ПРОЕКТ „ИЗКУСТВЕН ИНТЕЛЕКТ” – ОГРАНИЧЕНИЯ

Проф. Сергей Герджиков
Конференция „Съзнание”
Софийски Университет, 2014

Abstract

This article is a stock of ideas and theories about intelligence, provoked by the project "Artificial Intelligence (AI)" and offered by computer specialists and neurouscientists.

I formulate and discuss here some deep limitations to the AI, not cleared enough.

Biological limitation. Natural selection as a natural process *could not be implemented* — it is unique following ‘the way of life’, the life process, under always unique constellations of conditions. The robots with AI will not be in a position to follow such an undeterminable behavioural line under momentary conditions to be taken into account by the AI. Robots does not seek survival.

Sensory limitation. Any artificial intelligence as part of the world in which we live is perceived sensory in our body and interpreted as such. Monitoring and testing such as Turing test will produce data dependent on our body (and culturally dependent on our culture) on how AI is different from a man.

When recognising any sensory object, you must include the body senses. *Sensory objects cannot be considered as objects of an artificial intellect*. The engineers here, obviously, is totally confused when trying to write programs for ‘sensory qualities’ like a wet, hard, relief, not to mention the colors, smells and tastes. And how to design nice and pleasant, attractive and repulsive, rational and irrational?

Intuitively and experimentally (in behavior) *no AI de facto has senses*. Even if there is some ‘sensitivity’ resulting from his machine organization, it remains hidden to us. Of course, we observe the robot's behavior. This behavior, even externally resembling human’s, will differ from the last under tough enough tests. Intractable remains a matter such as: whether the robot with cameras and programs *sees* and what does he see? It should be clear that *the robot does not see things around it such as objects and their movements*. In his computer there is information processing of huge rows of data, instantaneously changing, grouped by specific sets of 0 and 1.

Keywords

Intelligence, artificial intelligence, imitations, information, adaptation, surviving, sensual objects

Резюме

Тук се формулират и обсъждат базови ограничения пред проекта AI, които не са изяснени в известните ми публикации върху AI.

Биологично ограничение. Селекцията като естествен процес, естественият отбор, не може да бъде имитирана – естественият отбор е уникално прокарване на пътя на живота, жизнения процес, при винаги уникални констелации от условия. В роботите няма жизнен процес и няма уникални моментни условия, които да се отчитат от AI. Изкуственият отбор няма за цел оцеляването.

Сетивно ограничение от Homo sapiens. Наблюдението и тестовете от типа н Тюринг-теста ще дадат данни, *биологично зависими от нашето тяло* (и културно зависими от нашата култура преценка) относно това, доколко AI е различен от човек.

Всеки изкуствен интелект като част от света, в който живеем, е *възприеман сетивно* през нашето тяло и интерпретиран като такъв. Когато обсъждаме какъвто и да е сетивен обект, трябва да включваме сетивото, тялото. Сетивните обекти не може да се приемат като обекти на един изкуствен интелект (робот, снабден с камери и компютър). Тук инженерите, очевидно, се объркват тотално, когато се опитват да пишат програми за „сетивни качества" като мокро, твърдо, релефно, да не говорим за цветове, миризми и вкусове. А как да проектираме приятно и неприятно, привлекателно и отвратително, рационално и ирационално?

Сетивно ограничение от AI. AI интуитивно и фактически няма сетива. Дори да има някаква сетивност, произтичаща от организацията му, тя остава скрита за нас. При използване на роботи с AI не можем да знаем какво роботът реално изпитва и прави. Разбира се, ние можем да наблюдаваме поведението на робота. Това поведение, външно наподобяващо човешкото, ще се отличава от последното при достатъчно труден тест. Нерешим остава въпрос като: дали роботът с камери и програми вижда и какво вижда? Ясно е, че роботът не вижда нещата около себе си като предмети и техни движения, осветявани всеки момент различно. В неговия компютър се обработват мигновено променящи се огромни редици от данни, групирани като специфични серии от 0 и 1.

Тези ограничения правят изкуствения интелект непостижим в обозримо време, в обозримата технология и в оборзимото различие между организми и машини.

Въведение

Тази статия е равносметка на идеи и теории за интелигентността, провокирани от проекта „Изкуствен интелект" и предлагани от компютърни специалисти и невроучени. Не се следва играта на понятия: „съзнание–реалност", „ум–тяло", „ментални състояния („Theory

of Mind", квалии)–мозъчни състояния", "умолект–език", „логика–език", и свързаните с тях дискусии, аргументи, контрааргументи и мисловни експерименти. Въпроси от типа: „Какво е мислене?", „Какво е интелект?" и „Може ли машината да мисли?" са неясно формулирани и пораждат необхватен хаос от спекулации. При емпирично поставяне техните модификации водят към *информационните теории и технологии* и *невронауката (неврологията)*.

