

# Atli Harðarson: Háskólamenntun og gervigreind

Erindi á fundi um gervigreind og háskólakennslu sem haldinn var

á Háskólatorgi 21. febrúar 2023 á vegum Nýsköpunarstofu menntunar

Svo lengi sem ég man eftir mér hafa vélmenni verið hluti af framtíðinni. Ég var barn að aldri þegar ég las um möguleika á að vélmenni ynnu menn í skák, svo komu vélar sem mátuðu stórmeistara en þær voru ekki kallaðar vélmenni heldur tölvur. Í teiknimyndum um framtíðina voru vélmenni sem unnu verk eins og að ryksuga. Nú eru sjálfvirkar ryksugur algengar en þær eru ekki kallaðar vélmenni. Sama má segja um vélar sem aka bílum í borgarumferð eða afgreiða í verslunum. Þær eru ekki nein „menni“ heldur bara sjálfkeyrandi bílar eða afgreiðsluvélar. Í hvert sinn sem einhver gerð „vélmenna“ hefur orðið algeng hefur fólk séð hvað hún er algerlega laus við alla mennsku og gefið henni nafn sem endar á einhverju öðru en „menni“.

Svo lengi sem ég man eftir mér hefur hugmyndin um gervigreind líka verið á kreiki. Hún var mótuð fyrir miðja síðustu öld og sett fram af Alan Turing (1950) í grein sem heitir *Computing Machinery and Intelligence*. (Íslensk þýðing mín á þessari grein birtist í tímaritinu *Hug* árið 1995).

Turing áleit að á þessari öld, þeirri tuttugustu og fyrstu, yrðu til vélar sem hugsuðu og hefðu raunverulegt vit. Nýlegar tækniframfarir sem gera vélum kleift að þýða og semja texta og svara alls konar spurningum kunna að benda til að hann hafi haft nokkuð fyrir sér. Sagan um skáktölvur, ryksugur og bíla gefur samt tilefni til að spyrja hvort vélmenni verði alltaf í framtíð og nái því aldrei að verða hluti af neinni samtíð. Hvort líkingin við mannveru verður ekki tóm fjarstæða um leið og við sjáum vélarnar að verki.

## Frá Joseph Weizenbaum til Cathy O'Neil

Frá dögum Alans Turing hafa bjartsýnismenn og svartsýnismenn um gervigreind togast á. Sumt sem bjartsýnismenn hafa sagt um aukna sjálfvirkni og hagræði sem af henni leiðir er augljóslega rétt. En þeir svartsýnu hafa líka slegið keilur. Einn þeirra er Joseph Weizenbaum. Hann var frumkvöðull í gervigreindarforritun og bjó til frægan spjallbotta sem var kallaður Eliza. Þetta var árið 1966. Eliza blessunin gerði lítið annað er endurtaka það sem fólk sagði og jánka því. Hún var eiginlega hálfgerður skjallbotti og kryddaði mál sitt með frösom á borð við „segjum tvö, þetta hefur mér svo iðulega fundist líka.“ Mörgum viðmælendum sem kunnu lítil skil á forritun þótti hún ákaflega vitur og skilningsrík og við það varð höfundur hennar nokkuð brugðið.

Áratug eftir að Eliza var fyrst ræst sendi Weizenbaum (1976) frá sér bókina *Computer Power and Human Reason*. Þar var hann afar gagnrýninn á yfirlýsingar gervigreindarfræðinga um dásemdir þessarar tækni og sagði meðal annars að jafnvel þótt það sem bjartsýnustu talsmenn

hennar sögðu reyndist satt öðluðust vélar hvorki menningar tilfinningar né dómgreind. Þær yrðu siðlaus vitsmunaskrímsli og við ættum að fara varlega í að færa þeim vald yfir lífi okkar.

Þótt tæknin færði mönnum vissulega ódýrari og fullkomnari iðnvarning með aukinni sjálfvirkni óttaðist Weizenbaum að hún ógnaði líka frelsi okkar. Sumt af því sem hann sagðist óttast hefur síðar orðið að veruleika. Um þetta er fjallað í merkilegri bók eftir Cathy O'Neil (2016) sem heitir *Weapons of Math Destruction*. Höfundurinn er stærðfræðingur. Hún vann á tímabili við að forrita fjáröflun vogunarsjóða og síðan við að reikna út hvernig best borgaði sig að bauna auglýsingum á fólk gegnum samskiptamiðla. Vegna þessara starfa fór hún að hugsa um hvernig sjálfvirk söfnun og úrvinnsla gagna er notuð í sívaxandi mæli til að fylgjast með fólki, hafa áhrif á það og úthluta lífsgæðum. Hún sneri þá við blaðinu og tók að vara við hættum sem fylgja sjálfvirkri stjórnsýslu og vélrænu mati á einstaklingum.

