КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2024.25.468486
УДК 004.056.5:004.896

Іосіфов Євген Анатолійович
аспірант кафедри інформаційної та кібернетичної
безпеки імені професора Володимира Бурячка
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0000-0001-6203-9945
y.iosifov.asp@kubg.edu.ua

Соколов Володимир Юрійович
к.т.н., доцент
доцент кафедри інформаційної та кібернетичної
безпеки імені професора Володимира Бурячка
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0000-0002-9349-7946
v.sokolov@kubg.edu.ua

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ, ТЕХНОЛОГІЙ, СЕРВІСІВ ТА ПЛАТФОРМ ДЛЯ РОЗПІЗНАВАННЯ ГОЛОСОВОЇ ІНФОРМАЦІЇ В СИСТЕМАХ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Анотація. У статті проведено комплексний порівняльний аналіз методів, технологій а також розглянуто сучасні підходи до використання технологій розпізнавання мови та обробки природної мови (NLP) у контексті національної безпеки та інформаційної безпеки. Розглянуто ключові аспекти використання технологій для моніторингу комунікацій, виявлення підозрілої активності та застосування у сфері розвідки та контррозвідки, роль у забезпеченні кібербезпеки, можливості біометричної ідентифікації за голосом, етичні та правові аспекти, технологічні виклики. Постановка проблеми акцентує увагу на викликах, пов'язаних із широким впровадженням технологій розпізнавання мови та NLP, зокрема недостатня точність алгоритмів, що створює ризики для надійності систем безпеки. Також підкреслено важливість вирішення етичних та правових питань, пов'язаних із приватністю громадян та можливим зловживанням технологіями для масового нагляду. У роботі наведені приклади систем для забезпечення цілей кібербезпеки, таких як системи масового прослуховування та аналізу, системи цільового моніторингу, платформи аналізу соціальних мереж, системи біометричної ідентифікації та інші. У розділі результатів дослідження представлено високорівневу структуру систем захисту від загроз, яка охоплює канали загроз та рівні захисту. Розглянуто складність сучасних загроз, які можуть інтегруватися в декілька каналів одночасно, зокрема з використанням голосової інформації. Деталізовано місце та роль голосової інформації у структурі захисту від загроз, акцентовано на важливості інтеграції різних систем та платформ для забезпечення комплексної безпеки. Розглянуто два підходи до побудови системи безпеки, яка працює з голосовою інформацією: агрегування максимально можливої інформації з існуючих систем та створення системи під кожну конкретну проблему. Проведено порівняльний аналіз цих підходів, визначено їх переваги та недоліки а також описано обмеження та ризики застосування методів розпізнавання голосової інформації, зокрема надійність та точність технологій, наявність даних для тренування моделей, вартість впровадження, питання конфіденційності та приватності, безпеки даних, використання у військовій та розвідувальній діяльності, етичні питання, ризики підробки голосу та штучних голосів.

Ключові слова: Natural Language Processing; аудіодані; розпізнавання голосової інформації; автентифікація; глибоке навчання; машинне навчання; обробка тексту; кібербезпека; інформаційна безпека.

ВСТУП

У сучасному світі, де інформаційні технології відіграють ключову роль у забезпеченні національної безпеки, розпізнавання голосової мови та обробка природної мови (від англ. Natural Language Processing, NLP) є потужними інструментами, які мають значний вплив на безпеку держави. Ці технології дозволяють автоматизувати аналіз великих обсягів інформації [1], [2], що надходить у вигляді голосових повідомлень або текстових даних, що є критично важливим для збору розвідувальних даних, моніторингу громадської безпеки та забезпечення правопорядку а також відкривають нові можливості для аналізу та запобігання потенційним загрозам. Але ці технології також породжують низку технічних, етичних та правових викликів.

Актуальність дослідження зумовлена зростаючою роллю цифрових комунікацій у сучасному суспільстві та необхідністю ефективного захисту державних інтересів в інформаційному просторі. Розвиток технологій розпізнавання мови та NLP дозволяє автоматизувати процеси моніторингу та аналізу великих обсягів даних, що є критично важливим для своєчасного виявлення та реагування на потенційні загрози [3].

Метою даного дослідження є комплексний аналіз проблематики застосування технологій розпізнавання голосової мови та NLP у контексті забезпечення державної безпеки. У роботі розглянуті такі ключові аспекти:

- використання технологій для моніторингу комунікацій та виявлення підозрілої активності;
- застосування у сфері розвідки та контррозвідки для аналізу іншомовних джерел та виявлення прихованих повідомлень;
- роль у забезпеченні кібербезпеки, зокрема у виявленні фішингових атак та аналізі потенційно шкідливого контенту;
- можливості та обмеження біометричної ідентифікації за голосом;
- етичні та правові аспекти використання цих технологій у контексті балансу між безпекою держави та приватністю громадян;
- технологічні виклики, пов’язані з підвищенням точності розпізнавання та обробки мови в різних умовах.

Дане дослідження має на меті не лише висвітлити поточний стан розвитку та застосування цих технологій, але й окреслити перспективи їх подальшого вдосконалення та інтеграції в системи забезпечення державної безпеки. Особлива увага буде приділена аналізу потенційних ризиків та розробці рекомендацій щодо їх мінімізації.

Постановка проблеми. У сучасних умовах зростаючої залежності держав від цифрових технологій, застосування технологій розпізнавання голосу та NLP стає важливою складовою забезпечення національної безпеки. Незважаючи на значний прогрес у розробці таких технологій, їх широке впровадження супроводжується рядом викликів. По-перше, недостатня точність алгоритмів у різних мовних і акустичних умовах може призвести до помилкових результатів, що в свою чергу створює ризики для надійності систем безпеки. По-друге, обробка великих обсягів даних вимагає значних обчислювальних ресурсів, що може бути обмеженням для оперативного моніторингу загроз. По-третє, етичні та правові аспекти, зокрема питання приватності громадян, викликають занепокоєння через можливе зловживання такими технологіями для масового нагляду. Недостатнє вирішення цих викликів може знизити ефективність державних систем безпеки та призвести до небажаних наслідків як для національної безпеки, так і для громадянського суспільства.

Аналіз останніх досліджень і публікацій. Попередня робота [4] полягала у вивченні сучасних підходів до побудови та тренування мовних моделей для вирішення різних завдань NLP, таких як машинний переклад, розпізнавання мови, пошук інформації, аналіз тональності, підсумовування тексту та інші. Основна увага приділяється нейронним мережам, механізмам вбудовування, архітектурам енкодера та декодера, а також застосуванню трансформерів для паралелізації обчислень. Дослідження [5] присвячене порівнянню основних підходів у NLP та розпізнаванні мови. Автори досліджують вимоги до наборів даних для тренування текстових та мовних моделей, порівнюють основні інструменти і техніки [6], а також описують останні тренди в цій сфері. А в роботі [7] розглядаються структури систем автоматичного розпізнавання мови, включаючи гібридні та кінцеві моделі. Описуються переваги і недоліки кожної системи [8], а також проводиться порівняння вимог до тренувальних даних і обчислювальних ресурсів на прикладі реальних моделей.

Але системи штучного інтелекту та розпізнавання мови можуть відстежувати активність на предмет підозрілої поведінки, надаючи своєчасні сповіщення фахівцям з безпеки та адміністраторам безпеки, таким чином покращуючи можливості виявлення загроз та намірів використання потенційних вразливостей [9], [10].

Розпізнавання мови, інтегроване з глибоким навчанням і технологією блокчейн, може використовуватися для біометричного контролю доступу, забезпечуючи безпеку процесів автентифікації та ідентифікації [11]. Цей підхід допомагає зменшити ризики, пов’язані з безпекою даних, конфіденційністю та витоком інформації. Системи розпізнавання мови можуть виявляти аномалії в поведінці користувачів після надання доступу за біометричними даними (в тому числі, голосу), забезпечуючи тим самим додатковий рівень безпеки. Це особливо корисно для виявлення несанкціонованого доступу або незвичайних дій в системі.

