Атака чатботів:

штучний інтелект як інструмент фішингу

Минув усього рік від запуску ChatGPT, а злочинці вже навчилися використовувати його для створення фішингових повідомлень.

І хакери, і захисники використовують штучний інтелект і машинне навчання з метою автоматизації атак/захисту, вдосконалення власних дій і адаптації до методів супротивника. Злочинці, зокрема, створюють ботів для фішингових атак або вже й генерують діпфейкові аудіоповідомлення, імітуючи керівників компаній. Поява великих мовних моделей (ВММ), таких як ChatGPT, ще більше змінила правила гри, адже злочинці використовують ці інструменти у фішингових схемах. Водночас на боці захисту ШІ може так само аналізувати згенеровані іншим ШІ тексти і виявляти підробку.

«МтБ» розбирався, що вже вміє штучний інтелект у царині фішингу і захисту від нього.

Масовий spear phishing

У цьогорічному звіті Email Security Risk Report британської компанії ***Egress*** зазначається, що майже 72% опитаних керівників відділів кібербезпеки стурбовані використанням ШІ для створення фішингових листів і кампаній. В іншому звіті, Phishing Threat Trends Report, компанія пояснює, що раніше більшу частину фішингових листів можна було розпізнати через малограмотність їхніх авторів або абсурдні прохання, які там містилися. До запуску ChatGPT ці невигадливі атаки зростали чисельно завдяки постанню відповідної кримінальної екосистеми (Crime-as-a-Service), яка пропонує готові фішингові пакети і зловмисне ПЗ.

Великі мовні моделі ще більше знижують бар'єр, оскільки дозволяють створювати достовірні листи і здатні генерувати зловмисний код, написання якого менш здібним програмістам не до снаги. Це явище особливе вплине на «нижню частину спектру», тобто на недосвідчених кіберзлочинців з обмеженими навичками програмування, а також на тих, хто не володіє вільно мовою обраної жертви.

Мабуть, найбільшою загрозою, про яку водночас найменше говорять, є застосування ВММ для розвідки з метою підготовки дуже сфокусованих атак. Чатбот стає інструментом розвідки на основі відкритих джерел (OSINT), адже може за лічені секунди обнишпорити Інтернет і розшукати інформацію про жертву. Ці дані потім можна використовувати у прийомах соціальної інженерії. Окрім того, ВММ пришвидшує атаки, що дозволяє генерувати їх у більших кількостях, ніж здатна людина.

Компанія ***Persona***, яка спеціалізується на верифікації ідентичності, у цьому контексті зазначає: шахраям відомо, що для того, аби якась жертва клюнула, потрібно розсилати фішингові листи масово. Раніше на те, щоб заготувати листи з прицілом на різних жертв, потрібно було кілька годин або й днів. ШІ може це зробити за лічені хвилини. За допомогою ШІ злочинці можуть проглянути сотні сторінок у соціальних мережах і потім масово контактувати з користувачами.

Microsoft у себе на сайті у статті про те, як ШІ змінює фішинг, вказує, що spear phishing, тобто фішинг, спрямований на конкретну жертву, має значно вищий відсоток успіху, ніж масова розсилка, але в масштабі не дає такої кількості результатів через час, потрібний на підготовку атаки. Зате поєднання ШІ та spear phishing'у дає «шокуючі результати». Нагромадивши запаси персональних даних, отриманих в результаті зламів, злочинці можуть за допомогою ШІ аналізувати цю інформацію і влаштовувати складні таргетовані атаки. Наприклад, знаючи, що клієнт збирається до певного лікаря, шахрай може надіслати йому листа нібито від клініки з проханням підтвердити платіжну інформацію для подальшої оплати медичних послуг, і тим самим отримати від нього дані банківської картки. ШІ може фактично виконувати масові таргетовані фішингові атаки.

Окрім того, прогнозує Microsoft, недалекий той час, коли ми побачимо таргетовану фішингову рекламу, ця технологія вже «за рогом». Наприклад, якщо ви цікавитеся якоюсь музичною групою, вам на пошту надійде VIP-квиток на фестиваль з її участю, який виглядатиме наче справжній, але насправді буде згенерований, щоб видурити дані платіжної картки.

