

Gervigreindarlæsi – gagnrýnin notkun er forsenda gagnlegrar gervigreindar

Notkun gervigreindar eykst stöðugt og samkvæmt nýlegri skoðanakönnun segist helmingur Íslendinga nota vinsælasta gervigreindarforritið. En hvað er átt við með gervigreind? Orðið hefur verið notað um marga mismunandi virkni í tölvum, til dæmis í tengslum við tillögur að skemmtiefni á streymisveitum, sjálfkeyrandi bíla, greiningu á röntgenmyndum, vélþýðingar og það sem hefur fengið mesta athygli undanfarin misseri: spjallmenni. Þetta eru ólík fyrirbæri og eiga kannski fátt annað sameiginlegt en að vera tengd við gervigreind. Fjölmargar mismunandi skilgreiningar á gervigreind eru til – sumar svolítið galgopalegar eins og „gervigreind er það sem fyrirtæki sem gera út á gervigreind kalla gervigreind.“ Aðrar eru hagnýtar eins og sú sem skýrsla ESB um gervigreind frá 2018 notar: „Með gervigreind er átt við kerfi sem geta greint umhverfi sitt eða annað ílag og tekið ákvarðanir um aðgerðir – að einhverju leyti sjálf – til að ná tilteknum markmiðum.“ Enn aðrar skilgreiningar eru afmarkandi, t.d. sú sem Marvin Minsky notaði þegar hann sagði gervigreind vera „vísindasvið sem fjallar um gerð véla sem geta framkvæmt aðgerðir sem krefðust greindar ef þeir væru unnar af mönnum“.

Þegar talað er um gervigreind í þessum pistli er þó fyrst og fremst átt við það sem fólk hefur í almennu tali undanfarin ár kallað *gervigreindina*, með greini. Það er gervigreind sem byggir á myndandi tauganetslíkönum og getur með afar kraftmikilli (og orkufrekri) mynsturgreiningu lært að mynda nýja texta, myndir, tónlist og kvikmyndir. Þessi geta líkananna byggir á ógrynni gagna sem þau hafa verið látnin greina – texta, tal, myndir, myndbönd og tónlist, svo eitthvað sé nefnt.

Notkun þessarar gerðar gervigreindar, *myndandi gervigreindar*, hefur breiðst afar hratt út enda felast ótal tækifæri í notkun hennar. Hún hefur stórbætt vélþýðingar á stuttum tíma; forritarar geta flýtt fyrir sér með því að láta gervigreind leysa einföld verkefni og jafnvel líka flókin; gervigreind hefur verið notuð til að búa til myndir í auglýsingar; kennarar í skólum nota gervigreind sér til aðstoðar og nemendur til að hjálpa sér að skilja námsefnið betur og jafnvel stundum til að létta á álagi. En gervigreind getur líka verið notuð í vafasömum tilgangi eins og til að blekkja fólk, hrekkja eða klekkja á því með falsi, til dæmis með framleiðslu á falsfréttum og öðrum óreiðuupplýsingum, fölsuðum myndum eða myndböndum, eða textum til að auðvelda fjárplógsstarfsemi og svik. Og þótt ætlunin sé ekki að gera neitt vafasamt geta ýmsar aðrar hættur líka fylgt notkun gervigreindar.

Flest höfum við heyrt fjölmörg dæmi, bæði góð og slæm, um það hvernig gervigreind hefur reynst. Fólk hefur fundið bót meina sinna eftir að hafa notað gervigreind til að greina sjúkdóma, fólk hefur getað nýtt sér gervigreind til að auka skilning sinn á flóknum viðfangsefnum í námi, og svo eru sumir hæstánægðir með öll fínú lögin sem tónlistarstreymisveitan Spotify kynnir fyrir þeim – lög sem það hefði kannski aldrei heyrt ef gervigreindin hefði ekki valið þau fyrir þá. En svo eru líka dæmi um fólk sem hefur fengið ranga sjúkdómsgreiningu og það valdið því tjóni, fólk sem fær rangar útskýringar á því sem það þarf nauðsynlega að skilja, að ekki sé talað um þá sem eru hundóánægðir með öll þessi hræðilegu gervigreindarlög sem Spotify velur á spilunarlistana.

Gervigreindarlæsi snýst um að geta notað gervigreind með gagnrýnum hætti. Læsi er í þessu samhengi getan til að geta skilið. Hugtakið gervigreindarlæsi tengist þannig hugtökum á borð við upplýsingalæsi, tölvulæsi, fjölmiðlalæsi og vísindalæsi. Þjálfun í gervigreindarlæsi er þjálfun í að skilja hvernig gervigreind virkar, hverjir eru styrkleikar tækninnar og takmarkanir. Gervigreindarlæsi er að geta metið hversu vel við getum treyst tækninni fyrir þeim verkefnum sem við þurfum eða viljum nýta hana til.

