

МАШИННО ОБУЧЕНИЕ И DEERFAKE



Основна идея

Машинното обучение позволява на компютърните системи да откриват зависимости в данни и да подобряват резултатите си чрез примери. Deepfake използва такива технологии за създаване на синтетични изображения, звук и видео.

1. Какво е машинно обучение?

Машинното обучение (Machine Learning) е направление в изкуствения интелект, при което компютърните системи се обучават чрез данни, вместо всяко действие да бъде описано предварително с отделно правило.



Елемент	Роля
Данни	Примери, чрез които моделът се обучава.
Характеристики	Признаци в данните, които са полезни за задачата.

Модел	Система, която открива зависимости.
Обучение	Процес на настройване на модела.
Проверка	Оценяване с нови, невиджани примери.
Пример За разпознаване на котки моделът се обучава с много означени снимки на котки и други животни. След това се проверява с нови снимки.	

1.1. Обучение с примери



При контролираното обучение примерите имат правилен отговор, наречен етикет. Моделът сравнява своите предположения с правилните отговори и постепенно намалява грешките.

- обучаващи данни - използват се за настройване на модела;
- тестови данни - използват се за независима проверка;
- етикет - правилният клас или отговор;
- точност - колко често моделът дава правилен резултат.

Важно
Ако моделът запомни обучаващите примери, но не се справя с нови данни, той не е научил задачата достатъчно добре.

Задача 1. Quick, Draw!

Отворете [Quick, Draw!](#) и наблюдавайте как системата разпознава рисунките. Запишете два успешни и един неуспешен опит.



1.2. Разпознаване на образи, текст и звук

Машинното обучение може да работи с различни видове данни.

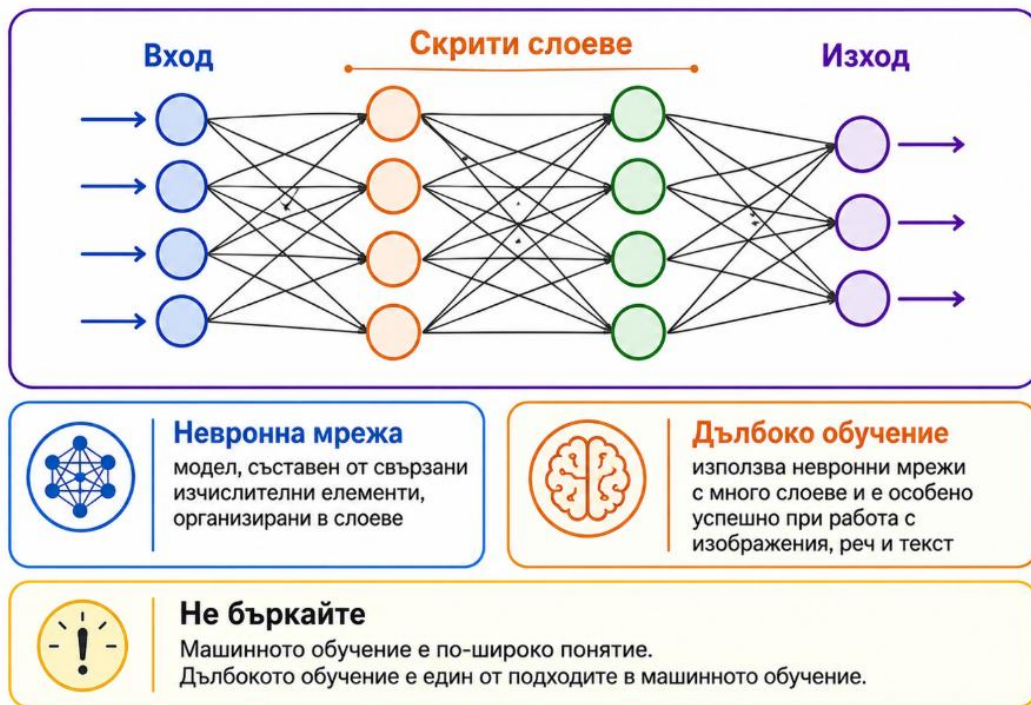
Вид данни	Примерна задача
Изображения	разпознаване на предмети, лица и медицински образи
Текст	класификация, превод, търсене на сходни думи
Звук	разпознаване на реч и музика
Числови данни	прогнози и откриване на необичайни стойности

Задача 2. Сравни човек и машина

Опишете как човек разпознава нарисуван предмет и как според вас го разпознава обучен модел.

