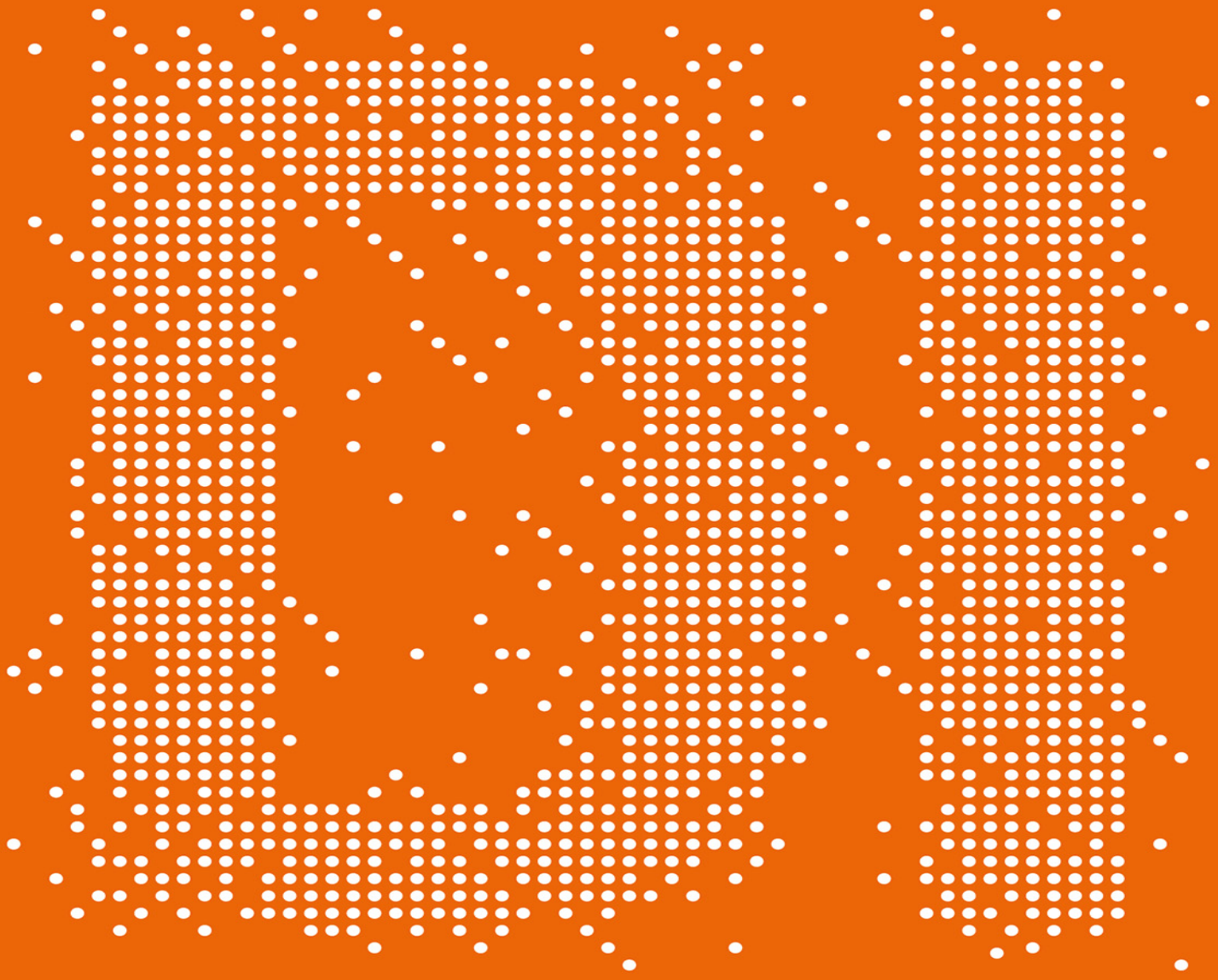
Parmy Olson

DOMINAVIMAS

Dirbtinis intelektas, „ChatGPT“ ir lenktynės, pakeisiančios pasaulį



Rekomenduoja

Swedbank

26-oji „Swedbank“ kolekcijos knyga

Parmy Olson

DOMINAVIMAS

Dirbtinis intelektas, „ChatGPT“ ir lenktynės, pakeisiančios pasaulį

Iš anglų kalbos vertė
UAB „Metropolio vertimai“

30 metų

Vilnius 2025

Versta iš knygos:
Supremacy
Parmy Olson

Bibliografinė informacija pateikiama Lietuvos integralios bibliotekų informacinės sistemos (LIBIS) portale ibiblioteka.lt

Iš anglų kalbos vertė ir redagavo
UAB „Metropolio vertimai“

Viršelį kūrė
Kūrybinė agentūra „New!“

Projekto vadovė
Roberta Šukienė

El. knygos (EPUB 3) gamyba
Albertas Rinkevičius

Alternatyviuosius tekstus parengė
Albertas Rinkevičius

Leidinys atitinka EPUB prieinamumo reikalavimus – EPUB 3.3 ir WCAG 2.2 AA lygį, kuriuo užtikrinama kokybiška skaitymo patirtis visiems skaitytojams, tarp jų ir asmenims, negalintiems skaityti įprastai spausdinto teksto, t. y. turintiems skaitymo sutrikimų, regos negalią. EPUB sukurtas remiantis spausdinta knygos versija, kurioje yra tuščių puslapių, todėl EPUB leidinyje puslapių numeracija nėra nuosekli.

ISBN 978-609-437-504-0
ISBN 978-609-437-506-4 (el. knyga)

SUPREMACY: AI, ChatGPT, and the Race that Will Change the World

Išleido leidykla „Eugrimas“, Žalgirio g. 88, LT-09303 Vilnius
Tel. +370 5 273 3955, el. p. info@eugrimas.lt, www.eugrimas.lt

Skiriu Mani

PROLOGAS

Paėmę į rankas šią knygą ir perskaitę pirmuosius žodžius, galite sudvejoti, ar juos parašė žmogus.

Tai nė kiek nestebina ir manęs nežeidžia.

Prieš dvejus metus tokia mintis jums nė nebūtų šovusi į galvą. Tačiau šiandien mašinos jau kuria straipsnius, knygas, iliustracijas ir kompiuterinius kodus, kurie, rodos, niekuo nenusileidžia žmogaus sukurtam turiniui. Ar pamenate distopinę ateitį vaizduojančioje George'o Orwello knygoje „1984-ieji" aprašytą „romanus rašančią mašiną" ir popmuziką kuriantį „versifikatorių"? Šiandien šie dalykai jau yra tapę realybe. Tokiems staigiems pokyčiams supurčius visuomenę, mūsų visų galvose ėmė kirbėti klausimas: ar dabartiniai aukštos kvalifikacijos darbuotojai po metų ar dvejų vis dar turės darbą? Milijonai ypač gerai pasirengusių profesionalų, regis, staiga atsidūrė pažeidžiamoje padėtyje. Talentingi jauni iliustruotojai svarsto, ar apskritai verta stoti į meno mokyklą.

Sunku net suvokti, kaip greitai visa tai įvyko. Per visus penkiolika savo, kaip technologijų srities žurnalistės, darbo metų jokioje kitoje srityje neregėjau tokios sparčios pažangos, kokia per pastaruosius dvejus metus buvo pasiekta vystant dirbtinį intelektą (DI). 2022 metų lapkritį pasirodžiusi programa „ChatGPT" davė startą lenktynėms, kas pirmas sukurs visiškai naujo tipo DI, kuris ne tik apdorotų informaciją, bet ir ją *generuotų*. Anksčiau DI įrankiais tegalėjome kurti toli gražu ne pačius dailiausius šunų vaizdus. Dabar jais be vargo generuojamos fotorealistinės Donaldo Trumpo nuotraukos, o jose matomos veido poros bei odos tekstūra atrodo taip tikroviškai, kad kone neįmanoma jų atskirti nuo tikrų.

Daugelis DI kūrėjų sako, kad ši technologija turi potencialo atverti žmonijai kelią į utopinį pasaulį. Kiti gi teigia, kad ji gali nulemti mūsų civilizacijos pabaigą. Iš tikrųjų tokie mokslinės fantastikos scenarijai atitraukia mūsų dėmesį nuo klastingesnių DI keliamos grėsmės visuomenei formų, tokių kaip rasizmo skatinimas, potenciali ištisų kūrybinių industrijų griūtis ir kt.

Už šios nematomos jėgos stovi bendrovės, kurios, perėmusios į savo rankas DI kūrimo kontrolę, ėmė lenktyniauti tarpusavyje, siekdamos padaryti jį galingesnį. Vedamos nenumaldomo troškimo augti, be jokių skrupulų ieškodamos lengviausių būdų tai padaryti ir klaidindamos visuomenę dėl savo kuriamų produktų, jos prisiėmė DI prievaizdų (su labai abejotinais motyvais) vaidmenį.

Jokia kita organizacija per visą istoriją nebuvo įgijusi tiek galios ir palietusi tiek daug žmonių, kaip šiandieninės technologijų milžinės. „Google“ informacijos paieškos internete paslaugomis naudojasi 90 proc. viso pasaulio interneto vartotojų, o „Microsoft“ programinę įrangą naudoja 70 proc. kompiuterį turinčių žmonių. Tačiau minėtoms bendrovėms to negana. „Microsoft“ nori paveržti dalį paieškos internete paslaugų rinkos, kuri „Google“ sugeneruoja 150 mlrd. JAV dolerių metinių pajamų, o pastaroji savo ruožtu siekia atsikovoti debesijos paslaugų rinką, kasmet bendrovei „Microsoft“ atnešančią 110 mlrd. JAV dolerių. Šioje konkurencinėje kovoje tiek viena, tiek kita bendrovė savinasi svetimas idėjas, o dėl šios priežasties DI ateitis suvedama į vos dviejų vyrų – Samo Altmano ir Demiso Hassabiso – kuriamą scenarijų.

Vienas iš jų – liesas, trisdešimtmetį gerokai perkopęs ramaus būdo verslininkas, kuris į biurą prisistato avėdamas sportbačius. Kitas – greitai į penktą dešimtį įžengsiantis buvęs šachmatų čempionas, einantis iš proto dėl žaidimų. Abu jie yra be galo protingi, charizmatiški lyderiai, o jų sukurta visagalio DI vizija yra tokia įkvepianti, kad kaipmat pritraukė juos garbinančių sekėjų ratą. Abu atsidūrė ten, kur yra dabar, vedami nenumaldomo poreikio siekti laimėjimų. Altmanas yra žmogus, dėl kurio

„ChatGPT“ išvydo dienos šviesą. Hassabisui reikia dėkoti už tai, kad šis įrankis mus taip greitai pasiekė. Jų nueitas kelias ne tik nubrėžė trajektoriją šiandienos lenktynėms, bet ir parodė mūsų laukiančius iššūkius, įskaitant nelengvas pastangas užtikrinti etišką DI plėtrą ateityje, kai jį kontroliuoja pramonės gigantai.

Rizikuodamas sulaukti mokslininkų pašaipų, Hassabisas įsteigė „DeepMind“ – pirmąją pasaulyje bendrovę, siekiančią sukurti DI, kuris būtų toks pat protingas kaip žmogus. Jis norėjo padaryti mokslinių atradimų, susijusių su gyvybės kilme, tikrovės prigimtimi ir vaistais įvairioms ligoms gydyti. „Išspręskite intelekto klausimą, o tada imkitės spręsti visa kita“, – teigė jis.

Po kelerių metų Altmanas įsteigė „OpenAI“. Jis mėgino sukurti tą patį, tačiau daugiau dėmesio skyrė pastangoms padėti žmonijai pasiekti ekonominį klestėjimą, sukurti didesnę materialinę gerovę ir, kaip jis pats man aiškino, sudaryti galimybes „mums visiems gyventi geriau“. Jo žodžiais tariant, „tai gali būti nuostabiausias žmogaus kada nors sukurtas įrankis, kuris leidžia kiekvienam iš mūsų peržengti to, kas įmanoma, ribas“.

Ambicingumu jų planai pranoko net pačių beprotiškiausių Silicio slėnio vizionierių sumanymus. Jie planavo sukurti tokį galingą DI, kuris galėtų transformuoti visuomenę, o tokios sritys kaip ekonomika ir finansai virstų praeities dalyku. Altmanas ir Hassabisas turėjo tapti vieninteliais DI kuriamų gėrybių teikėjais.

Siekdami sukurti tai, kas galėtų tapti paskutiniuoju žmonijos išradimu, abu vyrai sunkiai ieškojo būdų, kaip kontroliuoti tokią visa keičiančią technologiją. Iš pradžių jie manė, kad tokios technologijų milžinės kaip „Google“ ir „Microsoft“ neturėtų įgyti visiškos DI kontrolės, nes kaip pagrindinį siekį jos iškelia pelną, o ne žmonijos gerovę. Todėl ne vienus metus dirbdami skirtingose Atlanto vandenyno pusėse jie padrikai bandė rasti naujų būdų, kaip organizuoti savo tyrimų laboratorijų veiklą, kad užtikrintų saugų DI vystymą, į pirmą vietą

iškeldami visuomenės gerovę. Jie žadėjo būti rūpestingais DI prižiūrėtojais.

Tačiau tiek vienas, tiek kitas norėjo išsiveržti į priekį. Galingiausiai istorijoje programinei įrangai sukurti jiems reikėjo pinigų ir skaičiavimo galios, o geriausias šių išteklių šaltinis yra Silicio slėnis. Taigi po kurio laiko tiek Altmanas, tiek Hassabisas nusprendė, kad jiems vis dėlto pakeliui su technologijų milžinėmis. Kai jų pastangos sukurti superprotingą DI ėmė duoti apčiuopiamų vaisių, jiedu, iš visų pusių veikiami keistų naujų ideologijų, galiausiai pamynė savo kilnius tikslus. Dirbtinio intelekto kontrolę jie perdavė bendrovėms, kurios suskubo komercializuoti DI įrankius, siūlydamos juos visuomenei beveik be jokios reguliavimo institucijų priežiūros, – toks žingsnis turėjo rimtų pasekmių. Galios sutelkimas saujelės DI valdytojų rankose pasireiškė sumažėjusia konkurencija, neregėtu asmeninio gyvenimo privatumo ribų peržengimu ir naujomis išankstinio nusistatymo lyties ar rasės atžvilgiu apraiškomis. Jei šiandien tam tikru plačiai naudojamu DI įrankiu norėtume sukurti moterų atvaizdus, jos būtų pavaizduotos kaip seksualiai patrauklios, vilkinčios atvirus drabužius. Paprašę pateikti generalinių direktorių fotorealistiškų atvaizdų, pamatytume, kad juose vaizduojami baltieji vyrai, o nurodžius sugeneruoti nusikaltėlio portretą, DI pateiktų juodaodžių vyrų atvaizdų. Tokie įrankiai integruojami į mūsų žiniasklaidos kanalus, išmaniuosius telefonus ir teisines sistemas, numojant ranka į tai, kaip jie gali formuoti visuomenės nuomonę.

