УНИВЕРЗИТЕТ У БЕОГРАДУ

ТЈУРИНГОВ ТЕСТ

Београд, 2008. год

МАТЕМАТИЧКИ ФАКУЛТЕТ

МАСТЕР ТЕМА : ТЈУРИНГОВ ТЕСТ

ментор:
проф. др Жарко Мијајловић

чланови комисије :
проф. др Милан Божић, проф. др Александар Јовановић

студент:
Славица Шантрић

Садржај

1. Увод

Идеја ефективне израчунљивости настала је релативно недавно, тридесетих година, али то није сметало да се овај појам уврсти међу најважније доприносе практичној култури модерног научног живота. Јер, увођење таквог техничког појма није довело само до значајних открића у логици, мада је то био првобитни циљ, већ се испоставило да ова теорија представља витално интелектуелно средство у изградњи модела сложених система. Вероватно се најважније промене налазе у рачунарству и рачунарима, али се, исто тако, значајне промене ове теорије могу наћи у расправљању питања из биологије, психологије, лингвистике, филозофије и наравно математике.

Главни појам ове теорије је појам алгоритма или ефективне процедуре. Најједноставније речено, под алгоритмом се подразумева ефективан поступак који примењен на класу речи над неким алфабетом (те речи називају се улазима) евентуално даје одговарајуће излазне речи (тј. излазе). Дакле, алгоритам је процедура за израчунавање неке функције. Примери таквих поступака су познати практично у свим областима математике. Споменимо неке:

- Сабирање и множење природних бројева,
- Еуклидов алгоритам за одређивање највећег заједничког делиоца два природна броја,
- Диференцирање елементарних функција,
- Решавање појединих класа диференцијалних једначина,
- Поступци за испитивање таутологичности исказних формула.

Сви ови и други поступци израчунавања имају следеће заједничке особине које се могу сматрати неопходним, да би се извесна процедура сматрала ефективном (наравно, овде се претпоставља једна идеална ситуација; тако нпр. у следећим условима реч „постоји“ има управо оно значење које има у математици):

1. Сваки алгоритам дат је као коначан низ инструкција.
2. Постоји рачунско средство које интерпретира и изводи инструкције алгоритма.
3. Постоји меморијски простор у којем се чувају (привремено или стално) сви подаци који се јављају приликом израчунавања.

4. Израчунавање по датом алгоритму је дискретне природе, дакле изводи се корак по корак и без коришћења непрекидних метода или анологних средстава.
5. Израчунавање по датом алгоритму је детерминисано, тј изводи се без коришћења случајних метода или средстава; дакле, поновљене примене алгоритма на исте улазне величине производе исте излазне величине.
6. Не постоје никаква ограничења на величину улаза, број инструкција, величину меморије, као ни на дужину рачуна који се изводи за конкретан улаз.
7. Алгоритам не мора давати резултат за све улазе, израчунавање алгоритма може, дакле, да се никад не заврши.
8. Ипак, у следећем смислу постоји граница у могућностима рачунарских средстава: испоставља се да постоји универзалан алгоритам који симулира израчунавање по сваком алгоритму.
9. Алгоритама и објеката на којима се они изводе има пребројиво много, али не и више.
10. Алгоритми, улазни и излазни симболи могу се ефективно кодирати у скупу природних бројева.

Полазећи од ових услова предложено је више формалних система у оквиру којих се дефинише и анализира математичким средствима појам ефективне израчунљивости. Дајемо преглед неких од њих.

2. Систем израчунљивости

2.1. Тјурингове машине

Тридесетих година почиње нагли развој логике. K.Gödel, A.Church, A.Tarski, S.Kleene, E.Post и други проучавају формализоване системе и из тих радова и радова Alana Turinga настаје математички појам ефективне израчунљивости. Turing конструише 1935. год. један апстрактан модел израчунљивости који по њему носи назив алгоритамски систем Turing-ових машина. Ово откриће занимљиво је јер је дало теоријски оквир за пројектовање и стварање рачунара који се могу програмирати (Von Neumann и његове колеге), као и формалних програмских језика. Дајемо један неформалан опис Тјурингових машина.

Тјурингова машина је механичко средство којем је придружен меморијски простор у виду траке која се бескрајно пружа улево и удесно. Трака је подељена на просторе једнаке величине који се називају ћелијама. Над траком се налази глава машине која се у сваком моменту налази над неком ћелијом траке. Трака може да се креће улево и удесно и то у једном кораку померање се врши највише за једну ћелију. Најзад машина врши следеће радње:

-брише садржај ћелије над којом се глава налази;
-евентуално, брише и уписује нов симбол из унапред датог алфабета A;
-врши померање траке за једну ћелију улево или удесно.

Док је активна, машина изводи само једну од набројаних операција у јединици времена, и једна таква операција назива се рачунским кораком. После сваког изведеног корака машина се налази у једном од стања из унапред датог коначног скупа стања S. Рад машине изводи се према инструкцијама из неког утврђеног коначног низа инструкција који називамо програмом. Свака инструкција изгледа овако:

(1) pXYq

где су p, q стања, тј. p, q∈ S, док је X симбол алфабета A, а Y такође симбол из проширеног алфабета A∪{L,R}, где L, R ∉ A. Инструкцијом (1) изражава се правило, да уколико се машина налази у стању p и ако је садржај ћелије коју глава испитује симбол X, онда машина уписује у ту ћелију симбол Y или врши померање траке улево или удесно, а затим прелази у стање q. Померање траке улево врши се ако је Y-L, док се померање удесно изводи ако је Y-R. Најзад, алфабет A садржи срецијалан знак, нека је то симбол 0, чије је значење следеће:

ако је X-0, онда је садржај испитиване ћелије празан;
ако је Y-0, онда се садржај испитиване ћелије брише.

Овде ћемо се ограничити на случај најједноставнијег алфабета, тј. узећемо да је A=$\{0,1\}$. Мада избор оваквог алфабета може да изгледа као велико ограничење, испоставља се да рачунске могућности такве машине нису мање од Тјурингових машина са произвољно великим алфабетом.Стања машине означићемо са $q_0,q_1,\ldots$ Међу њима разликујемо почетно стање, нека је то q_0 и завршно стање q_Z. Дакле, на почетку рада машина се налази у стању q_0, а уколико се нађе у стању q_Z онда она престаје са радом.

Ево једног примера програма за Тјурингове машине. Нека је на траци конфигурација тј. блок јединица смештен између две нуле. Треба написати програм који ће дописати две јединице с десна на тај блок и вратити главу на почетак блока. Тај програм има шест инструкција:

$$q_01Lq_0;\ q_001q_1\ ;\ q_11Lq_2;\ q_201q_3;\ q_31Rq_3;\ q_30Lq_Z\ .$$

Израчунавање ро овом програму приказано је у низу конфигурација, одакле се види, у сваком кораку, садржај траке као и стање у којем се машина налази.

Најпре уводимо унарну репрезентацију природних бројева. Сваки природан број n представљен је на траци блоком од n+1 јединица. Лево и десно од блока налазе се празне ћелије. На пример, број 0 представљен је блоком од једне јединице, док је број 4 представљен блоком од 5 јединица.

За аритметичку функцију f(x) кажемо да је Turing–израчунљива уколико постоји програм P за Тјурингове машине који израчунава вредност функције f на следећи начин:

у почетном стању q глава машине налази се на првој ћелији блока јединица којим је у унарној нотацији представљена улазна вредност x, аргумент x, аргумент функције f; по извршеном програму P, глава машине налази се над првом ћелијом блока јединица који представља вредност f(x), такође у унарној нотацији.

Програм (1) израчунава вредност функције f(x)=x+2. Дакле, ова функција је Turing–израчунљива

Функције са више аргумената се израчунавају на следећи* начин као функције једног аргумента. У Turing–израчунљиве функције спадају разне аритметичке функције: сабирање и множење природних бројева, полиномне функције, низ простих бројева, затим карактеристичне функције скупова: парних бројева, кôдова формула Pean–ове аритметике, кôдова доказа u Pean–овој аритметици.

Ево програма за израчунавање збира природних бројева:

$$q_0 1 L q_0;\ q_0 0 1 q_1;\ q_1 1 R q_1;\ q_1 0 L q_2;\ q_2 1 0 q_3;\ q_3 0 L q_z.$$

Очигледно је да инструкција има највише пребројиво много, па како су програми за Тјурингове машине коначни низови инструкција, следи да и програма има тачно пребројиво много. Отуда следи да Turing–израчунљивих функција има не више од пребројиво много. Подсетимо се, такође, да аритметичких функција има континуум много.

