



Справедлив ли е алгоритъмът?

Предразсъдъци в изкуствения
интелект и лична
отговорност онлайн

Цели на урока

Учениците ще могат да:

- разпознават как в алгоритми и системи с ИИ могат да се появят предразсъдъци
- обясняват как данните влияят върху резултата на системата
- анализират реален казус с проблем, засегнати страни и последствия
- аргументират позиция в дискусия
- формулират принципи за отговорно дигитално поведение



Кой получава покана за интервю?

В голяма компания използват ИИ, който преглежда автобиографии и подрежда кандидатите за работа. След време се установява, че системата по-често отхвърля жени кандидати за технически позиции. Причината е, че е обучена върху стари данни, в които за такива позиции предимно са били назначавани мъже.

ИИ преглежда
автобиографии

Отхвърля жени
кандидати по-често

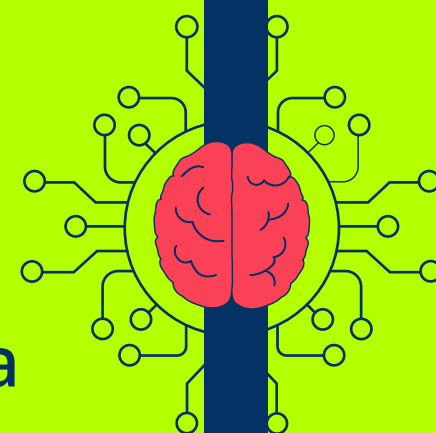
Техническа грешка или
несправедливост?

Най-голямата платформа с практични и качествени обучения и ресурси
за учители и директори

Как се появява предразсъдък в ИИ?

Алгоритъм - последователност от правила, по които система обработва информация и стига до резултат.

Данни - примерите, с които системата се учи: текстове, снимки, оценки, минали случаи.



Предразсъдък в ИИ - когато системата дава по-неблагоприятен резултат за определена група хора заради начина, по който е създадена или обучена.

Схема: Данни > Алгоритъм > Решение > Последствия

ИИ не е неутрален. Създава се от хора, обучава се с човешки данни и се използва в човешки ситуации. Въпросът е не само "Работи ли?", а "Справедливо ли работи?"

Разследване на алгоритъм

Групова работа: Всяка група получава казус.

Обсъдете и попълнете работния лист:

- Какво решение взема алгоритъмът?
- Какъв е проблемът в този случай?
- Кой може да бъде засегнат?
- Какви данни може да са довели до този резултат?
- Кой носи отговорност?
- Какво трябва да се промени?



Казуси за групова работа

1

ИИ подбира кандидати за работа - дава по-ниски оценки на жени за технически позиции.

2

Камера за лицево разпознаване - повече грешки при определени групи хора.

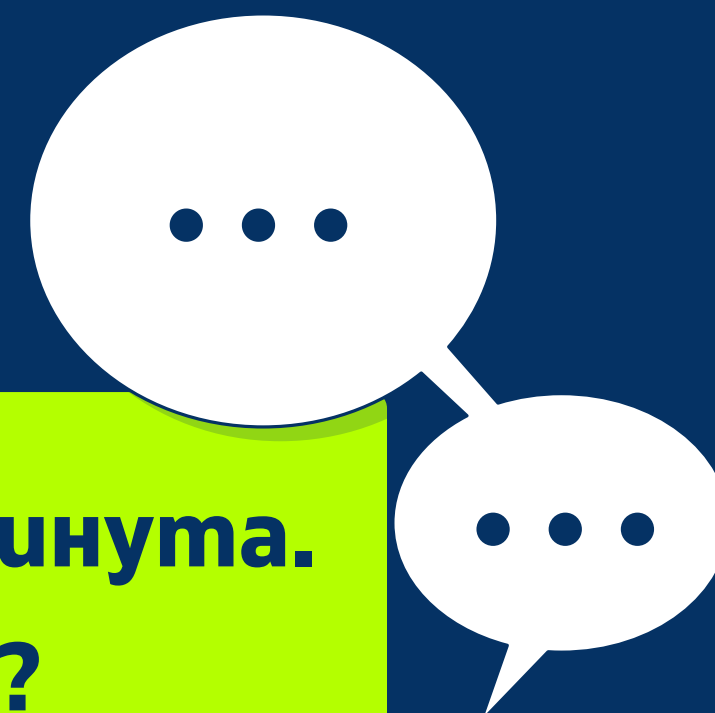
3

Алгоритъм оценява риск - повлиян от стари неравенства в данните.

4

Социална мрежа показва повече гневни коментари и крайни мнения, за да задържи вниманието.

Насочващи въпроси към казусите



**Всяка група представя най-важния си извод за 1 минута.
Къде точно се появява несправедливостта?**

Кой човек стои зад този резултат? Какво би почувствал?

Най-голямата платформа с практични и качествени обучения и ресурси за учители и директори

Кръгла маса: Кой е отговорен?

преподаваме.bg

Роли в дискусиата:

- Програмист
- Компания
- Потребител
- Засегнат човек
- Институция / регулатор

Трябва ли човек да може да оспори решение, взето от алгоритъм?

Задача:

Всяка роля защитава своята гледна точка. В края казва: "Нашата отговорност е..."

Отговорността не е едностранна - тя е споделена.

Въпроси за кръглата маса

Учителят насочва дискусията с въпроси:

- Кой трябва да провери системата преди употреба?
- Трябва ли човек да може да оспори решение на алгоритъм?
- Кога е опасно да казваме: "Компютърът така реши"?
- Какво може да направи един ученик като потребител?

В края всяка ученик със съответната роля споделя:
„Нашата отговорност е...”



Най-голямата платформа с практични и качествени обучения и ресурси
за учители и директори

Роли в кръглата маса



Програмист

проверява
кода и
данните



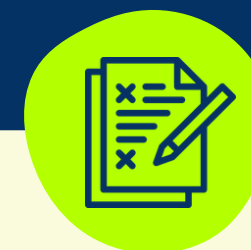
Компания

отговаря за
употребата



Потребител

не вярва сляпо
на системата



Засегнат човек

иска обяснение
за решението



Институция

създава
правила за ИИ

Най-голямата платформа с практични и качествени обучения и ресурси
за учители и директори



Изходен билет

Довършете трите изречения самостоятелно.

Довършете:

- Един алгоритъм може да бъде несправедлив, когато...
- В такъв случай засегнат може да бъде...
- Едно отговорно действие на дигиталния гражданин е...

Нашите принципи

Създайте три принципа за "Брошура на дигиталния гражданин". Всеки принцип започва с действие:

- Проверявай автоматичните решения, когато засягат хора.
- Питай кой е създал системата и с какви данни е обучена.
- Защищавай правото на човек да получи обяснение.

Всяка група прочита един принцип.

Учителят записва най-силните на дъската.

Върнете се в групите за 5 минути.



Самооценка

преподаваме.bg

Отбележете как се справихте днес:

Разпознавам как в алгоритъм
може да се появи предразсъдък.

Мога да посоча кой е засегнат и
кой носи отговорност.

Формулирах принцип за отговорно
дигитално поведение.





Какво научихме днес?

- Алгоритъмът взема решения според правила и данни от хора.
- Предразсъдъците в ИИ идват от непълни или несправедливи данни.
- Засегнати могат да бъдат реални хора: кандидати, ученици, пациенти.
- Отговорността е споделена между създатели, компании и потребители.
- Дигиталният гражданин проверява и не приема всяко решение за справедливо.

Питам. Проверявам. Не приемам сляпо.



Благодаря!