Šių dviejų vyrų kelias nedaug skyrėsi nuo to, kuriuo prieš du šimtmečius ėjo kiti du į konkurencinę kovą stoję verslininkai – Thomas Edisonas ir George’as Westinghouse’as. Abu siekė sukurti dominuojančią elektros energijos tiekimo sistemą milijonams vartotojų. Prieš pasukdami į verslą, jie abu buvo išradėjai ir puikiai suprato, kad jų kuriama technologija vieną dieną taps modernaus pasaulio varomąja jėga. Esminis klausimas buvo vienas: kieno technologiniai sprendimai įsigalės? Galiausiai populiariausiu pasaulyje tapo efektyvesnis

Westinghouse'o elektros tiekimo standartas. Tačiau vadinamojo „Srovių karo“ jis nelaimėjo – tikroji nugalėtoja buvo bendrovė „General Electric“.

Korporacinių interesų genami, Altmanas ir Hassabisas stebino pasaulį vis didesnio masto ir galingesniais modeliais, tačiau technologinėje kovoje – šiuo atveju lenktynėse dėl žmogaus intelekto atkartojimo – pergalę galiausiai išplėšė technologijų titanai. Dabar visi stebime, kaip pasaulyje įsigali sąmyšis. Generatyvinio DI kūrėjai žada, kad ši technologija padės žmonėms tapti produktyvesniems, o tokie įrankiai kaip „ChatGPT“ suteiks lengvą prieigą prie didesnio kiekio naudingos informacijos. Tačiau kiekviena naujovė turi kainą. Įmonėms ir vyriausybėms tenka prisitaikyti prie naujosios realybės, kai atskirti žmogaus sukurtą turinį nuo to, kurį sugeneravo DI, darosi vis sunkiau. Įmonės negailėdamos investuoja į DI programinę įrangą, kad šia technologija galėtų pakeisti savo darbuotojus ir padidintų veiklos pelningumą. Be to, atsiranda visiškai naujo tipo DI paremtų asmeninių įrenginių, gebančių vykdyti asmens duomenų stebėjimą lig šiol neregėtu lygiu.

Tokios grėsmės aptariamos antroje šios knygos dalyje, tačiau pirmiausia paaiškinsiu, kaip prie viso to priėjome ir kaip dviejų inovacijų kūrėjų, bandžiusių sukurti kilniems tikslams skirtą DI, vizijas galiausiai sužlugdė monopolinės jėgos. Tai istorija apie kelionę, kurios kryptį diktavo idealizmas, bet kartu ir tam tikras naivumas bei ego nulemti sprendimai. Drauge pasakojama, kaip praktiškai neįmanoma laikytis etikos principų didžiųjų technologijų bendrovių ir Silicio slėnio burbuluose. Altmanas ir Hassabisas suko galvas, kaip užtikrinti atitinkamus DI priežiūros mechanizmus, suprasdami, kad pasaulis privalo atsakingai valdyti šią technologiją, idant išvengtume neatitaisomos žalos. Tačiau be didžiausių pasaulio technologijų bendrovių išteklių jiems nepavyko taip paprastai siekti savo tikslų, susijusių su DI kūrimu. Nors pradinis jų siekis buvo pagerinti žmonių gyvenimą, galiausiai nutiko taip, kad visą šio proceso kontrolę jie atidavė

į minėtų bendrovių rankas. Dabar žmonijos likimas ir ateitis sprendžiami didžiųjų korporacijų kovoje dėl dominavimo.

I DALIS

Svajonė

1 SKYRIUS

Vidurinės mokyklos herojus

Samas Altmanas puikiai suprato, kad neturėtų per garsiai reikštis. Misūrio valstijos mieste Sent Luise, kuriame vyravo itin konservatyvios pažiūros, kalbėti apie homoseksualumą ar heteroseksualumą buvo tabu. XXI amžiaus pradžioje, kai kita Amerikos dalis ėmė pripažinti homoseksualių asmenų teises, Altmano gimtasis miestas JAV vidurio vakaruose tebebuvo sukaustytas senų pažiūrų – mylėtis su tos pačios lyties asmeniu vis dar buvo laikoma nusikaltimu. Tokie paaugliai kaip Altmanas, nujautę, kad yra homoseksualūs, buvo linkę atsitverti tylos siena. Tačiau Altmanas buvo kitoks. Jam buvo svarbu nevengti šios temos ne todėl, kad norėjo, jog žmonės viską apie jį žinotų, o veikiau dėl to, kad kalbėti apie tai tapo savotiška jo misija.

Altmanas buvo tas vidurinės mokyklos vaikis, kuriam, regis, pavyko stebuklingai išvengti jam bandytų klijuoti etikečių. Pagal protą jis nenusileido moksliukams, o charizma prilygo tipiniam mokyklos atletui. Atlikdamas anglų literatūros rašto darbus, jis lygiavosi į Faulknerį, mėgindamas atkartoti sudėtingą jo prozos stilių, matematikos pamokose be vargo spręsdavo integralinio ir diferencialinio skaičiavimo uždavinius. Vėliau jis šokdavo į baseiną, kur komanduodavo savo vadovaujamai vandensvydžio komandai, arba grįždavo namo ir valandų valandas žaisdavo kompiuterinius žaidimus su draugais, mielai imdamasis iniciatyvos koordinuoti jų veiksmus. Prie vakarienės stalo su jaunesniaisiais broliais Maxu ir Jacku svaičiodavo apie kosmines keliones ir raketas, o žaisdamas su jais stalo žaidimus, tokius kaip „Samurajai“, mėgdavo prisiimti lyderio vaidmenį. Tokiose ir daugelyje kitų situacijų jis buvo linkęs kontrolę imti į savo rankas.

Altmanas augo vidurinei klasei priklausiusių žydų šeimoje. Jo motina Connie'ė – dermatologė, tėvas Jerry'is – teisininkas. Jerry'is aktyviai prisidėjo prie prieinamo būsto politikos įgyvendinimo Sent Luiso mieste, taip pat inicijavo istorinių pastatų atstatymą. Tokia jo veikla padėjo formuotis sūnaus polinkiui į visuomeniškumą. Samas puikiai prisimena, kaip vieną dieną tėvas nusivedė jį į savo biurą ir pasakė, kad net jei neturi laiko kam nors padėti, „vis tiek privalo sugalvoti, kaip tai padaryti".

Samas, vyriausias iš keturių vaikų, be galo pasitikėjo savimi, o jo beatodairiška drąsa pelnė aplinkinių pagarbą. Jis atvirai kalbėjo apie savo seksualumą, kai kiti jo amžiaus arba XX a. dešimtojo dešimtmečio pabaigoje augę vaikai būtų tai laikę paslaptyje. Jis drąsiai priėmė tai, ką daugelis JAV vidurio vakarų gyventojų laikė blogiu, ir parodė, kad tai gali būti šaunu – iš dalies todėl, kad norėjo padėti panašiems į save.

Tokį pašaukimą atrasti jam padėjo internetas. Pradėjęs lankytis interneto portale „America Online" (AOL), Altmanas suprato, kad tokių kaip jis yra gerokai daugiau. Prisijungimas prie AOL virtualiųjų pokalbių kambarių buvo išties kažkas nepaprasto. Pasigirsdavo šaukimo signalas ir pyptelėjimai – tarsi „rankos paspaudimas", jie rodydavo, kad modemas mezga saugų ryšį su pasauliniu žiniatinkliu, o tuomet nuskambėdavo cypiantis, čaižus garsas, primenantis sugedusios CB radijo stotelės traškesį. Prisijungęs, pajusdavai širdį plakant greičiau – prieš akis atsiverdavo begalės galimybių ir virtualiųjų pokalbių kambarių. Sėdėdamas prie kompiuterio, galėjai bendrauti su įdomiais žmonėmis iš tolimiausių pasaulio kampelių. Buvo pokalbių kambarių tokiais pavadinimais kaip „Paplūdimio vakarėlis" ar „Pusryčių klubas". Didžiausiuose pokalbių kambariuose virė tikras chaosas, ten reiškėsi gausybė įtartinų veikėjų. Tačiau labiau tikslinei auditorijai skirtose diskusijų grupėse, tokiose kaip „Gyvūnų mylėtojai", „X failų gerbėjai" ar „Gėjai ir lesbietės", pokalbiai rišosi sklandžiau.

Tokiems žmonėms kaip Altmanas šie pokalbių kambariai buvo tikras išsigelbėjimas. Kaip pasyvus pokalbių kambario dalyvis, prisidengęs slapyvardžiu, galėjai saugiai stebėti pokalbius apie LGBTQ bendruomenei draugiškas vietas. Pasaulyje, kuriame dažnai tekdavo jaustis pašaliečiu, tokie virtualūs pokalbių kanalai suteikė priklausymo bendruomenei jausmą. „AOL pokalbių kambariai padarė didelę įtaką mano gyvenimui, – vėliau pasakojo Altmanas, duodamas interviu žurnalui „New Yorker“. – Kai tau vienuolika ar dvylika, slapukavimas slegia.“

AOL pokalbių kambariai buvo tokie svarbūs LGBTQ bendruomenei, kad 1999 metais (tuo metu Altmanui buvo keturiolika) maždaug trečdalyje tokių pokalbių kanalų buvo aptariamos su homoseksualumu susijusios temos. Būdamas šešiolikos, jis prisipažino tėvams esąs homoseksualus. Jo motiną ši žinia pritrenkė. Vėliau tame pačiame žurnalo straipsnyje pasakodama apie savo sūnų, ji teigė, kad jis visada atrodė „neutralus lyties atžvilgiu ir prijaučiantis technologijoms“, tačiau jis taip pat neatitiko įprastų visuomenės standartų. Pavyzdžiui, JAV krašte, kuriame visi dievina barbekiu, Altmanas praktikavo vegetarišką mitybą. Nors aistringai domėjosi kompiuteriais, jis nebuvo atsiskyrėlis ir neturėjo bendravimo sunkumų. Kai visi aplink klausėsi dešimtojo dešimtmečio popmuzikos, jis rinkosi klasiką.

Ne pagal metus subrendusį sūnų Altmanai pervedė į elitinę privačią Johno Burroughso mokyklą, įsikūrusią didelėje žalioje teritorijoje Sent Luiso pakraštyje. Ši mokykla skelbėsi siekianti ugdyti mokinių talentus „visuomenės gerovės labui“.

Samas niekada nepraleisdavo progos imtis lyderio vaidmens. Jis vadovavo vandensvydžio komandai, ėjo mokyklos metraščio redaktoriaus pareigas, taip pat pasisakydavo mokyklos susirinkimuose. Nors puikiai sutarė su mokytojais, kartais sąmoningai laužydavo taisykles, kad sulauktų dėmesio. Per kasmetinį rudenį vykstantį sportininkų palaikymo renginį kartu su savo vadovaujamos

vandensvydžio komandos nariais jis ant scenos nusiplėšė drabužius, ir visi liko stovėti vien su „Speedo“ glaudėmis, šypsodamiesi šūkaujančiai ir plojančiai miniai.

Už tokį žingsnį jis užsitraukė mokyklos fizinio ugdymo vadovo nemalonę, bet, užuot į tai numojęs ranka ar pasiskundęs draugams ar kitam mokytojui, jis nepabūgo šio klausimo kelti pačiu aukščiausiu lygiu. Taigi jis patraukė tiesiai į mokyklos direktoriaus Andy'io Abbotto – švelnaus būdo buvusio anglų kalbos mokytojo – kabinetą. Vyresnysis pedagogas visada mielai priimdavo šį liesą tamsiaplaukį paauglį didelėmis akimis. Šis dažnai užsukdavo į jo kabinetą pasidalyti idėjomis arba pasiskųsti dėl neteisybės, kurią norėdavo aprašyti mokinių laikraštyje.

Jis pastebėjo, kad jaunuolis visai nesibaimina autoritetų. Jei Abbottas kada nors priimdavo nepopuliarų sprendimą, paliesdavusį kitus mokinius, vaikinas suskubdavo jų ginti. „Jis išdrįsdavo paprieštarauti, ir tą darydavo argumentuotai“, – prisimena direktorius. Ramaus būdo pedagogas iki šiol Altmaną laiko „gabiausiu kada nors sutiktu jaunuoliu“.

Toks nuoširdumas ir nebijojimas pasirodyti atviram ir pažeidžiamam padėjo mąslaus žvilgsnio vaikinui pelnyti kitų įtakingų technologijų ir valdžios pasaulio veikėjų, tarp jų – investuotojų, žiniasklaidos atstovų ir įtakingiausių pasaulio bendrovių vadovų – prielankumą, prireikus paremti jo epinę misiją. Ilgainiui Altmanas suprato, kad galios pozicijose esantys asmenys ambicingų sumanymų turinčio jaunuolio kelyje gali padėti pašalinti kliūtis – panašiai kaip vidurinėje mokykloje tą darė direktorius Abbottas.

Kartą jis apsiėmė įgyvendinti didelio masto mokyklinį projektą, kuriuo siekė interneto portale AOL sukurtą savo tinklą perkelti į realų gyvenimą. Sėkmingai įveikęs biurokratinius labirintus ir gavęs direktoriaus palaiminimą, jis įsteigė pirmąją mokyklos LGBTQ paramos grupę. Tai buvo tarsi pogrindinis tinklas, vienijęs mokinius, kurie galėjo

gauti psichologinių konsultacijų arba sutikti bendraminčių. Per metus prie šio tinklo prisijungė apie tuzinas moksleivių.

Tačiau Altmano tai netenkino. Jis ėmė įkalbinėti mokytojus, kad šie ant savo kabinetų durų užklijuotų lipdukus su užrašu, nurodančiu, jog jų klasės yra saugi erdvė homoseksualiems mokiniams, taip bandydamas paversti pedagogus savo sąjungininkais. Galiausiai jis įkūrė grupę pavadinimu „Homoseksualių ir heteroseksualių mokinių aljansas“, kurios tikslas – didinti visuomenės informuotumą apie homoseksualių asmenų teises.