Система автоматичного розпізнавання мови (від англ. Automatic Speech Recognition, ASR), яка описана в [12], вразливі до ворожих атак. Дослідження підкреслює потребу в надійних механізмах захисту, таких як згладжування сигналу і навчання на атаках зловмисника, для захисту систем ASR від таких загроз. Інтеграція розпізнавання мови зі штучним інтелектом і машинним навчанням дозволяє створювати системи в [10], які не тільки розпізнають мову, а й генерують та інтерпретують інформацію, пов’язану з безпекою. Це допомагає представляти критичні висновки про безпеку в зрозумілій для користувачів формі.

МЕТОДИКА ДОСЛІДЖЕННЯ

У роботі використано наступні методи: порівняльний аналіз методів, технологій, сервісів та платформ для розпізнавання голосової інформації; огляд і аналіз сучасних систем забезпечення інформаційної безпеки, що використовують розпізнавання мови та NLP; аналіз останніх досліджень і публікацій у сфері розпізнавання мови; критичний аналіз обмежень та ризиків застосування цих технологій у контексті національної безпеки; порівняння підходів до побудови систем безпеки, що працюють з голосовою інформацією; аналіз етичних та правових аспектів використання технологій розпізнавання голосу; узагальнення та формулювання рекомендацій щодо впровадження та розвитку цих технологій у системах забезпечення інформаційної безпеки.

Об’єкт дослідження — розпізнавання голосової інформації в системах забезпечення інформаційної безпеки. Предмет дослідження — методи, технології, сервіси та платформи для розпізнавання голосової інформації.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Високорівнева структура систем захисту від загроз

Сучасні загрози характеризуються своєю складністю і інтегрованістю в декілька каналів одночасно. Атаки з використанням голосової інформації можуть бути виконані як ізольовані, наприклад з використанням рацій або телефонів, так і інтегровані в інші канали, такі як соціальні мережі або з частковим використанням електронної пошти як додаткового каналу передачі інформації.

На рис. 1 зображена спрощена структура виникнення і захисту від кіберзагроз, що охоплює кілька ключових компонентів для забезпечення безпеки організацій і державних установ, а саме *канали загроз*, через які можуть виникати загрози і *рівні захисту*, що можуть бути застосовані для кожного з каналів загроз. Оскільки елементи рівнів захисту, зазвичай, не пов’язані між собою — це ускладнює виявлення і захист від мультиканальних загроз.

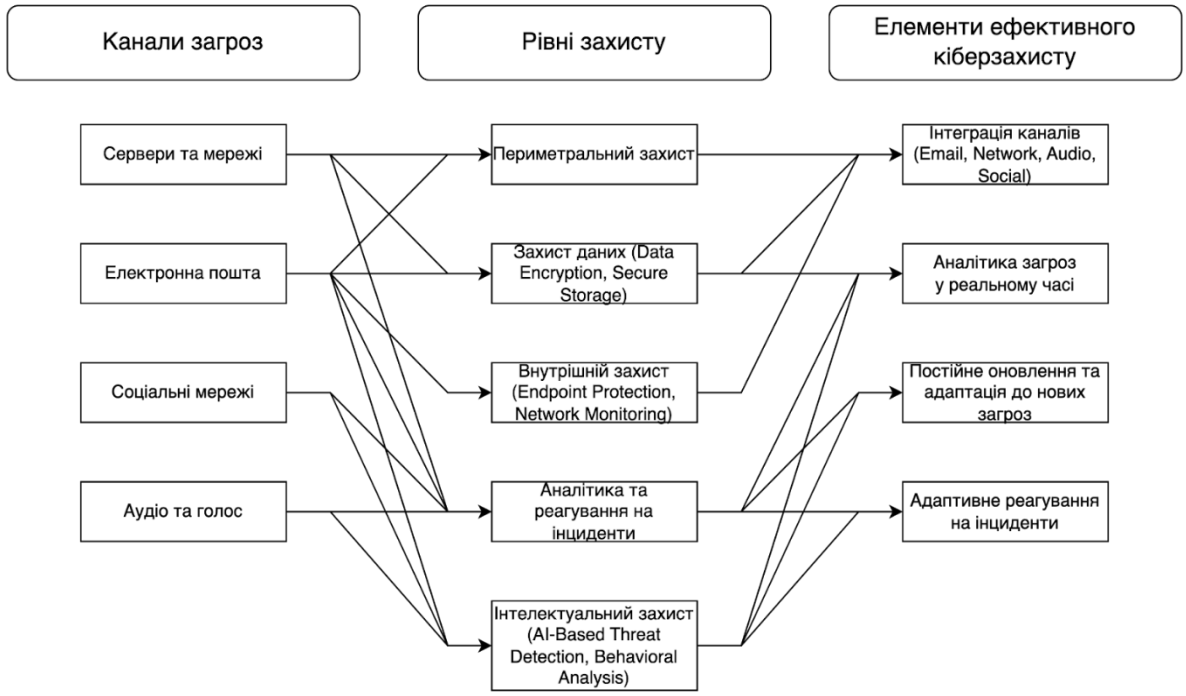

Рис. 1. Формування каналів загроз, рівнів захисту та елементів кіберзахисту

Для виявлення сучасних, мультиканальних загроз потрібен інтегрований рівень захисту, що відображається у блоці елементи ефективного кіберзахисту і демонструє, як різні рівні захисту можуть бути інтегровані в єдину систему для комплексного забезпечення безпеки.

ISSN 2663 - 4023

Місце та роль голосової інформації у структурі захисту від загроз

Одним з найскладніших для аналізу і водночас найбільш натуральним є канал голосової інформації. Аудіоінформація вбудована в усі сфери життя людини і тим чи іншим чином в більшість каналів придатних для атаки. Розглянемо детальніше рівні загроз і можливі системи захисту від них. На рис. 2 представлена деталізація систем і підходів для різних рівнів відображає (зеленим) системи пов'язані з захистом або ризиком атак використовуючи аудіо канал.

Рис. 2. Елементи кібербезпеки в системах розпізнавання голосової інформації

На державному рівні для боротьби з загрозами на основі аудіо даних використовується цілі спектри систем, платформ і підходів, які відіграють важливу роль у забезпеченні кібербезпеки та захисту від загроз у сучасному інформаційному просторі. Кожен з елементів цієї схеми виконує свою унікальну функцію, яка доповнює інші, створюючи таким чином комплексну структуру захисту.

Розглянемо системи масового прослуховування та аналізу, які є критичними для виявлення потенційних загроз на ранніх етапах. Вони дозволяють державним та правоохоронним органам отримувати доступ до великого обсягу даних з різних джерел комунікації та аналізувати ці дані з метою виявлення підозрілих патернів або загроз. Ці системи часто використовуються для боротьби з тероризмом та іншими видами злочинної діяльності. Але як видно з назви (масового) ці системи доволі обширні, що відображається у вартості їх роботи (обчислювальні ресурси, електрика, тощо). Тому в паралель використовуються Системи цільового моніторингу орієнтовані на спостереження за конкретними об'єктами або групами людей. Вони використовуються для стеження за підозрюваними особами, забезпечення безпеки важливих подій або об'єктів, а також для запобігання можливим загрозам у реальному часі. Ці системи дозволяють зосередитися на конкретних цілях і забезпечують детальний аналіз їх діяльності. Для забезпечення безпеки важливих подій, об'єктів або систем додатково виористовуються Системи біометричної ідентифікації, що використовують

використовують біометричні дані, такі як відбитки пальців, розпізнавання обличчя або голосу, для підтвердження особи користувачів. Вони забезпечують високий рівень безпеки у порівнянні з традиційними методами аутентифікації, такими як паролі. Ці системи широко застосовуються у правоохоронних органах, банківському секторі та для контролю доступу до важливих об'єктів.

