

Predgovor

Na najbolj splošni ravni lahko sodoben elektronski računalnik obravnavamo kot edinstven spoj matematične teorije, elektrotehnike in programiranja.

Lahko rečemo, da je čistim zamislim matematične logike pomagal vdihniti praktično, otipljivo življenje, ki današnji čas z računalniško tehniko dejansko prežema do obisti.

Turingove zamisli predstavljajo eno od izhodišč, ki lahko bralcu s skoraj neprimerljivo jasnostjo osvetlijo marsikatero značilnost sveta, v katerem živimo.

Matematična in logična izhodišča

Danes so poznani kot Turingovi stroji. Gre za namišljene stroje, ki po navodilih lahko spreminjajo znake, zapisane na neskončnem traku.

Turing v izbranih delih večkrat praktično, na roke, prikaže zamisel, ker še ni imel dostopa do sodobnega elektronskega računalnika (urejanje neurejenega stroja s pomočjo omejitvenih posegov, šahovski program). Sam prevzame vlogo stroja in natančno oponaša njegovo delovanje, kar imenuje »papirni stroj«.

Pokazala so, da lahko izdelamo popoln računski stroj - računalnik, ki je lahko zelo preprost in kljub temu vsesplošno uporaben.

Izdelava splošnega elektronskega računalnika

Po tehničnih načrtih Maxa Newmana in dešifrirnih postopkih Alana Turing so nato v sodelovanju s telekomunikacijskim inženirjem Tommyem Flowersom izdelali še bolj zmogljivo elektronsko napravo za dešifriranje kod »Tunny«, prvi elektronski računalnik na svetu Colossus (1943).

Turingu pogosto pripisujejo izključne (ali vsaj prevelike) zasluge za razvoj elektronskega računalnika. Njegov doprinos je vsekakor temeljen, vendar je treba posebej poudariti, da tudi omejen.

Sodobni elektronski računalniki so bili (tudi na začetku petdesetih let, ko so nastala izbrana dela), še vedno težko dostopni. Lahko bi jih prešteli na prste.

Izdelava elektronskega posnetka možganov

Izhajal je iz predpostavke, da je popolna mehanizacija umskih zmožnosti možna, in celo verjel, da bodo elektronski računalniki lahko nekoč presegli zmožnosti človeka.

Obravnava jih kot otroke (nepopisane liste), ki jih lahko vzgojimo, vendar ne moremo v celoti vnaprej predvideti njihovih značilnosti, dejanj in rezultatov, ko enkrat odrastejo.

Predlagal je, da večjo učljivost in prilagodljivost strojev okolju dosežemo z uvajanjem (strojnih ali programskih) posnetkov neurejenih mrež živčnih celic ter uvajanjem nekakšne strojne občutljivosti za dražljaje in posege od zunaj.

*

V preizkusu spraševalec preprosto poskuša ugotoviti, ali odgovore na postavljena vprašanja podaja človek ali stroj.

Posebej izpostavi splošnost (univerzalnost) elektronskih računalnikov, ki jo razume kot zmožnost oponašanja vseh drugih vrst računskih strojev s programom. Problem izdelave inteligentnega stroja tako prenese na problem izdelave ustrezne programske opreme za elektronski računalnik.

Naposled predlaga, da elektronske računalnike z ustrezno programsko opremo in učnimi postopki lahko spremenimo v inteligentne stroje. Postopek primerja z odraščanjem in vzgojo otrok ter uvaja različne zamisli za strojno učenje, od zagotavljanja primerne nabora osnovnih navodil in podatkov (znanja), do postopkov za inteligentno obravnavo podatkov in učenje iz izkušenj s pomočjo »nagrad« in »kazni«.

1. Igra posnemanja

Namesto da nadaljujemo s tem problematičnim pristopom, predlagam, da izvirno vprašanje nadomestimo z drugim sorodnim vprašanjem, ki je izraženo bolj jasno in nedvoumno.

Problem v novi obliki lahko opišemo kot igro, ki se imenuje »*igra posnemanja*«.

Sedaj lahko zastavimo novo vprašanje: »Kaj se bo zgodilo, ko bo stroj prevzel vlogo igralca A v tej igri?« oziroma: »Ali se bo spraševalec v tem primeru odločal napačno tako pogosto, kot se odloča napačno, ko poteka igra med moškim in žensko?« Na ta način smo nadomestili prvotno vprašanje: »Ali stroji lahko razmišljajo?«

2. Kritika problema v novi obliki

Možno je, da bi se to nekoč lahko zgodilo, vendar tudi če bi nam bila ta iznajdba na voljo, bi bil poskus narediti »razmišljajoči stroj« bolj človeški s tem, da bi ga oblekli v takšno umetno meso, nesmiseln.

Nočemo kaznovati stroja zaradi nezmožnosti, da bi blestel na lepotnih tekmovanjih, niti nočemo kaznovati človeka za poraz v dirki z letalom.

Ob tem se lahko vprašamo, ali stroji ne morejo izvajati nečesa, kar bi pravzaprav lahko opisali kot razmišljanje, četudi je zelo drugačno od tega, kar počne človek? Ta pomislek je zelo upravičen, vendar lahko odvrnemo, da nas ne bi smel vznemirjati, če kljub temu lahko sestavimo stroj, ki zadovoljivo sodeluje v igri posnemanja.

