Лабораторія Цифрової Безпеки

МІЖНАРОДНИЙ ФОНД ВІДРОДЖЕННЯ

ПРЯМУЄМО РАЗОМ

Віртуальні друзі:
чи(м) нам загрожують голосові асистенти?

Тетяна Авдєєва

Аналітичний звіт підготовлено за підтримки Європейського Союзу та Міжнародного фонду «Відродження» в рамках спільної ініціативи «Європейське Відродження України». Звіт представляє позицію авторів і не обов'язково відображає позицію Європейського Союзу чи Міжнародного фонду «Відродження»

Увімкнути улюблену пісню? Знайти найближчу аптеку? Вимкнути світло у кімнаті чи поставити будильник? Все це наразі можна робити за допомогою віртуальних асистентів. Ба більше, деякі голосові асистенти вже спричиняли конфузи - наприклад, Siri відправляла повідомлення без погодження з користувачами, а іноді - навіть з заблокованого телефона від осіб, які не є власниками пристрою (просто як реакція на стандартне “hey, Siri!”). Голосовий асистент Gemini, що має замінити Google Assistant, вже здатен контекстуально жартувати (щоправда не завжди добре). Його здатність взаємодіяти з середовищем та генерувати інформацію справді вражає. Використання віртуальних асистентів популярне не тільки для вирішення побутових задач: їх доступність і фізична поширеність часто використовується і для проведення опитувань та досліджень. Якщо помножити це на рівень розвитку технологій цифрових аватарів, цілком можна отримати реалістичні діпфейки, віртуальних асистентів, яких плутають з реальними людьми, або маніпулятивних голосових помічників. Крім того, більшість таких розробок все ж належать приватним компаніям. Тож питання стандартів постає як ніколи гостро, особливо в турбулентні періоди.

У відповідь на ці виклики, Бразилія забороняє створювати віртуальних асистентів, що схожі на конкретних осіб чи взагалі є антропоморфними (людиноподібними) у передвиборчі періоди. Спеціальні правила використання віртуальних асистентів нещодавно були розроблені у Рекомендації Ради Європи щодо ШІ в місцях позбавлення волі та закладах пробації. Зокрема, пропонується використовувати ці технології у сфері охорони здоров'я і освіти, для реабілітації і корекції поведінки. Самі віртуальні асистенти згадуються і в Акті про цифрові послуги (DSA) та Акті про цифрові ринки (DMA) (які нещодавно набрали чинності).

Щоб уникнути жорсткого регулювання та заборон, компанії намагаються розробляти правила на рівні саморегулювання. Наприклад, Hulu, Netflix і Spotify прямо зазначають у своїх політиках приватності про те, що вони можуть отримувати персональні дані, якщо людина взаємодіє із сервісом за допомогою віртуального асистента, а деякі з них навіть мають власних віртуальних помічників. У політиках компанії запевняють, що діяльність відповідає Загальному регламенту про захист даних (далі - GDPR). Для перевірки відповідності законодавству про захист даних навіть розробили віртуального асистента Olivia. Виходить такий собі оксюморон - віртуальний помічник перевіряє відповідність інших віртуальних помічників вимогам щодо приватності.

Тим часом, Міністерство закордонних справ вже запускає цифрового аватара для озвучування офіційних повідомлень. Хтозна - може за кілька років голограма зможе не тільки поширювати наперед заготовлені заяви, а й самостійно взаємодіяти з аудиторією. Зрештою, далеко ходити не потрібно - держава вже оголосила, що застосунок "Дія" скоро матиме свій інструмент ШІ - **віртуального помічника "Надія"**. За задумкою, він буде консультувати користувачів додатка, пояснюючи технічні аспекти або, наприклад, радячи людині, де знайти найближчий ЦНАП. Технічна частина розробляється разом із OpenAI та Microsoft, втім поки що система далека від ідеалу. І хоча запуску проєкту ще не відбулося, швидкість розвитку технологій навряд змусить нас чекати надміру довго, перш ніж перший державний віртуальний помічник почне свою роботу. Тож, постає логічне запитання - а чи є взагалі якісь стандарти, що регулюють діяльність таких систем?

Хто ховається за маскою "Siri"?