Оскільки масовий моніторинг і прослуховування не є достатнім в поточному світі де соціальні мережі відіграють значну роль у поширенні інформації та формуванні суспільної думки, важливим елементом також є платформи аналізу соціальних мереж, що спеціалізуються на моніторингу та аналізі активності в соціальних мережах. З їх допомогою можна виявляти дезінформаційні кампанії, маніпуляції громадською думкою, а також слідкувати за підозрілими акаунтами або групами. На більш низькому рівні роботи з системами масового і цільового прослуховування а також з платформи аналізу соціальних мереж використовуються інструменти лінгвістичного аналізу. Вони допомагають аналізувати текстові дані для виявлення ключових слів, настроїв або прихованих повідомлень. Вони використовуються як у кібербезпеці, так і в контексті розвідки та контррозвідки. Ці інструменти дозволяють отримувати цінну інформацію з текстових джерел, таких як соціальні мережі, електронна пошта або інші види комунікацій.

Системи моніторингу збирають і структурують інформацію з різних каналів і джерел для виявлення атак і зловмисників по зібраним даним використовуються Системи виявлення аномалій, які є важливою частиною будь-якої стратегії кібербезпеки. Вони використовують методи машинного навчання та аналізу даних для виявлення незвичайної активності у мережевому трафіку або поведінці користувачів. Виявлення аномалій може бути першим кроком у виявленні кібератак або інших загроз, які не були зафіксовані традиційними засобами безпеки.

Як було зазначено вище, сучасні методи і вектори атак викоритовують одночасно багато каналів і джерел, тож дані і виявлені аномалії повинні бути інтегровані в більш високорівневі інтегровані платформи кібербезпеки, що об'єднують кілька різних компонентів захисту, таких як захист від кіберзагроз, аналітика загроз у реальному часі та можливості реагування на інциденти. Ці платформи здатні забезпечити цілісну картину безпеки для організацій, дозволяючи їм контролювати всі аспекти кібербезпеки з одного місця. Вони важливі для великих організацій, які потребують комплексного підходу до управління безпекою та ефективного реагування на різноманітні загрози.

Таким чином, всі ці системи і платформи взаємодіють між собою, створюючи комплексну структуру, яка забезпечує ефективний захист як від кіберзагроз, так і від фізичних загроз з боку зловмисників. Вони відіграють ключову роль у сучасних стратегіях безпеки, як на рівні організацій, так і на державному рівні.

Приклади систем для забезпечення цілей кібербезпеки

Системи масового прослуховування та аналізу:

- ECHELON — глобальна система радіоелектронної розвідки;
- PRISM — програма збору та аналізу даних електронних комунікацій;
- XKeyscore — система пошуку та аналізу глобальних інтернет-даних.

Системи цільового моніторингу:

- Carnivore/DCS1000 — система моніторингу електронної пошти;
- Stingray— пристрої для перехоплення мобільних комунікацій.

Платформи аналізу соціальних мереж:

- Palantir — платформа для аналізу великих даних та виявлення зв'язків;

– Babel Street — система моніторингу та аналізу соціальних медіа.

Системи біометричної ідентифікації:

– IDENT— система біометричної ідентифікації США;
– VoiceGrid — система розпізнавання голосу для правоохоронних органів.

Інструменти лінгвістичного аналізу:

– VADER (Valence Aware Dictionary and sEntiment Reasoner) — інструмент для аналізу настроїв;
– LIWC (Linguistic Inquiry and Word Count) — програма для аналізу тексту.

Системи виявлення аномалій:

– NIST (National Institute of Standards and Technology) Anomaly Detection — система для виявлення аномалій у великих наборах даних;
– IBM QRadar — платформа для виявлення загроз та аномалій у мережевому трафіку.

Інтегровані платформи кібербезпеки:

– IBM i2 Analyst's Notebook — платформа для аналізу та візуалізації даних;
– Splunk — платформа для аналізу машинних даних та виявлення загроз.

Варто зазначити, що детальна інформація про багато систем, що використовуються державними органами, часто є засекреченою. Крім того, технології постійно розвиваються, і нові системи та підходи з'являються регулярно.

Підходи до побудови системи безпеки, яка працює з голосовою інформацією

Є два підходи до побудови системи безпеки сфокусованій на голосовій інформації:

1. Агрегування максимально можливої інформації з максимальної можливої кількості існуючих систем та продуктів (див. табл. 1), передбачає встановлення і інтеграцію даних із різноманітних джерел, таких як аудіосистеми, системи відеоспостереження, сенсори руху, системи ідентифікації та інші засоби моніторингу. Метою є створення єдиної платформи, яка може зібрати якомога більше інформації для всебічного аналізу, виявлення аномалій, загроз та прийняття рішень у системі безпеки.

Таблиця 1

Порівняльний аналіз підходу агрегації інформації

Переваги	Недоліки
1. *Широкий спектр даних*, тому отримується найбільш повна картина того, що відбувається, оскільки система аналізує дані з різних джерел. 2. *Висока точність*, бо інтеграція даних з різних систем зменшує ймовірність помилкових тривог, оскільки система має більше інформації для аналізу. 3. *Універсальність* дозволяє використо-вувати даний підхід до різних типів загроз та середовищ.	1. *Вартість*, бо агрегування даних з різних систем вимагає значних фінансових та технічних ресурсів, але підтримка в актуальному стані ще більш дорогий процес аніж первинна інтеграція. 2. *Інтеграція* великої кількості систем може бути технічно складною і вимагати багато часу для налаштування та підтримки. 3. *Обмеження релевантних даних* за рахунок значних обмежень у кастомізації того, які саме дані і в якому форматі збираються, з якими параметрами і якими моделями. 4. *Конфіденційність*, бо деякі системи є хмарними та пропрієтарними із закритим вихідним кодом і архітектурою.

2. Створення системи під кожну конкретну проблему передбачає розробку вузькоспеціалізованих систем безпеки, що зосереджуються на вирішенні конкретних завдань або проблем. Наприклад, для захисту від голосових загроз у кол-центрі може

бути розроблена система, яка фокусується лише на розпізнаванні аномалій у голосах клієнтів або голосової ідентифікації/аутентифікації.

Цей підхід також несе в собі характеристики так званої «клаптикової» інтеграції, але в даному підході нівелюються недоліки першого підходу пов'язані з гнучкістю, швидкістю і адаптивністю системи.

Таблиця 2

Порівняльний аналіз підходу індивідуальних систем

Переваги	Недоліки
1. *Ефективність*, оскільки системи спеціально налаштовані на виявлення конкретних загроз. 2. *Нижча вартість* системи за рахунок менших затрат на розробку та впровадження, оскільки вони мають менший обсяг даних для обробки. 3. *Простота реалізації*, бо вузькоспеціа-лізовані системи простіше налаштовувати та підтримувати.	1. *Обмежена функціональність*, бо така система може бути ефективною лише в певних сценаріях і не зможе забезпечити загальний захист. 2. *Низька адаптивність* до нових загроз може потребувати додаткових налаш-тувань або навіть повної модернізації. 3. *Фрагментація даних*, бо використання різних систем для різних завдань може призвести до проблем з координацією та управлінням всією системою безпеки.

Системи аналізу намірів і загроз у голосовій інформації повинні бути реалізовані у вигляді *рекомендаційної* або *консультативної* системи. Рішення приймаючі системи та консультативно-вирішучі системи не підходять для вирішення задачі, оскільки вони можуть допускати помилки, які можуть призвести до серйозних наслідків.

