

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет інформаційної безпеки та електронних комунікацій
(повне найменування факультету)

Кафедра інформаційної безпеки та наноелектроніки
(повне найменування кафедри)

Пояснювальна записка
до дипломного проєкту (роботи)
магістр
(ступінь вищої освіти)

на тему Аналіз систем захисту мовної інформації
(назва теми)

Виконав(ла): студент(ка) IIм курсу,
групи БК-712м
Спеціальності 125 Кібербезпека
(код і найменування спеціальності)

Освітня програма (спеціалізація)
Системи технічного захисту інформації,
автоматизація її обробки
АНАРМЕТОВА В.К.
(ПРИЗВИЩЕ та ініціали)

Керівник ЛІЗУНОВ С.І.
(ПРИЗВИЩЕ та ініціали)

Рецензент МОРОЗ Г.В.
(ПРИЗВИЩЕ та ініціали)

2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет інформаційної безпеки та електронних комунікацій

Кафедра інформаційної безпеки та наноелектроніки

Ступінь вищої освіти магістр

Спеціальність 125 Кібербезпека

(код і найменування)

Освітня програма (спеціалізація) Системи технічного захисту інформації, автоматизація її обробки

(назва освітньої програми (спеціалізації))

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри ІБтаН

Андрій КОРОТУН

«___» _____ 2023 року

З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТА(КИ)

АНАРМЕТОВОЇ Вікторії Костянтинівни

(ПРИЗВИЩЕ, ім'я, по батькові)

1. Тема проєкту (роботи) Аналіз систем захисту мовної інформації

Analysis of data protection systems

керівник проєкту (роботи) к.т.н., доцент, ЛІЗУНОВ Сергій Іванович,

(науковий ступінь, вчене звання, ПРИЗВИЩЕ, ім'я, по батькові)

затверджені наказом закладу вищої освіти від «28» листопада 2023 року №473

2. Строк подання студентом проєкту (роботи) 11.12.2023

3. Вихідні дані до проєкту (роботи) категоризація загроз конфіденційності та класифікація методів захисту мовленнєвої інформації

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) загрози конфіденційності мовленнєвої інформації, методи захисту мовленнєвої інформації, оцінка конфіденційності

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, кількість слайдів, плакатів)

Презентація доповіді (в MS PowerPoint), 11 слайдів.

6. Консультанти розділів проєкту (роботи)

Розділ	ПРИЗВИЩЕ, ініціали та посада консультанта	Підпис, дата	
		завдання видав	прийняв виконане завдання
1 – 3	ЛІЗУНОВ С.І., доцент кафедри ІБтаН	04.09.2023	05.12.2023
Нормоконтроль	КОРОЛЬКОВ Р. Ю., доцент кафедри ІБтаН		08.12.2023

7. Дата видачі завдання «04» вересня 2023 року.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проєкту (роботи)	Строк виконання етапів проєкту (роботи)	Примітка
1.	Збір та аналіз інформації про загрози конфіденційності мовленнєвої інформації	04.09.23 – 18.09.23	виконано
2.	Систематизація літературних джерел	19.09.23 – 24.09.23	виконано
3.	Складання і затвердження наукового завдання	25.09.23 – 27.09.23	виконано
4.	Формування та уточнення наукового завдання	27.09.23 – 28.09.23	виконано
5.	Класифікація атак на мовленнєву інформацію	29.09.23 – 15.10.23	виконано
6.	Оцінка конфіденційності	16.10.23 – 31.10.23	виконано
7.	Оформлення графічної частини	01.11.23 – 19.11.23	виконано
8.	Оформлення ПЗ	20.11.23 – 30.11.23	виконано

Студент(ка)

(підпис) Вікторія АНАРМЕТОВА
(Ім'я ПРИЗВИЩЕ)

Керівник проєкту (роботи)

(підпис) Сергій ЛІЗУНОВ
(Ім'я ПРИЗВИЩЕ)

АНОТАЦІЯ

Пояснювальна записка до магістерської роботи: 69 с., 20 рис., 159 джерел,.

АТАКИ, БЕЗПЕКА, ЗАГРОЗИ, ЗАХИСТ ІНФОРМАЦІЇ,
КОНФІДЕНЦІЙНІСТЬ, МОВЛЕННЄВІ ТЕХНОЛОГІЇ

Мета роботи: описання методів захисту конфіденційності мовленнєвої інформації.

Об'єкт та предмет дослідження: об'єктом дослідження виступає мовленнєва інформація; предметом дослідження є класифікація загроз та методів захисту цієї інформації.

Методи дослідження: описово-аналітичний.

Результати: результатом дослідження є вичерпна категоризація загроз конфіденційності та класифікація методів захисту відповідно до того, чи стосуються вони приватного повідомлення чи додаткової інформації.

Рекомендації щодо впровадження: робота носить фундаментальний характер, проте деякі її результати можуть бути використані при розробці нових методів захисту мовленнєвої інформації.

Практична цінність: представлені напрямки для подальшого розвитку, у яких удосконалення методів захисту є найбільш важливими.

ABSTRACT

Explanatory note to the master's thesis: 69 p., 20 figure, 159 sources.

**ATTACKS, COMMUNICATION TECHNOLOGIES, INFORMATION
PROTECTION, PRIVACY, SECURITY, THREATS**

The purpose of the work: description of methods of protecting the confidentiality of speech information.

Object and subject of research: the object of research is speech information; the subject of the study is the classification of threats and methods of protecting this information.

Research methods: descriptive and analytical.

Results: The result of the study is a comprehensive categorization of privacy threats and a classification of protection methods according to whether they concern a private message or additional information.

Recommendations for implementation: the work is of a fundamental nature, but some of its results can be used in the development of new methods of protecting speech information.

Practical value: directions for further development are presented, in which the improvement of protection methods is most important.

ЗМІСТ

	С.
Вступ	7
1 Загрози конфіденційності мовленнєвої інформації	9
1.1 Форми мовленнєвої інформації	9
1.2 Класифікація атак на мовленнєву інформацію	11
2 Методи захисту мовленнєвої інформації	21
2.1 Зменшення додаткової інформації	21
2.2 Покращення продуктивності	27
2.3 Обмеження доступу до особистих повідомлень	29
2.4 Обмеження доступу до відтвореного аудіо	33
3 Оцінка конфіденційності	36
3.1 Об'єктивні показники	36
3.2 Диференціальна конфіденційність	38
3.3 Сприйняття і психологія конфіденційності	41
3.4 Наслідки конфіденційності в мовленнєвих технологіях	44
3.5 Правовий та соціальний аспекти	46
Висновки	50
Перелік джерел посилання	52

ВСТУП

За останні роки якість і використання мовних інтерфейсів стрімко зросли. Як і з будь-якою новою технологією, швидкий прогрес також виявив небезпеки, зокрема загрози конфіденційності, які вона очевидно створює. Незахищені користувачі піддаються таким загрозам, як переслідування, алгоритмічні стереотипи, переслідування та завищення цін. Таким чином, дослідники, постачальники послуг і уряди мають стимул захищати користувачів не тільки тому, що це етично, але й тому, що це сприяє кращим продуктам і довгостроковому бізнесу.

Мовлення є способом спілкування і, таким чином, за своєю суттю містить велику кількість інформації. Мова містить також широкий спектр побічної інформації, як-от стан здоров'я та емоції, а також фізичну, психологічну та соціальну ідентичність, більшість з яких є приватною інформацією. З точки зору інформаційного вмісту конфіденційність є багатогранним питанням у всіх сферах мовленнєвих технологій.

Вивчення та покращення конфіденційності в мовленнєвих технологіях є важливим, оскільки порушення конфіденційності можуть мати серйозні наслідки. Загальний шаблон рішення зазвичай завжди однаковий: мінімізація передачі та зберігання, а також доступу до конфіденційної інформації, яка не має відношення до послуги, яку бажають користувачі. Тоді це «лише» питання про те, як досягається така мінімізація, яка інформація є релевантною та як визначити, що хоче користувач (надати інформацію та забезпечити контроль над службами та загрозами).

Основними сферами використання мовних технологій є телекомунікації та інтерфейси людина-комп'ютер. Що стосується телекомунікацій, то підходи до етичних, правових і технологічних питань, пов'язаних із конфіденційністю в телекомунікаціях через стаціонарні телефони, є усталеними, і відкриті дискусії в першу чергу стосуються того,

як існуюче регулювання та нагляд слід розширити, щоб охопити також мобільний зв'язок і голосовий зв'язок, протоколи через Інтернет (VoIP) [5–7]. Таким чином, основна увага в роботі приділяється інтерфейсам людина-комп'ютер, де комп'ютер обробляє мовні сигнали, але загрози комунікаційним програмам також розглядаються, де це важливо. Проте тут фокус обмежується акустичним мовленнєвим сигналом, також відомим як голос, оскільки обробка природної мови є окремою та значною мірою незалежною областю, яка потребує окремої обробки (наприклад, [8]).

Конфіденційність тісно пов'язана з проблемами безпеки, і часто їх важко розрізнити. Тут ми встановлюємо баланс, розглядаючи сценарії конфіденційності, коли агент має законний доступ до деяких особистих мовних даних користувача, але використовує їх для цілей, які суперечать або отримують доступ поза межами очікувань або вподобань користувачів. Наприклад, голосовий інтерфейс можна використовувати для керування домашньою автоматизацією, але якщо постачальник послуг ділиться цими голосовими даними також з рекламодавцями всупереч уподобанням користувачів, то це є порушенням конфіденційності. Отже, тут також виключаються проблеми безпеки, такі як підробка ідентифікаційної інформації, глибокі фейки, а також атаки на пристрої чи мережі, оскільки ці загрози більше пов'язані з безпекою, а не з конфіденційністю.

Останні роботи в цій галузі включають технічний огляд [9], а також науково-популярний огляд [10] цієї галузі, а кілька докторських дисертацій мають відповідні резюме [11–14]. Конфіденційність широко обговорюється в інших галузях науки, наприклад, у сусідній галузі обробки природної мови [8], у статистичній теорії приватності [15], у соціальних науках [16] і психології [17]. Існує також ретельне мета-дослідження всіх галузей науки, які обговорюють конфіденційність [18].

З огляду на вищенаведене, тема роботи є актуальною.

1 ЗАГРОЗИ КОНФІДЕНЦІЙНОСТІ МОВЛЕННЄВОЇ ІНФОРМАЦІЇ

1.1 Форми мовленнєвої інформації

Мовленнєва інформація може бути представлена у двох формах (див. рис. 1.1): по-перше, коли передається приватне повідомлення, воно становить загрозу конфіденційності, коли всупереч уподобанням користувача воно використовується несподіваним чином або передається третій особі. По-друге, оскільки мова містить широкий спектр приватної інформації, яка складно згрупована в акустичний сигнал, нескладно витягнути лише потрібне повідомлення. Таким чином, зазвичай мовна інформація завжди містить супутню побічну інформацію, об'єднану в приватне повідомлення. Це загроза конфіденційності, коли така приватна побічна інформація передається небажаному одержувачу разом із приватним повідомленням (позначено червоними стрілками та знаками оклику). Згідно з визначенням, припускаємо, що законному одержувачу можна довіряти побічну інформацію та що будь-яке її небажане використання позначається як небажана послуга.

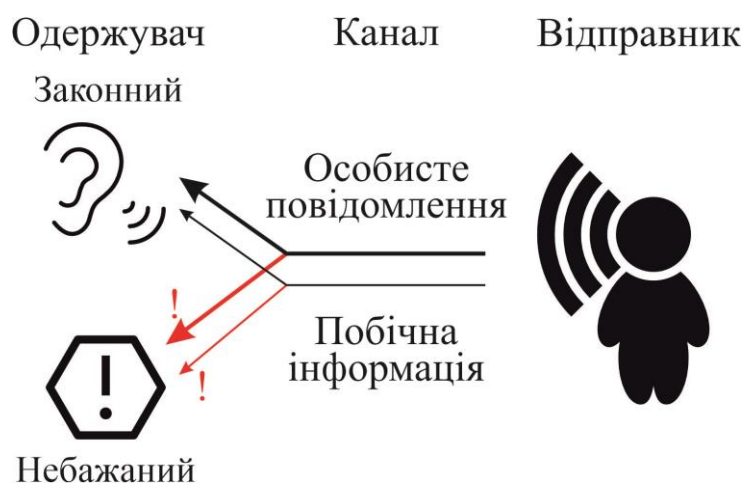


Рисунок 1.1 – Модель високорівневої загрози мовленнєвої взаємодії, де приватне повідомлення надсилається через канал законному одержувачу, але супутня побічна інформація додається до цього повідомлення

Різниця між двома формами інформації робить важливу різницю в підходах до пом'якшення. Коли приватне повідомлення розкрито, єдині доступні рішення – це або знизити точність, або обмежити доступ до приватного повідомлення, наприклад, за допомогою криптографії. Проте, коли розкривається побічна інформація, як додаткові методи також можна використовувати обробку сигналу для кращого видалення, заміни або спотворення такої побічної інформації.

Зауважимо, що рівень загрози не залежить від маршруту інформації, якщо вміст, який надходить до небажаного одержувача, однаковий. Проте поверхня атаки або канал, через який відбувається витік інформації небажаному одержувачу, має великий вплив на вибір пом'якшення. Тип небажаного одержувача, будь то послуга, пристрій чи особа, впливає головним чином на його потенційну здатність витягувати та використовувати особисту інформацію в підлий спосіб. Основна відмінність у типі відправника, особи, пристрою чи служби полягає в їхній різній здатності видаляти побічну інформацію з мовного сигналу перед передачею.

Загрози конфіденційності спікерів можна розділити на два типи; висновок властивостей/атрибутів і повторна ідентифікація. Різниця полягає в тому, що враховуючи одного відомого мовця, ми можемо пов'язати нову приватну інформацію чи атрибути з цим мовцем за допомогою висновку про властивості. Для порівняння, аналізуючи (анонімні) мовленнєві дані невідомого мовця, ми можемо визначити особу найбільш ймовірного мовця в базі даних багатьох мовців. Таким чином, єдина відмінність полягає в тому, що при висновку про властивості ми призначаємо атрибути одному мовцю, тоді як при повторній ідентифікації ми розглядаємо багато мовців і призначаємо атрибути одному (або підмножині) з них. Однак, якщо ми розглядаємо фізичну ідентичність (наприклад, ім'я в паспорті) як властивість або атрибут мовця, то повторна ідентифікація означає, що ми змогли зробити висновок про властивість мовця. У цьому сенсі повторне ототожнення є одним із конкретних типів висновку про властивість.

1.2 Класифікація атак на мовленнєву інформацію

Атаки можна класифікувати відповідно до кількості інформації, доступної для зловмисника, такої як інформація про спікерів і про навчені моделі. Доступ до навчальних даних, а також зразків мовлення певного користувача (відомих як дані про реєстрацію) є особливо корисним для зловмисника. Наприклад, у завданні анонімізації під час конкурсу VoicePrivacy 2022 метою було замінити (псевдонімізувати) особу мовця, але зберегти всі інші мовленнєві характеристики, наприклад мовну інформацію. Анонімні речення відомі як пробні висловлювання. Тоді сценарії атаки були класифіковані як [19]:

- 1) Незахищений: анонімізація не виконується; зловмисники мають доступ до оригінальних даних випробування та реєстрації.
- 2) Неосвічений зловмисник: пробні дані анонімні, але зловмисники не знають про це, тому вони використовують вихідні дані для реєстрації.
- 3) Лінива поінформованість: дані про випробування та реєстрацію доступні зловмисникам, але анонімні за допомогою іншого псевдодоповідача.
- 4) Напівінформований: як і у випадку з ледачим інформуванням, зловмисники також можуть навчити свою модель за допомогою тієї самої системи анонімізації, але з різними псевдомовцями.

