту. Вбудовані в агентну модель, вони дозволяють агентам приймати рішення на основі накопиченої інформації та оцінок ризику у реальному часу. І НКК у свою чергу дозволяють якісно моделювати взаємозалежність факторів, що формують поведінку агентів, наприклад, рівень довіри, відчуття ризику, мотивацію до дотримання політик. Це дозволяє точно налаштовувати поведінкові правила без надмірної жорстокості моделей.

Говорячи про переваги агентно-орієнтованого моделювання можна виділити високий ступінь реалізму за рахунок моделювання взаємодії між агентами, гнучкість у формуванні сценаріїв, включаючи рідкісні чи гіпотетичні інциденти, можливість проведення аналізу типу «що буде якщо…» без ризику для реальної системи, а також практичну користь для навчання персоналу та підтримки прийняття рішень.

В якості обмежень варто сказати про врахування якісних даних та експертної калібровки для якісної роботи моделі, ризик надмірної складності і не повної прозорості при використанні само-навчальних агентів, валідація моделей може бути складною через нефіксовану або емерджентну природу результатів (результати часто не є фіксованими, тобто при повторних симуляціях навіть за однакових початкових умов, система може демонструвати різні виходи), а також високі обчислювальні витрати при великій кількості агентів або складних сценаріях.

2.4 Використання штучного інтелекту для автоматизації когнітивного моделювання

Застосування штучного інтелекту дозволяє автоматизувати процес побудови, навчання та оптимізації когнітивних моделей, що в свою чергу значно підвищує швидкість розробки та точність симуляцій складних

когнітивних процесів. У сучасних дослідженнях ШІ виступає не лише як інструмент обчислення, а як активний компонент когнітивних систем, який здатний адаптуватись до нових умов.

Одним із ключових напрямів автоматизації є використання машинного навчання (ML) для виявлення патернів у поведінці користувачів, прогнозування реакцій або формування рішень на основі історичних даних. Наприклад, нейронні мережі здатні імітувати роботи пам’яті та механізмів уваги, наближаючи модель до реальних когнітивних процесів.

Також, згадане вище, підкріплювальне навчання (RL) дозволяє агентам навчатись через взаємодію із середовищем, де вони отримують нагороду або покарання за певні дії, поступово оптимізуючи власну стратегію. Це дає змогу моделювати процеси прийняття рішень, притаманні як людині, так і складним кіберсистемам.

Крім того, технології обробки природної мови (NLP) використовуються для відтворення мовного мислення та аналізу семантичних зв’язків у комунікації агентів. У когнітивному моделюванні це відкриває можливості для дослідження діалогових систем, інтерпретації намірів користувачів або зловмисників у кіберпросторі.

Створення достовірних та гнучких когнітивних моделей вимагає подолання обмежень традиційних підходів, а саме суто символічного та суто нейронного. Нейро-символічний ШІ поєднує в собі переваги обох шляхом об’єднання глибокого навчання та здатності виявляти закономірності, з символічними методами (наприклад, здатність міркувати, узагальнювати). Ця ідея також узгоджується зі згаданою вище моделлю подвійного процесу когніції, яка розрізняє дві взаємодоповнювальні системи мислення: швидке, інтуїтивне й емоційне, яке відображає роботу глибинних нейронних мереж, що здатні розпізнавати образи і приймати рішення на основі великих обсягів даних, та повільне, свідоме, логічне й аналітичне, яке відповідає символьному рівню обробки інформації, де відбувається планування, дедукція та перевірка гіпотез.

Для побудови надійний когнітивних систем потрібна гібридна архітектура, що поєднує в собі ці два рівні мислення. Вона повинна спиратись на попередні знання, логічні правила та гнучкі механізми навчання, щоб дозволити системі не лише аналізувати, а і пояснювати власні рішення.

Одним із центральних елементів цих процесів на сьогоднішній день є автоматизація абдукції, тобто перехід від спостереження до формування пояснювальної гіпотези. Особливу роль в цьому напрямку відіграють великі мовні моделі (ВММ), які здатні не лише аналізувати великі обсяги текстових або числових даних, а й пропонувати нові гіпотези, які покращують точність прогнозів у складних завданнях класифікації, діагностики чи поведінкового аналізу. Крім того, ВММ можуть виконувати функцію аналітичного компасу, виявляючи прогалини у знаннях, демонструючи, де існуючі моделі не справляються з передбаченням або поясненням даних [42]. Це дозволяє дослідникам ефективніше спрямовувати ресурси, концентруючись на сферах з найменшою науковою визначеністю.

Перевагами використання штучного інтелекту є забезпечення більшої масштабованості, адаптивності та прогностичності когнітивних моделей. В такому разі системи здатні самостійно пристосовуватись до нових умов, моделювати непередбачувану поведінку, вирішувати проблеми обсягу обчислень, суб’єктивності та ефективно працювати в умовах невизначеності.

Водночас існують і виклики. По-перше, автоматизовані моделі потребують великої кількості якісних даних. По-друге, алгоритми глибокого навчання часто мають низьку інтерпретованість, що ускладнює їх використання в критичних системах. По-третє, виникає етичний аспект, питання довіри до автономних рішень, контроль за їх діями та запобігання потенційним зловживанням.

Висновки до другого розділу

Нечіткі когнітивні карти (НКК) становлять інтелектуальний та гнучкий підхід до моделювання. Вони забезпечують ефективний динамічний аналіз,

дають змогу працювати з якісною невизначеністю та формалізувати експертні знання.

Говорячи про НКК, то там застосування штучного інтелекту підвищує точність та гнучкість. Методи еволюційного оптимізування використовуються для автоматичного налаштування ваг зв’язків на основі емпіричних даних. Крім того, нейронні мережі можуть навчатись прогнозувати зміни у вузлах карти, фактично виконуючи роль динамічного оновлення в режимі реального часу.

Байєсівські мережі відзначаються своєю здатністю інтегрувати експертні знання, працювати з неповними або обмеженими даними та обчислювати апостеріорні ймовірності, що робить їх ще одним дуже цінним інструментом. Використання динамічних БМ ще більше підсилює їх можливості, дозволяючи переходити від статичних оцінок до активного прогнозування.

Втім, практична реалізація таких систем вимагає уважного врахування обчислюваних обмежень, наприклад проблеми експозиційного зростання таблиць умовних ймовірностей, що веде до так званого «вибуху CPT». Окрім цього, експерти з ризиків мають зважати на компроміс між швидкістю обчислень і можливою похибкою при оцінюванні малоймовірних подій.

Традиційна побудова БМ передбачає ручне визначення структури та ймовірнісних розподілів, але штучний інтелект робить цей процес значно швидше та ефективніше, а також може вирішувати проблему складності та величини обчислень.

Третім розглянутим не менш ефективним інструментом є агентно-орієнтоване моделювання (АОМ). Завдяки йому можливо перейти від пасивних, сигнатурних систем виявлення до активного, поведінкового аналізу загроз, моделюючи складну динаміку взаємодії між користувачами, захисними системами та зловмисними агентами. Такий підхід дозволяє виявити не лише окремі інциденти, а й системні патерни, що формують макрорівень кіберризиків.

Використання штучного інтелекту (ШІ) у когнітивному моделюванні значно підвищує можливості кожного з цих підходів. Алгоритми машинного навчання, нейронні мережі та еволюційні методи дозволяють автоматизувати побудову моделей, оптимізувати параметри, виявляти приховані закономірності та робити прогнози з вищою точністю. Інтеграцію ШІ з НКК, БМ і АОМ сприяє переходу від статичних експертних моделей до динамічних, адаптивних когнітивних систем, здатних самостійно вчитись та приймати рішення в умовах невизначеності.

Розділ 3 РОЗРОБКА ТА ЗАСТОСУВАННЯ КОГНІТИВНИХ СЦЕНАРІЇВ У СИСТЕМАХ УПРАВЛІННЯ РИЗИКАМИ КІБЕРБЕЗПЕКИ

3.1 Формування сценаріїв кіберінцидентів на основі когнітивних моделей

Як практична реалізація формування сценаріїв кіберіцидентів на основі когнітивних моделей у цій роботі було використано НКК для сценарію фішингу.

Як початковий етап необхідно сформувати постановку задачі та подумати про збір даних. Нашою метою у даному випадку є моделювання сценарію фішингової атаки на організацію. Для цього збираються дані про типові показники, а саме частота надходження фішингових листів, ймовірність того, що співробітник відкриє його (залежить від навчання і обізнаності), наявність фільтрів спаму тощо. На основі наявних даних та думок експертів визначаються ключові концепти (вузли моделі).

Визначені концепти в нашому випадку:

- C1: довіра до відправника. Рівень довіри користувача до адресата листа (зазвичай збільшується, якщо лист від знайомої особи);
- C2: фільтрація електронної пошти. Ефективність спам-/фішинг-фільтра, що зменшує кількість зловмисних листів, які потрапляють у поштову скриньку;
- C3: MFA (багатофакторна аутентифікація). Наявність або відсутність додаткового рівня захисту при вході;
- C4: безпекова обізнаність користувача. Рівень навчання та підозрілості користувача;
- C5: доставка фішингового листа. Факт того, що фішинговий лист успішно надійшов на пошту користувача;

- C6: відкриття шкідливого вкладення або посилання. Дія користувача відкрити прикріплений файл або перейти за фішинговим посиланням;
- C7: компрометація системи. Успішна атака, де зловмисник отримує доступ до облікового запису або виконує код у системі жертви;
- C8: латеральне поширення. Поширення шкідливого програмного забезпечення або користувацьких облікових даних мережею організації після первинного проникнення;
- C9: швидкість реагування. Оцінка наскільки швидко команда реагує на підозрілі події;
- C10: складність фішингової атаки. Відображення того наскільки фішингова кампанії персоналізована, правдоподібна і технічно обхідна;
- C11: дотримання політик безпеки. Наприклад, використання політик доступу, політик оновлень;
- C12: частота та якість тренінгів і навчань для персоналу. Проведення заходів спрямованих на навчання користувачів;
- C13: рівень здібностей зловмисника. Опис технічних ресурсів, навичок, мотивації.

У НКК кожен концепт являються змінними моделі, а їхня взаємодія формалізується ваговими коефіцієнтами на ребрах. У наведеному прикладі концепти охоплюють впливи соціальної інженерії, технічного захисту та дій користувача.

Далі ми створюємо матрицю зв'язків НКК, яку можна уявити як таблицю ваг w_{ij}, де кожен елемент позначає силу впливу концепту C_i на концепт C_j. Вагові коефіцієнти задаються в діапазоні [-1;1] і також можуть бути описані лінгвістично, наприклад, «слабкий», «помірний», «сильний», «критичний/дуже сильний».

В нашому випадку значеннями ваг можуть бути:

1) C1 → C6, вага +0,8 (сильний позитивний вплив), адже висока довіра стимулює відкриття вкладення;

2) C2 → C5, вага -0,9 (дуже сильний негативний вплив), адже ефективний фільтр майже повністю блокує доставку фішингового листа;

3) C5 → C6, вага +0,6 (помірно позитивний вплив), адже якщо лист доставлено і користувач або довіряє, або необізнаний, зростає ймовірність відкриття листа;

4) C4 → C1, вага -0,7 (сильно негативний вплив), адже висока обізнаність зменшує довіру до підозрілих повідомлень;

5) C4 → C6, вага -0,5 (помірно негативна), адже обізнаний користувач рідше відкриває невідомі повідомлення;

6) C6 → C7, вага +0,9 (дуже сильний позитивний вплив), адже якщо вкладення відкрите, майже гарантована настає компрометація (особливо за відсутності MFA);

7) C3 → C7, вага -0,8 (сильний негативний), адже активована MFA суттєво знижує ймовірність компрометації (навіть при відкритті вкладення);

8) C7 → C8, вага +0,8 (сильний позитивний вплив), адже компрометація одного комп'ютера призводить до розповсюдження атаки мережею;

9) C10 → C1, вага +0,6 (помірно позитивний вплив), адже більш персоналізовані листи, які містять, наприклад, ім'я/роль/контекст, підвищують ступінь довіри;

10) C10 → C2, вага -0,5 (помірно негативний вплив), адже більш витончені техніки можуть обходити фільтри, отже знижувати їх ефективність;

11) C9 → C8, вага -0,7 (високий негативний вплив), адже швидке реагування може зупинити розгортання атаки і зменшити можливість тривалої компрометації;

12) C7 → C9, вага +0,5 (помірно позитивний вплив), адже поширення атаки може прискорити швидкість реагування;

13) C11 → C3, вага +0,7 (високий позитивний вплив), адже політика стимулює впровадження MFA;

14) C11 → C2, вага +0,5 (помірний позитивний вплив), адже політики визначають тонкі налаштування;

15) C12 → C4, вага +0,8 (високий позитивний вплив), адже навчання напряму впливає на обізнаність;

16) C12 → C1 та C12 → C6, вага -0,5 (помірний негативний вплив), адже тренінги знижують «наївну» довіру до отриманих листів та вкладень, а також їх відкриття;

17) C13 → C10, вага +0,9 (дуже високий позитивний вплив), адже чим більша обізнаність і навички зловмисника, тим правдоподібніша і складніша атака.