Рекомендаційні системи надають користувачам можливість самостійно приймати остаточні рішення на основі аналізу даних, що знижує ризик помилкових дій з боку автоматизованої системи. Консультативні системи, в свою чергу, дозволяють отримувати поради та пропозиції щодо можливих дій, залишаючи за користувачем максимальний контроль над кінцевим вибором.

Вибір на користь таких підходів пов'язаний з необхідністю мінімізувати можливість хибних спрацьовувань, які можуть виникнути через складність інтерпретації голосових даних. Системи, що приймають остаточні рішення, можуть помилятися в критичних ситуаціях, що робить їх менш надійними у порівнянні з системами, які лише надають рекомендації або поради.

Таким чином, системи аналізу голосової інформації повинні фокусуватися на підтримці користувача в прийнятті рішень, генеруючи нотифікації, надаючи їм необхідні дані та рекомендації, але залишаючи останнє слово за людиною. Це забезпечить більшу гнучкість і знизить ризик негативних наслідків від помилкових дій системи.

Реалізація рекомендаційної системи виявлення зловмисних намірів і атак в голосових даних має включати:

- збір даних з різних джерел;
- попередню обробку та фільтрацію даних;
- застосування алгоритмів машинного навчання та NLP;
- аналіз результатів експертами;
- інтеграцію з іншими системами безпеки.

Для побудови таких систем існує ряд платформ, моделей, прототипів та програм, що працюють з голосовими даними. Основними системами та платформами для розпізнавання голосу є:

- Nuance Communications Dragon — система розпізнавання голосу, яка використовується у багатьох галузях, включаючи медицину та юриспруденцію [13];
- Google Speech-to-Text — потужна хмарна платформа для перетворення голосу в текст, що підтримує численні мови та акценти [14];
- Microsoft Azure Speech Services — хмарний сервіс, який пропонує функції розпізнавання та синтезу мовлення з можливостями персоналізації моделей для конкретних сценаріїв використання [15];
- Amazon Transcribe — сервіс від Amazon Web Services для автоматичного транскрибування голосових файлів у текст з можливістю інтеграції в різні бізнес-процеси [16];
- IBM Watson Speech-to-Text — рішення від IBM для розпізнавання мовлення, яке підтримує кілька мов і спеціалізується на інтеграції з іншими AI-сервісами Watson [17].

Також слід зазначити існуючі прототипи та інструменти для досліджень:

- Kaldi — відкрита платформа для створення систем розпізнавання мовлення, що активно використовується в наукових дослідженнях та експериментальних проєктах [18];
- DeepSpeech (Mozilla) — відкрита система розпізнавання мовлення на основі глибокого навчання, яка має за мету забезпечити доступність технологій розпізнавання голосу для всіх [19];
- Julius — система розпізнавання мовлення з відкритим вихідним кодом, яка використовується у різних проєктах з обробки мовлення [20];
- Pocketsphinx — легка система розпізнавання мовлення, що підходить для інтеграції у мобільні та вбудовані системи [21].

Окремим підвидом є біометричні системи ідентифікації за голосом:

- BioID Voice Recognition — системи біометричної ідентифікації, які використовують голосовий відбиток для підтвердження особи [22];
- Agnitio Voice ID — рішення для біометричної ідентифікації за голосом, яке використовується у безпеці та правоохоронних органах [23].

Крім того, для аналізу емоцій та інтонацій існують окремі інструменти:

- Beyond Verbal — платформа, що аналізує емоційний стан людини на основі її голосу [24];
- Affectiva — система аналізу емоцій, яка може використовувати голос для оцінки емоційного стану у реальному часі [25].

А також існують інші інструменти для обробки голосових даних, наприклад:

- VoxSigma — система для автоматичного транскрибування та аналізу аудіо-файлів, що використовується для моніторингу ЗМІ та юридичних досліджень [26];
- Speechmatics — сервіс для розпізнавання голосу, який підтримує транскрибування на кількох мовах та використовується у медіа, телекомунікаціях та інших галузях [27].

Цей перелік включає як комерційні, так і відкриті рішення, що дозволяє вибрати відповідну технологію залежно від конкретних потреб та ресурсів організації.

Приклад системи, створеної для виявлення злочинних намірів, зображений на рис. 3.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

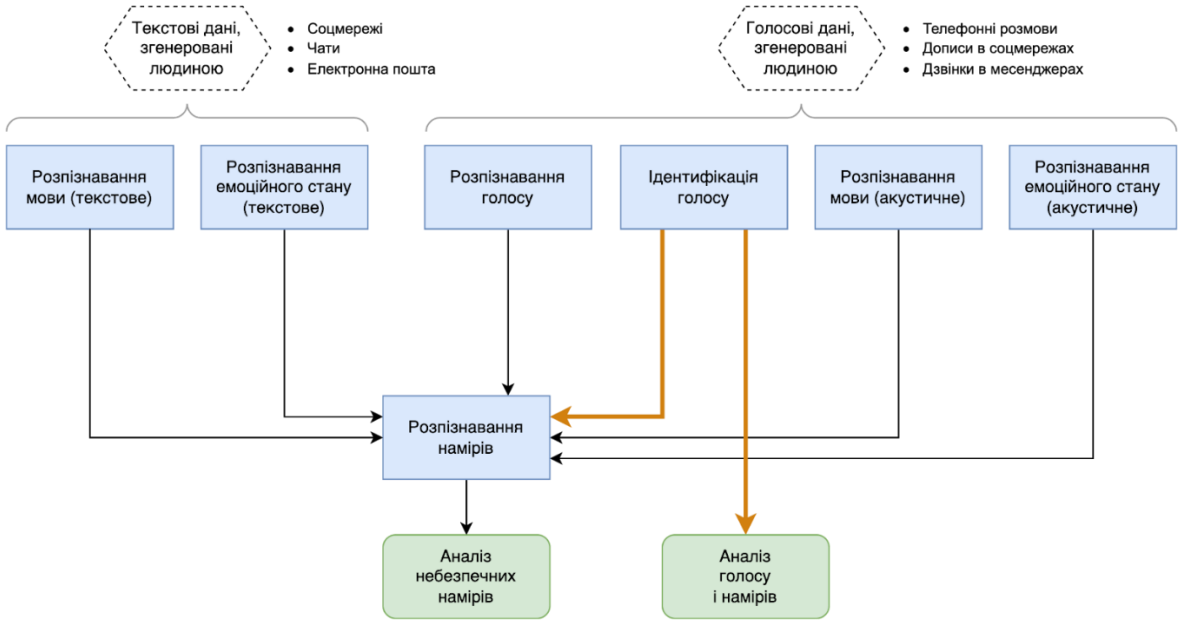

Рис. 3. Схема і порядок взаємодії систем розпізнавання голосової і текстової інформації для аналізу намірів

На рис. 3 зображено схему процесу аналізу текстових та голосових даних для виявлення небезпечних намірів. Текстові та голосові дані, згенеровані людиною, проходять через різні етапи розпізнавання: мови, емоційного стану (як текстового, так і акустичного), голосу та ідентифікації голосу [28]. На основі цих даних здійснюється розпізнавання намірів, що дозволяє проводити аналіз небезпечних намірів або голосу з додатковою оцінкою можливих загроз.

Переваги застосування методів розпізнавання голосової інформації

Технології розпізнавання голосу та NLP значно *підвищують ефективність* роботи з великими обсягами даних. Вони дозволяють швидко обробляти текстову та голосову інформацію, виконуючи це набагато швидше, ніж люди. Крім того, автоматизація рутинних завдань, таких як транскрибування дзвінків, сортування електронної пошти або аналіз звітів, сприяє значному скороченню часу, витраченого на повторювані операції.