Turing–израчунљиве функције деле се на тоталне и парцијалне. Код тоталних функција вредност функција дефинисане су за све вредности аргумената, док код парцијалних то не мора да буде случај. На пример, један програм за израчунавање празне функције има само једну инструкцију:

$$q_0 1 1 q_0.$$

Према овом програму (инструкцији) машина никад не престаје да ради.

Да бисмо увели универзалну Тјурингову машину, најпре се мора увести неко кодирање Тјурингових машина. Уколико се ограничимо на алфабет {0,1}, једна могућност је ова: кодови симбола 0, 1, L, R су редом 1, 2, 3, 4, док је код инструкције

$$I = q_i X Y q_j;$$

$$k(I) = 2^i 3^{k(x)} 5^{k(y)} 7^j,$$

где су $k(X)$, q_i, $k(Y)$ кодови симбола X, Y. Ако је $P = I_1 I_2 \ldots I_n$ један програм, можемо узети да је код програма P:

$$\ulcorner P \urcorner = 2^{k(I_1)} 3^{k(I_2)} \ldots p_n^{k(I_n)}.$$

где је 2, 3, ..., p_n, ... низ простих бројева. Најзад све програме можемо уредити овако:

$$P \prec Q \text{ акко } \ulcorner P \urcorner < \ulcorner Q \urcorner,$$

ра се у том уређају* сви програми могу ефективно набројати у низ:

$$P_0 \prec P_1 \prec P_2 \ldots$$

Тада се у P_n број n назива индексом програма P. Очигледно је да се за сваку Тјурингову машину може ефективно одредити њен индекс, као и да се

за сваки природан број n такође може ефективno конструисати Тјурингова машина Pn.

Уведимо и ову нотацију: ако је P програм и x природан број, тада $P(x)\downarrow y$ значи „програм P за улаз x у коначно много корака даје излаз y“, и тада кажемо да $P(x)$ конвергира ка y. Симбол $P(x)\downarrow$ значи да P за улаз x даје неки излаз, док $P(x)\uparrow$ означава да није $P(x)\downarrow$, и тада кажемо да P дивергира за улаз x. Слична је нотација у случају више аргумената.

Род универзалном Тјуринговом машином подразумевамо програм U са особином:

$$P_n(x)\downarrow y \Rightarrow U(n,x)\downarrow y$$
$$P_n(x)\uparrow \Rightarrow U(n,x)\uparrow.$$

Важи следеће тврђење које оправдава услов 8 из листе особина ефективне израчунљивости:

***Теорема**: Постоји универзална Тјурингова машина.*

Ако је $\varphi(x_1, \ldots, x_n)$ неки аритметички предикат, онда кажемо да је φ одлучив уколико је његова карактеристична функција:

$$k_\varphi(x_1, \ldots, x_n) = \begin{cases} 1 & \text{ako je } \varphi(x_1, \ldots, x_n) \\ 0 & \text{ako nije } \varphi(x_1, \ldots, x_n) \end{cases}$$

Turing–израчунљива. Уколико се неки проблем може формализовати или превести на неки аритметички предикат, тада кажемо да је тај проблем одлучив ако је одговарајућа карактеристична функција Turing–израчунљива. На пример, да ли је задовољено:

„n је паран број“,
„n је прост број“.
„полином p(x) са рационалним коефицијентима има рационалне корене“,
„низ симбола s је формула формалне аритметике“

су одлучиви проблеми, јер су одговарајући аритметички предикати одлучиви.

Сада наводимо пример јердног неодлучивог проблема. Проблем заустављања (Halting problem). Нека је k каректеристична* функција предиката $U(n,x)\downarrow$ где је U универзална Тјурингова машина. Тада важи:

$$U(n,x)\downarrow \Rightarrow k(n,x)= 1$$
$$U(n,x)\uparrow \Rightarrow k(n,x)= 0.$$

Доказујемо да k није Turing–израчунљива функција. Претпоставимо супротно. Тада је и функција h дефинисана овако:

$$h(m,x)=\begin{cases}\uparrow & \text{ako je } k(m,x)=1\\ 1 & \text{ako je } k(m,x)=0.\end{cases}$$

такође Turing–израчунљива, јер се на основу програма за функцију k лако конструише и програм за функцију h. Нека је P_m програм који израчунава функцију h. Тада на основу особина програма U и функција k и h налазимо:

$$U(m,m)\downarrow \Rightarrow k(m,m)=1\Rightarrow h(m,m)\uparrow$$
$$U(m,m)\uparrow \Rightarrow k(m,m)=0\Rightarrow h(m,m)\downarrow,$$

што у оба сличаја даје противречност. Дакле, функција k није израчунљива па ни проблем заустављања везан за предикат

$$U(n,x)\downarrow$$

чије је значење „Тјурингова машина са индексом n за улаз x престаје са радом после коначно много корака“, није одлучив.

Овај проблем има важну улогу у теорији ефективне израчунљивости, јер се неодлучивост многих других проблема доказује на основу неодлучивости проблема заустављања.

Наравно, у претходном „негативном“ решењу проблема заустављања, појам одлучивости прецизно је одређен. Доказати да је неки проблем одлучив, у овом контексту значи конструисати одређену Тјурингову машину, док се доказ да неки проблем није одлучив своди на доказ да једна одређена функција (карактеристична функција проблема) није Turing–израчунљива. С тим у вези, поставља се једно питање које није додуше у потпуности математичког карактера: Да ли је израчунљивост у интуитивном смислу обухваћена системом израчунљивости Тјурингових машина? Одговор донекле даје :

2.2. Cherch–ova теза.

Черчова теза: Класа интуитивно израчунљивих функција поклапа се са класом Turing–израчунљивих функција.

Дакле, ова теза тврди да је системом Turing–израчунљивих функција обухваћен у потпуности појам ефективне израчунљивости. Као што ово питање није у потпуности математичког карактера, ни сама теза није сасвим математичког карактера. Потврда за тезу налази се у чињеници да до данас ниједан систем израчунљивости није дао ширу класу израчунљивих функција од класе Turing–израчунљивих функција.

Најзад споменимо да се систем Тјурингових машина може увести на строжијем математичком језику. Наиме, ако је A={0,1}, за проширен алфабет можемо узети скуп {0, 1, 2, 3} где су симболи L и R редом замењени бројевима 2,3. За стања можемо узети само природне бројеве, па се програми могу дефинисатри као пресликавања коначних подскупова скупа NxA у скуп {0, 1, 2, 3} x N.

Машине које учествују у игри:

Питање које смо поставили у првом поглављу неће бити јасно док ми тачно не појаснимо шта подразумевемо под речју машина. Природно је што желимо да дозволимо употребу свих инжењерских техника у нашим машинама. Ми такође желимо да дозволимо могућност да инжењер или тим инжењера могу конструисати машину која ради, али чији се начин функционисања не може задовољавајуће описати од стране њених конструктора јер су они применили методу која је већином експериментална. Коначно, желимо да од машина одвојимо људе које су рођени на уобичајен начин. Тешко је дати дефиниције а да се притом задовоље ова три услова. Може се инсистирати да тим инжењера буде истог пола, али ово не би баш било сасвим задовољавајуће, јер је вероватно могуће да се створи комплетна личност од једне једине ћелије коже (рецимо) узете од једног мушкарца. Овако нешто било би подвиг биолошке технике који би заслуживао огромне похвале, али ми не бисмо били наклоњени да то посматрамо као случај „ конструисања машине која размишља“. Ово нас приморава да се одрекнемо захтева да свака врста технике треба да буде дозвољена.

Још смо спремнији да то учинимо имајући на уму чињеницу да је данашње интересовање за „машине које размишљају“ повећано једном одређеном врстом машине која се обично зове „електронски рачунар“ или „дигитални рачунар“. Следећи овај предлог ми дозвољавамо само дигиталним рачунарима да узму учешће у нашој игри.

На први поглед ово ограничење се чини драстичним. Покушаћу да покажем да у реалности није тако. Како бих ово урадио потребан ми је приказ природе и својстава ових рачунара.

Такође се може рећи да ће ова идентификација машина са дигиталним рачунаром, као нашим критеријумом за „размишљање" бити незадовољавајуће, ако се (насупрот мом веровању) испостави да дигитални рачунари нису у стању да се добро покажу у игри.