Перш ніж аналізувати різні регуляторні акти, давайте з'ясуємо, що саме вважають віртуальним асистентом. Найбільш міжнародним регуляторним актом, який прямо згадує про віртуальних асистентів є DMA. Зокрема, стаття 2 документу вказує, що віртуальні асистенти можуть бути частиною ключових сервісів платформ, а отже - підпадають під його дію. Далі стаття 2(12) дає і визначення віртуального асистента - воно є достатньо технічним, втім дає зрозуміти, про які саме системи йдеться. Тож про кого ми говоримо?

> **Віртуальний асистент** - програмне забезпечення, яке обробляє вимоги, завдання або запитання (включно з тими, що базуються на аудіо-, візуальних або письмових даних, жестах чи рухах), і яке на їх основі надає доступ до інших служб чи керує підключеними пристроями.

На відміну від чат-ботів, які прицільно виконують команди, є менш точними та сильно залежать від вихідних параметрів, віртуальні асистенти мають широке поле для застосування. Так, вони можуть вирішувати побутові завдання, не обмежені конкретними варіантами рішень, є більш ефективними та можуть взаємодіяти з особою, щоб уточнити завдання чи запит. Також варто пам'ятати про різницю між віртуальними асистентами і цифровими аватарами. Хоч іноді асистенти і можуть мати людиноподібну форму (наприклад, віртуальне обличчя або форму тіла), це не є обов'язковою характеристикою.

Перший віртуальний асистент Audrey був розроблений у 1952 році разом з появою систем розпізнавання голосу. На сьогодні існує досить багато прикладів віртуальних асистентів. Не всі з них є популярними чи поширеними - деякі застосовуються секторально, тобто лише в певних сферах життя. Серед них, наприклад, віртуальні асистенти, які використовують у сфері забезпечення правопорядку: ELLA (надання нетермінових сервісів у відділку поліції), Zenext (дистанційне виконання завдань, в тому числі і критичних, органами правопорядку), Apollo AI (внутрішня організація робочого простору). Серед найпопулярніших віртуальних асистентів же можна згадати:

- **Alexa** - віртуальний асистент від компанії Amazon, який є одним з найпоширеніших додатків свого виду і має “режим шепотіння” - тобто посилені здатності розпізнавання голосу у різних форматах;
- **Siri** - віртуальний асистент від компанії Apple, який працює на девайсах з операційною системою iOS та macOS;
- **Google Assistant** - віртуальний асистент від Google (скоро буде замінений Gemini), який здатен мати двосторонні розмови і значною мірою грунтується на технологіях ШІ.

Крім цього, елементи віртуальних асистентів можна знайти у системах “розумних будинків” та “розумних міст”, у системах безпеки і багатьох інших застосунках. Важливо пам'ятати, що такі асистенти не обов'язково мають забезпечувати будь-яку потребу - достатньо того, щоб вони могли виконувати свої функції, маючи три складові: програмне забезпечення, вхідні дані та доступ до інших пристроїв/програм.

Віртуальні асистенти міжнародного законодавця: у пошуках стандартів

Якщо DSA регулює переважно змістовну частину зобов'язань платформ і не містить окремих положень про віртуальних асистентів, то **DMA** встановлює загальну ринкову рамку і суб'єктів. Щоб підпадати під юрисдикцію ЄС і дію DMA відповідно, компанії, що надають сервіси віртуальних асистентів мають відповідати критерію достатньої кількості користувачів - тобто бути достатньо активними на території ЄС. Так, вони мають перетинати позначку 45 мільйонів користувачів щомісяця. Втім, у випадку з віртуальними асистентами - як визначити, що таке активний користувач? Особа, яка має встановлений додаток? Особа, яка використовує його щодня? Додаток до DMA вказує,

що використання віртуального асистенту хоча б один раз на місяць буде достатньо, щоб цей критерій було виконано. При цьому, використанням буде вважатися запитання чи введення даних в асистента, або ж будь-яка інша взаємодія з ним.