Проект Изкуствен интелект(AI) – кратка равносметка

Тюринг, Гьодел и непълнотата

През 1950 г. Тюринг публикува статията 'Computing Machinery and Intelligence', в която развива проекта си за тест, в който компютърът евентуално е неразличим от човек в общуването си с екзаминатор. Тук се формулират и някои философски и логически аргументи и възможни възражения. През 1951 темата е продължена с 'Intelligent Machinery, A Heretical Theory'. Дискусията се подема от медиите и Тюринг дава серия публични радио-лекции по AI (Intelligent Machinery, A Heretical Theory', 'Can Digital Computers Think?) и дискусията 'Can Automatic Calculating Machines Be Said to Think? През 1953 г. Тюринг публикува статия върху компютърния шахмат. (по Copeland 2004, 356)

Статията от 1950 г. започва с описание на тестово разпознаване на мъж и жена от екзаминатор, като жената (участник А) се опитва да го заблуди, че е мъж. (Turing 2004, 441). Сега Тюринг пита: „Какво ще се случи, ако машина заеме мястото на А в тази игра? Ще сгреши ли екзаминаторът толкова често, колкото ако играта се играе между мъж и жена? Тези въпроси заместват въпроса „Могат ли машините да мислят?" (ibidem) В играта се включва и трети участник, 'foil', който се опитва да помогне на екзаминатора. Компютърът ще отговаря с "не" на въпроси от типа: „Ти машина ли си?" Темите на разговор са най-разнообразни, те се задават предварително. Способността да се играе имитационната игра е критерий за „мислене". (ibid., 442, 443) „Аз вярвам, че след около 50 години (2000 г. – С. Г.) един програмиран компютър с памет около 10^9 бита (прибл. 80 мегабайта – С. Г.) ще играе имитационната игра толкова добре, че един среден екзаминатор ще има не повече от 70 % шанс да направи вярна идентификация след пет минути разговор." (ibid., 449) Тюринг отчита и възраженията, които се повдигат срещу тази постановка, вкл. аргументите „от съзнание" и от „континуалност на нервната система": „Нервната система определено не е машина с дискретни състояния. Малка грешка в информацията от нервния импулс, входящ в неврона, може да причини голяма дивергентност на изходния импулс. Може да се възрази, че щом нещата стоят така, не може да се очаква да се имитира поведението на нервната система от дискретна система. Вярно е, че една дискретна система трябва да е различна от континуална

машина. Но ако се придържаме към условията на имитационната игра, екзаминаторът няма да може да даде каквото и да е предимство на тази разлика." (ibid., 456). Очевидно, ако нервната система принципно се различава от компютър, това ще даде отражение в имитацията. Точно затова тестът на Тюринг е много по-добър в разпознаването на интелигентното поведение от всеки предложен теоретичен критерий.

Машината на Тюринг, която по дефиниция решава формални задачи с краен брой дискретни стъпки, или теоретичният модел на компютъра, все пак има ограничения (и според самия Тюринг). Тези принципни граници се коренят в непълнотата на формалните системи, доказана от Гьодел. Установено е съответствие на формалната система с машина на Тюринг. Тогава наличието на неразрешими (недоказуеми и неопровержими за една формално-аксиоматична система) пропозиции технически е съответно на спирането на универсалната машина на Тюринг или ирелевантността на даваните от нея отговори. Детерминираните стъпки на компютрите, аналогични на логическите изчисления, ще изпуснат безкрайно множество от истинни твърдения или интелигентни решения. На тази база се развихря, както обикновено, нестихващ дебат на чисто теоретична и спекулативна основа по въпроса: Могат ли машините да отговарят на въпроси подобно на хората? Тюринг привежда теореми аналогични на тези на Гьодел за неопределеността на формалните системи. Това са теоремите на Нани. „Проблемът за „зацикляне" (the halting problem) е наречен така (и изглежда за пръв път поставен) от Мартин Дейвис. Пропозицията, че проблемът за „зацикляне" не може да бъде решен от изчислителна машина е известен като 'halting theorem'. Невъзможно е да се знае дали на една машина на Тюринг ще й бъде необходимо безкрайно много време, за да извърши някои математически операции. Тук математическите доказателства са априорни относно реалния тест на Тюринг и създаването на интелигентни компютри. Те не могат да имат силата на фактическото разиграване на теста, ограничен във времето.

През миналия век са прости компютърни програми като Eliza, които могат да имитират разговорната реч и вследствие на това да заблудят повечето неподозиращи хора и да ги накарат да повярват, че си говорят с човек. (Повечето хора например по време на разговор използват само няколко стотин думи и съсредоточават вниманието си върху малък брой теми.) Но досега не е написана нито една компютърна програма, която може да заблуди хора, които си поставят точно определената задача да се опитат да установят в коя кутия е човекът и в коя машината (по Каку 2010).

Top–down стратегия

Подходът ‘отгоре-надолу’ към изкуствения интелект е известен и като GOFAI от англ. “Good Old-Fashioned AI” или „добрия старомоден изкуствен интелект“. Това значи *формулиране на изкуствения интелект в програми*. Задачата се определя като програмиране на правилата, по които се разпознават образи, разбира се реч и се ориентираме в пространството. (Да отбележим, че проектът не е за машина, която изчислява аритметични функции, нито за персонален компютър, който имаме на бюрото си.)

