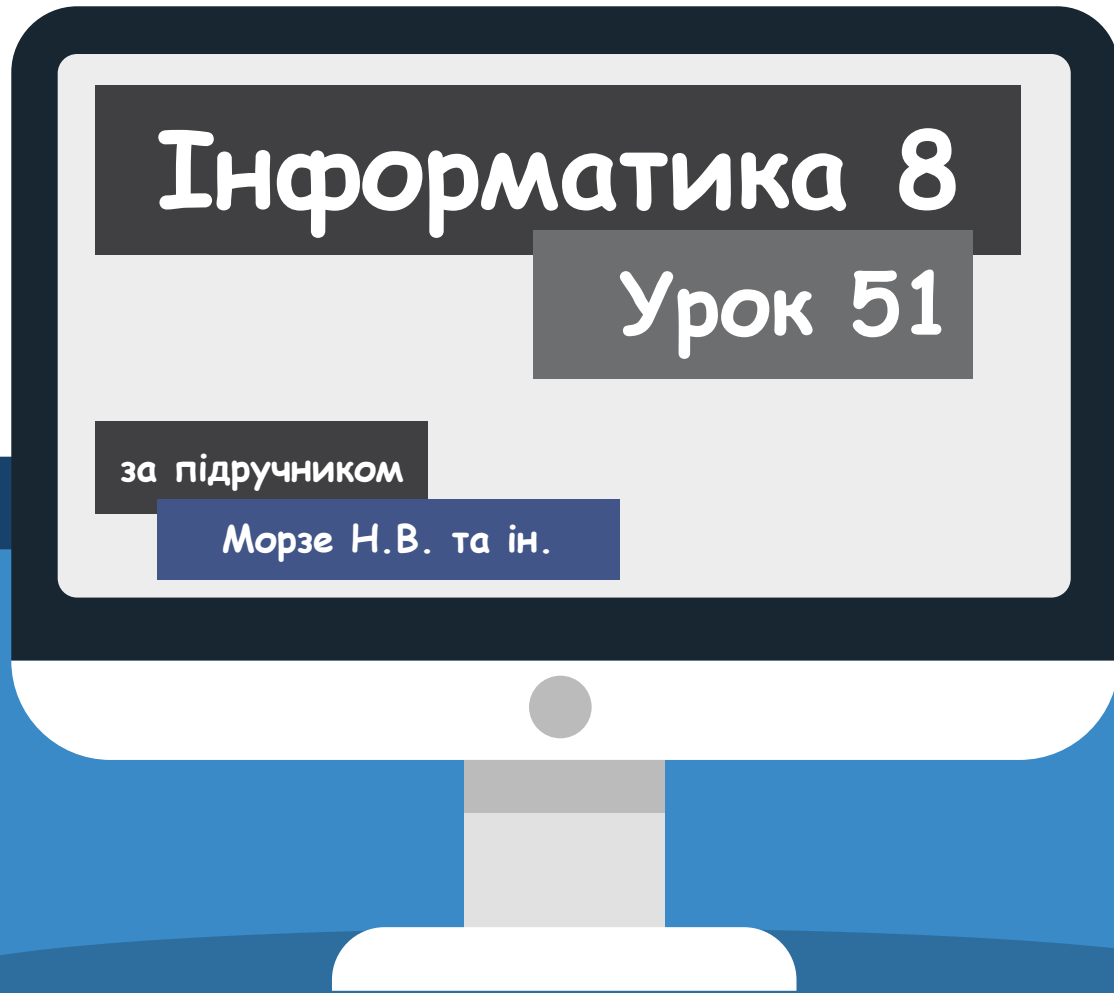


Інструменти для виявлення недостовірних даних



Нова українська школа



Як виявити факт втручання в дані?

Мотиваційний вступ

- Уявіть, що ви дивитеся відео, де відомий актор розповідає щось дуже дивне, або бачите фото, на якому ваш друг начебто побував на Місяці. Це може видатися правдою, але це не більше, ніж добре створений обман – дипфейк.

Завдання

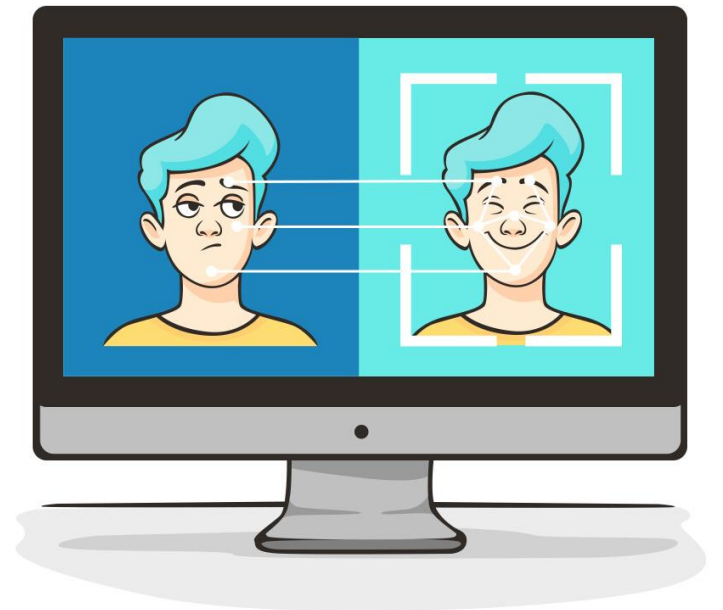
- Розуміючи, як працюють дипфейки і як їх виявляти, ви не лише захищаєте себе від обману, а й підтримуєте збереження правди у світі, де технології змінюють наше життя. Хто знає, можливо, ваші знання одного дня стануть ключовими для того, щоб викрити велику брехню!