Звичайно, це не враховує абсолютно усі можливі варіанти, проте в рамках навчальної роботи наведена більш спрощена модель з найбільш поширеними варіантами сценаріїв. Таким чином, матриця ваг кодує припущення про причинно-наслідкові зв'язки. Така матриця є основою для подальшого аналізу та імітації сценаріїв.

Матриця представлена в таблиці 3.1.

Таблиця 3.1

Матриця зв'язків ваг

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**
C1	0	0	0	0	0	+0,8	0	0	0	0	0	0	0
C2	0	0	0	0	-0,9	0	0	0	0	0	0	0	0
C3	0	0	0	0	0	0	-0,8	0	0	0	0	0	0
C4	-0,7	0	0	0	0	-0,5	0	0	0	0	0	0	0
C5	0	0	0	0	0	+0,6	0	0	0	0	0	0	0
C6	0	0	0	0	0	0	+0,9	0	0	0	0	0	0
C7	0	0	0	0	0	0	0	+0,8	+0,5	0	0	0	0
C8	0	0	0	0	0	0	0	0	0	0	0	0	0
C9	0	0	0	0	0	0	0	-0,7	0	0	0	0	0
C10	+0,6	-0,5	0	0	0	0	0	0	0	0	0	0	0
C11	0	+0,5	+0,7	0	0	0	0	0	0	0	0	0	0
C12	-0,5	0	0	+0,8	0	-0,5	0	0	0	0	0	0	0
C13	0	0	0	0	0	0	0	0	0	+0,9	0	0	0

Створена нечітка когнітивна карта за заданими параметрами зображена на рисунку 3.1.

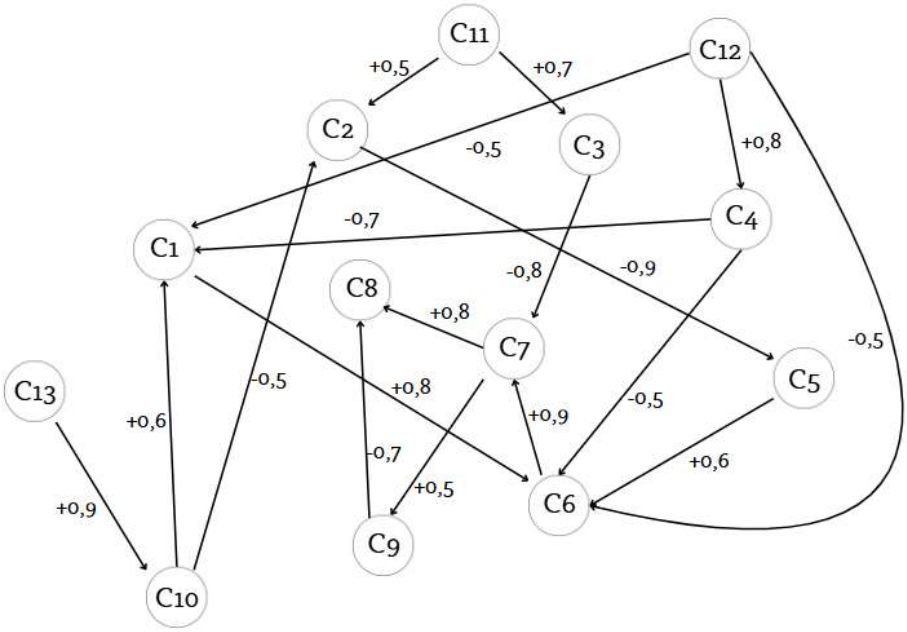


Рис.3.1 НКК для фішингового сценарію

Таким чином було показано різні можливі сценарії в залежності від визначених концептів. Наприклад, на зображені чітко видно, що зі зростанням довіри (С1) результат відкриття шкідливого вкладення або посилання (С6) стрімко зростає.

Окрім того ми можемо зробити динаміку сценарію в часі. Розглянемо умовну послідовність подій:

1. Розвиток довіри міг здійснитись через соціальну інженерію і контекст С1 зростає, з тим самим посилюючи у майбутньому концепт С6.
2. Користувач отримав лист. У початковий момент користувач отримує фішинговий лист (концепт С5 стає активним). Якщо С2 (фільтрація) слабкий, то лист доходить. Довіра до відправника (С1) може бути вже високою, наприклад 0,9.
3. Відкриття вкладення. За високих С1 і С5 концепт С6 активується і його значення зростає. Це ключовий момент атаки.

4. Якщо це стається, за умови неактивної MFA (концепт C3), то по відкриттю вкладення приходить в дію концепт C7 (компрометація системи) і зловмисник отримує доступ. Значення C7 стрімко зростає.
5. Після цього, якщо немає швидкої реакції (концепт C9), відбувається поширення загрози і зростання концепту C8, шкідливе ПЗ рухається по мережі.

Таким чином значення концептів змінюється за ланцюжком «C1↑ → C6 ↑ → C7↑ → C8↑» тобто «зростання довіри → відкриття фішингового листа → компрометація системи → поширення». На кожному етапі обчислюються нові активності концептів за базовою формулою НКК (оновлення активацій на основі суми зважених впливів попереднього кроку). Якщо на якомусь етапі ключовий концепт (наприклад, C6) залишається низьким або нульовим, подальші етапи можуть не будуватись. Таким чином модель може показувати динаміку інциденту в часі.

Окрім цього, модель також дозволяє аналізувати альтернативні гілки подій в залежності від зміни окремих концептів. Прикладами альтернативних сценаріїв можуть бути:

1. Активована MFA (C3), якщо ввімкнена багатофакторна аутентифікація це не виключає ризик, проте може суттєвого його знизити. Тобто вплив C3 на C7 значний і негативний (-0,8) і є вірогідність, що це сценарій зупиниться на цьому етапі.

2. Висока обізнаність користувачів (C4), адже за підвищеного рівня безпеки користувач менше довіряє незнайомим листам, що впливає на концепт C1, і в свою чергу на пряму на C6. Таким чином користувач може проігнорувати лист і не дійти до етапу компрометації.
3. Якщо є ефективна фільтрація (високий концепт C2), то доставка лист (C5) матиме низьку активацію, є висока вірогідність, що лист просто не потрапить до поштової скриньки користувача і подальші події не відбудуться.

Таким чином нечітка когнітивна карта дозволяє швидко модифікувати початкові умови «що якщо» і аналізувати наслідки цих змін. Зміна значення одного концепту породжує нові гілки розвитку подій і демонструє ефективність захисних заходів.

Загалом, описаний приклад демонструє процес формування сценарію фішинг атаки з використанням НКК: від визначення концептів та їхніх взаємозв'язків до динамічного аналізу альтернативних варіантів перебігу інциденту.

Детальніше про розрахунки оцінки ризику представлено в пункті 3.2.

3.2 Оцінка ризиків на базі сценарного аналізу з урахуванням когнітивних факторів

Для розрахунків стан концептів змінюється ітеративно відповідно до правила:

$$C_j(t+1) = f(C_j(t) + \sum_{i=1} C_i(t) W_{ji}) \quad (3.1)$$

Де $C_j(t) \in (0,1)$ – активація концепту j в ітерації t, $f(\cdot)$ – функція активації (в моделі сигмоїда $f(x) = \frac{1}{1+e^{-x}}$), що обмежується значеннями в (0,1).

Початкові значення C(0) задалися раніше експертно на основі даних (наприклад, рівень довіри в середньому, ефективність фільтру тощо). Для сценарного аналізу вводяться модифікації цих початкових значень.

Сценарії визначаються як набір початкових змін (наприклад, «фішинговий лист доставлено», «атакуючий має високий рівень навичок» тощо). Значення концептів C7 (компрометація) та C8 (латеральне поширення) є основними вихідними метриками.

Для моделювання і проведення симуляцій було використано програму FCM Expert. FCM Expert – це спеціалізоване програмне забезпечення призначене для створення та симуляцій нечітких когнітивних карт. Модель вже побудованої початкової НКК в FCM Expert зображена на рис 3.2.

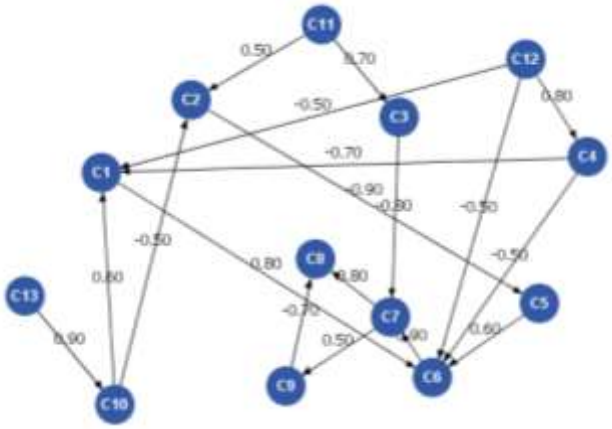


Рис. 3.2 НКК за початковими параметрами

Як попередні дані задані такі значення: $C(0) = [C1 = 0{,}4; C2 = 0{,}7; C3 = 0{,}5; C4 = 0{,}5; C5 = 0{,}1; C6 = 0{,}05;\ C7 = 0{,}02; C8 = 0{,}01; C9 = 0{,}5; C10 = 0{,}3; C11 = 0{,}6; C12 = 0{,}4; C13 = 0{,}3)$. У сценаріях нижче будуть змінюватись компоненти, які розглядаються, а інші залишатись базовими. Приклад процесу задання значення для вузлів наведено на рис 3.3.

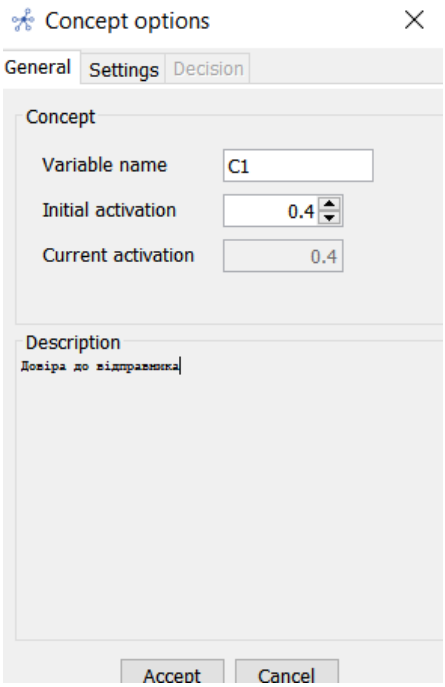


Рис. 3.3 Налаштування вузла C1 (довіра до відправника)