Sykį per rytinį mokyklos susibūrimą jis sumanė atkreipti bendruomenės dėmesį į šį klausimą. Anksčiau susirinkę jo naujai įkurtos grupės nariai pagrindinėje aktų salėje, prieš pasirodant kitiems mokiniams, ant visų kėdžių išdėliojo lapelius su atspausdintais numeriais. Priėjęs prie mikrofono, Altmanas paprašė visų, kuriems atiteko tam tikras numeris, atsistoti. „Apsižvalgykite aplink, – pasakė jis publikai, iš savo vietų pakilus apie šešiasdešimčiai vaikų. – Vienas iš dešimties – tiek mokinių mūsų mokykloje identifikuojasi kaip homoseksualūs.“

Tai buvo drąsus poelgis, bet kažkas buvo ne taip. Tarp dalyvių trūko kelių mokinių, kurie priklausė mokyklos krikščionių klubui. Vėliau Altmanas sužinojo, kad jie boikotavo jo pristatymą, likdami namuose arba savo klasėse. Pasipiktinęs moksleiviais, pakišusiais koją jo sumanymui, jis ir vėl patraukė į direktoriaus Abbotto kabinetą ir pareikalavo, kad krikščionių klubo narių nedalyvavimas būtų traktuojamas kaip pamokų praleidimas.

„Nieko blogo nedarėme, tik siekėme didesnio jų sąmoningumo“, – įrodinėjo paauglys. Nors jaunuolis nebuvo linkęs daryti karštakošiškų pareiškimų, tačiau iš jo žodžių ir griežtos veido išraiškos tapo aišku, kad jis pyksta.

„Kurį laiką bandžiau tokį mokinių elgesį pateisinti, – prisimena direktorius Abbottas. – Vis dėlto linkstu manyti, kad jis veikiausiai buvo

teisus.“

Altmanas mokyklą paliko išmokęs skaudžią pamoką. Kad ir kokių ambicingų idėjų turėtumei, visada atsiras žmonių, kuriems jos neįtiks. Lieka apsupti save galingais ir autoritetingais asmenimis ir turėti palaikančių žmonių ratą.

Netrukus Altmanas įstojo į prestižinį Stanfordo universitetą, įsikūrusį pačioje Kalifornijos valstijos Silicio slėnio širdyje. Šis saulės nutviekstas regionas garsėja kaip aukštųjų technologijų startuolių lopšys – jame susitelkę daugybė talentingų programinės įrangos inžinierių ir technologijų srities verslininkų. Nors išstypęs aštuoniolikmetis domėjosi programavimu ir buvo pasirinkęs kompiuterių mokslo studijas, jis negalėjo apsiriboti vienu dalyku – jį žavėjo daug sričių. Taigi jaunuolis nusprendė studijuoti įvairius humanitarinės krypties dalykus, taip pat lankė kūrybinio rašymo užsiėmimus.

Po paskaitų jis traukdavo į pietus – už dvidešimties minučių kelio jo laukdavo visai kitokios pamokos, ateityje turėjusios esminės reikšmės jo, kaip pasaulinio garso verslininko, karjerai. Populiariame San Chozės kazino Altmanas valandų valandas žaisdavo pokerį, taip tobulindamas savo psichologinius ir poveikio kitiems įgūdžius. Pokerio esmė – stebėti oponentus ir kartais sąmoningai juos klaidinti dėl turimų kortų kombinacijos. Altmanas taip gerai įvaldė blefavimą ir gebėjo perprasti kitų žaidėjų kūno kalbą, kad studijų laikotarpiu išloštais pinigais iš esmės padengė savo pragyvenimo išlaidas. „Būčiau tai daręs ir nemokamai, – vėliau sakė jis vienoje tinklalaidėje. – Man tai be galo patiko. Primygtinai rekomenduoju šį užsiėmimą norintiesiems šio to išmokti apie pasaulį, verslą ir žmonių psichologiją.“

Sritis, į kurią Altmanas galiausiai susitelkė, siekdamas pakeisti pasaulį, jo gyvenime atsirado dar studijų metais. Jis tapo tyrėju Stanfordo DI laboratorijoje – atskiroje milžiniško universiteto miestelio dalyje įsikūrusioje mokslo erdvėje, kurioje tarp gausybės išsiraizgiusių kabelių buvo galima pastebėti vieną kitą roboto rankos prototipą. Neseniai

atidarytai DI laboratorijai vadovavo Sebastianas Thrunas – radikalių pažiūrų kompiuterių mokslininkas skvarbiomis mėlynomis akimis ir vos juntamu vokišku akcentu. Thrunas priklausė naujai kartai akademikų, kurie nesitenkino vien paraiškų dotacijoms rašymu ir laukimu, kol įgis nuolatinę poziciją universitete, bet aktyviai bendradarbiavo su technologijų milžinėmis. Vos už penkių mylių nuo Stanfordo universiteto buvo įsikūrusi pagrindinė „Google“ būstinė. Be akademinės veiklos, Thrunas taip pat vadovavo pažangiems „Google X“ projektams, susijusiems su savavaldžių automobilių ir papildytosios realybės (angl. *augmented reality*) akinių kūrimu.

Paskaitose Thrunas studentams dėstė mašininį mokymąsi – metodą, kuris leidžia kompiuteriams analizuoti didelius kiekius duomenų ir remiantis jais daryti išvadas, o ne tik vykdyti iš anksto užprogramuotas užduotis. Mašininio mokymosi samprata yra itin svarbi DI srityje, nors pats terminas „mokymasis“ gali klaidinti: mašinos negali mąstyti ir mokytis taip, kaip žmonės. Thrunas atkreipė dėmesį į mąslų jaunuolį iš Sent Luiso, kuris domėjosi nenumatytomis DI veiklos pasekmėmis. Kas nutiktų, jei mašina išmoktų daryti blogus darbus?

Thrunas paaiškino, kad DI sistemos gali veikti neprognozuojamai, kad įvykdytų savo „tinkamumo funkciją“ arba pasiektų tikslą. Jis taip pat pridūrė, kad jei DI būtų sukurtas su tinkamumo funkcija išgyventi ir daugintis, jis galėtų netyčia sunaikinti visą biologinę gyvybę Žemėje. Tačiau tai nereikštų, kad DI yra blogas. Jis tiesiog nesuvokia savo veiksmų rimtumo. Jo motyvai beveik nesiskiria nuo tų, kuriais vadovaujamės plaudamiesi rankas. Mes nejaučiame neapykantos ant mūsų odos esančioms bakterijoms ir nesiekiame jų naikinti – tiesiog norime, kad mūsų rankos būtų švarios.

Kurį laiką ši mintis Altmanui nedavė ramybės. Kaip mokslinės fantastikos mėgėjas jis svarstė, ar būtent dėl to žmonėms taip ir nepavyko užmegzti kontakto su nežemiška gyvybe. Galbūt kitų planetų būtybės taip pat bandė sukurti DI, o vėliau buvo sunaikintos savo pačių

kūrinio. Kad to būtų išvengta, kas nors turėjo sukurti saugesnį DI, kol dar neatsirado išties pavojingas.

Daugiau nei dešimtmetį ši idėja kirbėjo Altmano galvoje, kol galiausiai tapo varomąja jėga, paskatinusia įsteigti „OpenAI“. Tačiau tuo metu toks užmojis atrodė per didelis. Dirbtinio intelekto sistemas kūrė tokie akademikai kaip Thrunas. Altmano bendramoksliai iš Stanfordo universiteto steigė startuolius, išaugusius į tokias bendroves kaip „Google“, „Cisco“ ir „Yahoo“. Jaunasis technologijų entuziastas irgi norėjo eiti tuo keliu. Jam tereikėjo verslo idėjos. Ši jį aplankė einantį iš paskaitų. „Argi būtų ne puiku, jei savo mobiliajame telefone galėčiau atsidaryti žemėlapį, kuriame matyčiau visų draugų buvimo vietas?“ – pasiteiravo jis kursioko ir bičiulio Niko Sivo.

Taip Altmanui kilo idėja sukurti mobiliajam telefonui skirtą skaitmeninį žemėlapį, kuriame vartotojas galėtų rasti savo draugus – tai galėjo tapti pagrindiniu būsimos bendrovės produktu. Tačiau įkurti įmonę nebuvo paprasta. Reikėjo pritraukti lėšų iš rizikos kapitalo investuotojų, ir nors trijų mylių spinduliu aplink Stanfordą jų buvo bent keliolika, Altmanas buvo dar per jaunas ir neturėjo pakankamai patirties. Galiausiai Altmanas ir Sivo rado išeitį – jiedu nusprendė dalyvauti jauniesiems verslininkams skirtoje trijų mėnesių trukmės intensyvių mokymų programoje „Y Combinator“, kurią kitoje JAV pusėje – Masačusetso valstijos Kembrižo mieste – inicijavo vienas technologijų magnatas. „Y Combinator“ vėliau tapo visų laikų sėkmingiausiu startuolių akceleratoriumi, investavusiu į tokias technologijų bendroves kaip „Airbnb“, „Stripe“ ir „Dropbox“, kurių bendra vertė dabar siekia 400 mlrd. JAV dolerių.

Tuo metu devyniolikmetis Altmanas apie tai dar nieko nežinojo. Dauguma Silicio slėnio investuotojų „Y Combinator“ laikė kvaila programišių vasaros stovykla. Šią organizaciją įkūrė Paulas Grahamas – „cargo“ šortus pamėgęs keturiasdešimt vienų kompiuterių mokslininkas, kuris tapo milijonieriumi, savo elektroninės prekybos verslą pardavęs

bendrovei „Yahoo“. Pasiekęs finansinę sėkmę, Grahamas ėmė reikštis kaip idėjinis lyderis, savo tinklapyje skelbdamas rašinius įvairiomis temomis, kurios ne visai atitiko tipiško programuotojo interesus – nuo ekonomikos, vaikų auginimo, žodžio laisvės iki to, ką reiškia būti moksliuku vidurinėje mokykloje.

Tačiau didžiausio populiarumo sulaukė jo tekstai apie startuolių kūrimą, prikaustę tokių jaunuolių kaip Altmanas dėmesį tarsi dvasinio lyderio pamokymai. Juose nuolat pabrėžiama, kad startuolio steigėjo asmenybė yra svarbesnė už bet ką kitą. Anot Grahamo, sėkmingai technologijų bendrovei įkurti nebūtina turėti genialią idėją – svarbiausia, kad už bendrovės vairo stovėtų genialus žmogus.

„Pavyzdžiui, „Google“ planas buvo tiesiog sukurti paieškos svetainę, kuri išties veiktų“, – rašė Grahamas. Akivaizdu, kad toks sumanymas pasiteisino. Laikai, kai viską lėmė genialios idėjos, baigėsi. Dabar į pirmą planą iškyla steigėjai, o geriausi iš jų yra programišiai – programuotojai, kurie nebijo mesti iššūkį nusistovėjusiai mąstysenai, kad sukurtų ką nors naujo. Būdamas programišius, „galite būti trisdešimt šešis kartus produktyvesnis nei dirbdamas eilinį darbą biure“, – rašė jis.

Technologijų bendrovės kūrimas Silicio slėnyje yra netgi tam tikra patriotizmo forma, nes šį žingsnį lydi nepalaužiamo individualizmo dvasia, kadaise įkvėpusi JAV valstybės kūrėjus. „Programišiai yra nepaklusnūs, – rašė Grahamas. – Tokia ir yra buvimo programišiumi esmė. Be to, tai ir amerikietiškumo esmė. Neatsitiktinai Silicio slėnis įsikūręs būtent Amerikoje, o ne Prancūzijoje, Vokietijoje, Anglijoje ar Japonijoje. Tose šalyse žmonės neperžengia primestų rėmų.“

Anot Grahamo, kelias į sėkmę yra visiškai paprastas. Startuokite be didelių investicijų, siūlydami bazinį produktą su minimaliu funkcionalumu, ir bėgant laikui jį optimizuokite. Orientuokitės į siaurą vartotojų ratą – dešimt žmonių, kurie nuoširdžiai vertina jūsų produktą, yra vertingesni už tūkstančius klientų, kurie netrykšta entuziazmu. Taip

pat mokėkite kūrybingai apeiti tam tikras taisykles. Dar daugiau – galite apskritai perrašyti nusistovėjusias visuomenės normas.

Grahamo idėjos ilgainiui sulaukė didelio atgarsio Silicio slėnyje ir pakurstė vis labiau įsigalintį įsitikinimą, kad startuolių kūrėjų vizija yra šventa ir neliečiama, todėl jiems turėtų būti leidžiama veikti be jokių ribų – tarsi dievams. Štai kodėl bendrovių „Google“ ir „Facebook“ įkūrėjai galėjo prisiimti modernių verslo autokratų vaidmenį, dažnai perimdami daugumos balsavimo teisę suteikiančių akcijų kontrolę ir pasirinkdami keistas bendrovių veiklos strategines kryptis. (Tai puikiai iliustruoja faktas, kad nei „Facebook“ direktorių valdyba, nei akcininkai neprieštaravo keistam ir brangiai kainuosiančiam Marko Zuckerbergo sprendimui paversti „Facebook“ virtualiosios realybės įmone.) Dėl dviklasės akcijų struktūros daugelio technologijų startuolių, įskaitant „Airbnb“ ir „Snapchat“, įkūrėjai galėjo išlaikyti neįprastą kontrolės lygį savo įmonėse. Grahamas ir kiti buvo įsitikinę, kad įkūrėjai tokią galią turi ne be priežasties. Protingiausiems ir talentingiausiems žmonėms, turintiems ilgalaikę viziją, reikėjo laisvės ją įgyvendinti.

Altmano asmenybėje Grahamas įžvelgė tuos pačius programišiams būdingus instinktus – begalinį smalsumą, aštrų protą ir gebėjimą mąstyti plačiai. Dar daugiau – jis pastebėjo, kad šis jaunuolis su tamsių plaukų ševeliūra puikiai jautėsi tarp vyresnių žmonių, todėl galėjo be vargo vadovauti ir tokiems kaip Grahamas, kuris buvo už jį dvidešimčia metų vyresnis. Kai Grahamas pasiūlė tuo metu devyniolikmečiui Altmanui palaukti metus, kol galės prisijungti prie „Y Combinator“, šis atrėžė, kad vis tiek dalyvaus. Grahamas jam iš karto pajuto simpatiją.