Що таке дипфейк?

Назва **дипфейк** (англійською *deepfake*) походить від поєднання інших англійських слів — *deep learning* та *fake*. У буквальному перекладі це означає «глибинне навчання» та «підробка, дезінформація».

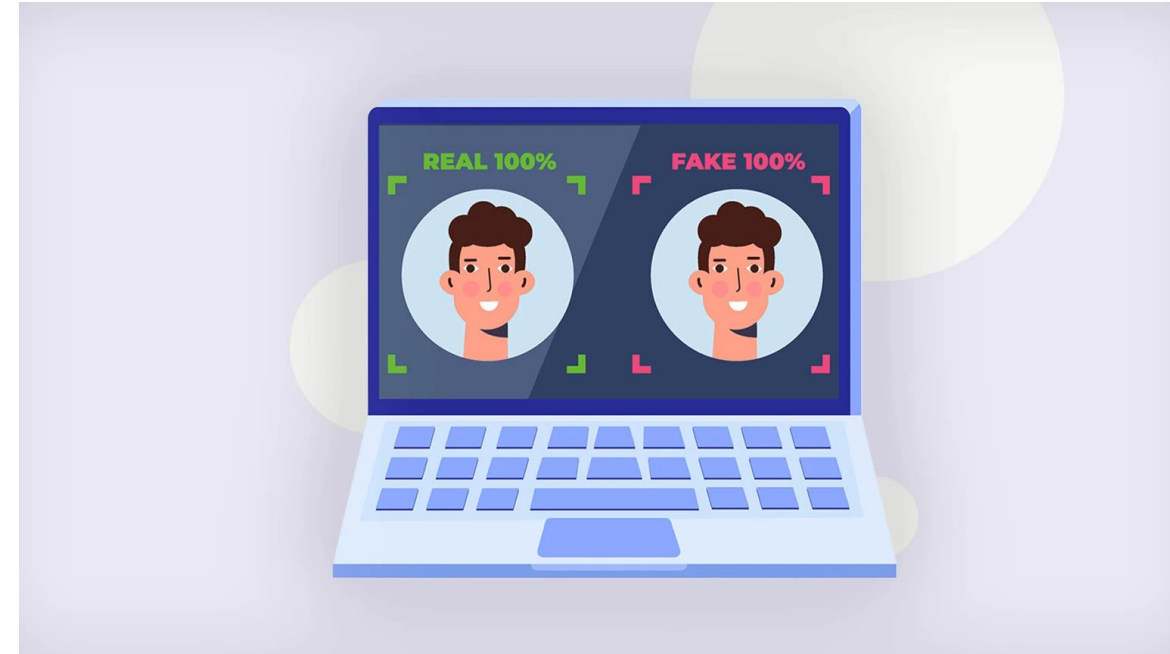


Дипфейк — це методика синтезу зображень людей, що реалізується за допомогою штучного інтелекту.



Що таке дипфейк?

Також дипфейками називають і створені фальшиві відео, зображення, аудіо або інший цифровий контент із реальними людьми «у головних ролях».



Такі матеріали можуть бути надзвичайно реалістичними. Підробки, створені за допомогою сучасних технологічних засобів, іноді майже неможливо візуально відрізнити від реального відео, фото чи аудіо.

Що таке дипфейк?

Прикладом редагування фото є сервіси заміни облич, наприклад:

AKOOL

akool.com

Reface

reface.ai

Зазначені технології використовують методики комп'ютерного зору та алгоритми глибокого навчання, такі як генеративні змагальні мережі (GANs), щоб замінити обличчя однієї людини обличчям іншої на фото чи відео.

Що таке дипфейк?

Наприклад, заміну фото Джека Горобця з відомого фільму на фото Жанін Генрікс зображено на малюнку.



