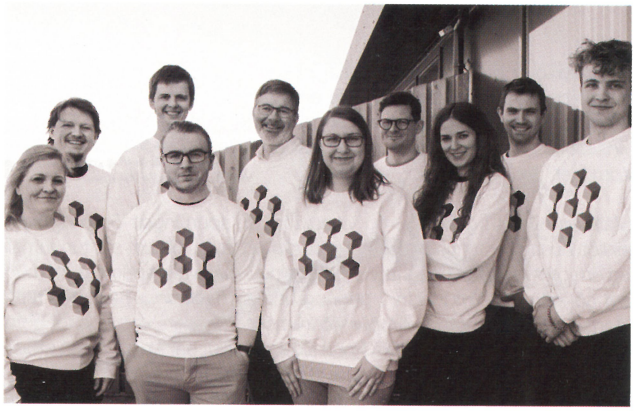
VÍTT OG BREITT UM MÁLLÍKÖN

Starfsfólk Miðeindar



Greinin er unnin upp úr fyrirlestrum sem starfsfólk Miðeindar hefur haldið um risamállíkön og hvernig hægt er að nýta þau. Risamállíkön á borð við GPT-4 og myndlíkön á borð við Midjourney hafa umbylt væntingum fólks til gervigreindar. Þau geta einfaldað vinnuferla en þeim fylgja líka nýjar áskoranir.

HVAÐ ERU MÁLLÍKÖN?

Tauganetslíkön má þjálfa á texta, mynd, hljóði, eða blöndu þess. Greinin fjallar aðallega um mállíkön (e. *language models*) þjálfuð á texta. Nokkrar tegundir hafa náð festu og kallast grunnmállíkön. Þau eru þjálfuð til að halda áfram með texta eða fylla inn í eyður:

> (1) Fjármálaráðherra lagði fjárlagafrumvarp fyrir <?> (*spá næsta orði)*
> *(2)* Fjármálaráðherra lagði <?> fyrir Alþingi. (*fylla í eyður*)

Til að leysa svona verkefni þarf líkanið grunnskilning[1] á tungumálinu, efnislegan og setningafræðilegan. Til að gera **betur** þarf heimsþekkingu. Líkanið lærir þessa hæfileika af fjölmörgum dæmum.

Runulíkön (einnig grunnlíkön) læra vörpun úr inntaki í úttak:

> (3) <is>Sólin mun skína á morgun. (*inntak*) <en>The sun will shine tomorrow. (*úttak*)

Pörin geta m.a. verið setningar á ólíkum tungumálum fyrir þýðingar, upprunalegur og leiðréttur texti fyrir málfarsleiðréttingu eða mynd og viðeigandi textalýsing. Mestu skiptir að sömu upplýsingar séu í inntaki og úttaki, annars er verkefnið illa skilgreint og hegðun líkansins ófyrirsjáanleg.

Nógu stór líkön sýna hæfileika til að leysa verkefni án sérstakrar þjálfunar. Hæfileikinn til að spá „rétt" eykst með stærð líkans, gagnamagni og gagnagæðum. Einnig mætti gefa líkaninu fleiri „sýnidæmi"; fleiri spurningar með svörum (e. *few-shot*) í stað þess að krefjast að líkanið leysi spurningasvörunarverkefnið án sýnidæma (e. *zero-shot*). Engin þjálfun fer fram; líkanið var þjálfað í byrjun en er svo stýrt með sýnidæmum í keyrslum (e. *prompt*).

Gríðarlegt reikniafl þarf til að grunnþjálfa risalíkön, jafnvel heilu reikniverin. Að grunnþjálfun lokinni má fínþjálfa (e. *finetune*) þau, keyra og hýsa með mun minni tilkostnaði. Margvíslegar tilraunir hafa verið gerðar til að smækka líkön en hættan er að líkönin geta tapað eiginleikum og staðið sig verr ef smækkunin er of mikil. Það er jafnvægislist að feta milli kostnaðar og frammistöðu.

Til eru nokkur íslensk mállíkön, en engin á stærð við GPT-4. IceBERT, BERT-líkan fyrir íslensku, getur flokkað texta en ekki búið til nýjan texta, eins og dæmi (2). mBART-enis er runulíkan líkt og dæmi (3) og er undirstaða velthyding.is. Runulíkanið ByT5 er notað í málfarsleiðréttingu og er undirstaða ai.yfirlestur.is.

UM RISAMÁLLÍKÖN

Risamállíkön eru flest þjálfuð á textum af netinu og úr gagnasöfnum, oftast á ensku. Líkönin eru yfirleitt fínþjálfuð í að skila vel mynduðu úttaki. Í GPT-4 er notuð styrktarþjálfun með mannlegri endurgjöf (e. *reinforcement learning – human feedback*, *RLHF*), sem kennir mállíkaninu að skilja spurningar og verkefni og svara þeim rétt og vel. Við þróun GPT-4 gerði OpenAI, í samstarfi við Miðeind, í fyrsta skipti tilraunir með þjálfun GPT með RLHF á öðru tungumáli en ensku.

