ALTERNATOR
Misliti znanost.

Avtomatizacija (lahko) osvobaja misel

16. 10. 2025

Številka: 20/2025

Avtorica:

- Ana Rebeka Kamšek



Harvardske računarke pri delu, okoli leta 1900. Vir: Observatorij Univerze Harvard, Wikimedia Commons (https://en.wikipedia.org/wiki/File:Astronomer_Edward_Charles_Pickering%27s_Harvard_computers.jpg) , javno dobro (https://creativecommons.org/licenses/by/4.0/)

Kaj se zgodi z mislijo, ko del svojih nalog prepustimo strojem? V znanosti je avtomatizacija obdelave podatkov zrasla iz potrebe: podatkov je preveč, časa za ročno obdelavo premalo. Z njo se znebimo rutinskega dela, v zameno pa od nas zahteva jasne korake, nadzor nad kakovostjo in premišljeno razlago rezultatov.

Od ročnega računanja do prvih oblik avtomatizacije

Na prelomu iz devetnajstega v dvajseto stoletje je na harvardskem observatoriju delovala skupina, ki se je je prijelo poimenovanje *Harvard computers* (https://platestacks.cfa.harvard.edu/women-at-hco) . Besedo *computer* danes razumemo kot poimenovanje za računalnik, takrat pa je pomenila človeka, ki računa. V skupini so namreč delovale ženske, ki so opravljale vlogo človeških računalnikov.

Na observatoriju so nebo snemali na steklene fotografske plošče. Že to je bil pomemben korak v astronomiji, ki se je dolgo

časa zanašala na opazovanje v živo in sprotno zapisovanje na roko. Takratni direktor harvardskega observatorija Edward Charles Pickering je skupino ustanovil prav zaradi naraščajoče zbirke fotografskih plošč. Harvardske računarke (glej naslovno sliko članka) so na ploščah ročno merile položaje zvezd in njihov izsev, vnašale zvezde v kataloge, preučevale zvezdne spektre ter zapisovale opombe o zanimivostih na slikah.

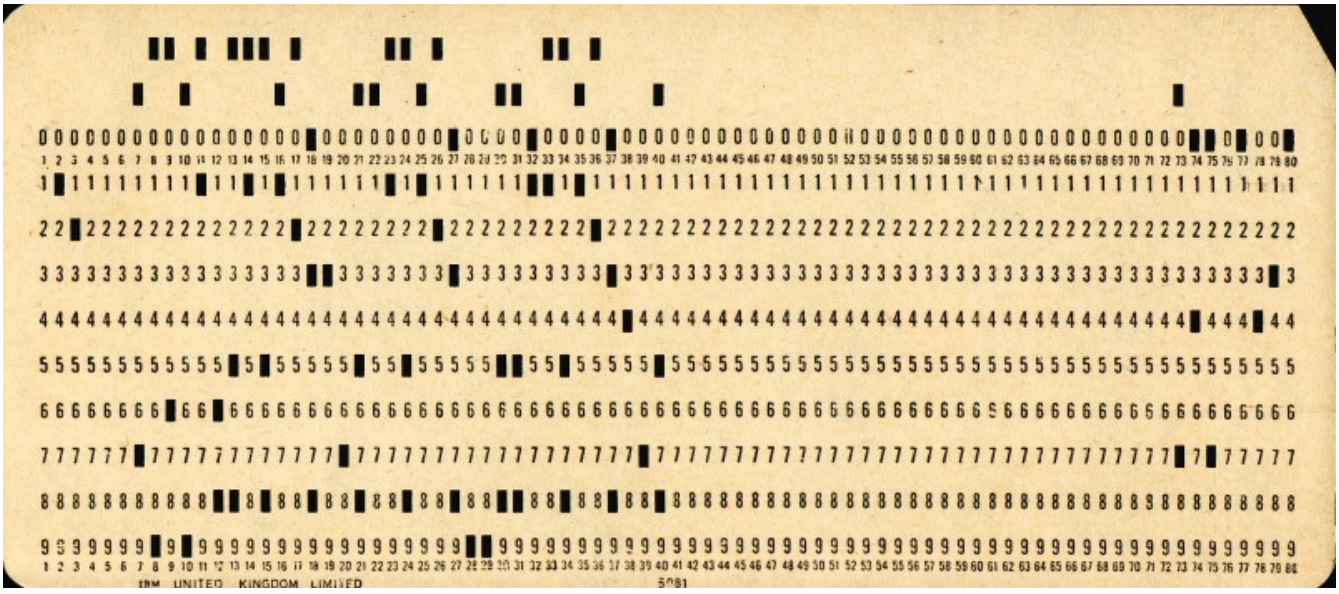
Delo je bilo ponavljajoče. Zahtevalo je osredotočenost, napete oči, ni pa ponujalo prav dosti ustvarjalnosti. Vseeno je bistveno prispevalo k napredku astronomije, saj je vodilo do novih velikih katalogov zvezd in drugih pomembnih korakov. Annie Jump Cannon je skupaj z vodjo observatorija ustvarila harvardsko shemo za razvrščanje zvezd v razrede na podlagi njihove temperature in spektralnega tipa, Henrietti Swan Leavitt pa se lahko zahvalimo za odkritje razmerja med izsevom zvezde in časom, v katerem se izsev spremeni.