Очевидно, що зі збільшенням інформації завдання зловмисника стає легшим, а точність моделей зловмисника підвищується. Крім того, такі сценарії атак можуть бути розроблені відповідно до конкретного випадку використання. Наприклад, якщо завдання полягає в анонімізації емоційного стану або стану здоров'я, ми можемо визначити різні сценарії атаки залежно від того, наскільки зловмисник володіє інформацією про категорії емоцій і стану здоров'я, які використовуються в анонімізації.

Рис. 1.2 ілюструє модель атаки для вимірювання ступеня конфіденційності [20]. Ми припускаємо, що є деякі приватні дані, які анонімні, щоб їх можна було використовувати в довіреному завданні. Зловмисник має доступ до анонімних даних, які тут називаються «публічними даними». Зауважте, що загальнодоступні дані не обов'язково є відкрито доступними для будь-кого, але це лише вказує на те, що дані достатньо вільно доступні, щоб зловмисник мав до них доступ. Зловмисник також має доступ до деяких інших даних про спікерів, знайдених з іншого джерела (знайдені дані), що допомагає отримати приватну інформацію. Знову ж таки, термін «знайдені дані» є вільним терміном і позначає будь-які дані, доступні зловмиснику. Зловмисник використовує будь-які доступні (знайдені) дані та анонімні загальнодоступні дані, щоб отримати приватну інформацію всупереч уподобанням користувачів. Анонімний потік даних позначається пунктирними лініями, а атака – червоними лініями та знаком оклику.

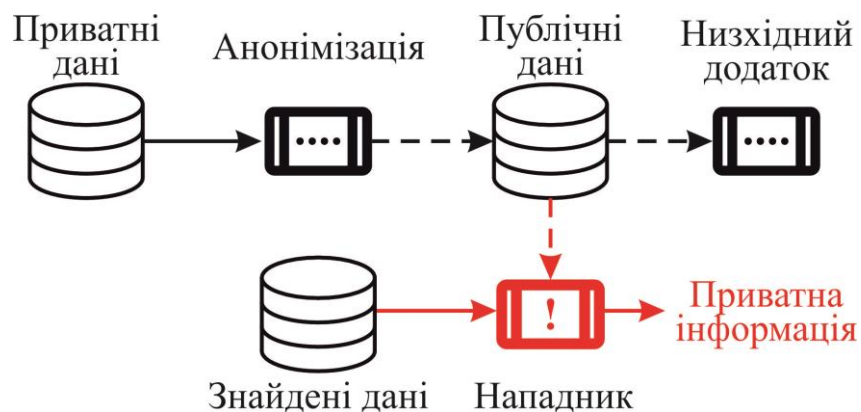


Рисунок 1.2 – Модель атаки для оцінки анонімізації із збереженням конфіденційності, коли приватні дані анонімізуються для видалення приватної інформації, а анонімні дані публікуються

Класифікуємо сценарії відповідно до поверхні атаки, звідки витягується інформація (див. рис. 1.3); канал між хмарними службами, інтерфейс користувача до периферійного пристрою або до хмарної служби,

від локальної мережі, акустичний або через спільний інтерфейс користувача локального пристрою. Розглядаємо лише випадки, коли частина технології отримує або передає інформацію, і виключаємо спілкування між людьми без пристроїв. Також припускаємо, що пристрої та мережеві з'єднання захищені таким чином, що лише авторизовані служби можуть спілкуватися з ними. Таким чином, загрози, на яких ми зосереджуємося, проілюстровані пронумерованими суцільними червоними лініями на рис. 1.3.

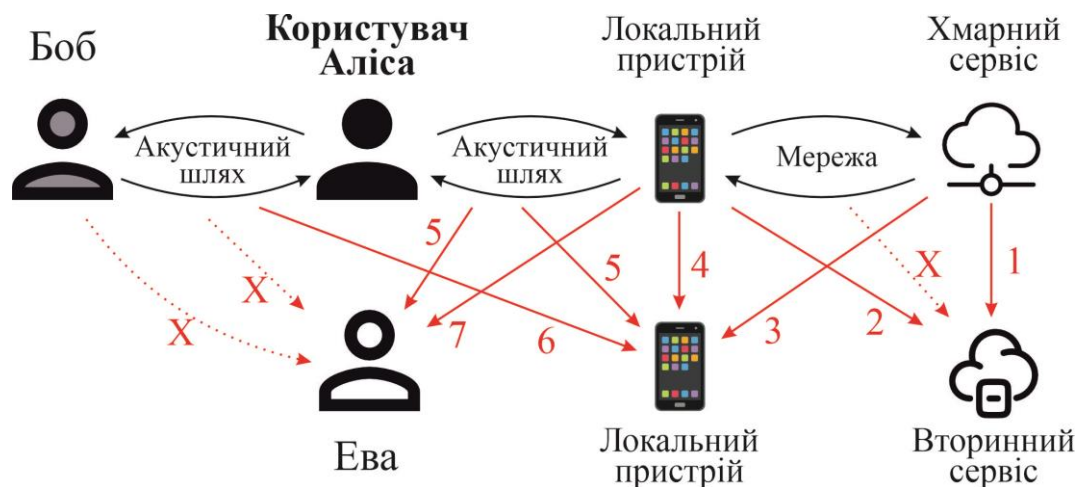


Рисунок 1.3. – Загрози конфіденційності користувача Аліси під час розмови з іншою особою Бобом або локальним пристроєм, але коли інформація витікає або передається через акустичні шляхи, периферійний пристрій, мережу або хмарний сервіс іншій особі Єва, пристрій або послуга, що суперечить уподобанням Аліси (червоні стрілки). Кожна загроза позначена номером. Пунктирні лінії, позначені «X», вказують на загрози поза поточним діапазоном і не розглядаються тут

1) Витік у хмару: припустімо, що користувач отримує доступ до віддаленої хмарної служби через периферійний пристрій (див. рис. 1.4). Таким чином, хмарний сервіс має законний доступ до приватного повідомлення користувача. Однак хмарний сервер може потім використовувати це приватне повідомлення або пакетну додаткову інформацію для інших цілей, ніж ті, що запитує користувач, витягти більше

інформації, ніж очікувалося, поєднати її з іншою інформацією або поділитися інформацією з третьою стороною. Наприклад, хмарний сервер голосового помічника може неналежним чином ділитися інформацією з рекламодавцем.

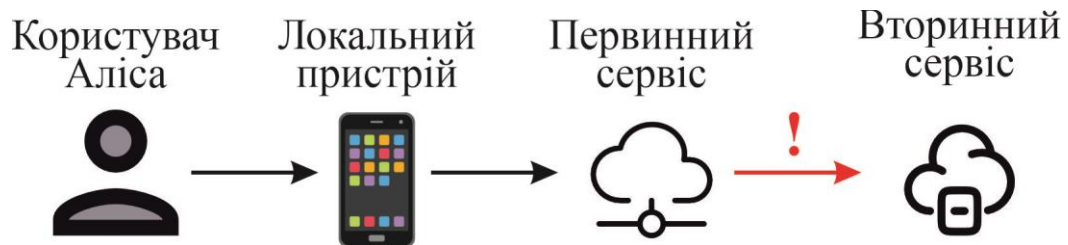


Рисунок 1.4 – Сценарій загрози «Витік хмари», коли користувач Аліса отримує доступ до (основної) хмарної служби за допомогою периферійного пристрою, але інформація передається до вторинної служби всупереч уподобанням (червона стрілка та знак оклику)

2) Помилкова активація: пристрої з мовними інтерфейсами можуть чути всі розмови в тому самому акустичному просторі, що й пристрій, і тому потребують механізмів для визначення того, які висловлювання призначені для мовного інтерфейсу. Популярним підходом є використання певного висловлювання, відомого як слово пробудження, для початку всіх взаємодій з інтерфейсом [21]. Тоді слово пробудження схоже на елементарний пароль, який запобігає активації інтерфейсу, коли мова не спрямована на пристрій. На жаль, розробка детекторів слів пробудження нетривіальна, і вони час від часу допускать помилки. Іноді вони можуть пропускати слово пробудження, коли воно промовляється (хибнонегативне) або помилково сприймати інший непов'язаний звук як слово пробудження (хибнопозитивне). Хоча хибні спрацьовування дратують користувача, коли служба не активується, хибні спрацьовування потенційно становлять серйозну загрозу конфіденційності. У деяких відомих випадках мовні інтерфейси активувалися зі звуків на телебаченні, щоб купувати небажані

предмети за вказівкою користувачів, а приватні розмови користувачів передавалися третім особам [22, 23].

3) Доступ до хмари: рис. 1.5 ілюструє загрозу, коли користувач Аліса отримує доступ до хмарної служби через периферійний пристрій і де зберігається деяка особиста інформація Аліси. Потім другий користувач Єва може отримати доступ до тієї самої хмарної служби через інший пристрій і потенційно отримати доступ до збереженої приватної інформації, всупереч уподобанням Аліси (рис. 1.6). Ця поверхня атаки подібна до сценарію Cloud Leak, головною відмінністю якого є одержувач, який тут є особою, тоді як у сценарії Cloud Leak це автоматизований агент.



Рисунок 1.5 – Сценарій загрози «Помилкова активація», коли користувач Аліса отримує доступ до периферійного пристрою, але інформація передається до хмарної служби всупереч налаштуванням (червона стрілка та знак оклику)



Рисунок 1.6. – Сценарій загрози «Доступ до хмари», де користувач Аліса використовує хмарний сервіс через периферійний пристрій, де він

потенційно зберігається та надається спільно з користувачем Євою, всупереч уподобанням Аліси (червона стрілка та знак оклику)

Прикладом цієї загрози є медичні записи, які можуть бути зібрані від пацієнтів до централізованої бази даних. Дослідник, який має дозвіл на доступ до бази даних, може потенційно отримати приватну інформацію понад очікувану [24].

4) Автентифікація WASN: Якість аудіо вловлювання мовлення, а також зручність використання голосових інтерфейсів можна покращити, використовуючи всі доступні підключені пристрої з мікрофонами, які знаходяться в одній кімнаті чи акустичному просторі [25–27]. Така співпраця може бути реалізована за допомогою акустичних сенсорних мереж, де кілька незалежних пристроїв одночасно вловлюють мову, і ці канали об'єднуються для отримання високоякісного сигналу (рис. 1.7) [11, 13]. Делікатним питанням є авторизація; Яким пристроям дозволено обмінюватися інформацією один з одним? Один підхід полягає в припущенні, що пристрої, які знаходяться в одній кімнаті або акустичному просторі, можуть чути той самий сигнал [13, 28]. Таким чином, присутність в одному просторі вже є неявним дозволом на участь у спільному прийомі сигналу. Щоб захистити конфіденційність, необхідно визначити, які пристрої знаходяться в одному акустичному просторі. Пристрої в іншій кімнаті можуть належати одній компанії або родині, і вони можуть бути підключені до однієї мережі, але все одно вони знаходяться поза сферою поточного обговорення. Без відповідних механізмів авторизації пристрої за межами кімнати можуть отримати доступ до приватної мови всередині кімнати.

5) Витоки мовного інтерфейсу та обговорення. Рисунок 1.8 ілюструє сценарій, коли пристрій або особа підслуховує обговорення між користувачем Алісою та локальним пристроєм. Так само на рис. 1.9 зображено дискусію між двома особами, підслухану за допомогою локального пристрою. Визначальною властивістю обох сценаріїв є витік на

акустичному шляху, що вимагає, щоб підслухувач був фізично присутній у тому самому акустичному просторі, що й приватне спілкування. Різниця між цими витоками та помилковою активацією (рис. 1.5) полягає в тому, що тут перехоплювач збирає дані неналежним чином, тоді як у помилковій активації активується неправильна функція.

Є чотири різні випадки, які слід розглянути залежно від того, чия мова чується, Аліса (або Боб) чи локальний пристрій, і хто є підслухувачем, Єва чи інший локальний пристрій. У випадку, коли Єва підслуховує мову Аліси, ми припускаємо, що люди мають навчене усвідомлення того, хто з інших людей присутні в тій самій кімнаті, і відповідно регулюють свою мову; Тоді або Аліса не заперечує, що Єва підслуховує її мову (несуттєва інформація), Аліса може змінити свій стиль мови на шепіт або змінити вміст так, щоб Єва не чула нічого приватного (зменшена передача інформації), або вона може піти до іншої кімнати для продовження взаємодії приватно (змінений акустичний канал).



Рисунок 1.7 – Сценарій загрози «Автентифікація WASN», коли користувач використовує послугу через розподілену сенсорну мережу всередині кімнати, але інший пристрій, поза кімнатою в тій самій мережі, приєднується до розподіленої сенсорної мережі та ділиться інформацією всупереч уподобанням (червона стрілка і знак оклику)



Рисунок 1.8 – Сценарій загрози «Витік мовного інтерфейсу», де користувач Аліса отримує доступ до периферійного пристрою (потенційно також включає віддалену службу), але інша місцева особа Єва або пристрій у тій самій кімнаті чи акустичному просторі підслуховує мовленнєву взаємодію всупереч налаштування (червона стрілка та знак оклику)

б) Спільний пристрій: багато практичних сценаріїв включають кілька користувачів, які спільно використовують один або кілька пристроїв. Наприклад, сім'я може поділитися розумною колонкою або телевізором, кімната для переговорів в офісі може мати смарт-пристрої, а пункти обслуговування клієнтів можуть мати спільний телефон між черговими. Кожен із цих пристроїв може з часом збирати конфіденційну інформацію про користувача(ів) і потенційно може ділитися нею з іншими користувачами (рис. 1.10). Примітно, що цей сценарій підкреслює, що витік відбувається на відстані в часі, тоді як загрози, що виникають на просторовій відстані, часто отримують більшість уваги.

Ця модель загроз чітко демонструє, що пристрої та служби, якими користуються кілька користувачів, повинні використовувати контроль доступу та керування авторизацією, якщо вони зберігають будь-яку конфіденційну інформацію. Це також, очевидно, пов'язано з багатьма загрозами безпеці – слід запобігати несанкціонованому доступу до пристроїв – але вони виходять за межі цієї роботи.