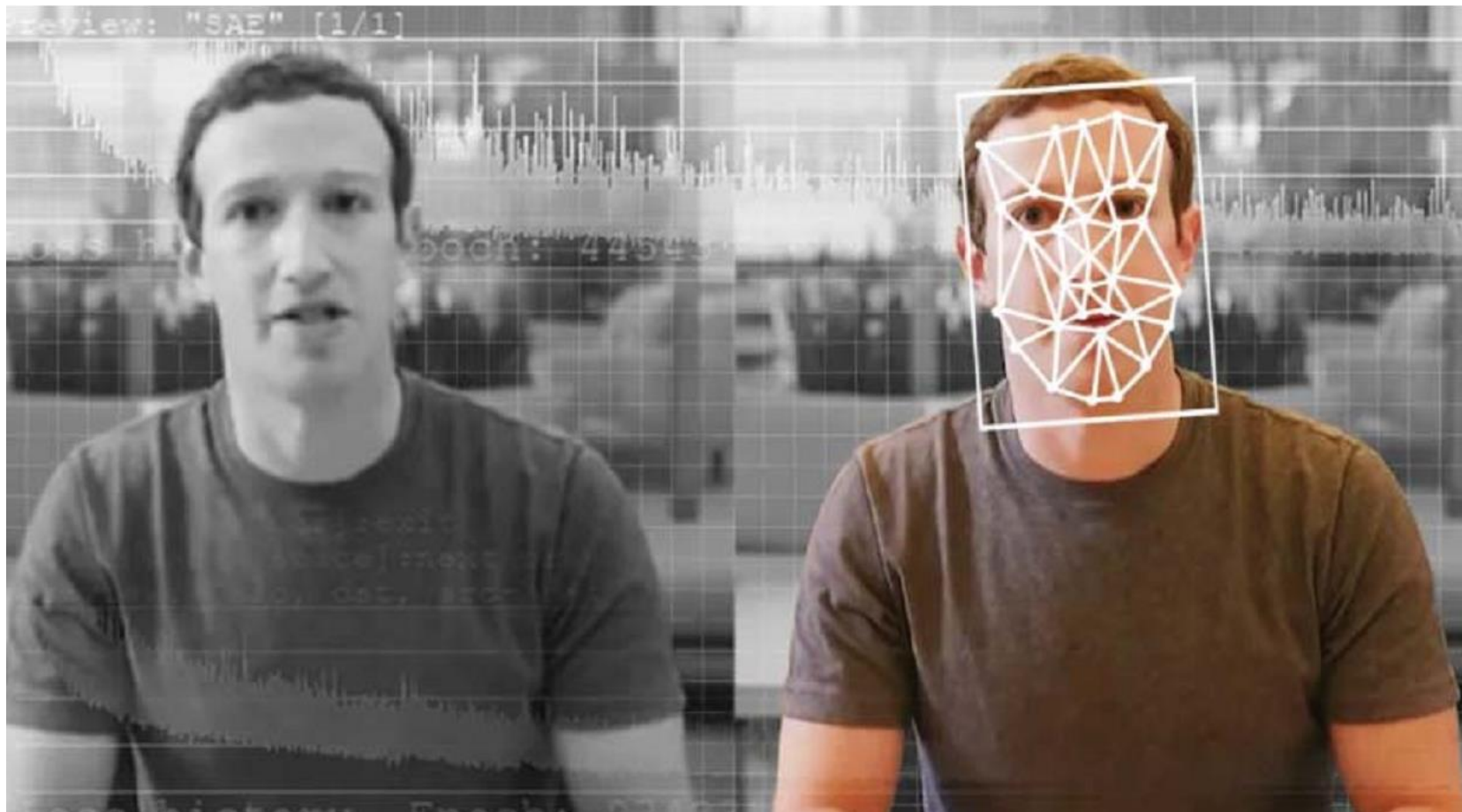
Як створюють дипфейкове відео?

Програми створення дипфейкового відео дають змогу повністю відтворити зовнішність людини. Для цього штучний інтелект аналізує велику кількість фото певної людини, збирає дані та відтворює зображення.



Як створюють дипфейкове відео?