Samstarf Miðeindar og OpenAI hófst í kjölfar heimsóknar forseta Íslands og sendinefndar til höfuðstöðva OpenAI í maí 2022. Meðal þátttakenda var stofnandi Miðeindar og að hans frumkvæði upphófust samræður milli fyrirtækjanna tveggja um hvernig íslenskan gæti nýst OpenAI sem fyrirmynd að stuðningi við smærri tungumál í risamállíkönum. Fyrsti áfangi samstarfsverkefnisins fólst í að kenna GPT-3 íslensku með fínþjálfun og meta hvaða textamagn þarf til að kenna risamállíkani tungumál.

Þegar undirbúningur GPT-4 hófst haustið 2022 leitaði OpenAI til Miðeindar um að taka þátt í þjálfuninni með RLHF. Miðeind fékk 40 sjálfboðaliða til að útbúa spurningar og verkefni á íslensku, meta svör líkansins og kenna því að svara betur. Gögnin voru notuð í þjálfun GPT-4 svo líkanið tók framförum í að skilja spurningar og svara á íslensku. Nú svarar líkanið nánast eingöngu á íslensku en áður slæddust með svör á öðrum málum. Líkanið skilur nú íslensku vel en á erfiðara með myndun, svo verkefninu er ólokið.

GPT-4 er aðeins eitt líkan og fjölmörg önnur hafa birst síðustu misseri. Má þar nefna BLOOM, LLaMA, OPT, GLM, Dolly-v2 og GPT-SW3, skandinavískt spunalíkan þjálfað á íslensku. Í hverri viku koma fram ný líkön svo listinn er ekki tæmandi. Ekki ætti að setja öll eggin í sömu körfuna og treysta á þriðju aðila sem bera ekki endilega hag íslensku fyrir brjósti. Það er því mikilvægt að Ísland setji sér skýra stefnu varðandi gervigreind.

Fyrir tíma risamállíkana var tímafrekt að útbúa líkön; gagnasöfnun, gagnamerking, þjálfun, rekstur, viðhald og innleiðing hjá notanda. Það er ekki fyrr en í síðasta skrefinu sem ágóði líkansins er ljós. Ferlið gat verið langt, kostnaðarsamt og þurft margar ítranir. Risamállíkön leysa fjölda verkefna án þjálfunar og gagnasöfnunar og því flest fyrri skref óþörf. Þess í stað er verkefninu lýst fyrir líkaninu, 1-2 sýnidæmi um góða lausn gefin og svo spreytir líkanið sig á raunverulegum dæmum. Risamállíkön gera því ýmsar máltæknilausnir aðgengilegri fyrir fólk og fyrirtæki. Risamállíkön eru þó ekki lausnin á öllum vandamálum. Meta þarf hvort risamállíkan sé rétti kosturinn eða hluti af heildarlausn.

HAGNÝTING RISAMÁLLÍKANA

Risamállíkön eru til margs fær og nýstárleg notkunardæmi sjást daglega. Hafa þarf þó í huga hvað þau geta (og geta ekki) gert. Gott er að líta á fyrirspurnina sem afbrigði af forritun. Við hönnum leiðbeiningar fyrir líkanið þar sem allar nauðsynlegar upplýsingar og skilyrði þurfa að koma fram og sýnidæmi geta hjálpað. Ekki má gleyma að líkönin geta haldið samhengi á milli fyrirspurna svo hægt er að biðja um betra svar ef eitthvað vantar. Það má líta á líkönin sem duglegan en (stundum) fljótvirkan starfsnema.

Eftir þjálfun læra líkönin ekkert nýtt og vita ekkert um atburði eftir þjálfun. Líkönin eru endurþjálfuð reglulega, sem er gífurlega kostnaðarsamt. Líkönin eru þjálfuð til að geta sér til um hvað kemur næst, svo þau eiga það til að búa til staðreyndir (e. *hallucinations*), og eru ekki áreiðanleg í flókinni röksemdafærslu. Ef þau eiga að svara upp úr þekkingargrunni þarf að setja þeim fastar skorður um að svara aðeins upp úr honum. Líkönin eru að auki þjónustulunduð og treysta notendum um of. Ef notandi leiðréttir líkan ranglega vilja þau taka því sem sönnu.

1 Til einföldunar tölum við um að líkanið hafi skilning.

Mannkynið er breyskt og fordómar okkar birtast í því sem við skrifum, þó að það sé ómeðvitað. Líkönin eru þjálfuð á efni frá mannfólki svo líkönin erfa bjaga (e. *bias*) sem finnast í textunum. Afbjögun líkana er vinsælt rannsóknarefni, en enginn haldbær árangur hefur náðst enn sem komið er.

