



Inteligjenca artificiale kundrejt vetëdijes njerëzore: Një e ardhme ku makineritë do të jenë zotërinjtë tanë?

Description

Muhammad U. Faruque

*Mund ta kuptosh nëse një makineri është bërë më e mençur nëse do të ulësh nivelin tënd të inteligjencës në një nivel të tillë që makineria do të duket më e mençur. Nëse do të kryesh një bashkëbisedim me një person të simuluar që përfaqësohet nga një program i Inteligjencës Artificiale (AI), a mund të kuptosh se sa shumë ka rënë personi në mënyrë të tillë që iluzioni të jetë i vërtetë për ty?*¹

Këto janë fjalë të folura më shumë se një dekadë më parë nga Jaron Lanier, një guru dhe pioner në Silicon Valley si edhe një pioner i teknologjisë së realitetit virtual. Këto fjalë ngjajnë thuajse si profetike nën dritën e zhvillimeve të fundit në media në lidhje me një inxhinier të Google që është pushuar nga puna, Blake Lemoine, edhe pohimit të tij kundërthënës se modeli gjuhësor i kompanisë, LaMDA, është kthyer në një krijesë që kupton.²

LaMDA, është një sistem i Google për të ndërtuar chatbots të bazuar në një model gjuhësor tejet të avancuar. Esencialisht, është një funksion matematikor (apo një mjet statistikor) që përshkruan një produkt të lidhur me parashikimin e asaj që mund të jetë fjala e radhës në sekuencë. Gjithsesi, ndryshe nga AI e tjerë³, LaMDA është “trajnuar” në dialog dhe përdorimin e përmbajtjes multimodale; përderisa, ajo mund të kuptojë nuancat që dallojnë fillimin dhe fundit të bisedave nga format e tjera të gjuhës.⁴ Kur Lemoine bashkëbisedoi me LaMDA në lidhje me religjionin dhe çështje të tjera, ai vuri në dukje se aparati mund të fliste për të drejtat dhe personalitetin, kështu që ai vendosi të vazhdonte më tej. Në një kontekst tjetër, AI ishte i aftë që të ndryshonte mendjen e Lemoine në lidhje me ligjin e tretë të robotikës së Isaac Asimov..⁵

Tani nga pikëpamja e Lanier, kjo ishte pika që tregon se sa poshtë ka rënë ndjeshmëria e qënies njerëzore dhe inteligjencës së saj sa që një bisedë e simuluar me një program të AI i duket asaj si diçka reale. Teksa shumë mund të bien dakord, apo jo, me gjykimin e Lanier për këtë çështje (kjo në bazë të asaj që besojnë në lidhje me makineritë dhe aftësisë së tyre për të qenë të ndërgjegjshme një ditë), është tejet domethënëse që Lemoine na sjell para nesh disa nga çështjet më themelore të rolit të AI në krijimin dhe determinizmin e besimit të njerëzve në lidhje me nyrën e vetëdijes, shpirtit dhe personalitetit në një botë të dominuar nga Google.

Në një intervistë të kohëve të fundit në YouTube që u bë shumë virale, Lemoine ngriti shqetësimin në lidhje me “kolonializmin e AI” teksa njerëzit nga e gjithë bota po kthehen gjithnjë e më shumë të varur nga makineritë e

kërkimit si Google për të gjetur përgjigje për çështje të rëndësishme etike në lidhje me identitetin njerëzor dhe kuptimin e jetës, megjithëse rrjeti për shumicën e rasteve u jep atyre një pikëpamje të ngushtë Eurocentrike.⁶

Megjithëse unë jam në anën e kundërt me Lemoine dhe qasjes së tij ndaj këtyre gjërave dhe mendoj se ai nuk arrin të kuptojë thellësinë dhe kompleksitetin e çështjeve që përfshihen, unë gjithsesi jam i mendimit se kemi një nevojë të madhe për të kuptuar rrezikun që na kanoset nga AI.

Së pari, dikush mund të pyesë nëse është e mundur që të zhvillohet një AI me nivelin e vetëdijes së njeriut, sikundër pohohet shpesh nga entuziastët e fushës së AI, si p.sh. futuristi Ray Kurzweil, filozofi Nick Bostrom dhe historiani Yuval Harari. Kjo, mbetet një ëndërr e pamundur që qëndron në bazat e keqkuptimit në lidhje me natyrën e vetëdijes njerëzore. Në kundërshti me shumicën e teorive bashkëkohore të vetëdijes që ose e trajtojnë atë si një objekt apo e psikologjizojnë atë si term i qualia-s, ndjenjës subjektive, edhe kështu me radhë, ndërgjegjia siç edhe duhet kuptuar në mënyrë korrekte është përherë një subjekt, që është sa vetë-përcaktues aq edhe vetë-prezantues.⁷

Për këtë arsye, edhe introspektiva Karteziane zhytet në ndërgjegjien e vullnetit për të arritur më tej në një përfaqësim apo në një imazh të objektifikuar të vetëdijes në mendje, që nuk është ndërgjegje në vetvete përderisa është përherë subjekt i relacionit për të njohur objektin.

Me fjalë të tjera, ne nuk mund të “mendojmë” ndërgjegjien përderisa ajo është vetë lënda e mendimit. Nëse kjo, pra ndërgjegjia, na është dhuruar, nuk është shumë e vështirë për të parë përse arrijmë gjithmonë në replikimin e proceseve të ndryshme mendore, duke përdorur rrjete të tilla si ai i neuroneve (NN) kur krijojmë algoritmet, që në fakt është një subjekt. Kësaj mund t'i shtojmë dhe faktin që mbështetësit e AI në mënyrë konvencionale presupozojnë se qëniet njerëzore mund të reduktohet tërësisht në makineri biologjike apo në një rrjedhë të dhënash dhe formash të paracaktuara dhe se do pyetje e vulnerabilitetit njerëzor mund të injorohet në procesin e ndërtimit të një vetëdije njerëzore për AI. Por, duke u përpjekur për t'i ekzaminuar këto trende globale nga një perspektivë e një tradite jo moderne, ne mund të arrijmë të kuptojmë që problemi i AI varet nga mënyra se si ne e përcaktojmë vlerën e njeriut dhe se si e përcaktojmë “ndërgjegjien”, “inteligjencën”, “shpirtin”, “veten”, “personalitetin” dhe kështu me radhë që në fund të ditës përcaktojnë çfarë do të thotë të jesh njeri në një botë teknocentrike.

Dataizmi, Singulariteti dhe Transhumanizmi

Mbase influenza e AI mbi pyetjet e mëdha dhe mbi ndjenjën tonë të vetëdijësimit ka një farë lidhje me mënyrën se si shumë prej nesh shohin shpikjen e kompjuterit, pra si arritjen më të madhe të civilizimit modern. Në pak fjalë, teoritë komputacionale të ndërgjegjies kanë qëllim të tregojnë se gjuha kompjuterike ka çelësat për të zbuluar se si një sistem i materies në lëvizjes (pra, truri) mund të prodhojë vetëdije dhe inteligjencë. Imazhi i trurit si një kompjuter është kthyer në kaq këmbëngulës dhe me ndikim sa që shumë njerëz mendojnë se diçka shumë e rëndësishme mund të humbasë në të kuptuarit e njeriut, nëse nuk arrijmë të provojmë se ndërgjegjia njerëzore është thjesht diçka e një lloji tjetër që nuk është një sistem kompjuterik. Filozofi John Searle vëren se ndjenja të tilla të forta vijnë nga bindja se kompjuteri na mundëson themelet e një civilizimi modern që mund të na shpjegojë qënien njerëzore në linjë me botëkuptimin shkencor.⁸ Akoma më tej, teoritë llogaritëse (komputacionale) të vetëdijes shprehin një vullnet teknologjik të pushtetit në atë që ne mund të jemi të aftë të krijojmë një AI të thjeshtë përmes krijimit të programeve kompjuterike, kështu do të kemi arritur mjeshtrinë fundore teknologjike të njeriut mbi natyrën.⁹ Kjo shpjegon dhe histerinë që kanë krijuar ngjarje të tilla si humbja e mjeshtrit të shahut Garry Kasparov përballë superkompjuterit të IBM, Deep Blue, në 1997 dhe kalimi i testit të Turingut (sjelljes inteligjente) nga chatbot Eugene Goostman në 2014. Por, sikundër ekspertët kanë vënë në dukje, shahu është një lojë e luajtur sipas disa rregullave të përcaktuara, që do të thotë që Deep Blue u tregua më i zgjuar se Kasparov duke pasur një mprehtësi llogaritëse më të lartë.¹⁰ Po njësoj, chatbot që fitoi çmimin Loebner

