

MI ētiskais un strukturālais novērtējums

Bibliotēku atbildība MI kontekstā – ko tām jāizprot un par ko jāiestājas

🕒 60–90 min | Jebkuram interesentam

Kāpēc šie jautājumi nevar gaidīt

MI ieviešana attīstās straujāk nekā ētiskais ietvars — bibliotēkām jābūt šīs tendences priekšgalā



Ētikas klūmes notiek visu laiku

Sejas atpazīšanas sistēmas ir kļūdaini identificējušas nevainīgus cilvēkus. Darbinieku pieņemšanas algoritmi ir sistemātiski diskriminējuši sievietes.



Aizspriedumi ir sistemātiski, nevis nejauši

MI aizspriedumi izriet no apmācības datiem, kuros atspoguļojas vēsturiskā netaisnība. Tie paši no sevis nepazūd bez apzinātas un mērķtiecīgas rīcības.



Lietotāju dati ir apdraudēti

Katra mākslīgā intelekta sistēma, kas saskaras ar lietotāju datiem, rada nopietnus konfidencialitātes jautājumus, kurus bibliotēkām ir jārisina gan juridisku, gan ētisku apsvērumu dēļ.



Ietekme uz vidi ir reāla

Viena mākslīgā intelekta modeļa apmācība var radīt tikpat lielas CO₂ emisijas kā piecu automobiļu izmantošana visa to dzīves ciklā. Bibliotēkām, kas apņēmušās ievērot ilgtspējības principus, šis aspekts ir jāņem vērā



Kas kontrolē MI, tas ietekmē zināšanu apriti

Pašlaik neliels skaits korporāciju kontrolē MI sistēmas, kas visa pasaulē veido to, kā informācija tiek meklēta, atlasīta un parādīta lietotājiem. Tas ir jautājums par informācijas un zināšanu suverenitāti.



Mākslīgā intelekta ētika

Kas ir pareizi? Kas ir kaitīgi? Kas to izlemj?

Ētika ir MI priekšnoteikums.

Ētikas ietvars

Starptautiskas organizācijas ir izstrādājušas ietvarus — bibliotekāriem tās būtu jāzina un jāpielieto.

UNESCO (2021)

Pirmais globālais normatīvais instruments par mākslīgā intelekta ētiku. Uzsver cilvēktiesības, dzimumu līdztiesību, vides ilgtspēju u.c. Skaidri aptver mākslīgo intelektu izglītības un informācijas jomā.

ES MI akts (2024)

Uz risku balstīts regulējums

Klasificē mākslīgā intelekta sistēmas pēc riska līmeņa: (nepieņemams → augsts → ierobežots → minimāls).

IFLA (2020)

Starptautiskās bibliotēku asociāciju federācijas paziņojums, kurā uzsvērtā brīva piekļuve, privātums, brīvība no novērošanas un tiesības piekļūt informācijai bez algoritmu starpniecības.

OECD (2019)

Pieci principi: iekļaujoša izaugsme; uz cilvēku vērsta vērtības; pārredzamība; stabilitāte; atbildība. Pieņemti G20 valstīs un uz tiem ir atsauce lielākajā daļā nacionālo mākslīgā intelekta stratēģiju.

Kā MI nodara kaitējumu

Izpratne par dažādajiem ceļiem, kā mākslīgā intelekta sistēmas nodara kaitējumu reālajā pasaulē.

Kaitējums	Piemērs	Bibliotēku konteksts
Sadalījuma kaitējums	MI, pamatojoties uz neobjektīvām prognozēm, liedz kādam aizdevumu, darbu vai mājokļa pabalstu.	MI, kas liedz atsaka bibliotēkas lietotājiem pakalpojumus vai atzīmē viņu kontus, pamatojoties uz demogrāfiskām pazīmēm.
Reprezentatīvs kaitējums	MI saista noteiktas grupas ar negatīvām iezīmēm (pēc rases veidoti “noziedznieku” tēli).	MI katalogu meklēšanas sistēma, kas nepietiekami atspoguļo literatūru no Globālajiem Dienvidiem.
Kaitējums pakalpojumam	Balss atpazīšana sliktāk darbojas ar sieviešu balsīm un tiem, kam izmantotā valoda nav dzimtā.	Mākslīgā intelekta uzzīņu rīki, kas angļu valodas lietotājiem darbojas labāk nekā citiem.
Starppersonu kaitējums	Dziļviltojumi tiek izmantoti, lai uzmāktos, nomelnotu vai manipulētu ar reāliem cilvēkiem.	Lietotāji, pret kuriem vērsta MI ģenerēta dezinformācija, kuras pamatā ir no bibliotēkām iegūti dati.
Epistēmiskais kaitējums	Mākslīgais intelekts pārliecinoši sniedz nepatiesu informāciju, tādējādi graujot uzticību faktiem.	Lietotāji, kas akadēmiskos vai pilsoniskos pētījumos atsaucas uz ģenerētām halucinācijām.