Завдяки автоматичному аналізу комунікацій технології можуть ефективно розпізнавати загрози, такі як терористичні змови або кіберзагрози. Це дозволяє значно *підвищити рівень безпеки*. Крім того, біометрична аутентифікація на основі голосу забезпечує надійний захист від несанкціонованого доступу, що є важливим аспектом сучасної кібербезпеки.

Голосові помічники та чат-боти відіграють важливу роль у *покращенні якості обслуговування* користувачів. Вони забезпечують швидку та ефективну допомогу, знижуючи навантаження на операторів. Також NLP інтерфейси полегшують взаємодію з технологіями для людей, які стикаються з труднощами при використанні традиційних інтерфейсів.

Розпізнавання голосу дозволяє трансформувати голосові повідомлення у текст для подальшого аналізу за допомогою NLP. Це знаходить своє застосування, наприклад, у записах дзвінків у *службах підтримки клієнтів* (в чому числі IVR). Аналіз тексту на основі NLP допомагає виявляти ключові слова, настрої та тенденції, що особливо корисно для аналізу соціальних мереж та вивчення громадської думки. Чат-боти та

голосові помічники автоматизують процес спілкування з користувачами, забезпечуючи відповіді на запити та надання необхідної інформації, як це роблять такі помічники, як Siri, Alexa, Google Assistant.

Серед прикладів сучасних технологій можна відзначити Google Assistant, Amazon Alexa та Apple Siri, які використовують розпізнавання голосу та NLP для взаємодії з користувачами. IBM Watson надає інструменти для аналізу тексту та розпізнавання голосу у бізнес-додатках, а Microsoft Azure Cognitive Services пропонує набір сервісів для розробки додатків із можливостями NLP та розпізнавання мовлення.

Також важливо відзначити NLP моделі, такі як Google BERT, яка використовує трансформери для розуміння контексту слів у реченнях. OpenAI GPT є генеративною моделлю для NLP, здатною генерувати тексти та виконувати інші завдання NLP. Mozilla DeepSpeech пропонує відкриту модель для розпізнавання мовлення на основі нейронних мереж, що також є важливим інструментом у цій галузі.

Слід також зазначити фактори, що впливають на вартість кінцевої системи:

1. Інфраструктура (сервери, хмарних сервіси, обчислювальні потужності).
2. Ліцензії та програмне забезпечення (комерційні моделі та платформи).
3. Розробка, налаштування та інтеграцію технологій в існуючі системи.
4. Персонал (залучення, навчання та зарплата).
5. Безпека (програмне та апаратне забезпечення).

Зазанечені вище фактори безпосередньо впливають на вартість проєктів, яка наведена в табл. 3. Особливо великий вплив мають такі фактори як інфраструктура, розробка та налаштування, а також заходи безпеки, що є критичними для різних масштабів впровадження технологій. Наведено оцінку вартості проєктів різного розміру, пов’язаних із впровадженням технологій розпізнавання мовлення та NLP. Таблиця демонструє три категорії проєктів: малий, середній та великий, кожен з яких відрізняється сферою застосування та орієнтовною вартістю реалізації. Вартість варіюється від декількох тисяч доларів для простих чат-ботів або голосових помічників до кількох мільйонів для масштабних систем, впроваджених у державних або великих комерційних структурах.

Таблиця 3

Оцінка вартості проєктів різного розміру

Розмір проєкту	Сфера застосування	Орієнтовна вартість, тис. дол.
Малий	Впровадження простих чат-ботів або голосових помічників	Одиниці — десятки
Середній	Розробка та інтеграція систем розпізнавання мовлення та NLP для бізнес-цілей	Десятки — сотні
Великий	Впровадження масштабних систем для державних чи великих комерційних структур	Тисячі

Розпізнавання голосу та NLP є потужними інструментами, які можуть значно підвищити ефективність і безпеку держави. Однак їх впровадження потребує значних інвестицій та ретельного планування для забезпечення конфіденційності, надійності та безпеки даних. Поширені моделі, такі як BERT та GPT, надають великі можливості для аналізу та розуміння тексту, а сучасні імплементації показують високу ефективність цих технологій у різних сферах.

Обмеження реалізацій методів розпізнавання голосової інформації

Основними обмеженнями систем є в першу чергу є вузкість застосування і по друге швидкість адаптації новітніх технологій. З розвитком технологій і особливо з появою новітніх методів генерації тексту та синтезу голосу, кіберзлочинці отримали нові інструменти для здійснення більш складних та ефективних атак. Однією з найбільш небезпечних сучасних загроз є атаки, пов'язані з видаванням себе за іншу людину, також відомі як спуфінг або імперсонація.

Новітні *моделі клонування голосу* на основі глибокого навчання, такі як Deepfake Voice або інші генеративні моделі, дозволяють створювати високоточні копії голосу реальних людей. Ці моделі здатні відтворювати інтонації, манери мовлення і навіть емоції. За наявності достатньої кількості аудіозаписів голосу людини зловмисники можуть створити голосовий клон. Такі технології використовуються для атак на високопоставлених осіб або бізнес-лідерів, коли зловмисники видають себе за них у телефонних розмовах або через голосові повідомлення. Це може призвести до виконання неправомірних фінансових транзакцій, отримання конфіденційної інформації або скомпрометування корпоративної безпеки. Наприклад, відомі випадки, коли компанії зазнавали значних збитків через фальшиві дзвінки від імені керівників, які «наказували» перевести гроші на рахунки хакерів.

Поєднання технологій клонування голосу і *методів соціальної інженерії* робить атаки ще більш ефективними. Соціальна інженерія базується на психологічних маніпуляціях, коли жертва вводиться в оману з метою виконання певних дій. Використовуючи згенеровані голоси, зловмисники можуть переконати жертв у тому, що вони спілкуються з відомою їм особою, що підвищує ймовірність успішної атаки. Ці атаки можуть націлюватися не лише на фінансові транзакції, а й на доступ до конфіденційної інформації, зокрема паролів, внутрішніх документів або систем безпеки. Вони також можуть використовуватися для дезінформації та поширення фейкових новин, що створює загрози на рівні державної безпеки.

З розвитком *технологій біометричної аутентифікації* все більше систем використовують голос як засіб підтвердження особи. Проте ці системи також можуть бути вразливими до атак з використанням згенерованих голосів. Голосові паролі, які раніше вважалися надійним засобом захисту, тепер можуть бути обмануті за допомогою технологій клонування голосу. Хакери можуть отримати доступ до облікових записів, банківських рахунків або інших критичних ресурсів, використовуючи підроблені голоси. Це ставить під загрозу не лише окремих користувачів, а й цілу інфраструктуру компаній та державних організацій, де голосова аутентифікація використовується для доступу до конфіденційних даних або систем.

Для захисту від таких атак необхідно вживати комплексні заходи:

- підвищення обізнаності персоналу через освітні програми для співробітників та користувачів для вчасного розпізнавання потенційних загроз і повідомлення про підозрілі випадки;
- мультифакторна аутентифікація, включаючи фізичні фактори, такі як токени або біометричні дані інших типів (відбитки пальців, розпізнавання обличчя і голосу);
- використання штучного інтелекту для виявлення фальсифікацій для аналізу голосових команд і виявлення аномалій або ознак підробки;
- постійне оновлення систем захисту для протидії новим типам атак.

Для успішної протидії таким загрозам необхідні інноваційні підходи до захисту та постійний розвиток систем кібербезпеки.

Ризики застосування методів розпізнавання голосової інформації

Технології розпізнавання голосу та NLP мають великий потенціал для підвищення ефективності державної безпеки. Однак їх використання вимагає уважного підходу до питань конфіденційності, безпеки даних, надійності, етики та захисту від зловживань. Для запобігання ризикам, пов'язаним із цими технологіями, держави повинні впроваджувати чіткі регуляції та забезпечувати високий рівень захисту.