3. Stroji, vključeni v igro

Želimo tudi dopustiti možnost, da bi inženir ali skupina inženirjev izdelala *stroj, ki deluje, vendar ga njegovi izdelovalci ne bi mogli zadovoljivo opisati*, ker so uporabili pristope, ki še niso dovolj preizkušeni.

Nekdo bi morda lahko vztrajal, da morajo skupino inženirjev sestavljati izključno osebe enega spola, vendar to ne bi bilo povsem zadovoljivo, saj je verjetno mogoče vzgojiti popolnega posameznika tudi iz ene same kožne celice, recimo kožne celice moškega. To bi bil dosežek biotehnike, ki bi si zaslužil najvišjo pohvalo, vendar ga še ne bi smatrali kot primer »*izdelave razmišljajočega stroja*«.

Kratek odgovor na to je, da ne sprašujemo, ali bi se vsi digitalni računalniki dobro odrezali v igri. Niti ne sprašujemo, ali bi se v igri dobro odrezali računalniki, ki so trenutno na voljo. *Sprašujemo, ali si lahko predstavljamo računalnike, ki bi se lahko zelo dobro odrezali v igri.*

4. Digitalni računalniki

Zamisel za digitalni računalnik lahko pojasnimo z izjavo, da je to stroj, namenjen izvajanju postopkov, ki bi jih lahko izvedel človeški računalnik.

Navodila te vrste so zelo pomembna, ker omogočajo, da se zaporedje postopkov ponavlja, dokler ni izpolnjen določen pogoj, pri čemer se pri vsaki ponovitvi brez spremembe upoštevajo ista navodila.

Zanimiva različica zamisli za digitalni računalnik je »digitalni računalnik z zmogljivostjo za naključne izbire«. Včasih tak stroj opisujejo celo kot napravo s *svobodno voljo* (sam tega izraza ne bi uporabil).

5. Univerzalnost digitalnih računalnikov

Strogo gledano, takšnih strojev pravzaprav ni. Vse se dejansko premika neprekinjeno ali zvezno. Vendar obstaja veliko vrst strojev, ki jih lahko priročno obravnavamo kot stopnjevite stroje z jasno ločenimi stanji.

Sistem »vesolja kot celote« je takšen, da lahko že zelo majhne napake v začetnih pogojih močno vplivajo na kasnejša stanja. Premik enega samega elektrona za milijardinko centimetra bi lahko pomenil razliko med tem, da bi človeka leto kasneje zasul plaz in tem, da bi se uspel rešiti. Bistvena lastnost mehanskih sistemov, ki jih imenujemo »stopnjeviti stroji«, je, da pri njih to ni mogoče.

Vse postopke lahko izvedemo z istim digitalnim računalnikom in ustrezno programsko opremo. Posledično so vsi digitalni računalniki v nekem smislu enakovredni.

6. Nasprotna mnenja

Verjamem, da bo čez približno petdeset let mogoče programirati računalnike s pomnilno zmogljivostjo okoli 10^9 (1 gigabit = 125 megabajtov) tako dobro, da bodo igro posnemanja igrali že zelo uspešno. Povprečni spraševalec takrat ne bo imel več kot 70 odstotkov možnosti, da po petih minutah spraševanja pravilno loči med človekom in strojem.

Prvotno vprašanje »Ali stroji lahko razmišljajo?« se mi zdi preveč nesmiselno, da bi si zaslužilo resno obravnavo.

Splošno mnenje, da znanstveniki neizogibno napredujejo od enega dobro uveljavljenega dejstva do drugega dobro uveljavljenega dejstva, brez vpliva kakršnihkoli nepotrjenih domnev, je povsem napačno. Če je jasno, katera so dokazana dejstva in katere so uporabljene domneve, ne more nastati škoda. Domneve so izjemno pomembne, ker predlagajo koristne nove smeri raziskav.

(1) Teološki pomislek

Po mojem mnenju je večja razlika med živim in neživim kot med človekom in drugimi živalmi.

Vendar, ali ne bi morali verjeti, da ima Vsemogočni svobodo podeliti dušo slonu, če tako želi? Seveda bi pričakovali, da bi to moč uporabil na način naravne mutacije, ki bi slonu zagotovila možgane, ustrezno izboljšane za izpolnitev tega namena. Povsem enak zagovor bi lahko uporabili v primeru strojev.

S poskusom izdelave inteligentnih strojev si ne bi človek nič bolj nespoštljivo in nasilno prilastil Božje moči in svobode za ustvarjanje duš, kot pa pri naravnem razmnoževanju otrok; prej smo v obeh primerih le orodja Božje volje in sredstva za zagotavljanje prebivališč dušam, ki jih ustvarja.

(2) Pomisleki, ki izhajajo iz brezbržnosti

»Posledice obstoja mislečih strojev bi bile preveč grozne. Zato upamo in verjamemo, da tega v resnici ne zmorejo.«

Radi verjamemo, da je človek na nek prefinjen način nadrejen nad preostalim stvarstvom.