Takrat so računarke – človeški računalniki – izračune izvajale ročno, a ravno njihov pristop je postal eden od temeljev avtomatizacije. Postopke so na observatoriju standardizirali in razčlenili na jasne, ponovljive korake. Tudi danes, ko ima vlogo fotografske plošče kamera, vlogo zvezka pa datoteka, bistvo obdelave podatkov ostajajo ponovljivi, dobro definirani koraki.

Ta poklic je vztrajal še nekaj časa. Med organizacijami, ki so zaposlovale človeške računalnike, je bila tudi ameriška vesoljska agencija NASA (https://www.nasa.gov/centers-and-facilities/jpl/when-computers-were-human/) . Petdeseta in šestdeseta leta prejšnjega stoletja so bila razburljiv čas za raziskovalne odprave v vesolje in zgodnjih poskusov z raketami ne bi bilo brez zanesljivih ekip računark. V tistih časih so že obstajali zgodnji mehanski računalniki, a niso bili pretirano zmogljivi, hkrati pa so zasedli veliko prostora. Računarke so tako na roke določale in sproti preverjale tirnice letenja vesoljskih plovil. Poleg izjemnih dosežkov na vesoljskem področju je NASA s tem utirala pot tudi na družbenem področju, saj je manj zastopanim skupinam odpirala vrata v tehnične poklice.

Naraščajoča količina dela in podatkov na različnih področjih je marsikoga navdihnila za iskanje rešitev, ki bi ljudi razbremenile ročnega računanja. Ena od priložnosti se je ponudila že ob koncu devetnajstega stoletja. Takrat je ameriški popis prebivalstva trčil ob problem: ročna obdelava zbranih podatkov se je tako vlekla, da so številke zastarale, še preden bi jih lahko pametno uporabili. Inženir Herman Hollerith je kot rešitev predlagal luknjane kartice (https://doi.org/10.1111/j.1542-734X.1992.1504_43.x) .

Luknjane kartice so pravzaprav še starejše. Že v začetku devetnajstega stoletja so jih uporabljali v tekstilni industriji, kjer so Jacquardove statve (https://doi.org/10.1080/08832323.1965.10118580) z njihovo pomočjo izdelovale tkanine z zapletenimi vzorci in tako razbremenile delavce ročnega upravljanja z nitjo. Hollerith je luknjane kartice uporabil v kontekstu obdelave podatkov – vsak odgovor na vprašanje ob popisu prebivalstva je pomenil točno določen položaj na kartici. Uradnik je preluknjal ustrezno polje in kartico oddal v bralnik. Tako je obdelal tudi več deset kartic na minuto. Sistem je bil hiter, ponovljiv in dosleden, obsežni popis prebivalstva pa je dobil tekmovanje s časom.



Slika 1: Luknjana kartica, na kateri je zapisana vrstica programa v jeziku Fortran. Foto: Pete Birkinshaw, vir: Wikimedia Commons (https://en.wikipedia.org/wiki/File:Used_Punchcard_(5151286161).jpg)

Kartice so hitro presegle kontekst popisa prebivalstva in se znašle v uporabi tudi v zavarovalnicah, vojski, na železnicah in drugje. IBM-ova 80-stolpčna kartica je postala standard – kos kartona, ki je predstavljal podatek (Slika 1). Na mnogih je pisalo: »Ne prepogibaj, ne natikaj in ne poškoduj« (angl. *Do not fold, spindle or mutilate*). Ni bila šala: vsaka guba je lahko preprečila uspešno uporabo kartice. Po več desetletjih široke rabe so jih naposled pri shranjevanju in obdelavi podatkov začeli izpodrivati magnetni trakovi in prvi elektronski računalniki. Do devetdesetih let so se luknjane kartice dokončno poslovile, želja po učinkovitih rešitvah za delo s podatki pa je živela naprej.