Sögurnar sem Cathy segir minna sumar á framtíðarhrollvekjur eins og 1984 eftir George Orwell (1949/2001) þar sem Stóri bróðir fylgdist með öllu og minnstu víxlspor dugðu til að menn ættu sér engrar viðreisnar von. Sá „stóri bróðir“ sem Cathy lýsir er ekki mennskur einstaklingur heldur fjöldi meira og minna samtengdra upplýsingarkerfa sem enginn hefur yfirsýn yfir og fáir vita hvernig virka. Sum þeirra ráða því hvort einstaklingur fær lán og þá á hvaða kjörum, önnur hvað tryggingar hans kosta, enn önnur hvort hann fær vinnu, skólavist eða reynslulausn úr fangelsi. Svo eru líka kerfi sem fylgjast með hverjir eru veikir fyrir með þeim hætti að auðvelt sé að selja þeim einhverja „snákaolíu“. Í hvert sinn sem við notum Facebook og Google eru skráðar upplýsingar um okkur sem er bætt við gagnasöfn slíkra kerfa.

Ein af fyrstu sögunum í bókinni segir frá matskerfi til að meta kennara í höfuðborg Bandaríkjanna, Washington. Þeim sem fengu lakastan vitnisburð hjá vélunum var sjálfkrafa sagt upp. Það átti á hreinsa til.

Mat á kennara réðist af reikniverki sem tók meðal annars tillit til árangurs nemenda hans á samræmdum prófum. Þetta voru krossapróf sem var svarað með blýanti. Eitthvað var um að skólar svindluðu og leiðréttu röng svör nemenda sinna áður en þau voru send í yfirferð. Kennslukona sem Cathy segir frá var rekin. Hún kenndi fimmta bekk, var vel liðin og álitin standa sig afar vel í starfi af fólki sem til þekkti. En árið áður voru nemendur hennar í fjórða bekk og þá var að líkindum átt við úrlausnir þeirra. Eftir eitt ár hjá umræddum kennara virtist þeim því hafa farið aftur. Matsmaður af holdi og blóði hefði sennilega áttað sig á hvers kyns var en það gerðu vélarnar ekki. Kennslukonan reyndi að kvarta. Þeir sem hún náði sambandi við vissu ekki hvernig kerfið virkaði og gátu ekki útskýrt hvernig „falleinkunn“ hennar var reiknuð. Það var ekkert hægt að gera. Vitsmunaskrímsli af því tagi sem Weizenbaum varaði við hafði bara prentað út uppsagnarbréf og það var enginn kostur á skýringum, endurupptöku eða áfrýjun.

Svo kemur hver sagan af annarri. Ansi margar þeirra eru líkar að því leyti að sjálfvirku kerfin ráðast helst á þá sem eru varnarlausir eða fátækir. Vitvélar sem fyrirtæki nota til að meta umsækjendur um atvinnu hafa til dæmis upplýsingar um yfirdrátt á bankareikningum inni í

sínu formúluverki. Þetta verður vítahringur þar sem atvinnuleysi veldur skuldasöfnun og skuldirnar draga úr líkum á að fá vinnu.

## Fylgjur skólans og glíma hans við tæknina

Frá upphafi menntabyltingarinnar fyrir 100 til 200 árum – þegar skólaganga varð stór hluti af lífi nær allra – hefur menntakerfið átt sér tvær fylgjur, aðra illa og hina góða. Sú illa er martröð þar sem kennarinn miðlar, nær óvirkir nemendur gangast undir próf til að mæla hve vel þeir tóku við og niðurstöðurnar eru svo notaðar til að flokka þá og skipa á bása. Góða fylgjan er draumurinn um skapandi skóla þar sem allir njóta sín sem jafningjar og eru virkir leitendur og veitendur.

Meðal þeirra sem ræða gervigreind og skólastarf má finna dæmi um bláeyga og svolítið gleiðgosalega bjartsýni um að tölvutæknin víki martröðinni frá og láti góðu draumana rætast. En við höfum líka efasemdaraddir og veruleikinn sjálfur er beggja blands – enda draga skólar sem til eru bæði dóm af góðu fylgjuni og þeirri vondu.