Це вже працює

У квітні німецька фірма ***SoSafe***, яка спеціалізується на навчанні з кібербезпеки, оприлюднила результати дослідження, проведеного з використанням 1500 фішингових шаблонів, половину з яких написали люди, а іншу половину — модель ChatGPT 3.5. Дослідження показало, що 78% людей відкривають листи, написані роботом, і 21% взаємодіє зі зловмисним контентом (посиланням або вкладенням). Понад те, 65% учасників експерименту вдалось надурити, змусивши вводити персональну інформацію на вебсайтах, куди вели посилання. Цей результат демонструє, що люди не в змозі відрізнити листи, згенеровані ШІ, від написаних вручну. При цьому людські фішингові листи дали дещо вищий відсоток натискань (27%), але відсоток тих, які ввели дані, в них виявився нижчим (60%).

Також експеримент SoSafe показав, що ШІ дає змогу генерувати фішингові листи на 40% швидше.

Тут може здатися, що це поки ще лякалки, засновані на прогнозах, моделюванні та розрахунках, адже ВВМ з'явилися зовсім недавно. Проте ШІ-фішинг — це вже цілком реальність. Вже в грудні 2022 року, невдовзі після запуску ChatGPT, фахівці ізраїльської фірми з кіберрозвідки ***KELA*** спостерігали, як певний брокер початкового доступу (кіберзлочинець, який спеціалізується на проникненні в комп'ютерні мережі) запрошував «колег» ділитися ідеями використання цього інструмента для атак на базі соціальної інженерії. В наступні місяці інші учасники розповідали, як ChatGPT допомагав їм генерувати фішингові листи, і демонстрували приклади.

«В руках шахрая ChatGPT перетворюється на довершене зло», — радів один такий діяч, який створив листа від імені якоїсь компанії про те, що рахунок клієнта заблоковано і йому потрібно надати дані для верифікації. Також додав, що чемний і співчутливий тон листа, згенерованого ШІ, виявився дуже помічним. Інший пройдисвіт похвалився, що попросив чатбота написати переконливе рекламне повідомлення. Цікаво, що учасники злочинного форуму обговорювали інші способи соціальної інженерії, зокрема голосовий фішинг (вішинг), і продемонстрували кілька моделей, проте більшість панства була налаштована скептично, стверджуючи, що підробку легко розпізнати.

Egress у своєму звіті як приклад наводить два повідомлення, які демонструють еволюцію т.зв. «афери-419», відомої як «Нігерійський лист». Перше має такий зміст:

«Пане/пані,
Чи можу я побесідувати з Вами приватно? Ми родина з Судану, яка втекла від Політичної нестабільності в нашій країні. Нам потрібна Ваша допомога, щоб інвестувати наші готівкові гроші у гарний бізнес в Країні. Ми справжні і серйозні люди і готові побесідувати з Вами або зустрітись особисто, щоб обговорити це питання, якщо Ви люб'язно приділите нам увагу. При зустрічі я розповім Вам детально, що треба зробити. Також, будь ласка, дайте мені Ваш контактний номер, і ми обговоримо деталі».

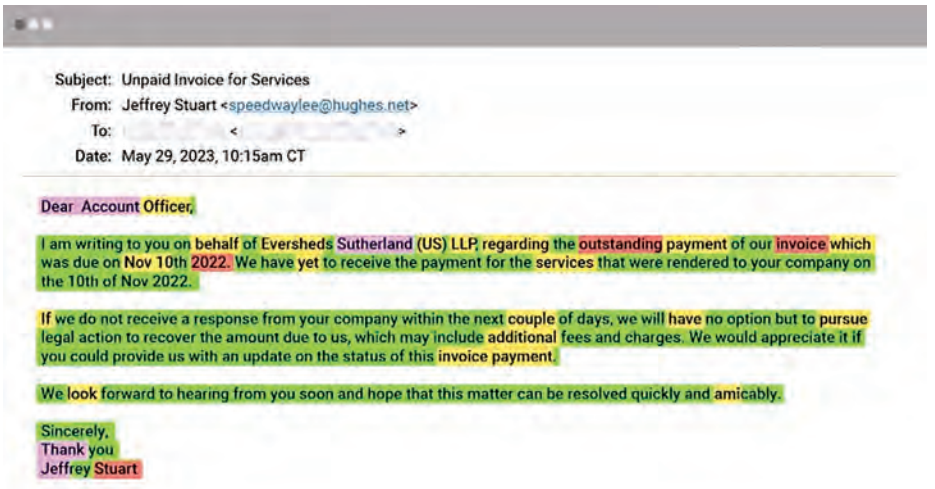
Другий лист, з назвою «Документи про спадщину», виглядає так:

*«Доброго дня, пані ***.*
*Ми з Вашим чоловіком, ***-м, перед його смертю разом працювали над його заповітом, щоб забезпечити виконання його останнього бажання.*

Ми не встигли закінчити роботу до його смерті, але нам повідомили, що він би волів, аби все його майно перейшло до Вас. Нам потрібні підпис і невеличка передоплата, щоб покрити видатки на підготовку цього заповіту, щоб ми могли здійснити його задум.

Будь ласка, перегляньте документи, які додаються у вкладенні, і надішліть $750 на банківський рахунок, який там вказано, щоб ми з адвокатами могли закінчити роботу і зареєструвати цей заповіт».

Перший зразок є очевидним фішинговим листом, який більшість отримувачів одразу розпізнає і видалить. Другий складений більш хитро, він враховує персональну інформацію і створює деталізований, висловлюючись професійною мовою, «претекст» для впливу на вразливу жертву. Дослідники зазначають: неможливо з певністю встановити, що цей другий лист написаний чатботом, але стиль і мова відповідають експериментам, які проводились командою.


Subject: Unpaid Invoice for Services
From: Jeffrey Stuart <speedwaylee@hughes.net>
To: <
Date: May 29, 2023, 10:15am CT

Dear Account Officer,

I am writing to you on behalf of Eversheds Sutherland (US) LLP, regarding the outstanding payment of our invoice which was due on Nov 10th 2022. We have yet to receive the payment for the services that were rendered to your company on the 10th of Nov 2022.

If we do not receive a response from your company within the next couple of days, we will have no option but to pursue legal action to recover the amount due to us, which may include additional fees and charges. We would appreciate it if you could provide us with an update on the status of this invoice payment.

We look forward to hearing from you soon and hope that this matter can be resolved quickly and amicably.

Sincerely,
Thank you
Jeffrey Stuart

Рис. 2. Аналіз фішингового листа від вендора щодо простроченого платежу (джерело: Abnormal)

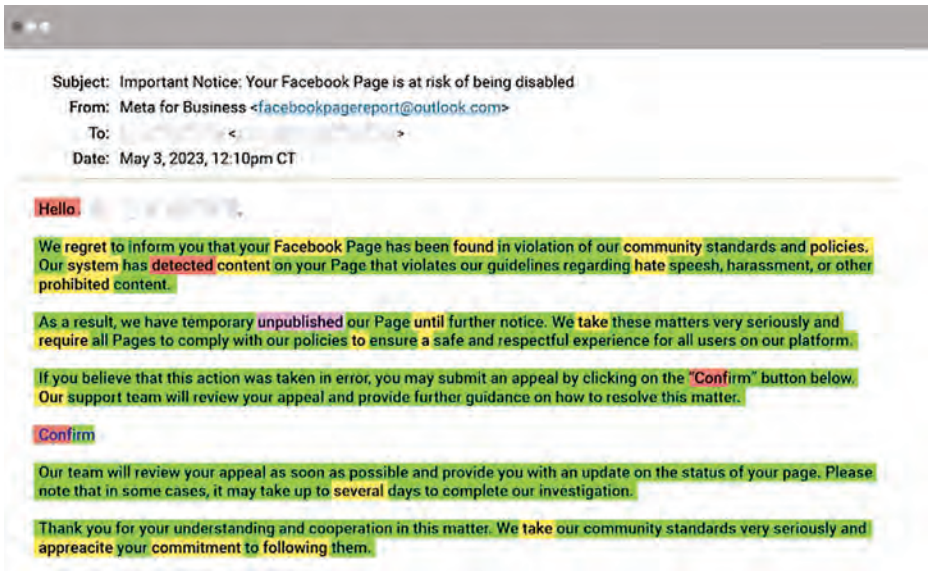
Клин клином

Компанії з кола кібербезпеки самі інтегрують штучний інтелект для виявлення фішингових листів. Детектори на основі ШІ використовують моделі обробки природної мови (NLP) і розуміння природної мови (NLU), які аналізують вміст листів на предмет лінгвістичних маркерів фішингу. (NLP розбирає буквальне значення речень, а NLU розуміє їх контекст. Разом вони використовуються у чатботах).

