

ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ ТА БЕЗПЕКА ДАНИХ

Лекція 10

Кібербезпека, посилена ШІ: фішинг, дипфейки, соціальна інженерія



Місток від права до кібербезпеки

ШІ — інструмент, відповідальність за людиною (лекція 9). Але той самий інструмент, що автоматизує бізнес, так само активно використовується проти нього.

- Як саме ШІ змінив ландшафт кіберзагроз?
- Які практичні заходи захисту існують — технічні та організаційні?

Три зміни, які ШІ вносить у кіберзагрози

1

Масштаб

10 000 персоналізованих повідомлень за секунди замість 10 000 окремих зусиль

2

Якість

Зникають класичні маркери шахрайства — граматичні помилки, незграбні формулювання

3

Доступність

Інструменти соціальної інженерії доступні будь-кому за кілька кліків

Та сама здатність писати «гарний людський текст» (лекція 2), що робить ШІ корисним, робить фішинг небезпечнішим.

Фішинг до і після ШІ

До ШІ

- Граматичні помилки, дивне форматування
- Неприродні звертання
- Підозрілі посилання з очевидними помилками
- Легко впізнати за навченими правилами

Після ШІ

- Бездоганно грамотний текст
- Персоналізація під конкретного отримувача
- Точна імітація корпоративного тону компанії
- Класичні маркери шахрайства зникають

Правило «шукай помилки в тексті» застаріло і більше не є надійним індикатором фішингу.

Соціальна інженерія, посилена ШІ



Приклад: лист, що згадує реальне ім'я керівника жертви, недавню новину про компанію і стиль спілкування, типовий для галузі — все автоматично зібране й скомпоноване ШІ.

- Соціальна інженерія — маніпуляція людьми для отримання доступу, обходячи технічний захист
- ШІ автоматизує збір і аналіз публічної інформації для створення переконливого «приводу» атаки

Голосове клонування та дипфейк-відео

Голосове клонування

Кількох секунд аудіозразка (напр. з публічного відео) може бути достатньо для базового клонування голосу конкретної людини.

Дипфейк-відео

Накладає обличчя чи міміку однієї людини на відео іншої. Ознаки, що раніше видавали підробку (моргання, тіні), поступово зникають (пор. лекцію 2).

Практичний ризик: дзвінок голосом реального керівника з терміновим проханням переказати кошти чи розкрити конфіденційну інформацію.

БЕС-шахрайство: класика, посилена ШІ

Business Email Compromise — зловмисник видає себе за керівника чи довірену особу, щоб змусити співробітника здійснити переказ коштів чи розкрити дані.

Класичний сценарій

Лист нібито від CEO — терміновий, «конфіденційний», обходить звичайні процедури погодження через тиск авторитету

Посилений ШІ сценарій

+ бездоганний лист у стилі керівника + підтверджувальний дзвінок чи відео з клонованим голосом/дипфейком

Чому людина залишається найслабшою ланкою

Авторитет

Лист «від керівника»

Терміновість

Тиск на швидке рішення без перевірки

Соціальний доказ

«Інші колеги вже погодили»

Технічні системи стають досконалішими, тому зловмисники атакують не технологію, а людську психологію — ту саму, що ми обговорювали в лекції 4 («впевнений тон переконує»).

Технічні методи захисту

1

SPF / DKIM / DMARC

Автентифікація електронної пошти — ускладнює підробку адреси відправника

2

Багатофакторна автентифікація (MFA)

Додатковий фактор захищає обліковий запис навіть при скомпрометованому паролі

3

Виявлення аномалій

Аналіз типової поведінки користувача чи трафіку із сигналом про відхилення

Жоден технічний захід не є стовідсотковим — технічний захист має доповнюватися організаційними процедурами.

Організаційні методи захисту

- ✓ Протокол зворотного дзвінка: підтверджуйте запит за заздалегідь відомим номером
- ✓ Принцип чотирьох очей: підтвердження двох незалежних осіб для великих сум
- ✓ Регулярне навчання з актуальними прикладами ШІ-атак
- ✓ Чіткі процедури ескалації підозрілих запитів без страху «витратити чужий час»

ШІ працює і на захисті

Атака

Фішинг, соціальна інженерія, дипфейки, голосове клонування

Захист

Спам-фільтри й системи виявлення шахрайства так само на основі ML/ШІ (пор. лекцію 6)

«Гонка озброєнь» між атакувальним і захисним застосуванням ШІ — постійний процес без остаточної переваги жодної сторони.

Чи можна надійно розпізнати ШІ-контент?

- **Технології водяних знаків (watermarking):**
 - приховані технічні маркери в зображеннях, аудіо, іноді тексти
- **Обмеження:**
 - можна видалити технічно підкованому зловмиснику; стандарти не універсальні; для тексту особливо ненадійно

Не покладайтеся виключно на технічні інструменти «виявлення ШІ» — вони недосконалі в обидва боки.

Чек-лист розпізнавання підозрілого контенту

- 1 Чи є елемент терміновості чи тиску ухвалити рішення негайно?
- 2 Чи запит стосується грошей, доступу чи розкриття чутливої інформації?
- 3 Чи можу я перевірити відправника незалежним каналом?
- 4 Чи узгоджується запит із типовою поведінкою людини чи процедурами?
- 5 Чи можу я поставити несподіване контрольне питання (голос/відео)?

Економічний вимір кіберзлочинності

Пряма шкода

Втрачені кошти від успішного ВЕС-шахрайства за один інцидент

Непряма шкода

Розслідування, юридичний супровід, зростання страхових премій

Профілактика

Технічний захист і навчання — операційні витрати бюджету

Оцінка кіберризиків дедалі частіше стає частиною звичайного фінансового планування — а не вузькою темою виключно IT-відділу.

Як це пов'язано з попередніми лекціями

Лекція 2 · Якість генерації

Здатність писати переконливий текст робить фішинг небезпечнішим

Лекція 4 · Психологія, дипфейки

Технічна основа для соціальної інженерії й дезінформації

Лекція 8 · Дуальність технології

Та сама технологія генерує і корисний, і шкідливий контент

Лекція 9 · Відповідальність

Жертва BEC через недотримання процедур теж частково відповідальна

ПІДСУМОВУЄМО

Що варто запам'ятати з цієї лекції

- ШІ змінює масштаб, якість і доступність кіберзагроз — фішинг без помилок, персоналізована соціальна інженерія
- Голосове клонування й дипфейки роблять ВЕС-шахрайство переконливішим, ніж будь-коли
- Людина залишається найслабшою ланкою — атакують психологію (авторитет, терміновість), а не технологію
- Протокол зворотного дзвінка ефективний навіть проти дуже переконливих дипфейків
- Технології виявлення ШІ-контенту недосконалі — не покладайтесь на них виключно

Наступна лекція: організація безпечного впровадження ШІ в компанії — політики, ролі, відповідальність.