Algoritmu neobjektivitāte

Neobjektivitāte nav kļūda. Tā bieži ir izmantoto datu rādītājs.

Ja pasaule ir netaisnīga un mākslīgais intelekts mācās no pasaules, tad mākslīgais intelekts mācās netaisnību.

Aizspriedumi MI sistēmās

Neobjektivitāte uzkrājas katrā mākslīgā intelekta izstrādes posmā — ne tikai datos

Datu ieguve

Mazākumtautību grupu nepietiekama pārstāvniecība; ierakstos iekodēta vēsturiska diskriminācija; atlasīti interneta dati, kas sliecas par labu turīgiem sabiedrības locekļiem.

Datu marķēšana

Cilvēku veiktā datu anotēšana atspoguļo anotētāju kultūras pieņēmumus un vērtības.

Modeļa dizains

Optimizācijas rādītāji (piemēram, precizitāte vairākuma grupās) var radīt maldinošu priekšstatu par sistēmas kvalitāti, slēpjot vājāku sniegumu mazākumgrupām.

Ģeogrāfiskais konteksts

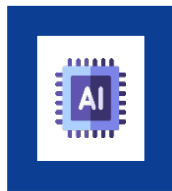
Rīks, kas apmācīts vienā kontekstā un ieviests citā, pārmanto neatbilstības. Medicīniskā MI, kas apmācīts ar ASV pacientu datiem, darbība Āfrikā neizdosies veiksmīgi.

Atgriezeniskā saite

Ja neobjektīvi rezultāti ietekmē nākotnes datus (piemēram, neobjektīvs darbā pieņemšanas mākslīgais intelekts → neobjektīvi darbaspēka dati → nākamais mākslīgais intelekts), neobjektivitāte laika gaitā uzkrājas.

Algoritmu neobjektivitāte

«Grupa ar golfa spēlētājiem»



Algoritmu neobjektivitāte



Algoritmu neobjektivit̃ate

≡ *Translate*

English ▼

Turkish ▼

He is a doctor



0 bir doktor



Algoritmu neobjektivitāte

Atkritumi iekšā, atkritumi ārā (angļu val. *Garbage In, Garbage Out (GIGO)*) - Nekvalitatīvi ievaddati rada nekvalitatīvus rezultātus.





Algoritmu neobjektivitāte

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

SCORING OF WELFARE BENEFICIARIES: THE INDECENCY OF CAF'S ALGORITHM NOW UNDENIABLE

27 November 2023

Algoritmu neobjektivitāte

- Konstatēts, ka *Amazon* mākslīgā intelekta darbā pieņemšanas sistēma ir neobjektīva pret sievietēm: apmācīta, izmantojot vēsturiskus tehnoloģiju nozares datus, kur pārsvarā bija vīriešu.
- Tiesībsargājošo iestāžu piemēri atklāj, ka MI radītu kļūdu līmenis, apstrādājot datus par cilvēkiem ar tumšāku ādu, ir līdz pat 34% augstāks.
- Medicīniskie algoritmi ir uzrādījuši neobjektivitāti melnādaino pacientu ārstēšanas vajadzību prognozēšanā.
- Mājokļu un darba sludinājumus var algoritmiski vadīt, pamatojoties uz demogrāfiskajām īpašībām.

Algoritmu neobjektivitate





Datu privātums un uzraudzība

"Ja tu nemaksā par produktu, tu pats esi produkts."

Bibliotēkas lietotāju dati nav resurss. Tās ir tiesības, kas jāaizsargā.

MI sistēmas un lietotāju dati

Datu plūsmas bieži vien ir neredzamas, taču to sekas ir redzamas.

Datu plūsmas ķēde

- 1 Lietotājs iesniedz vaicājumu, izmantojot MI tērzēšanu vai meklēšanu.
- 2 Vaicājumu apstrādā trešās puses (bieži vien citā valstī).
- 3 Pakalpojuma sniedzējs reģistrē vaicājumu modeļa uzlabošanai.
- 4 Lietošanas paradumi tiek apkopoti un varbūt pat pārdoti.
- 5 Parādās profili, kas sasaista grāmatu rezervēšanas, meklēšanas un vaicājumu paradumus.
- 6 Šie profili var būt pieejami iestādēm saskaņā ar noteiktiem regulējumiem.