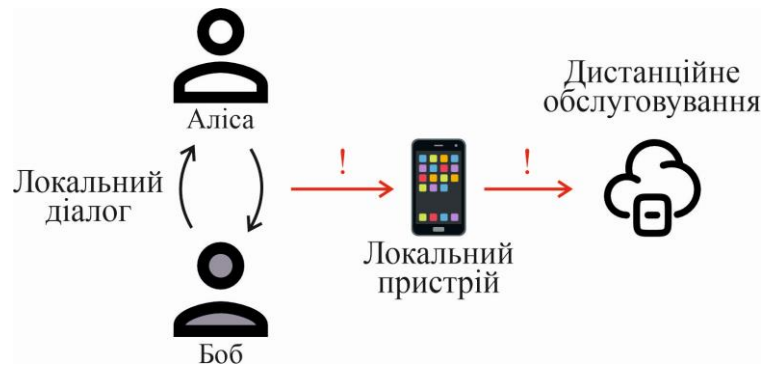


Рисунок 1.9 – Сценарій загрози «Витік обговорення», де діалог між двома користувачами, Алісою та Бобом, підхоплюється локальним пристроєм і потенційно передається віддаленій службі всупереч налаштуванням (червона стрілка та знак оклику)



Рисунок 1.10 – Сценарій загрози «Спільний пристрій», де користувач Аліса використовує периферійний пристрій, де інформація зберігається та ділиться з іншим користувачем Євою, всупереч уподобанням Аліси (червона стрілка та знак оклику)

7) Рівні довіри: Сценарій Cloud Leak на рис. 1.4 представляє собою послідовність автоматизованих агентів – периферійних, первинних і вторинних хмарних служб – де було б бажано, щоб з кожним кроком у послідовності зменшувався обсяг інформації, яка використовується спільно. Така мінімізація потенційної поверхні атаки є доцільною з точки зору управління ризиками. Однак, у той же час, можна інтерпретувати це як рівень довіри. Користувач може безпосередньо контролювати локальний акустичний простір і крайовий пристрій, і, відповідно, ми можемо очікувати, що користувач має найвищий рівень довіри до крайового пристрою. З

кожним наступним кроком у послідовності рівень контролю буде зменшуватися, а також рівень довіри.

Очевидно, що довіра є поняттям, яке залежить принаймні від психологічного, соціального та культурного контексту, тому його важко визначити. Одне з можливих визначень довіри таке. Якщо користувач Аліса деякий час спостерігав за агентом, він може сформулювати прогноз щодо поведінки агента. Якщо розподіл ймовірностей передбачення вузький, то Аліса має високу впевненість у цьому передбаченні. Тоді ми говоримо, що Аліса довіряє агенту. Таким чином, ми визначаємо, що прогноз високої впевненості дорівнює довірі. Якщо передбачення також полягає в тому, що агент поводить себе вигідно для Аліси, тоді вони сприймають агента як заслуговує довіри. Однак це залежить від здатності Аліси передбачати дії. Якщо агент дійсно поводить себе вигідно, то він об'єктивно заслуговує довіри.

2 МЕТОДИ ЗАХИСТУ МОВЛЕННЄВОЇ ІНФОРМАЦІЇ

2.1 Зменшення додаткової інформації

Основним підходом до зменшення приватної додаткової інформації в мовних сигналах є обробка сигналу. Він має дві протилежні цілі, що відповідають відповідно корисності та конфіденційності: довірене завдання обробки або аналізу, коли певна категорія інформації витягується з законною метою, і завдання захисту від загрози, коли зловмисний оператор намагається витягти приватну інформацію поза законною метою. Якість довіреного завдання зазвичай відповідає класичній методології обробки мовлення, тоді як захист від завдання загрози можна досягти шляхом видалення, заміни або спотворення приватної інформації. Головне питання полягає в тому, як інформація для довіреного завдання відокремлена від іншої приватної інформації.

Найпоширенішим підходом для захисту від цього сценарію є обмеження потоку інформації від локального периферійного пристрою до хмарного сервера (рис. 2.1). Таке обмеження або скорочення потоку інформації може бути реалізовано шляхом видалення, заміни або спотворення приватного вмісту, як показано на рис. 2.1, а. Крайній пристрій може бути повністю від'єднаний від хмари чи мережі та обробляти інформацію лише локально, як на рис. 2.1, б [29]. Роблячи це, ми повинні припустити, що локальний пристрій має достатню ємність і програмне забезпечення для виконання запитуваних завдань, що він не скомпрометований і що обмеження потоку інформації є достатнім, щоб гарантувати, що приватна інформація не витікає.



а – передача приватної інформації до хмари зменшується;

б – передача приватної інформації до хмари запобігається.

Рисунок 2.1 – Два альтернативні засоби захисту від сценарію «витоку хмари»

Два основні підходи до видалення приватної інформації ґрунтуються на інформаційному вузькому місці або так званій змагальній моделі. Інформаційні вузькі місця засновані на продавлюванні бажаної інформації через метафоричне вузьке місце, так що немає місця для проходження побічної інформації [30]. Часто такі моделі дотримуються структури автокодувальника, який навчений реконструювати вхідний сигнал із переданого сигналу.

Змагальні моделі (які не слід плутати з генеративними змагальними моделями) у свою чергу використовуються при навчанні моделей машинного навчання, щоб переконатися, що жодна приватна інформація не може бути виведена з переданої інформації. Він базується на паралельному моделюванні довірених і загрозливих завдань, але передане повідомлення оптимізується таким чином, що довірене завдання успішно виконується, а загрозливе — не вдається. Узагальнення ідеї інформаційного вузького місця полягає в тому, щоб розділити мову на кілька незалежних потоків даних. Це дозволяє програмам вибирати конфіденційну інформацію, яку слід передати в кожному випадку.

1) Інформаційне вузьке місце: щоб відокремити лише приватне повідомлення та відкинути будь-яку побічну інформацію, за допомогою підходу інформаційного вузького місця мовленнєва інформація

пропускається через вузьке місце настільки щільно, що лише законне повідомлення може пройти через [31, 32] (див. рис. 2.2). Таким чином, цей підхід забезпечує захист від зловмисника, який має доступ до вихідних даних вузького місця. Проблеми полягають у розробці вузького місця, яке є достатньо щільним, щоб приватна стороння інформація була відкинута, а модель структури та методології навчання оптимізували б якість легітимного повідомлення.

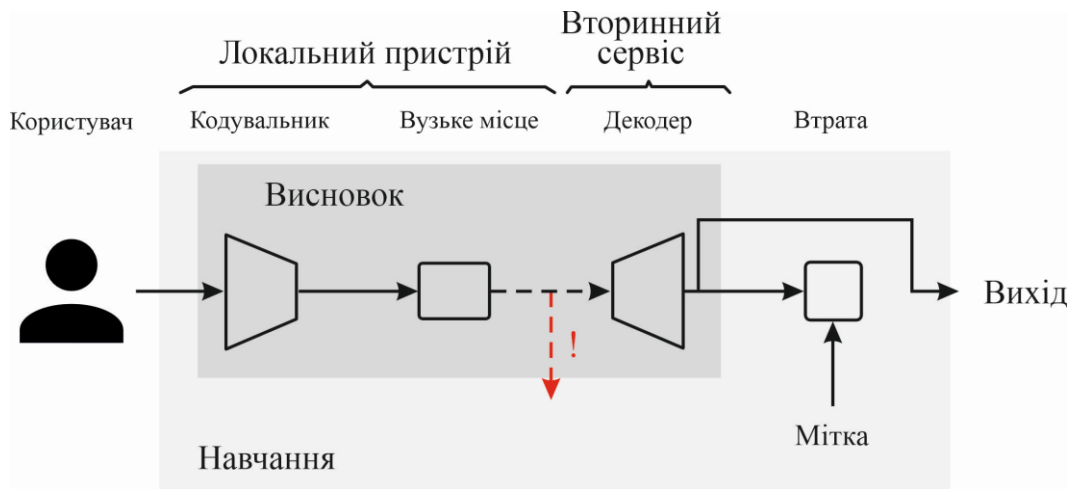


Рисунок 2.2 – Навчання методу аналізу мовлення зі збереженням конфіденційності за принципом інформаційного вузького місця. Точка атаки вказана червоною стрілкою та знаком оклику

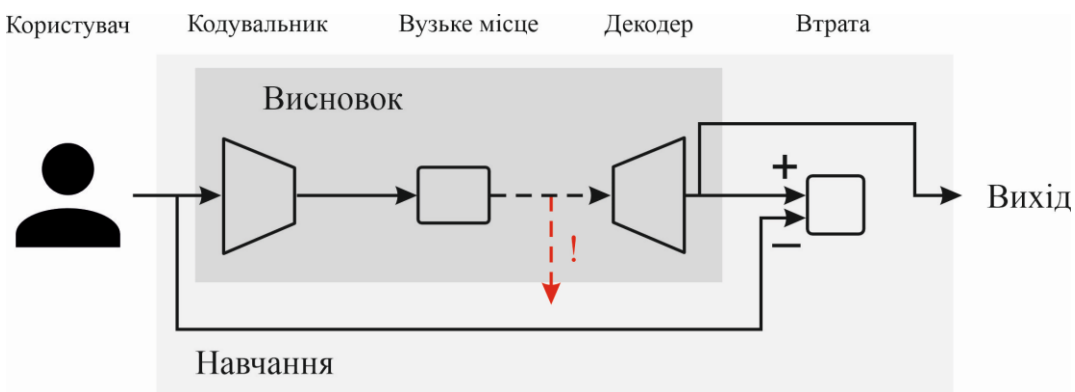


Рисунок 2.3 – Навчання структури автокодувальника, що зберігає конфіденційність, як приклад принципу інформаційного вузького місця.

Точка атаки вказана червоною стрілкою та знаком оклику

Інформаційний вміст у вузькому місці можна зменшити або шляхом зменшення рангу сигналу (зменшення розмірності даних, що проходять через вузьке місце, наприклад [33]), або квантуванням і кодуванням сигналу з низьким бітрейтом (наприклад, [9, 34]). Потім компроміс між точністю реконструкції та ентропією вузького місця визначає ступінь конфіденційності.

Паралельно, коли використовується модель машинного навчання, модель має бути навчена для найкращого компромісу між корисністю та конфіденційністю. Зазвичай це передбачає мінімізацію втрати точності приватного повідомлення. Наприклад, якщо завдання полягає в тому, щоб отримати текстовий вміст за допомогою автоматичного розпізнавача мовлення, то частота помилок цього розпізнавача повинна бути мінімізована. Подібним чином, у програмах, де вихідний сигнал відтворюється для слухачів, ми повинні використовувати як об'єктивні методи оцінки якості, такі як POLQA і PESQ [35, 36], так і суб'єктивні тести на прослуховування, такі як P.800 і MUSHRA [37, 38].

Часто використовуваним підходом до реалізації вузького місця є структура автокодувальника, що складається з кодера, вузького місця та декодера, де метою є реконструкція вхідного сигналу з виходу вузького місця [33, 39, 40] (див. рис. 2.3). Однак для збереження конфіденційності вузьке місце має бути достатньо щільним, щоб оригінальний мовний сигнал не міг бути повністю реконструйований, щоб він нагадував оригінал за формою сигналу. Це ускладнює розробку функції втрат, оскільки вихідні дані недостатньо схожі на вхідні. Одним із рішень, подібних до навчання представленням, є передача вихідних даних автокодера знову кодеру та порівняння функцій вузьких місць [41, 42].

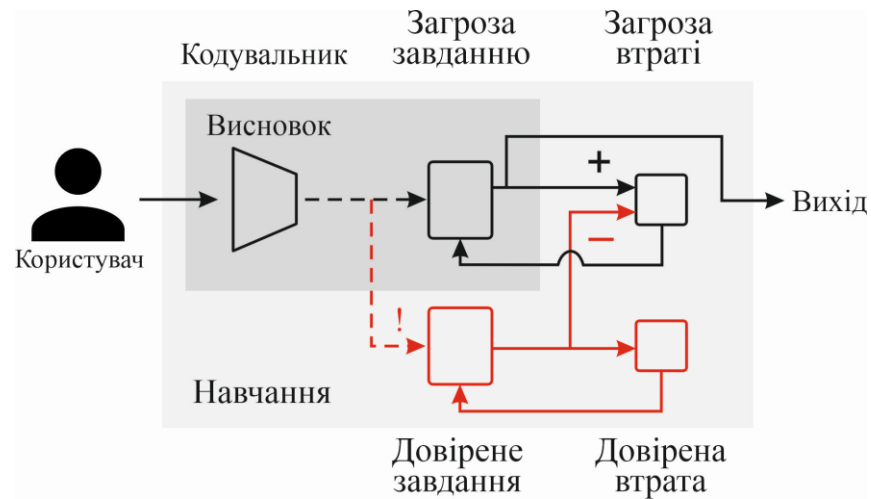


Рисунок 2.4 – Навчання методу аналізу мовлення із збереженням конфіденційності із змагальним підходом. Довірене завдання має право видобувати певну інформацію, тоді як завдання загрози (намальоване червоним) витягує іншу особисту інформацію. Під час навчання довірене завдання конкурує із завданням суперницької загрози, таким чином у блоці кодувальника видаляється вся особиста інформація. Під час логічного висновку завдання загрози можна ігнорувати

2) Змагальний підхід: Альтернативою зменшенню розміру інформаційного вузького місця є використання змагальної моделі під час навчання, щоб мінімізувати побічну інформацію у вузькому місці [43, 44]. Рисунок 2.4 ілюструє структуру моделі, де довірені та загрозливі завдання відповідають відповідно корисності та конфіденційності, які, у свою чергу, відповідно відповідають вилученню приватного повідомлення та побічної інформації. Завдання загрози незалежно оптимізовано для вилучення приватної додаткової інформації з вихідних даних вузького місця. Зберігаючи завдання загрози фіксованим, кодер і довірені завдання спільно оптимізовані, щоб максимально підвищити точність приватного повідомлення та мінімізувати точність приватної додаткової інформації. Таким чином, тренування змагальності поєднує в собі максимізацію та мінімізацію точності приватної побічної інформації. Це змушує кодер мінімізувати приватну побічну інформацію, що проходить через вузьке

місце, оскільки завдання загрози намагається максимізувати вилучення приватної побічної інформації.

Перевагою змагальної конфігурації є те, що розробник може конкретно вибрати, який тип загрози та категорію інформації буде видалено з даних. Система також може бути оптимізована наскрізно, щоб усі компоненти були спільно оптимізовані для найкращої продуктивності. Це може привести до ефективної моделі з точки зору компромісу між обчислювальними витратами та якістю результату.

Основна проблема змагального навчання полягає в тому, що воно не дає жодних теоретичних гарантій чи заходів конфіденційності. Найкраще, що він може зробити, це продемонструвати ступінь, до якого обрана змагальна модель не змогла витягнути приватну побічну інформацію з вибраної категорії приватної інформації. Таким чином, ми можемо очікувати, що якийсь інший зловмисник з кращою моделлю все одно зможе витягти приватну побічну інформацію, і що інші категорії приватної інформації потенційно присутні у вихідних даних вузького місця.

3) Роз'єднання: інформаційні вузькі місця можна узагальнити, щоб охопити кілька каналів вузьких місць, де кожне вузьке місце представляє роз'єднане представлення (рис. 2.5). Тобто кожен канал представляє окремий і незалежний атрибут сигналу, так що ступінь анонімності може бути вибраний для кожного каналу відповідно до випадку використання, наприклад. [45–47].

Основна проблема з розплутуванням полягає у визначенні обмежень, з якими інформація надходить у відповідний канал. Наприклад, модель можна навчити зіставляти певні канали з мітками в наборі даних (наприклад, [46, 47]). Крім того, використовуючи навчання представленням, ми можемо обмежити канали на основі об'єктивних критеріїв, таких як часовий діапазон, тобто тривалість часу, протягом якого дані інтегруються (наприклад, [41, 48, 49]).