През 50-те и 60-те години на 20 в. са създадени компютри, които могат да играят дама и шах, да събират разпилени детски кубчета и т.н. Напредъкът е впечатляващ и се прогнозира, че след няколко години роботите да надминат по интелигентност хората.

През 1969 г. в Станфордския изследователски институт е създаден робота SHAKEY. Това е малък компютър, поставен на комплект от колела и има инсталирана камера. Камерата е в състояние да прави оглед на стаята, а компютърът анализира и идентифицира предметите в нея. Опитва се да се придвижва около тях. Тук започват засечките.

Този подход използва постиженията на теоретици като Дейвид Мар. Изчисляват се входните пиксели от камерата и се преобразуват според програма в геометрични форми. В резултат се създават роботи, на които са им необходими часове, за да се придвижат през специална стая, в която има предмети само с прави линии, т.е. квадрати и триъгълници. Една плодова мушица, в чийто мозък има само около 250 000 неврона и която притежава нищожно малко от изчислителната мощност на тези роботи, може без усилие да се придвижва в три измерения, извършвайки смайващи лупинги във въздуха, докато тези роботи се загубват в две измерения.

Дейвид Мар използва аспекти от резултатите на биолози като Джеймз Гибсън, например определянето на формите в реалната среда: „Тези геометрични твърди тела съдържат огромна сложност, но те напълно могат да се анализират в термините на три компонента: *лица*, *ръбове*, и *върхове*.” (Gibson 1986, 29) Проблемът обаче е много по-дълбок. Твърдото тяло се различава от мекото, от течността и от газа, но това не може да се формализира в термините на пикселни разпределения, в каквито се анализират ръбове, лица и върхове. А повечето тела в реална среда са съвсем неправилни.

Камерите подават двумерна и динамична информация за тези релефи, а програмата разпознава лицата, ръбовете и върховете, заедно със сенките, които трябва да водят към третото измерение. Ограничението е, че програмите, обработващи данните от камерата, фактически обработват поредици от пиксели, които определят като прави линии, триъгълници и други правилни фигури. Огромно предизвикателство, както пише Дейвид Мар (1982) е тримерната визия според осветеността и сенките. (Цветовете се изключват.)

При това една статична картина, която се описва от примерно от 100 милиона пиксела, при завъртане на робота или преместване на предмета, се губи тотално и трябва да се описва новата картина, т.е. да се започне отначало. Това зрящите същества правят без усилия. Много време и грешки отнема в ранното детство ученето да виждаме и разпознаваме предмети, повърхности и среди, при това с помощта на всички сетива.

На практика нашите най-усъвършенствани роботи от типа Top-Down като роботите-скитници на планетата Марс притежават интелекта на насекомо. В прочутата Лаборатория за изкуствен интелект в Масачузетския Технологичен Институт експерименталните роботи се сблъскват с трудности при възпроизвеждането на постижения, които са по силите дори на буболечки, като откриването на скривалища и разпознаването на опасности.

Нито един робот не би могъл да разбере и най-обикновен детски разказ, който се чете пред него. Някои специалисти обясняват неуспеха с това, че роботите могат да виждат, и то много по-добре от хората, но не разбират това, което виждат (Каку 2000). Реално те не виждат, защото виждането е именно *разпознаване на обектите от средата*, а не просто регистрация на светлинното разнообразие. Обектите в средата са: на фона на небе и въздух, трева и хълм, отпред расте дърво, а отдясно има скала. Проблемът е по-дълбок от самата сложна форма. Проблемът е в това, че сетивните системи на организмите в реалното им придвижване в околната среда са адаптирани да разпознават обектите в тяхното „биологично значение", като неопределени множества от индивидуални обекти, които правят преминаването възможно или невъзможно, определят траекторията до ценни ресурси или избягването на опасности. Животните избягват твърдите тела, през които не могат да преминат, а през някои те преминават. Те избягват по преценка за вреда или полза различни среди и повърхности по пътя на придвижването си. Те се движат към биологично значими цели. Роботите няма как да имат цел да се запазят или репродуцират, като обменят вещества, енергия и информация, неутрализирайки ентропията и извършвайки специфична и винаги насочена към ре-синтез активност.

Интелектът в правила

С оглед на тези проблеми следващата задача е формулирането на „разпознаването" или на „здравия разум" *в серия от правила*. Най-амбициозният опит е CYC на Дъглас Ленат (1984). Подобно на „Проекта Манхатън" за атомната бомба, CYC трябвало да бъде „Проектът Манхатън" за изкуствения интелект, финалният тласък, който да доведе до създаването на истински изкуствен интелект.

Дъглас Ленат е признат специалист по изсуствен интелект. Той разработва области като машинно учене (machine learning) (програмите AM and Eurisko), познавателна

репрезентация, и „онтологично инженерство”("ontological engineering") – неговата CYC програма в MCC и в Сйкорп (Cycorp). Той също публикува критика на конвенционалната еволюционна теория на случайните мутации на основата на своя опит с програмата Eurisko. Ленат представя естествения език във формален, използвайки програмата CYC и един огромен речник от думи. Използва екстензия на първо-редно предикатно изчисление. ОТ МИТ го критикуват и измислят единицата „микроленат” за измерване на термини – неговият речник е тнвърде голям, за да се използва на практика.