Окрім самого по собі визначення таких технологій та кола сервісів, що потрапляють у сферу дії акту, DMA встановлює три основних обов'язки для великих компаній (і, відповідно, віртуальних асистентів, які вони можуть пропонувати як частину свого сервісу):

- Стаття 6(3)(2) зобов'язує компанію уможливити зміни налаштувань за замовчуванням під час першого використання віртуального асистента (це, наприклад, може запобігти автоматичному надсиланню повідомлень, яке часто є дефолтною функцією, і в цілому уможливити встановлення альтернативних віртуальних асистентів);
- Інший обов'язок передбачений статтею 6(7)(1) - забезпечення інтероперабельності між програмним забезпеченням та обладнанням, доступ до якого можливо отримати за допомогою асистента. Це, зокрема, передбачає ефективний обмін даними, що робить технологію корисною та зручною для користувача;
- Стаття 6(10) покладає зобов'язання надавати користувачам ефективний доступ до даних, які згенеровані чи обробляються, коли сервіс використовується.

У статті 6(4) DMA намагається вирішити проблему так званого власного преференціювання (від англ. "*self-preferencing*"), коли, до прикладу, Google Assistant буде віддавати перевагу товарам Google при рекомендаціях, а Siri рекомендуватиме продукцію Apple. І зворотну проблему - коли компанії пропонують єдиною опцією власного віртуального асистента. Іншою новелою, що стосується забезпечення чесної конкуренції, є спроба уникнути перехоплення сервісів компанією, яка пропонує віртуального асистента. Наприклад, коли Apple робить Siri вбудованою функцією, що унеможливлює встановлення будь-яких інших асистентів (так званий "sherlocking"). На додачу до DMA, деякі зобов'язання передбачені DSA - як-от заборона на цільову (таргетовану) рекламу для дітей. Відповідно, за порушення вимог застосовуються досить високі штрафи.

Окрім документів, що регулюють скоріше інфраструктурні питання, змістовне регулювання є у старому-доброму **GDPR**. Враховуючи, що сам по собі акт не здизайнований спеціально, щоб покривати питання віртуальних

асистентів, не так багато норм насправді стосуються подібних технологій. Втім, на рівні принципів і загальних вимог все ж можна виокремити певні “правила гри”. Перш за все, важливо згадати про обов'язок згоди на обробку даних перед тим, як така обробка почалася. Крім того, стаття 22 передбачає право не бути підданим автоматизованим рішенням у індивідуальних кейсах. Це означає, що результат роботи технології не має ґрунтуватися виняткова на профайлінгу чи автоматизованій обробці даних, якщо ця діяльність має юридичні наслідки для особи, без можливості отримати сервіси альтернативно. Також до компаній-власників віртуальних асистентів є вимога повідомляти про інциденти з даними протягом 72 годин після такого інцидента.

Втім, GDPR має ще одне важливе положення - і воно стосується оцінки впливу технологій чи їх складових на безпеку і захист персональних даних (стаття 35(1)) у випадку, якщо такі технології є високоризиковими. Що принципово - така оцінка має відбуватися до початку обробки персональних даних. При цьому, варто згадати і про обов'язок повідомляти наглядові органи у сфері захисту даних про результати таких оцінок ризиків. Проте, як помітно, вимоги все ще є досить загальними та радше рамковими, адже віртуальні асистенти не бралися до уваги при напрацюванні GDPR.

У відповідь на практичну проблему, Європейська рада із захисту даних (EDPB) напрацювала **Керівництва 02/2021 щодо віртуальних голосових асистентів**. Наразі цей документ є одним із найбільш деталізованих регуляторних актів у сфері віртуальних асистентів. Керівництва містять як загальні стандарти та нагадування про застосовність вимог, на кшталт прав на доступ до даних, їх виправлення чи видалення, до сфери віртуальних асистентів, так і більш специфічні проблеми. До них належать ідентифікація користувачів за голосом (та безпекові проблеми, які супроводжують це питання), особливі вимоги щодо профайлінгу (з огляду на те, що більшість віртуальних асистентів надають персоналізовані послуги), обробка біометричних даних та персональних даних дітей тощо. Що важливо, Керівництва наголошують на наявності у віртуальних асистентів функцій машинного навчання через постійну потребу бути у режимі тіньової активності - по суті, слухати, чи не пролунають заповітні активаційні слова “hey, Siri” чи “okay, Google”, які переведуть асистента у активний режим. Тож, після остаточної фіналізації тексту Акту про штучний інтелект, варто очікувати, що його вимоги будуть застосовні й до віртуальних асистентів.