**Створена
«цифрова маска»
накладається на
відео, при цьому
також
відбувається
синхронізація
голосу, жестів,
рухів.**

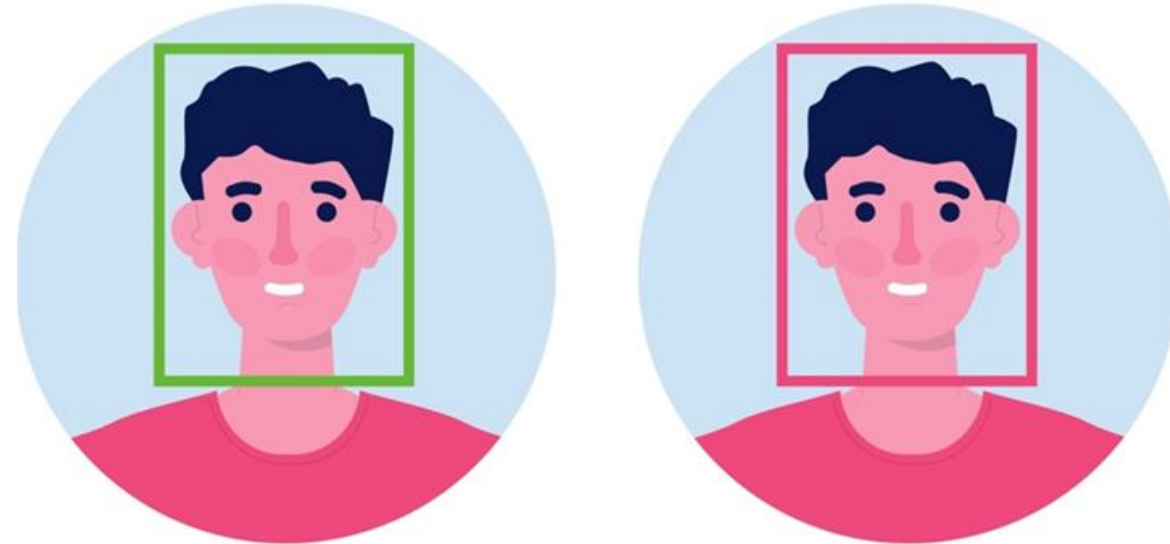


Як створюють дипфейкове відео?

Щоб зробити відео із заміною обличчя, потрібно кілька кроків:

1) необхідні тисячі різних знімків обличчя двох людей, створених за допомогою алгоритму штучного інтелекту, який називається кодером.

Кодер знаходить та вивчає подібності між цими двома об'єктами і зводить їх до спільних загальних рис, стискаючи зображення в процесі;

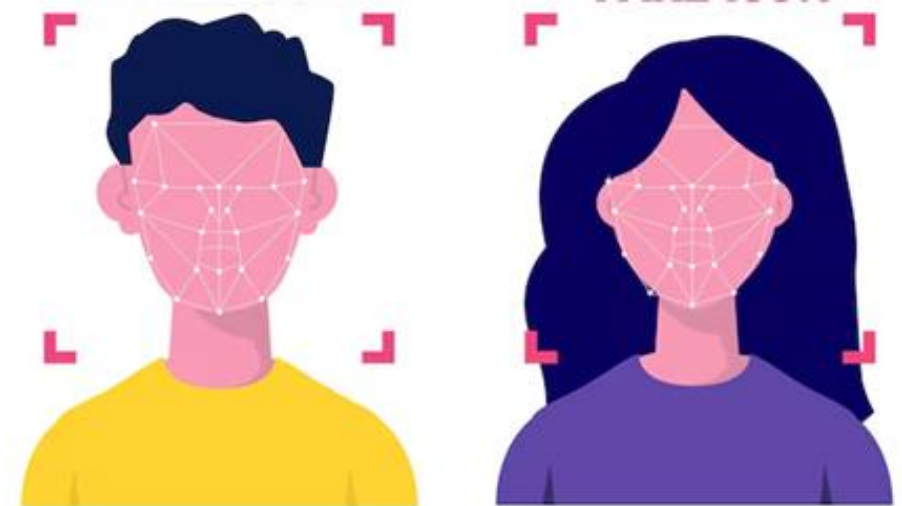


Як створюють дипфейкове відео?

Продовження...

2) використати алгоритм штучного інтелекту, який називається декодером і відновлює обличчя, перетворені у стиснені зображення.

Цей процес відбувається таким чином: оскільки обличчя різні, тренується один декодер для відновлення обличчя першої особи, а інший декодер — для відновлення обличчя другої особи;

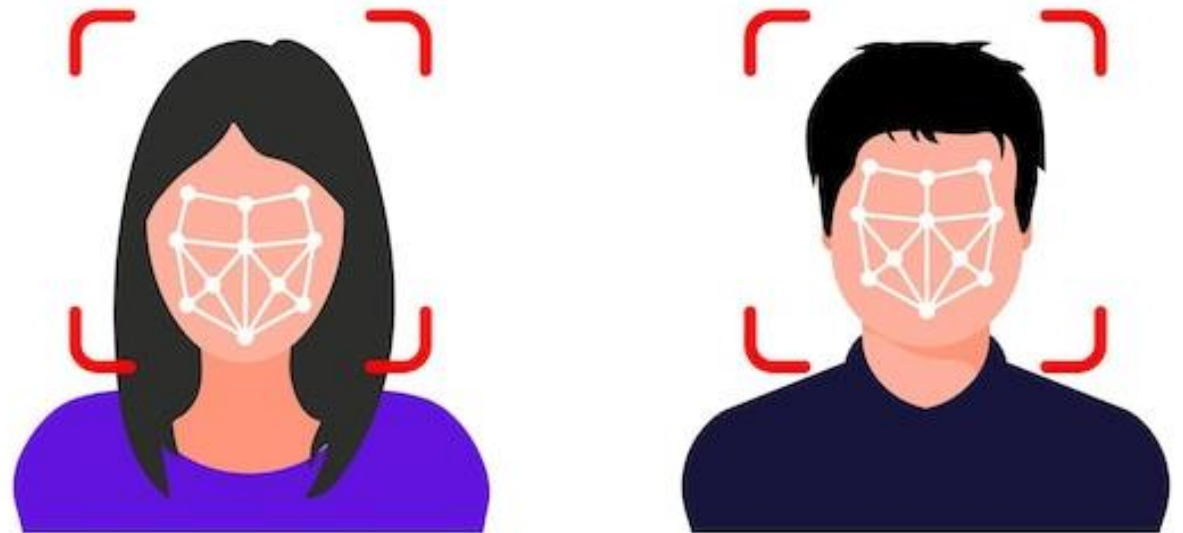


Як створюють дипфейкове відео?