Í árdaga tölvutækni í skólum gerðu menn eins og Seymour Papert (1980) sér vonir um að tæknin ýtti undir skapandi starf. Ykkur rámar ef til vill í að Papert var einn af þeim sem þróuðu forritunarmálið Logo á sjöunda áratug síðustu aldar. Hann taldi að með því að bjóða börnum að forrita tölvur mætti auka áhuga þeirra á ýmsum námsgreinum og gera þeim kleift að gera sínar eigin uppgötvanir og læra að hugsa rökrétt og skynsamlega eins og vísindamenn. Þessi draumur hans hefur ekki ræst. Veruleikinn er nær því sem sagnfræðingurinn Larry Cuban (1993) lýsti í grein sem heitir *Computers Meet Classroom; Classroom Wins*. Þar segir hann að þegar ný tækni og gamlar skólahefðir glíma þá hafi skólinn betur. Með tölvum í kennslustofum, myndvörpum og PowerPoint festist miðlun til óvirkra viðtakenda í sessi og þegar allir urðu sitengdir varð hægt að halda henni áfram jafnvel þótt nemendur væru í sóttkví heima hjá sér. Tæknin hefur fram til þessa ekki stuggað burt martröðinni um skóla þar sem nemendur eru lítt virkir viðtakendur og kerfið flokkar þá eftir því hve vel tekst að mata þá.

En hvað með nýjustu tækni og tækni framtíðarinnar? Hvað með gervigreind sem getur þýtt milli tungumála, samið ritgerðir um alls konar efni og svarað spurningum á skriflegum prófum?

Hér held ég að geti brugðið til beggja vona.

## Háskólar og gervigreind

Á undanförunum áratugum hefur vöxtur menntakerfa mest verið á háskólastigi. Háskólar gegna orðið veigameira hlutverki en framhaldsskólar við flokkun fólks inn á lagskiptan vinnumarkað. Þeir gera þetta að verulegu leyti með prófum. Fjölmennustu kúrsarnir í háskólum heimsins eru á netinu og nemendur gangast undir rafræn próf. Sumum kennslubókum fylgir raunar hugbúnaður sem bæði semur próf að miklu leyti og fer yfir þau. Með slíkum tækjum er hægt að munstra á fólk einingar í milljónavís með fremur lítilli vinnu kennara. En hvað ef nemandinn sem þreytir próf hefur aðgang að vél sem getur svarað? Getur hann þá ekki gramsað til sín

einingar svo hundruðum skiptir án neinnar fyrirhafnar? Verður svona verksmiðjuframleiðsla á einingum og gráðum þá ekki marklaus með öllu og sjálfhætt?

Ég veit að sumir háskólar reyna að forðast þetta með því að tryggja að nemendur séu vélarvana og sambandslausir í prófum. Væntanlega styttest í að þeir verði með ósýnilegar vélar innan í eigin líkama og það þarf þá ansi nærgöngult eftirlit til að koma í veg fyrir „svindl.“ Það má láta sig dreyma um að vegna þessa endurheimtum við háskóla þar sem samræða og leit að þekkingu eru í fyrirrúmi fremur en flokkun á fólki inn á vinnumarkað. Slíkri flokkun verður væntanlega hætt ef hún verður með öllu óframkvæmanleg. Kannski öðlast hugsjónin um universitas sem samfélag leitandi huga þá nýtt líf.

Það væri samt heldur barnalegt að bjóða alla heimsins gervigreind velkomna á þeim forsendum að hún geri þá hlið háskólanna sem sver sig í ætt við martröðina annað hvort ómögulega eða svo marklausa að hún hætti að vaxa eins og hún hefur gert. Á meðan við enn höfum löggilt starfsheiti og prófgráður sem þýða eitthvað þurfum við að flokka fólk. Og eins og Cathy O’Neil bendir á í bók sinni frá 2016 er hægt að nota gervigreind til þess arna, og líka til að fylgjast með fólki og safna um það upplýsingum sem segja meira um einkalíf þess, þankagang, hæfni og lundarfar en einkunnir á prófum munu nokkurn tíma gera. Tæknin sem ef til vill gerir einkunnir ómögulegar getur gert miklu ágengari og ranglátari mannjöfnuð hættulega ódýran og auðvöldan.

Hvað kemur út úr glímu háskólanna við gervigreindina vitum við ekki – en við skulum að minnsta kosti vera minnug þess að þegar skólinn glímdi við tölvuna féll bannsett martröðin kannski í augnablik á annað hnéð en ekkert meira en það.

## Rit

Cuban, L. (1993). Computers meet classroom: Classroom wins. *Teachers college record*, 95(2), 185–210.

O’Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown publishers.

Orwell, G. (2001). *Nineteen eighty-four*. <http://gutenberg.net.au/ebooks01/0100021.txt> (Upphaflega gefin út 1949).

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman & co.