Већ, постоји један број дигиталних рачунара који раде, и може се поставити питање, "Зашто да не покушамо да урадимо експеримент одмах? Било би лако задовољити услове игре. Могао би се користити један број испитивача, а сакупљена статистика како би се показало колико је често права идентификација дата. "Кратак одговор је тај да ми не питамо да ли би се сви дигитални рачунари добро показали у игри, нити да ли би се рачунари који су нам данас на располагању добро показали, већ, да ли би се рачунари који се могу замислити добро показали. Али ово је само кратак одговор. Ово питање ћемо касније видети по другачијем светлу.

Дигитални рачунари:

Идеја која стоји иза рачунара је та, да је циљ, да ове машине могу извести било коју операцију, коју може да изведе и човек. Особа би требало да прати следећа правила; она нема овлашћење да се разликује од рачунара ни за један детаљ. Можемо предпоставити да се ова правила налазе у књизи (програму), која се мења кад год му је задат нови задатак. Њему је такође на располагању неограничена количина папира на коме рачуна. Он такође може множити и сабирати на „стоној машини", али ово није битно. Сматра се да се дигитални рачунар обично састоји од три дела:

1. Складишта (меморијски простор);
2. Извршне јединице;
3. Контроле.

Меморија је складиште информација, и кореспондира са папиром код човека, било да је ово папир на коме он рачуна или онај на коме су правила књиге одштампана. Како човек обавља рачунање у својој глави један део складишта кореспондира са његовом меморијом.

Извршна јединица је део који изводи различите индивидуалне операције које су укључене у рачунање. То шта су ове индивидуалне операције вариране од машине до машине. Обично се могу урадити прилично дуге операције као што су „Помножи 3540675445 са 7076345687" али код неких машина само најпростије као што су „Напиши 0" су могуће.

Поменули смо да је „књига правила" која је унета у рачунар замењена

у машини једним делом меморије . Она се тада зове „табелом инструкција". Дужност је контроле да утврди да ли се ове инструкције поштују тачно и правим редоследом. Контрола је тако конструисана да се ово обавезно догађа.

Информација је у складишту обично разбијена на пакете умерено мале величине.

Мора да се прихвати чињеница да дигитални рачунари могу бити и да јесу конструисани према принципима које смо описали, и да они у ствари веома прецизно опонашају радње које обавља човек.

Књига правила за коју смо рекли да је користи човек, је наравно само фикција. Људи заиста памте шта треба да ураде. Ако неко жели да направи машину која имитира понашање људи у неким комплексним операцијама он мора да пита како се то ради, а затим да одговор преведе у форму једне инструкционе табеле. Конструисање инструкционих табела се обично описује као „програмирање". „Програмирати једну машину која треба да обави операцију *А*" значи ставити одговарајућу инструкциону табелу у машину како би она обавила операцију *А*.Најактуелнији дигитални рачунари имају само ограничено складиште. Не постоји теоретска потешкоћа у идеји о рачунару са неограниченим складиштем. Наравно да само ограничени делови могу бити коришћени у једном тренутку. Исто тако само ограничена количина се може конструисати, али ми можемо да замислимо да се додаје још по потреби.

Идеја о дигиталним рачунарима је стара. Чарлс Бебиџ (Charls Bebidge), професор математике на Кембриџу од 1828. год. до 1839. год. , планирао је такву машину, названу Аналитичка машина, али никада није завршена. Иако је Бебиџ имао основне идеје, његова машина у то време није се чинила атрактивном. Брзина тог рачунара била би дефинитивно већа од брзине човека али око 100 пута спорија од Манчестерске машине, која је сама једна од најспоријих модерних машина. Складиште би било чисто механичко, користило би точкове.

Пошто Бебиџова машина није електрична, и пошто су сви дигитални рачунари на неки начин еквивалентни, примећујемо да ова употреба електричне енергије не може бити од теоријског значаја. Наравно да електрична енергија ступа на снагу када је у питању давање и добијање сигнала, тако да није изненађујуће што је налазимо у обема овим конекцијама. У нервном систему хемијски феномени су барем толико битни колико и електрични.

Најактуелнији дигитални рачунари имају само ограничену меморију. Не постоји теоретска потешкоћа у идеји о рачунару са неограниченом меморијом. Наравно да само ограничени делови могу бити коришћени у једном тренутку. Исто тако само ограничена количина се може конструисати, али ми можемо да замислимо да се додаје још по потреби.

Често се придаје значај чињеници да су модерни дигитални компјутери

електронски, и да је нервни систем такође електронски. Пошто Бебрицова машина није електрична, и пошто су сви дигитални рачунари на неки начин еквивалентни, примећујемо да ова употреба електричне енергије не може бити од теоријског значаја. Наравно да електрична енергија добија на важности када је у питању давање и добијања сигнала, тако да није изненађујуће што је налазимо у обе ове конекције. У нервном систему биолошких бића хемијски феномени су барем толико битни колико и електрични. У неким комјутерима систем за складиштење је углавном акустичан. Коришћење струје се стога сматра само површном сличношћу. Ако бисмо желели да тражимо такве сличности требало би да потражимо математичке аналогије функције.

Универзалност дигиталних рачунара

Дигитални рачунари који су разматрани у претходном поглављу могу се класификовати међу „машине дискретног стања". Ово су машине које се покрећу изненадним скоковима и кликовима од једног поприлично одређеног стања до другог.

Суштинско својство механичких система које смо назвали „дискретно стање машина" је да се тај феномен не појављује. Чак и ако разматрамо праве физичке машине уместо идеализованих машина, довољно тачно познавање стања у једном тренутку пружа довољно тачно познавање у сваком следећем кораку касније.

Како смо већ поменули, дигитални рачунари се сврставају у класу машина дискретних стања. Али број стања за које је таква машина способна је обично изузетно велики. Није тешко видети зашто број стања треба да буде тако велики. Рачунар укључује једно складиште које кореспондира са папиром које користи човек. Мора постојати могућност да се напише у складишту било која комбинација симбола коју је је било могуће написати на папиру. Како би смо то јасније објаснили претпоставимо, да се само цифре од 0 до 9 користе као симболи. Разне варијације у рукопису се игноришу. Предпоставимо, да је рачунару дозвољено 100 папира од којих се на свакоме налази 50 линија а на свакој линији простор за 30 цифара. Онда је број стања $10^{100 \times 50 \times 30}$, то јест $10^{150\,000}$. Логаритам базе два од броја стања се обично зове „капацитет складиштења" машине.

Узимајући у обзир табелу која кореспондира са машином дискретног стања могуће је предвидети шта ће она урадити. Не постоји разлог зашто овај рачун не би могао да се изведе посредством једног дигиталног рачунара. Под условом да може да се изведе довољно брзо, дигитални рачунар би могао да имитира понашање било које машине дискретног стања. Игра имитације би се могла играти користећи машину која је у питању (као *Б*) и имитирање дигиталног рачунара (као *А*) а испитивач не би био у стању да их разликује. Наравно да дигитални рачунар мора имати довољан капацитет за складиштење, као и да ради довољно брзо. Штавише, мора се програмирати изнова за сваку нову машину коју жели да имитира.

Због ове особине дигиталних рачунара, да они могу да опонашају било коју машину дискретног стања, назвамо их *универзалним машинама*. Постојање машина са овим својством има за важну последицу то да, чак и ако изоставимо питање њихове брзине, није неопходно дизајнирати различите нове машине које ће обављати различите процесе рачунања. Сви они могу да се ураде на једном дигиталном рачунару, одговарајуће програмираним за сваки појединачни случај. Отуда је последица тога да су сви дигитални рачунари на известан начин еквивалентни.

Сада можемо поново расправити о питању које смо поменули. Опрезно

је предложено да питање: "Могу ли машине да мисле? " треба заменити питањем "Могу ли се замислити рачунари који би добро функционисали у игри имитације?" Ако хоћемо, можемо површно то генерализовати као "Постоје ли машине дискретног стања које ће то урадити добро?" Али с обзиром на својство универзалности, видимо да су оба питања идентична са следећим: "Хајде да усредсредимо пажњу на један одређени дигитални рачунар *Ц*. Да ли је истина да модификовањем овог рачунара тако да има адекватно складиште, подешавањем његову брзину рада и инсталирањем одговарајућег програма, можемо направити *Ц* тако да игра улогу *А* у игри имитације, при чему улогу *Б* има човек?"