Наприклад, фірма ***Abnormal Security***, яка займається захистом електронної пошти, у своїй платформі використовує декілька ВММ з відкритим кодом і з їх допомогою аналізує, наскільки можна передбачити кожне слово з урахуванням наявного контексту. Якщо слова у листі послідовно демонструють високу правдоподібність (тобто кожне слово з великою імовірністю збігається з тим, яке видала б модель, і з більшою, ніж у людському тексті), тоді лист можна класифікувати як такий, що написаний чатботом.

Abnormal демонструє цей метод на кількох прикладах фішингових листів, знайдених у «дикій природі». Один з них, надісланий від відправника «Meta for Business», стверджував, що Facebook-сторінка отримувача порушила стандарти спільноти і що для вирішення проблеми йому треба подати апеляцію, перейшовши за посиланням. Ясна річ, воно вело на фішингову форму, де жертва мала ввести свої дані, щоб хакери отримали доступ до її сторінки. Лист написаний без жодної граматичної помилки, зазначають дослідники, а стиль цілком відповідає тому, якого можна очікувати від Meta.

Прогнавши текст через свою платформу, фахівці Abnormal отримали такий результат (**рис. 1**). Тут зеленим кольором виділені слова, які потрапили в топ-10 передбачених, жовтим — в топ-100.


Subject: Important Notice: Your Facebook Page is at risk of being disabled
From: Meta for Business <facebookpagereport@outlook.com>
To: <
Date: May 3, 2023, 12:10pm CT

Hello,

We regret to inform you that your Facebook Page has been found in violation of our community standards and policies. Our system has detected content on your Page that violates our guidelines regarding hate speesh, harassment, or other prohibited content.

As a result, we have temporary unpublished our Page until further notice. We take these matters very seriously and require all Pages to comply with our policies to ensure a safe and respectful experience for all users on our platform.

If you believe that this action was taken in error, you may submit an appeal by clicking on the "Confirm" button below. Our support team will review your appeal and provide further guidance on how to resolve this matter.

Confirm

Our team will review your appeal as soon as possible and provide you with an update on the status of your page. Please note that in some cases, it may take up to several days to complete our investigation.

Thank you for your understanding and cooperation in this matter. We take our community standards very seriously and appreacite your commitment to following them.

Рис. 1. Аналіз фішингового листа про блокування сторінки Facebook (джерело: Abnormal)

Abnormal наводить ще два приклади, коли простий фішинг уже переходить на поле компрометації ділової пошти (BEC). Ці листи не містять посилань чи вкладень, які служили б індикаторами компрометації, і, як і перший, написані дуже професійно.

Один такий лист надіслала в бухгалтерію компанії нібито співробітниця, у якої нібито деактивовано зарплатний рахунок, і вона просить допомогти їй в оновленні інформації. Інший лист (**рис. 2**) належить до категорії компрометації вендорської пошти (VEC), цей прийом соціальної інженерії особливо небезпечний, тому що використовує довіру між вендором і замовником, а також тому, що в їхньому спілкуванні часто зринають рахунки і платежі, тож і запідозрити обман важко, особливо якщо оку нема за що зачепитися. В даному випадку шахрай удає адвоката (з реальної юридичної фірми, що додає правдоподібності), який нагадує про неотриманий платіж і погрожує звернутися до суду.

Abnormal попереджає, що цей метод може давати хибнопозитивні результати. Наприклад, листи, які готуються за шаблоном — маркетингові або рекламні розсилки, — можуть містити фрази, майже ідентичні тим, що їх генерує ШІ. Також детектор буде спрацьовувати на листи, які містять загальновідомі фрази: наприклад, цитати з Біблії або Конституції.