Наразі ж реалізація функцій у сфері захисту даних та моніторинг дотримання стандартів значною мірою залежить від незалежності і загальних

спроможностей наглядових органів на національному рівні. І тут, в Україні, наприклад, постає логічна проблема - що робити за відсутності оновленого законодавства про захист персональних даних? Наразі, **Закон України "Про захист персональних даних"** пропонує ще менше механізмів для захисту даних в контексті віртуальних асистентів, ніж GDPR. По суті, єдиним насправді дієвим положенням залишається вимога отримувати згоду особи на обробку будь-яких персональних даних. Втім, функція активності за замовчуванням, можливість надсилати повідомлення без погодження і багато інших викликів залишаються ігнорованими через застаріле регулювання.

У випадках, коли законодавство не здатне вирішити проблему захисту прав людини, часто вдаються до напрацювання етичних кодексів та стандартів (принцип етики за замовчуванням). Етичні дилеми у сфері віртуальних асистентів неодноразово аналізувалися академіками, представниками індустрії, правозахисниками та незалежними експертами. На основі таких аналізів напрацьовано десятки рекомендацій - принципів розробки і застосування асистентів. Давайте розглянемо декілька найбільш поширених та **фундаментальних принципів**:

- **Дотримання чинного законодавства**. Хоч наразі відсутнє специфічне регулювання віртуальних асистентів у більшості контекстів, норми чинного законодавства все ще застосовні до правових відносин, що виникають внаслідок їх використання. Зокрема, це стосується чинних законів про захист персональних даних, обов'язків у сфері інтелектуальної власності та протидії монополіям тощо. На практиці, наприклад, це передбачає, що набори даних для тренування системи відповідають вимогам у сфері прав людини, компанія не уникає оподаткування, захищає конфіденційну інформацію тощо. Застосовність таких регуляцій також означає, що відповідні правові механізми для, наприклад, відновлення порушених прав також застосовуються - зокрема, норми щодо захисту прав споживачів;
- **Прозорість та здатність до пояснення**. Прозорість у роботі віртуальних асистентів не означає, що компанія мусить розкривати "технічні секрети" того, як асистент працює. Це також не означає, що компанії достатньо просто оприлюднити технічну інформацію і на цьому зупинитися - не пояснюючи, як саме ті чи інші технічні засоби впливають на права. Перш за все, принцип прозорості означає повідомлення користувачів про взаємодію з автоматизованою системою. По-друге, користувачів мають повідомити, як саме система працює - зрозумілою і простою мовою, у стислому форматі і отримавши на підставі таких пояснень

згоду на обробку даних. Зрештою, користувач має розуміти, до яких саме наслідків призведуть ті чи інші його дії: наприклад, що на практиці означатиме дозвіл на автоматичне надсилання повідомлень системою;

- **Справедливість та неупередженість**. По суті, серцевиною цього принципу є вимога запобігати дискримінації та захищати права людини належним чином, уникаючи як свідомого порушення прав вразливих чи маргіналізованих груп, так і несвідомого упередження у алгоритмах (які, наприклад, іноді можуть віддзеркалювати інституційну дискримінацію). Для цього багато експертів наголошує на необхідності залучати як можна більше різних стейкхолдерів до процесу розробки системи, особливо якщо вона орієнтована на використання у публічному секторі. В таких випадках, розробник має виправляти прогалини і проблемні аспекти. Крім того, цей принцип більш прицільно передбачає відсутність упереджень у рекомендаціях від віртуальних асистентів - не лише у значенні уникнення дискримінації, а і у площині захисту від порушенням правил ринкової конкуренції;
- **Безпека і захищеність**. Безпеку слід розуміти у широкому сенсі, включно з правовим та технічним захистом. І якщо в правовому полі йдеться переважно про доступ до даних, то технічний аспект має враховувати вразливість систем до різного роду кібератак, шкідливого програмного забезпечення, фішингу тощо. Також важливим є розташування серверів, на яких зберігаються персональні дані на території держав, що не мають високого індексу порушення прав людини. При цьому, цифрова безпека - завдання і заслуга не лише компанії-розробника. Так, компанії мають регулярно нагадувати користувачам про цифрові небезпеки і сприяти підвищенню цифрової грамотності;
- **Захист приватності**. Як вже зазначалося раніше, загальні правила щодо захисту персональних даних застосовні і до віртуальних асистентів. Втім, важливо розуміти, що специфіка таких технологій передбачає, що персональні дані мають захищатися не лише в момент використання віртуального асистента, а на всіх етапах життєвого циклу системи - тобто і в момент розробки, і коли система видаляється;
- **Відповідальність та нагляд**. Поняття відповідальності у цьому випадку розглядається досить багатовимірно: йдеться про юридичну, дисциплінарну, а також репутаційну відповідальність. За відсутності юридичного механізму притягнення відповідальних за помилки, компанії мають усвідомлювати, як саме виправляти прогалини та розробити ефективні засоби правового та позаправового захисту для користувачів - портал для скарг, строки обробки скарг, наслідки