To lažje dosežemo s tolažbo, ki bi jo morda morali poiskati v zamisli kot je selitev duš.

(3) Matematični pomisleki

Najbolj znan med temi izsledki je Gödelov izrek (1931), ki pravi, da je v dovolj močnem logičnem sistemu vedno mogoče oblikovati izjave, ki jih v okviru tega sistema ni mogoče niti dokazati niti ovreči, razen če je sistem že sam po sebi neskladen.

Če je prilagojen tako, da daje odgovore na vprašanja kot v igri posnemanja, bodo vedno obstajala nekatera druga vprašanja, na katera bo bodisi podal napačen odgovor, bodisi sploh ne bo podal odgovora, ne glede na to, koliko časa mu damo na razpolago.

Skratka, morda obstajajo ljudje, ki so pametnejši od vsakega posameznega stroja, vendar lahko obstajajo še drugi stroji, ki so pametnejši in tako naprej.

(4) Pomisleki, ki izhajajo iz problema zavesti

»Šele, ko bi stroj lahko napisal sonet ali sestavil koncert zaradi misli in čustev, ki jih čuti sam v sebi, in ne zaradi nekega naključnega izbora znakov, bi se lahko strinjali, da je tak stroj enakovreden možganom. Tak stroj torej ne le piše, ampak tudi ve, da je zapisal.«

»Dejansko je to zelo skrajn solipsističen pogled. Morda je to tudi najbolj logičen pogled, vendar otežuje posredovanje razumskih in znanstvenih idej.«

»Vendar mislim, da teh skrivnosti ni nujno potrebno rešiti, preden lahko odgovorimo na vprašanje, ki smo si ga zastavili v tem članku.«

(5) Pomisleki, ki izhajajo iz omejitev ali prikrajšanosti strojev

Običajno nobena izmed teh trditev ni podprta z dejstvi. Verjamem, da so večinoma utemeljene na sklepih preprostega znanstvenega posploševanja ali indukcije.

Zdi se mi, da ta pomislek temelji na neustreznem razlikovanju dveh vrst napak. Lahko jih imenujemo »napake delovanja« in »napake sklepanja«. Abstraktni stroji so matematične izmišljotine in ne fizični predmeti. Po opredelitvi niso sposobni napraviti napak v delovanju. V tem smislu lahko resnično rečemo, da »stroji nikoli ne delajo napak«. Napake sklepanja pa lahko nastanejo le, kadar ima rezultat stroja neko vsebino oziroma pomen. Trditev, da stroj takšne napake ne more narediti, je očitno neutemeljena.

V tem smislu lahko stroj nedvomno postane predmet svojega lastnega razmišljanja. Lahko se na primer uporablja za pomoč pri izdelavi lastnih programov ali pa za napovedovanje učinkov, ki bi jih imele spremembe njegovega lastnega ustroja. Z opazovanjem rezultatov svojega vedenja lahko svoje programe prilagodi tako, da naslednjič učinkoviteje dosežejo določen cilj. To so dejanske možnosti za bližnjo prihodnost in ne le utopične sanje.

(6) Pomislek Lady Lovelace

To ne pomeni, da ni mogoče izdelati elektronske opreme, ki bo 'mislila sama zase', ali, če uporabimo izraz iz biologije, takšne v kateri bi lahko vzpostavili pogojni refleks, ki bi služil kot osnova za 'učenje.'

Stroji me pogosto presenetijo.

Pogled, da stroji ne morejo izzvati pristnega presenečenja, je po mojem mnenju posledica zmotnega sklepanja, ki so mu še posebej podvrženi filozofi in matematiki. Temelji namreč na predpostavki, da v trenutku, ko je določeno dejstvo zajeto ali predstavljeno v misli, istočasno postanejo dostopne vse posledice tega dejstva. Ta predpostavka je sicer lahko koristna v številnih okoliščinah, vendar prehitro pozabimo, da je pravzaprav napačna.

(7) Pomislek, ki izhaja iz zveznosti živčnega sistema

Človeški živčni sistem zagotovo ni stopnjeviti stroj.

Zveznost živčnega sistema bi lahko navedli kot resen pomislek. Pod vprašaj namreč postavlja možnost, da bi stopnjeviti stroj lahko posnemal delovanje živčnega sistema in dejavnost, kakršna je razmišljanje.

Digitalni računalnik ne more natančno napovedati, kakšne odgovore bo na posamezne probleme podal diferencialni analizator, vendar povsem zadovoljivo lahko poda ustrezno vrsto odgovora.

(8) Pomisleki, ki izhajajo iz nedoločnosti pravil vedenja

Res pa je tudi, da se ne moremo zlahka prepričati, da ni zaključenih zakonov obnašanja, ki bi zajela vse posamezne primere obnašanja, kot se lahko prepričamo o tem, da ni zaključenih pravil ravnanja, za vsako mogočo situacijo.

Edini način, ki ga poznamo za iskanje takšnih zakonov, je znanstveno opazovanje in zagotovo ni okoliščin, v katerih bi lahko upravičeno dejali: »Dovolj smo iskali. Takšnih zakonov ni.«

Ni mogoče, da bi se lahko iz teh odgovorov kdorkoli naučil dovolj o programu, da bi lahko napovedal odgovore na še nepreizkušene vrednosti.