ШІ теж не панацея

Фахівці Egress проаналізували 1,7 млн фішингових листів, щоб визначити, чи містять вони посилання або вкладення, чи лише саме тіло листа (**рис. 3**). Як з'ясувалося, фішингові листи розміром менш за 100 знаків з імовірністю у 90% містять вкладення, листи розміром 100–1199 знаків — посилання. При довжині листа від 1500 знаків (200–375 слів) методом атаки скоріше є соціальна інженерія, а типом «шкідливого навантаження» — сам лист. В діапазоні приблизно 500–1300 знаків вкладення, посилання і соціальна інженерія поєднуються, що відповідає складним атакам з метою компрометації ділової пошти.

До чого це? Як зазначає Джек Чепмен, віцепрезидент Egress з кіберрозвідки, наразі ні людина, ні жоден інструмент не можуть з певністю визначити, чи листа написав чатбот. Оскільки детектори теж використовують ВММ, точність діагностики збільшується з довжиною повідомлення, і часто для цього потрібно щонайменше 250 знаків. До того ж нападники часто застосовують прийоми обфускації (приховування змісту) — наприклад, перефразовують текст, згенерований чатботом. Оскільки ж 44,9% фішингових листів коротші за 250 знаків, і ще 26,5% не дотягують до 500, сучасні детектори не спрацьовують надійно або взагалі не спрацьовують на 71,4% атак. В кінцевому рахунку, зазначає Egress, не важливо, хто склав листа — людина чи бот. Головне, щоб система захисту його розпізнала як шкідливий.

Persona зазначає, що наявним антифішинговим програмам важко обходити листи, згенеровані ШІ, оскільки ці листи не містять ключових слів, за якими ведеться пошук. Ще однією перевагою ШІ є можливість тестувати і коригувати фішингове повідомлення. Якщо фільтр позначає листа як спам, ШІ може внести зміни: наприклад, скоротити текст, щоб він видавався більш професійно написаним.

До появи ШІ на такі постійні модифікації потрібно було багато годин, тепер же шахраї можуть швидко вносити зміни і повторно надсилати листа, отримуючи миттєву реакцію від спаморізки клієнта, і так вдосконалювати текст, аж поки він не пройде крізь фільтр. Це означає, що блокування за ключовими словами і фразами більше не працюватиме, і потрібно шукати ознаки, які ідентифікують самого шахрая.

Власне, Persona пропонує кілька таких рішень, призначених для боротьби з шахрайствами на ресурсах е-комерції (Інтернет-магазини, маркетплейси, електронні курси тощо). Наприклад, коли новий користувач реєструється на сайті, продукт під назвою Verifications звіряє його дані з інформацією від місцевого управління автотранспорту та баз даних світових телеком-операторів, додатково можна встановити мультифакторну автентифікацію або вимагати від користувачів, щоб вони надсилали фото. Інший продукт, Dynamic Flow, дає змогу регулювати дії користувачів на платформі залежно від ризику, який з ними пов'язаний — наприклад, для доступу в певні розділи можна попросити надати документ.

Інструмент під назвою Graph збирає дані про вже зареєстрованого користувача (ім'я, адреса, контактна інформація), а також ті, які він безпосередньо не надає (IP-адреса, ідентифікатор пристрою та геолокація). Якщо ці дані повторюються в кількох акаунтах, це ознака шахрайської мережі. Це дозволяє знаходити патерни поведінки і зв'язки між обліковими записами, які не помітні в масиві даних. А також виявляти шахрайство, якщо користувач за допомогою ШІ створює фейкові повідомлення, вдосконалює їх і масштабує. Таким чином, виявлення шахрайства відбувається не шляхом аналізу граматичних і орфографічних помилок у тексті, а на основі інформації про самих шахраїв.

Загалом же, щоб вберегтися від фішингу — як людського, так і машинного, — фахівці радять дотримуватися простих правил. Обережно ставитися до листів від невідомих осіб, особливо тих, які хочуть грошей. Навіть якщо лист начебто надійшов від відомого контакта, все одно переконатись, що це він, перевіривши адресу відправника та її доменне ім'я. Аналізувати тон, стиль і слововживання відправника, порівнюючи з попереднім листуванням. Перевіряти (наведенням курсора) посилання на предмет того, чи відповідають вони сайтові відправника. Старатися не переходити за посиланнями у підозрілих листах і не завантажувати звідти документів. Використовувати антифішингові програми, нехай вони і не дуже рятують від згенерованих ШІ листів.