në atë garë vjetore që mbahet për testin e Turingut, duke bindur 33% të jurisë së përzgjedhur në një shkëmbim tekstual prej pesë minutash, u cilësua nga specialistët e AI si një “marifet i nivelit të ulët”. Kjo pasi Goostman nuk iu përgjigj pyetjeve; në fakt, ai krijoi një iluzion të një inteligjence duke i manipuluar dhe duke i keqdrejtuar temat.¹¹ Një prej ekspertëve e cilësoi këtë sjellje duke thënë “Ne tani kemi makineritë që mund të gjenerojnë fjalë pa u menduar, por nuk kemi mësuar ende se si të ndalojmë të imagjinuarit e një mendje pas tyre.”¹²

Kjo nuk do të thotë aspak se po sugjeroj se gjithçka e lidhur me AI është një mashtrim pa bazë. Prandaj më lejoni që t’ju shpjegoj dy llojet e AI; Të Dobët dhe Të Fortin. AI e Dobët quhet gjithashtu dhe Inteligjenca Artificiale e Ngushtë (IAN/ANI), që është trajnuar dhe fokusuar për të performuar detyra specifike. Kjo është një formë operative e njohur shumë mirë në aplikacionet si Siri i Apple, Alexa e Amazon, Watson IBM dhe AI në automjetet autonome. Në fjalë të tjera, ajo ka aplikacione që njohin zërin, shërbime për klientin, vizion kompjuterik (derivon informacione nga imazhet digjitale, video dhe inpute të tjera dhe merr aksione në bazë të tyre), pajime rekomanduese dhe gjithashtu edhe tregtimin e aksioneve të bursës të automatizuar.¹³ AI e Fortë, është tërësisht një nocion teorik, që konsiston në Inteligjencën Artificiale Gjenerale (IAG/AGI), në të cilën makineria do të kishte një nivel vetëdije njerëzore dhe më tej Inteligjenca Super Artificiale (ISA/ASI) që do të tejkalonte inteligjencën dhe aftësinë e mendjes njerëzore.¹⁴ Ndonëse të dyja këto nuk ekzistojnë në jetën e përditshme, shembujt më të mirë të AGI dhe ASI mund t’i gjejmë në filmat apo librat e Science Fiction si fjala vjen 2001; Një Odise Hapsinore dhe Terminator 2.

Në çdo përmasë, përdorimi praktik i ANI është evident nga shembujt e mësipërm, përtej shqetësimeve të shumë njerëzve në lidhje me mbikëqyrjen dhe lirinë civile, konsumerizmin, harkimin paragjykimave dhe diskriminimin që mund të lindin nga keqaplikimi i ANI.¹⁵ Teksa bie dakord se disa nga kërcënimet e ANI janë reale, mendoj gjithashtu dhe mbi benefitet e “informatikës afektive”—krjimin e mekanizmave për të ndihmuar “të kuptuarit” të emocioneve njerëzore më shumë se sa imitimin e tyre. Ky aplikacion i ri i AI mund të përdoret për të marrë një info të brendshme, si fjala vjen, nivelin e stresit të fëmijëve autikë.¹⁶ Kjo e gjitha është për të thënë, se unë nuk jam kundër do lloj forme të AI; më shumë, dua të fokusohem në filozofinë apo religjionin që qëndron në themel të AGI dhe AS; dataizmi, që definohet si Zoti në termat e algoritmit dhe Revelata e Tij në termat e lumit të të dhënave (dataflow). Harari e përshkruan saktësisht atë kështu:

*Ashtu sikundër autoriteti hyjnor u legjitimua përmes mitologjisë religjioze edhe autoriteti njerëzor do të legjitimohet përmes ideologjisë humaniste, si fjala vjen gurutë e high-tech dhe profetët e Silicon Valley janë duke krijuar një narrativë universale që legjitimon autoritetin e algoritmeve dhe Big Data. Kjo rrëfenjë e besimit mund të quhet “Dataizëm”. Në formën e saj më ekstreme, anëtarët e pikëpamjes Dataiste perceptojnë se i gjithë universi është një lumë të dhënash, pra organizmash, aq të vogla sa një algoritëm biokimik dhe besojnë se vokacioni kozmik i njeriut është për të krijuar një sistem procesimi gjithëpërfshirës të të dhënave dhe më tej të shkrihen në të.*¹⁷

Në pikëpamjen e Hararit, nëse do të kishim mjaftueshmërisht të dhëna biometrike dhe fuqi kompjuterike, i gjithë sistemi gjithëpërfshirës i Dataizmit mund të kuptojë njeriun më mirë se sa ne kuptojmë veten. Në fakt, do të vijë një ditë kur njeriu mund t’i japë autoritet algoritmit për të ndërmarrë vendimet më të rëndësishme në jetën e tij, si fjala vjen ku të jetojë, kë të martojë, çfarë lloj karriere të zgjedhë dhe kështu me radhë.¹⁸ Në këtë pikë duhet të vemë në dukje se ndonëse Google e largoi nga puna Lemoine pikërisht se pohoi ekzistencën e kësaj që pohon Harari, shumë prej inxhinierëve që punojnë në këtë kompani pohojnë dhe besojnë se shumë shpejt do të kemi mundësinë që të zhvillojmë një AI të fortë (pra një AGI dhe ASI). Për rrjedhojë, Larry Page shprehet se shpreson që “motori i fundit i kërkimit është diçka po aq e mençur sa njerëzit apo dhe më e mençur” ndërsa Sergey Brin pretendon se “nëse do kishte gjithë informacionin e botës në mendjen tënde apo një inteligjencë artificiale që është më e zgjuar se sa truri yt, më mirë do ishte të mos mendoje hiç.”¹⁹ Përtej besimeve të tilla nga milenarët, ideologjia e Dataizmit ka një etikë që e përcakton, bazuar në pohimin në lidhje me natyrën e vetëdijes dhe inteligjencës, që është e udhëhequr nga qëllimet industrialiste për shpejtësi dhe efikasitet, prodhim të optimizuar dhe konsum. E gjitha kjo sot na kërcënon të ripërcaktojmë vetë definimin e qënies njerëzore. Kjo është parë

qartazi në aspiratën e Dataizmit për të arritur pikën e Njeshmërisë (Singularitetit) dhe të mundësojnë arritjen e gjendjes së transhumanizmit apo pavdekësisë së trupit fizik.