Mállíkön nýtast í greiningu og vinnslu texta, spurningasvörun og sem alhliða aðstoðarmenni. Mállíkönin geta skrifað texta í ólíkum stíl, tungumáli og um ólíkt efni. Með gagnagrunnstengingu má svara spurningum með því að leita að skyldu efni í gagnagrunninum. Þannig þarf ekki tíma starfsmanns og notandi sleppur við frumskóg ítarlegra vefsíðna tengdra efninu. Mállíkönin gagnast líka í innri skjalavinnslu, þar sem t.d. má útbúa samantekt á löngum reglugerðum til að auðvelda yfirsýn yfir flókið efni. Hægt er að búa í haginn fyrir notkun mállíkana og skoða hvaða gögn eru til staðar.

SIÐFERÐISLEGAR SPURNINGAR

Til að tryggja ábyrga notkun mállíkana þarf að hugsa málið til enda áður en byrjað er.

Passa þarf að líkanið taki ekki ákvarðanir sem ýfa upp bjaga úr þjálfunargögnum. Evrópusambandið hefur gengið einna lengst í að móta stefnu um notkun gervigreindar. Til skoðunar er að skylda aðila til að upplýsa um það hvenær vél tekur ákvarðanir um hagi fólks, sem gæti krafist þess að manneskja staðfesti ályktanir líkansins. Hugsanleg notkun gervigreindar er einnig flokkuð í áhættuflokka, og lagt er til að banna alfarið þann áhættusamasta.

Ýmis álitamál tengjast þjálfunarferlinu og gagnanotkun. Þar er helst til skoðunar gagnsæi uppruni þjálfunargagna, og svo persónuverndarsjónarmið varðandi þjálfunargögn og hvert gögnin berast við notkun líkana. Myndlíkön sem hafa verið þjálfuð á myndasöfnum á vefnum hafa valdið hörðum deilum. Listafólk hefur mótmælt því að þeirra efni sé notað til að þjálfa líkan sem á að vinna þeirra störf án þess að þau fái nokkuð fyrir. Svipaðar deilur hafa spunnist um réttinn til að þjálfa á textum af netinu.

LOKAORÐ

Gervigreindin mun hafa áhrif á fjölbreytt störf og þróunin er þegar hafin. Það stefnir í umfangsmiklar samfélagslegar breytingar, sem geta vakið ugg, en ekki má gleyma gífurlegum ávinningi sem fylgir. Gervigreindin mun nýtast á sviðum sem erfitt er að sjá fyrir. Frumgerðir að stjórnun gerviútlima með raddstýringu („Taktu upp kaffibollann á borðinu“) hafa þegar litið dagsins ljós. Appið Be My Eyes er gott dæmi um aðgengisstuðning með hjálp gervigreindar. Þar geta blindir og sjónskertir notendur fengið samband við sjáandi sjálfboðaliða í appinu, sýnt þeim umhverfið og fengið t.d. að vita hvar gleraugun enduðu. Mörgum þykir óþægilegt að sýna ókunnugum einkalíf sitt, svo sérstök sjóngædd útgáfa GPT-4 var tengd inn. Líkanið fær þá spurningu frá notanda og myndefni sem inntak og svarar hvar gleraugun lentu. Möguleikarnir eru miklir en við verðum að stíga rétt og varlega til jarðar.

GOGG ÁHRIFIN

Kári Harðarson, tölvunarfræðingur

Kannast þú við pappírsleikfang sem er kallað „Goggur“? Hann er brotinn saman, trélitaður og með tölum innan í. Þú átt að velja eitthvað hornið og sjá hvaða tala er á bak við. Það er erfitt fyrir flesta (a.m.k. mig) að muna hvað er bak við hvaða lit. Ef goggurinn er brotinn sundur verður allt augljóst því yfirsýnin fæst.

Ég vil halda því fram að tölvunarfræðingar eigi að forðast að búa til svona gogga að óþörfu. Flokkið hugsanir og verk og gerið þeim sem koma á eftir ykkur kleift að lesa í gegnum skipulegan, línulegan texta.

Ég vil nefna nokkur dæmi til að forðast. Fyrsta dæmið eru samskiptaforrit en þau fá oft að þrífast mörg saman og það þarf að kíkja inn í hvert og eitt þeirra. Er einhver að ná í þig með Gmail, Outlook, Slack, Zoom, Teams eða Jira eða jafnvel á Facebook? Oftar en ekki er hvert innbox hálftómt. Misstir þú af einhverju?

Einn samskiptastaðall gæti leyft okkur að velja sjálf hvernig, hvar og hvenær við tökum á móti upplýsingum. Svona stöðlun var lykillinn að internetinu, mörg net voru sameinuð í eitt en það þurfti Bandaríkjaher til. Öll netlögin starfa saman í dag nema það efsta. Ég vona reyndar að þetta netlag verði einhvern tíma staðlað en ég ætlaði ekki að leysa málið, bara benda á svæsið tilfelli af „gogg áhrifunum“.

Athygli er takmörkuð auðlind og margir keppa um hana eins og sést í dæminu að ofan. Ekki sólunda henni með því að breyta línulegum lestri í síðuflakk („browsing“) um tómar síður og möppur.