повторних помилок тощо. Також слід розробити ефективний механізм компенсації шкоди.

Перелік принципів явно не є вичерпним і часто залежить ще й від сфери використання віртуальних асистентів та їх функціональних можливостей. Секторальні етичні стандарти теж можуть бути предметом дискусії (як-от, у сфері охорони здоров'я, у сфері захисту публічного порядку тощо). Втім, сама по собі наявність принципів ще не робить їх виконання, по-перше, повсюдним (адже багато компаній ігнорують ці самонапрацьовані стандарти через відсутність механізму відповідальності і неможливість технічно перевірити їх дотримання), по-друге, часто проблеми виникають на практиці без жодного злого умислу - навіть за рахунок здатності системи до самонавчання. Тож, з чим наразі стикається людство, коли кличе Alexa?

Говоріть тихіше, вас можуть почути!

Проблематичних питань, пов'язаних із використанням віртуальних асистентів наразі є досить багато. Навіть не говорячи про невщухаючий дискурс довкола права генеративного ШІ на свободу вираження та супутні питання захисту авторського права, найбільш кричущими на порядку денному залишаються питання захисту приватності (у практичному розумінні), одвічні проблеми дискримінації автоматизованими системами, а також безпекові питання.

Приватність. Одним з основних викликів у сфері приватності є відповідь на запитання "хто і коли слухає?". Як вже зазначалося, технічно система постійно перебуває у режимі тіньового слухання, аби мати змогу розпізнати запит користувача і перейти в активний режим. Втім, таким чином система може отримувати і надмірну кількість інформації - зокрема, і чутливої. Далеко не всі компанії розкривають, чи зберігаються дані, отримані в режимі тіньового слухання на серверах компанії. Якщо це дійсно так - це означає, що жодна розмова, в якій поруч з вами лежить телефон - не є конфіденційною. Принаймні, для розробника віртуального асистента.

Збір даних є лише частиною проблеми, адже більшість великих компаній зазначають в умовах користування, що вони будуть співпрацювати з національними органами правопорядку. Що це означає на практиці? Наприклад, журналісти розслідувачі мають зустріч з журналістським джерелом у питаннях антикорупційних розслідувань. Намагаючись перешкодити їм або ж дізнатися, звідки журналісти отримують інформацію,

органи правопорядку звертаються до компанії, щоб отримати записи з голосового асистента - і от, вони вже знають усю інформацію.

Іншою поширеною проблемою є автоматичне надсилання повідомлень віртуальними асистентами. Інода система не запитує додаткового дозволу чи не проговорює повідомлення повторно, просто надсилаючи його іншому користувачу. Більшість ситуацій є радше смішними, проте часом трапляються і серйозні конфузи. Наприклад, якщо небажане повідомлення система надішле босу. Інша проблема - нездатність системи чітко розпізнавати голосові команди (відповідно, повідомлення містять неправильні слова). І тут може бути багато причин, одна з яких - брак біометричних даних для тренування системи. Як отримувати достатню кількість чутливих даних і при цьому не порушувати національне законодавство - те ще питання.

І хоча більшість компаній намагається дизайнувати віртуальних асистентів дотримуючись принципу "приватності за замовчуванням", часто цього не досить. Однією з причин є здатність до машинного навчання - і, відповідно, здатність системи змінюватися самостійно. Іншою проблемою є середовище, в якому система функціонує - наприклад, коли віртуальний асистент активується голосовими командами, чи здатен він ефективно розрізняти голоси дітей та дорослих? На практиці, виявляється, що ні. Це означає, що обробка даних дитини без згоди батьків щоразу порушує правила захисту персональних даних. І поки що ефективного рішення цієї проблеми немає. Водночас, переслуховування записів з голосових асистентів реальними людьми для того, щоб виправляти помилки і покращувати роботу асистентів, було визнане порушенням приватності і надміру високим ризиком порушення конфіденційності. Це сталося після рішення німецьких наглядових органів у сфері захисту даних, які з'ясували, що такі повторні прослуховування записів людьми часто тягнуть за собою поширення записів іншим особам - в жарт або через наявність знайомства.