1.3. Невронни мрежи и дълбоко обучение

Невронната мрежа е модел, съставен от свързани изчислителни елементи, организирани в слоеве. Тя е вдъхновена от начина, по който биологичните неврони предават сигнали, но не е точно копие на човешкия мозък.



Дълбокото машинно обучение (Deep Learning) използва невронни мрежи с много слоеве. То е особено успешно при работа с изображения, реч и текст.

Не бъркайте

Машинното обучение е по-широко понятие. Дълбокото обучение е един от подходите в машинното обучение.

1.4. Качество и пристрастност на данните



Качеството на резултата зависи от качеството на данните.

Проблем с данните		Възможен резултат	
1	 Твърде малко примери	 ниска точност	
2	 Едностранчиви примери	 пристрастни решения	
3	 Грешни етикети	 моделът научава погрешни зависимости	
4	 Стари данни	 резултатите не отразяват настоящата ситуация	



Казус

Система за разпознаване на лица е обучена почти само със снимки на една възрастова група.

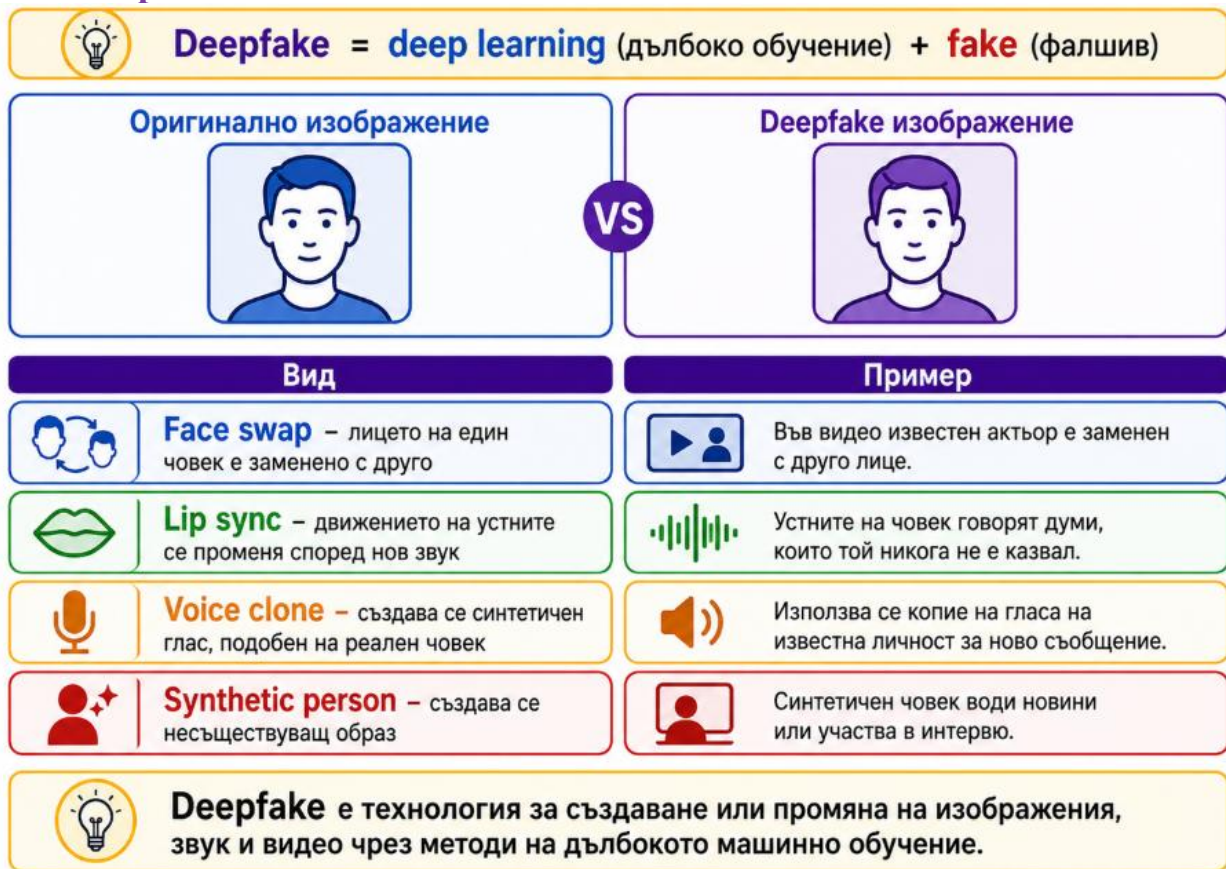
Какъв проблем може да възникне?

