Deepfake-детектив: як не дати ШІ себе обманути 🕵️

Воркшоп з цифрової безпеки для тих, хто не хоче стати маніпуляцією. Навчіться розпізнавати підробки, які можуть змінити вашу реальність.

Твої очі тобі брешуть?

Сьогодні ШІ може підробити голос за 3 секунди запису. А обличчя? Ще швидше. Технологія вже настільки досконала, що професійні детективи помиляються.

Питання до залу: Хто готовий поставити 100 грн, що це відео — реальне?

https://www.tiktok.com/@tom.cruise3492/video/7606967385001643295

Анатомія цифрової брехні

Що таке Deepfake?

Це коли нейромережа «натягує» обличчя або голос однієї людини на дії іншої. ШІ навчається на тисячах фото і відео, а потім створює переконливу підробку.

1

Діпфейк — це технологія, заснована на штучному інтелекті, що дозволяє замінювати обличчя та голоси людей на відео та аудіо з практично абсолютною реалістичністю. Спочатку діпфейки використовувалися в розважальній сфері: замінити обличчя відомого актора своїм в улюбленому фільмі або підставити голос знаменитості, щоб розіграти друзів, — весело ж!

2

Однак сьогодні діпфейки стають серйозною загрозою: їх використовують для шахрайства, політичної дезінформації та навіть шантажу. Уявіть, як легко повестися на такий «гачок», коли вам дзвонять із голосом близької людини й в сльозах вимагають гроші, щоб урятувати життя. Страшно? Ще б пак

5 ознак того, що перед тобою «робот» 🤖

Тренуйте свій зір! Ось візуальні і аудіо-сигнали, які викрадають фейки:

1

Очі

Немає кліпання або дивний погляд — очі виглядають «мертвими» і нерухомими

2

Зуби

Виглядають як одна біла смуга без деталей — класичний «монозуб» ШІ

3

Вуха та окуляри

ШІ часто «зажовує» дрібні деталі, вони виглядають розмитими або викривленими

4

Тіні

Світло падає не так, як на фон — фізичні закони порушуються

5

Голос

Роботизована інтонація або занадто чистий звук без природних пауз

У що ми віримо дарма?

Міф: «Я не зірка, мене не підроблять»

Реальність: Шахраям байдуже! Вони полюють на твоїх друзів, батьків, колег. Достатньо 5 хвилин відео зі соцмережі.

Міф: «Антивірус розпізнає фейк»

Реальність: Твій мозок — єдиний надійний фільтр. Немає програми, яка на 100% визначить deepfake. Критичне мислення важливіше за антивірус.

Час розслідування! 🔎

Практикум «Анатомія фейку » — перевірте свої навички детектива на реальних прикладах.

01

Кейс № 1
https://youtu.be/6VhX-s8Xb74

02

Кейс №2: «Папа в хайповому пуховику»
Контент: Pope Francis in a Puffer Jacket

03

Кейс № 3 · Boston Dynamics Do You Love Me

Не існує універсального способу з першого погляду точно визначити, чи є відео реальним, чи згенерованим ШІ. Найкраще, що можна зробити, - не вірити беззастережно всьому, що ви бачите в інтернеті. Якщо щось здається підозрілим або "неприродним", найімовірніше, так воно і є.
В умовах, коли мережа заповнюється ШІ-контентом сумнівної якості, найкращий захист - уважний перегляд. Варто звертати увагу на спотворений текст, об'єкти, що зникають, і рухи, що порушують закони фізики. І якщо ви все ж іноді потрапляєте на фейк - не варто звинувачувати себе: навіть експерти не завжди можуть розпізнати підробку

Інструменти для перевірки

- Google Lens для пошуку оригіналів
- Deepware для аналізу відео
- Верифікація джерел

Стань творцем дінфейку (безпечно!)

Творчий блок: «Зламай систему»— познайомтеся з інструментами, які використовують для створення deepfake. Зрозуміти технологію — значить знати, як їй протистояти.

Reface

Додаток для заміни обличчя в відео.
Створіть мема зі своїм обличчям на відомій персоні.

ElevenLabs

Створення голосу за зразком.
Напишіть 3 секунди — отримайте відтворення своєї інтонації.

Мета практикуму

Зрозуміти, наскільки це швидко і просто.
Якщо ви можете це зробити за 5 хвилин — шахраї тим більше.

Твій Цифровий Бронежилет

Як не стати жертвою?

Секретне слово

Домовся з батьками, друзями про пароль для екстрених дзвінків. Якщо хтось подзвонить і скаже, що це ти, але не знає пароль — це фейк.

Правило 30 секунд

Не вір першій емоції! Перевір джерело, подивись на контекст, не панікуй. Шахраї розраховують на швидку реакцію.

Приватність

Закрий профілі в соцмережах, обмеж свій цифровий слід. Чим менше фото у публічному доступі — тим складніше ШІ створити твій клон.

Запам'ятай: Перевіряй джерела, не поспішай з висновками, обмежуй доступ до особистої інформації.

Дякую, Детективе! 🤝

Ви вже не просто користувач — ви цифровий детектив, який бачить за межами пікселів.

QR-коди для тебе

- Анкета зворотного зв'язку

create.kahoot.it

Kahoot!

Create interactive quizzes, polls, presentations, and more to engage your audience.

Будь розумнішим за алгоритм!