Avtomatizacija v dobi masovnih podatkov

Ko so nekoč harvardske računarke in kasneje Nasini človeški računalniki razvijali metode za obvladovanje velikih količin podatkov, so postavljali temelje avtomatizacije in sistematične obdelave podatkov. Danes se soočamo s podobnim izzivom, le da z današnjimi instrumenti lahko zajamemo še več podatkov, kot jih lahko pregledamo in obdelamo ročno. Ogromne količine podatkov imajo denimo v velikih projektih na področju astronomije (https://doi.org/10.3986/alternator.2023.12) in fizike osnovnih delcev (https://doi.org/10.3986/alternator.2020.28) .

Od leta 2013 do pomladi 2025 je v vesolju deloval vesoljski observatorij Gaia (https://www.esa.int/Science_Exploration/Space_Science/Gaia) Evropske vesoljske agencije. Njen cilj je bil ustvariti tridimenzionalni zemljevid naše galaksije. Gaia je zbrala ogromne količine podatkov o sestavi galaksije, poziciji nebesnih teles, njihovi hitrosti in svetlosti. Skozi leta je kartirala več kot milijardo zvezd in drugih objektov ter opravila več kot tri bilijone opazovanj. Evropska vesoljska agencija je podatke izdajala postopoma in sproti, vsi podatki pa naj bi bili javno dostopni v roku petih let.

V nasprotju z občasnim objavljanjem večjih arhivov je Observatorij Vere Rubin (https://rubinobservatory.org/about) v Čilu zasnovan za spremljanje sprememb v vesolju skorajda v živo. Po načrtovanem začetku delovanja v letu 2025 naj bi vsakih nekaj noči posnel celotno južno nebo. Iz posnetkov bodo astronomi lahko ustvarili najobsežnejši vesoljski film vseh časov, njegova vsebina pa bo omogočila študij temne energije in temne snovi, manjših teles v Osončju, opazovanje sprememb na nebu zaradi različnih dogodkov ter kartiranje Rimske ceste. Vsako noč bo sistem zaznal okoli deset milijonov dogodkov na nebu, kar je približno stokrat več kot v prejšnjih sistematičnih opazovanjih. Med posnetkom in zaznavo spremembe na nebu bo minila le minuta.

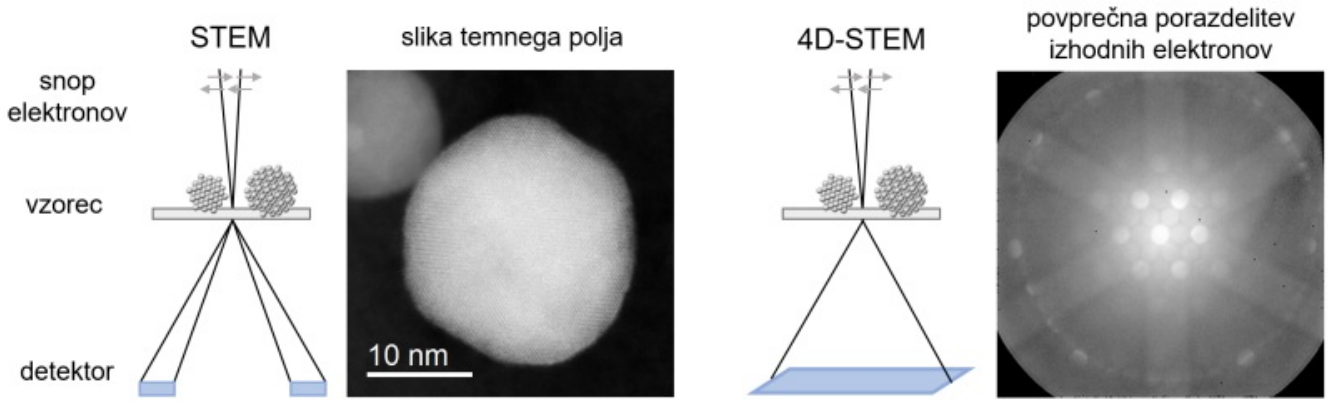
Dober primer nujne avtomatizacije najdemo tudi v fiziki osnovnih delcev. V velikem hadronskem trkalniku (https://home.cern/science/accelerators/large-hadron-collider) (LHC) pod mejo med Švico in Francijo v eni sekundi zabeležijo bilijon trkov delcev, ki potujejo skoraj s svetlobno hitrostjo. Burno dogajanje v tem 27-kilometrskem obroču korak za korakom odgovarja na najbolj temeljna vprašanja o našem vesolju. Množico dogodkov in dragocenosti v njej obdela mreža računalnikov (angl. *Worldwide LHC Computing Grid* (https://wlcg-public.web.cern.ch/) oz. WLCG), ki so razpršeni po svetu na več kot 170 lokacijah v 42 državah. Vsak dan obdelajo več kot dva milijona opravil, saj hranijo, posredujejo in analizirajo podatke, ki jih ustvari veliki hadronski trkalnik. Ne glede na to, za kako enostavne ali kompleksne izračune gre, pri tolikšni količini podatkov avtomatizacija ni več vprašanje.