Налаштування параметрів симуляції показано на рис. 3.4

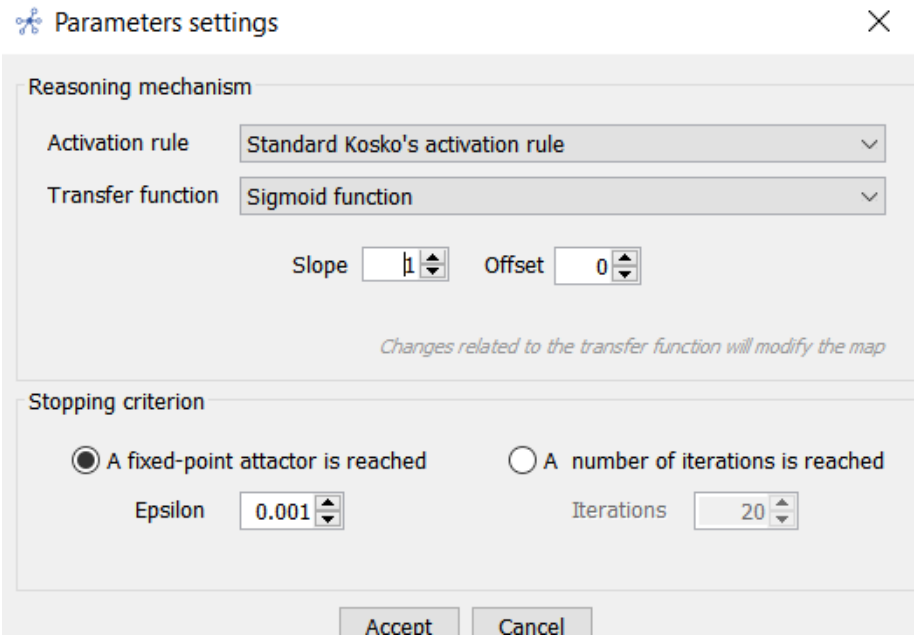


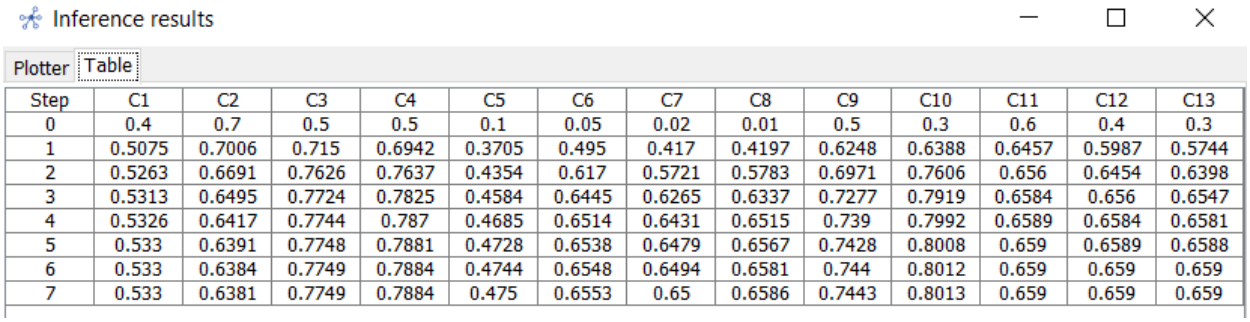
Рис. 3.4 Параметри симуляції FCM Expert

Як правило активації вибрано стандартне правило Коско, тому що воно враховує всі фактори через ваги і є стандартним правилом для когнітивних карт. Функція активації типовою та рекомендованою для НКК є саме сигмоїда, адже вона моделює когнітивну природу сприйняття ризику. Тобто нечіткі когнітивні карти імітують людське мислення (як користувачі оцінюють ризик, формують довіру, реагують на небезпеку, приймають рішення під фішинговою атакою в нашому випадку), і людські реакції нелінійні, що можна відтворити за допомогою сигмоїди. Вона також забезпечує, що значення лишались у фізично можливому діапазоні, була можлива інтерпретація результатів. Slope – 1, тому це забезпечує помірну чутливість, плавну реакцію вузлів на зміни факторів. Як параметр зупинки симуляції вибрано пункт «a fixed-point attactor is reached (epsilon 0.001)», бо так симуляція зупиняється в стабільному стані, ми отримуємо точні фінальні ризики.

У цій роботі будуть представлені розрахунки для ключових концептів, що формують траєкторію атаки: C12 → C4 і C10 → C1 і C5 → C6 → C7 → C8, тобто як тренінги впливають на обізнаність, як вона і складність атаки у свою чергу впливають на довіру, як довіра і доставка листа впливають на його

відкриття, як відкриття впливає на компрометацію і як компрометація впливає на латеральне поширення).

Запускаючи симуляцію на початковому етапі, не змінюючи нічого ми отримуємо результат, зображений на рисунку 3.5.



Step	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
0	0.4	0.7	0.5	0.5	0.1	0.05	0.02	0.01	0.5	0.3	0.6	0.4	0.3
1	0.5075	0.7006	0.715	0.6942	0.3705	0.495	0.417	0.4197	0.6248	0.6388	0.6457	0.5987	0.5744
2	0.5263	0.6691	0.7626	0.7637	0.4354	0.617	0.5721	0.5783	0.6971	0.7606	0.656	0.6454	0.6398
3	0.5313	0.6495	0.7724	0.7825	0.4584	0.6445	0.6265	0.6337	0.7277	0.7919	0.6584	0.656	0.6547
4	0.5326	0.6417	0.7744	0.787	0.4685	0.6514	0.6431	0.6515	0.739	0.7992	0.6589	0.6584	0.6581
5	0.533	0.6391	0.7748	0.7881	0.4728	0.6538	0.6479	0.6567	0.7428	0.8008	0.659	0.6589	0.6588
6	0.533	0.6384	0.7749	0.7884	0.4744	0.6548	0.6494	0.6581	0.744	0.8012	0.659	0.659	0.659
7	0.533	0.6381	0.7749	0.7884	0.475	0.6553	0.65	0.6586	0.7443	0.8013	0.659	0.659	0.659

Рис.3.5 Результат симуляції з початковими даними заданими експертами

При симуляції НКК з початковими значеннями всіх концептів спостерігається поступове збільшення ключових показників атак та захисних факторів у системі. На початковому кроці значення довіри до відправника (C1 = 0,4), обізнаності користувача (C4 = 0,5) та відкриття листа (C6 = 0,05) є низькими, що відображає стартовий стан перед активною фішинговою атакою. Після запуску симуляції система починає еволюціонувати відповідно до вагових зв’язків між вузлами. Уже на першому кроці видно суттєві зміни, більшість концептів зростають, оскільки на них впливають інші вузли через позитивні або негативні ваги. Подальші кроки показують зростання більшості показників, але темп змін поступово сповільнюється. Це свідчить про природній процес згладжування динаміки, коли взаємні впливи поступово урівноважуються. На пізніх кроках симуляції система досягає стабілізації, де значення концептів перестають істотно змінюватись і різниця між кроками мінімальна. Це означає, що модель досягла атрактора, тобто точки, у якій усі причинно-наслідкові впливи збалансовані, і подальша еволюція системи не змінює її поведінку.

Для першого досліджуваного сценарію ми змінимо лише один концепт – регулярність проведення тренінгів (C12) для того, щоб дослідити як навчання користувачів впливає на фішингову атаку, а саме на довіру до листа,

ймовірність його відкриття та успішність атаки (компрометація системи та поширення).

В такому випадку ми змінюємо концепт C12 на максимальне значення, 1. Результати нової симуляції представлені на рисунку 3.6.

Inference results

Plotter Table

Step	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
0	0.4	0.7	0.5	0.5	0.1	0.05	0.02	0.01	0.5	0.3	0.6	1.0	0.3
1	0.4329	0.7006	0.715	0.7858	0.3705	0.4207	0.417	0.4197	0.6248	0.6388	0.6457	0.7311	0.5744
2	0.4752	0.6691	0.7626	0.7975	0.4354	0.5575	0.5557	0.5783	0.6971	0.7606	0.656	0.675	0.6398
3	0.5089	0.6495	0.7724	0.7921	0.4584	0.6136	0.61	0.6306	0.7261	0.7919	0.6584	0.6626	0.6547
4	0.5246	0.6417	0.7744	0.7895	0.4685	0.6384	0.6328	0.648	0.7371	0.7992	0.6589	0.6599	0.6581
5	0.5303	0.6391	0.7748	0.7888	0.4728	0.649	0.6429	0.6544	0.7415	0.8008	0.659	0.6592	0.6588
6	0.5322	0.6384	0.7749	0.7885	0.4744	0.6531	0.6473	0.6569	0.7432	0.8012	0.659	0.6591	0.659
7	0.5328	0.6381	0.7749	0.7885	0.475	0.6547	0.6491	0.658	0.744	0.8013	0.659	0.6591	0.659
8	0.533	0.6381	0.7749	0.7885	0.4752	0.6552	0.6499	0.6585	0.7443	0.8013	0.659	0.659	0.659

Рис. 3.6 Результат симуляції при зміні концепту C12

Графік динаміки при симуляції з C12=1 показано на рисунку 3.7.

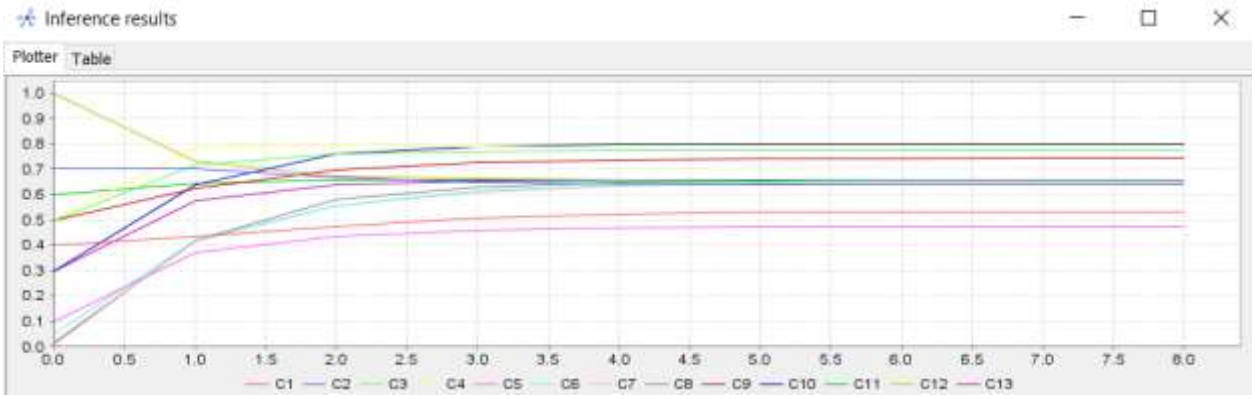


Рис. 3.7 Графік динаміку симуляції з C12=1

Порівняння отриманих даних зі стартовою симуляцією показує, що збільшення C12 викликає помітні зміни вже на перших і середніх ітераціях. Уже на першому кроці добре видно, що C4 (обізнаність) та C6 (ймовірність відкриття) реагують спадом, це добре ілюструє, що тренінги підсилюють користувацьку уважність та критичність до листів. Як результат, на ранніх етапах C1 (довіра до відправника) також знижується, що зменшує ймовірність неправильних рішень з боку користувача.

Далі, ці зміни передаються по ланцюгу впливів, і ми можемо бачити, що нижчі показники довіри й відкриття листа зменшують ризик компрометації (C7), а отже на проміжних кроках спостерігається уповільнення росту C7 та C8 (латерального поширення). Це відповідає логіці того, що чим краще

навчений користувач, тим менша ймовірність того, що атака просунеться від етапу доставки до етапу внутрішнього проникнення.

Водночас на фінальних кроках двох симуляцій (C12 = 0,4 та C12 = 1) вони обидві сходяться до дуже близьких значень. Це очікуваний результат для нечіткої когнітивної карти, оскільки модель працює як динамічна система, що прямує до атрактора, тобто стану рівноваги в якому всі взаємні впливи вузлів компенсуються. У цій точці система більше не реагує на зовнішні зміни, і результати стабілізуються незалежно від початкових умов. Але хоч кінцеві значення концептів дійсно стають однаковими (через математичну природу моделі), зміни на ранніх та середніх кроках чітко демонструють вплив користувачів на ключові етапи атаки, і ми можемо зробити висновок, що регулярні якісні тренінги знижують довіру до листа, обмежують ймовірність його відкриття, зменшують ризик компрометації та гальмують подальше поширення загрози. Саме ця динаміка показує, що навчання є ефективним превентивним заходом, навіть якщо система з часом рухається до стабільного стану.

У наступному сценарії ми залишаємо вже отримані дані, але змінюємо пункти, які впливають на саму фішингову атаку та майстерність її реалізації, а саме рівень знань зловмисника (C13 стає 0,8) та складність фішингової атаки (C10 стає також 0,8). Проводимо нову симуляцію результат якої показаний на рисунках 3.8-3.9.