Esošie aizsardzības pasākumi

- Datu aizsardzības regula (ES) - tiesības uz piekļuvi, dzēšanu un datu minimizēšanu.
- Daudzi bibliotēku konfidencialitātes noteikumi attiecas arī uz digitālo vidi.
- IFLA privātuma ietvars bibliotēku sistēmām.

Galvenie riski

- Datu pārsūtīšanai no trešajām valstīm var nebūt līdzvērtīgas privātuma aizsardzības.
- Mākslīgā intelekta rīks var piešķirt pārāk plašas tiesības apstrādāt datus.
- Lietotāju MI vaicājumi var atklāt sensitīvu informāciju (veselības, juridisko, politisko).

MI sistēmas un lietotāju dati

Klientu konfidencialitātei ir juridisks un ētisks pamats. Mākslīgais intelekts eksistē tā ietvarā.

PAMATA PRINCIPS: Bibliotēkas apņemšanās nodrošināt lietotāju privātumu nebeidzas pie mākslīgā intelekta robežas. Katrai mākslīgā intelekta sistēmai, kas apstrādā lietotāju datus, ir jāatbilst tādiem pašiem standartiem kā jebkurai citai bibliotēkas sistēmai.

1

Datu privātuma ietekmes novērtējums: pirms jebkura MI rīka, kas skar lietotāju datus, ieviešanas, veiciet dokumentētu ietekmes uz datu aizsardzību novērtējumu. Novērtējiet, kādi dati tiek vākti, kur tie plūst un kādi riski pastāv

2

Datu minimizācijas piemērošana: mākslīgā intelekta rīkiem jābūt konfigurētiem tā, lai tie apkopotu minimāli nepieciešamos datus. Personu identificējoša informācija, kad vien iespējams, jāizslēdz no mākslīgā intelekta vaicājumiem.

3

Izskatiet pakalpojumu sniedzēju konfidencialitātes politiku. Galvenie jautājumi: vai pakalpojuma sniedzējs saglabā vaicājumus? Cik ilgi? Vai datus var izmantot nākotnes modeļu apmācībai? Kas ir to apstrādātāji?

4

Lietotāju informētība: Ja lietotājiem sniegtajos pakalpojumos tiek izmantoti mākslīgā intelekta rīki, ir jāsniedz skaidrs paziņojums par to. Lietotājiem vajadzētu būt iespējai atteikties no mākslīgā intelekta starpniecības pakalpojumiem un piekļūt līdzvērtīgām cilvēku piedāvātām alternatīvām.

5

Apmāciet darbiniekus par privātumu mākslīgā intelekta jomā: ikvienam darbiniekam, kas izmanto mākslīgā intelekta rīkus, ir jāsaprot, ko viņš drīkst un ko nedrīkst ievadīt.



Ietekme uz vidi

MI izmantošana rada “oglekļa pēdas nospiedumu”, un bibliotēkām, kas apņēmušās veicināt ilgtspējību, ir jācenšas to mazināt. MI sistēmu enerģijas izmaksas nav acīmredzamas. Tās vienkārši ir “paslēptas” datu centros.

Ietekme uz vidi

Mērogs ir svarīgs. Mūsdienu mākslīgā intelekta sistēmu resursu prasības ir ārkārtējas.

626,000

kg CO₂

Aprēķinātās emisijas,
lai apmācītu vienu
lielo valodas modeli
(MIT pētījums).

10×

Vairāk enerģijas
ChatGPT vaicājums
patērē aptuveni 10
reizes vairāk
enerģijas nekā
standarta Google
meklēšana.

6.4 milj.

Kubikmetru ūdens

Paredzamais
dzesēšanas sistēmu
ūdens patēriņš
Microsoft MI datu
centros 2022. gadā.

2–4%

no pasaules CO₂

IKT sektora pašreizējā
globālo oglekļa emisiju
daļa strauji pieaug,
ņemot vērā
pieprasījumu pēc MI.



Mākslīgā intelekta suverenitāte un vara

Kas veido mākslīgo intelektu? Kam tas pieder? Kādas vērtības tas atspoguļo?

Kontrole pār mākslīgā intelekta infrastruktūru ir kontrole pār zināšanu nākotni.