Продовження...

3) щоб виконати заміну обличчя, просто подають закодовані зображення в «неправильний» декодер.

Наприклад, стиснене зображення обличчя людини А надходить у декодер, навчений на особі В. Декодер потім реконструює обличчя особи В з виразами та орієнтацією обличчя А.



Музею Сальвадора Далі знадобилося 1000 годин для машинного навчання, щоб створити їхню дипфейк версію Сальвадора Далі. За допомогою цього відвідувачі музею можуть дізнаватися більше про мистецтво від самого митця.



Для чого створюються дипфейки?

Здебільшого про дипфейки прийнято думати як про інструмент дезінформації, чорного піару та введення в оману для якихось недобрих справ. Наприклад, їх використовують шахраї та злочинці, шантажуючи фальшивими матеріалами своїх жертв (і цими жертвами можуть бути звичайні люди, необов'язково відомі).



Для чого створюються дипфейки?

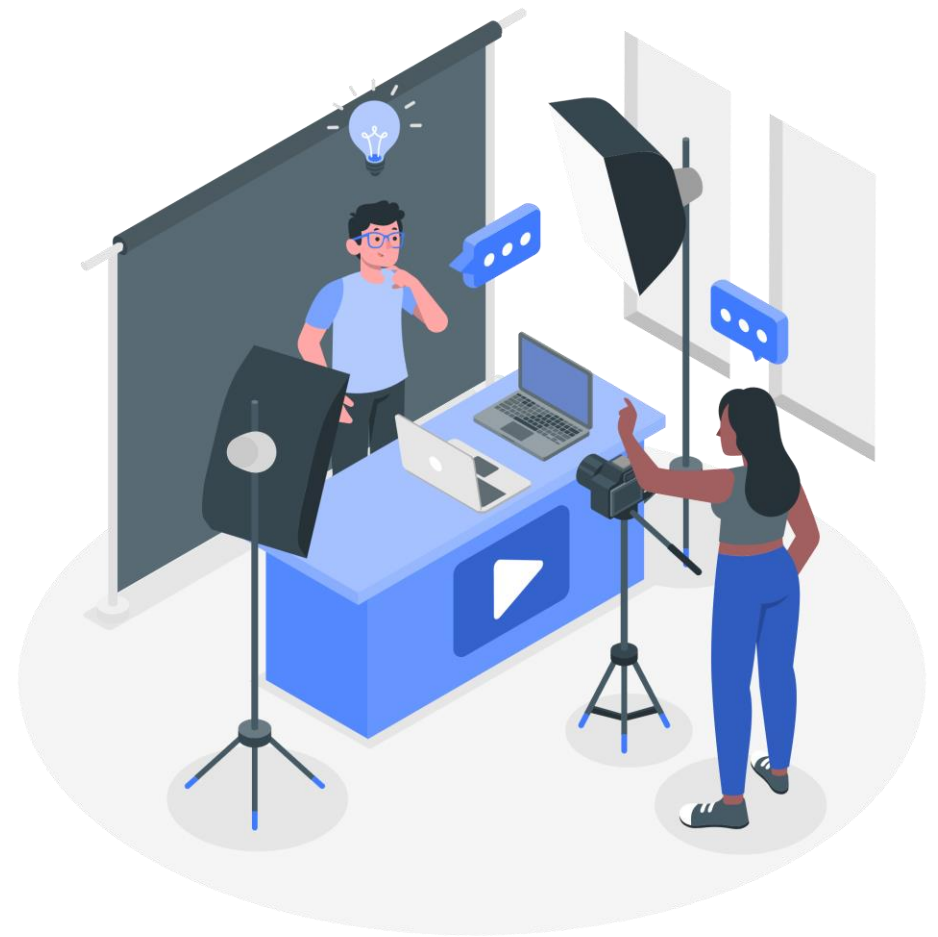
Проте технологія може використовуватися і з цілком «мирною» метою — наприклад, у сфері кіно. Вона дає змогу «замінити» актора у небезпечних трюках або «оживити» зірок минулого.



Для чого створюються дипфейки?

*Цікавою можливістю є перспектива **дипфейкового дубляжу**.*

Відеоряд не потрібно дублювати — обличчя акторів можна буде редагувати так, ніби вони говорять різними мовами. Наразі це лише творчі плани, але ж усе, зрештою, починається з творчих планів.



Як розпізнати дипфейк?

Розпізнати медіапідробку може бути дійсно складно, проте можливо.