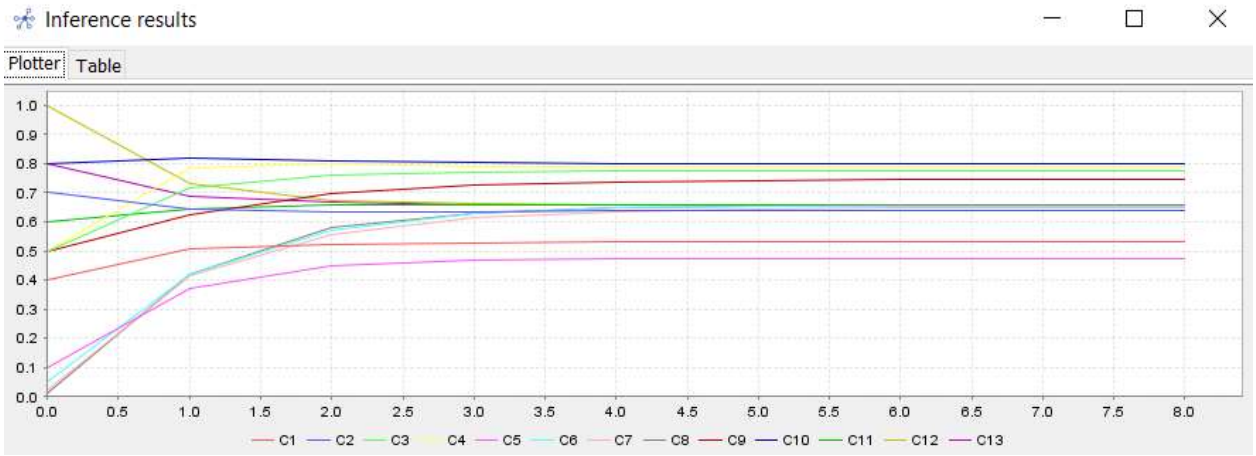


Рис. 3.8 динаміка симуляції при зміні параметрів вузлів C13 і C10

Inference results

Plotter | Table

Step	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
0	0.4	0.7	0.5	0.5	0.1	0.05	0.02	0.01	0.5	0.8	0.6	1.0	0.8
1	0.5075	0.6457	0.715	0.7858	0.3705	0.4207	0.417	0.4197	0.6248	0.8205	0.6457	0.7311	0.69
2	0.521	0.636	0.7626	0.7975	0.4476	0.5721	0.5557	0.5783	0.6971	0.8087	0.656	0.675	0.666
3	0.5276	0.6364	0.7724	0.7921	0.4688	0.6274	0.6131	0.6306	0.7261	0.8035	0.6584	0.6626	0.6606
4	0.5309	0.6373	0.7744	0.7895	0.474	0.6464	0.6364	0.6486	0.7374	0.8019	0.6589	0.6599	0.6594
5	0.5323	0.6378	0.7748	0.7888	0.4751	0.6527	0.6454	0.6551	0.7419	0.8014	0.659	0.6592	0.6591
6	0.5328	0.638	0.7749	0.7885	0.4753	0.6547	0.6486	0.6575	0.7436	0.8013	0.659	0.6591	0.6591
7	0.533	0.638	0.7749	0.7885	0.4753	0.6553	0.6497	0.6583	0.7442	0.8013	0.659	0.6591	0.6591
8	0.533	0.638	0.7749	0.7885	0.4753	0.6554	0.6501	0.6586	0.7444	0.8013	0.659	0.659	0.659

Рис. 3.9 результат симуляції при зміні параметрів вузлів C10 і C13

Порівнюючи цю симуляцію з попередньою можна чітко побачити, що зміна, а саме збільшення C10 і C13, спричинило збільшення початкової довіри користувачів (C1) і впливу на відкриття листа (C6), оскільки більш складна та правдоподібна атака підвищує ризик помилки користувача. Тобто навіть з урахуванням тренінгів, які ми лишили з минулої симуляції, це не настільки сильно вплинуло на неуспішність виконаної атаки, хоча все ж і має стримувальний ефект, що можна чітко відстежити по тому, що динаміка змінилась не критично раптово, а більш плавно та регульовано.

Якщо провести таку саму симуляцію, але з початковим значенням вузла C12, для того, щоб порівняти динаміку зміни, то цей результат можна побачити на рисунку 3.10.

Inference results

Plotter | Table

Step	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
0	0.4	0.7	0.5	0.5	0.1	0.05	0.02	0.01	0.5	0.8	0.6	0.4	0.8
1	0.5818	0.6457	0.715	0.6942	0.3705	0.495	0.417	0.4197	0.6248	0.8205	0.6457	0.5987	0.69
2	0.5717	0.636	0.7626	0.7637	0.4476	0.6309	0.5721	0.5783	0.6971	0.8087	0.656	0.6454	0.666
3	0.5497	0.6364	0.7724	0.7825	0.4688	0.6575	0.6295	0.6337	0.7277	0.8035	0.6584	0.656	0.6606
4	0.539	0.6373	0.7744	0.787	0.474	0.6591	0.6464	0.652	0.7393	0.8019	0.6589	0.6584	0.6594
5	0.5349	0.6378	0.7748	0.7881	0.4751	0.6575	0.6502	0.6574	0.7432	0.8014	0.659	0.6589	0.6591
6	0.5336	0.638	0.7749	0.7884	0.4753	0.6563	0.6507	0.6587	0.7443	0.8013	0.659	0.659	0.6591
7	0.5332	0.638	0.7749	0.7884	0.4753	0.6558	0.6506	0.6589	0.7445	0.8013	0.659	0.659	0.6591

Рис. 3.10 Результат симуляції з початковим значенням C12, та новими C10,C13

На рисунку 3.10 чітко видно, що без урахування підвищеного навчання користувачів можливість успішної компрометації системи та подальшого поширення зростає ще більше.

Таких симуляцій можна проводити ще багато, щоб дослідити кожен сценарій і динаміку змін залежно від нових встановлених значень вузлів і їх впливів на інші.

3.3 Моделювання поведінки атакуючих і внутрішніх користувачів у сценаріях

Моделювання поведінки людей, як внутрішніх користувачів, так і атакуючих, є критичним елементом аналізу фішингових загроз, адже на відміну від технічних вразливостей, фішинг використовує когнітивні особливості людини (довіру, неуважність, емоційний стан, загальну обізнаність з кіберризиків). Саме тому його успішність залежить не лише від технічної складності зловмисника, але й від того, як користувач інтерпретує отриманий лист, наскільки він довіряє відправнику та які дії здійснює після відкриття повідомлення.

Проведені раніше симуляції чітко демонструють, як зміна одного з таких когнітивних факторів, наприклад, довіра користувача або його рівень навчання, здатна трансформувати всю динаміку фішингової атаки та кінцевий рівень ризику. І саме тому, що людина найчастіше стає первинною точкою входу в систему і її реакції впливають на успішність атаки, необхідно змоделювати портрети користувачів та атакуючих, щоб чіткіше розуміти їх можливу поведінку.

Важливо розуміти, що внутрішній користувач – це набір поведінкових профілів, кожен з яких має свої власні мотивації. У робочому контексті мотиваційними чинниками можуть бути:

1) виконання завдань, адже користувач прагне швидко виконати робочі обов’язки, і це може підвищити готовність відкривати листи та вкладення без детальної перевірки;

2) економія часу, а саме бажання мінімізувати витрати часу на рутинні дії (скоріше відкрити лист без перевірки);

3) соціальна мотивація, яка може проявлятись у прагненні відповідати на листи від керівництва чи колег, наприклад, коли фішинговий лист імітує вже створені внутрішні комунікації;
4) винагорода або цікавість, тому що відкриття листів часто може бути через цікаві теми або обіцянки певної винагороди (наприклад, подарунки, платіжки або бонуси).

У моделі ці мотивації опосередковано впливають на вузли, що визначають поведінку, а саме C1 (довіра до відправника) чи C6 (ймовірність відкриття). Наприклад, висока мотивація швидко виконати напряму підвищує C1 та C6.

Рівень обізнаності у свою чергу визначається знанням правил кібербезпеки, навичками розпізнавання фішингу та звичкою перевіряти підозрілі листи. Він формується завдяки:

- тренінгам і досвіду, тому що регулярне навчання, як було показано на практичній симуляції, підвищує обізнаність (C4);
- практичному досвіду, бо особи, що вже стикались з фішингом, за правилом будуть більш обережні;
- культурою організації, адже політики та процедури впливають на практичну обізнаність.

У побудованій НКК обізнаність (концепт C4) має негативний вплив на C1 і C6, тобто зменшує наївну довіру і ймовірність відкриття листа чи вкладення. Симуляції наочно показують, що збільшення C12 (навчання) призводить до підвищення C4 на ранніх кроках і тимчасового зниження C1 та C6.

Довіра до відправника є ключовим когнітивним параметром. Вона може формуватися через:

1) контекст повідомлення, що включає в себе, наприклад, ім’я відправника чи тему;

2) історію взаємодій, адже лист від колеги викликає менше підозр, ніж повністю невідомий;
3) обізнаність, тому що критичне мислення і знання значно зменшують автоматичну довіру.

Можна побачити, що в НКК C1 є вузлом з великою кількістю входів (C4, C12, C10). Вплив C10 (складність атаки) може підвищити довіру, якщо атакуючий робить лист правдоподібним, в той час як C4 та C12 її знижують. Це дає змогу відображати реалістичні випадки, де навіть при високій обізнаності дуже добре спрямоване повідомлення може змусити довіритись.

Типовими реакціями користувача при отримані підозрілого листа можуть бути миттєве відкриття, що характерне для користувачів з низькою обізнаністю або високою мотивацією, перевірка відправника, здебільшого при середній обізнаності, запит до IT/колег, що очікувано при високій політиці і культурі звітності, або просто ігнорування чи видалення, коли довіра низька і обізнаність висока.

У симуляцій такі рівні відслідковуються через зміну концепту C6 (відкриття) і його подальшому впливі на C7 (компрометація). Показники часової динаміки, значення на різних кроках, демонструють, які реакції переважають на практиці.

Не можна сформувати лише портрет користувача, знання про зловмисника також суттєво впливають на моделювання.

Зазвичай атакуючий у фішинговому сценарії переслідує одну або кілька цілей, це можуть бути:

- отримання облікових даних, що є ключовою метою для подальшого проникнення;
- виконання фінансових операцій або шахрайство;
- розповсюдження шкідливого ПЗ або відкриття «дверей» у мережі;
- збір інформації для подальших спрямованих атак.

Ці цілі відображаються у вузлах НКК як С7 (компрометація) і як опосередковані ризики С8 (латеральне поширення), тобто успішність атаки досягається при високих значеннях цих вузлів.

Однієї цілі не достатньо, у зловмисника також має бути стратегія для успішної атаки. Стратегія атакуючого формує вибір тактики для досягнення цілі:

1. Широкі масові кампанії, в яких роблять ставку на велику кількість потенційних жертв. У моделі це підвищує С5 (доставка фішингового листа) завдяки широкому охопленню;
2. Персоналізовані повідомлення для конкретних людей або ролей, що підвищує ймовірність обходу фільтру (вплив на С2 через С10, а також підвищення довіри С1);
3. Багатофазні атаки, в яких спочатку здобувається доступ, потім ескалуються привілеї і далі йде рух латерально (від С7 до С8).

У створеній НКК С10 моделює складність атаки та правдоподібність, С13 – здібності зловмисника. Значні підвищення С10 та/або С13 підсилюють негативний вплив на С2 (фільтрацію) і позитивно впливають на С1 через правдоподібність, що симулюється у сценаріях.

Тактика обирається виходячи з поставлених цілей та доступних ресурсів, і також там можуть використовуватись таймінг та повторність (часті повідомлення можуть обернутись зниженням пильності, якщо користувачі звикнуть).

Окрім того, сучасні атакуючі адаптуються, вони аналізують зворотний зв’язок (наприклад, які листи проходять, а які ігноруються) і змінюють тактику, маніпулюють соціальною інженерією, темами, часом відправлення тощо. У НКК адаптивність можна відобразити як динамічну зміну С10 і С13 між сценаріями або як зворотний зв’язок, що підсилює С10 за умови низької ефективності фільтрів (С2). У вже виконаних симуляціях це представлено як варіації С10/С13 та відповідні відгуки в моделі. Ключовим результатом

виконаних у попередньому розділі симуляцій є те, що поведінкові параметри користувачів та характеристики атакуючого спільно визначають динаміку ризику фішингу, і ці взаємозв'язки можна змоделювати і проаналізувати за допомогою НКК.