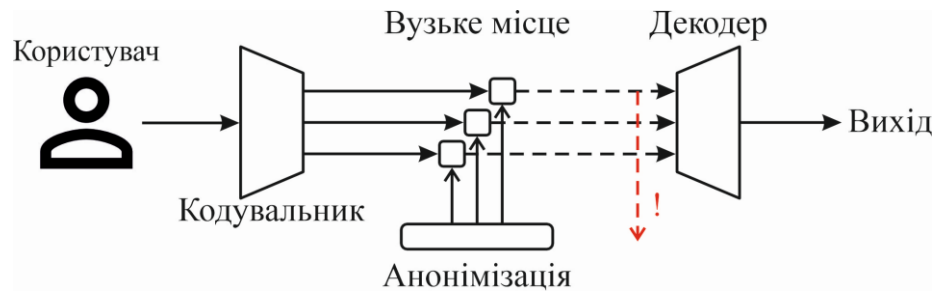


Рисунок 2.5 – Обробка мовлення із збереженням конфіденційності через розплутування, де мова розкладається на незалежні потоки інформації, кожен з яких представляє окрему категорію інформації, і де рівень анонімізації можна вибрати окремо для кожної категорії. Точка атаки вказана червоною стрілкою та знаком оклику

2.2 Покращення продуктивності

У багатьох випадках використання, таких наприклад, як помилкова активація, справжньою причиною є неадекватна продуктивність служби. Наприклад, якщо детектор пробуджуючих слів спрацьовує неправильно (помилково спрацьовує), то система почне прослуховувати розмову, коли вона не повинна була, таким чином порушуючи конфіденційність. Оскільки неадекватний детектор слів пробудження спричиняє порушення конфіденційності, найкращим рішенням є вдосконалення детектора слів пробудження. Таким чином, лікування спрямоване на причину, а не на симптоми.

Однак цей підхід має дві основні проблеми. По-перше, підвищення продуктивності часто вимагає збільшення обчислювальної потужності та споживання інших ресурсів. Це не тільки фінансово дорого, але й негативно впливає на навколишнє середовище [50]. Зокрема, балансувати обчислювальне навантаження легше на хмарному сервері, і таким чином існуючі ресурси можна використовувати ефективніше. По-друге, навіть із

покращеною продуктивністю мовні інтерфейси завжди матимуть випадкові помилки. Тобто конструкція мовної технології, що зберігає конфіденційність, повинна включати кілька рівнів захисту, особливо коли мова йде про операції з великими наслідками [51]. Скажімо, коли активовано детектор пробуджуючих слів, то перед виконанням будь-яких дій, які потенційно можуть порушити конфіденційність, система може вимагати, щоб мовець перебував у тій самій кімнаті, або була перевірена особа мовця, або додатковий крок підтвердження «Ви впевнені?» » або подібні. Інтрुзивність таких додаткових засобів захисту повинна відображати серйозність потенційного порушення, щоб захист не сприймався як надмірно нав'язливий і не зменшував корисність послуги.

Підвищення продуктивності також може означати, що системі потрібні додаткові функції. Наприклад, у випадку, якщо другий локальний пристрій підслуховує взаємодію Аліси з Бобом (витік обговорення) або основним пристроєм (витік мовного інтерфейсу), обов'язковим захистом конфіденційності є те, що вторинний пристрій або знає, що він не є призначеним одержувачем мовлення (аналіз мовлення), або сповіщає користувача про його присутність (розробка інтерфейсу користувача). Це підкреслює важливість довіри до пристроїв, присутніх в акустичному просторі. Додатковий локальний пристрій може чути все, що вимовляється в локальному просторі. Таким чином, користувач повинен мати високий рівень довіри до пристрою та постачальника послуг, щоб він був достатньо компетентним, щоб знати, коли він є частиною розмови, щоб він був достатньо доброзичливим, щоб захистити конфіденційність користувачів, а також повідомити користувача про діяльність, яка потенційно порушує конфіденційність [52, 53].

2.3 Обмеження доступу до особистих повідомлень

Кожне мовленнєве застосування передбачає передачу необхідного законного повідомлення, яке, швидше за все, буде приватним. Оскільки передавання такої приватної інформації неминуче, це наражає користувача на загрозу його конфіденційності. Оскільки нам потрібна мовна програма, загрози можна пом'якшити, зменшивши точність інформації шляхом додавання шуму, зробивши інформацію недоступною, наприклад шифрування, або вибір не надсилати нічого, а застосовувати обробку лише локально, відому як обробка краю.

1) Зменшення точності та диференціальної конфіденційності: багато програм вимагають лише середніх показників населення, а не інформації про конкретних осіб. Наприклад, кол-центру може знадобитися знати розподіл за статтю, але не потрібно знати гендерну ідентичність будь-якого конкретного клієнта. Цю інформацію можна отримати за допомогою простої стохастичної схеми, відомої з області диференціальної конфіденційності [54]. Зокрема, клієнт спочатку кидав монету, щоб визначити, чи варто йому говорити брехню чи правду. Коли клієнт бреше, клієнт знову кидає монету, щоб визначити, яку статтю повідомити (тут ми використовуємо лише двійкові гендерні категорії для стислості). Це дає 75% ймовірність того, що було повідомлено справжню особу, і 25% ймовірність того, що вона була помилковою [55]. Іншими словами, цей підхід відповідає спотворенню сигналу або додаванню шуму до сигналу для захисту конфіденційності (рис. 2.6).

Тоді клієнт завжди може стверджувати, що він бреше, так що ця система дає правдоподібне заперечення. Одночасно, враховуючи досить велику вибірку, ми завжди можемо оцінити справжній розподіл із шумової вибірки. Зауважте, що хоча цей підхід дає правдоподібне заперечення, інформація все одно статистично корелює з правдивою інформацією.

Враховуючи кілька зашумлених фрагментів інформації, все ще можливо відновити точну приватну інформацію про мовця. Таким чином, надзвичайно важливо кількісно визначити ступінь захисту за допомогою заходів диференційованої конфіденційності.



Рисунок 2.6 – Захист приватності шляхом додавання шуму до приватних даних відповідно до ідеї диференціальної приватності

Наведений вище приклад легко узагальнює будь-яку категоричну інформацію, таку як політична чи релігійна приналежність, а також постійні параметри, такі як вік мовця чи інші біометричні характеристики. З евристичної точки зору зниження точності безперервних параметрів таким чином стає схожим на додатковий шум, який можна виміряти за допомогою співвідношення сигнал/шум.

2) Криптографія та безпечні обчислювальні системи: якщо передана інформація зашифрована, її не можна використовувати для зловмисних цілей без доступу до ключа. Дивно, однак, можливо обробляти зашифровану інформацію за допомогою гомоморфного шифрування, так що зашифрований результат обчислення можна повернути та відкрити [24, 56, 57] (рис. 2.7). Це може бути застосовано, наприклад, для виділення спектральних особливостей мовлення або розпізнавання мовлення в хмарі без доступу до мовного сигналу за допомогою простого тексту [58, 59]. Тоді периферійний пристрій шифрує мовний сигнал і надсилає зашифровані дані в хмару, яка витягує зашифровану інформацію та повертає її на крайній пристрій, який може відкрити зашифрований результат. Хоча гомоморфне шифрування в принципі забезпечує прекрасне рішення для забезпечення конфіденційності,

воно часто пов'язане з непомірно високою ціною обчислень. Кількість виконаних операцій і переданих бітів експоненційно зростає зі складністю початкової проблеми.



Рисунок 2.7 – Захист конфіденційності шляхом шифрування зв'язку та обробки. Сіра зона вказує на зашифрований домен, а червоні подвійні стрілки зі знаком оклику вказують на захищений потік даних

Іншим підходом є захищене багатостороннє обчислення (MPC), де кілька сторін можуть обчислювати спільну функцію, нічого не відкриваючи одна одній [60]. MPC забезпечує набагато менші накладні витрати на обчислення та зв'язок, у той час як можна одночасно продемонструвати, що надається незв'язність, незворотність та відновлюваність біометричної інформації [61]. Його можна застосувати, наприклад, для розпізнавання мовця в хмарі, так що модель динаміка користувача не розкривається в хмарі, а модель розпізнавання не розкривається користувачеві [62].

3) Розподілене навчання: більшість передових засобів обробки мовлення сьогодні використовує машинне навчання, яке має бути користувачами, де пристрій відстежує людей, присутніх у кімнаті, таким чином, щоб особиста інформація не передавалася неавторизованим слухачам (червона пунктирна лінія та знак оклику). навчаються за допомогою великих баз даних мовлення. Дані найкращої якості точно відповідають сценаріям використання послуг. Запис взаємодії користувачів із їхніми пристроями стає привабливим для навчання вдосконалених моделей, оскільки він точно відповідає варіанту використання. Це становить значну загрозу

конфіденційності, оскільки такий необмежений запис може захопити широкий спектр приватної інформації, включаючи всі взаємодії з пристроєм, але потенційно також будь-яку та всю мову поблизу пристрою.

Розподілене навчання — це підхід до моделей навчання без необхідності централізованого збору даних [63]. Моделі навчаються на локальних периферійних пристроях, і лише оновлення моделі обмінюються між вузлами та/або центральним сервером (див. рис. 2.8). Оскільки необроблені дані ніколи не залишають периферійний пристрій, це відповідає підходу до захисту конфіденційності, коли потік інформації обмежений.

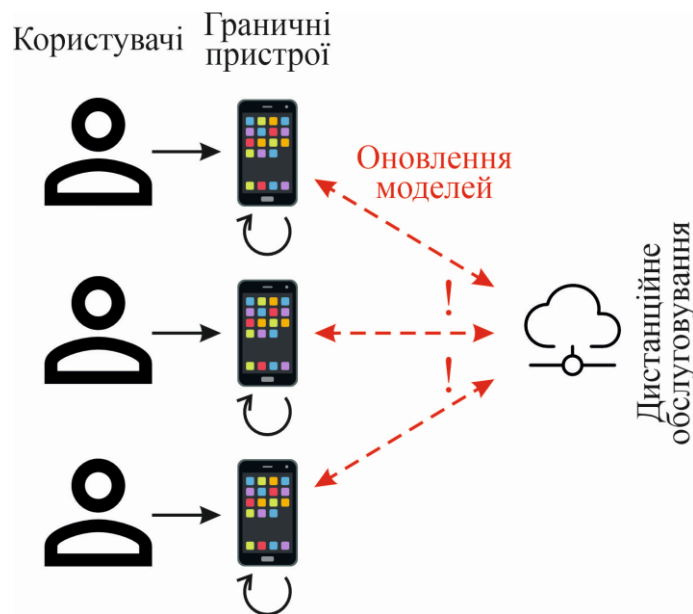


Рисунок 2.8 – Інтегроване навчання як приклад розподіленого навчання, де пристрої навчають моделі локально та передають лише оновлення моделі в хмару. Червоні пунктирні стрілки зі знаком оклику вказують на зменшення потоку інформації

Інтегроване навчання є одним із різновидів розподіленого навчання, де центральний хмарний сервер збирає, об'єднує та перерозподіляє оновлення моделі. Він використовувався, наприклад, для розпізнавання мовців та емоцій, моделювання мови, а також для неконтрольованої оцінки мікрофонних кластерів у сенсорних мережах [64–71].

Загалом, розподілене навчання є багатообіцяючим підходом, навіть якщо він має кілька проблем. По-перше, створення, навчання та тестування системної архітектури набагато складніше, ніж звичайне машинне навчання. По-друге, оновлення моделі є специфічними для моделі і не можуть бути легко повторно використані, якщо оновлено структуру моделі. Навчання, накопичене під час навчання, фактично втрачається кожного разу, коли модель оновлюється, що також ставить під загрозу справедливе порівняння конкуруючих підходів. По-третє, навіть якщо розподілене навчання покращує конфіденційність, це не є гарантією конфіденційності, оскільки оновлення моделі також можуть містити конфіденційну інформацію [72].

2.4 Обмеження доступу до відтвореного аудіо

Коли користувач взаємодіє з мовним інтерфейсом, голосова відповідь мовного інтерфейсу може містити особисту інформацію. Цю приватну інформацію можуть підслухати інші користувачі в тому самому акустичному просторі, що становить загрозу конфіденційності. Теоретично було б можливо ідентифікувати та відстежувати користувачів у тому самому акустичному просторі та передавати приватну інформацію лише тоді, коли вона не становить загрози конфіденційності (рис. 2.9). Це дає змогу користувачам Loudspeakers Local device Alice довіряти локальному пристрою та його здатності відстежувати та визначати рівні авторизації людей, які присутні в акустичному просторі. Однак такі системи ще не були широко опубліковані.

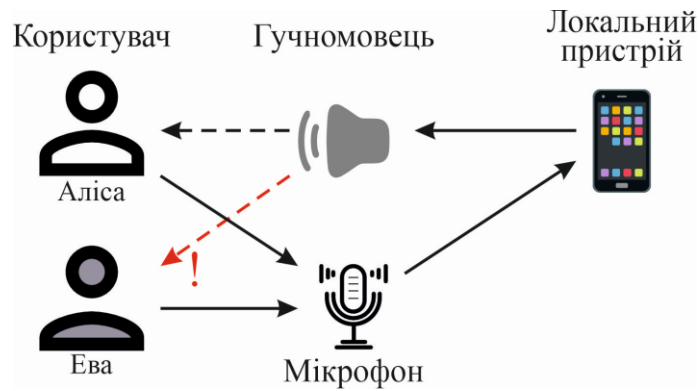


Рисунок 2.9 – Захист від витоку відтворення за допомогою відстеження авторизації користувачів, де пристрій відстежує людей, присутніх у кімнаті, щоб особиста інформація не передавалася неавторизованим слухачам (червона пунктирна лінія та знак оклику)

На рис. 2.10 представлено інше рішення, яке використовує конструктивні та деструктивні перешкоди між гучномовцями для створення звукових зон, де приватна інформація є, відповідно, зрозумілою або спотвореною (див., наприклад, [73–75]). Вибираючи місце в просторі, де мова є зрозумілою, і припускаючи, що ми знаємо, де знаходиться цільовий користувач, таким чином ми можемо обмежити доступ до приватної інформації лише для цільового мовця. Зауважте, що звукові зони з деструктивними перешкодами дійсно отримують часткове спостереження приватного повідомлення, але воно (сподіваємось) спотворюється настільки, що стає нерозбірливим. Таким чином, цей підхід є ще одним методом, який використовує спотворення приватного повідомлення для збереження конфіденційності. Основні завдання цього підходу полягають у тому, щоб зробити зону конструктивного звуку достатньо великою, щоб вона дозволяла невеликі рухи голови, і зробити деструктивну інтерференцію рівномірною всюди, щоб не було ізолюваних точок з конструктивною інтерференцією за межами бажаної звукової зони. Однак перевага цього підходу полягає в тому, що він вимагає відстеження лише цільового слухача, що хоч і складно, але набагато простіше, ніж відстежувати всіх людей у кімнаті.