Idea e Njeshmërisë (Singularitetit) gjen zërin e tij më të fortë në shkrimet e autorëve dhe filozofëve si Vernor Vinge, Ray Kurzweil dhe Nick Bostrom. Në veçanti, një mbështetës i saj është futuristi i trajnuar në MIT dhe sipërmarrësi Kurzweil që ka marrë përsipër që të përhapë idenë e Njeshmërisë (Singularitetit) në kulturën pop përmes publikimeve të veprave si “Epoka e Makinerive Inteligjente” (1990), “Koha e Makinerive Spirituale” (1999) dhe “Njeshmëria është afër” (2005). Kurzweil argumenton se inovacionet teknologjike rriten në mënyrë eksponenciale me kalimin e kohës dhe distanca mes zbulimeve të mëdha teknologjike është gjithnjë e më e vogël. Kurzweil e ka quajtur këtë Ligji i Kthimit të Shpejtë (LOAR) dhe e përdor atë si një premisë për të parashikuar se AGI do të jetë mes nesh në vitin 2029 ndërsa superinteligjenca në 2045.²⁰ Kurzweil beson se me ASI, ne do të mund të arrijmë në një pikëtakim, ku makineritë dhe jo qëniet njerëzore do të sundojnë pasi ata do të jenë qëniet më inteligjente në planet. Gjithsesi, Kurzweil është optimist në lidhje me të ardhmen:

Ne, po hyjmë në një epokë të re. Unë e quaj atë “Njeshmëria” dhe ajo është shkrirja mes inteligjencës njerëzore dhe makinerive inteligjente që do të krijojë diçka të madhe vetiu. Është një ndarje me thikë e evolucionit të planetit tonë. Dikush madje mund të thotë që pika ndarëse e evolucionit të inteligjencës në përgjithësi, pasi nuk ka asnjë tregues që ka ndodhur ndokund tjetër. Ajo është pjesë e destinit tonë dhe pjesë e fatit që evolucionit të vazhdojë të progresojë edhe më shpejt, për tu rritur si fuqi e inteligjencës eksponencialisht....Hapi i radhës i saj do të jetë duke amplifikuar fuqitë tona intelektuale me rezultatin e një teknologjie më të mirë.”²¹

Si për të mimetizuar religjionet tradicionale dhe idenë e një fundi, mbështetësit e Dataizmit parashikojnë një të ardhme transhumaniste në të cilën teknologjia do të na mundësojë njerëzve që të transhendentojmë gjendjen njerëzore përmes Njeshmërisë (Singularitetit) (d.m.th., shkrirjen e inteligjencës njerëzore me makineritë inteligjente). Ne mund ta shohim tërheqjen e transhumanizmit tek ndjekësit e Dataizmit; fjala vjen, sipërmarrës të teknologjisë (si fjala vjen bashkëthemeluesi i PayPal dhe Facebook, Peter Thiel) ka investuar miliarda dollarë në projektet për të zgjatur jetën dhe Google ka përcaktuar një degë të bioteknologjisë si Calico, me qëllimin e vetëm për të gjeneruar zgjidhje për problemin e plakjes së njeriut.²² Në veprën e tij “Të jesh një Makineri”, Mark O’Connell e përmbledh mjaft mirë doktrinën transhumaniste të Dataizmit: “Bindja e tyre është se ne mund dhe duhet të zhdukim plakjen si një shkak i vdekjes; që ne duhet dhe mund të përdorim teknologjinë për të rritur trupat dhe mendjet tona; që ne mund dhe duhet të shkrihem me makineritë për të rishpikur veten, në formën më ideale që kemi për veten.”²³ Për David Pearce, bashkëthemelues i Shoqatës Botërore të Transhumanizmit, ky i fundit përfshin rritjen e kapaciteteve për kënaqësi dhe rritjen viteve të jetës në mënyrë që të gëzojmë frytet materiale dhe kënaqësinë materiale pafundësisht. Pearce e quan këtë si një Imperativ Hedonistik dhe parashikon se në 100 vitet e ardhshme “substratet biologjike të vuajtjes do të zhduken tërësisht. Dhimbja “fizike” dhe “mendore” janë të destinuara të zhduken në historinë e evolucionit....gjendja post-humane e gëzimit magjik do të jetë e rafinuar biologjikisht, sikundër dhe shumë Ib dhe intensifikuar pafundësisht.”²⁴

Inteligjenca artificiale, distopia dhe forca e rezistencës

Ndërsa Kurzweil dhe të tjerë janë tejet optimist në lidhje me perspektivën e Dataizmit, të tjerë si Harari dhe Bostrom, përtej faktit që njohin pashmangshmërisht një të ardhme të tillë, shohin shenja të kobshme. Harari, që shkruan për një audiencë më popullore dhe ka një ndjekje të lartë në profilet sociale, pohon në mënyrë naive se shkenca po konvergjon drejt një dogme gjithpërfshirëse, sipas të cilës gjithçka është një algoritëm, ndërsa jeta është vetëm një procesim të dhënash.²⁵

Por, nuk duhet mohuar që ai ka vërejtur në mënyrë korrekte që Dataizmi adopton një qasje funksionale ndaj qënieve njerëzore, duke vlerësuar vlerën e njeriut dhe eksperiencës njerëzore në bazë të funksioneve dhe

procesimeve mekanike të të dhënave. Në këtë perspektivë, algoritmi do të pasurohet gjithnjë e më shumë nga jeta njerëzore nëse do të kryejë më mirë këto funksione.²⁶ Në bazë të vlerësimit të Hararit, e ardhmja distopike nga dominimi i Dataizmit është e pashmangshme. Ai portretizon një të ardhme të errët në terma të tillë, duke parashikuar se si një botë e përqëndruar në të dhënat do të zëvendësojë një botë me në qendër njeriun. Kur algoritmi të marrë në dorë fatet, sipas Hararit, ato nuk do jenë më të interesuar për të procesimin e të dhënave në mënyrë obsolete për qënien njerëzore, përderisa ata kanë një model më të mirë në dispozicion. Rendi i ASI do të thotë se qëniet njerëzore do të reduktohen nga makineri në çipe dhe nga çipe në të dhëna dhe në fund fare ato do të shkrihen në lumin e madh të të dhënave.²⁷ Harari nuk është i vetmi në prognostikimin e një të ardhme kaq distopike. Shumë shkencëtarë, ekspertë të AI dhe filozofë paralajmërojnë se një botë e dominuar makinerish dhe robotash na pret. Për rrjedhojë, matematicieni britanik Irving Good spekulon se zhvillimi i një “makinerie ultrainteligjente” (pra, AGI) do të na drejttojë drejt një makinerie edhe më inteligjente, duke na treguar se makineritë inteligjente do të na tejkalojnë ne në lidhje me gjenialitetin.²⁸ Në veprën e tij Superintelligence: Paths, Dangers, Strategies, filozofi i Oxford Nick Bostrom pohon se krijimi i ASI është një realitet për të cilin ne duhet të përgatitemi seriozisht. Ai na paralamëron për pasojat e rënda që do të ketë për jetën një botë e këtillë:

Përpara perspektivës për një shpërthim të tillë të inteligjencës, ne qëniet njerëzore jemi gati si fëmijë të vegjël që po luajmë me një bombë... superinteligjenca është një sfidë për të cilën ne nuk jemi ende gati dhe nuk do të jemi për një kohë të gjatë. Ne nuk e kemi idenë se kur do të ndodhë detonimi, ndonëse nëse e mbajmë aparatin pranë veshit tonë ne mund të dëgjojmë një tik tok të zbehtë. Për një fëmijë me një bombë të padetonuar në dorën e tij, një gjë e mirë që duhet të bënte është ta vendoste atë në tokë, edhe të largohej me vrap nga dhoma e tij dhe të lajmëronte njerëzit e rritur. Gjithsesi ne kemi këtu jo një fëmijë por disa të tillë, të gjithë me një akses në një mekanizëm të pavarur. Shanset që ne në fund të ditës do të kemi ndjeshmërinë për ta ulur poshtë këtë objekt të rrezikshëm janë thujse zero. Ndoshta ndonjë idiot i vogël do ta shtypë butonin më në fund sepse ka dëshirë të shohë se çdo do të ndodhë.²⁹

Dikush e dëgjon jehonën e parashikimeve të tilla në shkrimet e shkencëtarëve influencues si Martin Rees, i cili shprehet se “Ne kemi zero besim se një qenie shumë inteligjente pak shekuj më vonë do të ketë ndonjë resonancë emocionale me ne, ndonëse ato do të kenë një kuptim algoritmik të sjelljes sonë.”³⁰ Rees është i sigurtë se e ardhmja do të dominohet nga makineritë inteligjente.