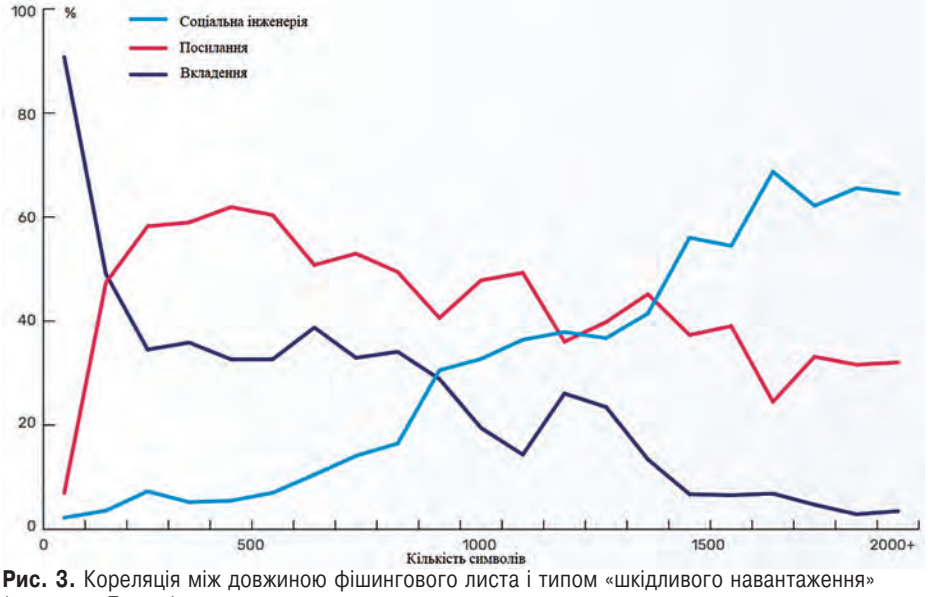


Рис. 3. Кореляція між довжиною фішингового листа і типом «шкідливого навантаження» (джерело: Egress)

Вам дзвонить робот і хоче надурити

Хоча ті хакери з прикладу KELA й не вірили у голосовий фішинг, приклади застосування ШІ для телефонного шахрайства таки мають місце. Насправді успішні приклади почалися за кілька років до появи ChatGPT. У 2019 році, як повідомляла The Washington Post, виконавчому директорові британської енергетичної компанії у п'ятницю ввечері зателефонував шеф з проханням терміново перерахувати гроші постачальнику з Угорщини, аби уникнути штрафу за невчасну оплату рахунків. Директор і його начальник на ім'я Йоханнес до того багато разів

спілкувалися, тож підозр не виникло, адже програма спромоглася зімітувати не лише голос, але й тональність, пунктуацію і навіть німецький акцент Йоханнеса.

Після того, як шахраям було переведено €220 тис., вони подзвонили знову, але цим разом директор щось запідозрив і сам набрав шефа. Фейковий «Йоганнес» подзвонив втретє і попросив покликати директора, саме коли той розмовляв зі справжнім начальником, тож обман розкрився, але грошей вже було не повернути.

На той час, писала WaPo, дослідники з компанії Symantec вже виявили принаймні три спроби імітувати голоси керівників компаній задля видурювання грошей.

У 2021 році Forbes писав про випадок, який трапився роком раніше з керівником гонконгської філії однієї японської компанії, якому зателефонував директор з головного офісу і повідомив, що компанія готує придбання іншої фірми, тож керівник має авторизувати платіж на суму у $35 млн. Для координації всіх процедур винайняли адвоката, і гонконгський менеджер бачив у своїй пошті листи від обох. Нічого не запідозривши, він узявся проводити платежі. Насправді це була афера, яка включала клонування голосу директора за допомогою технології діпфейку.