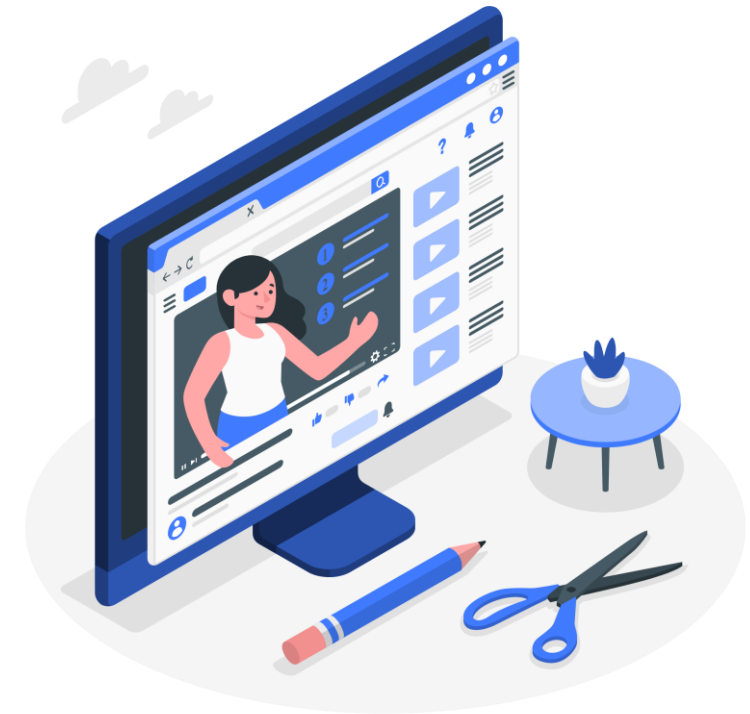
По-перше, важливо критично мислити. Завжди треба перевіряти інформацію, послуговуватися надійними джерелами, не поширювати сумнівні дані.



Як розпізнати дипфейк?

По-друге, деякі ознаки дипфейкових матеріалів можна помітити візуально. Якщо розглядати відео покадрово, можна побачити таке:

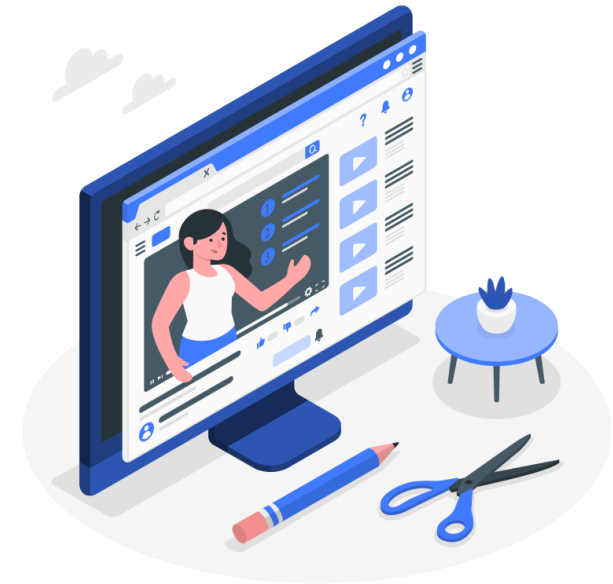
- ✓ обличчя не завжди пропорційно «накладається» на тіло, краї «маски» не всюди рівні;
- ✓ біля очей і брів можуть пролягати дивні тіні;
- ✓ неприродний відтінок кольору шкіри;



Як розпізнати дипфейк?