Мотото на Ленат е „Интелектът – това са 10 милиона правила“. През 1984 г. Ленат предрича, че след десет години – през 1994 г., CYC ще съдържа между 30 и 50 процента от информацията за „консенсусната реалност“. Към 2005 г. CYC още не е близо до постигането на тази цел. Както установяват специалистите от „Сайкорп“, трябва да бъдат програмирани милиони и милиони кодирани линии, за да се приближи един компютър до здравия разум на четиригодишно дете. Към 2008 г. последната версия на програмата CYC съдържа само незначителните 47 000 понятия и 306 000 факти. “Хората научават тези закони, защото ние, колкото и да е досадно, продължаваме да се потопяваме в околната среда през целия си живот, като усвояваме спокойно законите на физичната и биологична среда, но роботите няма как да го правят. Дори толкова просто нещо като посочването на разликата между една отворена врата и прозорец може да бъде дяволски трудно за един робот.” (Каку 2010)

Мичио Каку също изпуска спецификата. Тя не е в дългото учене, а в това, че „консенсусната реалност” не е формулируема в правила. Тя се синтезира и ресинтезира непрекъснато, селектирана и напасвана по уникален начин за всеки индивид и в огромната си част е несъзнавана.

В същото време продължават успехите на изчислителните машини в области като счетоводство и шахмат. През 1997 г. световният шампион по шахмат Каспаров и компютърът Deep Blue с процесор Intel Pentium, програмиран от шахматисти специално за този противник, играят равностойно серия интересни сеанса.

Роботи, програмирани да се придвижват в опростена среда, се правят и днес, и то с добър успех. Добър пример е японският АСИМО (Advanced Step in Innovative MObility, - Усъвършенствана стъпка в иновативната мобилност). Той се рекламира като „разпознаващ хора, предмети и жестове. Може да ходи, да тича, да рита топка, да носи табла, да бута количка, да се здрависва, да говори и дори да танцува.″ В тези описания се вижда цялото недоразумение, с което се идентифицира разпознаването на хора и говоренето, стига нещо да е имитирано. Ако разпознаването е чрез добре определени маркери, то е по силите на една несложна програма. Придвижването на ASIMO е наистина впечатляващо, но роботът никога не се придвижва в неопределена предварително и необхванатa от програмите му среда.

Направени са наистина големи стъпки към движение на хунамоиден корпус в права линия, в изкачване и слизане по стълби и обръщане. Не се показва придвижване на робота през различни помещения с неочаквани препятствия.

2013. „ICub, хуманоидния робот в лабораторията за роботика IDSIA в Швейцария, се опитва да достигне за чаша син. За да направи това, той трябва да планира и контролира движението на всички си стави в унисон. ICub е с доказани възможности за успешно изпълнение на следните задачи:

- пълзи, използвайки визуални указания с оптични маркер на пода;

- решава сложни 3D лабиринти;

- стрелба с лък; учи се да стреля в центъра на мишената;

- мимики, в израз на емоции;

- силов контрол, използване на проксимална сила/въртящ момент с помощта на датчици;

- използване на малки обекти като топки, пластмасови бутилки, и т.н;

- избягване на сблъсък в рамките на не-статични среди, както и на самосблъсък."

Новото поколение роботи мълчаливо, но видимо изоставя проекта AI в неговия силен вариант: като кандидат за решаване теста на Тюринг. Програмирани двигателни и поведенчески задачи развиват скромните постижения преди две-три десетилетия. Очевиден е напредъкът в програмираната кинематика, понякога аналогична на тази на човешкото тяло.

Ограничения на Top-down стратегията

Защо проектът е толкова успешен в изчислителен аспект, с който са свързани и постиженията на програмите за игра на шах и подобни, а се проваля в сферата на „неправилното" човешко и животинско поведение? На този проект му липсва перспективата на живото (вчастност речево) поведение, което е имплицирано в постановката на Тюринг. В действителност теоретиците и инженерите на AI си дават сметка за физиологията и дори твърдят, че я моделират. От друга страна, самите невроучени до началото на 90-те години също споделят изчислителния модел. Така или иначе биологията е твърде обща, а изчислението е ясно формулирано като функция, която може да работи в изследването на кортекса.

Истината, както се потвърждава в гореописаните неуспехи, е, че биологичната адаптация, състояща се в постоянно решаване на проблеми, е *задача неизчислителна*. Най-абстрактно казано: едно решение не се извежда от един проблем, а се синтезира. Дори доказателството на теорема, в които теоремата се извежда, не е изчисление и не извежда

резултата от проблема, а от *подбрани* предпоставки и дефиниции. Селекцията като форма на създаване на информация е и това, което става в природата.