(9) Pomisleki, ki izhajajo iz možnosti izvenčutnih zaznav

Žal pa so statistični dokazi vsaj za telepatijo prepričljivi. Ko jih enkrat sprejmete, je le še nekaj korakov do verovanja v duhove in bogove.

Če so možne izvenčutne zaznave, se lahko zgodi karkoli.

Zahteve igre posnemanja bi izpolnili tako, da bi udeležence postavili v ločene »sobe, odporne na telepatijo«.

7. Strojno učenje

Če ostanemo pri tej prisposodbi, lahko postavimo vprašanje: »Ali je mogoče izdelati stroj z zmogljivostjo 'nad kritično maso'?«

Namesto da bi poskušali sestaviti program, ki oponaša um odraslega, zakaj ne bi raje poskusili sestaviti takšnega, ki oponaša um otroka? Če bi tak program nato izpostavili ustreznemu učnemu postopku, bi ga lahko vzgojili podobno kot otroške možgane.

Stališče, da »stroj lahko naredi le to, kar mu znamo predpisati z jasnimi navodili za izvedbo,« se v tej luči zdi nenavadno. Večina programov, ki jih lahko vstavimo v stroj, bo namreč povzročila, da bo stroj počel stvari, ki jih sploh ne moremo razumeti ali pa jih smatramo za povsem naključne.

*

Turing tu zelo jasno izrazi svoje prepričanje, da bodo elektronski računalniki lahko »zelo natančno posnemali dejavnosti človeškega uma«.

Turing opozarja, da je treba iznajti učinkovit način za ocenjevanje in določiti jasna merila za ustrezno in neustrezno delovanje stroja, če želimo z uvajanjem nagrad in kazni pospešiti razvoj umetne inteligence.

Predlaga, da se računalnik že s preprostim beleženjem lastnih stanj lahko začne spominjati in da lahko z obravnavo značilnosti teh stanj začne razmišljati o samem sebi. To pa sta že dve bistveni znamenji inteligence.

Predavanje družbi v Manchesteru (1951)

Menim, da lahko izdelamo stroje, ki bodo zelo natančno posnemali dejavnost človeškega uma. Občasno bodo delali napake, občasno pa bodo prišli do novih, zelo zanimivih zaključkov. Na splošno bodo njihovi rezultati in ugotovitve vredni pozornosti ravno toliko kot rezultati in ugotovitve človeškega uma.

Brez koncepta, ki ustreza nekakšnemu psihološkemu »načelu užitka«, si ne moremo predstavljati napredka na tem področju. Zagotovo bi bilo nekaj takega zelo smiselno vpeljati v delovanje stroja. Predlagam da uporabimo dva ključa, ki predstavljata užitek in bolečino.

Zelo verjetno bo metoda strojnega razmišljanja hitro presegla naše omejene zmožnosti. Lahko pričakujemo, da bodo stroji na neki točki prevzeli nadzor, podobno, kot je to opisal Samuel Butler v delu »Erewhon«.

*

Turing meni, da je elektronski digitalni računalnik med vsemi obstoječimi stroji najbolj vreden primerjave s človeškimi ali živalskimi možgani.

Izrecno izpostavi problem zaostanka na področju pomnilnikov in programerskih pristopov, ki so ključni za razvoj računalništva in umetne inteligence.

Ob koncu predavanja lepo nakaže, da raziskave umetne inteligence nimajo samo poseben tehnološki, temveč tudi širši, antropološki pomen, saj zahtevajo boljše razumevanje naše lastne telesne, umske in čustvene narave.

Predavanje za tretji program BBC (15. maja 1951)

Mislim tudi, da: »če dejansko obstaja stroj, ki ga lahko primerjamo z možgani, potem je to ravno digitalni računalnik.«

Mislim, da je povsem verjetno, da bomo konec stoletja stroje programirali že tako dobro, da bodo na primer odgovarjali na vprašanja in bo skoraj nemogoče ugotoviti, ali odgovore podaja človek ali stroj.

Veseli bi morali biti, ko nas stroj preseneti, podobno kot smo veseli, ko učenec stori nekaj, česar ga nismo izrecno naučili.

*

Precej razpravljajo predvsem o tem, kaj pravzaprav pomeni razmišljati in ali lahko trdimo, da tudi elektronski računalniki počnejo nekaj podobnega razmišljanju.

Dotaknejo se tudi sposobnosti učenja, posploševanja, povezovanja in ocenjevanja zamisli, sposobnosti logičnega sklepanja in podobno. Po njihovem mnenju so vse to značilnosti, ki jih lahko upravičeno pričakujemo od umetne inteligence.

Precej razpravljajo o možnostih, da bi stroji čutili in želeli. Zelo pa jih zanima tudi možnost, da bi stroji delovali in se odzivali v skladu s svobodno, individualno voljo.

Pogovor za tretji program BBC (10. januar 1952)

TURING: Zamisel za preizkus je, da se stroj z odgovarjanjem na vprašanja pretvarja, da je človek. Uspel bo le, če se bo pretvarjal dovolj prepričljivo.