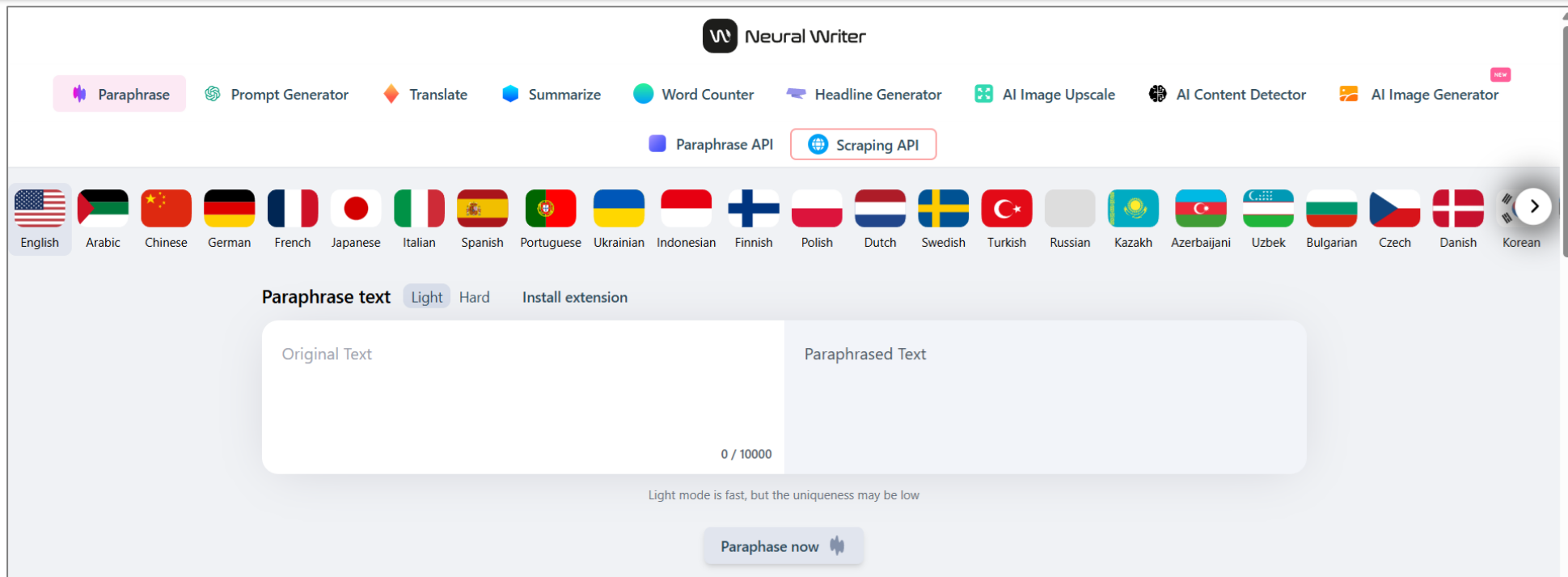
Продовження...

- ✓ окуляри чи інші великі аксесуари можуть мати відблиски, котрі не гармоніюють із загальною картинкою;
- ✓ на дипфейкових відео волосся часто виглядає неприродно: може бути «розмитим» чи мати неправильну лінію росту;
- ✓ невідповідні міміка та рухи (зокрема рух очей!).



Як розпізнати дипфейк?

По-третє, існують програми, які допомагають боротися з дипфейками.
Наприклад, детектор *NeuralWriter* аналізує слова та фрази в тексті, який ви вставляєте в нього.



Як розпізнати дипфейк?

Він звертає увагу на різні характеристики тексту, такі як стиль, складність і вживання слів. Після розпізнавання детектор порівнює ці характеристики з тим, як зазвичай пишуть люди і як пише ШІ.

Якщо текст схожий на те, що може генерувати ШІ, детектор дає нам натяк, що це може бути автоматично згенерований текст.



Neural Writer

neuralwriter.com



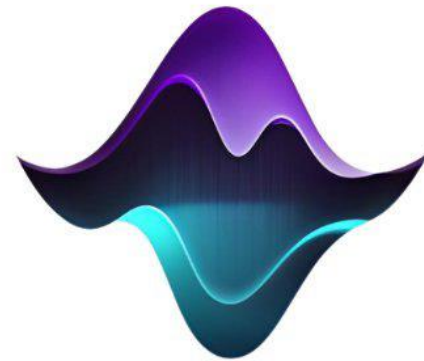
Як розпізнати дипфейк?

Для перевірки справжності аудіозаписів можна використовувати спеціальні програми для перевірки аудіо, наприклад:

PlayHT



AI Voice Detector



AI or not



Як розпізнати дипфейк?

Щоб виявити втручання штучного інтелекту в зображення, можна ретельно придивитись до зображення та виявити ознаки підробки.

Наприклад, на зображенні, яке поширилось мережею наприкінці грудня 2023 року, маємо:

1) на передньому плані – дівчина з розпущеним волоссям і шевронами, прапор України на шевроні дівчини перевернутий;



Як розпізнати дипфейк?

Продовження...

2) у військової, що бере участь в активних бойових діях, напрочуд ідеальні укладка та макіяж, наче вона щойно вийшла із салону;

3) військовослужбовці вдягнені в російський однострій.



Як розпізнати дипфейк?

А можна скористатись спеціальними сервісами виявлення підробленого зображення, наприклад **Fotoforensics**.

Те саме — з підробкою відео.
Наприклад, можна скористатись сервісом **Deerware**.



Як розпізнати дипфейк?

Також можна пошукати сервіс у бібліотеці відповідних інструментів **Detectortools**.