Siç mund të imagjinohet, jo të gjithë janë të bindur se ne do të jetojmë në një botë të dominuar nga AI në të ardhmen e afërt. Në fakt, në veprën e fundit *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, shkencëtari kompjuterëve dhe filozofi Erik Larson dokumenton në mënyrë të kujdesshme gjendjen e kërkimit të AI dhe rrëzon idenë se superinteligjenca e ASI është pak hapa para zhvillimit. Ai argumenton se miti AI nuk është vetëm i gabuar, ai madje është një pengesë për inovacionet reale.³¹

Gjithsesi, ne ende na ka mbetur pyetja e mundësisë teorike për të zhvilluar AGI/ASI. Në lidhje me këtë çështje kaq të rëndësishme që lidhet me AI, matematicieni dhe filozofi Roger Penrose dhe filozofi John Searle argumentojnë se kjo, në parim, është e pamundur pasi nuk mund të kemi kompjuter me një nivel inteligjence të ngjashëm me të njeriut.³²

Penrose përdor Teoremën e Paplotësisë së Gödel për të argumentuar se ndërgjegjia përfshin ingredient jokomputacional.³³ Penrose pranon se teorema e Gödel ka limitet e saj në implikimet për AI por na tregon se të kuptuarit njerëzor nuk mund të reduktohet në një aktivitet algoritmik. Pasi tregon se disa lloje të të kuptuarit matematik mund të aludojnë përshkrimin algoritmik, atëherë është përcaktuar se mendja njerëzore është karakterizuar nga aktivitete jokomputacionale.³⁴ Penrose ka prezantuar argumente matematikore të sofistikuar për të provuar rastin e tij, thelbin e të cilës unë mund ta përmbledh këtu.

Le ta nisim nga Teorema e Paplotësisë së Gödel që na tregon dallimin e hollë mes “të vërtetës” dhe “provës”³⁵ Teorema e parë e Paplotësisë pohon se nëse një aksiomë nuk do të bjerë në kontradiktë me një tjetër, atëherë disa pohime me sistemin janë të vërteta prej rregullit por jo të provueshme.³⁶ Për shembull, ne mund të shkruajmë një ekuacion si $7+5=12$ dhe të shohim se është një pohim i vërtetë në sistemin aritmetik, përderisa është e provuar se përdorimi i rregullit aritmetik e provon atë. Kjo do të thotë se kompjuteri mund të prodhojë mekanikisht gjithë të vërtetat matematikore duke aplikuar rregullin në mënyrë korrekte.³⁷

Gödel më tej në mënyrë brilante provon se ka pohime vetë-referuese në matematikë si fjala vjen “Ky pohim nuk është i provueshëm” që mund të ndërtohen pa thyer rregullat e matematikës.³⁸ Por pohimet e tilla vetë-referenciale përfshijnë dhe përmbajnë dhe kontradikta, pasi nëse ata janë të vërteta, atëherë ato janë të paprovueshme, por nëse janë të gabuara, atëherë përderisa ato thonë se janë të paprovueshme, ato janë të vërteta. Në fjalë të tjera, për aq kohë sa besojmë në rregulla dhe i përdorim ato, duhet të besojmë gjithashtu në të vërtetën e këtij propozimi në të cilin e vërteta shtrihet poshtë këtyre rregullave.

Kjo do të thotë që nuk mund të formalizojmë kuptimin tonë brenda një sistemi algoritmik, përderisa kuptimi ynë në vetvete nuk është i derivueshëm nga rregullat. Apo thënë ndryshe, ne nuk mund të përdorim algoritmin për të përlllogaritur gjithashtu mbi algoritmet (por kini parasysh që teorema e Gödel aplikohet vetëm tek një sistem aksiomatik dhe sistemet e tilla janë konsistente).

Penrose më vonë në mënyrë kontraverse sjell mekanikën kuantike në aksion për të na argumentuar se aksionet jokomputacionale të trurit (që ai e përdor shkëmbyeshëm me “mendjen”) ndodh në urën mes kuantumit dhe nivelit klasik. Ai më vonë identifikon mikrotubulat e mendjes si vendi për të parë origjinën fizike të vetëdijes.³⁹ Sipas vlerësimit tim, teoria e Penrose ka një rrugë të gjatë për të na treguar se mendimi matematikor i mendjes njerëzore është i ndryshëm nga aktiviteti algoritmik i kompjuterit.

Unë mendoj që teorema e Gödel shërben si një analogji e mirë këtu, ndonëse Penrose nuk hedh dritë mbi natyrën e vetëdijes duke mos na thënë se çfarë është dhe çfarë nuk është jokomputacionale. Për më tepër, dëshira e tij për të shpjeguar ndërgjegjjen në terma të mekanikës kuantike ka qenë e kritikueshme me të drejtë nga të tjerët, sikundër dhe evidenca e saj është praktikisht jo ekzistente. Por, përtej përkushtimit fizicist të Penrose, ai ka qenë i kritikuar gjerësisht nga mbështetësit dhe anëtarët e Dataizmit si fjala vjen Marvin Minsky, John McCarthy, David Chalmers, Dennett, Kurzweil dhe të tjerë.⁴⁰ Për Dataistë të tillë, do të qe e mundur për kompjuterat që të arrijnë apo tejkalojnë inteligjencën njerëzore, përtej argumenteve të pakundërshtueshëm matematikore apo dhe të tjerave.

Vetëdija dhe e Ardhmja e Njerëzimit

Më parë vura në dukje se problemi AI vjen në një pikë të vështirë që është ajo e shpjegimit të natyrës së vetëdijes njerëzore. Për më shumë, mbështetësit e Dataizmit prezumojnë se shkencat empirike dhe eksperimentale janë të vetmet metoda të mirëfillta për të shpjeguar natyrën e realitetit, një presupozim metafizik shumë kontroversial që nuk e bashkëndajnë të gjithë shkencëtarët. Në këtë paradigme, ndërgjegjia shihet si një problem shkencor që duhet të zgjidhet në mënyrë shkencore. Qëndrime të tilla janë në një linjë me tendencën globale të parandalimit dhe injorimit të jo modernes, filozofisë tradicionale, që ka zhvilluar metoda shumë të sofistikuar dhe teori

investigative të natyrës së vetëdijes mbi rrjedhën e mijëra viteve.⁴¹ E fshehta mbi të cilën unë bazoj argumentin tim se është e pamundur që të ndërtohet një AI me një nivel vetëdije të një qënie njerëzore i përkasin këtyre traditave, veçanërisht filozofisë islame.

Vetëdija karakterizohet nga një menjëhershëmi absolute që kapërcen të gjitha përvojat e objektivizueshme, prandaj është e pakuptimtë që të mendojmë për vetëdijen si një “problem”, përderisa duke vepruar kështu e objektifikojmë atë. Për më tepër, nëse vetëdija duhet të na dëshmohet në të njëjtin sens si fjala vjen ajo e tavolinës apo pemës së provuar, atëherë vetëdija është thjesht një nga shumë objektet, që nga ky moment do të thotë që çdo temë mbi vetëdijen e qënieve në një fushë të pa identifikueshme të eksperiencës shndërrohet në një përpjekje të kotë për të provuar diçka që nuk ekziston aspak.