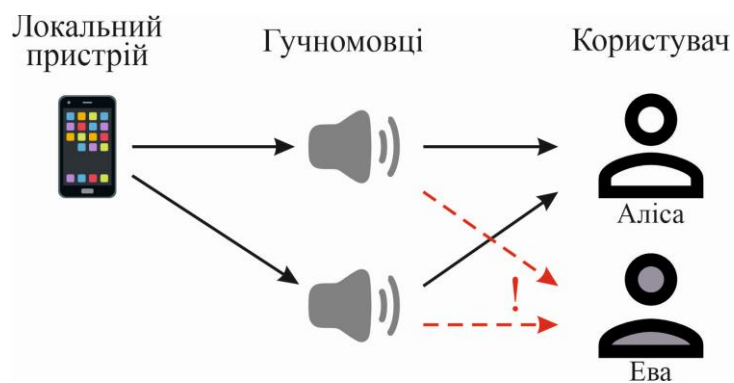


Рисунок 2.10 – Захист від витоку відтворення за допомогою звукових зон, де конструктивні перешкоди між динаміками використовуються для збереження розбірливості для користувача Аліса, а деструктивні перешкоди спотворюють її для користувача Єва (червоні пунктирні лінії та знак оклику)

3 ОЦІНКА КОНФІДЕНЦІЙНОСТІ

Щоб оцінити продуктивність будь-яких методів збереження конфіденційності, нам потрібні показники продуктивності як для корисності, так і для конфіденційності, що відповідають відповідно довіренім і загрозливим завданням. Показники продуктивності довіреної задачі визначаються програмою, вони відповідають типовим процедурам звичайної методології обробки мовлення [76, 77] і тому не потребують подальшого обговорення тут. Треба лише зазначити, що часто існує компроміс між показниками корисності та конфіденційності, тому немає сенсу оцінювати лише один, але належні експерименти завжди повинні оцінювати компроміс. Тоді мета полягає у визначенні показників ефективності для вимірювання ступеня конфіденційності, що забезпечується методом.

3.1 Об'єктивні показники

Показники, застосовні до наведеної вище моделі атаки, включають:

- Рівень частоти помилок (EER) розглядає злоумисника, який приймає рішення, застосовуючи порогове значення до оціночної функції, де поріг вибирається таким чином, щоб помилкові позитивні та помилкові негативні результати були рівними [20, 78]. Збільшення EER означає, що злоумисник зробив більше помилок і конфіденційність покращилася.

- Незалежна від застосування логарифмічна функція вартості відношення правдоподібності $C_{\min} \log$ узагальнює EER, розглядаючи оптимальні порогові значення для всіх можливих попередніх ймовірностей і всіх можливих функцій вартості помилок [20, 79].

- Зв'язуваність визначається як (логарифмічна) ймовірність того, що два набори даних є екземплярами з одного або іншого походження [80].
- Взаємну інформацію можна використовувати для кількісного визначення того, скільки нової інформації про мовця ми отримуємо з нового набору даних [81].

Всі вищезазначені показники мають дві помітні слабкості, спільні з підходом диференційованої конфіденційності. По-перше, вони не захищають від майбутніх атак. Це особливо очевидно для EER, який вимірює продуктивність однієї реалізації зловмисника. Зрозуміло, що більша, досконаліша модель могла б зробити потужнішу атаку. Це також стосується вимірювань на основі правдоподібності, де визначення метрик не залежать від моделі, але покладаються на моделі розподілу ймовірностей, що залежать від конкретних випадків. Коли ці статистичні моделі вдосконалюються, можуть бути виявлені нові недоліки.

По-друге, кожна з наведених вище метрик пов'язана з окремими спостереженнями. Навпаки, мова — це безперервний потік інформації, який дає серію спостережень. З кожним спостереженням ми можемо зменшити довірчі інтервали, якщо різні випадки мають різні розподіли ймовірностей. Таким чином, повторні спостереження, з усією впевненістю, порушують конфіденційність для всіх помітних характеристик, коли кількість спостережень є достатньо високою.

Більш повна оцінка об'єктивних показників конфіденційності надається в [82], а виклик VoicePrivacy дає практичний приклад того, як застосовувати ці показники [19].

3.2 Диференціальна конфіденційність

Анонімність і конфіденційність ніколи не є абсолютними. Навіть якщо поточні методи та наявні на даний момент набори даних не дозволяють нам зробити висновок про щось приватне з анонімного набору даних, немає гарантії, що користувач буде захищений також, коли нова методологія буде представлена в майбутньому або коли стануть доступними нові пов'язані набори даних. Тому ми можемо стверджувати лише про те, наскільки добре захищені користувачі. Диференціальна конфіденційність є такою теорією та методологією для характеристики ступеня захисту конфіденційності для користувачів [54, 83].

Диференціальна конфіденційність працює з базою даних із приватною інформацією, такою як вік n користувачів, з якої ми обчислюємо середній вік m_n . Таку статистику населення зазвичай можна вважати анонімною, якщо розмір сукупності достатньо великий. Однак припустимо, що новий користувач Аліса зареєстрований у базі даних, а середній вік оновлено до m_{n+1} . Якщо потім зломисник отримує доступ до двох середніх значень, m_n і m_{n+1} , а також до кількості користувачів до Аліси n , тоді можна тривіально знайти вік Аліси як $m_{n+1}(n+1) - m_n n$. Іншими словами, відстежуючи зміни в анонімній інформації, все одно можна розкрити приватну інформацію.

Однак для захисту конфіденційності ми можемо додати шум до окремих вимірювань, і це дає певний захист від продемонстрованої вище диференціальної атаки. Таким чином, обчислення статистики популяції стає рандомізованим алгоритмом M . Формально ми говоримо, що рандомізований алгоритм M забезпечує ϵ -диференціальну конфіденційність, якщо для всіх наборів даних D і D' , що відрізняються одним користувачем, і підмножини вихідного $S \subseteq \text{Range}(M)$,

$$\ln \Pr[M(D) \in S] \leq \epsilon + \ln \Pr[M(D') \in S]. \quad (1.1)$$

Тут $\Pr[\cdot]$ відноситься до ймовірності та втратою приватності ϵ . Чим менше, тим краще конфіденційність.

Інтерпретація цього визначення полягає в тому, що воно обмежує вплив будь-якого окремого користувача на загальну логарифмічну ймовірність виходу меншим, ніж ϵ . Оскільки логарифм правдоподібності характеризує ентропію, таким чином відповідає кількості приватної інформації, доступної зловмиснику. Хоча диференціальна конфіденційність визначається на основі членства в наборі даних, її можна застосувати до будь-якого атрибута мовця. Іншими словами, якщо зловмисник зацікавлений у будь-якому конкретному атрибуті мовця, тоді алгоритм із ϵ -диференційною конфіденційністю надасть щонайбільше ϵ -nats інформації про цей атрибут. Тут одиниця ϵ -nat відповідає природним одиницям інформації, визначеним як натуральний логарифм правдоподібності, тоді як біти відповідають логарифму правдоподібності за основою 2.

Перевага диференційної конфіденційності полягає в тому, що вона забезпечує точну математичну основу для аналізу ступеня конфіденційності. Однак він обов'язково завжди базується на статистичній моделі, яка наближено відповідає базовій системі. Навіть якщо диференціальна конфіденційність таким чином дає точні відповіді, їх надійність все одно завжди залежить від точності базових статистичних моделей. Тим не менш, за допомогою диференційної конфіденційності ми можемо, наприклад, проаналізувати ефект наявності двох паралельних джерел інформації з диференційною конфіденційністю ϵ_1 і ϵ_2 відповідно.

Оскільки значення ϵ_k представляють втрату конфіденційності, то сукупна втрата конфіденційності від двох джерел інформації становить щонайбільше їх суму $\epsilon_1 + \epsilon_2$ [83].

Інформаційні вузькі місця мають помітну паралель із різною конфіденційністю [31]. Там, де інформаційні вузькі місця обмежують кількість інформації, яку дозволено пройти, рівняння (1.1) також кількісно визначає кількість витоку інформації. Різниця полягає в тому, що вузьке місце містить і передбачуване приватне повідомлення, і витік інформації, тоді як рівняння (1.1) містить лише витік інформації. Якщо відомий бітрейт передбачуваного приватного повідомлення, то, очевидно, розмір витоку вузького місця можна визначити кількісно. Однак ця інтерпретація базується на двох неявних припущеннях. По-перше, вузьке місце має бути квантовано, як у векторно-квантованому варіаційному автокодері (VQ-VAE) [84], оскільки вузьке місце з безперервним значенням теоретично може містити необмежену кількість інформації, оскільки будь-який N -вимірний простір може бути відображається на 1-вимірний скаляр за допомогою кривих із заповненням простору [85]. По-друге, рівняння (1.1) виконується лише для одного спостереження, тоді як часовий ряд дає повторні спостереження. З кожним новим спостереженням ми отримуємо нову інформацію. Ми можемо очікувати, що такі повторювані спостереження будуть корельованими, але, незважаючи на це, маючи достатню кількість спостережень, ми можемо відрізнити будь-які різні розподіли. Таким чином, у своїй простій формі диференціальна конфіденційність не дає жодного захисту, якщо ми можемо спостерігати часовий ряд протягом досить тривалого часу.

Методологія диференціальної конфіденційності лише нещодавно була введена в технологію мовлення, серед іншого розпізнавання мовлення із збереженням конфіденційності [68], розпізнавання емоцій [64] та анонімізація мовця [86]. Оскільки математична строгість диференціальної конфіденційності має очевидні переваги, ймовірно, що застосування цієї методології збільшиться.

3.3 Сприйняття і психологія конфіденційності

Люди схильні мати глибоке почуття власності на деякі речі, як матеріальні, так і нематеріальні. Зокрема, люди мають почуття власності на інформацію про себе [87, 88]. Такі почуття певною мірою відірвані від матеріальних наслідків порушення приватного життя. Наприклад, навіть якщо публікація аудіозапису, що розкриває таємні інтимні стосунки, не матиме прямих економічних наслідків, вона може завдати психологічної шкоди та зашкодити особистим стосункам. Наслідки загроз для приватного життя тоді неодмінно стають питанням психології та соціальних наук. Щоб якісно чи кількісно виміряти такі ефекти, нам також потрібні суб'єктивні тести з користувачами.

Водночас люди мають усталені соціальні правила щодо приватності між людьми [16]. Крім того, користувачі також мають тенденцію антропоморфізувати технологію, тобто ставитися до пристроїв і послуг як до людей [89, 90], так що вони, ймовірно, припускатимуть, що пристрої застосовують людські соціальні правила до конфіденційності. Зверніть увагу, що такі соціальні правила пов'язані з людською поведінкою та продуктивністю. Це не показує, чи вони застосовні до надлюдської продуктивності, якою можуть володіти комп'ютери, наприклад здатності інтегрувати інформацію в масивні бази даних і зберігати точні записи подій, що минули давно. Тоді те, як машини мають поводитися, особливо щодо надлюдських можливостей, є питанням не лише дизайну користувацького інтерфейсу, а й моральної психології [91]: нам потрібні дискусії на суспільному рівні про те, як мають поводитися автоматизовані служби та як вони є. дозволено поводитися.

Незалежно від того, чи дійсно система зберігає конфіденційність чи ні, деякі користувачі можуть сприймати систему як загрозливу, а інші можуть не помічати загроз, які вона створює [92]. Очевидно, що користувачі не люблять

користуватися службами, які вони сприймають як загрозливі. Розуміння того, як люди сприймають конфіденційність, цікаво саме по собі, але таке розуміння має важливе значення для розробки ефективних інтерфейсів користувача.

До сприйняття та досвіду конфіденційності можна підійти двома альтернативними способами. Ми можемо проводити дослідження користувачів, де люди взаємодіють або з машинами, або один з одним. Розрізнення важливе в тому сенсі, що взаємодія між людьми базується на соціальних правилах, які добре встановлені та розроблені протягом тривалого часу. Таким чином, ми можемо очікувати, що вони будуть стабільними з часом, тоді як взаємодія людини з комп'ютером постійно розвивається, оскільки люди навчаються більше. Крім того, використовуючи дослідження взаємодії між людьми, ми можемо дізнатися про ефекти, пов'язані з продуктивністю, схожою на людину, але, ймовірно, не можемо покладатися на спостереження ефектів, пов'язаних із продуктивністю надлюдини. Дослідження взаємодії людини з людиною, однак, завжди є проксі, якщо справжньою метою є розробка інтерфейсу людина-комп'ютер.

Таким чином, дослідження взаємодії людини та комп'ютера безпосередньо характеризують бажане явище. Однак компроміс полягає в тому, що розуміння людьми конфіденційності з пристроями може не відображати справжній рівень конфіденційності, а висновки, зроблені на основі думок користувачів, можуть не відображати їхні справжні уподобання, а ці уподобання є рухомими цілями. У міру того, як люди набувають більше досвіду роботи з технологіями та дізнаються про них більше, їхнє ставлення до них і їх сприйняття змінюються.

Незважаючи на ці недоліки, обидва типи експериментів є важливими для покращення нашого розуміння конфіденційності та вдосконалення технологій. Наприклад, дослідження між людьми показали, що люди по-різному відчують приватність у різних акустичних середовищах; галаслива їдальня може краще маскувати звуки та захищати від прослуховування, ніж

реверберуючий коридор [12, 13, 93–95], а багатомісні палати для пацієнтів у лікарнях мають проблеми з конфіденційністю [96]. Подібним чином дослідження «людина-машина» показали, що коли чат-бот активно повідомляє про вибір, пов'язаний з конфіденційністю, це покращує досвід конфіденційності для користувачів [97], голосові інтерфейси з невідомими функціями викликають страх у користувачів і зменшують збереження послуг [98–100] і порушення конфіденційності дуже шкодять довірі до послуг і знижують бажання користувачів користуватися послугами, але таку довіру можна відновити [101].

Для конфіденційності, яку можна використовувати, важливо, щоб інформація про конфіденційність надавалася легко, про зміни в ступені конфіденційності повідомлялося негайно, і щоб користувачі могли контролювати рівень конфіденційності [97, 99, 101–103]. Для порівняння, візуальні інтерфейси можуть використовувати світло або піктограми для моніторингу, а тактильні інтерфейси – для контролю рівня конфіденційності. Хоча звук можна використовувати для моніторингу стану системи за допомогою озвучення [104, 105], це не є широко поширеним у сигналізації конфіденційності. Це компроміс між наповненням акустичного простору інформацією та здатністю й схильністю слухової системи людини блокувати монотонні звуки. Щоб бути ефективним, звук має бути сприйнятим, коли користувач свідомо хоче перевірити рівень конфіденційності, а зміни рівня конфіденційності повинні викликати відповідні та відповідні значні зміни в звуковому ландшафті, щоб користувачі свідомо реєстрували такі зміни. Крім того, користувацькі інтерфейси повинні надавати правильну інформацію про рівень конфіденційності та дозволяти контролювати його [10, 99].

У будь-якому випадку вкрай важливо, щоб послуги були розроблені таким чином, щоб вони відображали справжній ступінь конфіденційності. Зауважте, що цілком можливо розробити системи, які передають високий рівень конфіденційності, навіть якщо система не поважає конфіденційність користувача. Насправді постачальники послуг мають короткострокові

стимули використовувати такі підходи до проектування, якщо вони покращують загальне задоволення користувачів. Однак такі підходи до дизайну відомі як темні шаблони або оманливі шаблони дизайну, і вони вважаються неетичними [106]. Шляхом обману такі темні шаблони змушують користувачів вірити, що вони в безпеці, хоча насправді вони зазнають зловживання. Таким чином, хороший дизайн конфіденційності повинен активно спілкуватися та дозволяти контролювати справжній ступінь конфіденційності. Така практика проектування є не тільки етичною, але й винагороджує постачальників послуг у довгостроковій перспективі, покращуючи утримання цих послуг [101].