Hægt er að tala um nokkur leiðarstef í gervigreindarlæsi:

- Hvernig virkar gervigreind?
- Hvernig getum við notfært okkur gervigreind og hvenær ættum við ekki að gera það?
- Eru einhver lagaleg eða siðferðisleg álitamál sem við þurfum að fást við þegar við notum gervigreind?

Skoðum hvers vegna þetta skiptir máli.

1. Að skilja hvernig gervigreind virkar

Þegar við skiljum á hverju tæknin byggir eigum við auðveldara með að meta hvenær hún getur gagnast okkur og hvenær ekki. Við áttum okkur á því að gervigreindin er ekki alviturt töfratæki heldur hugbúnaður sem notar flókin stærðfræðilíkön til að greina mynstur í gögnum sem eru alltaf að einhverju leyti ófullkomin. Gervigreindarlíkön læra þessi mynstur ekki af því að þau skilji heiminn svo vel – þau skilja í rauninni ekki neitt – heldur með því að stilla af tölur í risastóru vigurrými með því að bera saman ógrynna sambærilegra gagna, t.d. texta eða mynda, í ferli sem er kallað þjálfun. Þegar líkönin verða nógu stór geta þau svo framkallað efni sem er sambærilegt við það sem þau voru mótuð af í þjálfunarferlinu. Útskýringar á því sem býr að baki tækninni geta auðveldað notendum að átta sig á getu hennar og takmörkunum.

2. Hvernig getum við notað gervigreind okkur til gagns – og hvenær er hún líkleg til að vera til ógagns?

Það er lykilatriði að geta lagt gagnrýnið mat á hvort og hvenær notkun gervigreindar hentar og hvað sú notkun getur haft í för með sér. Sama hvað um er að ræða, þá er möguleiki á að það sem gervigreindarlíkan myndar sé rangt, ónákvæmt, hlutdrægt eða villandi, þótt það sé sannfærandi. Það er vegna þess að líkönunum er inngróið að líkja eftir gögnunum sem þau eru þjálfuð á. Það efni getur endurspeglað gamlar hugmyndir um samfélagið þar sem kynjamisrétti var meira en nú eða kynþáttafordómar samþykktari. Það getur líka boðið upp á misnotkun eins og þegar Microsoft þjálfði spjallmenni á gögnum af Twitter en þurfti að slökkva á því tæpum sólarhring síðar vegna þess að spjallmennið var farið að dreifa nasískum áróðri á samfélagsmiðlinum.

Notendur þurfa líka að vera gagnrýnir á allt sem gervigreindin myndar. Þeir þurfa að geta sannreynt það sem skiptir máli, sérstaklega það sem getur haft afleiðingar fyrir þá eða aðra í námi, starfi eða daglegu lífi. Þeir þurfa að gæta sín á leiðandi spurningum, því mörg líkananna eru þjálfuð til að þóknast notandanum og gervigreindin gæti þannig byggt undir ranghugmyndir og misskilning, frekar en að efla og fræða notandann. Það er því mikilvægt að átta sig á hvernig best er að setja fram spurningar eða verkefni þegar við notum gervigreind. Til að auðvelda okkur að skilja það sem gervigreindin myndar getur t.d. reynst vel að brjóta flókin verkefni niður í minni hluta og ganga úr skugga um að gervigreindin leysi hvern hluta verkefnisins eins og ætlast er til.

3. Áhrif gervigreindar á fólk, samfélag og umhverfi

Gervigreindin krefst gríðarlega mikillar orku og á síðustu árum hefur orkunotkun vegna gervigreindar aukist. Áætlað er að árið 2028 muni orkunotkun gervigreindartölvuvera í Bandaríkjunum samsvara um 22% af orkunotkun allra heimila í landinu. Það gerist á sama tíma og við þurfum að draga úr orkunotkun til að vinna gegn loftslagsbreytingum. Í því ljósi hljótum við að gera þá kröfu að

ávinningurinn af tækninni sé umtalsverður fyrir fólk og samfélög og gagnist í baráttunni gegn loftslagsbreytingum.

Þegar við notum gervigreind í hvers kyns sköpun, þurfum við að hafa í huga á hverju niðurstaðan byggir – til dæmis hvort hún brjóti mögulega í bága við höfundarrétt. Ógrynni gagna þarf við þjálfun gervigreindarlíkana. Ekki er alltaf gefið upp hvaðan gögnin koma og komið hefur í ljós að við smíði margra líkana hefur höfundaréttarvarið efni verið notað í óleyfi. Slík notkun getur verið skaðleg fyrir höfunda efnisins og fjölmörg dómsmál hafa verið höfðuð vegna þessa. Ætti það að hafa áhrif á það hvernig og hvort við notum tiltekin gervigreindarlíkon?