3.4 Впровадження когнітивних сценаріїв у процес прийняття рішень (SOC, CERT)

Використання когнітивних сценаріїв у діяльності команд SOC та CERT відкриває можливість значно підвищити ефективність реагування на кіберінциденти.

Першим кроком інтеграції є використання НКК для пріоритезації інцидентів. Когнітивні моделі дозволяють оцінити, які фактори найбільше впливають на успішність атаки, а також виокремити події, що мають найбільший потенціал ескалації. Наприклад, якщо симуляція показує, що поєднання високої довіри користувача та збільшення складності атаки створює високий рівень ймовірності відкриття фішингового листа чи вкладення, то SOC може автоматично підвищувати критичність інцидентів, пов'язаних з такими шаблонами. Це в свою чергу забезпечує ресурсну економію та покращує точність реагування.

Другим елементом впровадження є формування автоматизованих когнітивних профілів користувачів і типових моделей поведінки. На основі НКК SOC/CERT можуть створювати поведінкові карти ризиків, наприклад, визначати співробітників, які є більш вразливими до соціальної інженерії, а також генерувати сценарії, як ті чи інші когнітивні характеристики, наприклад, низька обізнаність чи висока довіра, впливатимуть на можливість компрометації. Це формує персоналізований підхід до управління ризиками в організації.

Наступним рівнем є застосування когнітивних сценаріїв для сценарного прогнозування. SOC та CERT можуть використовувати результати симуляцій

для побудови «що-якщо» моделей, які тестують стратегії реагування на різні типи атакуючих тактик. Наприклад, у нашому випадку фішингової кампанії модель допомагає оцінити, як зміна інтенсивності атак, зміна рівня складності листів або адаптація атакуючого впливатимуть на користувачів. Це вже дає змогу вибудовувати стратегії випереджувального реагування, а не лише постфактум реакції на інциденти.

Особливо важливою складовою інтеграції є підтримка процесів прийняття рішень у реальному часі. На основі когнітивної карти можна налаштувати автоматичні тригери, які попереджатимуть аналітиків SOC про можливу ескалацію інциденту ще до того, як компрометація стане очевидною технічно. Наприклад, якщо система виявила поведінку, типову для користувача з високою ймовірністю довіри до підозрілих листів, інцидент може бути позначений як більш критичний навіть за відсутності прямої компрометації. Такі індикатори значно прискорюють реакцію та мінімізують наслідки атаки.

Ще один аспект – це інтеграція когнітивних сценаріїв у навчальні програми та тренінги для SOC/CERT. Результати моделювання дають змогу адаптувати навчання під реалістичні сценарії, засновані на реальній поведінці людей та типових тактиках зловмисників. Це робить тренувальні симуляції значно ефективнішими, адже вони враховують не лише технічні прояви атаки, але й людський фактор.

У комплексі таке впровадження створює багаторівневу системи оцінки ризиків, де технічні індикатори доповнюються когнітивними сигналами, а реакція базується на поведінкових патернах, підтверджених моделюванням. Це дозволяє SOC та CERT оперативніше виявляти загрози, точніше прогнозувати розвиток інцидентів та ефективніше управляти ризиками пов’язаними з фішинговими атаками та іншими соціально-інженерними загрозами.

3.5 Автоматизація моніторингу та коригування ризиків за допомогою когнітивних моделей

Автоматизація процесів моніторингу та оцінки ризиків у кібербезпеці є необхідною через динамічність атак, зростання обсягів даних та обмежень людських ресурсів. Нечіткі когнітивні карти, використані у цій роботі, є не лише інструментом для аналізу сценаріїв, але й можуть бути інтегровані у автоматизовані системи для безперервного відстеження змін ризикових факторів, адже завдяки своїй здатності моделювати взаємозв'язки між технічними та когнітивними концептами, НКК можуть стати основою для інтелектуальних модулів, що динамічно оцінюють ситуацію та ініціюють коригувальні дії без прямої участі аналітика.

Важливо врахувати, що автоматизація роботи з когнітивними моделями потребує низки технічних особливостей, які забезпечують коректне функціонування системи та її здатність адаптивно реагувати на всі ці зміни у поведінці користувачів та зловмисників.

Першим кроком автоматизації є створення каналу, який забезпечує постійне оновлення значень концептів НКК на основі реальних даних із інфраструктури, до таких даних можуть належати, наприклад, метрики електронної пошти (кількість підозрілих листів, частота спрацьовування антифішингових фільтрів, повторювані шаблони надсилання), поведінкові показники користувачів (частота відкриття листів, успішність проходження тренінгів, типові помилки), а також технічні параметри системи (спрацювання MFA, швидкість реагування, індикатори можливого латерального поширення). Отримані дані автоматично інтерпретуються і трансформуються у значення концептів НКК, що дозволяє моделі відображати реальний стан ризиків у режимі, наближеному до реального часу.

Ще одна важлива особливість це те, що можливість запуску обчислень у режимі реального часу передбачає, що модель має бути технічно реалізована у вигляді сервісу або мікросервісної компоненти, яка працює з мінімальною

затримкою. Найчастіше для цього використовують Rest API або gRPC API, через які SOC отримує оновлені оцінки ризику після кожної події.

Після отримання актуальних значень система може автоматично запускати процес семантичного висновку НКК, тобто модель самостійно обчислює, як зміна одного фактору, наприклад, зниження ефективності MFA або збільшення підозрілих листів, впливає на подальші елементи атаки, а саме довіру користувача, ймовірність відкриття листа, ризик компрометації та латеральне поширення. Таким чином, відбувається автоматичне прогнозування розвитку атаки, а наявність атрактора в моделі забезпечує стабілізацію результатів, що в свою чергу робить прогноз передбачуваним та системним.

Окремо варто виділити потребу в модулі автоматичного коригування вагових коефіцієнтів, бо, оскільки поведінка користувачів та атакуючих змінюється, ваги між концептами не можуть постійно залишатись статичними. Технічно це може бути реалізовано за допомогою алгоритмів навчання, наприклад, генетичні алгоритми, градієнтний спуск на основі історичних подій або адаптивне коригування залежно від частоти інцидентів. Система має зберігати історію змін ваг та автоматично перевіряти, чи не призводять ці зміни до деградації точності.

На основі впливів між концептами модель може визначати рівень ризику та автоматично видавати попередження. Наприклад, при збільшені складності атаки формувати сигнал про ймовірне зростання ризику компрометації. Важливо, що НКК враховує взаємодію людських факторів та технічних параметрів, як вже було сказано раніше, тому її попередження відрізняються від традиційних алертів більш комплексним трактуванням ситуації, що дозволяє отримувати більш структуровані повідомлення із зазначенням причинно-наслідкових ланцюжків. Наприклад, «зниження довіри $\rightarrow$ зростання відкриття листів $\rightarrow$ збільшення ризику компрометації». Така інтерпретованість є цінною для швидкого прийняття рішень.

Важливим етапом автоматизації є не лише виявлення проблеми, але й ініціація коригувальних дій. НКК можуть формувати рекомендації, які базуються на причинно-наслідкових залежностях. Наприклад, якщо модель визначає підвищений рівень компрометації через зменшення обізнаності, вона може запропонувати автоматичне планування додаткового тренінгу, або якщо ризик зростає через технічний фактор (недостатня ефективність фішингового фільтра тощо), то вона може запропонувати посилення політик безпеки або перевірку конфігурації. Умовно, якщо НКК показує, що значення концепту «компрометація» перевищує 0,7, то SOC запускає автоматизовані дії: блокування акаунту, примусову зміну паролю, обмеження доступу чи створення інциденту для аналітика. Такий підхід забезпечує адаптивну корекцію ризиків, де модель виступає не лише детектором, а й рекомендаційною системою.

Також важливою технічною особливістю є необхідність забезпечення стійкості та масштабованості. Когнітивна модель повинна обробляти велику кількість запитів без деградації продуктивності, а також мати можливість працювати в кластерному середовищі. Це включає в себе контейнеризацію, оркестрацію, розподілений кеш та логування виконання симуляцій для подальшого аналізу SOC та CERT. Докери дозволяють ізолювати модель, забезпечити її однакову поведінку у різних середовищах, швидко масштабувати або оновлювати компоненти. Оркестрація для того, щоб система могла обробляти велику кількість паралельних запитів та автоматично розподіляти навантаження. Для збереження проміжних результатів симуляцій, значень концептів або попередньо розрахованих сценаріїв, і зменшення кількості повторних обчислень у систему додають розподілений кеш. І логування, бо кожен запуск когнітивної моделі має фіксуватись у журнал, включаючи початкові дані, послідовність змін концептів, фінальний стан та час виконання. Це потрібно для аудиту, аналізу інцидентів, подальшого навчання системи та оцінки її коректності.

Усі ці технічні аспекти забезпечують можливість використовувати нечіткі когнітивні карти не лише як дослідницький інструмент, але і як реальну операційну складову системи кіберзахисту, яка автоматично аналізує поведінку користувачів і зловмисників, оцінює ризики та запускає реакції у відповідь на зміну середовища.

Висновки до третього розділу

У цьому розділі було показано як симуляції з використанням нечітких когнітивних карт демонструють, що зміна одного концепту системи впливає на інші концепти у ланцюговій взаємодії. Наприклад, зміни у довірі користувача, складності атаки або здібностях зловмисника відображаються на ймовірності відкриття листа, компрометації системи та поширенні фішингової атаки. Система показує динаміку взаємопов’язаних змін, де початкові значення поступово стабілізуються на фінальних кроках, досягаючи атрактора. Це підкреслює, що нечіткі когнітивні карти ефективно відображають складні взаємозв’язки між когнітивними та технічними факторами і дозволяють моделювати потенційні ризики у кібербезпеці.

Інтеграція НКК у SOC/CERT також дає можливість удосконалити систему пріоритезації сповіщень. Наприклад, фішингові листи, отримані користувачами з низьким рівнем обізнаності або підвищеною довірою, можуть автоматично отримувати вищий пріоритет для оперативної перевірки. Так само сценарії, у яких симуляція показує високу ймовірність компрометації, можуть бути помічені як критичні незалежно від технічного рівня загрози.

Автоматизація, у свою чергу, робить когнітивну модель частиною активного циклу реагування, перетворюючи її з інструмента аналізу на адаптивний механізм, який здатен самостійно виявляти критичні зміни і ініціювати захисні дії. Такий підхід не лише підвищує якість оцінки ризиків, а й забезпечує організацію гнучким, інтелектуальним і стійким інструментом протидії сучасним фішинговим атакам.

Розділ 4 ОБМЕЖЕННЯ, ВИКЛИКИ ТА ПЕРСПЕКТИВИ РОЗВИТКУ КОГНІТИВНОГО МОДЕЛЮВАННЯ

4.1 Технічні та методологічні виклики

Попри значний потенціал когнітивного моделювання у сфері кібербезпеки, цей підхід стискається з низкою технічних та методологічних обмежень, які впливають на точність результатів, масштабованість рішень та можливість їх практичного впровадження.

Ключовим чинником, що спонукає застосовувати когнітивне моделювання, є нестача достовірних і достатньо повних даних про реальні кіберінциденти. Більшість компаній уникають розголошення подробиць атак, щоб не зазнати репутаційних втрат, тому значна частина випадків або не фіксується офіційно, або подається в сильно узагальненому вигляді. Така закритість робить неможливим побудову точних статистичних моделей, які потребують великого масиву історичних спостережень та верифікованих даних. У результаті організації змушені переходити до більш гнучких, але менш формально точних методів, серед яких особливе місце займає сценарний аналіз із використанням нечітких когнітивних карт. Цей підхід дозволяє частково компенсувати брак даних за рахунок експертних оцінок та моделювання взаємозв’язків між людськими, організаційними та технічними факторами [43].