TURING: Takoj, ko lahko v možganih prepoznamo delovanje vzrokov in učinkov, se nam to ne zdi več razmišljanje, ampak nekakšno dolgočasno, nedomiselno delo. S tega vidika bi lahko opredelili razmišljanje kot miselne dejavnosti, ki jih ne razumemo. Če je to res, potem izdelati razmišljajoči stroj pomeni samo izdelati takšnega, ki bo počel zanimive in nerazumljive stvari.

NEWMAN: Če želite, lahko v uri ali dveh, sestavite program za stroj v Manchestru, ki bo preučil vse možne igre šaha od začetka do konca in na ta način našel najboljše možne poteze. Vendar bo trajalo več milijard let, da boste ta program tudi v celoti izvedli.

*

Predstavi pa celo zamisel za izdelavo celovitega robotskega posnetka človeškega bitja z elektronskimi možgani. Zaradi težavnosti te naloge predlaga, da najprej raziščemo, kaj lahko naredimo samo z elektronskimi možgani.

Turing je sklepal, da urejanje človeških možganov z vzgojo in vadbo poteka po postopkih, ki jih je mogoče dovolj natančno opisati (čtetudi niso stopnjeviti stroj) in v obliki programa prenesti na stopnjevite stroje, kot so elektronski računalniki.

Odpira pa tudi zelo zanimivo možnost poljubnih, odprtih raziskovanj, s katerimi bi stroj (podobno kot otrok, ki spoznava svet) lahko prišel do lastnih, izvirnih odkritij, spoznanj in sposobnosti, ki so bistvene za razvoj umetne inteligence.

1. Uvod

Enak problem predstavlja možnost, da bi nas nadomestila druga živalska vrsta. Tudi ta možnost nedvomno obstaja in je prav tako neprijetna kot prejšnja.

To je še ena izmed številnih nejasnih primerjav, ki izvirajo iz pomanjkanja izkušenj s stroji, ki bi morda dejansko lahko po lastni volji ustvarjali iz sebe in proizvajali raznolike stvari. Bistvo stroja naj bi bilo preprosto v tem, da, dokler deluje, vedno znova ponavlja isto stvar.

Človek pa lahko po drugi strani s svojimi umskimi zmožnostmi vedno odkrije še boljše metode za reševanje problemov in tako učinkovito »preseže« metode, ki so na voljo strojem.

2. Zavrnitev nekaterih pomislekov

Teh pomislekov ne moremo popolnoma prezreti, saj je tudi zamisel »*intelligence*« sama po sebi bolj čustvena kot matematična.

Vendar ta zahteva pravzaprav sploh ne pogojuje prisotnosti *intelligence*.

Ko igramo partijo proti takšnemu namišljenemu stroju, dobimo podoben občutek, kot bi merili svoje zmožnosti z nečim *živim*.

3. Vrste strojev

Ne moremo resno razpravljati o tem, kako bi lahko izdelali inteligentne stroje, če prej ne uvedemo primernih tehničnih izrazov za opis različnih vrst strojev.

Vse stroje lahko obravnavamo kot zvezne, vendar jih je pogosto koristno in možno obravnavati kot stopnjevite.

Možgani so verjetno zvezni upravljalni stroj, vendar so zelo podobni stopnjevitim strojem.

Različne ureditve ali konfiguracije strojev

Vendar pa je lastnost »biti stopnjevit«, le pomoč in orodje za teoretičnega raziskovalca, in ne služi evolucijskemu namenu. Zato ne moramo pričakovati, da bi nam pri tem lahko pomagala kar narava sama in bi ustvarila resnično »stopnjevite« možgane.

Če ima stroj končno število N možnih stanj, potem pravimo, da ima pomnilno zmogljivost, ki je enakovredna $\log(N)$ bitov.

Pomnilna zmogljivost stroja bolj kot karkoli drugega določa zapletenost njegovega delovanja.

Logični računski stroji

Stroj lahko prebrani znak spremeni. Prebrani znak določa in vpliva na delovanje stroja, medtem ko drugi znaki, zapisani drugje na traku, v danem trenutku ne vplivajo na njegovo delovanje.

Pomen univerzalnega stroja je jasn. Ne potrebujemo neskončnega števila različnih strojev za različne naloge. *Zadostuje en sam stroj*. Namesto inženirskega problema izdelave različnih strojev za različne naloge nas čaka pisarniški problem »programiranja« univerzalnega stroja za izvajanje teh nalog.

Iz praktičnih izkušenj vidimo, da logični računski stroji lahko storijo vse, kar je mogoče opisati »vsaj približno« ali »povsem mehansko«. To dejstvo je že dovolj potrjeno, da med logiki obstaja soglasje o pravilnosti in primernosti uporabe izraza »izračunljiv z logičnim računskim strojem«, ko govorimo o takšnih problemih.

Praktični računski stroji

Če določen problem, ki ga rešujemo, zahteva več kot 10170 korakov, skoraj z gotovostjo lahko pričakujemo napačen odgovor, zato je pametno, da naloge omejimo na manjše število korakov.