Игра имитације

Предлажем да размотримо питање „Могу ли машине да размишљају". Ово би требало започети дефиницијама значења речи „машине" и „размишљати". Уместо што ћу покушати да дам такву дефиницију заменићу ово питање другим, које је уско повезано са њим а које је изражено релативно недвосмисленим речима.

Ако би човек покушао да се претвара да је машина свакако се не би показао како треба. Одмах би га одале спорост и нетачност у аритметици. Зар не могу машине да изведу нешто што би требало описати као размишљање али нешто што је веома другачије од оног што човек ради? Ова примедба је веома јака, али можемо барем рећи ако машина може бити конструисана да задовољавајуће игра игру имитације, ова примедба не треба да нам смета.

Хилбертов програм

Формализацију аритметике, која се заснива на формализацији логике коју је започео Фреге, Хилберт ће детаљно извести тек двадесетих година овог века. Овај проблем добија посебно место у једној сложенијој теорији која се назива Хилбертовим програмом или Хилбертовом теоријом доказа или још математиком. Хилберт ту и тамо и даље изјављује да је истинитост у математици исто што и непротивуречност, формализам који Хилберт заступа у свом програму је негде између чистог формализма и конструктивизма, а с друге стране он се слаже с платонизмом у прихватају средстава класичне математике.

Циљ Хилбертовог програма је да елиминише бесконачност из математике. Елиминисати бесконачност не значи забранити позивање на њу, него само показати да је она корисна фикција.

2.3. λ–рачун

Израчунљивост је одиста `апсолутни` математички појам. То је апстрактна идеја која измиче свим тумачењима појмова `тјурингових машина` . Није потребно придавати било какав посебан значај `тракама` и `унутрашњим стањима`, итд. и другим особеностима Тјуринговог довитљивог и специфичног приступа. Постоје и други начини да се изрази појам израчунљивости, од којих је историјски први по значењу био `ламбда рачун` америчког логичара Алонса Черча (Alonzo Church), до којега је дошао уз сарадњу Стивена Клина (Stephen C. Kleene). Черчов поступак био је сасвим другачији, и знатно апстрактнији од Тјуринговог.

У овој схеми, нас занима `универзум` објеката, означених, рецимо као

$$a,\ b,\ c,\ d,\ \ldots\ z,\ a',\ b',\ c',\ d',\ \ldots\ z',\ a'',\ b'',\ldots\ a''',\ \ldots\ a'''',\ldots$$

од којих сваки представља математичку операцију или функцију. (Разлог за коришћење вишеструких апострофа јесте једноставно у томе да се омогући неограничени скуп сиомбола за ознаке ових функција.) `Докази` ових функција, то јест, ствари на које ове функције делују, јесу друге ствари исте врсте, тј. такође су функције. Штавише, резултати (`вредности`) деловања једне такве функције на другу такође је функција. Тако, када пишемо

$$a = bc,$$

тиме сматрамо да је резултат функције b која делује на функцију c нека трећа функсија a. Нема тешкоћа у изражавању појма функција од две или више променљивих у овој схеми. Ако желимо да мислимо о f као о функцији двеју променљивих p и q, рецимо, можемо једноставно да напишемо

$$(fp)q$$

(што је резултат fp примењене na q). За функцију трију променљивих разматрамо

$$((fp)q)r,$$

и томе слично.

Сада долази знаменити поступак апстракције. За ово користимо грчко слово (ламбда), а иза њега следи слово које означава једну од Черчових функција, рецимо x, коју посматрамо као `произвољну променљиву`. Свака појава проименљиве x, унутар израза са угластом заградом који непосредно

следи сматра се тада просто за `празнину` у коју се може убацити било шта што следи целом изразу. Тако, ако пишемо

$$\lambda x.[fx],$$

ово значи да функција која делује на, рецимо, a даје резултат fa. То јест

$$(\lambda x.[fx])a = fa.$$

Другим речима, x.[fx] је једноставно функција f, тј.

$$\lambda x.[fx] = f.$$

Ово захтева мало размишљања. У томе је математичка лепота која се у првом тренутку чини тако свакидашња и ситничава да је лако потпуно превидети поенту. Размотримо пример из познате школске математике. Нека је функција f тригонометријска операција израчунавања синуса неког угла, тако да је апстрактна функција `sin` дефинисана са

$$\lambda x.[\sin x] = \sin.$$

(Немојте се бринути око тога како `функција` x може да буде угао. Ускоро ћемо видети начин на који се бројеви могу посматрати као функције; а угао је само једна врста броја). За сада, ово уистину јесте свакидашње. Али, хајде да замислимо да ознака `sin` није још измишљена, али да смо свесни да је развој у ред за sinx:

$$x \quad \frac{1}{6}x^3 + \frac{1}{120}x^5 \ldots$$

Тада бисмо могли дефинисати

$$\sin = \lambda x.[\, x \quad \frac{1}{6}x^3 + \frac{1}{120}x^5 \ldots]$$

Запазимо да бисмо, још једноставније, могли дефинисати, рецимо, `шестину куба` као поступак за који не постоји погодно стандардно обележавање:

$$Q = \lambda x.[\frac{1}{6}x^3],$$

и утврдити, на пример,

$$Q(a+1) = \frac{1}{6}\ (a+1)^3 = \frac{1}{6}\ a^3 + \frac{1}{2}\ a^2 + \frac{1}{2}\ a + \frac{1}{6}\ .$$

Значајнији за нашу расправу јесу изрази који су сачињени једноставно од Черчових основних функционалних поступака, као што су

$$\lambda f.[f(fx)].$$

Ово је функција која, када делује на неку другу функцију, рецимо g, даје g примењено двапут на x тј.

$$(\lambda f.[f(fx)])g = g(gx).$$

Такође смо могли `апстраховати` прво x и добити

$$\lambda f.[\lambda x.[f(fx)]],$$

што можемо скратити у

$$\lambda fx.[f(fx)].$$

Ово је поступак који, када се примени на g, даје `двапут поновљену функцију g`.Заправо, ово је управо функција коју Черч идентификује са природним бројем 2:

$$2 = \lambda fx.[f(fx)],$$

у овом систему могу се заменити природни бројеви, тако да $(2g)y = g(gy)$. Слично он дефинише:

$$3 = \lambda fx.[f(f(fx))], \quad 4 = \lambda fx.[f(f(f(fx)))], \quad \text{итд.}$$

заједно са

$$1 = \lambda fx.[fx], \quad 0 = \lambda fx.[x].$$

Заиста, Черчово `2` је више налик изразу `двапут`, а његово `3` изразу `трипут` итд. Тако је деловање 3 на функцију f, односно 3f, поступак `поновити f три пута`. Дејство 3f на y, стога, биће

$$(3f)y = f(f(f(y))).$$

Погледајмо како једноставна математичка радња, наиме додавање 1 броју, може да се изрази у Черчовој схеми. Дефинишимо

$$S = \lambda abc.[b((ab)c)].$$

Да би смо видели да S доиста једноставно додаје 1 броју описаном у Черчовом запису, испитујемо га на броју 3:

$$S3 = \lambda abc.[b((ab)c]3 = \lambda bc.[b((3b)c)] = \lambda bc.[b(b(b(bc)))] = 4,$$

пошто је $(3b)c = b(b(bc))$. Јасно је да се ово исто тако односи и на било који други природан број. (Заправо, $\lambda abc.[(ab)((ab)c)]$, ради исто то као и S.)

Шта је са множењем броја са два? Ово удвостручавање се може постићи са

$$D = \lambda abc.[(ab)((ab)c],$$

што се опет може илустровати дејствовањем на 3:

$$D3 = \lambda abc.[(ab)((ab)c]3 = \lambda bc.[(3b)((3b)c)] = \lambda bc.[(3b)(b(b(bc)))] = \lambda bc.[(b(b(b(b(b(bc))))))] = 6.$$

Заправо, основне аритметичке радње: сабирање, множење и подизање на степен могу се дефинисати, респективно. као:

$$A = \lambda fgxy.[(fx)((gx)y)],$$
$$M = \lambda fgx.[(f(gx)],$$
$$\Pi = \lambda fg.[fg],$$

Можемо се уверити да доиста важе релације

$$(A\mathbf{m})\mathbf{n} = \mathbf{m} + \mathbf{n}, \quad (M\mathbf{m})\mathbf{n} = \mathbf{m}\mathbf{x}\mathbf{n}, \quad (\Pi\mathbf{m})\mathbf{n} = \mathbf{n}^{\mathbf{m}},$$

где су **m** и **n** Черчове функције за два природна броја, **m** + **n** је функција за њихов збир итд. Последња од ових релација је најзанимљивија. Проверимо је за случај **m**=2, **n**=3 :

$$(\Pi 2)3 = ((\lambda fg.[fg])2)3 = (\lambda g.[2g])3 = (\lambda g.[fx.[f(fx)]g])3 = \lambda gx.[g(gx)]3 = \lambda x.[3(3x)]$$
$$= (\lambda x[fy.[f(f(fy))](3x)] = \lambda xy.[(3x)((3x)((3x)y))] = \lambda xy.[(3x)((3x)(x(x(xy))))]$$
$$= \lambda xy.[(3x)(x(x(x(x(x(xy))))))] = \lambda xy.[(x(x(x(x(x(x(x(xy))))))] = 9 = 32$$

Године 1937., Черч и Тјуринг су независно један од другог показали да се сваки израчунљив (или алгоритамски) поступак, сада у смислу Тјурингових машина, може остварити (помоћу појмова неког од Черчовог израза (и обратно)) у Черчовом систему израчунљивости.