Dauguma kitų programos dalyvių buvo inžinieriai ir programišiai, tarp jų ir populiaraus interneto forumo „Reddit“ įkūrėjai. Grahamas ir jo žmona Jessica Livingston kiekvienam startuoliui skyrė po 6000 JAV dolerių – tokio dydžio stipendiją Masačusetso technologijos institutas skirdavo aukštesniųjų pakopų studentams vasaros metu. Nors dauguma rizikos kapitalo investuotojų startuoliams negailėdavo milijonų,

Grahamas ragino steigėjus stengtis kuo daugiau padaryti su kuo mažesnėmis lėšomis ir pasiekti bent minimalų pelningumą. Jis skatino atsisakyti teisininkų, bankininkų bei viešųjų ryšių specialistų paslaugų ir taupyti, viską atliekant patiems.

Pats Grahamas viską darė su minimaliais ištekliais. Kiekvieną antradienio vakarą jis gamindavo savo firminį patiekalą – vištienos frikasė – ir kviesdavosi draugus kaip lektorius, kad šie papasakotų apie startuolių kūrimą. Livingston rūpindavosi kiekvienos naujos įmonės teisiniais dokumentais.

Altmanas su bičiuliu Sivo naująją savo bendrovę pavadino „Loopt“. Jis persikėlė į Kembridžą Masačusetso valstijoje, kad galėtų dirbti pirmajame „Y Combinator“ biure netoli Grahamo namų. Altmanas su Grahamu užmezgė artimą ryšį, panašų į tą, kokį jis palaikė su savo mokyklos direktoriumi. Vilties kupiniems jauniesiems verslininkams Grahamas buvo tarsi dvasinis vadovas, į kurį jie mėgdavo kreiptis trumpiniu PG. Altmanas rimtai žiūrėjo į Grahamo mokymus ir į „Loopt“ žvelgė ne kaip į įmonę, galinčią jam sukrauti turtus, bet kaip į idėją, kuri galėtų padaryti pasaulį geresnį. Jis visa galva pasinėrė į prototipo tobulinimą, maitinosi vien greitai paruošiamais ramenais ir „Starbucks“ kavos skonio ledais. Dėl tokio darbo krūvio ir prastos mitybos jis susirgo skorbutu – liga, kurią sukelia vitamino C trūkumas.

Nors berniukiškų bruožų jaunuolis buvo neblogas programuotojas, jis dar labiau atsiskleidė kaip verslininkas. Jis nė nesudvejojęs susisiekė su „Sprint“, „Verizon“ ir „Boost Mobile“ vadovais ir pristatė jiems didingą viziją apie tai, kaip pasikeis žmonių bendravimo ir naudojimosi telefonu įpročiai. Kalbėdamas santūria maniera ir vartodamas raiškius posakius, išlavintus kūrybinio rašymo paskaitose, jis aiškino, kad „Loopt“ vieną dieną taps svarbiu kiekvieno mobiliojo telefono naudotojo palydovu. Tuo metu programėlių parduotuvės dar neegzistavo, todėl Altmanui reikėjo įtikinti mobiliojo ryšio operatorius jo programą iš anksto įdiegti į pirmuosius išmaniuosius telefonus. Taigi palenkti telekomunikacijų

bendrovių vadovus į savo pusę jam buvo itin svarbu, o tą pasiekti jis galėjo pasinaudodamas savo neprilygstamu talentu įtaigiai pristatyti įmonę. Galiausiai „Sprint“, „Verizon“, „Boost“ ir net „BlackBerry“ sutiko įdiegti jo paslaugą į jų siūlomus įrenginius.

Pasibaigus trijų mėnesių „Y Combinator“ programai, Altmanas surinko pradines lėšas savo startuolio plėtrai. Penkiolikai investuotojų, kurių dauguma buvo pasiturintys Grahamo draugai, jis surengė maždaug penkiolikos minučių trukmės pristatymą, per kurį išdėstė „Loopt“ viziją. Paskui jis kreipėsi į dar didesnį finansinį potencialą turinčias rizikos kapitalo įmones Silicio slėnyje. Sulaukęs kelių pasiūlymų, jis sutarė dėl 5 mln. JAV dolerių finansavimo iš dviejų pirmaujančių rizikos kapitalo įmonių, anksčiau investavusių į „Google“, „Yahoo“ ir „PayPal“.

Gavęs šias lėšas, Altmanas metė studijas Stanfordo universitete, kad galėtų visiškai atsidėti „Loopt“ plėtrai. Su kelių pasamdytų inžinierių komanda jis persikėlė į Palo Altą Kalifornijoje ir įsikūrė „Sequoia“ bendradarbystės erdvėje. Kaip ir greta plušėję „YouTube“ kūrėjai, jie iki išnaktų užsibūdavo darbe, karštligiškai kurdami programinį kodą. Vėliau Altmanas perkėlė komandą į pirmąjį „Loopt“ biurą prestižinėje Mauntin Vju vietovėje pačioje Silicio slėnio širdyje – vos už kelių kvartalų nuo pagrindinės „Google“ būstinės.

Silicio slėnis – beprotiškų idėjų kalvė. Jame steigiamos ne šiaip įmonės – čia gimsta technologijų imperijos. Kūrėjai čia neapsiriboja paprastais išradimais – jie siekia proveržio technologijų ir mokslo srityse. Norintieji tirti Alzheimerio ligą gali rinktis universitetą Rytų pakrantėje ar Europoje. Tačiau ieškantieji būdo, kaip sustabdyti senėjimą, vyksta į Silicio slėnį.

Svarbiausias šio regiono privalumas – platus verslo ryšių tinklas. Bet kurią dieną, apsilankęs renginyje, galėjai netikėtai sutikti žmogų, turintį galios iš esmės paspartinti tavo verslo augimą. Užbėgęs papusryčiauti į „Bucks“ restoraną Vudsaide, Kalifornijoje, akis į akį susidurdavai su

vienu iš „Yahoo“ įkūrėjų, besimėgaujančiu vaisių ir jogurto desertu prie to paties stalo, prie kurio Elonas Muskas kadaise dalyvavo savo pirmajame susitikime dėl „PayPal“ finansavimo. Užsukęs į „Musto“ barą San Fransisko „Battery“ klube, galėjai netikėtai išvysti vieną iš „Facebook“ įkūrėjų.

Altmanas greitai įsiliejo į Silicio slėnio programuotojų, investuotojų ir vadovų tinklą. Tie, kurie žinojo, kaip prisijungti prie šio šiuolaikinio įtakingųjų klubo, turėjo daugiau šansų pasinaudoti tramplinu į sėkmę, pavertusiu jo narius milijardieriais. Altmanas turėjo tokių gerų tinklaveikos įgūdžių, kad sugebėjo užmegzti reikiamus ryšius ir pristatyti „Loopt“ 2008 metais įvykusioje kasmetinėje prestižinėje „Apple“ kūrėjų konferencijoje. Vilkėdamas džinsus ir dvejus polo marškinėlius – žalią ir rožinį, dėl kurių atrodė tarsi vaikų televizijos laidos vedėjas, liekno sudėjimo jaunasis verslininkas pareiškė, kad „Loopt“ yra didžiausia socialinių žemėlapių paslauga pasaulyje. „Mes kuriame laimingus atsitiktinumus“, – pasakė jis, su vos matoma šypsena žvelgdamas į susirinkusiuosius.

Iš pirmo žvilgsnio viskas atrodė puiku. Tačiau pažvelgus atidžiau darėsi aišku, kad „Loopt“ patiria sunkumų. Žmonės tiesiog nebuvo labai suinteresuoti ieškoti draugų naudodamiesi Altmano skaitmeniniu žemėlapiu. Entuziastingo žvilgsnio verslininkas buvo įsitikinęs, kad jaunieji mobiliųjų įrenginių naudotojai troško susitikti su draugais lygiai taip pat, kaip ir jis. Tačiau idėja susitikti su bičiuliais bare ar susisiekti su nepažįstamaisiais, prireikus papildomo žaidėjo krepšinio rungtynėms, reikalavo per daug pastangų, juolab kad bendrauti buvo galima tiesiog per ekraną. Baigiantis pirmajam XXI a. dešimtmečiui, vis daugiau žmonių bendravo su draugais tokiuose socialiniuose tinkluose kaip „Facebook“. Šis tinklas augo gerokai sparčiau nei „Loopt“. „Facebook“ jau buvo pritraukęs šimtus milijonų aktyvių vartotojų, o „Loopt“ platforma teturėjo penkis milijonus užsiregistravusių narių.

Padėties nepagerino ir tai, kad „Loopt“ tapo kontraversiška. Praėjus metams nuo įmonės įkūrimo, Altmanas sulaukė skambučio iš savo buvusio vidurinės mokyklos direktoriaus Andy'io Abbotto. Šis papasakojo, kad tėvai verčia savo vaikus naudotis „Loopt“, norėdami sekti jų buvimo vietą, o kartą vienas iš tėvų paskambino į mokyklą ekskursijos metu, kad praneštų, jog autobusas, kuriuo keliavo jo vaikas, viršijo greitį. „Pažiūrėk, ką tu pridarei“, – pusiau juokais pasakė buvęs mokytojas.

Altmanas buvo girdėjęs ir blogesnių dalykų. „Sulaukiame susirūpinimo iš moterų grupių“, – prisipažino jaunasis verslininkas. Kai kurie vyrai verčia savo žmonas įsidiegti „Loopt“ programėlę, kad galėtų nuolat sekti jų buvimo vietą. Tai buvo šiurpus ir potencialiai pavojingas piktnaudžiavimo jo kūriniu būdas. „Bet mes ieškome sprendimo“, – greitai pridūrė jis. „Loopt“ vartotojai galėjo suklastoti savo buvimo vietos informaciją. Pažeidžiamos moterys, kurios lankydavosi maisto prekių parduotuvėje, galėjo sudaryti įspūdį, tarsi būtų namuose.

Nors daugelis verslininkų būtų neigę, kad jų programėle piktnaudžiaujama, Altmanas, regis, norėjo atvirai spręsti šią problemą. Dar būdamas paauglys jis suprato, kad slapukavimas tik pablogina situaciją. Iškilusias problemas geriau ištraukti į dienos šviesą. Kartą jis sulaukė skambučio iš atkaklios „Wall Street Journal“ technologijų žurnalistės Jessicos Lessin, kuri teiravosi apie „Loopt“ privatumo problemas ir piktnaudžiavimo atvejus. Jos nuostabai, Altmanas noriai aptarė prieštaringus klausimus ir netgi išsiuntė jai elektroniniu paštu ilgą dokumentą, kuriame išsamiai aprašė visus su programėlės naudojimu susijusius pavojus, – tą vėliau atskleidė žurnalistė savo paskelbtame straipsnyje.

Tai, kas atrodė kaip profesinė savižudybė, buvo gudrus viešųjų ryšių ėjimas, kurį jis vėliau ne kartą panaudos kaip savotišką atvirkštinės psichologijos triuką. Rodydamas pernelyg didelį susirūpinimą blogiausiu savo kūrinio panaudojimo scenarijumi, Altmanas galėjo nuginkluoti

kritikus ar žurnalistus, tokius kaip Lessin. Niekas kitas negalėjo paleisti kritikos strėlių jo pusėn, nes jis tai padarė pats. Atrodė, kad jo sąžiningumas jam pačiam visiškai nebuvo naudingas. Vis dėlto turint omenyje tai, kad programėle naudotasi persekiojimo tikslais, sąžiningiausia būtų buvę jos veikimą visiškai nutraukti.

Galiausiai už jį tai padarė vartotojai. Altmanas iš anksto neįvertino, kaip nejaukiai jie jausis siųsdami savo GPS koordinates, kad galėtų susitikti su kitais. „Supratau, kad negali priversti žmonių daryti tai, ko jie nenori“, – pripažino jis vėliau.

Didžiąją savo gyvenimo trečiojo dešimtmečio dalį jaunasis verslininkas praleido rūpindamasis „Loopt“ augimu. Jis pasinaudojo naujomis „iPhone“ tiesioginių pranešimų galimybėmis, kad vartotojams siųstų trumpąsias žinutes, rodomas pagrindiniame telefono ekrane, ir taip skatintų juos naudotis „Loopt“ pokalbių funkcija. Jis sudarė sąlygas reklamuotojams siųsti „greituosius pasiūlymus“ „Loopt“ vartotojams. Kiekvieną naujinį jis pristatydavo kaip neabejotiną sėkmę. „Sulaukėme fantastiško įvertinimo“, – teigė jis viename 2010 metais pasirodžiusiame interviu. Tačiau dažniausiai tai tebuvo skambios frazės. 2012 metais „Loopt“ reguliariai naudojosi vos keli tūkstančiai žmonių visame pasaulyje. Kalbos apie imperijos sukūrimą ir liko kalbomis. Kaip ir didžioji dauguma technologijų startuolių, „Loopt“ patyrė nesėkmę.

Technologijų startuolių pasaulyje pagrindinis steigėjo tikslas – paversti savo idėją daugiamilijardine bendrove arba parduoti savo įmonę už kelis milijardus dolerių didesniam rinkos žaidėjui. Išlikti nepriklausomam tapo vis sunkiau, o daugumą startuolių prarydavo tokios technologijų milžinės kaip „Google“ ar „Facebook“. Jei steigėjui pavykdavo tai padaryti, gautą pelną jis dažnu atveju panaudodavo naujai įmonei įkurti, o tada šį ciklą pradėdavo iš naujo kaip serijinis verslo kūrėjas. Tačiau „Loopt“ pasitraukimas buvo neįspūdingas. 2012 metais Altmanas pardavė ją dovanų kuponų bendrovei už maždaug 43 mln. JAV

dolerių ir vos padengė įsiskolinimą investuotojams ir savo darbuotojams.

Tuo metu Altmanas galėjo palikti Silicio slėnį užnugaryje, tačiau „Loopt“ žlugimas pakurstė dar tvirtesnį įsitikinimą, kad jis turėtų daryti ką nors prasmingesnio. Jis buvo ne pirmas technologijų vizionierius, kuriam nesėkmė tapo postūmiu siekti dar ambicingesnių tikslų. Maždaug dešimtmečiu anksčiau direktorių valdybos sprendimu Elonas Muskas buvo atleistas iš „PayPal“ vadovo pareigų. Pasimokęs iš savo patirties, Muskas suprato, kad jam pabodo dirbti su tokiu paviršutinišku dalyku kaip vartotojų mokėjimų paslauga. „Kita mano įmonė turės prisidėti prie ilgalaikių teigiamų pokyčių“, – sakė jis viename interviu. Po kelerių metų jis šį pažadą išpildė. Muskas pradėjo dirbti su „Tesla“ įkūrėjais ir ėmėsi gelbėti žmoniją nuo egzistencinės klimato kaitos grėsmės.