Ключовими аспектами викликів та ризиків якими супроводжується використання технологій розпізнавання голосу та NLP є:

1. *Надійність та точність*. Технології розпізнавання голосу та NLP можуть допускати помилки, що впливають на прийняття рішень. Неточні результати можуть призвести до неправомірних дій або пропущених загроз. Наприклад, неправильна ідентифікація голосу підозрюваного може призвести до хибних арештів або пропущених терористичних атак.

2. *Наявність даних*. Одним з ключових обмежень технологій розпізнавання голосу є залежність від якості та кількості даних. Для тренування моделей розпізнавання мовлення потрібні великі обсяги даних, які включають різноманітні зразки голосу з різних джерел, мов і діалектів. Недостатність або неповнота таких даних може призвести до зниження точності системи. Також варто враховувати питання різноманітності даних. Технології можуть демонструвати нижчу ефективність при роботі з мовами або діалектами, для яких доступно мало навчальних даних. Наприклад, якщо модель навчена переважно на англомовних зразках, вона може працювати менш ефективно з іншими мовами або з користувачами з акцентами.

3. *Вартість впровадження*. Впровадження технологій розпізнавання голосу є дорогим процесом, що потребує значних фінансових ресурсів. Основними статтями витрат є: інфраструктура: (необхідність інвестувати в потужне обладнання або хмарні ресурси для обробки великих обсягів голосових даних), програмне забезпечення та ліцензії (високі витрати на придбання ліцензій для комерційних моделей та платформ, які надають функції розпізнавання голосу та NLP), розробка та інтеграція (витрати на розробку індивідуальних рішень, налаштування моделей під конкретні завдання та їх інтеграцію з існуючими системами), персонал (залучення фахівців з NLP, машинного навчання та кібербезпеки, а також їх навчання й підтримка) і кібербезпека (витрати на захист даних та створення системи безпеки, яка зможе протистояти можливим загрозам, таким як злом або витік даних). Усе це робить технології розпізнавання голосу недоступними для малих і середніх підприємств або організацій з обмеженими бюджетами а також країн, що є одним із суттєвих обмежень їх широкого впровадження.

4. *Конфіденційність та приватність*. Збір та аналіз голосових і текстових даних можуть порушувати права на приватність громадян. Неправильне використання цих технологій може призвести до незаконного стеження та втручання в особисте життя. Наприклад, використання державою технологій для моніторингу телефонних розмов без належних правових підстав може порушувати права людини.

5. *Безпека даних*. Збереження та обробка великих обсягів голосових і текстових даних потребують високого рівня захисту. Витік або злом таких даних можуть мати серйозні наслідки для державної безпеки. Зокрема, злом баз даних, які містять розмови високопосадовців, може надати доступ до чутливої інформації ворогам держави.

6. *Використання у військовій та розвідувальній діяльності*. Технології можуть бути використані для збору розвідувальних даних, але водночас можуть стати об'єктом атак з боку противників, які прагнуть дезінформувати або зламати системи. Наприклад,

противник може використовувати NLP для створення дезінформаційних кампаній або зламу голосових командних систем.

7. *Етичні питання*. Використання технологій повинно враховувати етичні аспекти, зокрема недопущення дискримінації та забезпечення прозорості у використанні даних. Важливо розробляти алгоритми, які неупереджено ставляться до всіх груп населення, без дискримінації за ознаками раси, статі чи віку.

8. *Підробка голосу* та *штучні голоси*. Розвиток технологій підробки голосу (deepfake) створює ризики для безпеки, оскільки зловмисники можуть використовувати ці технології для створення фальшивих повідомлень або команд. Наприклад, підробка голосу високопосадовця може призвести до хибних наказів або дезінформації.

Дискусійні питання та виклики щодо впровадження технологій розпізнавання голосу в системах інформаційної безпеки

Актуальність і новизна технологій розпізнавання голосу та NLP є одними з найважливіших тем у сучасному світі, особливо з огляду на зростання ролі цифрових комунікацій та необхідність забезпечення національної безпеки. Ці технології відіграють ключову роль у багатьох аспектах державного управління, комерційної діяльності та навіть повсякденного життя, оскільки вони дозволяють автоматизувати аналіз великих обсягів інформації, що надходить у вигляді голосових повідомлень або текстових даних. Це стає критично важливим для таких завдань, як збір розвідувальних даних, моніторинг громадської безпеки та забезпечення правопорядку, а також для запобігання потенційним загрозам.

Актуальність технологій розпізнавання голосу та NLP обумовлена зростаючою роллю цифрових комунікацій у сучасному суспільстві. У світі, де інформація стає основним ресурсом, здатність швидко й точно аналізувати великі обсяги даних має вирішальне значення. Технології NLP та розпізнавання голосу дозволяють значно скоротити час, необхідний для обробки інформації, підвищуючи тим самим ефективність роботи державних та приватних організацій.

Однією з найбільш актуальних сфер застосування цих технологій є національна безпека. Використання NLP для моніторингу комунікацій дозволяє виявляти потенційно небезпечні ситуації на ранніх етапах. Наприклад, аналіз текстових або голосових повідомлень може допомогти у виявленні підозрілої активності, що є важливим у контексті боротьби з тероризмом та іншими загрозами державної безпеки.

У комерційному секторі технології розпізнавання голосу й NLP також знаходять широке застосування. Вони використовуються для підвищення якості обслуговування клієнтів, автоматизації рутинних завдань, таких як сортування електронної пошти або транскрибування дзвінків, а також для покращення маркетингових стратегій через аналіз соціальних мереж і відгуків клієнтів.

Новизна технологій розпізнавання голосу та NLP полягає у швидкому розвитку і вдосконаленні методів машинного навчання, які використовуються для підвищення точності та ефективності цих технологій. Однією з найбільш значущих новацій останніх років є розвиток моделей глибокого навчання, таких як BERT від Google і GPT від OpenAI, які дозволяють значно поліпшити результати NLP.

Ці моделі здатні розуміти контекст слів у реченнях, що дозволяє їм не тільки точніше інтерпретувати текст, але й генерувати нові повідомлення, що виглядають, як створені людиною. Такий прогрес відкриває нові горизонти для автоматизації у сферах, де раніше це було неможливо. Наприклад, сучасні чат-боти і голосові помічники стали

настільки «розумними», що можуть вести складні діалоги з користувачами, забезпечуючи їм підтримку практично у будь-яких питаннях.

Крім того, новітні технології дозволяють створювати персоналізовані системи взаємодії з користувачами. Наприклад, технології розпізнавання голосу можуть навчатися на індивідуальних особливостях мовлення конкретної людини, що дозволяє створювати більш точні та ефективні системи, які здатні реагувати на особливі вимоги користувача.

Попри великий потенціал і нові можливості, які надають технології розпізнавання голосу та NLP, їх впровадження супроводжується рядом викликів. Одним з головних викликів є питання конфіденційності та безпеки даних. Збирання та обробка великих обсягів голосових і текстових даних можуть створювати ризики для приватності користувачів, що викликає занепокоєння у громадськості та потребує відповідного законодавчого регулювання.

Ще одним важливим викликом є надійність і точність технологій. Незважаючи на прогрес у галузі, помилки в розпізнаванні голосу або аналізі тексту все ще можливі, що може призвести до серйозних наслідків, особливо у сферах, де точність має критичне значення, таких як державна безпека або охорона здоров’я.

Етичні питання також стають все більш актуальними у зв’язку з розвитком цих технологій. Наприклад, алгоритми, що використовуються для аналізу даних, можуть бути упередженими щодо певних груп населення, що може призводити до дискримінації.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Актуальність і новизна галузі технологій розпізнавання голосу та NLP є очевидними, оскільки вони відкривають нові можливості для підвищення ефективності, безпеки та якості послуг у різних сферах. Проте, їх впровадження вимагає ретельного підходу, враховуючи всі можливі виклики та ризики. Держави та компанії повинні зосередитися на розробці чітких регуляцій, які захищатимуть права громадян і забезпечуватимуть надійність та етичність використання цих технологій.