Качеството на резултата зависи от качеството на данните. Ако данните са непълни, едностранчиви или грешно означени, моделът може да дава несправедливи или неточни резултати.

Задача 3. Казус

Система за разпознаване на лица е обучена почти само със снимки на една възрастова група. Какъв проблем може да възникне?

2. Какво е Deepfake?



Deepfake е технология за създаване или промяна на изображения, звук и видео чрез методи на дълбокото машинно обучение. Резултатът може да изглежда или звучи така, сякаш реален човек е направил или казал нещо, което не се е случило.



Оригинално изображение Deepfake изображение

2.1. Как работи Deepfake?

КАК СЕ СЪЗДАВА DEERFAKE?



Компонент	Роля
 Генератор	създава нови изображения, звук или видео
 Детектор	различава реални от синтетични примери

 **GAN (Generative adversarial network)** – генеративно състезателна мрежа.

Един от известните подходи използва две взаимодействащи си невронни мрежи: генератор и детектор. Генераторът създава съдържание, а детекторът оценява доколко то прилича на реално.

Компонент	Роля
Генератор	създава нови изображения, звук или видео
Детектор	различава реални от синтетични примери
Обучение	двата компонента се подобряват чрез многократна проверка

Generative adversarial network (GAN) – генеративно състезателна мрежа. Технология, която се използва при създаването на Deepfake.

3. Приложения на Deepfake

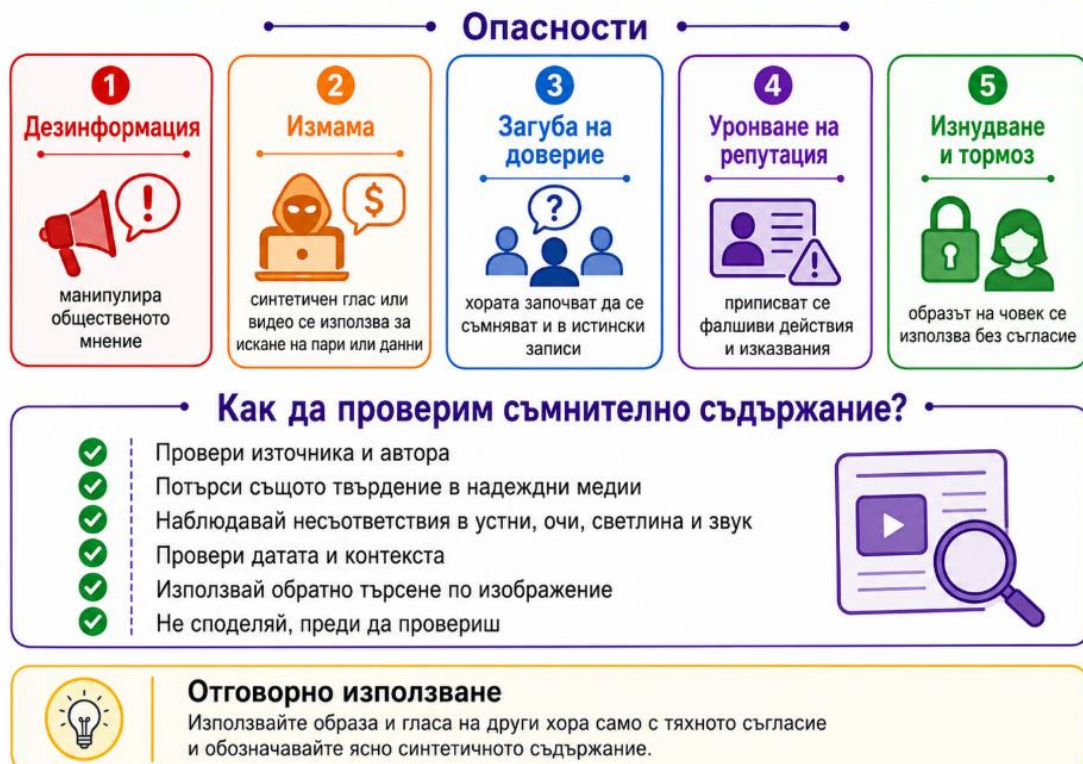


Задача 4. Полезно или рисково?