Предметите в средата имат биологичен смисъл и съизмеримост с тялото. Те се виждат и чуват, вкусват и осезават като тримерни среди и тела, които са *препятствия или преходи* към биологично значимите храни и вода, опасности и врагове, функционални артефакти. Едно програмирано поведение неизбежно ще блокира в непредвидени препятствия. Едно „самообучаващо се поведение", от друга страна, е неизбежно безразлично към „оцеляване" и зависимо от програми.

Нашият мозък не е програмиран да преобразува данни по алгоритъм от вход към изход. Очевидно ние правим и това, изчисляваме и обработваме информация. Но тези процеси са силно подпомогнати от знакови системи и далеч отстъпват на изчислителната мощ на компютрите дори от най-старото поколениие. Нашите мозъци са част от нашите тела, които пък са потопени в света, а не са „изправени срещу" света. В средата има електромагнитни вълни в огромен спектър, но зрението обработва изключително тясна лента от 300 до 700 nm. При това едни и същи основни цветове се получават от различни вълнови пакети. Слухът обработва вълни от 16 до 20 000 Hz. А какви обективни физични свойства обработва осезанието, вкусът или обонянието? Тези сетива са специализирани за детекция на полезни и вредни вещества, не като физически тела и среди и химически съединения. Следователно информацията, обработвана от сетивата, е вече кодирана чрез трансдукция в органични зависимости от оцеляването. За да я разкодираме и превърнем в обективно описание, трябва да познаваме кодирането или максимално да елиминираме биологичната специфика на сетивната информация чрез уреди. В науката това се прави в някаква степен, но не като трансцендиране на биологичните условия.

Критиците и някои от авторите на проектите за изкуствен интелект казват: „Роботите виждат много по-добре от нас, но не разбират това, което виждат." (Каку 2000) „Компютрите виждат прави и кръгове." Или: „Компютрите виждат 0 и 1."

Какво виждат компютрите реално? Това е излишен въпрос. Трябва да е ясно, че това, което правят тези машини с оптичната информация, няма нищо общо с усещанията и възприятията на организмите. То е представимо за нас в светли и тъмни геометрични форми. В компютрите се извършват изчисления на пиксели, генерират се цифрови изрази на геометрични форми и се насочват движения по програма. Няма място за „усещане".

Оптичните програми, снабдени с мощни видеокамери, имат драматично ограничение. Един мигновен кадър на зрителното поле може да се опише със стотици милиони пиксели. Само малко движение на робота – и кадърът се сменя. Малка промяна на осветеността – и пикселовата картина драстично се променя. Ние не се смущаваме, а веднага се ориентираме

в промяната. Компютърът не може да свърже единия кадър с другия, а изчислява отново. Няма как в изчислителната програма да се зададе алгоритъм, по който един кадър се свързва със следващия в динамични условия, аналогични поне елементарно на естествените.

Разпознаване на статични образи.

Четене на ръкописен текст. Можем ли да зададем буквата *к* като инвариантен патерн? Например една права линия и две пречупени, съединяващи се в средата на правата. Този модел няма да работи, защото ръкописно ние не изписваме прави, нито спазване напълно събирането на двете прави в средния район на вертикална права. Може да напишем *к* и без да съединяваме трите прави линии, а като оставяме кухина или като изписваме малка права напред в средния участък на вертикалната права и после две наклонени. С висока степен на вероятност една добра четяща програма може да се справи с един почерк, на този, който я използва. Но ако той е написал буквата по необичаен начин, тя няма да бъде разпозната.

На практика стотиците милиони индивиди, които пишат буквата к милиарди пъти, ще я напишат, разложена на пиксели, в *неограничен брой комбинации от пиксели*.

Аналогична е задачата за разпознаване на реч. Тя изглежда обаче по-дефинирана цифрово, защото променливите в една произнесена фонема са по-малко. Затова днес компютрите могат да изпълняват прости вербални команди, ясно произнесени и еднозначни.

Налице са преводни програми. Компютърът превежда по начина, по който вижда и чува, чете и изпълнява - както една кукла прави забавни неща, имитиращи живо същество. Компютрите могат да усвоят синтаксиса на един език (например да боравят с граматиката, формалната структура и т.н.), но не и неговата реална семантика (какво означават думите според речниците или как се използват).

Всички тези ограничения се коренят в биологичното интегрално измерение *оцеляване*. Компютърът не „оцелява" и няма „нужда" изобщо да прави каквото и да е. Ако му бъде зададено „оцеляване" като запазване на структурната цялост срещу увреди, износване, компютърът ще се самоизключи като гаранция за оцеляване. Но биологичните системи се възобновяват, ресинтезират, срещу спонтанното си разпадане.

Нашият мозък обаче не притежава подобна структура. Той не е програмиран и не преобразува еднозначно данни от вход към изход по определен алгоритъм.

Bottom-up стратегия

Тази стратегия е плод на пореден биологичен урок. Технологията се опитва да следва „логиката" на самия опит и на ученето чрез проби и грешки.