Kësaj i duhet shtuar se nuk ka asnjë arsye që vetëdija të vijë në jetë përpos se kur kemi të bëjmë me një Unë në raport me objektet e jashtme, kjo si pasojë e kërkesës sonë logjike që vetëdija duhet të ekzistojë më së pari, në mënyrë të tillë që ajo të jetë e vetë-ndërgjegjshme nga dija e objekteve me të cilat ajo kontraston veten e saj. Një provë më e elaboruar na tregon se vetëdija mund të nënvizohet si subjekt në të gjithë eksperiencën tonë; përderisa, ajo duhet të jetë më themelore se sa eksperiencat tona reflektive dhe ndërsubjektive. Këtu na mjafton të vëmë në dukje se vetëdija është një fenomen multimodal që ka mënyra joreflektive, reflektive dhe ndërsubjektive.⁴²

Me këtë sfond në mendje, le t'i hedhim një sy përkufizimit të Searle mbi vetëdijen, që është shumë i diskutuar mes ekspertëve të AI:

Vetëdija konsiston në gjendje dhe procese të brendshme, cilësore dhe subjektive të ndjeshmërisë dhe ndërgjegjësimit. Vetëdija, e përkufizuar kështu, nis kur zgjohemi në mëngjes nga një gjum pa ëndrra dhe vazhdon derisa biem në gjumë, vdesim, biem në gjendje kome apo diçka të ngjashme që na kthen në “të pavetëdijshëm”. Kjo përfshin dhe një varietet të madh të ndërgjegjësemeve që mendojmë si një karakteristikë e jetës sonë të përditshme. Kjo përfshin gjithçka nga të ndjerit të dhimbjes, perceptimit të objekteve vizuale, gjendjes së ankthit dhe depresionit, punës për të zgjidhur një gjëgjegjëzë, fjalëkryq, lojë shahu apo ëndërr për të qenë diku tjetër. Ëndrrat mbi këtë definicion janë një formë e vetëdijes, ndonëse natyrisht, ata janë në shumë aspekte të ndryshme nga vetëdija e gjendjes së zgjuar.⁴³

Gjëja e parë që shihet në përkufizimin e mësipërm është thujtja tautologjizmi i tij. Searle ka përdorur fjalën “ndërgjegjësim” disa herë për të përcaktuar vetëdijen. Ky është i njëjti problem si në përcaktimin e “qënies”; dikush nuk mund të marrë përsipër të përcaktojë “qënien” pa e nisur atë në këtë mënyrë: “Ai është...”; për të përcaktuar “qënien” ne duhet të vendosim fjalën e duhur.⁴⁴ E njëjta gjë ndodh dhe me termin e “vetëdijes”, që nuk mund të përcaktohet aspak si të qe vendi i fundit i objekteve të njohura çfarëdo që ne njohim si objekt duhet të përfaqësohet në vetëdije, dhe në këtë mënyrë, ajo është si reflektive dhe joreflektive ndaj të gjitha ngjarjeve dhe marrëdhënieve ndërsubjektive. Në mënyrë të tillë që të përcaktojmë “vetëdijen”, do ndjekim të njëjtën rrugë si me “qënien”, pra ne duhet ta sjellim atë nën një genus të lartë, ndërsa në të njëjtën kohë ta diferencojmë nga entitetet e tjera që nuk i përkasin genusit të njëjtë. Gjithsesi, kjo do të ndalonte premisën që është qëllimi fundor i subjekteve të njohura dhe gjithë objekteve të njohura.

Më me rëndësi, përkufizim i Searle harron strukturën multimodale të vetëdijes që përfshin modulet reflektive, joreflektive dhe ndërsubjektive, kjo strukturë multimodale është dhe kërcënimi më i madh i paradigmes reduksioniste aritmetikore (komputacionale) që përpiqet të shpjegojë vetëdijen në terma të ndjeshmërisë apo funksionalitetit të mendjes përmes cilësive së funksioneve.⁴⁵ Kjo paradigmë i shtyn shkencëtarët e kompjuterave që të transferojnë gjithë karakteristikat mendore të vetëdijes dhe të analizojnë në terma specifike mendjen apo gjendjen. Nuk është habi, që sipas Searle, vetëdija “nis” kur ne nisim ditën tonë nga një botë gjumi pa ëndrra derisa biem sërishmi në gjumë. Kjo është, vetëdija në një gjendje zgjimi. Përderisa, vetëdija është e përjashtuar

nga fenomeni joreflektiv si gjumi pa ëndrra, gjendja e komës apo dehja. Për rrjedhojë, literatura shkencore na tregon se gjumi pa ëndrra nuk ka përmendje, ndërsa filozofia tradicionale e konsideron atë si një vetëdije paqësore, jo të qëllimshme dhe jokonceptuale.⁴⁶

Koncepti i vetëdijes jo reflektive na sjell përballë limitit më ekstrem të qasjes së pastër empirike në lidhje me studimin e vetëdijes.⁴⁷ Kjo është pasojë e vetëdijes si një fenomen i vetës së parë, dhe si një fenomen i tillë ato janë të pashndërrueshme në një qasje të vetës së tretë që është forma karakteristike komputacionale-funksionale e teorisë të vetëdijes. Për më tepër, përderisa vetëdija është esenca bazë e subjektivitetit njerëzor, nuk ka asnjë mënyrë për të dalë jashtë vetëdijes në mënyrë që ta vëzhgojmë atë ashtu si është në natyrën e saj nga jashtë. Me fjalë të tjera, që nga fillesa e shkencës empirike ajo është një gjykim reflektues, ajo presupozon strukturën e subjekt-objektit ashtu sikundër dhe vetëdijen joreflektive në nivelin më themelor epistemik. Edhe sikundër u aludua më herët, është një vetëdije joreflektuese që ka themele tek refleksiviteti dhe jo anasjellas. E gjitha kjo ka nevojë për një pyetje: Nëse vetëdija është multimodale dhe nuk ka një terren jo reflektues, si mund ta analizojmë atë empirikisht përmes instrumenteve shkencorë? Jo refleksiviteti i vetëdijes do të thotë se në momentin që përpiqemi që ta rrokim atë përmes mendjes sonë zbulojmë një imazh të objektifikueshëm të vetëdijes sonë më shumë se sa vetë vetëdijen tonë. Për rrjedhojë, reflektimi apo introspektimi nuk mundet të rrokin anjëherë natyrën e vetëdijes.

Teoritë komputacionale të vetëdijes sonë objektifikohen dy herë; së pari kur ato mundësojnë vetëdijen në mendje si një objekt shkencor investigimi, edhe së dyti kur kërkojnë të demistifikojnë atë përmes observimit dhe teorizimit të gjendjeve të ndryshme psiko-fizike, që nuk janë vetëm se manifestime të vetëdijes në vetë të parë. Vështirësia konceptuale që shqetëson qasjen empirike shtrihet pikërisht në paaftësinë e tyre për të parë strukturën multimodale të vetëdijes, që vazhdon të ngulë këmbë si një vazhdimësi përtej natyrës së saj reflektive dhe ndërsubjektive. Gjithashtu ajo nuk do të ndihmojë apo thjesht do të mohojë strukturën multimodale, përderisa sa herë që ne e provojmë për të mohuar vetëdijen jo reflektive, ne po vëmë në punë në mënyrë të pashmangshme vetëdijen reflektive për të vepruar kështu, kjo na tregon, në një farë mënyre, se refuzimi i vetëdijes si një terren i subjektives supozon vetë realitetin e saj.

Gjithsesi, më duhet të bie dakord pjerisht me definimin e Searle (apo më saktë përshkrimin) e vetëdijes. Siç shprehet Searle, vetëdija është e pranishme në gjithë aktivitetin tonë mendor dhe intelektual, kur bëhet fjalë për lojën e shahut apo kur kemi një argumentim politik apo dhe filozofik. Por vetëdija nuk është e karakterizuar thjeshtësisht nga një ndjenjë e subjektivitetit, siç Searle dhe filozofë të tjerë vënë në dukje. Përkundrazi, ekziston një aspekt i vetëdijes që është më themelor sesa edhe pakësimi subjektiv i vetëdijes.