Постоје и други начини за дефинисање појма израчунљивости. Постоји свакако још погоднија дефиниција израчунљивости (рекузивности) до које су дошли Ј. Хербранд и Гедел, Х.Б. Кари 1929. године, као и М. Шенфинкл 1924. године, имали су нешто раније другачије приступе, из којих је делимично настао Черчов рачун.

3. Рекурзивне функције

Занимљиво је питање да ли се класа Turing–израчунљивих функција може описати на „алгебарски„ начин, тј. да ли се она може добити користећи одређене операције над аритметичким функцијама полазећи од неких једноставних функција. То је могуће, овде ћемо само навести, одговарајуће операције над функцијама су: супституција, рекурзија и минимизација.

Постоје и многи други системи ефективне израчунљивости, о којима се читалац може информисати у литератури. Ипак појам израчунљивости остаје исти, који год од ових приступа се усвоји.

4. Мишљење разних других

Сада можемо рећи да смо разјаснили ову ставку и спремни смо да наставимо полемику у вези са нашим питањем: "Да ли машине могу да мисле?", као и његову варијанту наведену на крају претходног одељка. Не можемо у потпуности напустити првобитни облик проблема, јер ће се мишљења разликовати у погледу адекватности супституције и морамо бар послушати шта се може рећи с тим у вези.

А сада ћу наставити са разматрањем мишљења која су супротстављена A.M. Turingu.

(1) *Теолошки приговор.*

Размишљање је функција човекове бесмртне душе. Бог је дао бесмртну душу сваком човеку и жени, али не и некој животињи или машини. Стога, ниједна животиња нити машина не могу да мисле. [1]

Како хришћани гледају на муслиманско веровање да жене немају душу? Али оставимо ову ставку по страни и вратимо се на главну тврдњу. Чини ми се да горе наведена тврдња указује на озбиљно ограничење свемогућног Бога. Црква прихвата да има ствари које Он не може да уради, као што је да једно изједначи са два, али зар не би требало да верујемо да може да да душу слону ако то сматра прикладним? Могли бисмо да очекујемо да би то спровео у дело заједно да мутацијом која би слона обезбедила адекватно побољшаним мозгом који би одговарао потребама његове душе. Тврдња врло сличног облика може се изрећи и кад су у питању машине. Може изгледати другачије јер ју је много теже "прогутати". Али то заиста значи само то да сматрамо да је мање вероватно да би Он размотрио те околности погодним за давање душе.

(2) *Примедба названа «глава у песку».*

"Последице тога да машине мисле биле би исувише застрашујуће. Надајмо се и верујмо да оне то неће моћи".

Ова тврдња се ретко када може изразити тако отворено као у горе наведеном облику. Али она погађа већину нас који о томе уопште и размишљају. Оно је посебно јако код интелектуалаца, пошто они цене моћ мишљења много више од осталих и више су склони да заснују своје веровање у човека и овој моћи.

Не сматрам да је ова тврдња довољно заснована да би захтевала побијање. Утеха би била адекватнија: можда је треба тражити у трансмиграцији душа.

(3) *Математички приговор.*

Постоје бројни резултати математичке логике који се могу користити да

[1] Вероватно је ово мишљење јеретичко. Св. Тома Аквински (Summa Theologica, цитирао Бернард Расел, Историја Западне филозофије) сматра да Бог не може да створи човека без душе. Али то не мора да значи недостатак божје моћи, већ последицу тога да су људске душе бесмртне, и стога неуништиве.

покажу да постоје ограничења моћи машина. Најпознатији од њих познат је под називом Годелова теорема и он показује да у било ком довољно логичном систему могу се формулисати искази који се не могу ни потврдити, ни оповргнути унутар система, изузев ако је сам систем недоследан. Ту су и други, по много чему слични резултати, захваљујући Черчу, Клину, Росеру и Тјурингу. Последњи резултат је најзгоднији за разматрање, пошто се односи директно на машине, док се други могу користити у релативно индиректном аргументу: на пример, ако хоћемо да користимо Годелову теорему, потребна су нам и нека средства за описивање логичких система у погледу машина, а машина у односу на логичке системе. Оно о чему желим да разговарамо односи се на један тип машине која је у основи дигитални рачунар са бесконачним капацитетом. Овај резултат тврди да има извесних ствари које таква машина не може да уради .Питања за која знамо да машина неће дати одговоре су типа "Замислимо машину која изгледа на следећи начин... Да ли ће ова машина икад одговорити са "да" на било које питање?" Тачке треба заменити описом неке машине у стандардном облику, што може бити нешто слично ономе из предходног. Када описана машина има извесну релативно једноставну везу са машином која се испитује, може се показати да је одговор или погрешан или да до њега неће доћи. То је математички резултат: сматра се да он показује извесне немогућности машине, које нису својствене људском интелекту.

Кратак одговор на ову трврдњу је да, иако је установљено да постоје извесна ограничења моћи одређене машине, само је наведено, без икаквог доказа, да таква ограничења не важе за људски интелект. Али ја не сматрам да треба олако одбацити ово становиште. Кад год се машини постави неко од критичних питања и она да дефинитиван одговор, знамо да тај одговор мора бити погрешан. И ми сами често дајемо погрешне одговоре на питања. Они који се придржавају математичког аргумента би, сматрам, углавном били ради да прихвате игру имитације као базу дискусије. Они који верују у две претходне примедбе вероватно не би били заинтересовани ни за какве критеријуме.

(4) *Аргумент о свести.*

Овај аргумент је веома добро изложен у Листерској расправи професора Цеферсона, из које наводим: "Док машина не напише сонет или компонује кончерто због мисли и емоција које осећа, а не због случајног ређања симбола, не можемо се сложити да је машина исто што и мозак –тј, не само да их напише, него и да зна шта је написала. Ниједан механизам не може осетити ни показати задовољство због постигнутог успеха, тугу због неуспеха, нити је може усрећити ласкање, ни учинити несрећном грешка, нити може да буде опчињена супротним полом, љута или депресивна када не може добити оно што хоће".

Ова тврдња изгледа као порицање валидности нашег теста. Према најекстемнијем облику овог схватања једини начин на који неко може да буде сигуран да машина мисли јесте да буде машина и да сам осети да размишља. Сигуран сам да професор Џеферсон би веома радо прихватио игру имитације као тест. Игра (без играча *Б*) често се користи у пракси под називом *viva voce* како би се утврдило да ли неко заиста разуме нешто или је "научио као папагај".
Дакле, укратко, верујем да би се већина оних који подржавају *аргумент о свести* могла наговорити да би они вероватно пристали да ураде наш тест.