Še bolj uporabne omejitve pa bi lahko dobili, če bi predpostavili, da mora med posameznimi koraki svetlobni val prepotovati vsaj 1 cm (to bi bil zelo majhen stroj) in da na odgovor lahko čakamo največ 100 let. Na ta način bi dobili omejitev največ 1020 korakov.

S pomočjo sistema, ki nekoliko spominja na telefonsko centralo, je na primer mogoče zelo hitro pridobiti določen podatek s »klicanjem« pomnilnega mesta, kjer je shranjen. Zamuda je v tem primeru lahko le nekaj mikrosekund.

Univerzalni praktični računski stroji

S tem ne mislim, da lahko šele izdelamo digitalni računalnik, ki bo zahtevane naloge izvedel, temveč, da lahko izvedemo vse te naloge zgolj z ustreznim programiranjem enega izmed digitalnih računalnikov (na primer računalnika ACE). Programiranje pa je čisto pisarniško delo.

Ta postopek lahko širimo na vse mogoče načine, vendar bomo vedno znova soočeni z nalogo, da nekje shranimo to pozitivno celo število nedoločene velikosti. Edina oziroma najboljša rešitev je verjetno, da za ta namen uvedemo poseben »trak«.

Če to storimo, pa v naši nadaljnji obravnavi brez težav lahko zavržemo vse druge oblike pomnilnika, saj poskušamo dokazati le zamisel in ne uporabno izvedbo.

Papirni stroji

Učinek računskega stroja lahko zaenkrat ustvarimo tako, da zapišemo določen nabor pravil in postopkov, ki jih mora nato izvesti človek.

To pripravo lahko imenujemo »*papirni stroj*«.

Človek je pravzaprav univerzalni stroj, če ga podvržemo strogim pravilom ravnanja in ga opremimo s papirjem, svinčnikom in radirko.

Delno naključni in navidezno delno naključni računski stroji

Opisane vrste stopnjevutih strojev lahko spremenimo tako, da na določenih mestih omogočimo uporabo več možnih postopkov, pri čemer stroj postopek izbere naključno. Tak stroj lahko opišemo kot »delno naključen«.

Če želimo posebej izpostaviti, da določen stroj ni delno naključen, ga lahko opišemo kot »strogo določen stroj«.

Včasih je stroj lahko strogo gledano, določen, vendar se na videz še vedno zdi delno naključen. To bi se zgodilo, če bi na primer uporabili številke poljubnega števila za določitev izbir delno naključnega stroja. Tak stroj lahko opišemo kot »navidezno delno naključen«.

4. Neurejeni stroji

Gibanje neurejenega stroja prvega reda z N enotami sčasoma seveda postane ponavljajoče, tako kot pri vsakem določenem stroju s končno pomnilno zmogljivostjo.

Neurejeni stroji prvega reda so zanimivi, ker jih lahko uporabimo kot preprosto ponazoritev živčnega sistema z naključno razporeditvijo živčnih celic.

Na to odgovorjam: če smo opredelitve podali pravilno (četudi suhoparno), ne nastopajo le kot opredelitve tega, kateri stroji pripadajo prvemu in kateri drugemu redu, temveč kot ocene verjetnosti, da določen stroj prvega ali drugega reda pripada določeni ureditvi. Če naključno izberete stroj prvega reda z določenim številom enot, je zelo malo verjetno, da boste izbrali stroj drugega reda.

5. Posegi v delovanje strojev: prilagodljivi in samo-prilagodljivi stroji

Obstaja skrajna oblika, pri kateri odstranimo dele stroja in jih zamenjamo z drugimi. To lahko opišemo kot »*poseg z izvijačem*«. Na drugem koncu lestvice je »*poseg s papirjem*«.

Kadar lahko temeljito spremenimo delovanje stroja, govorimo o »*prilagodljivem stroju*«. En stroj je lahko bolj prilagodljiv kot drug stroj. Včasih se lahko stroj sam spreminja ali pa spreminja svoja lastna navodila.

Kadarkoli notranji postopki stroja spremenijo vsebino pomnilnika, lahko seveda govorimo o stroju, ki se »*spreminja sam*«.

6. Človek kot stroj

Elektronska vezja, ki jih uporabljamo v računskih strojih, kot kaže, že imajo bistvene lastnosti živcev. Podatke lahko prenašajo iz enega mesta na drugo in jih shranjujejo. V primerjavi z živci imajo elektronska vezja le eno privlačno prednost, *hitrost*. Ta prednost pa je tako bistvena, da morda celo pretehta prednosti živcev.

Če bi želeli, da stroj sam odkriva stvari, bi mu morali omogočiti svobodno gibanje. To pa bi predstavljalo resno grožnjo za navadne državljane. Kljub temu, da je ta pristop za izdelavo inteligentnega stroja verjetno »zanesljiv«, se zdi preprosto *prepočasen in preveč nepraktičen*.

