

Torzítás – Hiba a mesterséges Intelligenciában

Képzeld el a következő helyzetet ...

... egy mesterséges intelligencia dönti el helyetted az általános iskola után, hogy melyik középiskolába jársz (és ebből következően milyen pályát válassz)!

Ez a mesterséges intelligencia megpróbálja megtalálni a megfelelő iskolát a környezetedben, az előzetesen megállapított adottságaid, illetve olyan jellemzőid alapján, mint a nemed, származásod, érdeklődési köröd és intelligenciád.

Ehhez, az ilyen tényezők nagy részét be kell táplálni egy programba adatok formájában, a lehető legtöbb diákra vonatkozóan. Így ezek a tényezők a sok adat révén összekapcsolhatók lesznek egymással, és a feltárt összefüggések alapján a lehető legpontosabb eredmény kalkulálható ki számodra.

Ám hirtelen ez a mesterséges intelligencia rendszer úgy dönt, hogy neked szakközépiskolába kell járnod, pedig soha nem is akartál! Vajon mi siklott félre?



1. feladat: Képzeld el, hogy pontosan ez történt egy lány/nőnemű barátoddal és az MI tévesen a rossz iskolába akarja őt küldeni.

- Gyűjtsetek össze csoportokban minél több olyan tulajdonságot, amivel rendelkezhet (hol él, mit szeret csinálni, milyen jellemvonásai vannak...)!
- Most írtok le, milyen előítéletek kapcsolódhatnak ezekhez a tulajdonságokhoz (pl. lányok/nők a műszaki tudományokban, szegény/gazdag környéken él, rendetlen, szemüveges, szőke, stb.)!
- Mit gondolsz, melyik iskolába irányítaná a rosszul programozott algoritmus a barátnődet? És Te melyik iskolába küldenéd őt szívesen?
- Minek kellene történnie ahhoz, hogy a mesterséges intelligencia program ne kövesse el újra ugyanazt a hibát?

Lehetséges, hogy nagy mennyiségű adat összegyűjtése során hibák csúsznak be. Lehet, hogy vannak már olyan diákok, akik nagyon hasonlítanak rád: mondjuk hasonlóan okosak, szeretnek logikai feladatokat megoldani és nagyon társaságkedvelők. Miután ezek a gyerekek mind műszaki iskolába akartak járni, az MI úgy dönt, hogy Te is a műszaki iskolát választanád.

Az ilyen hibákat torzítási hibáknak nevezzük. Ezek akkor fordulnak elő, amikor előítéletek kúsznak be a program által használt adathalmazokba.

2. feladat: Gondoljátok át kiscsoportokban, hogy miként lehetne elkerülni, hogy akár egyetlen gyermek is rossz iskolába kerüljön? Hogyan kerülhetők el az ilyen, ún. torzítási hibák?

Gondold végig: Honnan jönnek a rendelkezésre álló adatok? Ki dönti el, mely adatok kerülnek felhasználásra?

Ilyen algoritmusok már léteznek a való világban. Mesterséges intelligenciát használnak például annak eldöntésére, hogy valaki kaphat-e pénzügyi kölcsönt: ekkor a program az életkor, jövedelem és a lakóhely alapján értékeli ezeket az embereket. Ha azonban egy környéken már sok embernek magas az adóssága, a program dönthet úgy, hogy a kölcsönt igénylő személy kevesebb pénzt kap, annak ellenére, hogy nincs adóssága.

Látható tehát, hogy hihetetlenül nehéz olyan döntési algoritmust kifejleszteni, amely minden lehetőséget figyelembe vesz, és a lehető legigazságosabban dönt a betáplált adatok alapján. Vagyis ezek a programok csak annyira "gondolkodnak" igazságosan, amennyire azok az emberek, akik megtanítják ezeket a döntésekre. A gyakran emlegetett angol nyelvű mondás szerint: „**Garbage in, garbage out**” (szó szerint: ha szemét megy be, szemét is jön ki), azaz rossz adatból, csak hibás eredményekre juthatunk.



Mik azok a torzítási hibák?

Milyen példákat tudsz rá mondani?

Milyen típusú torzítási hibák lehetnek?