Kako velike količine podatkov ustvarjamo pri nas

Primerov, kjer ne moremo in ne želimo obdelati vseh podatkov na roke, ne najdemo le pri velikih mednarodnih projektih v astronomiji ali fiziki osnovnih delcev, omenjenih zgoraj – enak izziv se v malem ponovi tudi v naših laboratorijih. Zadošča že, da se odpravimo na Institut "Jožef Stefan" ali Kemijski inštitut v Ljubljani. Obe ustanovi namreč ponujata opremo na področju mikroskopije, s katero lahko v kratkem času ustvarimo komaj obvladljive količine podatkov.

Elektronska mikroskopija uporablja za opazovanje vzorcev elektrone. Tako lahko dosežemo bistveno boljšo ločljivost kot z optičnimi mikroskopi, ki delujejo s pomočjo vidne svetlobe. S slednjimi lahko razlikujemo le med točkama, ki sta oddaljeni vsaj dvesto nanometrov, kar ne zadošča za sodobne nanomateriale, ki jih sestavljajo enote, velike le do sto nanometrov. V vrstičnem presevnem elektronskem mikroskopu (angl. *scanning transmission electron microscopy* oz. STEM) fokusiran snop elektronov premikamo po vzorcu, potem pa z različnimi detektorji preštejemo, koliko elektronov je prišlo ven na drugi strani vzorca pod določenimi koti (Slika 2). Tako dobimo slike materialov, ki dosegajo tudi atomsko ločljivost.

Bolje od atomsko ločljivih slik ne gre. Ali pač? Namesto da zgolj preštejemo elektrone, ki pridejo iz vzorca pod vnaprej določenimi koti, lahko za vse elektrone zapišemo, kako so potovali. Elektroni po izhodu iz vzorca namreč tvorijo zanimive porazdelitve – uklonske vzorce, v katerih je shranjena informacija o kristalni strukturi, napetostih med atomi, usmerjenosti posameznih delov kristala, celo o notranjih električnih in magnetnih poljih. Medtem ko snop elektronov premikamo po vzorcu, na vsakem koraku posnamemo dvodimenzionalno porazdelitev izhodnih elektronov (Slika 2). Ker tudi premikanje po vzorcu poteka v dveh dimenzijah, tako sestavimo štiridimenzionalno množico podatkov, od koder izvira tudi ime metode – 4D-STEM (https://doi.org/10.1017/S1431927619000497) . Če lahko iz običajnih slik določimo velikost objektov in razdalje med njimi, vsaj kvalitativno pa lahko sklepamo tudi o razlikah v kemijski sestavi in debelini vzorca, lahko z metodo 4D-STEM na nanoskali določimo še bistveno več nians v kristalni strukturi.