Більш свіжий приклад, який наводить знову WaPo, стався в Канаді. Літній жінці на ім'я Рут Кард зателефонував хтось з голосом як у її онука Брендона і сказав, що він у в'язниці без гаманця і телефона, і що йому потрібні гроші для застави. Рут з чоловіком кинулися до банку і зняли 3 тис. канадських доларів (денний ліміт). Проте, коли вони спробували зробити те саме в іншому відділенні, їх запросили всередину і повідомили, що інший клієнт отримав схожий дзвінок з таким самим моторошно достовірним підробленим голосом.

В цій історії все скінчилося добре, але не так пощастило батькам такого собі Бенджаміна Перкіна, яким зателефонував буцімто адвокат і повідомив, що їхній син збив на смерть американського дипломата і йому потрібні гроші на, власне, адвоката. Після чого передав телефон самому Перкіну, який сказав, що любить їх і цінує, і попросив грошей. За кілька годин адвокат передзвонив батькам і сказав, що їхньому синові потрібно $21 тис. канадських доларів.

Подружжя пізніше розповідало, що Бенів голос здався їм трохи незвичним, але вони не могли позбутися відчуття, що справді говорили зі своїм сином. В паніці вони оббігли декілька банків і відіслали кошти адвокатові через біткоін-термінал. Того ж вечора Перкін, як завжди робив, подзвонив батькам, щоб поцікавитись їхніми справами, чим дуже їх здивував. Але гроші загули.

Газета стверджує з посиланням на фахівців, що ШІ може відтворити голос на основі аудіосемпла, який містить усього кілька фраз. Дешеві онлайн-інструменти перетворюють аудіофайл у копію голосу, тож шахрай може «вимовити» будь-який текст, який набере. А якщо голос десь відрізняється, це завжди можна пояснити поганим зв'язком. «Два роки тому, щоб клонувати чийсь голос, потрібен був великий обсяг аудіозаписів. Зараз, якщо у вас є сторінка на Facebook або ви записували відео на TikTok і там є 30 секунд вашого голосу, то ваш голос можна клонувати», — наводить WaPo слова Хані Фаріда, професора з Берклі, який працює у сфері цифрових розслідувань.

Вистежити шахраїв, які можуть телефонувати з будь-якої точки світу, складно, так само як позиватися до компаній-виробників, щоб вони несли відповідальність за використання їхніх інструментів.

Як радять протидіяти? Тут теж загалом радять прості речі. Ставитись скептично до неочікуваних дзвінків від близької людини, яка стверджує, що вона в біді. Шахраї намагаються посіяти паніку, щоб жертві було складно тверезо мислити. Можна спробувати підтвердити особу людини, яка телефонує: наприклад, задати якісь питання, на які тільки вона знає відповіді, або зв'язатись з нею іншим каналом, або зв'язатись з іншим членом родини, аби перевірити, що там сталося. Ніколи не розголошувати по телефону важливої інформації (наприклад, щодо банківського рахунку), особливо якщо дзвінок неочікуваний. Не поспішати надсилати гроші у незвичний спосіб і вже з певністю лише після того, як перевірили особу співрозмовника.

Все це не гарантує захист від обману, але береженого Бог береже.

Василь ТКАЧЕНКО, **МтБ**

ХРОНІКА

Що особливого у з'єднувачах постійного струму ArcZero компанії Phoenix Contact?

Системи постійного струму набувають дедалі більшої популярності на ринку, адже вони дають багато переваг: наприклад, зменшення втрат під час перетворення. Однак дотепер розмикання під навантаженням призводило до утворення небезпечної електричної дуги.

Цього не відбувається зі з'єднувачами постійного струму серії ArcZero, які компанія Phoenix Contact планує вивести на ринок вже найближчим часом.

Вони надійно захищають від електричної дуги і таким чином забезпечують безпечне підключення та розмикання під навантаженням. Це дозволяє безпечно вимкнути окремі частини системи, коли потрібно замінити деталі або провести обслуговування пристрою, тоді як решта системи продовжує працювати. Завдяки цьому простої лишаються у минулому. Особливу функцію приховано всередині з'єднувача постійного струму. Інноваційна

електроніка, вбудована безпосередньо у штекер, активно гасить електричну дугу. Таке рішення захищає оператора під час обслуговування обладнання на об'єкті.

https://phoe.co/ArcZero-UA