Насекомите например не се движат, като сканират околната среда и формулират изображението в трилиони трилиони пиксела, които обработват със суперкомпютри. Вместо това мозъците им са съставени от невронни мрежи, обучаващи се уетуерни машини, които бавно научават как да се движат в един враждебен свят чрез потопяване в него. Простите инсектоидни механични създания, които се потопяват в околната среда и се учат от грешките си, могат да се лутат успешно по пода на МТИ в продължение на няколко минути.

Това постигат и роботите, изпратени на Марс.

Разликата между този подход и подхода отгоре-надолу е разликата между обратна връзка и програма. В обратните връзки се моделира основен принцип на поведението и на живите същества: проба – грешка.

Кое е грешка и кое не е, се задава от програма в която се формулират задачи, например поддържане на някаква температура или *следване на някаква определена линия*. Работата е там, че реално кое е грешка и кое не е биологичен въпрос, а не въпрос на улучване или избягване. Веднъж избягването на едно препятствие може да е правилно, а следващият път – погрешно по една и съща програма поради промяна в незначителни условия. За организма пък едно избягване може да е грешка в един случай и правилно решение в друг.

Този подход все пак е един порядък по-адекватен от подхода на програмирането.

Един от проектите на Брукс е COG. **Cog** е проект на Групата за Хуманоидна роботика (Humanoid Robotics Group) на Масачузетския технологичен институт. Той е базиран върху хипотезата, че човешката интелигентност е базирана върху комуникацията с други човешки индивиди. Така се учат децата. Машината е построена така, че да реагира на сигналите на хората за това кое е правилно и кое не е. От нея се очаква да наподобява дете, което се учи.

COG изглежда като плетеница от жици, електрически вериги и различни приспособления, като изключим факта, че има глава, очи и ръце. В него *не са програмирани никакви закони на интелекта*. Вместо това той е проектиран да съсредоточава погледа си върху човека, който го обучава и който се опитва да го научи на прости умения. Една изследователка, която забременяла по време на експериментите, се обзаложила кой ще се учи по-бързо — COG или нейното дете до двегодишна възраст. Детето надминало много машината.

От 2003 г. проектът е прекратен. Днес Cog е изложен в музея на МИТ.

Bottom-up подходът е определено по-добър от Top-down, ако се цели имитиране на адаптивно поведение. Свидетелство за това е, че роботите на Марс не се провалят лесно, учейки се на релефи и избягване на препятствия. Защо и подходът „дъно-връх” не успява в имитирането на човешко поведение? На подхода *Bottom-Up* също му липсва биологична перспектива. Въпросът е: дали програмите за обучение за определени задачи имат същата

форма като обучението в оцеляване? Защото *ако нямат такава форма, а формата на неутрални задачи, то те няма да проявяват качествата на естествените интелекти*. На компютъра като система не му е нужно да се самообучава – на него трябва да му се предпише в програма да се самообучава в това, което прави. Самообучението, зададено в програма, има *радикално ограничение – то е насочено към определени класове операции, а не към интегрална функция като адаптация или оцеляване*. Тази именно функция или състояние-атрактор е, която еволюционно селектира успешните решения.

Ограничения на Bottom-Up стратегията

Еволюцията е пред технологиите и по степента на сложност на биологичните организми и мозъците. Големите невронни мрежи, инсталирани в роботи, могат да се състоят от десетки до стотици „неврони“, но човешкият мозък има над 100 милиарда неврона. Червей, чиято нервна система е картографирана от биолозите, има само над 300 неврона, една от най-простите, срещани в природата. Но между тези неврони има повече от 7 000 синапса. Колкото и да е проста, неговата нервна система е толкова сложна, че досега никой не е успял да изгради неин компютърен модел.

Един неврон се моделира от хиляди микротранзистори. През 1988 г. компютърни експерти предсказват, че до 2000 г. ще имаме роботи с около 100 милиона изкуствени неврона. В действителност невронна мрежа със 100 неврона се смята за изключителна.

Най-голямата ирония се крие във факта, че машините могат без усилие да изпълняват задачи, които хората смятат за „трудни“, като умножаването на големи числа или играта на шах, но се объркват напълно, когато поискате от тях да изпълнят задачи, които са изключително „лесни“ за човешките същества, като ходенето из стая, разпознаването на лица или клюкарстването с приятели.

Марвин Мински от МТИ, водещ автор и един от създателите на AI, обобщава проблемите по следния начин: „Историята на AI е много забавна, защото първите истински постижения бяха доказателства в областта на логиката и висшата математика. Но след това започнахме да се опитваме да правим машини, които да отговарят на въпроси, свързани с простичките разкази в учебниците за първи клас. Но и до днес няма машина, която да може да прави това.“ (цит. по Каку 2000)

В реалното поведение има по нещо и от двете посоки: нали мозъкът е до някаква степен генетично програмиран да прави определени неща. От друга страна, мозъкът е в началото празен откъм информация за средата извън инстинктите и гените, които ги определят. Още с първите часове за едно новородено започва ученето и приспособяването. Дали синтезът на двата подхода ще създаде желаната машина?

Бъдещето ще покаже.