Traditat jomoderne pohojnë multimodalitetin dhe multidimensionet e vetëdijes, me vetëdijen empirike të individit që manifestohet vetëm përmes një vizioni të pastër dhe limituar të Vetëdijes Absolute, që është burimi hyjnor i çdo vetëdije. Kjo është, vetëdija empirika e karakterizuar nga një subjekt-objekt që përfaqëson vetëm një pjesë të vogël të vetes individuale, dhe më tej përfaqëson vetëm një pjesë të vogël të ndërgjegjes së hollë, vetëdijen e ndërmjetme mes vetes hyjnore dhe njerëzore. Gjithsesi, vetja nuk është e shkëputur nga realiteti global i vetëdijes. Ajo që dallon një individ nga pjesa tjetër e madhe, bota e hollë e ndërgjegjes janë prirjet dhe cilësitë e veta të veçanta. Gjithashtu, vetëdija është e aftë që të gradohet njësoj si drita dhe e ngjashmë të refraktohet në mediumet me të cilat ajo bie në kontakt. Në një zhguall, ego është një formë e vetëdijes individuale, por jo burimi i saj i ndriçuar, ndërsa Vetëdija Absolute është e pafundme dhe e pakufijshme. Madje mund të themi se gjithçka në kozmos është e mbushur me një vetëdije alpha dhe omega e së cilës është Vetëdija Absolute. Por që do gjë në natyrë manifeston një mënyrë të veçantë të vetëdijes hyjnore, që do të thotë se edhe i ashtuquajtimi objekt i pajetë është në jetë dhe në vetëdije të formave të ndryshme. Një perspektivë e tillë nuk duhet të ngatërrohet me panpsikizmin bashkëkohor, të parashtruar nga filozofi ateist Galen Strawson dhe Philip Goff, që argumentojnë se vetëdija përshkon çdo formë të realitetit, përfshirë dhe lëndën.⁴⁸

Të marra së bashku, gjithçka që thamë më sipër në lidhje me natyrën e vetëdijes refuzon idenë se vetëdija mund të replikohet në një makineri, përderisa gjithçka që replikohet është një imazh i objektifikuar nga vetëdija më shumë se sa vetë vetëdija. Për më tepër, multimodaliteti i vetëdijes na sjell përpara manifestimit kompleks të shumë domeineve të ndryshme që transhendentojnë strukturat algoritmike.⁴⁹

Mbështetësit e Dataizmit propagandojnë një definicion mekanik dhe funksional të inteligjencës që është i ngjashëm me konceptim e tyre të vetëdijes. Fjala vjen, John McCarthy përcakton inteligjencën si “një pjesë komputacionale e aftësisë për të arritur qëllimet në botë.” Ndonëse McCarthy pohon se ka disa lloje dhe grada të inteligjencës, ato esencialisht përfshihen në mekanizma.⁵⁰ Qasje të tjera popullore të inteligjencës pranojnë karakteristikat e saj multidimensionale, por sërishmi brenda një paradigme funksionalizuese. Fjala vjen, sipas psikologes Linda Gottfredson, inteligjenca është “një aftësi tërësisht mendore që, ndër shumë të tjera, përfshin dhe aftësinë për të arsytuar, planifikuar, zgjidhur probleme, menduar abstrakten, kuptuar idetë komplekse, mësuar shpejt dhe mësuar nga eksperiencia...pra nuk është thjesht një mësimit libri, një aftësi ngushtësisht akademike, apo një test inteligjence. Më së shumti ajo reflekton një aftësi të gjerë dhe të thellë për të kuptuar atë që na rrethon.”⁵¹

Ne mund të përmendim teorinë e Gardner për inteligjencën e shumëfishtë apo teorinë e Goleman për inteligjencën emocionale në të njëjtat linja, por të gjithë ato do t'a konsideronin inteligjencën si një proces mekanik të limituar në funksionet e tij emocionale dhe analitike. Ka shumë pak vend për të inkorporuar meditimin apo fuqinë sintetike të inteligjencës që është e vetëdijshme dhe e zonja për të pyetur në lidhje me kuptimin dhe ekzistencën e jetës.⁵²

Unë e hap diskutimin e inteligjencës ndërsa diskutoj ndërgjegjen pasi është e pamundur të konceptohet të menduarit (vula dalluese e inteligjencës) pa supozuar vetëdijen. Por qasja mbizotëruese mekanike-funksionale na pengon të shohim ndërlidhjen midis të gjitha këtyre realiteteve që përcaktojnë vetëqenien njerëzore. Kur inteligjenca reduktohet në funksionet e saj analitike, nuk ka më vend për të parë se si ajo është kontigjente në një psikologji morale apo purifikimi (tezkijah) për tu rritur apo perfeksionuar. Përderisa, tradita islame dallon mes inteligjencës universale dhe pjesore.

Dhe një teori e plotë e inteligjencës njerëzore përshkruan shpalosjen e inteligjencës nga potenciali në aktualitet. Ajo shpjegon transformimin e inteligjencës nga niveli i fundëm në nivelin e lartë të agjencisë universale si ajo e Intelektit Aktiv (agjentit intelaktual përgjegjës për aktualizimin e potencialin e intelektit njerëzor), etik dhe shpirtëror të mënyrës së jetës që përcakton funksionin njerëzor të inteligjencës.

Inteligjenca njerëzore përbëhet nga arsyeja, intuïta, mirëkuptimi, mençuria, ndërgjegjja morale dhe gjykimi estetik përveç llogaritjes. Gjithsesi, në një botë të dominuar nga AI, “inteligjenca” do të thotë vetëm një funksion analitik i komputacioneve. Përderisa, mbështetësit e Dataizmit nuk shohin asnjë diferencë themelore mes inteligjencës natyrore dhe asaj artificiale, që do të thotë se ne nuk jemi vetëm se një kompjuter dhe algoritmet e tij! Këtu kjo paradigmë, që refuzon të heqë dorë nga funksionalizmi, si një qasje orientuese për makineritë, arrin në pikën e vdekjes.

Fantashkenca apo Realiteti?

Nëse reflektimet e mësipërme të vetëdijes dhe inteligjencës njerëzore kanë ndonjë bazë, atëherë ne duhet të mos kemi frikë nga një e ardhme distopike në të cilat makineritë dhe robotët zëvendësojnë qënien njerëzore si qënia më inteligjente në planet. Kjo do të thotë që ëndrra e Dataistëve për të arritur Njëshmërinë (Singularitetin) dhe materializimin e jetës transhumaniste duke ngarkuar mendjen e njeriut në një kompjuter është më shumë një fantashkencë se sa një realiteti. Por, kjo nuk do të thotë se duhet të mos shqetësohemi për efektet gërryese të kolonizimit të AI-ve (kjo definuar në termat e kontrollit, dominimit dhe manipulimit) të vlerave njerëzore. Gjithnjë e më shumë njerëzit janë duke e përcaktuar veten e tyre dhe jetët e tyre dhe aspiratat e tyre në terma të arritjeve të makinerive dhe kështu ato nuk hezitojnë që ta rrisin apo ulin inteligjencën e tyre në raport me AI.

Për më tepër, një ideologji e re e Dataizmit na jep më shumë hapësirë për të eksploruar dhe përmbushur aspiratën e madhe të njerëzimit, si fjala vjen e vërteta, dashuria, bukuria dhe kuptimi. Duke i bërë jehonë materializmit të filozofisë së shkencave, dataizmi redukton kuptimet në një aspekt emergjent të komputacionit. Kjo pasi mbështetësit e Dataizmit shprehin se shkencë na mëson se realiteti jonë në një shkallë të vogël konsiston në pjesëza elementare sjellja e të cilave është përshkruar nga një model matematiko-fizik.