За всяко приложение посочете при какви условия е приемливо и кога би станало подвеждащо.

4. Опасности от Deepfake

ОПАСНОСТИ ОТ ДЕЕРФАКЕ И ПРОВЕРКА



Deerfake може да причини вреда, когато се използва без съгласие или с цел заблуждаване.

4.1. Как да разпознаваме съмнително съдържание?

Няма един белег, който винаги доказва Deerfake. Най-надеждният подход е да се проверят едновременно източникът, контекстът и визуалните или звуковите несъответствия.

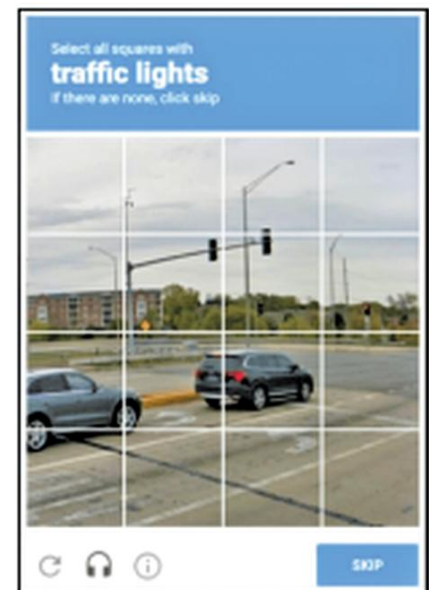
- неестествено движение на устните или лицето;
- разминаване между звук и картина;
- странни сенки, отражения или ръбове;
- необичайно мигане и изражения;
- липса на надежден първоизточник;
- сензационно твърдение без потвърждение.

Важно

Съвременните Deerfake материали може да нямат очевидни визуални дефекти. Затова проверката на източника е по-важна от търсенето само на странности в изображението.

4.2. САРТСНА и обучение на модели

САРТСНА е инструмент, създаден да различава хора от автоматизирани програми. Някои задачи за разпознаване на предмети също могат да подпомагат събирането на означени данни за обучение на системи.



Поверителност

При работа с онлайн услуги проверявайте какви данни се събират и за каква цел се използват.

5. Отговорно използване

- Използвайте образа и гласа на други хора само с тяхното съгласие.
- Обозначавайте ясно синтетичното или променено съдържание.
- Не създавайте материали, които могат да навредят или заблудят.
- Не споделяйте непроверени сензационни записи.
- Пазете личните снимки, видео и гласови записи.
- При съмнение потърсете надежден възрастен или институция.

Етичен принцип

Това, че технологията позволява нещо, не означава, че е правилно да го направим.

6. Практическа задача: проверка на съдържание

Задача 6. Анализирай публикация

Получавате видео, в което известен човек прави необичайно изявление.

1. Определете първоизточника.
2. Проверете дали надеждни медии публикуват същото.
3. Потърсете пълния оригинален запис.
4. Наблюдавайте картината и звука за несъответствия.
5. Не споделяйте, докато не сте сигурни.
6. Запишете кратък извод: надеждно, съмнително или невъзможно за потвърждение.

7. Обобщение

Понятие	Най-важното
Машинно обучение	моделите се обучават чрез данни и примери
Обучаващи данни	използват се за настройване на модела
Тестови данни	проверяват работата с нови примери
Невронна мрежа	модел от свързани слоеве
Дълбоко обучение	невронни мрежи с много слоеве
Deerfake	синтетично или променено изображение, звук или видео
GAN	генератор и детектор се обучават състезателно
Проверка	източник, контекст, потвърждение и технически белези
Най-важното Машинното обучение може да създава полезни системи, но същите технологии могат да бъдат използвани за манипулация. Затова са необходими критично мислене, проверка и отговорност.	

8. Провери знанията си

1. Какво е машинно обучение?
2. Каква е разликата между обучаващи и тестови данни?
3. Какво е невронна мрежа?
4. Какво наричаме дълбоко обучение?
5. Какво е Deerfake?
6. Каква е ролята на генератора и детектора?
7. Посочете две полезни приложения на Deerfake.
8. Посочете три опасности от Deerfake.
9. Как трябва да проверим съмнително видео?
10. Какво означава отговорно използване на синтетично съдържание?