Ограничения на изчислителната парадигма

Теориите и обясненията в изчислителната парадигма обособяват когнитивните функции от жизнения процес като цяло и поддържат утопични нагласи към проекта „изкуствен интелект". Потвърждение на това е, че обширните и детайлни разработки като тези на Мар и Гросбърг не водят до емпирични резултати в реалното описание на възприятията и интелигентното поведение и най-вече в създаването на изкуствен интелект, който би играл успешно имитационната игра на Тюринг.

Заслужава да се отбележи модела на обучение на Гросбърг като многоизмерна карта от произволно m-мерно входно пространство към произволно n-мерно изходно пространство. Моделирането на сетивната информация в многомерни виртуални пространства, в които се фиксират типовете входна и съответно изходна информация е забележително постижение, което унифицира анализа на вход-изходните съответствия в посока от сетивна към моторна информация. Пол и Патриша Чърчланд използват такъв модел в своите теоретични визии за работата на мозъчната кора (Churchland, PS 1986, Churchland, PM 1995).

Всички тези автори спадат към огромната изчислителна парадигма в неврологията, която обаче е сериозно разклатена след системните неуспехи по проекта „Изкуствен интелект" и главно от биологичните обяснения на главния мозък от невроучени като Антонио Дамасио, Рудолфо Линас, Уилиям Ута и много други. Марвин Мински, дългогодишен автор в тази област, се заема също с невронаука, като обаче препотвърждава изчислителния проект: „да предложи как вероятно работи човешкият мозък, и да се създаде машина, която може да чувства и мисли. Тогава можем да се опитаме да приложим тези идеи, за да разберем себе си и да развием Изкуствената интелигентност." (Minsky 2006, 7)

Така последните десетилетия в проекта за изкуствен интелект са период на пренасочване към естествения интелект с оглед бъдещи имлпементаци. Силна стъпка в това направление изглежда Теорията "Hierarchical Temporal Memory" на Джеф Хокинс. Но нейният анализ е тема на следваща публикация.

Реален Тюринг-тест

Тестът, който може да покаже сравнимостта на компютър и човек, несъмнено е нещо от рода на „теста на Тюринг".

Ето как Пол Чърчланд описва един съвременен реален Тюринг-тест:

„През декември 1993 г. годишното Тюринг-тест състезание се проведе в Сан Диего, организирано с любезното и ефективно домакинство от отделението по електроника на Дженеръл Дайнамикс Корпорейшън. Това е състезание, провеждано за имплементацията на известния тест за машинна интелигентност, предложен от британския математик и компютърен учен Алън Тюринг през 1951 г.” (Churchland 1995, 227). 229 “На всеки от осемте съдии беше дадено много внимателно определено време от петнайсет минути за разговор с другия край на телетипната връзка. След всеки такъв рунд всички те се превключваха към друга телетипна връзка към нов терминал, посрещайки нов кандидат и тема, и повтаряха разпитния процес. Те не знаеха колко от осемте кандидата бяха машини и колко от тях бяха хора. Това те самите трябваше да определят. Осем рунда и около два часа и половина по-късно, содиите трябваше да класират всички осем терминала в намаляващ ред на „проявено човешко поведение” ("apparent humanity") на непознатата единица, която срещаха на другия край на връзката. „Най-добрата машина” трябваше да бъде онази, която просто получеше най-високо общо класиране от осемте съдии.” (ibid., 228)

„Да отбележим, че да спечели това състезание, за всяка от машините не беше необходимо да заблуди съдиите, че тя е човек. Тя не трябваше да отблъсне всяка от човешките атаки, а само да победи другите машини. Доколкото имаше парична награда, нейните конкуренти бяха само другите програмирани компютри. И при това положение, все пак от съдиите се искаше да прокарат линия през своето класиране на определено ниво, с вероятните човешки същества над линията и вероятните машини под нея. На минали състезания някои от участвали машини наистина успяха да заблудят съдиите да повярват, че на другия край на линията е бил човек. На техническата конференция преди състезанието, философът-университетски професор Даниел Денет, председател на журито, описа процедурите, които да се следват от съдиите и реферите.” (ibid., 229-230).

„Какви бяха резултатите? Машините с „Либерална” срещу „Консервативна” програма изпревариха другите две машини с не много голяма разлика. Но победата на машина беше гарантирана от правилата. По-интересно е, че нито един от осемте съдии не беше заблуден, че някоя от трите машини участници е човек. В това отношение съдиите спечелиха 1.000, а програмирани машини спечелиха 0.000. Дори при ограничението за една тема на разговор, всяка една машина участник загуби този тест на Тюринг, срещу всеки един съдия. (ibid., 232-233)

„Искам да подчертая, че от тази история могат да се извлекат само два урока. Първият е: въпреки процъфтяващата индустрия на класическото програмиране AI продължава да произвежда много поразителни функционални системи, нищо далеч напомнящо на истински човешки интелект все още не е сред тях. Или във всеки случай не сред участниците в нашия конкурс. В този конкурс трите участника компютърни програми се оказаха написани с цел "да проявят интелигентност просто дотолкова, че да си тръгнат с Льобнерова награда,", а не с цел реално пресъздаване на човешката интелигентност.