Проте водночас серйозним методологічним викликом є саме те, що нечіткі когнітивні карти майже повністю залежать від експертних оцінок, як під час формування структури моделі, так і під час визначення вагових коефіцієнтів впливу між концептами [44]. Це створює ризик суб’єктивності, де кожен експерт інтерпретує взаємозв’язки по-своєму, може мати власні упередження, різний досвід, різні уявлення про поведінку користувачів чи атакуючих. У результаті ці упередження фактично вбудовуються у матрицю впливів і впливають на роботу всієї моделі. Навіть застосування групових методів отримання згоди (наприклад, метод Дельфі, агрегування думок) не дає

змоги повністю виключити людський фактор. Вони лише зменшує його вплив, але не усувають повністю.

Ця залежність від експертів також означає, що моделі критично визначаються якістю людей, які їх створюють, а саме їхньою компетентністю, досвідом у домені, здатністю мислити системно і умінням формувати причинно-наслідкові зв'язки. У складних кіберінцидентах, де поведінка атакуючих є динамічною та адаптивною, експерти можуть переоцінювати або недооцінювати певні фактори, що в подальшому призводить до спотворення результатів симуляцій.

Крім того, НКК складно автоматично валідувати, оскільки взаємозв'язки між концептами часто відображають когнітивні або поведінкові процеси, які неможливо прямо виміряти або перевірити за допомогою доступних даних. Тобто когнітивні моделі, в тому числі нечіткі когнітивні карти, спрощують людську поведінку до набору змінних і залежностей, які не завжди здатні охопити всю її складність. Реакція користувачів на ті самі фішингові атаки чи інші фактори можуть залежати від психологічного стану, емоцій, ситуативного контексту, соціального тиску чи попереднього досвіду, тобто факторів, які важко формалізувати. Аналогічно поведінка атакуючого включає адаптивність, креативність і здатність модифікувати техніки в реальному часі, що також важко інтегрувати в статичну або слабко-адаптивну модель. Це суттєво ускладнює масштабування та використання когнітивних моделей у реальних системах підтримки прийняття рішень.

Також рішення повинні бути не просто точними, а й прозорими та передбачуваними, оскільки від цього залежить рівень довіри, можливість перевірки та регуляторна відповідність. Розуміння моделі у сфері кібербезпеки включає два взаємопов'язані аспекти – інтерпретованість та пояснюваність. Інтерпретованість дає можливість встановити причину отриманого рішення, тоді як пояснюваність описує сам механізм мислення системи, логіку її обчислень та зв'язків між параметрами. Навіть попри те, що когнітивне моделювання забезпечує певний рівень структурної прозорості, воно не

завжди дозволяє детально відстежити, як саме первинні суб’єктивні ваги та зв’язки між концептами трансформуються у кінцевий прогноз.

Особливо чітко обмеження традиційних моделей проявляються у дослідженні розподілених інформаційних систем, де ризики змінюються під впливом великої кількості факторів [45]. Сучасна інфраструктура, зокрема екосистеми Інтернет речей, генерує потоки гетерогенних, динамічних та часто непередбачуваних даних, що не можуть бути адекватно оброблені статичними моделями. У таких умовах виникає потреба у створенні метрико-орієнтованих моделей, здатних гнучко реагувати на зміни середовища, автоматично перебудовувати структуру зв’язків та перераховувати вагові коефіцієнти відповідно до актуального стану інфраструктури [46].

Коли когнітивні карти зростають до масштабів десятків або сотень концептів, вони перестають бути легкими для обчислення структурами та перетворюються на складні нелінійні системи, що вимагають значних обчислювальних ресурсів. Зі збільшенням кількості вузлів автоматично зростає й кількість зворотних зв’язків між ними, а це означає, що кожна зміна одного параметру здатна впливати на інші, інколи багаторазово, через декілька ітерацій. У результаті симуляція такої КМ перетворюється на складний пошук нового стану рівноваги, а кожна ітерація розрахунку наближає систему до збіжності, однак чим більше карта, тим повільнішим стає цей процес. Додатковий рівень складності виникає тоді, коли КМ використовують нечіткі значення, нелінійні функції активації чи гібридні механізми обчислення. Нечітка логіка збільшує кількість проміжних станів, які потрібно врахувати, а нелінійність викликає можливість появи хаотичної поведінки або чутливості до мінімальних змін у вхідних даних. В такому випадку система змушена виконувати більше ітерацій для стабілізації, а це пропорційно збільшує витрати часу та обчислювальних потужностей. Коли когнітивна модель занадто велика або перевантажена сценаріями, час відгуку починає перевищувати допустимі межі, що ставить під сумнів можливість практичного

застосування надто складних КМ у реальних умовах, де час визначає ефективність реагування.

Крім того, у процесі моделювання часто необхідно проводити десятки чи сотні різних сценаріїв, від типових фішингових кампаній до складних ланцюгових атак. Коли сценарій потребує багатьох ітерацій та обробляється великою багатовимірною моделлю, система стикається з ризиком перевантаження. У хмарних або локальних інфраструктурах це може призвести до затримок, а інколи і до неможливості виконання пріоритетних завдань через конкуренцію за ресурси.

Когнітивні моделі все дедалі частіше застосовуються у зв’язці з іншими методами аналізу, а саме АОМ, алгоритмами навчання, прогнозними моделями або класичними підходами до оцінювання ризиків. Така мультикомпонентна інтеграція є логічним кроком для підвищення точності, масштабованості та адаптивності системи, проте об’єднання цих інструментів нерідко виявляється складним процесом через фундаментальні відмінності між ними.

По-перше, різні типи моделей ґрунтуються на відмінному математичному апараті, наприклад, когнітивні карти використовують вагові коефіцієнти та функції активації для моделювання причинно-наслідкових зв’язків, тоді як машинне навчання працює з багатовимірними статистичними просторами, а АОМ оперує автономними об’єктами зі своєю логікою поведінки. Це призводить до труднощів у поєднанні результатів, оскільки кожен інструмент генерує дані у власній структурі, асоціює їх з різними параметрами та вимагає специфічних умов для коректної інтерпретації.

По-друге, моделі можуть використовувати різні формати вхідних даних, наприклад, агентні симуляції генерують покрокові траєкторії поведінки, тоді як когнітивні карти оперують нормалізованими високорівневими концептами, і узгодження таких даних потребує складної попередньої обробки: нормалізації, стандартизації, семантичного узгодження значень та визначення того, які параметри репрезентують одну й ту саму сутність.

По-третє, інтегровані моделі часто мають різні часові горизонти прогнозування і якщо ці горизонти не узгодити, результати можуть стати суперечливими, тоді одна модель передбачатиме стрімке зростання ризику, а інша стабілізацію або зниження.

Таким чином, у разі некоректної інтеграції можуть виникати суперечливі висновки, дублювання або перекручення факторів, а інколи і повна втрата критичних сигналів.

Також варто зазначити, що на відміну від класичних методик ризик-менеджменту, когнітивне моделювання поки не має уставлених стандартів або загальноприйнятих формалізованих підходів. Це може призводити до великої варіативності у тому, як різні фахівці створюють та описують моделі. Відсутність стандартизації ускладнює обмін моделями, повторюваність досліджень, аудит безпеки та подальше технологічне впровадження.

4.2 Проблеми збору та якості даних для моделей

Однією з ключових передумов ефективності когнітивного моделювання є доступ до якісних, релевантних та достатньо повних даних, проте на практиці організації часто стикаються зі значними труднощами у зборі і структуризації інформації, необхідної для коректного функціонування когнітивних моделей. На відміну від класичних статичних методів, КМ потребують комплексного уявлення про взаємозв’язки між параметрами, поведінкою користувачів, характеристиками інфраструктури та контекстом атак.

Проблема неповноти даних є однією з критичних і водночас найскладніших у контексті побудови когнітивних моделей. У розподілених інформаційних системах дані не завжди надходять у повному, цілісному вигляді, адже частина подій може просто не потрапити до централізованих журналів, що може трапитись через технічні збої, нестабільність каналів передачі, неправильно налаштовані політики логування, помилки адміністраторів чи через відсутність необхідних агентів моніторингу. У

результаті чого утворюються так звані «сліпі зони» інфраструктури, тобто сегменти, де реальна активність не відображена у даних, і які часто можуть використовувати зловмисники. Такі прогалини суттєво ускладнюють роботу когнітивних карт, оскільки моделі покладаються на наявність причинно-наслідкових залежностей між концептами ризику, і якщо частина цих залежностей не відстежена або відсутня, то модель може сформувати викривлене уявлення про систему. Наприклад, відсутність одного-двох критичних логів здатна змінити знак впливу між концептами, зменшити видимість реальної загрози або, навпаки, створити ілюзію ризику там, де його немає.

Особливо небезпечною неповнота даних стає у сценаріях реального часу, адже у таких умовах навіть дрібна подія, не зафіксована системою, може бути ключовим фрагментом у ланцюгу атаки, наприклад, початковим сигналом фішингу, першою спробою сканування порту, одиничним аномальним запитом до внутрішньої служби чи короткочасною появою шкідливого процесу. Якщо дані не потрапляють у когнітивну модель, система не здатна правильно оцінити контекст, що різко знижує її ефективність та оперативність реагування.

Більше того, неповні дані ускладнюють не лише прогнозування, але й валідацію самої моделі. Аналітик може помилково вважати модель неточною або неправильною, хоча насправді саме вхідні дані є фрагментарними. Це створює додаткові труднощі для калібрування ваг, перевірки сценаріїв, визначення ключових концептів та кореляції логічних зв’язків між ними.

Важливо, щоб дані були не лише повними, а й актуальними та своєчасними. У динамічному середовищі застаріла інформація стає практично непридатною, адже моделі повинні відображати поточний стан інфраструктури, поведінки користувачів та атакуючої сторони. Якщо дані надходять із затримкою або не оновлюються належним чином, когнітивна карта може працювати з неактуальними впливами, неправильно оцінювати коефіцієнти та прогнозувати застарілі сценарії.

Не менш суттєвою є проблема шумів і недостовірних значень. У журналах безпеки часто містяться дублікати, некоректні поля, помилки формату або навмисно хибна інформація з атакуючого боку. Високий рівень шуму ускладнює процес побудови точних функцій активації та може створювати ефект хибно активованих вузлів, коли модель реагує на сигнали, що не несуть реальної загрози.

Також людські когнітивні упередження, зокрема упередження аналітиків та розробників моделей, неминуче формують сенс, який витягується з даних. Підтверджувальне упередження є одним із найнебезпечніших когнітивних викривлень, оскільки воно спонукає аналітика не шукати істину, а підлаштовувати дані під уже сформовану думку. У контексті моделювання інформаційних систем це проявляється у свідомому або несвідомому ігноруванні даних, що суперечать початковим очікуванням, та у вибірковому використанні фактів, які підтримують бажаний сценарій [47]. Наприклад, якщо аналітик вважає певний ризик незначним, він може оцінювати неоднозначні події як несуттєві або звертати увагу лише на ті лог-записи, які підтверджують його припущення.

Для когнітивного моделювання наслідки такого упередження є критичними. Коли знання, які закладаються у когнітивні карти, або оцінка результатів моделей ШІ формуються під впливом попередніх переконань, отримані висновки щодо ризиків втрачають об’єктивність та надійність.

4.3 Етичні та юридичні аспекти використання когнітивних моделей

Етичні та юридичні аспекти використання когнітивних моделей у кібербезпеці є критично важливими для забезпечення надійності та довіри до автоматизованих рішень. У сучасних умовах, коли організації все частіше використовують когнітивні моделі для прогнозування кіберризиків та підтримки рішень, питання прозорості, пояснюваності та дотримання нормативних вимог стають особливо актуальними. Законодавство про захист

персональних даних, міжнародні стандарти кібербезпеки та внутрішні політики компаній формують рамки, у яких має здійснюватися моделювання, а будь-яке їх порушення може привести до серйозних юридичних наслідків.

Серед етичних проблем особливе значення мають прозорість моделей та пояснюваність їх рішень, адже недостатня інтерпретованість результатів когнітивної моделі може призвести до втрати довіри користувачів чи аналітиків, а також до неправильного реагування на потенційні загрози. Багато складних когнітивних моделей, особливо у глибокому навчанні, функціонують як «чорні ящики», тобто їхні процеси прийняття рішень важко або інколи неможливо зрозуміти людині, що ускладнює перевірку їхньої справедливості, надійності та притягнення до відповідальності у разі помилок [48].