Дискримінація. Питання алгоритмічних упереджень не є новим. Насправді, воно становить одну з найбільших проблем у сфері технологій ШІ: постійно не вистачає даних, набори є нерепрезентативними, на додачу - інституційна дискримінація повсякчас просвічує крізь ніби-то нейтральні системи. Тож упередженість є досить серйозною проблемою навіть незважаючи на те, що технології переважно вважають більш нейтральними за людей. Які проблеми вже вдалося виявити на практиці, окрім нерепрезентативних наборів даних?

- **Етнічна дискримінація**. За рахунок відмінностей у тембрі голосу та інших голосових характеристиках, система дуже часто не розпізнає голоси представників нетипових для регіону мешканців. Наприклад, дуже часто віртуальні асистенти просто “не чують” афроамериканців та індійців через їх тембр голосу, не здатні ефективно розбирати слова чи постійно плутають їх при голосовому наборі. Іншою проблемою є погане розпізнавання системою регіональних і малопоширених мов. Так, неангломовні запити часто гірше обробляються, а повідомлення набираються з більшою кількістю помилок. Це сприяє тому, що менше неангломовного населення використовує віртуальних асистентів і, відповідно, не може отримувати переваги від такої технології. На додачу, віртуальні асистенти часто видають расистські пошукові результати чи навіть генерують расистські жарти (що стало однією з причин паузи Gemini);
- **Гендерна дискримінація**. Окрім очевидної неспроможності часом розпізнавати жіночі голоси чи неправильної ідентифікації гендеру особи (з чим в принципі сучасні системи впоратися не можуть), дослідження вказують навіть на те, що віртуальні асистенти посилюють онлайн-харасмент. Так, системи вдаються до явно сексистських жартів або ж видають пошукові результати, що містять дискримінаційну інформацію. Крім того, загальний дизайн більшості віртуальних асистентів залишає бажати кращого - зокрема, більшість асистентів мають жіночі голоси та назви. На думку багатьох дослідників, це дозволяє сексуалізувати такі системи і робить їх більш привабливими в очах користувачів, хоча очевидно, що аудиторією цих технологій буде здебільшого чоловіче населення. Наразі дискурс щодо гендеру віртуальних асистентів набирає обертів;
- **Дискримінація на підставі віку**. За рахунок нижчого рівня цифрової грамотності, старшим поколінням та дітям часто набагато важче використовувати віртуальних асистентів. Ба більше, проблема полягає і в недостатньому розумінні ризиків, породжених такою технологією, що в подальшому робить ці вікові категорії людей вразливішими. Проблема, втім, виникає через недотримання принципу прозорості і пояснюваності систем - якби розробники чітко давали зрозуміти, як віртуальний асистент функціонує - виникало б менше проблемних ситуацій. Також, дискримінацію старших осіб можна помітити, коли віртуальних асистентів використовують при прийомі на роботу - там системи просто не розглядають заявки від кандидатів у віці.

Водночас, віртуальні асистенти мають і позитивні сторони - наприклад, вони допомагають людям з інвалідністю ефективніше використовувати технології та користуватися суспільними благами. Проте виклики у сфері забезпечення рівності наразі відносно нівелюють позитивний вплив технологій, адже ті, хто значною мірою покладається на або залежить від віртуальних асистентів часто потерпає від цього - наприклад, це може призводити до впливу помічників на голосування осіб на виборах чи породження інших стереотипів, до яких схильна система.

Безпека. Питання цифрової безпеки є актуальним і поза тематикою віртуальних асистентів. Наприклад, питання безпечних паролів, двофакторної аутентифікації, уникання незнайомих листів та повідомлень (як можливого фішингу), використання незахищених мереж - є актуальними у будь-яких обставинах. Втім, є деякі особливості роботи, що роблять системи досить високоризиковими з безпекової точки зору. Давайте розглянемо декілька з них.