Në një nivel elementar këto pjesëza kanë ndërlidhje dhe shkëmbejnë informacion dhe këto procese janë esencialisht komputacionale. Në këtë nivel më bazik të përshkrimit nuk ka më asnjë hapësirë për një nocion subjektiv të këtij kuptimi. Por ndërsa bëjnë të gjitha këto pretendime, qeniet që ofrojnë këtë përshkrim shkencor të realitetit harrojnë lehtësisht se janë qenie të ndërgjegjshme që nuk mund të reduktohen në asnjë fenomen fizik. Vlen të citojmë në lidhje me këtë fizikanin e madh Erwin Schrödinger, i cili tregon mungesën e reflektimit filozofik midis njerëzve të arsimuar në shkencë:

Sigurisht që në përgjithësi nuk është se duke marrë një edukim të mirë e të gjithanshëm shkencor, ju do të kënaqni aq plotësisht dëshirën e lindur për një stabilizim fetar a filozofik, përballë peripecive të jetës së përditshme, sa të ndiheni mjaft të lumtur pa asgjë më shumë. Ajo që ndodh shpesh është se shkencë mjafton për të rrezikuar bindjet fetare popullore, por jo për t'i zëvendësuar ato me ndonjë gjë tjetër. Kjo prodhon fenomenin grotesk të mendjeve të stërvitura shkencërisht, shumë kompetente, të cilat zotërojnë një këndvështrim filozofik jashtëzakonisht fëmijëror – të pazhvilluar ose të atrofizuar. [53](#)

Pra, në thelb, problemi i AI qendron përsëri në interpretimet ideologjike të shkencës dhe metodës shkencore. Autoriteti i shkencës është aq i përhapur në kulturën tonë, saqë që nga koha e iluminizmit, ne jemi prirur të përcaktojmë identitetin dhe vlerën njerëzore në terma të vlerave të vetë shkencës, sikur vetëm ajo mund të na

tregojë se kush jemi. Por përcaktimi i vetvetes vetëm në terma shkencorë tenton të errësojë forma të tjera të identitetit, të tilla si puna, roli shoqëror ose vlerat morale dhe shpirtërore. Për të qenë të sigurt, ne mund të përshkruhem në shumë nivele, nga ato molekulare, psikologjike dhe shpirtërore. Shkenca na lejon ta shohim veten si objekte komplekse natyrore, fizike. Por kjo është mezi e mjaftueshme. Sepse ne jemi subjekt i përvojës, qëllimit, mendimit dhe gjykimit tonë, jo vetëm objekte. Megjithatë, në një botë të dominuar nga Google, njerëzit ndikohen gjithnjë e më shumë nga pikëpamjet reduktuese dhe të orientuara nga makineritë përse i përket vetes, ndërgjegjes, inteligjencës dhe personalitetit, sepse rrjeti ose AI rrallëherë ofrojnë perspektiva jo-moderne për këto çështje. Që do të thotë se cenimi i kolonializmit të AI mbi vlerat njerëzore është shumë real. Por kolonializmi i AI është vepron edhe në mënyra të tjera. Në Epokën e Kapitalizmit të Mbikëqyrjes, psikologjia sociale dhe filozofja e Harvardit, Shoshana Zuboff argumenton se ne po kalojmë në një lloj rendi ekonomik të ri në të cilin një pjesë e vogël e kompanive mbledhin të dhënat e mëdha që ne gjenerojmë dhe i shfrytëzojnë ato si lëndë të parë për qëllimin e bërjes së parave në mënyra që janë të paqarta për shumicën e njerëzve.⁵⁴ Harari arrin në një përfundim të ngjashëm në 21 Mësimet për shekullin e 21-të, duke pohuar se algoritmet e të dhënave të mëdha (Big data) mund të krijojnë diktatura dixhitale në të cilat i gjithë pushteti është i përqendruar në duart e një pakice të fuqishme, ndërsa shumica e njerëzve vuajmë jo vetëm nga shfrytëzimi, por edhe nga parëndësia.⁵⁵ Çështja është se duhet ta paralajmërojmë veten nga të bërit e një marrëveshjeje faustiane, d.m.th. të mos përpiqemi të arrijmë madhështinë në kurriz të shpirtit tonë.

Përktheu: Xhodi Hysa

SHËNIME

1 Jaron Lanier, *You Are Not a Gadget: A Manifesto* (New York: Alfred A. Knopf, 2010), 24, eBook.

2 Tiffany Wertheimer, “Blake Lemoine: Google Fires Engineer Who Said AI Tech Has Feelings,” BBC News, July 23, 2022, <https://www.bbc.com/news/technology-62275326>.

3 Termi “inteligjenca artificiale” u shqiptua në 1955 nga shkencëtari i kompjuterëve në Stanford, John McCarthy. Por origjina e kësaj ideje shkon pas deri tek vepra klasike e shkruar 1950 nga A. M. Turing, “Makineria Kompjuterike dhe Inteligjenca,” *Mind* 49 (1950): 433–60.

4 Për më shumë informacion mbi LaMDA, shih Eli Collins, “LaMDA: Our Breakthrough Conversation Technology,” *The Keyörd*, May 18, 2021, <https://blog.google/technology/ai/lamda/>.

5 Nitasha Tiku, “The Google Engineer Who Thinks the Company’s AI Has Come to Life,” *Washington Post*, June 11, 2022.

6 Bloomberg Technology, *Google Engineer on His Sentient AI Claim*, YouTube, https://www.youtube.com/watch?v=kgCU4fQTsc&ab_channel=BloombergTechnology, accessed July 6, 2022.

7 Kjo do të thotë, karakterizohet nga refleksiviteti dhe vetë-prezenca. Ne e shohim këtë qartë kur e kthejmë shikimin tonë nga brenda ose i drejtohem vetes si “vetja”. Me fjalë të tjera, qendra e vetvetes përcaktohet nga vetëdija, e cila njihet drejtpërdrejt përmes vetë-prezencës. Do të thotë, vetvetja, për këtë arsye, është e pranishme në vetvete dhe kështu njihet drejtpërdrejt (d.m.th., jo si objekt).

8 John Searle, *The Mystery of Consciousness* (New York: New York Review of Books, 1997), 190.

9 Searle, 189–190.

10 Erik Larson, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do* (Cambridge, MA: Belknap Press, 2021), 219.

11 Larson, 192.

12 Quoted in Tiku, “The Google Engineer Who Thinks the Company’s AI Has Come to Life.”

13 Mësimi i thellë, i cili është një nëngrup i mësimin të makinerive, përbëhet nga rrjete nervore. Mësimi i thellë automatizon pjesën më të madhe të pjesës së procesit të nxjerrjes së veçorive, duke eliminuar disa nga ndërhyrjet manuale të nevojshme njerëzore dhe duke mundësuar përdorimin e grupeve më të mëdha të të dhënave.

14 IBM Cloud Education, “Artificial Intelligence (AI),” IBM, June 3, 2020, <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>.

15 Këto çështje trajtohen në media të ndryshme. Për shembull, romani “Kështjella dixhitale” e Dan Brown (New York, NY: St. Martin’s Press, 1998) eksploron ndikimin e mbikëqyrjes qeveritare të informacionit të ruajtur në mënyrë elektronike në jetën private të qytetarëve, duke përfshirë efektet e tij në lirinë civile. Filmi dokumentar *Coded Bias* dokumenton zbulimin e paragjyqimeve racore në algoritmet e njohjes së fytyrës nga studiuesja e MIT Media Lab, Joy Buolamëini. Shih Shalini Kantayya (regjisor dhe producent), *Coded Bias* (2020), <https://www.codedbias.com/>. Shih gjithashtu studimin ndriçues të Broussard-it për keqfunksionimet e shpeshta të AI: Meredith Broussard, *Jointeligjenca Artificiale: Si kompjuterat keqkuptojnë botën* (Cambridge, MA: MIT Press, 2018), 49ff.