Pravila, s katerimi je zakodirano sporočilo, ustrezajo zakonom vesolja, prestrežena sporočila ustrezajo dokazom, ki so na voljo fiziku, ključi za razbijanje sporočil pa ustrezajo pomembnim konstantam, ki jih mora ta določiti. *Ujemanje je res veliko*, vendar lahko kriptografske probleme s stopnjevitimi stroji obravnavamo zelo enostavno, medtem ko fizikalnih problemov še ne moremo.

7. Strojno učenje

Precej nepošteno je pričakovati, da bi se stroj, ki je pravkar prišel iz tovarne, lahko enakopravno kosal z univerzitetnim diplomantom.

Lahko rečemo, da je človek kot stroj (kolikor ga lahko primerjamo s strojem), ki je podvržen številnim posegom. Pravzaprav je poseg pri človeku pravilo in ne izjema.

Če želimo izdelati inteligenten stroj in natančno (kolikor je to mogoče) sledimo človeškemu zgledu, bi morali začeti s strojem, ki ima malo zmožnosti za izvajanje zapletenih postopkov in urejeno odzivanje na ukaze (v obliki posegov). Nato bi poskušali s pomočjo ustreznih posegov, ki bi posnemali pouk, spremeniti stroj tako, da bi se lahko vedno bolj zanesli na njegove odzive.

8. Načini urejanja neurejenih strojev

Pri neurejenih strojih lahko postopno dosežemo takšno ureditev, da delujejo kot urejeni stroji za določen namen, vendar je v tem primeru potrebno vse posege od zunaj ustrezno omejiti, ko takšno ureditev enkrat dosežemo.

Primer prikazuje splošno značilnost strojev drugega reda, da lahko z ustreznimi začetnimi pogoji opravijo vsako zahtevano nalogo, če imajo na voljo dovolj časa in zadostno število enot. Zlasti pri neurejenih strojih drugega reda z dovolj enotami so možni takšni začetni pogoji, da postopno lahko dosežemo ureditev univerzalnega stroja z dano pomnilno zmogljivostjo.

Postopek vzpostavljanja začetnih pogojev, ki omogočijo, da stroj lahko opravi posebno, koristno nalogo, lahko imenujemo »urejanje stroja«. Urejanje je torej oblika »spremembe«.

9. Možganska skorja kot neurejen stroj

Razlike med jeziki, ki jih govorimo, niso posledica razlik v razvoju francosko govorečih in angleško govorečih delov možganov. So posledica tega, da so jezikovni deli enih in drugih možganov podvrženi različnim oblikam usposabljanja.

Vse to nakazuje, da je možganska skorja dojenčka kot neurejen stroj, ki ga je mogoče urediti z ustreznimi posegi ali usposabljanjem. Z urejanjem bi morda lahko povzročili spremembo takšnega neurejenega stroja v univerzalni stroj ali nekaj podobnega. Tak človek pa ne bi imel zdrave pameti in bi brez omahovanja upošteval tudi najbolj nesmiselne ukaze.

Posest možganske skorje (recimo) je sama po sebi neuporabna, če ne obstajajo mehanizmi in postopki, ki jo lahko ustrezno uredijo. Če bi volk z mutacijo pridobil organ primerljiv človeški možganski skorji, še ne bi imel prednosti pred drugimi živalmi.

10. Poskusi urejanja: mehanizmi užitka in bolečine

Učenje otroka je v veliki meri odvisno od uvajanja nagrad in kazni, kar nakazuje, da bi urejanje stroja morda lahko izvajali zgolj z dvema vhodoma: enim za uvajanje »užitka« ali »nagrade« (R ali spodbujevalni red) in drugim za uvajanje »bolečine« ali »kazni« (P ali omejevalni red).

Vpliv spodbujevalnega dražljaja (užitka) okrepi značaj, saj preprečuje njegovo spreminjanje, medtem ko vpliv preprečevalnega dražljaja (bolečine) zmoti značaj, saj spodbuja njegovo spreminjanje in omogoči, da so določene značilnosti značaja ponovno podvržene naključnim spremembam.

To pomeni, da bo učitelj s preudarnim uvajanjem spodbujevalnih in preprečevalnih posegov lahko dosegel, da se bo »značaj« približal želenemu in bo napačno delovanje stroja vedno manj pogosto.

11. Neurejeni stroji omejevalnega reda (P)

Ko s posegi dosežemo ureditev stroja, ki je ne opredeljujejo ustrezna navodila, stroj lahko naključno izbere navodila, jih začasno vnese v svoj opis in jih poskusno uporabi. Če se pojavi preprečevalni dražljaj (bolečina), se vsi začasni vnosi prekličejo. Če se pojavi spodbujevalni dražljaj (užitek), vsi začasni vnosi postanejo trajni.

V prikaz delovanja sem na opisan način (papirni) stroj uspešno uredil v univerzalni stroj.

Upal sem tudi, da bom morda lahko našel nekaj takšnega, kot so »nepravilni glagoli«, ki bogatijo jezik. Zdi se, da smo ljudje povsem zadovoljni, če stvari niso posebej strogo pogojene s splošnimi matematičnimi pravili. Sumim, da bi se neurejen stroj omejevalnega reda brez sistematičnega pomnilnika obnašal na precej podoben način zaradi naključno razporejenih pomnilnih mest.