- **Доступ до записів зустрічей**. Якщо організація онлайн-зустрічі здійснюється за допомогою віртуального асистента, якщо асистент робить короткий виклад зустрічі чи просто працює під час неї у фоновому режимі - він може записувати усю інформацію, про яку йдеться на такій зустрічі;
- **Доступ до фінансової інформації**. Наприклад, за допомогою віртуального асистента можна відкривати додатки, які є онлайн-маркетплейсами і шукати там потрібні товари. Це означає, що віртуальні асистенти (і відповідно компанії-розробники) мають доступ до банківської інформації і можуть використовувати її. Іншими особами, хто може це робити є зловмисники, які незаконно отримали доступ до віртуального асистента;
- **Доступ до інформації працівниками компаній-розробників**. Наприклад, працівники Google та Apple зазначали, що чули досить багато персональної чутливої інформації, коли перевіряли ефективність роботи голосових асистентів. Серед чутливих даних були записи до лікаря, медична інформація і навіть потенційні договори щодо торгівлі наркотиками;
- **Шкідливе програмне забезпечення**. Хоча подібні атаки потребують досить багато ресурсу - як людського, так і фінансового (для придбання спеціального обладнання), дослідження вказують, що вони не є абсолютно безуспішними. До того ж, найчастіше вони

користуються прогалинами в програмному забезпеченні (тож варто вчасно оновлювати свої пристрої);

- **Реагування на голоси інших людей**. Часто голосові асистенти не до кінця адаптуються до голосу власника пристрою (особливо якщо пристрій є нещодавно придбаним). Як наслідок, інші люди, які активують голосового асистента навіть на заблокованому пристрої, можуть давати йому базові команди, на кшталт відправлення повідомлення комусь.

Тож, поруч зі стандартними правилами цифрової безпеки, варто пам'ятати про додаткові ризики, породжені особливим функціоналом віртуальних асистентів. І такі ризики слід враховувати, коли особа належить до вразливих і маргіналізованих груп, може стати об'єктом переслідувань з боку держави тощо.

Висновки та рекомендації

Законодавчі процеси в Україні йдуть повним ходом. Втім, багато європейських стандартів все ще не було впроваджено. І на сьогодні це не те щоб є проблемою - всі зобов'язання щодо євроінтеграційних процесів мають відповідати суспільним реаліям і зважати на процеси, що відбуваються всередині самого ЄС з імплементацією багатьох актів (досить лише глянути на те, скільки практичних викликів наразі існує щодо DSA). Водночас, бажання розробляти і активно впроваджувати технології, які за декілька років будуть потрапляти у сферу дії цих документів ніхто не скасовував. Зокрема, інтенції зробити віртуального асистента “Надія” і використання віртуальних помічників держслужбовцями мають зважати на майбутнє регулювання і вже сьогодні брати до уваги європейські стандарти. З огляду на це, Лабораторія цифрової безпеки нагадує про необхідність:

- Розробити адекватну регуляторну рамку у сфері захисту персональних даних, а також напрацювати законопроєкти, спрямовані на імплементацію DSA, DMA, та відповідних секторальних регуляцій (як-от Акт про штучний інтелект);
- Надавати користувачам можливість змінювати налаштування віртуального асистента (і самостійно встановлювати налаштування при першому використанні), вимикати його за потреби/бажання, отримувати послуги у альтернативний спосіб, отримувати доступ до власних даних та реалізувати усі супутні права, пов'язані із персональними даними;

- Забезпечити належний рівень технічної безпеки віртуального асистента, зокрема не лише всередині системи, а і щодо фізичного зберігання даних, безпеки походження програмного забезпечення;
- Забезпечити можливість зворотного зв'язку з боку користувачів - зокрема, створити портал для скарг щодо порядку роботи системи, потенційних порушень чи збоїв;
- Дотримуватися рекомендацій щодо використання систем ШІ у публічному секторі при впровадженні віртуальних асистентів державними структурами чи у державним структурах;
- Регулярно здійснювати оцінку впливу діяльності віртуального асистента на права людини, змінювати технічні налаштування та інтерфейс у разі, якщо буде виявлено ризики порушення прав людини;
- Підвищувати рівень цифрової безпеки серед держслужбовців, які використовують віртуальних асистентів в роботі, та українського населення в цілому.