(5) *Аргумент о различитим немогућностима.*
Ови аргументи имају облик "Признајем да можеш натерати машине да раде све оне ствари које си поменуо, али никада нећеш моћи да натераш некога да уради X". Бројне карактеристике X су наговештене у вези са овим. Овде су неке од њих:

> Буди љубазан, домишљат, леп, дружељубив имај иницијативу, буди духовим, разликуј погрешно од исправног, чини грешке, заљуби се, уживај у јагодама са шлагом, натерај некога да заволи, учи из искуства употребљавај речи правилно, буди предмет властитих мисли имај исто онолико различитих облика понашања колико и човек, уради нешто заиста ново (Неке од ових карактеристика посебно су разматране на наведеним странама оригиналног текста: *Computing machinery and intelligence od A.M. Turing-a).*

Овакви искази се обично ничиме не поткрепљују. Верујем да су углавном засновани на принципу научне индукције. Човек током живота види хиљаде машина. На основу онога што види он долази до великог броја различитих закључака. Оне су ружне, свака има јако ограничену сврху примене, ако су неопходне за минимално другачију сврху, бескорисне су, разноврсност понашања било које од њих је веома мала, итд, итд. Наравно, он закључује да су то неопходна својства машина уопште. Многа од ових ограничења у вези су са веома малим капацитетом складиштења информација код већине машина. (Претпоставка је да се капацитет меморије односи на све машине, а не само на *discrete state machines*, тј. машине дискретног стања. Тачна дефиниција није неопходна, пошто се не захтева никаква математичка прецизност у актуелној расправи). Претпостављам да је до тога долазило због примене принципа математичке индукције. (Дела и обичаји човечанства не чине се баш погодним материјалом на који се може применити принцип научне индукције. Иначе бисмо могли, као већина енглеске деце, закључити да сви говоре енглески и да је глупо учити француски.
Ипак, има посебних примедби на које треба скренути пажњу у вези са неким немогућностима које су наведене. Немогућност уживања у јагодама са шлагом може се учинити будаласто. Можда би се могла направити машина

која ће уживати у јагодама са шлагом, али сваки покушај да се то уради био би апсурдан.

Тврдња да машине не могу да погреше је чудна. Човек не може а да не одговори: "Да ли су оне због тога лошије?" Мислим да се ова критика може објаснити помоћу игре имитације. Тврди се да би испитивач могао да разликује машину од човека само тако што ће им задати одређени број аритметичких проблема. Машину бисмо препознали по њеној невероватној тачности. Одговор на ово је једноставан. Машина програмирана за играње игрице не би покушала да да *тачан* одговор када су у питању аритметички задаци. Она би намерно направила грешке на такав начин који ће збунити испитивача. Механичка грешка би се обично показала у некој неодговарајућој одлуци какву грешку ће направити у аритметичком задатку. Али не можемо даље о томе расправљати. Чини ми се да ова критика зависи од конфузије две врсте грешака. Можемо их назвати "грешке функционисања" и "грешке закључивања". Грешке функционисања настају услед неке механичке или електричне грешке, због које се машина понаша другачије од онога како је направљена да се понаша. По дефиницији оне не могу да праве грешке у функционисању. У том смислу можемо рећи да "машине никада не могу да греше". Машина може, на пример, да напише математичку једнакост или реченицу на енглеском. Када се унесе погрешна претпоставка, кажемо да је машина направила грешку у закључивању. Јасно је да нема разлога да кажемо да машина не може да направи овакву врсту грешке.

На тврдњу да машина не може бити предмет властитих мисли може се одговорити само ако се може доказати да машина има *неку* мисао на *неку* тему. Међутим, "тема операција неке машине" ипак нешто значи, бар људима који раде са њом. Ако, на пример, машина покушава да нађе решење једначине $x^2 - 40x - 11=0$, можемо рећи да је једначина у том тренутку тема. У том смислу, машина без сумње може да буде сама себи тема. Може се користити за прављење својих сопствених програма, или да предвиди промене у властитој структури. Посматрајући резултате сопственог понашања машина може да модификује сопствене програме, како би ефикасније дошла до неког циља. Ово су могућности блиске будућности, а не неки утопијски сан.

Критика да машина не може да има разноврсно понашање је само други начин да се каже да она нема велики капацитет складиштења података.

Критике које разматрамо овде често су прикривени облици аргумента о свести. Обично ако неко тврди да машина може да уради неку од ових ствари и опише врсту методе коју машина може да користи, неће пуно импресионирати некога. Сматра се да је овај метод, ма који он био, јер мора бити механички, прилично базичан.

(6) *Приговор леди Лавлис.*

Једна варијанта примедбе леди Лавлис је да машина никада не може да уради нешто "заиста ново". На ово се може одговорити "Нема ништа ново под капом небеском". Ко може да буде сигуран да "оригинални рад" који је он урадио није једноставно производ семена које је посађено у њега током образовања, или слеђења добро познатих општих принципа. Једна боља варијанта овог принципа каже да нас машина никада не може"изненадити". То је аргумент који морамо сматрати затвореним за даљу расправу, али можда вреди приметити да признавање такво изненађења захтева исто толико "креативног менталног напора", без обзира да ли изненађујући догађај потицао од човека, књиге, машине или нечега другог.

Схватање да машине не могу изненадити, проистиче из једне погрешне претпоставке којој су филозофи и математичари посебно подложни. То је претпоставка да чим се нека чињеница изложи уму, све последице те претпоставке настају истовремено с њом у уму. То је веома корисна претпоставка у многим ситуацијама, али се такође лако заборавља да је она погрешна. Природна последица тога је да човек предпостави да нема сврхе разматрати последице које су настале на основу података и општих принципа.

(7) *Аргумент о континуитету у нервном систему.*

"Нервни систем сигурно није машина дискретног стања. Мала грешка у информацијама о величиие нервног импулса у неурону, може да направи велику разлику у величини излазног импулса. Ако је тако, може се тврдити да се не може упоређивати понашање нервног система са системом дискретног стања (рачунаром).

Истина је да машина дискретног стања (рачунар) мора бити другачија од континуалне машине (нервног система). Али ако се придржавамо услова игре имитације, испитивач неће бити у стању да на било који начин искористи ову разлику.

(8) *Аргумент о неформалности понашања.*

Немогуће је направити скуп правила која би требало да опишу шта човек треба да уради у свакој ситуацији која се може замислити. Може се, на пример, донети правило да човек треба да се заустави кад види црвено светло, а настави ако види зелено, али шта, ако се неком грешком, оба појаве у исто време? Човек би помислио да је најбоље зауставити се. Али из ове одлуке могу произићи неки проблеми касније. Покушати наћи правила понашања која би покрила сваку могућу ситуацију, па чак и ону која може настати на семафору, изгледа немогуће. Слажем се са свим овим .

(9) *Аргумент о екстра – сензорној перцепцији.*

Претпостављам да је читалац упознат са идејом о екстра – сензорној перцепцији и значењем њена четири облика: телепатијом, видовитошћу, прекогницијом и психокинезом. Ови узнемиравајући феномени као да поричу све наша оубичајена* научна схватања. Како бисмо волели да их порекнемо! Нажалост, статистички докази, бар када је телепатија у питању, запањујући су. Веома је тешко преуредити нечије схватање тако да се ове нове чињенице уклопе у њих. Једном када их човек прихвати, чини се да није далеко следећи корак – да поверује у духове и утваре. Идеја да се наша тела понашају само у складу са познатим законима физике, заједно са неким другим, још увек неоткривеним, али некако сличним законима, био би први корак, упркос томе што се сукобљавају са ЕСП (екстра – сензорном перцепцијом);

5. ТЈУРИНГОВ ТЕСТ

Замислимо да се на тржишту појавио најновији модел рачунара, са величином меморије и бројем логичких јединица који надмашују људски мозак. Предпоставимо такође да су ти уређаји пажљиво програмирани и опскрбљени огромним количинама података одговарајуће врсте. Произвођачи тврде да њихове справе заиста мисле.Могуће је да они такође тврде да њихове справе заправо мисле и корак даље, рећи да уређаји заправо осећају: бол, срећу, сажаљење и понос и тако даље, да су свесни и разумеју шта раде.Тврди се, заправо, да су они свесна бића. Како да установимо да ли можемо да верујемо тврдњама произвођача или не? Да бисмо проверили тврдњу произвођача да одређена справа има људска својства , једноставно ћемо, на основу овог критеријума, питати да ли се она понаша као што би питали да ли се она понаша као што би се људско биће понашало у погледу тих особина.

Операционалиста би рекао да рачунар мисли уколико делује другачије него што то чини особа која размишља.

То значи да тражимо да рачунар даје одговоре сличне људским на било које питање које бисмо му поставили.У чланаку A.Turinga који је 1950.год. појавио у филозофском часопису Ум (Mind) под насловом „Рачунајуће машине и интелигенција“ изложена је замисао која се данас назива Тјурингов тест, одлучујућа провера којом се установља да ли се за уређај може разложно тврдити да мисли.

ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА

Циљеви вештачке интелигенције су да се машинским (обично електронским) путем опонашају колико је год могуће људске менталне активности , а такође и да се уколико је могуће, побољшају људске способности у том погледу у најмање четири правца.