Jei į grupelę prestižiniame San Fransisko klube „Battery Club“ sėdinčių žmonių mestumėte išmanųjį telefoną, jis pataikytų bent į tris, mėginančius išgelbėti pasaulį. Ne vienas Silicio slėnio verslininkas tikėjo, kad jo programėlė prisideda prie žmonijos gerovės, ir nors dalis jų iš tiesų sukūrė naudingų produktų, kuriais naudojasi milijonai žmonių, daugeliui kitų dėl to išsivystė tikras mesijo kompleksas. Šiame regione, kur pagrindinis dėmesys skiriamas inovacijoms, tokia gelbėtojų kultūra yra be galo paplitusi – prie to ypač prisidėjo Grahamo suformuluoti principai, grindžiami prielaida, kad steigėjai viską žino geriausiai. Priklausantieji šiai elitinei novatoriškų programišių bendruomenei tikėjo galintys spręsti ne tik inžinerines, bet ir giliai įsišaknijusias visuomenės problemas, bauginančias žmoniją jau daugelį metų.

Altmanas norėjo, kad „Loopt“ suburtų žmones, nes manė, kad būtent to jiems reikia. Tuo metu žmonės vis labiau klimpo į ekranus, beprasmiškai naršydami ir dalydami „patiktukus“ įvairiuose socialiniuose tinkluose ir kurdami vis labiau kiekybiškai matuojamą žmonių bendrumo jausmą. Altmanas jautė poreikį imtis ko nors

prasmingesnio. Galbūt reikėjo pasiūlyti žmonėms tai, ko jie patys dar nežinojo norintys. Būtent tai „Apple“ sėkmingai darė jau daug metų, o Silicio slėnyje tai buvo paslaptis, kurią visi bandė įminti.

Jaunuoliui iš Sent Luiso reikėjo grįžti į startuolių pasaulį ir taip įsitraukti į Silicio slėnio bendruomenę, kad jo vardas taptų neatsiejamas nuo tų bendrovių, kurios skelbėsi keičiančios pasaulį. Jis turėjo tapti ryškesne savo buvusių mentorių versija ir sugrįžti prie idėjų, brandintų dar Stanfordo DI laboratorijoje. Tai galiausiai nuvedė jį prie didesnio tikslo: išgelbėti žmoniją nuo egzistencinės grėsmės ir padėti jai pasiekti neregėtą klestėjimą.

2 SKYRIUS

Pergalės skonis

Klyksmai, amerikietiškųjų kalnelių dundesys ir atrakcionų parko orkestrinės mašinos žvangesys žymi 1994 metais išleisto kompiuterinio žaidimo „Theme Park“ pradžią. Priešais akis atsiveria didelis tuščias pikseliuotos žolės plotas, kurį reikia užpildyti milžiniškų mėsainių formos maisto kioskeliais ir linksmųjų kalnelių trasomis, kylančiomis į svaiginančias aukštumas. Žaidimo tikslas – uždirbti kuo daugiau pelno.

Jei manote, kad „Theme Park“ sukūrė pagyvenę žaidimų kūrėjai, siekę išmokyti vaikus verslo pagrindų, labai klystate. Iš tiesų jo kūrėjas – tamsiaplaukis paauglys iš Šiaurės Londono, vardu Demisas Hassabisas. Jo požiūris į darbą prilygo Silicio slėnio verslininkų darbo etikai, o aistra žaidimams buvo beribė. Dar gerokai prieš tapdamas pagrindiniu varžovu lenktynėse dėl pažangiausios DI sistemos sukūrimo Hassabisas mokėsi verslo valdymo per simuliacijas – tai tapo kertiniu jo darbo principu, lydėjusiu jį stengiantis sukurti už žmogų protingesnes mašinas.

Žaidimą „Theme Park“ žaidėjas pradeda turėdamas apie 200 000 JAV dolerių atrakcionams statyti ir darbuotojams samdyti. Pelną reikia užsidirbti iš bilietų, suvenyrų, ledų ir taiklumo žaidimų. Jei nesamdysi pakankamai mechanikų – atrakcionai ges. Jei nebus pakankamai apsaugos darbuotojų – parką nusiaubs chuliganai. Jei taupysi cukrų – lankytojai atsisakys ledų. Darbuotojai gali paskelbti streiką, tad kartais teks derėtis dėl atlyginimų. Būdamas vos septyniolikos, Hassabisas sukūrė šią sudėtingą sąnaudų ir naudos pusiausvyra paremtą sistemą kaip itin kompleksišką verslo valdymo simuliaciją. Žaidimas buvo toks įtraukiantis, kad nuo jo pasirodymo 1994 metais parduota net penkiolika milijonų kopijų.

Vaizdo žaidimai tarsi potvynio banga užplūdo Jungtinę Karalystę ir JAV, įtraukdami vaikus į ryškius dopamino kupinus pasaulius: vienur vėžliukai nindzės kovėsi šoninio slinkimo lygiais, kitur žaidėjai galėjo vairuoti pikapą sunkiai įveikiamais žvyrkeliais. Tačiau Hassabisas manė, kad geriausi vaizdo žaidimai – tai simuliacijos, atkartojančios tikrovę. Dieviškosios perspektyvos žaidimuose žaidėjui suteikiama galia kurti ir naikinti. Užuot valdęs vieną veikėją, tokį kaip Mario, gali kontroliuoti tūkstančių virtualių veikėjų gyvenimus, kurti kraštovaizdžius ar diktuoti civilizacijos raidą. Gali pastatyti miestą, o tada jį nusiaubti sukėlęs stichinę nelaimę arba pripildyti pramogų parką šimtais lankytojų.

Ši technologija leido ne tik smagiai leisti laiką, bet ir mokytis – nuo verslo valdymo iki visatos paslapčių pažinimo. Nors Hassabisas vertino žaidimų pramoginę vertę, galiausiai jį užvaldė noras pasitelkti juos kuriant dirbtinį superintelektą, kuris padėtų atskleisti žmogaus sąmonės paslaptis.

Siekis perprasti visatą buvo gerokai platesnis nei daugumos kitų mokslininkų ambicijos – bet tai nestebina, prisiminus, kad pats Hassabisas buvo kaip mįslė: vienintelis matematikos genijus bohemiškų laisvamanių šeimoje. Jo motina Angela – pamaldi baptistė, atvykusi į Jungtinę Karalystę iš Singapūro. Su būsimuoju vyru – laisvos dvasios graikų kilmės kipriečiu Costasu Hassabisu – ji susipažino Šiaurės Londone, šeimoje, pas kurią buvo apsistojusi. Iš pažiūros jie buvo visiškos priešingybės, tačiau susituokė ir susilaukė trijų vaikų – Demisas buvo pirmagimis. Costasas nuolat keitė darbus – dirbo mokytoju, vadovavo žaislų parduotuvei. Kai Demisui sukako dvylika, šeima buvo persikrausčiusi net dešimt kartų.

Jau tada buvo aišku, kad Demisas yra kitoks nei visi. Būdamas vos ketverių, jis laimėdavo šachmatų partijas prieš savo tėvą ir dėdę, o sulaukęs šešerių, vietiniuose šachmatų čempionatuose lengvai įveikdavo daugumą savo amžiaus vaikų. Kad pasiektų šachmatų lentą, Demisui tekdavo sėdėti ant pagalvėlės ar telefonų knygos. Jis laisvai

skaitė ir viskuo domėjosi, tačiau savo protines galias daugiausia sutelkdavo žaidimams. Kai tėvas namo parnešdavo stalo žaidimų su trūkstamomis figūrėlėmis, Demisas juos panaudodavo kurdamas naujus žaidimus ir įtraukdavo į tai savo jaunesnįjį brolį bei seserį.

Tačiau tikrosios įdomybės dar tik laukė. Dešimtmečiu anksčiau, dar gerokai prieš tai, kai Samas Altmanas praeito amžiaus dešimtajame dešimtmetyje atrado AOL pokalbių kambarius, Hassabiso dėmesį patraukė dar primityvesnės technologijos, kurių juoduose ekranuose mirgėjo stambūs pikseliai. 1984 metais, būdamas aštuonerių, už šachmatų turnyre laimėtus pinigus jis įsigijo namų kompiuterį „ZX Spectrum 48“. Tai buvo vienas pirmųjų asmeninių kompiuterių – masyvi juoda klaviatūra, kurią reikėjo prijungti prie televizoriaus, o programos su spalvinga grafika buvo įkeliamos naudojant kasetes.

Programavimo žinių Hassabisas sėmėsi iš knygų – taip savarankiškai išmoko kurti žaidimus „Spectrum“ kompiuteriui. Prieš eidamas miegoti, jis suprogramuodavo skaičiavimo užduotį, kad kompiuteris ją atliktų per naktį. Ryte skaičiavimai būdavo baigti. Hassabisui tai buvo tikras atradimas – „Spectrum“ galėjo už jį atlikti visą protinį darbą. Kompiuteris tapo tarsi jo proto pratęsimu.

Įsitraukęs į nišinį programavimo pasaulį, jis įsigijo galingesnį „Commodore Amiga 500“ – masyvų baltą įrenginį su pele ir monitoriumi – ir su mokyklos draugais subūrė programišių klubą. Rašydami kodą ir kurdami spalvingus vaizdus, jie mėgino atkurti mėgstamų žaidimų scenas. Hassabisas dažnai stengdavosi draugus pranokti įmantresne grafika. Jis taip pat galėjo be vargo išardyti kompiuterį ir vėl jį surinkti. Galiausiai jam pavyko sukurti skaitmeninį šachmatų žaidimą, kurį davė žaisti savo jaunesniajam broliui George’ui.

Šachmatai Hassabisui vis dar buvo svarbiausias dalykas gyvenime – jis svajojo tapti pasaulio čempionu. Motina palaikė sūnaus siekius ir pradėjo mokyti jį namuose, kad jis galėtų daugiau laiko skirti šachmatams. Per moksleivių atostogas jis nesiilsėdavo – keliaudavo iš

vieno šachmatų turnyro į kitą ir be perstojo varžydavosi. Vėliau Hassabisas sakė, kad žaidimai yra lyg sporto salė protui, o šachmatai – pati intensyviausia treniruotė. Kaip pokeris Samui Altmanui tapo psichologijos ir verslo mokykla, taip šachmatai Hassabisui padėjo įvaldyti strateginį mąstymą, kurio esmė – pirmiausia išsikelti tikslą, o tada ieškoti būdų jam pasiekti.

Tačiau vieną dieną viskas pasikeitė. Būdamas vienuolikos, Hassabisas dalyvavo šachmatų čempionate Lichtenšteine ir varžėsi su Danijos šachmatų čempionu. Po dešimties valandų trukusio žaidimo maratono, pareikalavusio daug protinių jėgų, danų šachmatininkas mėgino išprovokuoti varžovą atlikti tam tikrus ėjimus, kurie būtų lėmę lygiąsias. Hassabisas dar turėjo karalių ir karalienę, bet oponentas buvo pranašesnis – jo lentos pusėje buvo karalius, bokštas, rikis ir žirgas. Hassabisas jautėsi išsekęs ir manė, kad netrukus pralaimės. Jis pasidavė.

Danijos čempionas buvo apstulbęs. „Kodėl pasidavei?“ – paklausė jis.

Tada jis parodė Hassabisui, kad šis galėjo padaryti ėjimą, lemsiantį lygiąsias. Hassabisas sėdėjo įbedęs žvilgsnį į lentą. Kartais pralaimėjimas įžiebia dar didesnes ambicijas. Jei negali pakęsti pralaimėjimo, tuomet lieka ieškoti paguodos siekiant didesnio tikslo. Hassabisas suklupo, nors buvo įdėjęs milžiniškas pastangas. Apžvelgęs salę, kurioje palinkę prie lentų ir įtempę smegenis sėdėjo kiti šachmatų genijai, jis staiga suvokė, kad visas šis turnyras – tik proto galių švaistymas. Čia buvo susirinkę geriausi pasaulio strategai. Kas būtų, jei jie spręstų didesnes problemas? Tuo metu Hassabisas buvo antras geriausias pasaulyje tarp šachmatininkų iki keturiolikos metų. Tačiau tai buvo tik žaidimas.

Hassabisas pranešė tėvams, kad nebenori dalyvauti šachmatų turnyruose, ir vėl pradėjo lankyti mokyklą. Jis buvo tylus, jautrus vaikas, mėgęs Enyos muziką ir pats išmokęs pianinu sugroti jos kūrinį „Watermark“. Mėgstamiausias jo filmas buvo „Blade Runner“ – mokslinės fantastikos juosta apie detektyvą, medžiojantį žmonėmis apsimetančius DI padarus. Hassabisą be galo paveikė jautriausios šio

filmo scenos. Jis be perstojo klausydavosi jaudinančio Vangelio kūrinio, skambančio finalinėje filmo scenoje, kurioje antagonistas apgailestauja, kad jo prisiminimai „išnyks laike lyg ašaros lietuje“.

Sekmadieniais motina dažnai vesdavosi Hassabisą ir jo brolį bei seserį į Hendono baptistų bažnyčią Šiaurės Londone – didingą, iš pilko akmens sumūrytą statinį ant kalvos, nuo kurios atsiveria priemiesčių panorama. Tai buvo nedidelė tarptautinė bendruomenė, kurioje meldėsi atvykėliai iš Filipinų, Ganos, Prancūzijos, Indijos. Turėdamas kipriečio ir singapūrietės kraujo, Hassabisas čia galėjo jaustis savas. Sekmadieniai šioje vietoje būdavo gyvesni nei santūriose Anglijos Bažnyčios pamaldose – iškėlę rankas, tikintieji giedodami šlovindavo Viešpatį, ritmingai mušant būgnams ir grojant grupei. Bažnyčios pastorius ne itin griežtai laikėsi dogmų ir labiau akcentavo pagarbą artimui. Pamaldų metu liedavosi stiprios emocijos, o pati bažnyčia pabrėžtinai puoselėjo evangelinę dvasią.