Крім того, моделі здебільшого ґрунтуються на експертних оцінках та великому масиві історичних і соціальних даних, які початково можуть бути упередженими або неповними, що створює ризик дискримінаційних або некоректних висновків і несправедливого профілювання. Таке дискримінаційне упередження щодо певних груп людей, наприклад, може проявлятись якщо модель використовується у системах найму на роботу.

Під час симуляцій, навчання моделей і прогнозування ризиків також варто враховувати конфіденційність персональних та організаційних даних, що обробляються, аби не порушити права користувачів та не створити додаткові уразливості. У практичній реалізації це означає застосування засобів шифрування при передачі та зберіганні даних, а також обмеження доступу до них лише для уповноважених аналітиків. Важливим є використання анонімізації або псевдонімізації персональних даних у тих випадках, коли повне відтворення конкретної інформації не є критичним для моделювання. Крім того, варто впроваджувати журналювання та аудит доступу до даних, що дозволяє відстежувати дії користувачів та своєчасно виявляти потенційні порушення. Неврахування конфіденційності також може створити додаткові уразливості, наприклад, витік або неправильна обробка організаційних даних

може бути використана зловмисниками для планування атак, створюючи нові точки ризику для інформаційної системи.

Юридичні аспекти включають дотримання вимог законодавства щодо обробки персональних даних, таких як загальний регламент про захист даних (GDPR) та національні нормативні акти, наприклад, Закон України «Про захист персональних даних», Закони України «Про інформацію», «Про електронні комунікації», «Про доступ до публічної інформації» тощо.

Важливо відзначити відповідальність розробників та організацій за можливі помилки моделей, некоректні прогнози або неправильні рекомендації, що можуть призвести до фінансових втрат або шкоди репутації. Розробники когнітивних моделей, які виступають в ролі провайдерів високоризикових систем, можуть бути притягнуті до відповідальності, наприклад, якщо помилка моделі спричинена порушенням обов'язку догляду або недбалістю. Порушення обов'язку догляду означає невиконання зобов'язань щодо належного контролю та забезпечення ефективних механізмів для виявлення та запобігання потенційним порушенням, а недбалість проявляється у неспроможності діяти з належною обачністю під час розробки та впровадження систем, наприклад, через ігнорування прийнятих у галузі стандартів і практик [49-50].

При розробці та впровадженні когнітивних моделей важливо приділяти увагу ліцензійним питанням і захисту прав інтелектуальної власності, оскільки алгоритми моделювання та програмні компоненти можуть бути об'єктами авторського права або патентів. Використання сторонніх рішень, бібліотек, наборів даних або кодових фрагментів без належної ліцензії чи дозволу власника може призвести до юридичної відповідальності, включаючи цивільні позови або фінансові санкції. Крім того, правильне оформлення ліцензії забезпечує прозорість у використанні алгоритмів, дозволяє захищати власні розробки та уникати суперечок із третіми сторонами. Тому на етапі проектування та експлуатації когнітивних систем необхідно чітко визначати

джерела всіх використаних алгоритмів і даних, перевіряти умови ліцензій та забезпечувати дотримання авторських і суміжних прав.

4.4 Перспективи інтеграції з новими технологіями (машинне навчання, Big Data)

Інтеграція когнітивного моделювання з новими технологіями, такими як машинне навчання та Big Data, відкриває широкі перспективи для підвищення ефективності та точності систем управління ризиками. Сучасні когнітивні моделі, побудовані на основі нечітких когнітивних карт та БМ, демонструють значну користь при сценарному аналізі та оцінці ризиків, однак їхня продуктивність і точність часто обмежується доступністю даних та суб’єктивністю експертних оцінок. Інтеграція з новими технологіями дозволяє подолати ці обмеження, забезпечуючи автоматичне навчання моделей, обробку великих обсягів даних та адаптацію моделей до швидкозмінного кіберсередовища.

Проте важливо зауважити, що функціонування когнітивних моделей, особливо тих, що базуються на машинному навчанні, критично залежить від доступності та якості Big Data. Для успішного застосування ML у кібербезпеці необхідно мати справу з великими та складними наборами даних, що створює низку викликів [51]. Основними викликами є забезпечення масштабованості, швидкості обробки та ефективності аналітичних процесів. Крім того, спільнота ML сьогодні стикається з екологічними викликами, пов’язаними з великим обсягом даних, вимагаючи переходу до більш стійких практик [52].

Машинне навчання здатне істотно підвищити ефективність когнітивних моделей, зокрема автоматично визначати вагові коефіцієнти впливу та коригувати їх на основі накопиченої історичної інформації, що знижує залежність від суб’єктивних оцінок експертів [54]. Таким чином інтеграція ML/Big Data з когнітивними моделями призводить до вимірюваного підвищення ефективності, швидкості та точності (дослідження показують, що

ML-алгоритми значно підвищують точність виявлення загроз порівняно з традиційними методами безпеки, а також зменшують кількість хибних спрацювань) [19]. Проте, разом із цим інтеграція ML у когнітивні моделі стикається з низкою викликів, серед яких відмінності у форматах даних, складність інтерпретації результатів та необхідність забезпечення прозорості рішень.

Big Data у свою чергу надає інструменти для глибокого аналізу великих, різнорідних та динамічних масивів даних, що особливо важливо для роботи з сучасними системами інформаційної безпеки, де кількість подій, взаємодій і загроз постійно зростає. У контексті когнітивного моделювання це означає можливість наповнення моделі не лише експертними оцінками, а й реальними даними з логів, мережевих потоків, систем автентифікації, поведінкових профілів користувачів та телеметрії безпеки. Такий підхід дає змогу значно підвищити рівень деталізації моделі та точніше оцінювати взаємозв’язки між концептами.

Завдяки масштабованим архітектурам обробки потокових даних, таким як Apache Kafka або Flink, стає можливим здійснювати аналіз і прогнозування ризиків у режимі реального часу. Це дозволяє оперативно виявляти аномалії, фішингові кампанії, ескалацію інцидентів або зміни у поведінці користувачів, які до цього могли залишатись непоміченими. Іншими словами, Big Data суттєво зменшує кількість сліпих зон у системі, підсилюючи когнітивну модель актуальною інформацією, що постійно оновлюється.

Проте це так само може бути пов’язане з низкою технічних викликів, наприклад, дані можуть бути розрізненими, неструктурованими, дубльованими або містити помилки, що потребує значних ресурсів для їх очищення, нормалізації та стандартизації. Також необхідно забезпечити цілісність та достовірність інформації, адже навіть невеликі викривлення можуть призвести до хибних залежностей у когнітивній моделі. Впровадження систем контролю якості даних, управління метаданими, а також механізмів

аудиту та верифікації стає критично важливим елементом інтеграції Big Data з когнітивним моделюванням.

Гібридні підходи, які поєднують когнітивне моделювання з машинним навчанням та технологіями Big Data, забезпечують синергію цих методів і створюють більш адаптивні та прогнозуючі системи управління ризиками.

Перспективи розвитку інтеграції когнітивного моделювання з цими технологіями включають застосування глибинного навчання, автоматизованого навчання на основі потокових даних та вдосконалення прозорості для підвищення довіри до рішень моделей. Очікуваний ефект від впровадження таких гібридних систем полягає у підвищенні точності оцінки ризиків, прискорені процесів прийняття рішень та підвищенні надійності систем управління безпекою, що особливо актуально для розподілених і великих інформаційних інфраструктур.

Висновки до четвертого розділу

У цьому розділі було визначено ключові обмеження, виклики та перспективи розвитку когнітивного моделювання в контексті оцінювання та управління ризиками. Аналіз технічних та методологічних аспектів показав, що хоча когнітивні моделі мають значний потенціал для формалізації причинно-наслідкових зв’язків і підтримки прийняття рішень, їхня ефективність значною мірою залежить від коректності структуризації системи, вибору математичного апарату та здатності моделі враховувати динамічність середовища.

Проблеми збору, якості та повноти даних залишаються одним із найбільших бар’єрів для побудови точних і надійних когнітивних моделей. Недостатність або низька якість даних, наявність інформаційних прогалин та залежність від експертних оцінок можуть знижувати точність моделювання та обмежувати можливість практичного застосування результатів у реальних системах безпеки.

Етичні та юридичні аспекти використання КМ також набувають дедалі більшого значення. Питання відповідальності за помилки моделі, прозорості алгоритмів, захисту даних та дотримання прав інтелектуальної власності формують важливу частину нормативного поля, у межах якого ці технології можуть застосуватись.

Попри зазначені виклики, перспективи розвитку когнітивного моделювання є надзвичайно широкими. Інтеграція з машинним навчанням та технологіями Big Data відкриває можливості для підвищення точності, адаптивності та масштабованості моделей. Поєднання експертних знань із даними великого обсягу та алгоритмічними методами навчання створює підґрунтя для появи гібридних систем, здатних працювати в режимі реального часу та забезпечувати глибоке і прогнозоване розуміння ризиків.

ВИСНОВКИ

Проаналізувавши теоретичні, методологічні та практичні аспекти когнітивного моделювання у системах управління ризиками було встановлено, що когнітивні методи відіграють ключову роль у підвищенні обґрунтованості та передбачуваності рішень у сфері кіберзахисту. Проведений аналіз довів, що традиційні підходи до оцінювання ризиків не завжди здатні повноцінно врахувати динамічність атак, людський фактор, нелінійність взаємодій та високий рівень невизначеності сучасного кіберсередовища. Саме тому інтеграція когнітивного моделювання у процес прийняття рішень є актуальним та перспективним напрямом, що підтверджується дослідженням [53, 55].

У ході роботи були проаналізовані існуючі методи когнітивного моделювання, а саме: нечіткі когнітивні карти, байєсівські мережі та агентно-орієнтовані моделі, і описані їх переваги, обмеження та можливості комбінованого застосування в оцінці ризиків. Додатково було обґрунтовано, що застосування штучного інтелекту та автоматизації здатне підвищити точність, адаптивність та масштабованість когнітивних моделей.

У практичній частині роботи було сформовано когнітивну модель ризику фішингових атак із чіткою структурою концептів, вагових зв’язків та сценарних переходів. На основі нечіткої когнітивної карти створено набір сценаріїв розвитку кіберінцидентів з урахуванням поведінкових, технічних і організаційних факторів. Проведені симуляції показали, як зміна ключових параметрів (обізнаності користувачів, складності атаки, навичок зловмисника) впливає на кінцевий рівень ризику. Результати підтвердили, що когнітивні моделі здатні не лише відображати динаміку системи, а й виявляти критичні точки управління, завдяки яким можна знижувати ймовірність успішного інциденту.

Під час моделювання було доведено, що підвищення обізнаності користувачів має значний довгостроковий вплив на формування довіри та

зменшення ризиків, тоді як технічні засоби захисту забезпечують короткочасний, але сильний стабілізаційний ефект. Моделювання поведінки атакуючих і внутрішніх користувачів показало, що соціальні та психологічні фактори є критично важливими при визначенні вразливостей організації, а застосування когнітивного сценарного підходу дозволяє передбачати типові траєкторії атаки ще на етапі її формування.

Окрему увагу було приділено можливостям інтеграції когнітивних сценаріїв у роботу SOC та CERT. У роботі доведено, що навіть часткове включення когнітивного аналізу у процеси моніторингу може покращити пріоритизацію подій, підтримку рішень аналітиків та сценарну готовність до складних атак. Практична модель показала, що когнітивні сценарії здатні виступати інструментом автоматизованого моніторингу та своєчасного коригування ризиків.

У останньому розділі було систематизовано технічні, методологічні, етичні та юридичні обмеження застосування когнітивного моделювання, а також визначено ключові напрямки розвитку технологій. Зокрема, показано, що інтеграція з машинним навчанням та Big Data є перспективним шляхом до підвищення точності та адаптивності когнітивних карт, однак потребує вирішення питання якості даних, стандартизації, забезпечення конфіденційності та відповідності законодавству.

Таким чином, результати дослідження підтверджують доцільність та ефективність застосування когнітивного моделювання у сценарному аналізі кіберризиків. Виконана робота дозволила:

1. Сформувати теоретичну базу для застосування когнітивних методів у кібербезпеці.
2. Обґрунтувати вибір нечітких когнітивних карт як оптимального інструменту моделювання ризиків.
3. Розробити та побудувати практичну когнітивну модель фішингових загроз.