3.4 Наслідки конфіденційності в мовленнєвих технологіях

Щоб отримати повне уявлення про наслідки конфіденційності в мовленнєвих технологіях, обговоримо як категорію приватної інформації, так і застосування такої інформації. Зауважимо, що це не повний список вмісту та категорій програм.

По-перше, зауважимо, що хоча всі представлені категорії є (потенційно) приватною інформацією, уся стійка інформація потенційно може бути використана для ідентифікації мовця. Наприклад, хоча поточний емоційний стан сам по собі не може ідентифікувати мовця, тенденція демонструвати емоційний стан може допомогти в ідентифікації. Однак, якщо ми хочемо перевірити, що спікер є тим, за кого себе видає, ми можемо використовувати лише інформацію, яку не можна змінити навмисно. Тобто стиль мовлення є особливим прикладом властивості, яку хороший актор озвучування може вільно вибирати, що робить його «легким» для змінювання з шахрайською метою.

Тоді розпізнавання особистості мовця стає природною відправною точкою для досліджень приватності. Можна спробувати розпізнати, хто говорить (розпізнавання мовця), перевірити, чи має мовець заявлену особу (перевірка мовця), кластеризувати аудіо в сегменти з одним мовцем (діарізація мовця), а також ми можемо розробити методи обману особи (спуфінг), напр. [33, 43, 44, 61, 62, 65, 66, 107–115]. Подібним чином, за допомогою перетворення голосу ми можемо анонімізувати особу мовця, замінивши її випадковою ідентичністю (анонімізація) або конкретною (псевдонімізація), напр. [20, 86, 108, 116–121]. Характеристика мовця є природним доповненням до ідентифікації мовця [82, 122]. Такі методи, пов'язані з ідентифікацією мовця, можна застосовувати, наприклад, для розпізнавання користувача пристрою, перевірки клієнта в банку або як голосовий аватар в онлайн-іграх. Так само анонімізацію та псевдонімізацію можна використовувати, щоб приховати особу від публічних ЗМІ та ігор.

Розпізнавання мовлення, як і перетворення мовлення в текст, є, ймовірно, найбільшою підсферою дослідження мовлення. Стосовно конфіденційності, він містить дві очевидні проблеми для конфіденційності. Ми можемо спробувати обмежитися побічною інформацією, такою як усунення всієї нетекстової інформації з потоку даних, напр. [68, 70, 123–127], або ми можемо використовувати обробку природної мови, щоб анонімізувати текстовий вміст, наприклад. [8, 128–132].

Конфіденційність є ще важливішою в постійно ввімкнених програмах, зокрема, виявлення wake-word, тобто коли інтерфейс запускається певним ключовим словом, як-от «Комп'ютер» у «Комп'ютер, світло вимкнено». Ця програма є більш чутливою саме тому, що вона завжди ввімкнена, оскільки користувачі не можуть вибрати, коли обробляти дані [28, 133–135]. Функція «завжди ввімкнена» також помітна в додатках для проживання з допоміжними особами, які можуть «моніторити щоденні справи людей, споживання калорій і режим сну, а також забезпечувати коучингові втручання для виховання позитивної поведінки» [136]. Такі голосові

інтерфейси навколишнього середовища часто реалізуються через мережі акустичних датчиків, які створюють свої власні проблеми [26, 28, 39, 67, 111, 137].

Покращення мовлення означає усунення фонових шумів і спотворень із бажаного мовного сигналу (приватного повідомлення) [76]. Це завдання може вплинути на конфіденційність двома способами. По-перше, середовище запису може розкрити приватну інформацію про мовця, а по-друге, фонові шуми, як і конкуруючий мовець, можуть містити приватну інформацію [138, 139]. Таким чином, послаблення фонових шумів і конкуруючих динаміків, а також усунення реверберації може покращити конфіденційність. Крім того, хоча використання сенсорної мережі для захоплення мовлення може підвищити корисність, воно також створює нові загрози [13, 67].

Мовлення зазвичай вважається біометричною інформацією в повному обсязі, включаючи всі категорії інформації, оскільки такі категорії можуть бути використані для ідентифікації особи [78, 82, 140–142]. Однак деякі обмежують термін біометрії, щоб позначати лише фізичні та поведінкові характеристики. До таких властивостей належать здоров'я [143], емоції [64] та гендерна ідентичність [43, 46, 134, 144], кожна з яких потребує окремого розгляду.

3.5 Правовий та соціальний аспекти

Конфіденційність у мовленнєвих технологіях має великий вплив як на індивідуальному, так і на суспільному рівнях, як уже обговорювалося раніше. Масштаби впливу на суспільство можна оцінити, згадавши скандал щодо конфіденційності Cambridge Analytica [145, 146], коли приватну інформацію було вилучено із соціальних мереж. медіа та використовуються для цільової політичної реклами. Як у мовленнєвих технологіях, так і в соціальних медіа

послуги працюють із величезною кількістю користувачів і включають взаємодію між кількома користувачами. Це наражає обидві сфери на однакові ризики. У випадку з Cambridge Analytica найвідомішим наслідком було те, що вона вплинула на результати виборів у великій демократичній країні. Іншим відомим скандалом із біометрією є випадок, коли нібито анонімні дані пацієнтів 10% австралійців були оприлюднені для громадськості, лише щоб пізніше було показано, що людей можна було легко ідентифікувати [147]. Таким чином, усі дані про здоров'я цих пацієнтів були оприлюднені всупереч уподобанням користувачів і з невідомими довгостроковими наслідками.

Хоча порушення, пов'язані із мовними технологіями з наслідками порівнянної величини ще не мали місця, ці паралелі підкреслюють потенційний вплив порушень. Скандали, безпосередньо пов'язані з мовними технологіями, включають прослуховування приватних осіб співробітниками та контрагентами постачальників послуг [1, 2, 148].

Зазначимо, що постачальники послуг і технологій мають чіткі короткострокові стимули, які суперечать уподобанням користувачів щодо конфіденційності. Ми можемо легко знайти приклади, коли користувачі не мають реального вибору щодо захисту своєї конфіденційності. Наприклад, припустімо, що всі друзі та родина користувача використовують певну платформу для соціальної та голосової взаємодії, навіть якщо її налаштування конфіденційності є невідповідними. Тоді окремий користувач стикається з вибором: або відключитися від своєї соціальної мережі, або порушити конфіденційність. Це стосується всіх користувачів із достатньо великою частиною їхніх спільнот, які беруть участь у цій платформі. Це версія дилеми ув'язненого, де жоден користувач не має стимулу переходити до рішення, яке буде найкращим для всіх.

На індивідуальному рівні приклади потенційних скандалів завдають явних збитків особам, зокрема психологічних [143]. Додаткові небезпеки включають сталкерів [149–151]. Хоча такі індивідуальні пошкодження є

«невеликими» на суспільному рівні, їх поширеність робить їхній спільний вплив значним [4].

Ці приклади демонструють невід’ємну потребу в регулюванні мовленнєвих технологій на рівні суспільства щодо конфіденційності. Уряди вже відреагували на цю потребу: Європейський Союз очолив процес із Загальним регламентом захисту даних (GDPR) [152, 153], а штат Каліфорнія незабаром після цього прийняв Закон Каліфорнії про конфіденційність споживачів (CCPA) 2018 року [154]. Хоча ці закони охоплюють лише невеликий відсоток населення світу, оскільки хмарні служби зазвичай працюють у всьому світі, вони потребують здатності дотримуватися місцевих законів. У багатьох випадках найпростіше застосувати найсуворіші закони до всіх користувачів, щоб найсуворіші закони сприяли конфіденційності всіх користувачів. Тому постачальники послуг широко прийняли вимоги GDPR і CCPA, і це, ймовірно, мало великий вплив також на користувачів, які не підпадають під дію цих правил.

Що стосується регулювання, важливим наслідком об’єктивних вимірювань конфіденційності є те, що інструменти та вимірювання нададуть лише статистичні характеристики конфіденційності, але вони ніколи не можуть дати абсолютну впевненість. Це різко контрастує з концепцією унікальної ідентифікації, яка використовується в юридичних документах, таких як Загальний регламент захисту даних (GDPR) Європейського Союзу [152], який явно не залишає місця для статистичної невизначеності. Це відображено, наприклад, у Керівних принципах для віртуальних голосових помічників Європейської ради із захисту даних, де зазначено, що: [142] «... голосові дані за своєю суттю є біометричними персональними даними. Як результат, коли такі дані обробляються з метою однозначної ідентифікації фізичної особи... обробка повинна мати дійсну правову основу...».

Це залишає тлумачення відкритим. Можна стверджувати, що неможливо отримати абсолютну впевненість у ідентифікації мовця, щоб GDPR ніколи не запускався. Також можна стверджувати, що всі голосові дані

містять особисту інформацію, яка може бути використана для унікальної ідентифікації особи, так що вся обробка повинна мати дійсну правову основу. Обидві інтерпретації доводять до абсурду, який говорить про те, що істина повинна бути десь посередині. Насправді GDPR на практиці вимагає [142], щоб процес проектування голосових помічників включав оцінку впливу на захист даних, де оцінюються ризики та наслідки, щоб розробник міг вжити належних заходів для збереження конфіденційності. Таким чином, автори GDPR чітко усвідомлюють, що неможливо дати абсолютні гарантії конфіденційності, але що оцінка впливу (тобто об'єктивні показники конфіденційності) обов'язково повинна ґрунтуватися на статистичних показниках, навіть якщо такі показники не були визначені.

Поки уряди регулюють конфіденційність, корпорації та неурядові організації також усвідомили, що належна конфіденційність — це можливість. Наприклад, Open Voice Network прагне розробити та стандартизувати відкриті технічні стандарти та етичні принципи голосової допомоги [155], а MyData Global прагне допомогти людям і організаціям отримати вигоду від персональних даних у спосіб, орієнтований на людину [156, 157]. У дослідницькому співтоваристві автор цієї статті брав участь у створенні спеціальної групи інтересів у Міжнародній асоціації мовленнєвої комунікації (ISCA), присвяченій «Безпеці та конфіденційності мовленнєвої комунікації» [158]. Наскільки нам відомо, це найбільша у світі спільнота, яка займається цією темою.

ВИСНОВКИ

Кожен агент – будь то людина чи пристрій – який бере участь у взаємодії, а також акустичні шляхи та мережеві з'єднання, через які вони з'єднані, є потенційною поверхнею атаки. Будь-який агент, який може взаємодіяти з іншим агентом або прослуховувати з'єднання, є потенційним перехоплювачем. Оскільки конфіденційність визначається як сценарій, коли агент авторизований з деяким доступом, але перевищує цей дозвіл (навмисно чи ненавмисно), неможливо просто розірвати з'єднання, а тому потрібні більш досконалі проекти і методології.

Конфіденційність і етика значною мірою перетинаються. Більшість потенційних порушень етики в мовленнєвих технологіях ґрунтується на порушенні конфіденційності. Це означає, що потрібна етична дискусія в усьому суспільстві про те, що дозволено щодо конфіденційності користувачів.

Хоча дослідження в цій галузі пожвавилися зовсім нещодавно, вже існує значна кількість досліджень. Дослідження, однак, ще не є зрілими, але знаходяться у фазі швидкого розвитку, і є важливі підсфери, які ще не бачили великої роботи. Це робить його плідною сферою для досліджень, оскільки можна очікувати, що в найближчі роки буде знайдено важливе розуміння.

Хоча управління отриманням інформованої згоди має встановлені традиції та найкращі практики для більшості типів інтерфейсів, системи мовлення, аудіо та навколишнього середовища є особливо унікальними. А саме, акустична інформація – це потік, що змінюється в часі. Зачитування багатосторінкової форми згоди перед початком взаємодії є надто нав'язливим і надто детальним. Вимоги до конфіденційності також змінюються з часом. Таким чином, згода повинна бути отримана відповідно до фактичної потреби. Крім того, що він буде більш зручним у використанні,

він також дозволить краще пов'язати вибір із реальними потребами, оскільки згода отримується лише тоді, коли вона дійсно потрібна.

Метрики для потокової передачі: доступні теоретичні метрики відображають конфіденційність щодо кінцевого набору даних, тоді як мова є відкритим потоком даних. Наслідком цього є те, що теоретично можна розпізнати будь-який приватний атрибут або ідентичність за умови, що він має унікальний розподіл ймовірностей і можливим є достатньо тривале спостереження. Тому потрібні методології для характеристики впливу тривалості спостереження на конфіденційність.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Lynskey D. 'Alexa, are you invading my privacy?' – the dark side of our voice assistants. *the Guardian*. URL: <https://www.theguardian.com/technology/2019/oct/09/alexa-are-you-invading-my-privacy-the-dark-side-of-our-voice-assistants>.
2. Hern A. Apple contractors 'regularly hear confidential details' on Siri recordings. *the Guardian*. URL: <https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings>.
3. Brewster T. Fraudsters cloned company director's voice in \$35 million heist, police find. *Forbes*. URL: <https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=56f20b367559>.
4. Levy K., Schneier B. Privacy threats in intimate relationships. *Journal of cybersecurity*. 2020. Vol. 6, no. 1. URL: <https://doi.org/10.1093/cybsec/tyaa006>.
5. Penders J. Privacy in (mobile) telecommunications services. *Ethics and information technology*. 2004. Vol. 6, no. 4. P. 247–260. URL: <https://doi.org/10.1007/s10676-005-5605-9>.
6. Wong R., Garrie D. Privacy in electronic communications: the regulation of voip in the EU and the united states. *SSRN electronic journal*. 2009. URL: <https://doi.org/10.2139/ssrn.1466153>.
7. Analysis of privacy in mobile telephony systems / M. Arapinis et al. *International journal of information security*. 2016. Vol. 16, no. 5. P. 491–523. URL: <https://doi.org/10.1007/s10207-016-0338-9>.
8. Mahendran D., Luo C., McInnes B. T. Review: privacy-preservation in the context of natural language processing. *IEEE access*. 2021. Vol. 9. P. 147600–147612. URL: <https://doi.org/10.1109/access.2021.3124163>.

9. Revisiting speech content privacy / J. Williams et al. 2021 *ISCA symposium on security and privacy in speech communication*. ISCA, 2021. URL: <https://doi.org/10.21437/spsc.2021-9> .
10. Backstrom T., Bruggemeier B., Fischer J. Privacy in speech interfaces. *VDE ITG News*. URL: <https://www.vde.com/resource/blob/1991012/07662bec66907573ab254c3d99394ec7/itg-news-juli-oktober-2020-data.pdf> .
11. Nelus A. Privacy-preserving audio features for classification and clustering in acoustic sensor networks. Bochum : Universitätsbibliothek, 2022. URL: <https://doi.org/10.13154/294-9064>.
12. Sneha D. Robust and efficient methods for distributed speech processing - perspectives on coding, enhancement and privacy. *Aaltodoc Repository :: Home*. URL: <https://aaltodoc.aalto.fi/items/1abb540f-0c97-4d7f-98b3-e88b15f08539> .
13. Pablo P. Z. Preserving speech privacy in interactions with ad hoc sensor networks. *Aaltodoc Repository :: Home*. URL: <https://aaltodoc.aalto.fi/items/6d85d36a-a4f4-4112-a5ad-fc33224c787e> .
14. Pathak M. A. Privacy-Preserving machine learning for speech processing. New York, NY : Springer New York, 2013. URL: <https://doi.org/10.1007/978-1-4614-4639-2> .
15. Challenges towards the next frontier in privacy / R. Cummings et al. 2023. 50 p. (Preprint. 2304.06929). URL: <https://doi.org/10.48550/arXiv.2304.06929>.
16. Petronio S. Boundaries of privacy: dialectics of disclosure. Suny Press, 2002.
17. Derlega V. J., Chaikin A. L. Privacy and self-disclosure in social relationships. *Journal of social issues*. 1977. Vol. 33, no. 3. P. 102–115. URL: <https://doi.org/10.1111/j.1540-4560.1977.tb01885.x> .