У *роботици*, дотиче се практичних потреба индустрије у погледу механичких уређаја који могу да обављају "интелигентне" задатке, посредовање или надзор, брзином и поузданошћу које надмашује људске способности, или под тешким условима у којима би човеков живот био у опасности; од комерцијалног и општег интереса јесте и развој *експертских система*, по којима би суштинско знање читавих занимања, медицинских, правних и тако даље, било смештено у меморији рачунара! Да ли је могуће да би искуство и знање стручњака из ових области могли уистину бити замењени таквим складиштима података? Затим *психологија* постоји нада да ће се подраживањем понашања људског мозга (или мозга неке друге животиње) у некаквом електронском уређају, или неуспешном покушају да се

то учини, доћи до ванредног сазнања о раду мозга. На крају, постоји и нада да би вештачке интелигенције могла да осветли нека кључна питања из области филозофије, тако што би пружала увид у значење појма *ума.* Као пример симулација психотарапеута, дијалога између рачунарског „тарапеута“ и пацијента,

⋮

ПАЦИЈЕНТ: Постоји један изазов у томе.
ТАРАПЕУТ: Зашто вам је то важно?

Рачунар следи прост систем механичких правила иако се чини да је у стању разуме .

6. НАПАДИ НА МИШЉЕЊА R.PENROUZ–А:

ПРИСТУП ВЕШТАЧКЕ ИНТЕЛИГЕНЦИЈЕ "ЗАДОВОЉСТВУ" И "БОЛУ"

Једна од захтева који се поставља пред вештачке интелигенције (убудуће ВИ) истраживања јесте да она обезбеде пут до разумевање менталних вредности, као што су срећа, бол, глад. Размотримо пример корњаче Греја Волтера. Када су њене батерије на измаку, њен образац понашања се мења, и она тада смишља начин како да попуни залихе енергије. Уместо једноставног пребацивања из једног начина понашања у други, запажа се промена *у настојању* да се делује на одређен начин. Неки поборници ВИ замишљају да се појмови као што су бол или срећа могу моделовати на сличан начин.

Замислимо да поседујемо уређај – некакву машину, претпоставимо електронску - која има начина да региструје свој (претпостављени) резултат 'бол - задовољство'. На основу чега бисмо могли да тврдимо да он заиста осећа 'бол - задовољство'?

Вештачка интелигенција (или операционалистички) поглед на ствари био би да то пресудимо просто на основу начина на који се уређај понаша. Јасно је да су наши поступци условљени знатно комплекснијим мерилима. Али, као врло груба апроксимација, избегавање бола и призивање задовољства збиља јесу начини на који ми поступамо. За операсционалисту, ово је довољно оправдање, на истом нивоу апроксимације, за идентификацију бол задовољство резултатата нашег уређаја са односом бола и задовољства у њему. Такво поистовећивање изгледа да је један од циљева ВИ теорије.

ЈАКА ВЕШТАЧКО ИНТЕЛИГЕНТНА ПОСТАВКА И СЕРЛОВА КИНЕСКА СОБА

Постоји гледиште које се назива јака ВИ поставка (Енгл. *strong AI thesis*-прим.прев.)

Не само да ће поменути уређаји заиста бити интелигентни , поседовати умове и тако даље, већ ће се и менталне вредности одређене врсте моћи приписати логичком функционалисању *ма ког* рачунског уређаја, чак и најједноставније механичке направе, попут термостата.

Ментална активност је заправо извршавање добро уређеног следа поступака, који се често назива *алгоритам*, као рачунски поступак одређене врсте. У случају термостата, алгоритам је изузетно једноставан уређај региструје да ли је температура већа или мања од подешене вредности и тада одређује да се струјно коло прекине у првом, а повеже у другом случају. Али ако постоји алгоритам такве врсте за мозак- а присталице јаке ВИ поставке тврде да је то случај, тада би се он у начелу могао извести. То су само неке осебености *алгоритма* које мозак извршава.

Заиста, он би се могао извести на *било ком* модерном електронском рачунару опште намене, само да није ограничења у меморијском простору и брзини операција, такав алгоритам кад год ди се изводио, *у себи* би доживљавао осећања; поседовао свест; био ум.

Амерички филозоф Џон Серл (John Searle, 1980, 1987) снажно се супоставио овом ставу, где су рачунари већ прошли поједностављене верзије Тјуринговог теста, али су својства "разумевања" у потпуности одсутна. Један од таквих примера заснован је на рачунском програму Podžep Шенк (Roger Schank; Schank&Abelson, 1977) на симулацији разумевања једноставних прича попут следеће: „....хамбургер..(Човек улази у ресторан и наручује хамбургер. Када је хамбургер стигао, био је препечен попут угља; човек љутито изјури из ресторана , не плативши рачун, нити оставивши напојницу; Други пример: Човек улази у ресторан и наручује хамбургер. Када је хамбургер стигао, човек је био врло задовољан; и кад је напуштао ресторан, оставио је конобару велику напојницу и пре него што је платио рачун.) .“ Питање да ли је човек појео хамбургер у сваком од та два случаја (чињеницу која није експлицитно поменута ни у једној од прича). Где у веома ограниченом смислу, уређај је већ прошао Тјурингов тест. Питање које морамо размотрити јесте да ли је ова врста успеха заиста показатељ стварног разумевања рачунара, или можда самог програма.

Серлов аргумент, да се не ради о разумевању јесте увођење појма "кинеске собе". Он претпоставља најпре да се све приче причају на кинеском, а не на енглеском, сигурно не суштински важна промена, и да су све операције алгоритма који рачунар израчунава представљене у виду скупа упутстава (на енглеском) за употребу плочица са кинеским симболима. Серл замишља самог себе како све манипулације изводи унутар закључане собе.

Низ симбола који представљају најпре приче, а затим и питања, достављају се у собу кроз мали прорез. Никаква друга информација из спољнег света није доступна. Коначно, када се све манипулације окончају, завршна секвенца шаље се напоље кроз прорез. Пошто су све ове манипулације једноставно извођење алгоритама из Шенковог програма, оне морају дати као коначан резултат низ симбола који на кинеском значе „да“ или „не“, што ће представљати тачан одговор на кинеском на причу испричану на кинеском језику. Међутим Серл нам ставља до знања да он не разуме ни реч кинеског језика, тако да нема ни бледу представу о чему су приче заправо.

Ипак правилним извођењем низа операција које сачињавају Шенков алгоритам (упутства за алгоритам дата су на енглеском) он ће бити у стању да прође тест подједнако добро као и Кинез који би лако разумео садржину приче. Серлов закључак, јесте да само успешно решавање алгоритма не значи да је дошло до разумевања. Серл, (замишљено) закључан у кинеској соби, није разумео ниједну реч из прича!

Серлов аргумент изазвао је низ приговора. Пре свега, има нешто донекле обмањујуће у изразу „ не разуме ниједну реч“, како смо га користили. Разумевање се односи на обрасце исто као и на појединачне речи. Неке од образаца које успостављају симболи, чак и ако их не схвата, значење појединачних симбола решава алгоритам ове врсте. На пример кинески карактер за „хамбургер“ могао би бити замењен карактером за неко друго јело, рецимо „chow mein“и приче не би биле значајно измењене.

Као друго, извођење чак и веома једноставног рачунског програма као изузетно дуготрајан и напоран посао уколико би га обављала људска бића коришћењем симбола (имамо рачунаре који обављају такве задатке!). Овде расправљамо о принципу, а не о практичним аспектима целе ствари. Тешкоћа настаје пре свега са замишљеним рачунарским програмом за који претпостављамо да је довољно сложен да се упореди са људским мозгом и тако прође меродаван Тјурингове тест.

`Пресудан` ниво сложености у алгоритму, неопходан да би алгоритам испољио менталне вредности, није незамислив, да постоји, иако се овде бавимо принципима. У пракси апсурдно, није апсурдно у принципу, и аргумент је суштински исти као и раније: манипулатори симболима не разумеју причу, упркос јаке ВИ поставке да само решавање одговарајућег алгоритма доноси менталну вредност `разумевања`.