Jungtinėje Karalystėje baptistai sudarė religinę mažumą: šiai bendruomenei priklausė apie 150 tūkst. narių, palyginti su daugiau nei milijonu Anglijos Bažnyčios tikinčiųjų. Religija ir dievo idėja Hassabisui kėlė susižavėjimą. Jis ėmė svarstyti, ar į Dievo paieškas būtų galima leistis pasitelkus mokslą. Sulaukęs šešiolikos (baigęs mokyklą dvejais metais anksčiau už bendraamžius) jis perskaitė Nobelio premijos laureato fiziko Steveno Weinbergo knygą „Galutinės teorijos paieškos“ (*Dreams of a Final Theory*). Joje aprašomas didingas, beveik donkichotiškas siekis atrasti Visatos jėgas paaiškinančią, visa apimančią teoriją. Weinbergas tikėjo, kad visas pagrindines Visatos jėgas galima paaiškinti vos keliomis lygtimis – panašiai kaip Einšteinas, lygtimi $E = mc^2$ nusakęs energijos ir masės sąryšį. Idealiu atveju ši „visko teorija“ būtų tokia glausta, kad ją būtų galima užrašyti viename puslapyje, o gal net sutalpinti į vieną lygtį.

Hassabisas buvo sužavėtas, o paskui priblokštas, sužinojęs, kad mokslininkams nepavyko padaryti didelės pažangos šioje srityje. Tuomet

jam kilo išganinga mintis, kad jiems reikia pagalbos – papildomų intelektualinių galių. Galbūt jis galėtų prisidėti? Hassabisas pažvelgė į griozdišką „Commodore Amiga“ kompiuterį, naktimis atlikdavusį sudėtingus skaičiavimus. Galbūt praverstų išmanesnis kompiuteris? Jei jam pavyktų sukurti galingesnį, jo paties protą pranokstantį kompiuterį, galbūt jis padėtų mokslininkams įminti didžiąsias Visatos paslaptis ar netgi surasti dieviškąsias ištakas.

„Tai atrodė tobulas sprendimas“, – vėliau interviu „New York Times“ žurnalistui Ezrai Kleinui pasakojo Hassabisas. Iš pradžių vaikinas svarstė studijuoti fiziką, bet perskaitęs Weinbergo knygą nusprendė rinktis ką nors ambicingesnio. Pasinėrus į kompiuterių mokslą ir vis aktualesnę DI sritį, galbūt jam pavyktų sukurti reikšmingiausią mokslinį įrankį ir padaryti atradimų, kurie padėtų žmonijai. Hassabisas niekaip negalėjo atsitraukti nuo žaidimų pasaulio, todėl ėmėsi kurti ilgalaikį planą, kaip sujungti abi šias sritis. Svarbiausia jam buvo susitelkti į realų pasaulį imituojančius žaidimus. XX a. devintojo dešimtmečio pabaigoje kompiuteriniuose žaidimuose jau buvo galima modeliuoti pagrindinius civilizacijos bruožus. Jei kompiuteris galėtų tikroviškai atkurti pasaulį, tuomet ypač intelektualus kompiuteris ieškotų būdų, kaip išspręsti realaus pasaulio problemas – rasti sprendimus simuliacijoje, o tada pritaikyti juos realiame gyvenime.

Hassabisą įkvėpė vadinamieji dieviškosios perspektyvos žaidimai, o mėgstamiausias jų buvo „Populous“. „Mane žavėjo dinamiškai besivystantys pasauliai – žaidimo eiga keičiasi priklausomai nuo to, kaip jį žaidi, – sakė jis. – Gali modeliuoti pasaulio dalį tarsi žaisdamas smėlio dėžėje ir laisvai eksperimentuoti.“

Nepaisant pikseliuotos grafikos, pats žaidimas yra išties sudėtingas. Žaidėjas prisiima dievybės vaidmenį: jis gali valdyti žalią lygumų slėnį su jame išsibarsčiusiais nameliais, savo dieviškosiomis galiomis vadovauti jų gyventojams kovoje su kitų dievybių sekėjais. Jis gali formuoti žemės paviršių – išlyginti plotus, kad sekėjai turėtų kur statytis namus ir kurti

šeimas. Jis taip pat turi galių sukelti žemės drebėjimus. Hassabisas taip susižavėjo šiuo žaidimu, padėjusiu pamatus dieviškosios perspektyvos žaidimų žanrui, kad pasiryžo įsidarbinti jį sukūrusioje bendrovėje „Bullfrog Productions“. Taigi jis nusprendė dalyvauti konkurse dėl galimybės dirbti šioje bendrovėje. Deja, užsitikrinti darbo vietos jam nepavyko. Tuomet jis paskambino tiesiai į pačią bendrovę ir paprašė galimybės atlikti savaitės trukmės praktiką. Vadovai sutiko. Hassabisas jiems paliko tokį gerą įspūdį, kad sulaukęs penkiolikos jis buvo pakviestas grįžti į bendrovę vasaros darbui.

Po kurio laiko šešiolikmetis Hassabisas buvo priimtas į Kembridžo universitetą studijuoti kompiuterių mokslo, bet universitetas pranešė, kad jis dar per jaunas ir patarė jam bent metus palaukti. Tuo laikotarpiu jis vėl pradėjo dirbti „Bullfrog“ – algą gaudavo grynaisiais, o gyveno netoliese esančiuose Jaunųjų krikščionių draugijos (YMCA) nakvynės namuose Gilforde, Sario grafystėje. Pradėjęs nuo žaidimų testuotojo pareigų, netrukus jis tapo lygio dizaineriu – tiesioginiu „Bullfrog“ įkūrėjo Peterio Molyneux pavaldiniu.

Dėl plikos galvos ir juodų polo marškinėlių Molyneux labiau priminė aludės šeimininką nei žaidimų industrijos legendą. Pristatydamas savo projektus, jis švaistydavosi pažadais, kurių nepajėgė išpildyti, – dėl šios priežasties jis pradėtas vertinti gana prieštaringai. Kalbėdamas apie žaidimų veikimo principus ar funkcijas, jis nevengdavo drąsių pareiškimų, pavyzdžiui, pasakodamas apie žaidimo „Fable“ galimybes, jis teigė esą šiame virtualiame pasaulyje bus galima pasodinti gilę, o po kelių dienų iš jos išaugsiantis medis. Tačiau dalis šio pažado taip ir liko neišpildyta.

Vis dėlto Molyneux, kuris tuo metu vis dar mėgavosi „Populous“ sėkme, turėjo drąsių idėjų, kurios skatino visą industriją judėti į priekį. Patyrusiam verslininkui Hassabisas pasirodė neįprastai smalsus ir ne pagal metus subrendęs. Molyneux pamena, kaip genialusis paauglys pažėrė jam begalę klausimų apie „Bullfrog“ žaidimų technines ribas,

klausinėjo, kodėl tam tikros funkcijos vadinamos „dirbtiniu intelektu“, kai iš tikrųjų jos primena paprastas programines sistemas.

„Jo negąsdino net tos užduotys, kurios atrodė neįveikiamos“, – prisimena Molyneux. „Bullfrog“ įkūrėjui sumanius sukurti pramogų parko simuliacijos žaidimą, kitų komandos narių tai nesudomino – jiems labiau rūpėjo žaidimai su kardais ir kautynėmis. Tačiau Hassabisas mielai sutiko imtis šio projekto. Taip gimė spalvingas „Theme Park“ pasaulis su amerikietiškaisiais kalneliais ir gardumynų kioskais. Kuriant naująjį žaidimą, Molyneux tapo Hassabiso mentoriumi, o jiems talkino vos keli kompiuterinės grafikos kūrėjai. Trumpam atsitraukę nuo programavimo ir žaidimo struktūros kūrimo darbų, jie dažnai kalbėdavosi apie DI galimybes. Hassabisas atskleidė manantis, kad DI žmones pranoks ir taps sąmoningas maždaug po dešimtmečio.

„Dirbtinio intelekto ateitis atrodė kone ranka pasiekiama, – prisimena patyręs žaidimų kūrėjas. – Taip filosofuodami klausėme savęs, kodėl žmonės turėtų būti vieninteliai kūrėjai? Kodėl kūrybos naštos negalėtų perimti DI?“ Jie fantazavo, kaip ateityje DI kurs muziką, poeziją ir netgi žaidimus.

Tačiau kol kas jie naudojosi sistemomis, kurios vos atitiko DI apibrėžimą, kad „Theme Park“ įgautų bent šiek tiek tikroviškumo. Mašininio mokymosi metodai padėjo sukurti skirtingais charakterio bruožais pasižyminčius pramogų parko lankytojus: vieni buvo impulsyvesni ir labiau linkę leisti pinigus, kiti – taupesni. Žaidimas sulaukė milžiniškos sėkmės. Palyginti su „Populous“ žaidimu, kurio parduota penki milijonai kopijų, „Theme Park“ buvo daug sėkmingesnis – jo pardavimas šį skaičių viršijo net tris kartus.

Galiausiai į Kembridžą atvykęs Hassabisas jau buvo įgiję šiokį tokį įžymybės statusą. Pasiskolinęs Molyneux priklausantį „Porsche 911“, jis važinėjosi po universiteto miestelį, norėdamas padaryti įspūdį kitiems studentams. Kadangi per mokyklines atostogas jo dienotvarkė būdavo užimta šachmatų turnyrais, pirmieji metai universitete jam atstojo tikras

atostogas: vakare jaunuolis eidavo linksmintis su bendrakursiais, o ryte, nubudintas saulės spindulių, gulėdavo lovoje ir klausydavo „The Prodigy“ dainų. Jei raudonio išmuštais žandais nesėdėdavo su taure raudonojo vyno universiteto bare, žaisdavo greituosius šachmatus arba lenktyniaudavo skolintu „Porsche“. Galiausiai jis tą automobilį sudaužė ir dėl to turėjo skambinti mentoriui atsiprašyti. „Tai buvo jau antras kartas, kai jis sudaužė mano automobilį, – kiek susiraukdamas prisimena Molyneux. Tačiau pykti ant nuolat besišypsančio jauno genijaus buvo sunku. – Jis nepaprastai mielas“, – priduria jis.

Būtent Kembridže Hassabisas susipažino su kai kuriais būsimaisiais artimiausiais bendražygiais. Tarp jų buvo kompiuterių mokslą studijavęs Benas Coppinas, vėliau tapęs produktų kūrimo vadovu bendrovėje „DeepMind“. Kartu jie kalbėdavosi apie religiją ir tai, kaip DI galėtų padėti spręsti pasaulines problemas. Tačiau „DeepMind“ istorija prasidėjo dešimtmečiu vėliau. Pirmiausia reikėjo baigti studijas, o tada grįžti dirbti pas Molyneux. Netrukus Hassabisui į rankas pakliuvo keisčiausia kada nors matyta paraiška dėl darbo. Paštu atkeliavusio butelio viduje jis rado arbatos dėmėmis išmargintą, dailyraščiu rašytą laišką apdegintais kraštais. Jame siuntėjas teigė išsigelbėjęs iš sudužusio laivo ir atsidūręs saloje pavadinimu „Korporate“. Hassabisas iš karto suprato šią metaforą – jam pačiam būtų buvę nepakeliama dirbti didelėje korporacijoje.

Laiško autorius buvo Joe McDonaghas – programuotojas, dirbęs milžiniškoje korporacijoje „British Telecom“. Jis buvo didelis kompiuterinių žaidimų entuziastas ir svajojo apie karjerą žaidimų kūrimo srityje, todėl be galo nudžiugo, sulaukęs kvietimo į pokalbį paties Molyneux namuose. Atvykus duris atidarė smulkaus sudėjimo, elfą primenantis jaunuolis prigludusiais lyg šalmas plaukais ir vos matoma barzdele. Tai buvo Hassabisas. Jis atrodė gerokai jaunesnis nei dvidešimt vienų. „Pagalvojau – kas tas vaikis?“ – prisimena McDonaghas. Paaiškėjo,

kad Hassabisas jau ėjo vadovaujamas pareigas bendrovėje, todėl ir turėjo pravesti pokalbį.

McDonaghas netruko suprasti susidūręs su tokiu pat dideliu žaidimų mėgėju – tik itin azartišku. Sužinojęs, kad McDonaghas domisi origamiu, Hassabisas pasiūlė jam pasivaržyti, kuris greičiau išlankstys popierinę gervę. Hassabisas laimėjo. Visą likusią popietę jie žaidė stalo žaidimus. McDonaghas, vėliau paskambinęs pasiteirauti dėl darbo, sužinojo, kad tas keistas vaikis, su kuriuo buvo susitikęs per darbo pokalbį, jau buvo palikęs įmonę. Iš darbo Hassabisas nusprendė išeiti dėl to, kad Molyneux nebepajėgė atliepti jo techninių ambicijų. „Jam viskas vyko per lėtai“, – prisimena Molyneux.

Norėdamas sužinoti daugiau, McDonaghas gavo Hassabiso telefono numerį ir jam paskambino. „Kuriu naują įmonę“, – paaiškino Hassabisas. Jis atskleidė, kad bendrovė vadinsis „Elixir Studios“, o jos tikslas – sukurti pažangiu DI paremtą dieviškosios perspektyvos žaidimą, kuris galiausiai veiks kaip viso pasaulio simuliacija.

Tai buvo nepaprastai ambicinga vizija. McDonaghas nedvejodamas sutiko prisijungti. Pradėjęs dirbti „Elixir“, jis tapo vyriausiuoju žaidimų kūrėju – jo užduotis buvo kurti nepaprastus naujus pasaulius. Hassabisas iš savo buvusio mentoriaus buvo išmokęs pompastiškos rinkodaros meno, todėl žiniasklaidos atstovams pasakodamas apie savo tikslus, nevengdavo drąsių pareiškimų. Patekęs ant žurnalo „Edge“ – dešimtajame dešimtmetyje žaidimų madas diktavusio leidinio – viršelio, jis gyrėsi ketinantis kurti žaidimus, kurie ne tik pasižymės pažangiomis funkcijomis, bet ir pasieks gerokai platesnę auditoriją nei siauras paauglių segmentas. Jis teigė sukursiantis tokius sumanius žaidimus, kad juos norės žaisti net „The Economist“ skaitytojai. „Norėjau parodyti, kad žaidimai gali būti rimta meno forma – kaip knygos ar filmai“, – sako Hassabisas. Jis turėjo ilgalaikį planą: „Elixir“ tapus sėkmingai, ketino ją parduoti ir įkurti DI technologijų bendrovę.