Notendur bera ábyrgð á því sem þeir búa til og deila með öðrum. Þeir þurfa að huga að því hvernig það sem þeir gera hefur áhrif á þá sjálfa og aðra. Það er auðvelt að nota tæknina til að búa eitthvað til sem er örlítið frábrugðið því sem áður hefur verið gert. Stundum getur það jafnvel fallið vel í kramið eins og gervigreindarlagið „Walk My Walk“ sem komst í efsta sæti vinsældalista í Bandaríkjunum nú í byrjun nóvember. Hvaða áhrif mun tæknin hafa á tónlistarheiminn? Gervigreind á heldur ekki í neinum vandræðum með að framleiða falsfréttir sniðnar að litlum hópum í áróðursskyni – hver verða langtímaáhrifin af því á lýðræðið? Við getum notað tæknina til að bæta lýsingu í ljósmyndum en líka til að breyta myndum af öðrum í þeim tilgangi að niðurlægja. Þess háttar tilvik hafa komið upp í skólum, í haust var fjallað um gervigreindarmyndband sem nemendur í grunnskóla gerðu af kennara og nemanda kyssast. Og þótt ætlunin sé að nota gervigreindina til gagns getur gagnrýnislaus notkun haft alvarlegar afleiðingar. Nærtækt dæmi um slíkt er þegar Lögreglan á höfuðborgarsvæðinu notaði gervigreind til að „skerpa“ mynd af mönnum sem hún leitaði að. Niðurstaðan varð sú að myndin sýndi menn sem litu allt öðru vísi út en þeir sem leitað var að. Allt þetta þurfum við að hafa í huga, ekki til að hræðast það heldur til að þekkja það, geta brugðist við því og hafnað því ef þau siðferðislegu viðmið sem við viljum fylgja krefjast þess.

Lögreglan harmar birtingu falsaðrar ljósmyndar

Lögreglan á höfuðborgarsvæðinu birti falsaða ljósmynd í tilkynningu í tengslum við rannsókn á eldsneytisþjófaði. Embættið harmar atvikið og ætlar að endurskoða verklagið.

Dorgerður Anna Gunnarsdóttir
31. júlí 2025 kl. 12:38, uppfært kl. 16:24



⚠ Þessi færsla er meira en 3 mánaða gömul



Mynd lögreglunnar gaf ekki rétta mynd af útliti mannanna - úr frétt á RÚV.is

Hér fyrir neðan pistillinn eru hlekkir í efni þar sem farið er ítarlega í gervigreindarlæsi, hvernig hægt er að kenna það og þjálfra sig og aðra í því. Þar sem ætlunin er að innleiða gervigreindartækni þarf gervigreindarlæsi að koma fyrst. Gagnrýnin notkun er lykillinn að gagnlegri notkun á gervigreind. Þjálfun í gervigreindarlæsi, með skýrum markmiðum, ætti því að vera lykilatriði í allri umræðu um notkun gervigreindar, ekki síst í skólakerfinu.

Ítarefni (á ensku) um gervigreindarlæsi:

[Understanding AI Literacy](#) – leiðbeiningar frá Stanford-háskóla um notkun gervigreindar fyrir leiðbeinendur og kennara

Mills, K., Ruiz, P., Lee, K., Coenraad, M., Fusco, J., Roschelle, J. og Weisgrau, J. (2024). [AI Literacy: A Framework to Understand, Evaluate, and Use Emerging Technology](#).

OECD (2025). [Empowering learners for the age of AI: An AI literacy framework for primary and secondary education](#).

Aðrar heimildir

EC Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. [Artificial Intelligence for Europe](#).

Gallup. [Gríðarleg aukning í notkun á gervigreind](#). 13. nóvember 2025.

Iðunn Andrésdóttir. [Gervigreindarmyndband af kennara og nemanda í sleik fór í dreifingu](#). Rúv.is, 23. október 2025.

Amy Kraft. [Microsoft shuts down AI chatbot after it turned into a Nazi](#). CBS News, 25. mars 2016.

Marvin L. Minsky. 1968. *Semantic information processing*. Cambridge, MA: MIT Press

James O'Donnell og Casey Crownhart. [We did the math on AI's energy footprint. Here's the story you haven't heard](#). MIT Technology Review, 20. maí 2025.

Þorgerður Anna Gunnarsdóttir. [Lögreglan harmar birtingu falsaðrar ljósmyndar](#). Rúv.is, 31. júlí 2025.