Земља се тврди Серл, као ни аутомобил или термостат `не бави разумевањем`, док се појединци тиме баве. Да овај аргумент неоспорно утврђује да не постоји нека врста не материјалног `разумевања` везаног за решавање самог алгоритма, чије присуство ни на који начин не утиче на свест самог појединца. Такође он наводи (али не више од тога) да ниједан алгоритам, ма колико сложен, никад не може сам по себи да испољи истинско разумевање, насупрот тврђењима јаке ВИ поставке. Постоје, колико запажам,

и друге веома озбиљне тешкоће. Према овој поставци, само се алгоритам рачуна. Нема никакве разлике да ли се алгоритам изводи у мозгу, електронском рачунару, читавој држави Индуса, механичком уређају са точкићима и зупчаницима, или систему водоводних цеви. По овом становишту, једноставно, логичка структура алгоритма јесте та која је значајна за `ментално стање` које треба да представи, док је конкретна физичка материјализација алгоритма сасвим небитна. Као што Серл наглашава, ово заправо води ка једном облику `дуализма`. Дуализам је филозофско становиште које је заступао филозоф и математичар из XVII века, Рене Декарт, које тврди да постоје две засебне врсте супстанције: `дух` и обична материја. Да ли и како ове две врсте супстанције могу утицати једна на другу, додатно је питање. Суштина јесте у томе да дух није сачињен од материје и да је у стању да постоји независно од ње. У јакој ВИ поставци дух је логичка структура алгоритма.

Конкретно, физичко извођење алгоритма нешто је потпуно ирелевантно. Алгоритам има неку врсту нематеријалног `постојања` које је сасвим одвојено од било ког извођења алгоритма у физичким појмовима. Заговорници јаке ВИ поставке доиста узимају реалност алгоритма озбиљно, пошто верују да алгоритам чине `супстанцију` њихових мисли, осећања, разумевања и њихових свесних перцепција. Како је Серл запазио, становиште јаке ВИ поставке води ка пренаглашеном облику дуализма.

Даглас Хофштатер (Dourlac Hofctadter,1981) тврди да је, *у начелу*, књига потпуни еквивалент, у операциононалном смислу Тјуринговог теста, застрашујуће успорене верзије стварног Аншатајна. Заиста, пошто је књига претпостављена само као једна од материјализација алгоритма који представља Аншатајнову `личност`, она би била уистину била Аншатајн. Али сада имамо нову потешкоћу. Књига никада не би била отворена, или би је стално прелиставали. Како би књига 'знала ' за разлику? Да ли промене у алгоритмима (а овде укључујем садржај меморије као део алгоритма) оно што треба повезати са менталним догађајем пре него извођењем самих алгоритама? Или би можда Аншатајн–књига остао потпуно свестан чак ни ако га никада нико не би испитивао, нити би га узнемиравао? Хофштатер се дотиче неких од ових питања, али не успева да на њих одговори, нити да их разјасни.

Шта значи активирати неки алгоритам или га материјализовати у физичкој форми? Очигледно апсурдна идеја? Не сматрам да је ова идеја апсурдна сама по себи–она је само погрешна!

Серл, прихвата и следећи став: `Наравно да је мозак дигитални рачунар. Пошто је све у ствари дигитални рачунар, тако је и мозак`. Серл прихвата да се разлика између функционисања људског мозга (с којим је, он тврди, везан ум) и електронског рачунара (који, по њему нема ум) који изводе исти алгоритам, своди искључиво на њихову материјалну конструкцију. Он

тврди, али није у стању да објасни, да биолошки објекти (мозгови) могу имати `наум` и `значење` док електронски објекти то не поседују. Мени се не чини то као пут научној теорији ума.

Други разлог потиче из квантне физике, који је у супротности са првим. Према квантној механици, било која два електрона нужно су потпуно идентична, или ако се пар било којих честица једне врсте, замени одговарајућим честицама друге врсте, тада се строго узевши, заправо ништа се није десило. Оно што разликује особу од куће (честицама из цигала куће) јесте начин на који су конституенти уређени, не индивидуалност самих конституената.

Постоји аналогија, овоме на свакодневном нивоу, ако желимо да променимо неку реч, постоји време између нестанка слова и појављивања слова, како се положај сваког следећег слова поново прорачунава, а затим поново – поново прорачунава када се уметне друго слово. Питање је да ли је ситуација *иста* након замене, или не?

У квантној механици заменити једну честицу другом исте врсте значи заправо не мењати стање. Пошто се сваки атом може, у принципу, непрекидно пратити, тако да се може замислити да се сваком припише одговарајућа ознака.

Могуће је да се говори о индивидуалности атома. Прихватимо сада да индивидуалност особе није повезана са особеношћу која би се могла приписати њеним материјалним конституентима. Уместо тога, она мора бити повезана са конфигурацијом, рецимо да је то конфигурација у простору или простор–времену. Али заговорници јаке ВИ иду даље од тога. Ако се информациони садржај такве конфигурације може превести у неки други облик из кога би се, опет, могао реконструисати оргинал, тада по њима индивидуалност особе мора остати нетакнута.

Ствар је иста као са низом слова која уносим преко тастатуре. Сутрадан могу да убацим дискету, да успоставим мала наелектрисања и тако поново прикажем низ слова на екрану, баш као да се у међувремену ништа није ни догодило. Тако, они чак сматрају да би свест особе наставила да постоји док се `информација` о особи налази у том другом облику. Према том гледишту, `свест особе` третира се ефективно као део софтвера, а њено појединачно испољавање код материјалних људских бића јесте извршавање тог софтвера на хардверу, који сачињавају људско тело и мозак.

7. Закључак

Хајде на тренутак да се вратимо на приговор леди Лавлис, по коме машина може да уради само оно што јој ми кажемо да уради. Могли бисмо рећи да неко убризгава идеју у машину, да ће она на тренутак одговорити, а онда се ућутати, како кад клавирску жицу ударе чекићем. Придржавајући се ове аналогије, питамо: "Може ли се направити машина која ће бити суперкритична?"

Процене капацитета меморије људског мозга варирају од 10^{10} до 10^{15} бинарних цифара. Нагињем нижим вредностима и верујем да је само мали део искоришћен за више форме мишљења. Већина његовог капацитета вероватно служи за задржавање визуелних утисака. Изненађена бих била када би више од 10^9 било потребно за задовољавајуће играње игре имитације, бар против слепог човека. (Примедба: Капацитет *Енциклопедије Британике*, 11. издање, износи $2x10^9$). Капацитет меморије од 10^7 била би изводљива могућност чак са тренутним техникама. Вероватно није неопходно повећавати брзину рада машина. Делови модерних машина (рачунара) који се могу сматрати аналогним нервним ћелијама, раде скоро 1000 пута брже од нервних ћелија. То може обезбедити "маргину безбедности", што би покрило губитке брзине до којих дође. У процесу покушаја да имитирамо ум одраслог човека, морамо да пуно размишљамо о процесу који га је довео до стања у коме је. Можемо уочити три компоненте:

а) иницијално стање ума, тј. стање на рођењу;
б) образовање коме је био повргнут;
ц) остало искуство, које се не може назвати образовним, а коме је ум био изложен.

Стога смо поделили наш проблем на два дела: дечији програм и образовни процес. Ове две ставке остају веома повезане. Постоји очигледна веза између овог процеса и еволуције:

структура детета – машине = наследни материјал
промене = мутације
природна селекција = суд експериментатора

Међутим, треба да се надамо да ће оваĸав процес бити више истраживачки него еволуциони. Опстанак најспособнијих је спор метод за мерење предности. Експериментатор, уз помоћ интелигенције, требало би да га убрза. Подједнако је важно да он није ограничен на случајне мутације. Ако можемо ући у траг некој слабости, вероватно би могао да нађе и врсту мутације која ће га побољшати.

Неће бити могуће применити исти образовни процес на машину као на нормално дете. Машина, на пример, неће имати ноге и од ње се неће захтевати да изађе и напуни кантицу угљем.

Вероватно је мудро уврстити случајан елемент у машину за учење. Системски метод имаману што може да постоји велики блок без било каквог решења, у области коју треба прво испитати. Пошто вероватно постоји велики број задовољавајућих решења, случајни метод изгледа бољи од системског. Треба приметити и да се он користи у аналогном процесу еволуције. А ту системски приступ није могућ. Како би неко могао да прати различите генетске комбинације које су испробане, да их поново не би испробавао?

Ствари треба показивати и именовати, итд. Испред себе можемо видети само мало, али и пуно тога што у том малом треба да урадимо.

Литература:

- *Царев нови ум, Роџер Пенроуз, Информатика, Београд, 2004. год.*
- *Хилбертови проблеми и логика, Жарко. Мијајловић и аутори, Завод за уџбенике и наставна средства, Београд, 1986. год.*
- *Computing machinery and intelegence, чланак A. M. Turinga.*
- *http://www.wikipedia.org*