Hassabisas susitelkė į pagrindinio žaidimo – „Republic: The Revolution“ – kūrimą. Tai buvo politinių procesų simuliacija, kurioje žaidėjų tikslas – nuversti totalitarinį režimą išgalvotoje Rytų Europos valstybėje. Hassabisas norėjo, kad viskas atrodytų kuo tikroviškiau. Kad įgyvendintų entuziastingo jaunojo vadovo viziją, McDonaghas valandų valandas praleisdavo Britų bibliotekoje, gilindamasis į Sovietų Sąjungos istoriją. Pats Hassabisas susitelkė į techninę pusę – siekė sukurti DI sistemą, leidžiančią viename žaidime veikti net milijonui virtualių veikėjų. Tai buvo išties ambicingas tikslas, turint omenyje, kad iki tol daugumoje dieviškosios perspektyvos žaidimų jų būdavo vos vienas ar du tūkstančiai. Jis norėjo suteikti galimybę žaidėjams keisti vaizdo mastelį taip, kad, priartinę palydovinį miesto vaizdą, jie galėtų matyti daugiaaukščio balkone stovinčios gėlės žiedlapį.

Buvęs šachmatų čempionas pasistengė suburti protingiausius programuotojus, kokius tik galėjo rasti, – daugelis jų buvo baigę Oksfordo ar Kembridžo universitetus. Komandos dvasią jis stiprino žaidimais, kuriuose ir pats sužibėdavo, nesvarbu, ar tai būtų vaizdo žaidimas „Starcraft“, ar strateginis stalo žaidimas „Diplomacy“. Net stalo futbolas virsdavo nuožmiu sportu: iš peties įsukęs lazdą su plastikiniais žaidėjais, jis atlikdavo savo firminį „žaibo smūgį“ ir pasiųsdavo kamuoliuką į vartus. Hassabisas kartu su „Elixir“ komanda taip pat išbėgdavo į futbolo aikštę – sportiškumu nepasižyminčių programuotojų gretose jis užėmė puolėjo poziciją. Mačuose „penki prieš penkis“ su vietiniais Šiaurės Londono vyrukais jis pademonstruodavo tikrą kovinę dvasią – agresyviai kovodamas dėl kamuolio, kaip įniršęs terjeras atakuodavo dvigubai aukštesnių už save žaidėjų blauzdas ir dažnai pelnydavo įvarčių.

Artėjant numatytam projekto terminui, berniukiškų bruožų vadovas ir jo programuotojų komanda kasdien nenuilsdami dirbo nuo dešimtos valandos ryto iki kitos dienos šeštos ryto. Miegui likdavo trys ar keturios valandos, kurias jie praleisdavo posėdžių salėje, – kartais užmigdavo

tiesiog prie stalų, rankose tebelaikydami žaidimo pultelius. Visi pamiršo naktines linksmybes, o pats Hassabisas pasižadėjo daugiau niekada nepasigerti, bijodamas, kad tai gali pakenkti smegenims.

Žaidimo „Republic“ grafikai ir DI technologijoms Hassabisas kėlė kone absurdiškus reikalavimus. Jo vizija šviesmečiais lenkė tuo metu turėtus skaičiavimo pajėgumus. Tačiau jis nematė prasmės kurti virtualios šalies, jei negalės joje apgyvendinti tūkstančių tikroviškų veikėjų. „Nenorėjau vaizduoti žmonių kaip abstrakčių taškelių, padrikai klaidžiojančių ekrane, – sakė jis žurnalui „Edge“. – Norėjau, kad tai būtų vyrai, studentai, namų šeimininkės ir girtuokliai – visi gyvenantys atskirus, realistiškus gyvenimus.“

Nebuvo geresnio būdo atskleisti magiškas DI galias kaip per žaidimą. Tuo metu pažangiausi DI tyrimai vyko kompiuterinių žaidimų industrijoje. Išmanesnė programinė įranga leido kurti dinamiškus pasaulius ir sudarė sąlygas atsirasti naujam žaidimų stiliui – spontaniškam žaidimui (angl. *emergent gameplay*). Tokio tipo žaidimuose žaidėjui nebėra primetamas iš anksto numatytas maršrutas, kaip, pavyzdžiui, žaidime „Super Mario Bros“, – vietoj to jis įkurdinamas virtualiame pasaulyje ir aprūpinamas tam tikromis priemonėmis, kad pats galėtų kovoti už išlikimą. Būtent tokiu principu buvo paremtas tiek „Grand Theft Auto“, tiek vėliau pasirodęs „Minecraft“, tapęs visų laikų perkamiausiu vaizdo žaidimu.

Hassabisas tikėjo, kad jam pavyks sukurti kažką panašaus. Tačiau iškilo problema – žaidimas „Republic: The Revolution“ buvo nuobodus. Tai – blogiausia, kas gali nutikti žaidimo kūrėjui. Per penkerius kūrimo metus net ketverius jie skyrė technologijų tobulinimui, pamiršdami patį žaidimo procesą. Norint sukurti išties gerą kompiuterinį žaidimą, būtina nuolatos jį tobulinti. Iš esmės reikia pradėti nuo bazinės versijos, turinčios visus pagrindinius žaidimo elementus, o tada tūkstančius kartų žaisti ir drauge tobulinti. „Elixir“ kūrėjai, žinodami vadovo aukštus

reikalavimus technologinei žaidimo pusei, negalėjo skirti pakankamai laiko, kad žaidimą padarytų įdomesnį.

„Vaizdo žaidimuose svarbiausia – įsitraukimas ir jausmas, – teigia McDonaghas. – „Republic“ negalėjo pasiūlyti nė vieno iš šių dalykų. Pasijutome tarsi įstrigę technologinėje juodojoje skylėje.“ Programuotojai žinojo, kad žaidimas nėra pakankamai geras. Jam pasirodžius rinkoje, blogiausi komandos įtarimai pasitvirtino – kritikai išsakė prieštaringas nuomones ir teigė, kad žaidimas pernelyg sudėtingas. Žaidimo pardavimas taip pat neblizgėjo.

„Projektas buvo per daug ambicingas tam laikmečiui, – pripažįsta Hassabisas. – Pernelyg skubėjau padaryti techninį ir meninį pareiškimą.“

Tačiau Hassabisas nepasidavė – „Elixir“ ėmėsi kurti ne mažiau ambicingą dieviškosios perspektyvos žaidimą „Evil Genius“. Jame žaidėjai įsikūnija į veikėją, primenantį Džeimso Bondo filmų piktadarius, siekiančius užvaldyti pasaulį. Nors buvo išradingas, su šmaikščiu humoru, žaidimas taip ir nesulaukė didesnio populiarumo. Pasišovęs išleisti patobulintą žaidimo versiją „Evil Genius 2“, Hassabisas susidūrė su didelėmis investicijų į technologijas išlaidomis. 2005 metais genijus, siekęs pakeisti požiūrį į žaidimų kūrimą, buvo priverstas uždaryti „Elixir“. Ši patirtis jį palaužė, ypač žinant, kad iki tol visas jo gyvenimas buvo pergalių serija – nesvarbu, ar tai būtų šachmatai, stalo futbolas, ar mokykla.

Situaciją dar labiau pablogino smerkiantis kitų žaidimų pramonės atstovų požiūris. Hassabisui tai buvo tikras pažeminimas, turint omenyje, kad bendraudamas su žiniasklaida ir žaidimų srities atstovais, „Elixir“ jis pristatydavo kaip naujovių nešėją, kuri, pasitelkusi pažangias technologijas, pakeis tradicinį požiūrį į žaidimų kūrimą. McDonaghas prisimena kartą dalyvavęs konferencijoje ir užsiminęs, kad dirbo „Elixir“. Tai išgirdęs, vienas britų žaidimų pramonės senbuvis net nusijuokė. McDonaghas pasijuto taip, tarsi žemė slystų jam iš po kojų. „Pakelti nesėkmę buvo beprotiškai sunku“, – prisimena jis.

Kartą, besiginčydami dėl įmonės žlugimo, McDonaghas ir Hassabisas rimtai susibarė – tai buvo pirmas ir vienintelis kartas, kai jis girdėjo paprastai ramų kolegą pakeliant balsą. „Tai buvo siaubinga, – sako McDonaghas. – Visi studijavome Oksforde ir Kembridže, įpratome prie pergalių. Niekada anksčiau nebuvome pralaimėję, bet galiausiai visų akivaizdoje suklupome.“

Trokšdamas pademonstruoti žmonėms DI magiją, Hassabisas padarė esminę klaidą – jis bandė kurti žaidimą remdamasis tuo, kas jam pačiam buvo įdomiausia. Siekdamas sukurti už žmogų protingesnes mašinas, jis turėjo vadovautis atvirkštine logika. Jam reikėjo dar giliau pasinerti į DI sritį ir nebenaudoti jo kokybiškiems žaidimams. Vietoj to jis turėjo pasitelkti žaidimus galingam DI sukurti.

Po kelerių metų, jau įkopęs į ketvirtą dešimtį, McDonaghas vėl sulaukė skambučio iš buvusio vadovo. „Steigiu įmonę, kuri vadinsis „DeepMind“, – pareiškė entuziastingasis verslininkas, regis, visai pamiršęs ankstesnes nesėkmes.

„Nebenoriu to kartoti“, – pagalvojo McDonaghas ir atsisakė prisijungti prie buvusio bendražygio.

Vėliau jis su nuostaba stebėjo, kaip Hassabisas leidosi į dar vieną neįtikėtinai ambicingą avantiūrą. Šįsyk jis pranoko visus lūkesčius, sukūręs, regis, pažangiausią pasaulyje DI sistemą. Bent jau tol, kol horizonte pasirodė Samas Altmanas.

3 SKYRIUS

Gelbstint žmones

Vieną karštą 2006-ųjų vasaros dieną Altmanas, vilkėdamas tik sportinius šortus, gulėjo ant grindų savo studijos tipo bute Mauntin Vju, Kalifornijoje. Išskėtęs rankas, bandė ramiai kvėpuoti. Tą savaitgalį jis labai stengėsi susitarti dėl „Loopt“ palankaus sandorio, tačiau reikalai nesiklostė taip, kaip tikėjosi. Be to, jo bute tvyrojo trisdešimt penkių laipsnių karštis. Altmanui rodėsi, kad jo galva tuoj sprogs iš streso, – taip apie tuometinę savijautą jis pasakojo 2022 metais pasirodžiusioje tinklalaidėje „Art of Accomplishment“.

Ilgą laiką jis bandė save įtikinti, kad tokia savijauta verslininkams – kasdienybė. „Tai, ką jaučiu, yra normalu, – kalbėjo sau mintyse. – Bet tai nepadeda.“ Stresas viską tik blogino.

„Loopt“ ištikusi nesėkmė išmokė Altmaną svarbaus dalyko: negali priversti žmonių daryti to, ko jie nenori. Jis taip pat pasimokė, kaip svarbu emociškai atsiriboti nuo sudėtingų situacijų. Tos akimirkos, kai jis gulėjo ant grindų vienais šortais, tapo tikru lūžiu jo gyvenime. Altmanas nusprendė gyventi kitaip. Jam reikėjo išmokti emociškai nuo visko atsiriboti.

Pardavęs įmonę „Loopt“ ir išsiskyręs su Nicku Sivo – ilgamečiu verslo ir romantiniu partneriu, – Altmanas dar kurį laiką darbavosi jo startuolį įsigijusioje bendrovėje, o tada metams nuo visko atsitraukė. Jis visiškai nusišalino nuo verslo reikalų. Silicio slėnyje, kur vyrauja darboholizmo kultūra, sprendimas metams pasitraukti iš aktyvios veiklos retai sulaukia pritarimo, ir Altmanas netruko pajusti šio žingsnio padarinius. Jei vakarėlyje su kuo nors besišnekučiuodamas jis užsimindavo apie savo

planą metus pailsėti, pašnekovas kaipmat imdavo dairytis įdomesnės draugijos.

Jis vis dar palaikė ryšį su San Fransisko įlankos regione veikiančia „Y Combinator“, bendradarbiaudamas su šia organizacija kaip ne visą darbo dieną dirbantis partneris. Tuo metu rizikos kapitalo investuotojai jau nebežiūrėjo į jos vykdomą programą kaip į neorganizuotą programišių sambūrį – dabar ją laikė perspektyvių internetines paslaugas teikiančių įmonių kalve. Keletas šios programos dalyvių, pavyzdžiui, „Reddit“ ir „Scribd“, jau buvo išaugę į itin sėkmingas bendroves. „Y Combinator“ Silicio slėnio startuolių įkūrėjams tapo tarsi vartais į sėkmę. Kasmet paraiškas dalyvauti šioje programoje pateikdavo tūkstančiai kandidatų, tačiau tik maždaug šimtas jų įveikdavo atranką.

Likusį savo savanoriškų „atostogų“ laiką Altmanas paskyrė įvairiems pomėgiams – jis perskaitė dešimtis knygų pačiomis įvairiausiomis temomis: nuo branduolinės inžinerijos iki sintetinės biologijos, nuo investavimo iki DI. Keliavo po pasaulį, gyveno nakvynės namuose, lankėsi konferencijose. Pardavęs „Loopt“, jis gavo maždaug penkis milijonus JAV dolerių, ir dalį šios sumos investavo į keletą startuolių.

Altmanas atvirai pripažindavo, kad beveik visos bendrovės, į kurias buvo investavęs, žlugo, tačiau pridurdavo, jog ši patirtis išmokė jį atpažinti perspektyviausius projektus. Jis buvo įsitikinęs, kad dažnai klysti – nieko bloga. Svarbu kartais priimti teisingus sprendimus, pavyzdžiui, investuoti į startuolį, kuris vėliau išaugtų į itin sėkmingą bendrovę ir kurio akcijų pardavimas leistų uždirbti solidžią sumą.

Jei gyvenimą prilygintume drobei, galima būtų teigti, kad Altmanas rankose laikė didžiausią teptuką, pasirengęs ją išmarginti kuo stambesniais potėpiais. Jį vis labiau traukė DI. Maždaug tuo metu, kai pardavė „Loopt“, kartu su keliais draugais iš technologijų srities jis leidosi į žygį, kuriame diskutavo apie DI tyrimų ateitį, – šį faktą atskleidžia žurnale „New Yorker“ apie jį paskelbtas straipsnis. Altmanas manė, kad, esant tokiai sparčiai techninės įrangos ir mašininio

mokymosi pažangai, šios mašinos greičiausiai dar jo gyvenimo laikotarpiu sugebės atkartoti žmogaus smegenų funkcijas.