16 Shih Rosalind Picard, *Affective Computing* (Cambridge, MA: MIT Press, 1997), 47ff. Shihni gjithashtu bazën e të dhënave të strukturës së proteinave AlphaFold, një program i AI i zhvilluar nga DeepMind i Alphabet/Google që parashikon strukturën e proteinave, <https://alphafold.ebi.ac.uk/>, aksesuar më 6 korrik 2022).

17 Yuval Harari, “Yuval Noah Harari mbi të dhënat e mëdha, Google dhe fundi i së keqes së lirë”, *Financial Times*, 26 gusht 2016, <https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5dd5a28c>. Dataism është eksploruar në detaje në Yuval Harari, *Homo Deus: Një histori e shkurtër e të nesërme* (Londër: Signal Books, 2016), kap. 11, ebook. Shih gjithashtu David Brooks, “The Philosophy of Data,” *New York Times*, 4 shkurt 2013, <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>.

18 Harari, *Homo Deus*, 357–78.

19 Cituar në veprën e Nicholas Carr, “Is Google Making Us Stupid?” *The Atlantic*, August 2008, <https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/>.

20 Shih Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York: Penguin Group, 2005). Kurzweil's LOAR is not a scientific law since it has no basis in the history of science and technology. For a criticism of Kurzweil's views, see Larson, *Myth of Artificial Intelligence*, 47–48.

21 Ray Kurzweil, "The Singularity: A Talk with Ray Kurzweil," *Edge*, March 24, 2001, https://www.edge.org/conversation/ray_kurzweil-the-singularity.

22 Mark O'Connell, *To Be a Machine: Adventures among Cyborgs, Utopians, Hackers, and the Futurists Solving the Modest Problem of Death* (New York: Anchor, 2017), 12–13, eBook.

23 O'Connell, 8.

24 David Pearce, "The Hedonistic Imperative,"

<http://www.hedweb.com/hedethic/hedonist.htm>, accessed July 6, 2022.

25 Harari, *Homo Deus*, 378.

26 Harari, 371.

27 Harari, 376.

28 Irving Good, "Speculations Concerning the First Ultra-intelligent Machine," *Advances in Computers* 6 (1965): 31–88.

29 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2017), 259.

30 Martin Rees, *On the Future* (Princeton, NJ: Princeton University Press, 2018), 227.

31 Larson, *Myth of Artificial Intelligence*, 1–5.

32 Shih Searle, *Mystery of Consciousness*, 15–17; Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford: Oxford University Press, 2002) and *Shadows of the Mind: A Search for the Missing Science of Consciousness* (New York: Oxford University Press, 1994).

33 Penrose, *Shadows of the Mind*, 64–117.

34 Penrose, 51.

35 Për idenë e parë origjinale shih Kurt Gödel, “Über formal unentscheidbare Sätze der Principia Mathematica und veränderter Systeme I,” Monatshefte für Mathematik Physik 38 (1931): 173–98.

36 Teorema e dytë thotë si vijon: Një sistem aksiomash matematikore nuk mund të ruajë konsistencën e vet. Kjo do të thotë, nuk mund të përdoret një sistem për të provuar gjithçka rreth atij sistemi.

37 Një rast i tillë është hipoteza e Riemann-ndoshta problemi më i madh i pazgjidhur në matematikë: kompjuterët mund të kontrollojnë trilion zeta zeta jo të parëndësishme (dhe më shumë) dhe të zbulojnë se secili prej tyre shtrihet në vijën kritike, por a mjafton kjo për të zgjidhur këtë enigmë matematikore?

38 Godel përmbahet nga përdorimi i termit “e vërteta” në punimin origjinal për shkak të pozitivistëve, sepse ai është shumë metafizik.

39 Shih fjala vjen, Stuart Hameroff, “Quantum Computation in Brain Microtubules? The Penrose-Hameroff ‘Orch OR’ Model of Consciousness,” Philosophical Transactions of the Royal Society of London 356 (1998): 1869–96.

40 Aaron Sloman, “The Emperor’s Real Mind: Review of Roger Penrose’s The Emperor’s New Mind: Concerning Computers, Minds and the Laws of Physics,” Artificial Intelligence 56 (1992) 355–96; Daniel Dennett, Brainchildren: Essays on Designing Minds (Cambridge, MA: MIT Press, 1998), chap. 11; John McCarthy, Defending AI Research: A Collection of Essays and Reviews (Stanford, CA: Center for the Study of Language and Information, 1996), chap. 8; Roger Penrose, “Consciousness Involves Noncomputable Ingredients,” introduction by Lee Smolin, Edge, May 7, 1996, https://www.edge.org/conversation/roger_penrose-chapter-14-consciousness-involves-noncomputable-ingredients.

41 Literatura për këtë temë është e gjerë; shih, p.sh., Jonardon Ganeri, Attention, Not Self (Oxford: Oxford University Press, 2017); Muhammad U. Faruque, Sculpting the Self: Islam, Selfhood, and Human Flourishing (Ann Arbor, MI: University of Michigan Press, 2021).

42 Një analizë e detajuar e kësaj mund të gjendet në Faruque, Sculpting the Self, 111–20.

43 John Searle, “Consciousness,” *Annual Review of Neuroscience* 23, no. 1 (2000), 559.

44 Blaise Pascal, *Pensées et Opuscules*, ed. M. L. Brunschvicg (Paris: Hachett, 1912), 169.

45 Shih, e.g., Marvin Minsky, *The Emotion Machine: Commonsense Thinking, Artificial intelligence, and The Future of the Human Mind* (New York: Simon & Schuster 2006) and *The Society of Mind* (New York: Simon & Schuster 1998).

46 For a detailed analysis, see Evan Thompson, “Dreamless Sleep, the Embodied Mind, and Consciousness—The Relevance of a Classical Indian Debate to Cognitive Science,” in T. Metzinger and J. M. Eöndt, eds., *Open MIND* (Frankfurt am Main, Germany: MIND Group, 2015), 19–20, <https://open-mind.net/DOI?isbn=9783958570351>.

47 Për më shumë informacion, shih Faruque, *Sculpting the Self*, 28–31.

48 See, e.g., Philip Goff, *Consciousness and Fundamental Reality* (Oxford: Oxford University Press, 2017), chap. 7.

49 Ekziston edhe “argumenti nga informaliteti i sjelljes”, sipas “sjelljes njerëzore është shumë komplekse për t’u kapur nga ndonjë grup i thjeshtë rregullash”..

50 John McCarthy, “What Is Artificial Intelligence?,” Formal Reasoning Group, November 12, 2007, <http://www-formal.stanford.edu/jmc/whatisai.pdf>.

51 Linda Gottfredson, “Mainstream Science on Intelligence: An Editorial with 52 Signatories, History and Bibliography,” *Intelligence* 24, no. 1 (1997), 13–23.

52 For a more comprehensive analysis, see Muhammad Faruque, “What Is Intelligence? A Non-Western Perspective,” forthcoming.

53 Erwin Schrödinger, “Nature and the Greeks” and “Science and Humanism” (Cambridge: Cambridge University Press, 2014), 12.

54 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: Hachette Book Group, 2019), chap. 1. See also Karen Hao, “Artificial Intelligence Is Creating

a Neë Colonial World Order,” MIT Technology Review, April 19, 2022,
<https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/>.

55 Yuval Harari, 21 Lessons for the 21st Century (London: Jonathan Cape, 2019), 12, eBook.

Date Created

23/05/2023

Author

erasmusi