18. The handbook of privacy studies / ed. by A. de Groot, B. van der Sloot. Amsterdam University Press, 2018. URL: <https://doi.org/10.1515/9789048540136> .
19. The voiceprivacy 2020 challenge: results and findings / N. Tomashenko et al. *Computer speech & language*. 2022. Vol. 74. P. 101362. URL: <https://doi.org/10.1016/j.csl.2022.101362> .
20. A comparative study of speech anonymization metrics / M. Maouche et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-2248> .
21. Yiming W. Wake word detection and its applications. *JScholarship | Home*. URL: <https://jscholarship.library.jhu.edu/items/e3dbcebd-4086-4cbd-aa7c-7ad191ff4158> .
22. Liptak A. Amazon's Alexa started ordering people dollhouses after hearing its name on TV. *The Verge*. URL: <https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse> .
23. Moscaritolo A. Amazon alexa sends family's private conversation to contact. *PCMAG*. URL: <https://www.pcmag.com/news/amazon-alexa-sends-familys-private-conversation-to-contact> .
24. Teixeira F., Abad A., Trancoso I. Patient privacy in paralinguistic tasks. *Interspeech 2018*. ISCA, 2018. URL: <https://doi.org/10.21437/interspeech.2018-2186> .
25. Das S., Bäckström T. Enhancement by postfiltering for speech and audio coding in ad hoc sensor networks. *JASA express letters*. 2021. Vol. 1, no. 1. P. 015206. URL: <https://doi.org/10.1121/10.0003208> .
26. Zarazaga P. P., Backstrom T., Sigg S. Acoustic fingerprints for access management in ad-hoc sensor networks. *IEEE access*. 2020. Vol. 8. P. 166083–166094. URL: <https://doi.org/10.1109/access.2020.3022618> .
27. Provable consent for voice user interfaces / S. Sigg et al. *2020 IEEE international conference on pervasive computing and communications workshops*

(*percom workshops*), Austin, TX, USA, 23–27 March 2020. 2020. URL: <https://doi.org/10.1109/percomworkshops48775.2020.9156182>.

28. Clustering-based wake word detection in privacy-aware acoustic sensor networks / T. Koppelman et al. *Interspeech 2022*. ISCA, 2022. URL: <https://doi.org/10.21437/interspeech.2022-842>.

29. Edge computing: vision and challenges / W. Shi et al. *IEEE internet of things journal*. 2016. Vol. 3, no. 5. P. 637–646. URL: <https://doi.org/10.1109/jiot.2016.2579198>.

30. Tishby N., Pereira F. C., Bialek W. The information bottleneck method. 1999. 16 p. (Preprint. 0004057). URL: <https://doi.org/10.48550/arXiv.physics/0004057>.

31. Tishby N., Zaslavsky N. Deep learning and the information bottleneck principle. *2015 IEEE information theory workshop (ITW)*, Jerusalem, Israel, 26 April – 1 May 2015. 2015. URL: <https://doi.org/10.1109/itw.2015.7133169>.

32. From the information bottleneck to the privacy funnel / A. Makhdoumi et al. *2014 IEEE information theory workshop (ITW)*, Hobart, Australia, 2–5 November 2014. 2014. URL: <https://doi.org/10.1109/itw.2014.6970882>.

33. Perero-Codosero J. M., Espinoza-Cuadros F. M., Hernández-Gómez L. A. X-vector anonymization using autoencoders and adversarial training for preserving speech privacy. *Computer speech & language*. 2022. Vol. 74. P. 101351. URL: <https://doi.org/10.1016/j.csl.2022.101351>.

34. A vector quantized variational autoencoder (VQ-VAE) autoregressive neural \$F_0\$ model for statistical parametric speech synthesis / X. Wang et al. *IEEE/ACM transactions on audio, speech, and language processing*. 2020. Vol. 28. P. 157–170. URL: <https://doi.org/10.1109/taslp.2019.2950099>.

35. Allan J., Berg J. Perceptual objective listening quality assessment (POLQA), the third generation ITU-T standard for end-to-end speech quality measurement part ii–perceptual model. *Journal of the audio engineering society*. 2020. Vol. 68, no. 3. P. 193–222. URL: <https://doi.org/10.17743/jaes.2020.0003>.

36. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs / A. W. Rix et al. *2001 IEEE international conference on acoustics, speech, and signal processing. proceedings*, Salt Lake City, UT, USA. URL: <https://doi.org/10.1109/icassp.2001.941023> .
37. P.800. Methods for subjective determination of transmission quality. Replaces P.80 ; effective from 1996-08-30. Official edition. 1996. 37 p.
38. BS.1534. Method for the subjective assessment of intermediate quality level of audio systems. Effective from 2015-10-23. Official edition. 36 p.
39. Malekzadeh M., Clegg R. G., Haddadi H. Replacement autoencoder: a privacy-preserving algorithm for sensory data analysis. *2018 IEEE/ACM third international conference on internet-of-things design and implementation (iotdi)*, Orlando, FL, 17–20 April 2018. 2018. URL: <https://doi.org/10.1109/iotdi.2018.00025> .
40. Speech enhancement based on deep denoising autoencoder / X. Lu et al. *Interspeech 2013*. ISCA, 2013. URL: <https://doi.org/10.21437/interspeech.2013-130> .
41. Neumann M., Vu N. T. Improving speech emotion recognition with unsupervised representation learning on unlabeled speech. *ICASSP 2019 - 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Brighton, United Kingdom, 12–17 May 2019. 2019. URL: <https://doi.org/10.1109/icassp.2019.8682541> .
42. Unsupervised speech representation learning using wavenet autoencoders / J. Chorowski et al. *IEEE/ACM transactions on audio, speech, and language processing*. 2019. Vol. 27, no. 12. P. 2041–2053. URL: <https://doi.org/10.1109/taslp.2019.2938863> .
43. Privacy-Preserving siamese feature extraction for gender recognition versus speaker identification / A. Nelus et al. *Interspeech 2019*. ISCA, 2019. URL: <https://doi.org/10.21437/interspeech.2019-1148> .

44. Privacy-Preserving variational information feature extraction for domestic activity monitoring versus speaker identification / A. Nelus et al. *Interspeech 2019*. ISCA, 2019. URL: <https://doi.org/10.21437/interspeech.2019-1703>.
45. Aloufi R., Haddadi H., Boyle D. Privacy-preserving voice analysis via disentangled representations. *CCS '20: 2020 ACM SIGSAC conference on computer and communications security*, Virtual Event USA. New York, NY, USA, 2020. URL: <https://doi.org/10.1145/3411495.3421355>.
46. Stoidis D., Cavallaro A. Protecting gender and identity with disentangled speech representations. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-2163>.
47. Zhou K., Sisman B., Li H. Vaw-Gan for disentanglement and recomposition of emotional elements in speech. *2021 IEEE spoken language technology workshop (SLT)*, Shenzhen, China, 19–22 January 2021. 2021. URL: <https://doi.org/10.1109/slt48900.2021.9383526>.
48. Niekerk B. V., Nortje L., Kamper H. Vector-Quantized neural networks for acoustic unit discovery in the zerospeech 2020 challenge. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-1693>.
49. Lotfidereshgi R., Gournay P. Biologically inspired speech emotion recognition. *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, New Orleans, LA, 5–9 March 2017. 2017. URL: <https://doi.org/10.1109/icassp.2017.7953135>.
50. Carbon emissions and large neural network training / D. A. P. Patterson et al. 22 p. (Preprint. 2104.10350 Search...). URL: <https://arxiv.org/abs/2104.10350>.
51. Backstrom T., Andreas Nautsch A. N. Comments to 'guidelines 02/2021 on virtual voice assistants' of the european data protection board (EDPB). *EDPB | European Data Protection Board*. URL: https://edpb.europa.eu/sites/default/files/webform/public_consultation_reply/edpb_comments.pdf.

52. Xie Y., Peng S. How to repair customer trust after negative publicity: the roles of competence, integrity, benevolence, and forgiveness. *Psychology and marketing*. 2009. Vol. 26, no. 7. P. 572–589. URL: <https://doi.org/10.1002/mar.20289>.
53. Chen S. C., Dhillon G. S. *Information technology and management*. 2003. Vol. 4, no. 2/3. P. 303–318. URL: <https://doi.org/10.1023/a:1022962631249>.
54. Dwork C., Roth A. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*. 2013. Vol. 9, no. 3-4. P. 211–407. URL: <https://doi.org/10.1561/04000000042>.
55. Warner S. L. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the american statistical association*. 1965. Vol. 60, no. 309. P. 63–69. URL: <https://doi.org/10.1080/01621459.1965.10480775>.
56. A guide to fully homomorphic encryption / F. Armknecht et al. 2015. 35 p. (Preprint. 1192). URL: <https://ia.cr/2015/1192>.
57. A survey on homomorphic encryption schemes / A. Acar et al. *ACM computing surveys*. 2018. Vol. 51, no. 4. P. 1–35. URL: <https://doi.org/10.1145/3214303>.
58. Thaine P., Penn G. Extracting mel-frequency and bark-frequency cepstral coefficients from encrypted signals. *Interspeech 2019*. ISCA, 2019. URL: <https://doi.org/10.21437/interspeech.2019-1136>.
59. Zhang S.-X., Gong Y., Yu D. Encrypted speech recognition using deep polynomial networks. *ICASSP 2019 - 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Brighton, United Kingdom, 12–17 May 2019. 2019. URL: <https://doi.org/10.1109/icassp.2019.8683721>.
60. Cramer R., Nielsen J. B., Damgård I. B. Secure multiparty computation and secret sharing. Cambridge University Press, 2015.
61. Privacy-preserving PLDA speaker verification using outsourced secure computation / A. Treiber et al. *Speech communication*. 2019. Vol. 114. P. 60–71. URL: <https://doi.org/10.1016/j.specom.2019.09.004>.

62. Towards end-to-end private automatic speaker recognition / F. Teixeira et al. *Interspeech* 2022. ISCA, 2022. URL: <https://doi.org/10.21437/interspeech.2022-10672>.
63. When machine learning meets privacy / B. Liu et al. *ACM computing surveys*. 2021. Vol. 54, no. 2. P. 1–36. URL: <https://doi.org/10.1145/3436755>.
64. Feng T., Peri R., Narayanan S. User-Level differential privacy against attribute inference attack of speech emotion recognition on federated learning. *Interspeech* 2022. ISCA, 2022. URL: <https://doi.org/10.21437/interspeech.2022-10060>.
65. Improving on-device speaker verification using federated learning with privacy / F. Granqvist et al. *Interspeech* 2020. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-2944>.
66. Woubie A., Backstrom T. Federated learning for privacy-preserving speaker recognition. *IEEE access*. 2021. Vol. 9. P. 149477–149485. URL: <https://doi.org/10.1109/access.2021.3124029>.
67. Nelus A., Glitza R., Martin R. Estimation of microphone clusters in acoustic sensor networks using unsupervised federated learning. *ICASSP 2021 - 2021 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Toronto, ON, Canada, 6–11 June 2021. 2021. URL: <https://doi.org/10.1109/icassp39728.2021.9414186>.
68. Sotto voce: federated speech recognition with differential privacy guarantees / M. Shoemate et al. 2022. 5 p. (Preprint. 2207.07816). URL: <https://arxiv.org/abs/2207.07816>.
69. Communication-Efficient agnostic federated averaging / J. Ro et al. *Interspeech* 2021. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-153>.
70. Federated learning for ASR based on wav2vec 2.0 / T. Nguyen et al. *ICASSP 2023 - 2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Rhodes Island, Greece, 4–10 June 2023. 2023. URL: <https://doi.org/10.1109/icassp49357.2023.10096426>.

71. Attribute inference attack of speech emotion recognition in federated learning settings / T. Feng et al. 2022. 12 p. (Preprint. 2112.13416). URL: <https://arxiv.org/abs/2112.13416>.
72. Gradient-Leaks: understanding and controlling deanonymization in federated learning / T. Orekondy et al. 2018. 18 p. (Preprint. 1805.05838). URL: <https://arxiv.org/abs/1805.05838>.
73. Donley J., Ritz C., Kleijn W. B. Improving speech privacy in personal sound zones. *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Shanghai, 20–25 March 2016. 2016. URL: <https://doi.org/10.1109/icassp.2016.7471687>.
74. Sound zones as an optimal filtering problem / J. K. Nielsen et al. *2018 52nd asilomar conference on signals, systems, and computers*, Pacific Grove, CA, USA, 28–31 October 2018. 2018. URL: <https://doi.org/10.1109/acssc.2018.8645268>.
75. Wallace D., Cheer J. Combining background noise and artificial masking to achieve privacy in sound zones. *Computer speech & language*. 2021. P. 101285. URL: <https://doi.org/10.1016/j.csl.2021.101285>.
76. Springer handbook of speech processing / ed. by J. Benesty, M. M. Sondhi, Y. A. Huang. Berlin, Heidelberg : Springer Berlin Heidelberg, 2008. URL: <https://doi.org/10.1007/978-3-540-49127-9>.
77. Introduction to speech processing: 2nd edition / T. Bäckström et al. Aalto University, 2022. URL: <https://speechprocessingbook.aalto.fi/>.
78. ISO/IEC 19795-1:2021. Information technology Biometric performance testing and reporting Part 1: principles and framework. Replaces ISO/IEC 19795-1:2006 ; effective from 2021-05-23. Official edition. Geneva. URL: <https://www.iso.org/standard/73515.html>.
79. Brümmer N., du Preez J. Application-independent evaluation of speaker detection. *Computer speech & language*. 2006. Vol. 20, no. 2-3. P. 230–275. URL: <https://doi.org/10.1016/j.csl.2005.08.001>.