Результати цього дослідження можуть бути корисними для фахівців у сфері національної безпеки, розробників технологій, а також для законодавців, які працюють над регулюванням використання цих технологій у державному секторі.

Розвиток цієї галузі продовжує набирати обертів, і майбутні інновації, безсумнівно, принесуть ще більше змін у наше життя, дозволяючи досягти нових висот у використанні технологій для вирішення найрізноманітніших завдань.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Dasgupta, S., Piplai, A., Kotal, A., & Joshi, A. (2020). A Comparative Study of Deep Learning based Named Entity Recognition Algorithms for Cybersecurity. In *2020 IEEE International Conference on Big Data*, 2596–2604. https://doi.org/10.1109/BigData50022.2020.9378482
2. Romanovskyi, O., et al. (2021). Automated Pipeline for Training Dataset Creation from Unlabeled Audios for Automatic Speech Recognition. In *Lecture Notes on Data Engineering and Communications Technologies*, 25–36. Springer International Publishing. https://doi.org/10.1007/978-3-030-80472-5_3
3. Tan, H., et al. (2022). Adversarial Attack and Defense Strategies of Speaker Recognition Systems: A Survey. *Electronics*. https://doi.org/10.3390/electronics11142183
4. Iosifova, O., Iosifov, I., Rolik, O., & Sokolov, V. (2020). Techniques Comparison for Natural Language Processing. In *Proceedings of the 2nd International Workshop on Modern Machine Learning Technologies and Data Science, I, vol. 2631*, 57–67.

5. Iosifov, I. Iosifova, O., Sokolov, V., Skladannyi, P., & Sukaylo, I. (2021). Natural Language Technology to Ensure the Safety of Speech Information. In *Proceedings of the Workshop on Cybersecurity Providing in Information and Telecommunication Systems II*, *vol. 3187(1)*, 216–226.
6. Iosifov, I., Iosifova, O., & Sokolov, V. (2020). Sentence Segmentation from Unformatted Text using Language Modeling and Sequence Labeling Approaches. In *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PICST), vol. 1*, 335–337. https://doi.org/10.1109/picst51311.2020.9468084
7. Iosifova, O., Iosifov, I., Sokolov, V., Romanovskyi, O., & Sukaylo, I. (2021). Analysis of Automatic Speech Recognition Methods. In *Proceedings of the Workshop on Cybersecurity Providing in Information and Telecommunication Systems, Vol. 2923*, 252–257.
8. Romanovskyi, O., et al. (2022). Prototyping Methodology of End-to-End Speech Analytics Software. In *Proceedings of the 4th International Workshop on Modern Machine Learning Technologies and Data Science*, *Vol. 3312*, 76–86.
9. Mahdavifar, S., & Ghorbani, A. (2019). Application of Deep Learning to Cybersecuri-ty: A Survey. *Neurocomputing*, *347*, 149–176. https://doi.org/10.1016/j.neucom.2019.02.056
10. Sedkowski, W., & Bierczyński, K. (2022). Perceived Severity of Vulnerability in Cybersecurity: Cross Linguistic Variegation. In *2022 IEEE International Carnahan Conference on Security Technology*, 1–4. https://doi.org/10.1109/iccst52959.2022.9896488
11. Mounnan, O., Manad, O., Boubchir, L., Mouatasim, A., & Daachi, B. (2022). Deep Learning-Based Speech Recognition System using Blockchain for Biometric Access Control. In *2022 9th International Conference on Software Defined Systems (SDS)*, 1–2. https://doi.org/10.1109/SDS57574.2022.10062921
12. Chen, Y., et al. (2021). SoK: A Modularized Approach to Study the Security of Automatic Speech Recognition Systems. *ACM Transactions on Privacy and Security*, *25*, 1–31. https://doi.org/10.1145/3510582
13. Poulter, C. (2020). Voice Recognition Software—Nuance Dragon Naturally Speaking. *Occupational Medicine*, *70(1)*, 75–76. https://doi.org/10.1093/occmed/kqz128
14. Wang, H. H. (2021). Speech Recorder and Translator using Google Cloud Speech-to-Text and Translation. *Journal of IT in Asia*, *9(1)*, 11–28. https://doi.org/10.33736/jita.2815.2021
15. The Cloud and Microsoft Azure Fundamentals. (2019). Microsoft Azure Infrastructure Services for Architects, *Portico*, 1–46. https://doi.org/10.1002/9781119596608.ch1
16. Chen, L., et al. (2018). IBM Watson: Cognitive Computing in Healthcare and Beyon, AI Magazine [dataset]. In *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.aws.transcribe
17. Pickering, J. (2024). Cosegmentation in the IBM Text-to-Speech System. *Speech and Hearing*. https://doi.org/10.25144/22372
18. Povey, D., et al. (2011). The Kaldi Speech Recognition Toolkit. In *IEEE Workshop on Automatic Speech Recognition and Understanding*.
19. Hannun, A., et al. (2014). Deep Speech: Scaling up end-to-end speech recognition (Version 2). *arXiv*. https://doi.org/10.48550/arXiv.1412.5567
20. Lee, A., Kawahara, T. (2009). Recent Development of Open-Source Speech Recognition Engine Julius. In *Asia-Pacific Signal and Information Processing Association, Annual Summit and Conference*, 131–137.
21. Huggins-Daines, D., et al. (2006). Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices. In *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, 1, 185–188. https://doi.org/10.1109/icassp.2006.1659988
22. Recognition of Citizens' Voice with Social Media. (2019). https://doi.org/10.4135/9781526486882
23. Agnitio Launches Voice Authentication for Android. (2012). *Biometric Technology Today*, *2012(5)*, 12. https://doi.org/10.1016/s0969-4765(12)70094-2
24. Beyond the Standard Model of Verbal Probing. (2005). *Cognitive Interviewing*, 87–101. https://doi.org/10.4135/9781412983655.n6
25. Kulke, L., Feyerabend, D., & Schacht, A. (2020). A Comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for Identifying Facial Expressions of Emotion. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00329
26. Vocapia Research SAS. (2024). *VoxSigma Speech to Text Software Suite*. https://www.vocapia.com/voxsigma-speech-totext.html
27. Ash, T., Francis, R., & Williams, W. (2018). The Speechmatics Parallel Corpus Filtering System for WMT18. In *Proceedings of the 3rd Conference on Machine Translation: Shared Task Papers*, 853–859. https://doi.org/10.18653/v1/w18-6472
28. Iosifov, I., Iosifova, O., Romanovskyi, O., Sokolov, V., & Sukailo, I. (2022). Transferability Evaluation of Speech Emotion Recognition Between Different Languages. In *Lecture Notes on Data Engineering and Communications Technologies*, 413–426. https://doi.org/10.1007/978-3-031-04812-8_35

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

Ievgen Iosifov
PhD. Student of Volodymyr Buriachok Department of Information and Cybersecurity
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0000-0001-6203-9945
y.iosifov.asp@kubg.edu.ua

Volodymyr Sokolov
PhD., Associate Professor
Associate Professor of Volodymyr Buriachok Department of Information and Cybersecurity
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0000-0002-9349-7946
v.sokolov@kubg.edu.ua

COMPARATIVE ANALYSIS OF METHODS, TECHNOLOGIES, SERVICES, AND PLATFORMS FOR SPEECH RECOGNITION IN INFORMATION SECURITY SYSTEMS