Вторият урок е, че хората, дори много интелигентните, блестящи индивиди, не са толкова надеждни, колкото бихме могли да очакваме при различаването на реалната човешка интелигентност от машинните симулации, поне в ситуацията на телетипна връзка. Те не са надеждни дори когато машинните симулации бяха съвсем слаби. Това отново повдига стария въпрос за теста, предложен от Алън Тюринг. Има ли този тест някаква реална значимост? Аз ще аргументирам, че той няма, и че тестовете с реална значимост трябва да са други." (ibid., 233-234)

Оставям заключението без коментар. Ясно е, че ако тестът „Тюринг" не е добра дискриминация, той би бил заменен за двайсетте изминали години.

Заключение за принципните ограничения на AI

Термично ограничение. Това ограничение е вече изтъквано (Каку 2000, 2013). Компютър, обработващ информация от реална среда в реално време, за да се ориентира в пространство-времето, харчи огромна енергия за охлаждане, за да не се разрушат неговите електронни мрежи. Един съизмерим с човешкия мозък компютър няма да може да поддържа работна температура, защото няма да му стига наличната енергия за охлаждане.

Непостижима сложност. Един неврон се моделира от хиляди транзистори. От колко транзистора може да се моделират 100 милиарда свързани по специфични начини неврона? Най-развитите изкуствени невронни мрежи са сложни от порядъка на стотици неврони, при това без отчет на специфичните връзки, плод на някакъв опит. Въпросният биологичен опит е резултат на милиони години естествен отбор и милиони ситуации на реално поведение в реална среда за оцеляване. Колко елементарни проби-грешки ще са нужни на такава проста невронна мрежа, за да постигне поведението на едно насекомо със същия брой неврони?

„Биологично ограничение". Селекцията като естествен процес, естественият отбор, не може да бъде имитирана – естественият отбор е уникално прокарване на пътя на живота, жизнения процес, при винаги уникални констелации от условия. В изкуствените системи няма жизнен процес и няма уникални условия. Изкуственият отбор няма за цел оцеляването.

Сетивно ограничение от Homo sapiens. Наблюдението и тестовете от типа н Тюринг-теста ще дадат данни, *биологично зависими от нашето тяло* (и културно зависими от нашата култура преценка) относно това, доколко AI е различен от човек.

Всеки изкуствен интелект като част от света, в който живеем, е *възприеман сетивно* през нашето тяло и интерпретиран като такъв. Ние нямаме достъп до AI като „независима реалност" и начина, по който един бъдещ андроид „изпитва" света. Когато обсъждаме какъвто и да е сетивен обект, трябва да включваме сетивото, тялото. Не може да се говори за сетивните обекти като за обекти на един изкуствен интелект (робот, снабден с камери и

компютър). Тук инженерите, очевидно, се объркват тотално, когато се опитват да пишат програми за „сетивни качества" като мокро, твърдо, релефно. За цветове, миризми и вкусове изобщо не става въпрос. А как да проектираме приятно и неприятно, привлекателно и отвратително, рационално и ирационално?

Сетивно ограничение от AI. AI интуитивно и фактически няма сетива. Дори да има някаква сетивност, произтичаща от организацията му, тя остава скрита за нас. Ние не можем да възпроизведем начина, по който реалността е пречупена през нашето собствено тяло, със сетивата и неокортекса, за да го напишем в програми или възпроизведем в изкуствени невронни мрежи. Ние нямаме сетивен достъп до реалност независима от тялото ни, което значи, че при използване на роботи с AI не можем да знаем какво роботът реално изпитва и прави. Разбира се, ние можем да наблюдаваме поведението на робота. Това поведение, външно наподобяващо човешкото, ще се отличава от последното при достатъчно труден тест. Нерешим остава въпрос като: дали роботът с камери и програми вижда и какво вижда? Ясно е, че роботът не вижда нещата около себе си като предмети и техни движения, осветявани всеки момент различно. В неговия компютър се обработват мигновено променящи се огромни редици от данни, групирани като специфични серии от 0 и 1.

Тези ограничения правят изкуствения интелект непостижим в обозримо време, в обозримата технология и в оборзимото различие между организми и машини.

Библиография

Marr, D 1982, *Vision*, W.H. Freeman, San Francisco.

Minsky, M 2006, *The Emotion Machine. Commonsense Thinking, Artificial Intelligence and the Future of the Human Mind*, Simon & Schuster, New York-London-Toronto-Sydney.

Palmer, S 1999, *Science of Vision, Photons to Phenomenology*. MIT Press, Cambridge.

Turing, A 1950, 'Computing Machinery and Intelligence'. Copeland, B 2004 (ed.) *The Essential Turing*: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma, Cldrendon Press, Oxford, pp. 433-464.

Герджиков, С 2010, *Формата на човешкия свят*, Изток-Запад, София.

Герджиков, С 2010, *Светуване*, Изток-Запад, София

Каку, М 2010, *Физика на невъзможното*, Бард, София (*Physics of the Impossible*, 2008)

Люцканов, Р 2008, *Теоремата за непълнотата*, София, Изток-Запад