Slika 2: Slika nanodelca iz platine in bakra ter povprečen uklonski vzorec elektronov, vir: Kamšek idr. 2025 (https://doi.org/10.1021/acsnano.4c12528)

Kaj pa je cena takega celovitega vpogleda v katodo Li-ionske baterije, nov magnetni material ali katalizator za rabo vodika v gorivni celici? Tipično območje vzorca, ki ga vrstico po vrstico skenira elektronski snop za eno množico podatkov, obsega 256 krat 256 korakov, na vsakem koraku pa detektor posname sličico s porazdelitvijo izhodnih elektronov. Za samo eno območje na vzorcu, ki nas zanima, denimo za en kovinski nanodelec v katalizatorju, tako posnamemo 65.536 sličic. Snemanje ne traja bistveno dlje kot minuto. Kako dolgo bi trajalo, da bi si vse sličice ročno samo ogledali, preden bi se sploh lotili njihove obdelave?

Brez avtomatizacije smo obsojeni na ročno označevanje signalov in merjenje v že tako manj intuitivnem svetu uklonjenih elektronov, kjer sličice nastajajo iz ponavljajočih se struktur v vzorcu, ki ga preučujemo. Ročno bi vložili več časa in energije, dobili bi slabši rezultat, hkrati pa tvegali tudi pristranskost pri zaznavanju šibkih signalov. Če nam uspe analizo avtomatizirati do te mere, da je zanesljiva in natančna, so rezultati očitni.

Ob takem tempu ustvarjanja novih podatkov trčimo ob še en izziv. En sam podatkovni niz na trdem disku zasede več gigabajtov prostora. Če želimo posneti več različnih območij vzorca ali pa nas zanima stanje pred in po vpeljanih spremembah, hitro dosežemo stotine gigabajtov. Hramba in obdelava takih količin podatkov presegata zmožnosti računalnikov, namenjenih pisarniški rabi. Zmogljiv mikroskop in detektor nista dovolj – za izkoriščanje vsega, kar nam podatki ponujajo, potrebujemo ustrezno računalniško infrastrukturo in premišljeno strategijo ravnanja s podatki.

Ponovljivost, avtomatizacija in odgovornost v znanosti

Kako danes v znanosti pristopimo h količini podatkov, ki ji ročno nismo več kos? Obdelavo raziskovalnih podatkov opredelimo kot zaporedje nedvoumnih korakov in v ta namen pogosto napišemo programsko kodo. Ta ne služi samo temu, da delo prepustimo računalniku, ampak zagotavlja tudi ponovljivost. Če hranimo kodo in podatke, lahko jutri, čez pol leta ali čez tri ponovno poženemo analizo in se prepričamo o njenem poteku.

Kodo priložimo (https://doi.org/10.1038/514536a) tudi objavi rezultatov v obliki znanstvenega članka. Tako omogočimo drugim, da lahko popolnoma razumejo potek obdelave podatkov, hkrati pa vlivamo dodatno zaupanje v svoje rezultate. Kolikor je mogoče in z ozirom na njihov obseg, priložimo tudi originalne podatke in jih objavimo v skladu z načeli FAIR (https://doi.org/10.1038/sdata.2016.18) . Zainteresirani uporabniki lahko te podatke najdejo, do njih dostopajo, jih združijo z drugimi podatki in procesi ter jih uporabijo v novem kontekstu. Objava podatkov ne koristi le javnosti, ki bi želela preveriti smiselnost rezultatov, ampak tudi avtorjem, saj nas javna hramba motivira, da jih predstavimo v urejeni obliki. Kljub vsemu je zagotavljanje ponovljivosti rezultatov v splošnem večplasten proces in presega zgolj prilaganje kode in podatkov objavi v znanstveni reviji.

Velike količine podatkov ponujajo globok vpogled v naravo, s seboj pa prinašajo odgovornost. Napake v razumevanju, naj bodo v razvrščanju zvezd v razrede ali pa v razumevanju uklonskih vzorcev na mikroskopu, lahko vodijo do zmotnih zaključkov in zamujenih odkritij. Avtomatizacija ni nadomestilo za razmišljanje, temveč njegov zaveznik: ko svoje delo razstavimo na jasne korake, pridobimo čas in miselni prostor za nova vprašanja, dvome in razlage. Če algoritmom slepo zaupamo, le zamenjamo eno breme z drugim. Če pa razumemo njihove meje in zmožnosti, nas osvobodijo rutinskih opravil in nas usmerijo k tistemu, zaradi česar smo v znanosti: izvirni misli.

https://www.alternator.science/sl/daljse/avtomatizacija-lahko-osvobaja-misel/