Abstract. The article provides a comprehensive comparative analysis of methods, technologies, and modern approaches to the use of speech recognition and natural language processing (NLP) technologies in the context of national security and information security. The key aspects of the use of technologies for monitoring communications, detecting suspicious activity and application in the field of intelligence and counterintelligence, the role in ensuring cybersecurity, the possibilities of biometric identification by voice, ethical and legal aspects, and technological challenges are considered. The problem statement focuses on the challenges associated with the widespread adoption of speech recognition and NLP technologies, in particular, the lack of accuracy of algorithms, which creates risks to the reliability of security systems. The author also emphasizes the importance of addressing ethical and legal issues related to the privacy of citizens and the possible misuse of technologies for mass surveillance. The paper provides examples of systems for cybersecurity purposes, such as mass listening and analysis systems, targeted monitoring systems, social media analysis platforms, biometric identification systems, and others. The results section of the study presents a high-level structure of threat protection systems that covers threat channels and levels of protection. The complexity of modern threats that can integrate into several channels simultaneously, in particular using voice information, is considered. The author details the place and role of voice information in the structure of threat protection, emphasizing the importance of integrating various systems and platforms to ensure comprehensive security. Two approaches to building a security system that works with voice information are considered: aggregation of the maximum possible information from existing systems and creation of a system for each specific problem. A comparative analysis of these approaches is carried out, their advantages and disadvantages are identified, and the limitations and risks of using voice recognition methods are described, including the reliability and accuracy of technologies, the availability of data for training models, the cost of implementation, issues of confidentiality and privacy, data security, use in military and intelligence activities, ethical issues, and the risks of voice fraud and artificial voices.

Keywords: Natural Language Processing; audio data; speech recognition; authentication; deep learning; machine learning; text processing; cybersecurity; information security.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Dasgupta, S., Piplai, A., Kotal, A., & Joshi, A. (2020). A Comparative Study of Deep Learning based Named Entity Recognition Algorithms for Cybersecurity. In *2020 IEEE International Conference on Big Data*, 2596–2604. https://doi.org/10.1109/BigData50022.2020.9378482
2. Romanovskyi, O., et al. (2021). Automated Pipeline for Training Dataset Creation from Unlabeled Audios for Automatic Speech Recognition. In *Lecture Notes on Data Engineering and Communications Technologies*, 25–36. Springer International Publishing. https://doi.org/10.1007/978-3-030-80472-5_3

3. Tan, H., et al. (2022). Adversarial Attack and Defense Strategies of Speaker Recognition Systems: A Survey. *Electronics*. https://doi.org/10.3390/electronics11142183
4. Iosifova, O., Iosifov, I., Rolik, O., & Sokolov, V. (2020). Techniques Comparison for Natural Language Processing. In *Proceedings of the 2nd International Workshop on Modern Machine Learning Technologies and Data Science, I, vol. 2631*, 57–67.
5. Iosifov, I. Iosifova, O., Sokolov, V., Skladannyi, P., & Sukaylo, I. (2021). Natural Language Technology to Ensure the Safety of Speech Information. In *Proceedings of the Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3187(1)*, 216–226.
6. Iosifov, I., Iosifova, O., & Sokolov, V. (2020). Sentence Segmentation from Unformatted Text using Language Modeling and Sequence Labeling Approaches. In *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PICST), vol. 1*, 335–337. https://doi.org/10.1109/picst51311.2020.9468084
7. Iosifova, O., Iosifov, I., Sokolov, V., Romanovskyi, O., & Sukaylo, I. (2021). Analysis of Automatic Speech Recognition Methods. In *Proceedings of the Workshop on Cybersecurity Providing in Information and Telecommunication Systems, Vol. 2923*, 252–257.
8. Romanovskyi, O., et al. (2022). Prototyping Methodology of End-to-End Speech Analytics Software. In *Proceedings of the 4th International Workshop on Modern Machine Learning Technologies and Data Science, Vol. 3312*, 76–86.
9. Mahdavifar, S., & Ghorbani, A. (2019). Application of Deep Learning to Cybersecuri-ty: A Survey. *Neurocomputing, 347*, 149–176. https://doi.org/10.1016/j.neucom.2019.02.056
10. Sedkowski, W., & Bierczyński, K. (2022). Perceived Severity of Vulnerability in Cybersecurity: Cross Linguistic Variegation. In *2022 IEEE International Carnahan Conference on Security Technology*, 1–4. https://doi.org/10.1109/iccst52959.2022.9896488
11. Mounnan, O., Manad, O., Boubchir, L., Mouatasim, A., & Daachi, B. (2022). Deep Learning-Based Speech Recognition System using Blockchain for Biometric Access Control. In *2022 9th International Conference on Software Defined Systems (SDS)*, 1–2. https://doi.org/10.1109/SDS57574.2022.10062921
12. Chen, Y., et al. (2021). SoK: A Modularized Approach to Study the Security of Automatic Speech Recognition Systems. *ACM Transactions on Privacy and Security, 25*, 1–31. https://doi.org/10.1145/3510582
13. Poulter, C. (2020). Voice Recognition Software—Nuance Dragon Naturally Speaking. *Occupational Medicine, 70(1),* 75–76. https://doi.org/10.1093/occmed/kqz128
14. Wang, H. H. (2021). Speech Recorder and Translator using Google Cloud Speech-to-Text and Translation. *Journal of IT in Asia, 9(1),* 11–28. https://doi.org/10.33736/jita.2815.2021
15. The Cloud and Microsoft Azure Fundamentals. (2019). Microsoft Azure Infrastructure Services for Architects, *Portico*, 1–46. https://doi.org/10.1002/9781119596608.ch1
16. Chen, L., et al. (2018). IBM Watson: Cognitive Computing in Healthcare and Beyon, AI Magazine [dataset]. In *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.aws.transcribe
17. Pickering, J. (2024). Cosegmentation in the IBM Text-to-Speech System. *Speech and Hearing*. https://doi.org/10.25144/22372
18. Povey, D., et al. (2011). The Kaldi Speech Recognition Toolkit. In *IEEE Workshop on Automatic Speech Recognition and Understanding*.
19. Hannun, A., et al. (2014). Deep Speech: Scaling up end-to-end speech recognition (Version 2). *arXiv*. https://doi.org/10.48550/arXiv.1412.5567
20. Lee, A., Kawahara, T. (2009). Recent Development of Open-Source Speech Recognition Engine Julius. In *Asia-Pacific Signal and Information Processing Association, Annual Summit and Conference*, 131–137.
21. Huggins-Daines, D., et al. (2006). Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices. In *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, 1, 185–188. https://doi.org/10.1109/icassp.2006.1659988
22. Recognition of Citizens' Voice with Social Media. (2019). https://doi.org/10.4135/9781526486882
23. Agnitio Launches Voice Authentication for Android. (2012). *Biometric Technology Today, 2012(5),* 12. https://doi.org/10.1016/s0969-4765(12)70094-2
24. Beyond the Standard Model of Verbal Probing. (2005). *Cognitive Interviewing,* 87–101. https://doi.org/10.4135/9781412983655.n6
25. Kulke, L., Feyerabend, D., & Schacht, A. (2020). A Comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for Identifying Facial Expressions of Emotion. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.00329
26. Vocapia Research SAS. (2024). *VoxSigma Speech to Text Software Suite.* https://www.vocapia.com/voxsigma-speech-totext.html

27. Ash, T., Francis, R., & Williams, W. (2018). The Speechmatics Parallel Corpus Filtering System for WMT18. In *Proceedings of the 3rd Conference on Machine Translation: Shared Task Papers*, 853–859. https://doi.org/10.18653/v1/w18-6472
28. Iosifov, I., Iosifova, O., Romanovskyi, O., Sokolov, V., & Sukailo, I. (2022). Transferability Evaluation of Speech Emotion Recognition Between Different Languages. In *Lecture Notes on Data Engineering and Communications Technologies*, 413–426. https://doi.org/10.1007/978-3-031-04812-8_35