12. Omejitve in spodbude

Pretvoriti možgane ali stroj v univerzalni stroj je najskrajnejša oblika omejevanja ali discipline. Toda omejevanje samo po sebi zagotovo ne zadošča za nastanek inteligence. Kar preostane, imenujemo spodbujanje ali spodbuda.

Ko bi ta pravila postala dovolj splošna, posegi ne bi bili več potrebni, in stroj bi »odrastel«. To lahko imenujemo »neposreden pristop«.

Drug pristop bi bil takšen, da bi začeli s povsem neurejenim strojem in bi poskusili vanj hkrati vnesti tako omejitve kot spodbude. Namesto da bi ga takoj poskušali z omejitvami preurediti v univerzalni stroj, bi ga lahko uredili tako, da bi se bolj odzival na spodbude. Mislim, da bi morali preizkusiti oba pristopa.

Intelektualna, genetska in kulturna vprašanja

Verjetno se ne bi zelo zmotili, če bi domnevali, da vse probleme lahko skrajšamo na to obliko. Ponovno bomo morali razmisliti o tem šele, ko se bo pojavil problem, ki očitno ne bo ustrezal tej obliki.

Morda jih lahko imenujemo »intelektualna iskanja«. Na kratko bi jih lahko opredelili kot: »iskanja, ki jih izvajajo možgani, kadar obravnavajo sestavljene pojave z določenimi, posebnimi značilnostmi«.

S tega vidika je morda treba iskanje novih življenjskih pristopov prej obravnavati kot vrsto iskanja, ki ga izvaja človeška skupnost kot celota in ne le posamezniki sami zase.

13. Inteligenca kot čustveni koncept

Obseg pojavov, ki jih razumemo kot izraze inteligentnega vedenja, ravno toliko določa naše lastno umsko stanje in njegova usposobljenost, kot dejanske lastnosti pojava, ki ga obravnavamo.

Isti pojav nekdo torej lahko obravnava kot izraz inteligence, nekdo drug pa morda ne, ker je odkril pravila in vzroke določenega vedenja.

C bo zelo težko ugotovil ali igra s človekom ali s »papirnim strojem«.

*

Želela sta doseči, da bi elektronski računalnik odigral zadovoljivo partijo šaha, da bi reševal običajne šahovske probleme in podajal poteze v sprejemljivem času (nekaj minut).

Kot vsi kasnejši šahovski programi se ukvarja predvsem s postopki napovedovanja naslednjih nekaj potez (lastnih in nasprotnikovih) ter s postopki izbire najbolj primernih potez.

Prispevek predstavlja natančna in jasna merila za ocenjevanje položajev in potez, nato pa podaja primer partije, ki jo je Turing s strogo uporabo predstavljenih meril in postopkov ročno odigral (tako kot bi jo odigral računalnik) proti znancu Alicku Glenniju.

Prispevek za knjigo »Faster than thought« (1953)

Če lahko v angleščini in po potrebi s pomočjo matematičnih simbolov povsem nedvoumno razložimo, kako je potrebno izvesti izračune za igranje šaha, potem je vedno mogoče programirati poljubni digitalni računalnik, da te izračune izvede, pod pogojem seveda, da je pomnilna zmogljivost računalnika zadostna.

Na vprašanje 6 bi lahko odgovorili: »Nikoli ne bomo z gotovostjo vedeli, če je to mogoče, tako kot nikoli ne bomo povsem prepričani, da tudi drugi ljudje čutijo tako kot mi sami.«

To dejstvo bi lahko obravnavali kot potrditev precej lahkotnega pogleda, da: »ne moremo programirati stroja, da bo odigral boljšo partijo šaha, kot jo lahko odigramo sami.« To izjavo bi morali primerjati z drugo, podobno, da: »nobena žival ne more pogoltniti druge živali, ki je težja od nje.« Obe izjavi sta, kolikor vem, neresnični.

*

Dejstvo, da je prvi šahovski program nastal že nekaj let pred izumom elektronskega računalnika, je zares neverjetno. Oblikoval ga je človek, ki je videl naprej v prihodnost. Razumel je, da je izdelava računalnika le še vprašanje časa in da bodo ti stroji sposobni igrati šah.

Friedel je nato obravnaval odstopanja, s katerimi so se soočili in pojasnil verjetne razloge zanje. Najbolj pomembno je dejstvo, da je Turing delal s papirjem in svinčnikom, in je očitno uporabljal svoj občutek pri izbiranju potez. Ne da bi se zavedal, je uporabljal metodo obrezovanja Alpha-Beta, ki sta jo izumila Allen Newell in Herbert A. Simon pol desetletja za njim (takrat so stroji že dejansko lahko izvajali programe). V bistvu Turing ni poskrbel za izračun vsake možne poteze v danem položaju. Nekatere je ocenil kot očitno slabše in se mu ni zdelo smiselno, da bi porabil veliko časa za ugotavljanje in potrjevanje kako slabe v resnici so. In to je bistvo metode Alpha-Beta.

»Iščete napako v programu, Frederic? Ne, ne, iskati bi jo morali v Alanu Turingu! Alanu ni bilo dosti mar za podrobnosti; zanimal ga je splošni princip.«