4. Провести симуляції та сценарний аналіз, що продемонстрували причинно-наслідкову динаміку атаки.
5. Встановити ключові закономірності впливу людських і технічних факторів на рівень ризику.
6. Визначити перспективи інтеграції когнітивних моделей в SOC/CERT та системи управління ризиками.
7. Виявити обмеження та напрямки подальших досліджень, зокрема у частині автоматизації, використання великих даних та дотримання юридичних і етичних вимог.

Оформлення результатів цього дослідження здійснювалося згідно з методичними рекомендаціями кафедри [56].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Гаркушенко, А., Шевченко, С., Жданова, Ю., & (2025). Когнітивний підхід в інформаційній та кібербезпеці. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 1(29), 854–866. https://doi.org/10.28925/2663-4023.2025.29.945.

[2] Gugerty, Leo. (2006). Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 50.

[3] G. Strube Cognitive Modeling: Research Logic in Cognitive Science URL: https://doi.org/10.1016/B0-08-043076-7/00588-X

[4] S. Shevchenko, et al., Information Security Risk Management using Cognitive Modeling, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, CPITS-II, vol. 3550 (2023) 297–305.

[5] What is cognitive AI?

URL:https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cognitive-ai

[6] Strategic Decision Making & Marr's Three Levels of Analysis

URL: https://marcbuehner.com/unlock-strategic-decision-making-with-marrs-3-levels-of-analysis/

[7] Джалладова, І. А.-к. і Камінський , О. Є. (2025) «Соціально-психологічна стійкість систем кібербезпеки», Сучасні інформаційні технології у сфері безпеки та оборони. Київ, Україна, 53(2), с. 43–50.

[8] Physical symbol system hypothesis

[9] Rescorla, Michael, "The Language of Thought Hypothesis", The Stanford Encyclopedia of Philosophy (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.)

[10] Types of cognitive models and their applications

URL: https://fiveable.me/introduction-cognitive-science/unit-7/types-cognitive-models-applications/study-guide/OFv0laWpP33k77hm

[11] Payne, Laura. "means-ends analysis". Encyclopedia Britannica, 1 Sep. 2023 URL: https://www.britannica.com/science/means-ends-analysis

[12] Soar (cognitive architecture)

URL: https://en.wikipedia.org/wiki/Soar_(cognitive_architecture)

[13] Connectionist vs Symbolic Models

URL: https://fiveable.me/key-terms/introduction-cognitive-science/connectionist-vs-symbolic-models

[14] Goel, Ashok K. 2021. “Looking back, looking ahead: Symbolic versus connectionist AI.” AI Magazine 42: 83–85.

[15] Chipman, Susan E. F. (ed.), The Oxford Handbook of Cognitive Science, Oxford Handbooks (2017; online edn, Oxford Academic, 3 Nov. 2014)

[16] Cognitive Architectures That Simulate Human Learning Stages

URL: https://next.gr/ai/ai-fundamentals/cognitive-architectures-that-simulate-human-learning-stages

[17] Marewski JN, Mehlhorn K. Using the ACT-R architecture to specify 39 quantitative process models of decision making. Judgment and Decision Making.

[18] Codex Y. A Comparative Analysis of Cognitive Architectures in Artificial Intelligence: Assessing Theoretical Basis, Assumptions, and Effectiveness in Achieving Human-Like Behavior. Yubetsu Codex Computer science

[19] Khan, Waseem & Ashoka, K & Razak, M & M V, Manoj Kumar & Naseer, Riffat. (2025). A Comprehensive Survey on Cognitive Cyber Security Analysis Using Machine Learning Approaches. IEEE Access. PP. 1-1. 10.1109/ACCESS.2025.3614388.

[20] What is Human-in-the-Loop (HITL) in AI & ML?

URL: https://cloud.google.com/discover/human-in-the-loop

[21] What is human-in-the-loop?

URL: https://www.ibm.com/think/topics/human-in-the-loop

[22] Кучаковська, Г., & Хорольська, К. (2025). Інтеграція Lean Canvas у підготовку IT-фахівців як механізму когнітивного спрощення та раціоналізації складних рішень у життєвому циклі IT-продукту. Кібербезпека: освіта, наука, техніка, 2(30), 305–315. https://doi.org/10.28925/2663-4023.2025.30.976

[23] Y. Kostiuk, et al., A system for assessing the interdependencies of information system agents in information security risk management using cognitive maps, in: 3rd International Conference on Cyber Hygiene & Conflict Management in Global Information Networks (CH&CMiGIN), Kyiv, Ukraine, vol. 3925, 2025, 249–264.

[24] Шевченко, С., Жданова, Ю., Складанний, П., & Петренко, Т. (2024). Нечіткі когнітивні карти як інструмент візуалізації сценаріїв реагування на інциденти в системах безпеки. Кібербезпека: освіта, наука, техніка, 2(26), 417–429. https://doi.org/10.28925/2663-4023.2024.26.707

[25] ДСТУ ISO/IEC 27005:2023 Інформаційна безпека, кібербезпека та захист конфіденційності. Настанова керування ризиками інформаційної безпеки (ISO/IEC 27005:2022, IDT)

[26] NIST Special Publication (SP) 800-30, Revision 1, Guide for Conducting Risk Assessments

[27] Застосування сценарного планування для підвищення ефективності стратегічного управління в умовах невизначеності / А. Д. Чикуркова та ін. Актуальні проблеми економіки. Київ, 2025. № 5 (287). С. 390-402.

[28] Tkachenko, Olha. (2019). Cognitive Modeling of Composite Systems. Digital Platform: Information Technologies in Sociocultural Sphere. 2. 11-19. 10.31866/2617-796x.2.1.2019.175650.

[29] Derevyanchuk, Oleksandr. (2023). Розробка моделі нечіткої когнітивної карти для створення STEM-проєктів у професійній підготовці майбутніх фахівців інженерно-педагогічних спеціальностей. Alfred Nobel University Journal of Philoligy. 26. 160-169. 10.32342/2522-4115-2023-2-26-16.

[30] Osoba, Osonde & Kosko, Bart. (2019). Causal Modeling with Feedback Fuzzy Cognitive Maps. 10.1002/9781119485001.ch25.

[31] Yu T, Gan Q, Feng G. Modeling time series by aggregating multiple fuzzy cognitive maps. PeerJ Comput Sci. 2021 Sep 20;7:e726. doi: 10.7717/peerj-cs.726. PMID: 34616897; PMCID: PMC8459780.

[32] Feyzioglu, Orhan & Buyukozkan, G. & Ersoy, M.S.. (2008). Supply chain risk analysis with fuzzy cognitive maps. 1447 - 1451. 10.1109/IEEM.2007.4419432.

[33] Poczeta K, Papageorgiou EI, Gerogiannis VC. Fuzzy Cognitive Maps Optimization for Decision Making and Prediction. Mathematics. 2020; 8(11):2059.

[34] Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Mateo, Calif.: Morgan Kaufmann Publishers; 1988.

[35] Wright S. Correlation and causation. Journal of Agricultural Research. 1921;20: 557–585.

[36] Pearl J. Causality: Models, reasoning, and inference. 2nd ed. Cambridge: Cambridge University Press; 2009. doi:10.1017/CBO9780511803161

[37] Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search. New York: Springer Verlag; 1993.

[38] Gao, A. (2021). Lecture 12 on Bayesian Networks [Конспект лекції]. Department of Computer Science, University of Toronto.

[39] Сидоренко М. О Навчання Байєсівської мережі гібридним алгоритмом max-min K2 / Сидоренко М. О. // Наукові записки НаУКМА. - 2012. - Т. 138 : Комп'ютерні науки. - С. 44-47

[40] Sheidaei, A., Foroushani, A. R., Gohari, K., & Zeraati, H. (2022). A novel dynamic Bayesian network approach for data mining and survival data analysis. BMC medical informatics and decision making, 22(1), 251.

[41] Milov, O. & Kostyak, M. & Milevskyi, Stanislav & Pogasiy, S.. (2019). ЗАСОБИ МОДЕЛЮВАННЯ ПОВЕДІНКИ АГЕНТІВ В ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМА. Системи управління, навігації та зв’язку. Збірник наукових праць. 6. 63-70. 10.26906/SUNZ.2019.6.063.

[42] Yangqiaoyu Zhou, Haokun Liu, Tejes Srivastava, Hongyuan Mei, and Chenhao Tan. 2024. Hypothesis Generation with Large Language Models. In Proceedings of the 1st Workshop on NLP for Science (NLP4Science), pages 117–139, Miami, FL, USA. Association for Computational Linguistics.

[43] Shevchenko, Svitlana & Zhdanova, Yuliia & Kryvytska, Olha & Shevchenko, Halyna & Spasiteleva, Svitlana. (2024). Fuzzy cognitive mapping as a scenario approach for information security risk analysis.

[44] Wang, Guan & Feng, Yimin & Guo, Rongbin & Liu, Yusheng & Zou, Qiang. (2025). A Token-FCM based risk assessment method for complex engineering designs. Journal of Mechanical Design. 147. 1-21. 10.1115/1.4069274.

[45] Gorbachuk, Vasyl & Dunaievskyi, Maksym & Golotsukova, Tamila & Rybachok, Dmytro. (2025). Standards and artificial intelligence for cybersecurity.

[46] Палко, Д., & Мирутенко, Л. (2024). МЕТОД КОМПЛЕКСНОЇ ОЦІНКИ РИЗИКІВ КІБЕРБЕЗПЕКИ В РОЗПОДІЛЕНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 2(26), 487–502.

[47] Кравченко, О & Безуглий, В. (2025). Вплив когнітивних викривлень на моделювання бізнес-процесів в інформаційних системах. Scientific Bulletin of UNFU. 35. 134-141. 10.36930/40350415.

[48] Gupta, B. (2024). Explainable AI in Cyber Defense: Why Transparency Matters More Than Ever. Journal of Cybersecurity, 15(3), 45-52.

[49] Citolino, C. (2024). Corporate director liability in the era of cybersecurity risks. URL: https://www.theregreview.org/2025/10/22/citolino-corporate-director-liability-in-the-era-of-cybersecurity-risks/

[50] Smith, Gregory, et al. Liability for Harms from AI Systems: The Application of U.S. Tort Law and Liability to Harms from Artificial Intelligence Systems. RAND Corporation, 2024

[51] Veksler VD, Buchler N, LaFleur CG, Yu MS, Lebiere C and Gonzalez C (2020) Cognitive Models in Cybersecurity: Learning From Expert Analysts and Predicting Attacker Behavior. *Front. Psychol.* 11:1049. doi: 10.3389/fpsyg.2020.01049

[52] Dritsas, E., & Trigka, M. (2025). Exploring the Intersection of Machine Learning and Big Data: A Survey. Machine Learning and Knowledge Extraction, 7(1), 13. https://doi.org/10.3390/make7010013

[53] Шевченко, С., Жданова, Ю., & Спасітєлєва, С. (2025). Модель формування когнітивних навичок спеціалістів з кібербезпеки. Кібербезпека: освіта, наука, техніка, 2(30), 280–291. https://doi.org/10.28925/2663-4023.2025.30.973

[54] Sukaylo, I., & Korshun, N. (2022). The influence of NLU and generative AI on the development of cyber defense systems. Cybersecurity: Education, Science, Technique, 2(18), 187–196. https://doi.org/10.28925/2663-4023.2022.18.187196

[55] Негоденко В., Шевченко С., Негоденко О., Золотухіна О. (2025). Інтеграція теорії катастроф у моделі прийняття рішень для систем управління інформаційною безпекою. Телекомунікаційні та інформаційні технології, 4(89), 20-28. https://doi.org/10.31673/2412-4338.2025.048903

[56] Жданова, Ю. Д., Складанний, П. М., & Шевченко, С. М. (2023). Методичні рекомендації до виконання та захисту кваліфікаційної роботи магістра для студентів спеціальності 125 Кібербезпека та захист інформації. https://elibrary.kubg.edu.ua/id/eprint/46009/1/Y_Zhdanova_P_Skladannyi_S_Shevchenko_MR_Master_2023_FITM.pdf