Керівництво для перевірки дотримання прав людини при створенні віртуальних асистентів

Дотримання прав людини є ключовим елементом на всіх етапах життєвого циклу віртуального асистента - починаючи від дизайну і завершуючи вимкненням такої функції. Зважаючи на декларовані наміри впроваджувати подібні технології у державному секторі, перед запуском проєктів потрібно пересвідчитись, що вони відповідають вимогам у сфері прав людини. Для цього Лабораторія цифрової безпеки розробила короткий чекліст:

1. Чи відповідає віртуальний асистент вимогам законодавства про захист персональних даних на всіх етапах життєвого циклу? ☐

- державні органи, органи місцевого самоврядування, а також підприємства/установи/організації, що розробляють віртуального асистента, уповноважені обробляти персональні дані осіб або отримали згоду на обробку даних від особи? ☐
- чи має особа право на доступ до даних, можливість виправити або видалити свої персональні дані, заперечувати проти їх обробки? ☐
- чи належним чином особа повідомлена про те, які дані використовувалися при створенні системи, а також які дані використовує віртуальний асистент під час своєї роботи, як особа може впливати на обсяг даних, що обробляються? ☐
- чи захищена особа від автоматизованого рішення, яке несе для неї правові наслідки, та чи має особа альтернативні способи отримання послуг? ☐
- чи надає віртуальний асистент можливість регулювати кількість даних, яку опрацьовує та зберігає система, а також вимикати функцію профайлінгу? ☐
  - чи навчається система на даних, отриманих від користувачів? якщо так, чи можливо вимкнути таку функцію? ☐
  - чи збиратимуться та оброблятимуться персональні дані дітей, та чи зможуть діти використовувати віртуального асистента для отримання послуг без нагляду дорослих? чи є функціонал для отримання батьківської згоди на обробку даних осіб до 18 років? ☐
  - чи віртуальний асистент записуватиме і зберігатиме інформацію про “фоновий шум”? чи є можливість автоматично фільтрувати і видаляти її? ☐

2. Чи віртуальний асистент тренований на достатньо репрезентативних даних? ☐

- чи відображає віртуальний асистент розмаїття корінних народів та національних меншин, що мешкають на території України, зокрема при пропонуванні результатів чи при підборі необхідної інформації? ☐
- чи доступні послуги віртуального асистента мовами корінних народів? ☐
- чи здатен віртуальний асистент ефективно розрізняти мовні діалекти та регіональні особливості української мови? ☐

3. Чи повідомляють особу (кожного користувача) про те, як саме функціонує система, принцип роботи віртуального асистента, переваги та ризики для прав людини? Чи є таке повідомлення попереднім, зрозумілим і чітким? ☐

4. Чи функція віртуального асистента не увімкнена за замовчуванням, і чи користувачі не наражають конфіденційну інформацію (журналістські чи адвокатські джерела, банківська та лікарська таємниця тощо) на небезпеку? ☐

5. Чи мають користувачі можливість корегувати налаштування віртуального асистента, змінюючи режим роботи (за замовчуванням/під час використання додатка), кількість даних, інтерфейс, можливість швидко отримати послугу чи консультацію від людини тощо? ☐

6. Чи залучалися різні стейкхолдери до процесу обговорення дизайну віртуального асистента, його функціоналу та потенційних ризиків для прав людини? ☐

7. Чи вжили розробники засобів для забезпечення цифрової безпеки? ☐

- існує спеціальний режим доступу для адміністраторів віртуального асистента, який убезпечує від неконтрольованого доступу третіх осіб? ☐
- чи в безпечному місці зберігаються дані осіб, які отримуються внаслідок використання віртуального асистента? ☐
- чи можна увімкнути віртуального асистента голосовою командою і якщо так - чи може це зробити інша особа, крім власника пристрою? чи є механізми запобігання зловживанням? ☐

8. Чи здійснюється регулярний людський нагляд за роботою системи та перевірки її ефективності, надійності та відповідності правам людини? ☐

9. Чи існує механізм скарг на несправності чи помилки віртуального асистента, та чи є він ефективним і доступним усім користувачам? ☐

Лабораторія
Цифрової
Безпеки