

Princípy dátovej vedy - 18: Etika

Samuel Rosa

Fakulta matematiky, fyziky a informatiky
Univerzity Komenského v Bratislave

Etika

Ethics is the [philosophical](#) study of [moral](#) phenomena. Also called **moral philosophy**, it investigates [normative](#) questions about what people ought to do or which behavior is morally right. Its main branches include [normative ethics](#), [applied ethics](#), and [metaethics](#).

zdroj: wiki

- ako by sme sa mali správať, čo je morálne správne

Kantov kategorický imperatív

- Konaj tak, aby to mohlo byť všeobecným pravidlom.
- Ľudia sú ciele, nie prostriedky.

Kantov kategorický imperatív

- Konaj tak, aby to mohlo byť všeobecným pravidlom.
- Ľudia sú ciele, nie prostriedky.

=====

- Act only according to that maxim by which you can also will that it would become a universal law.
- Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.
- Every rational being must so act as if he were through his maxim always a legislating member in a universal kingdom of ends.

V dátovej vede

If you torture the data long enough, it will confess to anything.

Ronald Coase



Obr.: [wikimedia](#)

V dátovej vede

Neklamať

- Nevymýšľať si, neskrývať dáta (alebo ich časti)
- Neskrývať výsledky, ktoré sa mi nepáčia
- Nezveličovať
- Kriticky sa pozerať na svoje výsledky (priznať aj obmedzenia a nedostatky)

V dátovej vede

Neklamať

- Nevymýšľať si, neskrývať dáta (alebo ich časti)
- Neskrývať výsledky, ktoré sa mi nepáčia
- Nezveličovať
- Kriticky sa pozerať na svoje výsledky (priznať aj obmedzenia a nedostatky)
- Vizualizácie na vyjasnenie, nie na zakrytie
- Skúmať a korektne prezentovať kvalitu svojich dát

Celkové výdaje státního rozpočtu na důchody v mld.Kč

(bez výdajů na správu)



Zdroj: Ministerstvo financí

Obr.: aktualne.cz

Ďalšie “concerns”

Privacy concerns:

- Zo zverejnených dát/výsledkov by nemali byť konkrétni ľudia identifikovaní
- GDPR
- Sú vôbec dáta získané “morálne”? (Google, Facebook, Amazon... zbierajú “všetko, čo sa dá”)

Bias concerns:

- Aby modely nezdedili naše predsudky (biased data → biased model)
- Zvýhodňovanie/znevýhodňovanie rôznych skupín (muži/ženy, farba pleti ...)

V modelovaní

- Použiť relevantné, nie len dostupné dáta
- Mala by existovať metrika úspechu (kedy je vlastne model “dobrý” a kedy nie)
- Korektné vyhodnotenie kvality (napr. nepozerať na “test set” vopred!)
- Nezveličovať
- Ideálne: malo by byť jasné, ako model došiel k rozhodnutiam
- Korelácia nie je kauzalita; keď sa niečo javí v modeli, nemusí to byť v realite

V inferencii

*There are three kinds of lies:
Lies, damned lies, and statistics.*

Benjamin Disraeli



Obr.: [wikimedia](#)

V inferencii

- NIE: p -hacking, data dredging: veľa hypotéz, prezentujem iba tie, čo vyšli
- NIE: HARKing (Hypothesis after the results are known):
“hypotéza” až po pohľade na dáta
- Korektná interpretácia (nezveličovať, korelácia $\neq$ kauzalita)

V AI

- (Large language models, “algoritmy” na sociálnych sieťach, AlphaZero, autonómne autá...)
- Agency issues
 - rozhodujú o ľuďoch namiesto ľudí
 - dá sa výstup AI overiť, dá sa voči nemu odvolať?
 - môže algoritmus strieľať zo zbraní vo vojne?
 - autonómne auto a nehoda
- “Berú ľuďom prácu”
- Deepfakes, dezinformácie