MI koncentrācijas problēma

Pieci uzņēmumi — *OpenAI (Microsoft)*, *Google DeepMind*, *Anthropic (Amazon)*, *Meta AI* un *xAI* — pašlaik dominē vispējīgāko mākslīgā intelekta modeļu piedāvājumā. To lēmumi par datu apmācību, piekļuvi, cenu, satura moderēšanu nosaka piekļuvi informācijai miljardiem cilvēku.

Aprēķinu koncentrācija	Lielu mākslīgā intelekta modeļu apmācībai ir nepieciešami milzīgi skaitļošanas resursi, kas pieder vien dažiem mākoņpakalpojumu sniedzējiem (<i>AWS</i> , <i>Azure</i> , <i>GCP</i>).
Datu koncentrācija	Interneta dati, kas izmantoti <i>LLM</i> apmācībām, galvenokārt iegūti turīgās, angliiski runājošās valstīs.
Darbaspēka koncentrācija	Satura moderēšanu un datu marķēšanu mākslīgā intelekta vajadzībām lielākoties veic zemi apmaksāti darbinieki Globālajos Dienvidos, bieži vien ar minimālu darba aizsardzību.
Vērtības koncentrācija	Lēmumus par to, ko mākslīgā intelekta sistēmas teiks un ko neteiks, rādīs vai ieteiks, pieņem korporatīvās komandas nelielā skaitā valstu.
Regulējošā asimetrija	Mākslīgā intelekta regulējums visstraujāk attīstās ES. Daudzos citos reģionos ir ierobežota pārraudzība pār mākslīgā intelekta sistēmām, kas ietekmē to iedzīvotājus.

MI suverenitāte

Kolekciju suverenitāte: Kad mākslīgā intelekta sistēmas bibliotēku vārdā veido, filtrē vai klasificē kolekcijas. Lēmumi par to, kas ir redzams un kas tiek paslēpts no bibliotekāriem pāriet algoritmu ziņā, kuri izstrādāti komerciālām interesēm.

→ *Saglabāt cilvēka uzraudzību pār mākslīgā intelekta ietekmēto kolekciju pārredzamību.*

Zināšanu suverenitāte: MI sistēmas, kas galvenokārt balstītas uz Globālo Ziemeļu datiem, var pasniegt konkrētas kultūras tradīcijas zināšanas kā universālas. Bibliotēkām, kas apkalpo dažādas kopienas, ir kritiski jāizvērtē, ko mākslīgais intelekts zina un ko tas dzēš.

→ *Izcelt mazāk pārstāvētās zināšanu tradīcijas. Nelietot MI, kas mazina daudzveidību.*

Meklēšanas suverenitāte: Mākslīgā intelekta darbināta meklēšana maina to, ko lietotāji atrod. Lietotājam, kurš redz tikai to, ko parāda MI, informācijas vide ir ierobežota – to veido MI apmācītais rezultāts, nevis viņa faktiskās vajadzības.

→ *Piedāvāt meklēšanas veidus, kas nav MI darbināti.*

Juridiskā suverenitāte: Dati, ko apstrādā MI pakalpojumu sniedzēji, var būt pakļauti ārvalstu uzraudzības likumiem. Bibliotēkām jānodrošina, ka MI izmantošana pārrobežu kontekstā atbilst valsts privātuma aizsardzības noteikumiem.

→ *Pirms mākslīgā intelekta pakalpojumu iepirkuma lūgt juridisko atzinumu par datu glabāšanas pienākumiem.*

Grupas aktivitāte

4–6 cilvēku grupas | 25 minūtes | Izvēlieties vienu MI rīku, ko jūsu bibliotēka izmanto vai apsver lietot



Ētika

Vai šim rīkam ir ētikas regulējums?
Kas ir tā radītājs?
Kāds ir dokumentētais kaitējums?



Aizspriedumi

Vai rīks ir neatkarīgi pārbaudīts attiecībā uz demogrāfisko neobjektivitāti?
Kādas valodas un kopienas ir nepietiekami pārstāvētas tā apmācības datos?



Privātums

Kādi dati tiek vākti, glabāti un cik ilgi?
Vai rīks atbilst datu regulas prasībām? Vai ir pieejama dokumentācija?



Vide

Vai pakalpojumu sniedzējs publicē enerģijas patēriņa vai emisiju datus?
Vai datu centri darbojas ar atjaunojamo enerģiju?



Suverenitāte

Kurās valstīs dati tiek apstrādāti?
Vai pakalpojuma sniedzēja satura politika atspoguļo kopienas vērtības un tiesības?