80. General framework to evaluate unlinkability in biometric template protection systems / M. Gomez-Barrero et al. *IEEE transactions on information forensics and security*. 2018. Vol. 13, no. 6. P. 1406–1420. URL: <https://doi.org/10.1109/tifs.2017.2788000>.
81. Mackay D. J. C. Information theory, inference & learning algorithms. Cambridge, UK : Cambridge University Press, 2003. 640 p.
82. Preserving privacy in speaker and speech characterisation / A. Nautsch et al. *Computer speech & language*. 2019. Vol. 58. P. 441–480. URL: <https://doi.org/10.1016/j.csl.2019.06.001>.
83. Dwork C. Differential privacy: a survey of results. *Lecture notes in computer science*. Berlin, Heidelberg. P. 1–19. URL: https://doi.org/10.1007/978-3-540-79228-4_1.
84. Low bit-rate speech coding with VQ-VAE and a wavenet decoder / C. Garbacea et al. *ICASSP 2019 - 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Brighton, United Kingdom, 12–17 May 2019. 2019. URL: <https://doi.org/10.1109/icassp.2019.8683277>.
85. Mokbel M. F., Aref W. G., Kamel I. Performance of multi-dimensional space-filling curves. *The tenth ACM international symposium*, McLean, Virginia, USA, 8–9 November 2002. New York, New York, USA, 2002. URL: <https://doi.org/10.1145/585147.585179>.
86. Differentially private speaker anonymization / A. Shahin Shamsabadi et al. *Proceedings on privacy enhancing technologies*. 2023. Vol. 2023, no. 1. P. 98–114. URL: <https://doi.org/10.56553/popets-2023-0007>.
87. Cichy P., Salge T.-O., Kohli R. Extending the privacy calculus: the role of psychological ownership. *ICIS 2014 proceedings*. 2014. Vol. 30.
88. Psychological ownership: a review and research agenda / S. Dawkins et al. *Journal of organizational behavior*. 2015. Vol. 38, no. 2. P. 163–183. URL: <https://doi.org/10.1002/job.2057>.
89. Anthropomorphism, agency, and ethopoeia / C. Nass et al. *INTERACT '93 and CHI '93 conference companion*, Amsterdam, The Netherlands, 24–29 April

1993. New York, New York, USA, 1993. URL: <https://doi.org/10.1145/259964.260137>.

90. Cornelius S., Leidner D. Acceptance of anthropomorphic technology: a literature review. *Hawaii international conference on system sciences*. 2021. URL: <https://doi.org/10.24251/hicss.2021.774>.

91. Thompson S. J. Machine law, ethics, and morality in the age of artificial intelligence. IGI Global, 2021. 266 p. URL: <https://doi.org/10.4018/978-1-7998-4894-3>.

92. Detecting distrust towards the skills of a virtual assistant using speech. / L. Pepino et al. 2020. 5 p. (Preprint. 2007.15711). URL: <https://arxiv.org/abs/2007.15711>.

93. Sound privacy: a conversational speech corpus for quantifying the experience of privacy / P. P. Zarazaga et al. *Interspeech 2019*. ISCA, 2019. URL: <https://doi.org/10.21437/interspeech.2019-1172>.

94. Intuitive privacy from acoustic reach: a case for networked voice user-interfaces / T. Bäckström et al. *2021 ISCA symposium on security and privacy in speech communication*. ISCA, 2021. URL: <https://doi.org/10.21437/spsc.2021-12>.

95. Perception of privacy measured in the crowd – paired comparison on the effect of background noises / A. Leschanowsky et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-2299>.

96. Speech privacy in multiple-bed patient rooms / J. Reinten et al. *Eindhoven University of Technology research portal*. URL: <https://research.tue.nl/en/publications/speech-privacy-in-multiple-bed-patient-rooms>.

97. Brüggemeier B., Lalone P. Perceptions and reactions to conversational privacy initiated by a conversational user interface. *Computer speech & language*. 2022. Vol. 71. P. 101269. URL: <https://doi.org/10.1016/j.csl.2021.101269>.

98. Yeasmin F., Das S., Backstrom T. Privacy analysis of voice user interfaces. *The 27th IEEE conference of open innovations association FRUCT*

(FRUCT27), Trento, 9 September 2020. URL: <https://zenodo.org/doi/10.5281/zenodo.4026513>.

99. Lau J., Zimmerman B., Schaub F. Alexa, are you listening?. *Proceedings of the ACM on Human-Computer Interaction*. 2018. Vol. 2, CSCW. P. 1–31. URL: <https://doi.org/10.1145/3274371>.

100. Lutz C., Newlands G. Privacy and smart speakers: a multi-dimensional approach. *The information society*. 2021. Vol. 37, no. 3. P. 147–162. URL: <https://doi.org/10.1080/01972243.2021.1897914>.

101. Kurz M., Brüggemeier B., Breiter M. Success is not final; failure is not fatal – task success and user experience in interactions with alexa, google assistant and siri. *Lecture notes in computer science*. Cham, 2021. P. 351–369. URL: https://doi.org/10.1007/978-3-030-78468-3_24.

102. So P. Voice content and usability. Book Apart, 2021. 142 p. URL: <https://abookapart.com/products/voice-content-and-usability>.

103. Lewis M. Designing for Human-Agent Interaction. *Ai magazine*. 1998. Vol. 19, no. 2. P. 67. URL: <https://doi.org/10.1609/aimag.v19i2.1369>.

104. Hildebrandt T., Hermann T., Rinderle-Ma S. Continuous sonification enhances adequacy of interactions in peripheral process monitoring. *International journal of human-computer studies*. 2016. Vol. 95. P. 54–65. URL: <https://doi.org/10.1016/j.ijhcs.2016.06.002>.

105. Auditory augmented process monitoring for cyber physical production systems / M. Iber et al. *Personal and ubiquitous computing*. 2020. URL: <https://doi.org/10.1007/s00779-020-01394-3>.

106. Willis L. E. Deception by design. *SSRN electronic journal*. 2020. Vol. 34. P. 76. URL: <https://ssrn.com/abstract=3694575>.

107. Privacy-Preserving voice anti-spoofing using secure multi-party computation / O. Chouchane et al. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-983>.

108. X-Vector singular value modification and statistical-based decomposition with ensemble regression modeling for speaker anonymization

system / C. O. Mawalim et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-1887>.

109. Privacy-Preserving speaker recognition with cohort score normalisation / A. Nautsch et al. *Interspeech 2019*. ISCA, 2019. URL: <https://doi.org/10.21437/interspeech.2019-2638>.

110. Adversarial disentanglement of speaker representation for attribute-driven privacy preservation / P.-G. Noé et al. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-1712>.

111. Woubie A., Bäckström T., Zarazaga P. P. The use of audio fingerprints for authentication of speakers on speech operated interfaces. *2021 ISCA symposium on security and privacy in speech communication*. ISCA, 2021. URL: <https://doi.org/10.21437/spsc.2021-2>.

112. Bai Z., Zhang X.-L. Speaker recognition based on deep learning: an overview. *Neural networks*. 2021. Vol. 140. P. 65–99. URL: <https://doi.org/10.1016/j.neunet.2021.03.004>.

113. Voice-Indistinguishability: protecting voiceprint in privacy-preserving speech data release / Y. Han et al. *2020 IEEE international conference on multimedia and expo (ICME)*, London, United Kingdom, 6–10 July 2020. 2020. URL: <https://doi.org/10.1109/icme46284.2020.9102875>.

114. Kinnunen T., Li H. An overview of text-independent speaker recognition: from features to supervectors. *Speech communication*. 2010. Vol. 52, no. 1. P. 12–40. URL: <https://doi.org/10.1016/j.specom.2009.08.009>.

115. Hansen J. H. L., Hasan T. Speaker recognition by machines and humans: a tutorial review. *IEEE signal processing magazine*. 2015. Vol. 32, no. 6. P. 74–99. URL: <https://doi.org/10.1109/msp.2015.2462851>.

116. Speech pseudonymisation assessment using voice similarity matrices / P.-G. Noé et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-2720>.

117. Pierre C., Larcher A., Jouvét D. Are disentangled representations all you need to build speaker anonymization systems?. *Interspeech 2022*. ISCA, 2022. URL: <https://doi.org/10.21437/interspeech.2022-10586>.
118. Voice privacy through x-vector and cyclegan-based anonymization / G. P. Prajapati et al. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-1573>.
119. Design choices for x-vector based speaker anonymization / B. M. L. Srivastava et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-2692>.
120. Privacy and utility of x-vector based speaker anonymization / B. M. L. Srivastava et al. *IEEE/ACM transactions on audio, speech, and language processing*. 2022. P. 1–13. URL: <https://doi.org/10.1109/taslp.2022.3190741>.
121. Speaker anonymization using x-vector and neural waveform models / F. Fang et al. *10th ISCA speech synthesis workshop*. ISCA, 2019. URL: <https://doi.org/10.21437/ssw.2019-28>.
122. Automatic classification of speaker characteristics / Phuoc Nguyen et al. *2010 third international conference on communications and electronics (ICCE 2010)*, Nha Trang, 11–13 August 2010. 2010. URL: <https://doi.org/10.1109/icce.2010.5670700>.
123. Aloufi R., Haddadi H., Boyle D. Configurable privacy-preserving automatic speech recognition. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-1783>.
124. Streaming On-Device End-to-End ASR System for Privacy-Sensitive Voice-Typing / A. Garg et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-3172>.
125. Evaluating the vulnerability of end-to-end automatic speech recognition models to membership inference attacks / M. A. Shah et al. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-1188>.

126. Privacy-Preserving adversarial representation learning in ASR: reality or illusion? / B. M. L. Srivastava et al. *Interspeech 2019*. ISCA, 2019. URL: <https://doi.org/10.21437/interspeech.2019-2415>.
127. Privacy attacks for automatic speech recognition acoustic models in A federated learning framework / N. Tomashenko et al. *ICASSP 2022 - 2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Singapore, Singapore, 23–27 May 2022. 2022. URL: <https://doi.org/10.1109/icassp43922.2022.9746541>.
128. Liu Y., Kapadia A., Williamson D. Preventing sensitive-word recognition using self-supervised learning to preserve user-privacy for automatic speech recognition. *Interspeech 2022*. ISCA, 2022. URL: <https://doi.org/10.21437/interspeech.2022-85>.
129. Enhancing speech privacy with slicing / M. Maouche et al. *Interspeech*. 2022. URL: <https://inria.hal.science/hal-03369137/document>.
130. Privacy guarantees for de-identifying text transformations / D. I. Adelani et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-2208>.
131. Novotney S., Gu Y., Bulyko I. Adjunct-Emeritus distillation for semi-supervised language model adaptation. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-27>.
132. Habernal I. When differential privacy meets NLP: the devil is in the detail. *Proceedings of the 2021 conference on empirical methods in natural language processing*, Online and Punta Cana, Dominican Republic. Stroudsburg, PA, USA, 2021. URL: <https://doi.org/10.18653/v1/2021.emnlp-main.114>.
133. Privacy-Preserving feature extraction for cloud-based wake word verification / T. Koppelman et al. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-262>.
134. Stoidis D., Cavallaro A. Generating gender-ambiguous voices for privacy-preserving speech recognition. *Interspeech 2022*. ISCA, 2022. URL: <https://doi.org/10.21437/interspeech.2022-11322>.

135. Yang C.-H. H., Siniscalchi S. M., Lee C.-H. PATE-AAE: incorporating adversarial autoencoder into private aggregation of teacher ensembles for spoken command classification. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-640>.
136. Haider F., Luz S. A system for real-time privacy preserving data collection for ambient assisted living. *Proc. Interspeech*. 2019. URL: https://isca-speech.org/archive/pdfs/interspeech_2019/haider19_interspeech.pdf.
137. O'Connor M., Kleijn W. B. Distributed summation privacy for speech enhancement. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-1977>.
138. Cancellation of local competing speaker with near-field localization for distributed ad-hoc sensor network / P. P. Zarazaga et al. *Interspeech 2021*. ISCA, 2021. URL: <https://doi.org/10.21437/interspeech.2021-1329>.
139. Rech S. Multi-Device speech enhancement for privacy and quality : MA thesis. Aalto University, 2022. 72 p. URL: <https://aaltodoc.aalto.fi/items/ffb212c8-df08-4493-a234-0cdc00f4db83>.
140. Adler A., Youmaran R., Loyka S. Towards a measure of biometric information. *2006 canadian conference on electrical and computer engineering*, Ottawa, Ont., 7–10 May 2006. 2006. URL: <https://doi.org/10.1109/ccece.2006.277447>.
141. The privacy ZEBRA: zero evidence biometric recognition assessment / A. Nautsch et al. *Interspeech 2020*. ISCA, 2020. URL: <https://doi.org/10.21437/interspeech.2020-1815>.
142. Guidelines 02/2021 on virtual voice assistants | european data protection board. *EDPB | European Data Protection Board*. URL: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-022021-virtual-voice-assistants_en.
143. PRIORIS: enabling secure detection of suicidal ideation from speech using homomorphic encryption / D. Natarajan et al. *Protecting privacy through*

homomorphic encryption. Cham, 2021. P. 133–146. URL: https://doi.org/10.1007/978-3-030-77287-1_10.

144. Nelus A., Martin R. Gender discrimination versus speaker identification through privacy-aware adversarial feature extraction. *Speech communication; 13th itg-symposium*, Oldenburg, 12 October 2018. URL: <https://ieeexplore.ieee.org/document/8578003>.

145. Schneble C. O., Elger B. S., Shaw D. The Cambridge Analytica affair and Internet-mediated research. *EMBO reports*. 2018. Vol. 19, no. 8. URL: <https://doi.org/10.15252/embr.201846579>.

146. Heikkilä M. A Cambridge Analytica-style scandal for AI is coming. *MIT Technology Review*. URL: <https://www.technologyreview.com/2023/04/25/1072177/a-cambridge-analytica-style-scandal-for-ai-is-coming/>.

147. Culnane C., Rubinstein B. I. P., Teague V. Health data in an open world. 2017. 23 p. (Preprint. 1712.05627). URL: <https://doi.org/10.48550/arXiv.1712.05627>.

148. Day M., Turner G., Drozdiak N. Amazon workers are listening to what you tell alexa. *Bloomberg*. URL: <https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio>.

149. Lu D., Heath C. How abusers are exploiting smart home devices. *VICE - Unbequemer Journalismus und Dokus zu allem, was wichtig ist auf der Welt*. URL: <https://www.vice.com/en/article/d3akpk/smart-home-technology-stalking-harassment>.

150. Hulsen S. Stalkers intimideren ex-partners via slimme camera's, lampen en alarmen. *RTL Nieuws*. URL: <https://www.rtlnieuws.nl/nieuws/nederland/artikel/5345441/stalkers-intimideren-ex-partners-slimme-cameras-lampen-en-alarmen>.

151. Use of stalkerware and spyware apps increase by 93% since lockdown began in the UK. Avast. URL: <https://press.avast.com/use-of-stalkerware-and-spyware-apps-increase-by-93-since-lockdown-began-in-the-uk>.
152. Regulation - 2016/679 - EN - gdpr - EUR-Lex. *The official portal for European data* | data.europa.eu. URL: <http://data.europa.eu/eli/reg/2016/679>.
153. The GDPR & speech data: reflections of legal and technology communities, first steps towards a common understanding / A. Nautsch et al. *Interspeech 2019*. ISCA, 2019. URL: <https://doi.org/10.21437/interspeech.2019-2647>.
154. Civil code - civ. *California Legislative Information*. URL: https://leginfo.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.
155. OVON trustmark/interoperability. *Open Voice Network*. URL: <https://openvoicenetwork.org/>.
156. MyData. *MyData*. URL: <https://www.mydata.org/>.
157. MyData – an introduction to human-centric use of personal data (3rd, revised edition). *The URN resolver of The National Library of Finland*. URL: <http://urn.fi/URN:ISBN:978-952-243-617-7>.
158. Security and privacy in speech communication. *SPSC | ISCA SIG-SPSC*. URL: <https://www.spsc-sig.org/>.
159. Friedman B., Felten E., Millett L. I. Informed consent online: a conceptual model and design principles. *Dada*. URL: <https://dada.cs.washington.edu/research/tr/2000/12/UW-CSE-00-12-02.pdf>