AI CONTENT
DETECTOR
TOOLS



Розгадайте ребус

rebus1.com / ua



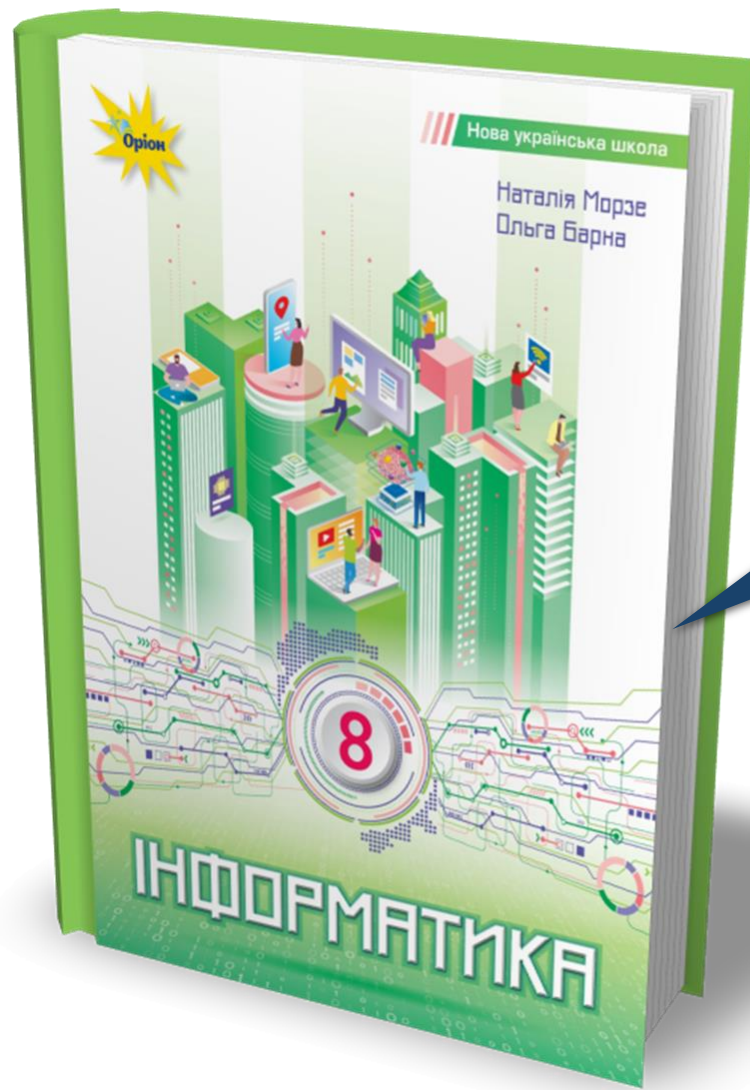
”””



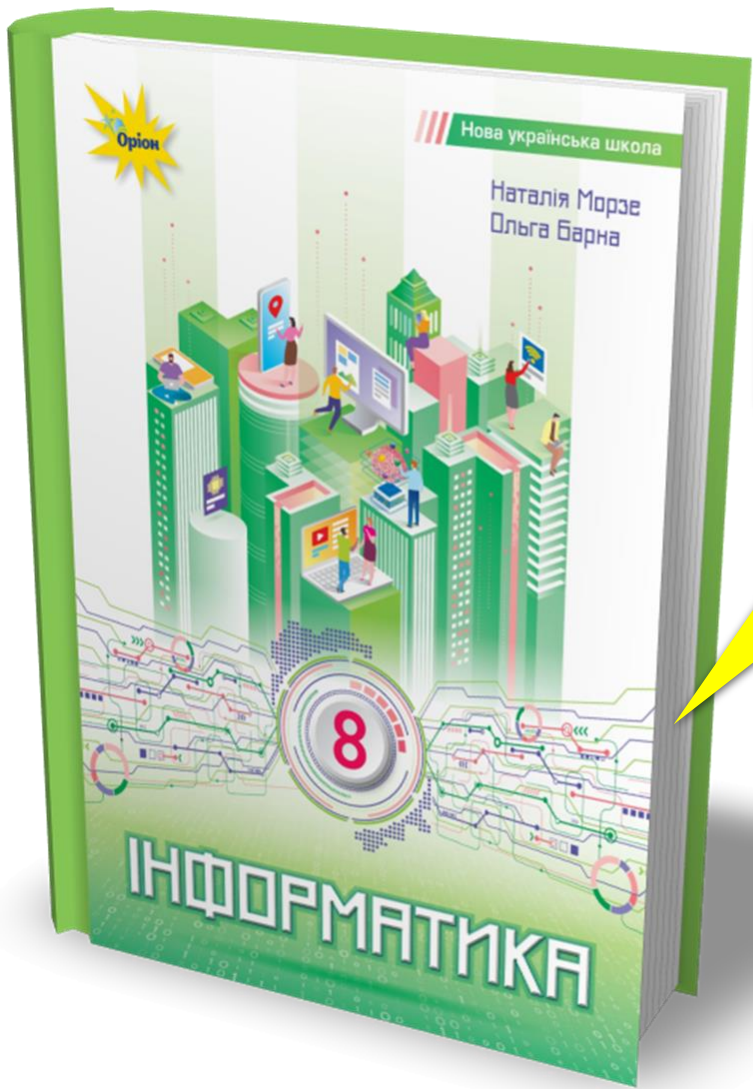
ФОТОМОНТАЖ



Домашнє завдання



Проаналізувати
с. 166-173



**Сторінка
167-173**



Інформатика 8

Урок 51

за підручником

Морзе Н.В. та ін.

Успіхів у навчанні!

Нова українська школа