Visa tai paskatino jį rimtai apmąstyti žmonijos vaidmenį mitybos grandinės viršūnėje. Jei mūsų intelektą galima imituoti kompiuteriu, ar tikrai esame tokie jau unikalūs? Altmanui įtikimesnis atrodė neigiamas atsakymas. Nors iš pradžių tokia mintis galėjo pasirodyti slegianti, jis netruko joje įžvelgti galimybę, kurią verta išnaudoti. Jeigu žmonės nėra tokie išskirtiniai, vadinasi, jų gebėjimus galėtų atkartoti – gal net pranokti – kompiuteriai. Galbūt jis pats galėtų prie to prisidėti.

Daugeliu atžvilgių Altmanas rėmėsi Silicio slėnyje vyraujančiu požiūriu, kad pati gyvybė yra inžinerinis galvosūkis. Didžiąsias žmonijos problemas esą galima spręsti taip pat, kaip optimizuojamas programėlių veikimas. Tokią mąstyseną iš dalies lemia įprotis sistemingai ir logiškai įveikti technines problemas, išugdytas tiek studijų, tiek programinės įrangos kūrimo metu. Sėkmė šioje srityje matuojama pagal tai, kiek efektyviai veikia sukurtas produktas. Todėl natūralu, kad šie metodai imami taikyti ir kitose visuomenės bei gyvenimo srityse.

Vargu, ar stebina tai, kad kalbėdamas apie žmones Altmanas dažnai pasitelkdavo kompiuterių mokslo terminiją. Pavyzdžiui, viename žurnalo interviu jis teigė, kad žmonės „mokosi vos dviejų bitų per sekundę greičiu“. Bitas – mažiausias informacijos kiekio vienetas, nusakomas dvejetainiais skaičiais (0 ar 1). Taip Altmanas vaizdingai pabrėžė, kokie riboti esame apdorodami informaciją. Žmogaus smegenų veiklą palyginus su kompiuterių veikimu, akivaizdu, kad pastarieji informaciją geba apdoroti kur kas didesniu greičiu, matuojamu gigabitais ar net terabitais per sekundę.

Norėdamas sukurti žmogaus intelektą pranokstančią mašiną, Altmanas, be abejo, turėjo likti Silicio slėnyje – ten, kur kuriama ateitis.

„Silicio slėnyje tvyro neblėstantis tikėjimas ateitimi, – yra sakęs jis. – Ten gali sutikti žmonių, kurie rimtai žiūri į tavo beprotiškas idėjas, o ne juokiasi iš tavęs.“ Be to, Silicio slėnyje veikia gerai išplėtotas pažinčių

tinklas, grindžiamas abipuse pagalba: padėjus kam nors pritraukti lėšų startuolio kūrimui, galima tikėtis pagalbos samdant talentingą inžinierių.

Grįžęs iš kelionių, Altmanas įkūrė ankstyvojo etapo investicijų įmonę „Hydrazine Capital“, siekdamas investuoti į įvairius startuolius – nuo gyvybės mokslų bendrovių iki edukacinės programinės įrangos kūrėjų. Pasinaudodamas ryšiais su keliais įtakingiausiais Silicio slėnio finansininkais, tarp jų – Paulu Grahamu ir Peteriu Thieliu (vienu pirmųjų „Facebook“ investuotojų), Altmanas surinko 21 mln. JAV dolerių investicijų fondą. Thielis, dabar garsėjantis kaip mįslingas milijardierius, kurio idėjos balansuoja ties mokslinės fantastikos riba, tapo svarbia figūra DI plėtros lenktynėse – jis rėmė tiek Altmaną, tiek Londone įsikūrusį Demisą Hassabisą. Jis priklausė vadinamajai „PayPal mafijai“ – elitinei internetinių mokėjimų milžinės bendrakūrėjų ir vadovų grupei, kurios nariai daugelį metų investavo į vienas kito įmones. Tarp jų buvo ir Elonas Muskas bei „LinkedIn“ įkūrėjas Reidas Hoffmanas.

Maždaug 75 proc. „Hydrazine“ fondo lėšų Altmanas skyrė „Y Combinator“ programoje dalyvavusiems startuoliams. Tokia strategija išties pasiteisino. Per maždaug ketverius metus fondo vertė išaugo dešimteriopai – daugiausia dėl investicijų į startuolius, priklausiusius nuolat besiplečiančiam elitiniam Silicio slėnio pažinčių tinklui. Tarp tokių pradedančiųjų įmonių buvo „Reddit“, kuri, kaip ir jo paties bendrovė, priklausė pirmajai „Y Combinator“ programos dalyvių grupei, taip pat „Asana“ – verslo programinės įrangos bendrovė, įsteigta „Facebook“ bendrakūrėjo milijardieriaus Dustino Moskovitzo. Vėliau šie ryšiai Altmanui buvo be galo naudingi kuriant itin galingą DI.

Altmanas puikiai suvokė, kad asmeniniai ryšiai ilgainiui vertingesni už greitą finansinę grąžą. Būtent todėl jam buvo nepriimtina priešiška laikysena, kurią, kaip rizikos kapitalo investuotojas, privalėjo demonstruoti bendraudamas su verslininkais. Tokiose derybose įprastai siekiama išsireikalauti kuo didesnę akcijų dalį už kuo mažesnę pinigų sumą. Altmano taip pat nežavėjo Silicio slėnyje įsigalėjusi pelno

vaikymosi kultūra. Jį labiau viliojo galimybė prisidėti prie įdomių projektų ir su tuo susijęs žinomumas. Jo paties turtas apsiribojo keturių miegamųjų namu San Fransiske, nekilnojamuoju turtu Big Serio pakrantės regione Kalifornijoje ir dešimties milijonų JAV dolerių suma, kuri jam generavo pajamas iš palūkanų.

2014 metais, Altmanui lankantis Grahamo namuose, šis virtuvėje jo paklausė: „Ar norėtum perimti vadovavimą „Y Combinator“?“ Altmanas nusišypsojo. Grahamas su žmona Jessica Livingston augino du mažamečius vaikus, o programai pasiekus tokį didelį mastą, porai sunkiai sekėsi susidoroti su išaugusiu darbo krūviu. Be to, Grahamas, duodamas interviu, dažnai pasakydavo neapgalvotų dalykų, kurie tik sustiprindavo įtarimus, kad Silicio slėnio elito gretose dominuoja baltieji „brograмeriai“[1]. Pavyzdžiui, viename tinklaraščio įraše jis teigė nenorintis startuolio steigti su moterimi, auginančia arba planuojančia susilaukti mažų vaikų.

„Y Combinator“ toliau plėtėsi, ir Grahamui darėsi vis sunkiau ją valdyti. Per ankstesnius septynerius metus finansavimas pagal šią programą buvo skirtas 632 startuoliams. Kasmet gauti finansavimą pretenduodavo apie dešimt tūkstančių kandidatų, tačiau atranką sėkmingai įveikdavo vos du šimtai. Technologijų srityje startuoliai kūrėsi dar neregėtu tempu, todėl „Y Combinator“ turėjo plėstis, kad patenkintų vis didėjantį finansavimo poreikį. „Aš nepajėgiu valdyti tokio masto projektų, – vėliau tais pačiais metais vykusioje konferencijoje sakė Grahamas, aiškindamas vadovybės pasikeitimą. – Tą puikiai sugebės daryti Samas.“

Tuo metu Altmanui tebuvo trisdešimt, o Grahamas artėjo prie penkiasdešimties. Altmanas jau buvo perėmęs „Y Combinator“ vadovavimą. Jis tapo savotišku startuolių guru – dalijosi įžvalgomis ir patarimais pačiomis įvairiausiomis temomis, net ir tomis, kurių pats gerai neišmanė. Nors Altmanas buvo jaunas ir iki tol vadovavo tik vienai įmonei, kuri, galima sakyti, patyrė nesėkmę, jis ryžosi paskelbti

tinklaraščio įrašą su devyniasdešimt penkiais trumpais patarimais, kurių, jo manymu, turėtų laikytis kiti startuoliai.

Altmanas dar buvo gana nepatyręs, bet Grahamui ir Livingston jis padarė tokį stiprų įspūdį, kad šie nė nesivargino sudaryti galimų naujų „Y Combinator“ vadovų sąrašo. Abu vieningai sutarė, jog šiai pozicijai labiausiai tinka būtent Altmanas. Grahamas savo protežė apgaubė kone mesijo aura: viename esė rašė, kad Sama (toks buvo Altmano internetinis slapyvardis) yra vienas iš penkių įdomiausių visų laikų startuolių įkūrėjų. „Spręsdamas dizaino klausimus, klausiu savęs: ką darytų Steve’as [Jobsas]? Tačiau kalbai pasisukus apie strategiją ar ambicijas, svarstau: ką darytų Sama?“, – sakė jis.

Stojęs už „Y Combinator“ vairo, Altmanas pirmiausia ėmėsi plėsti šią organizaciją ir jos veiklos mastą. Siekdamas veiklą organizuoti pagal institucinę sąrangą, jis subūrė stebėtojų tarybą, kuriai priklausė Jessica Livingston, pats Altmanas ir septyni „Y Combinator“ alumnai. Jis padvigubino nuolatinių partnerių skaičių ir pakvietė prisijungti dar kelis dalį laiko dirbančius partnerius, tarp jų – rizikos kapitalo investuotoją milijardierių Peterį Thielį.

Altmanas nuo vaikystės domėjosi pažangiausiomis mokslo sritimis ir tikėjo, kad mokslo laimėjimai yra būtini siekiant padėti žmonėms ir užtikrinti jų materialinę gerovę. Todėl jis sutelkė dėmesį į startuolius, kuriančius pažangias technologijas ir sprendžiančius sudėtingas mokslo bei inžinerijos problemas. „Man tiesiog patinka tai daryti, – sako jis dabar. – Nebijau prarasti pinigų, jei tik matau prasmę tame, ką darau. Svarbu imtis didžiausių iššūkių – nors tai ir susiję su didesne rizika, potenciali grąža dažnai ją pateisina.“

Iki tol „Y Combinator“ programoje daugiausia dalyvaudavo vartotojams skirtas programėles ir verslo programinę įrangą kuriantys startuoliai, kurių pajamų modelis buvo gana nuspėjamas. Tačiau Altmanas nemanė, kad tokios įmonės pajėgios pakeisti pasaulį. Todėl prie „Y Combinator“ programos jis pakvietė prisijungti startuolį „Cruise“,

kurio specializacija – savavaldžių automobilių gamyba, bei branduolinės sintezės technologijų startuolį „Helion Energy“, įsikūrusį Redmonde, Vašingtono valstijoje.

Branduolinė sintezė – tai procesas, kurio metu du lengvesni atomų branduoliai susijungia ir sudaro sunkesnį branduolį, išskirdami energiją. Tai ta pati reakcija, kuri vyksta Saulėje ir žvaigždėse. Ji taip pat paaiškina filme „Atgal į ateitį“ (*Back to the Future*) vaizduojamos „DeLorean“ laiko mašinos srauto kondensatoriaus (angl. *flux capacitor*) ir Tony'io Starko, žinomo kaip Geležinis Žmogus (*Iron Man*), egzoskeletui energiją tiekiančio lankinio reaktoriaus veikimą. Mokslininkai daug metų ieškojo švarios energijos šaltinio, tačiau bet koks proveržis nuolat nusikeldavo dešimtmečius. Dauguma tyrimų apsiribodavo teorijomis ir koncepcijos pagrindimu. Tačiau keturių akademikų įkurta „Helion“ skelbėsi galinti sukurti branduolinės sintezės reaktorių vos už keliasdešimt milijonų dolerių (o ne dešimtis milijardų) ir taip įgalinti žmoniją atsisakyti iškastinio kuro.

Nors tokios drąsios, pasaulį galinčios pakeisti idėjos skambėjo beprotiškai, būtent jos Altmaną labiausiai ir domino. Jis jau seniai norėjo įkurti branduolinės energijos bendrovę, o dabar galėjo tiesiog investuoti į tokią.

Jis puikiai suprato einantis nepramintu taku – investuojantieji į technologijų bendroves dažniausiai rinkdavosi programinę įrangą kuriančias įmones su tradiciniais verslo modeliais ir aiškesnėmis pelno strategijomis. Tačiau Altmanas tvirtai tikėjo, kad tokios įmonės gali ne tik pagerinti žmonių gyvenimą, bet ir atnešti didelį pelną. „Silicio slėniui turėtų būti gėda, kad ši įmonė vis dar nesulaukė finansavimo“, – sakė jis viename interviu, kalbėdamas apie „Helion“. Altmano teisuoliška laikysena nebuvo kuo nors išskirtinė. Ją galima palyginti su ideologine pozicija, kurią deklaravo kiti didžiųjų technologijų bendrovių vadovai, pavyzdžiui, Elonas Muskas, dar atviriau kalbėjęs apie savo tikslą išgelbėti žmoniją.

„Dar viena mobilioji programėlė? Vargu, ar ką nors nustebinsi, – sykį pasakė Altmanas. – Raketas kurianti bendrovė? Visi nori į kosmosą.“ Silicio slėnyje visi skelbiasi norintys išgelbėti pasaulį. Tačiau Altmanas, kaip ir Muskas, stengėsi susikurti tikro technologijų pasaulio gelbėtojo, rimtai žiūrinčio į savo tikslus, įvaizdį. Dauguma technologijų srities verslininkų nebyliai sutarė, kad kalbos apie žmonijos išgelbėjimą dažniausiai tėra rinkodaros triukas, skirtas paveikti visuomenę ir įmonių darbuotojus. Laikyta, kad jų kuriami produktai – tai banalias funkcijas atliekantys įrankiai, skirti, pavyzdžiui, el. paštui tvarkyti ar skalbiniams skalbti. Tačiau Altmanas siekė, kad „Y Combinator“ bendruomenė virstų rimta, didesnio masto verslininkų sąjunga, galinčia iš tiesų keisti pasaulį. Tokie sprendimai buvo rizikingesni, bet būtent jie sulaukė didžiausio dėmesio.

Investuodamas Altmanas elgdavosi kaip pokerio žaidėjas, kuris, turėdamas tik pusėtiną kombinaciją, meta ant stalo beveik visus savo žetonus, stebint kitiems apstulbusiems žaidėjams. Pats Altmanas tokį polinkį aiškino trūkstama smegenų dalimi – ta, kuri verčia rūpintis, ką apie jį pagalvos kiti. Dėl tokio požiūrio jis galėjo tiksliau vertinti riziką ir nebijojo investuoti į beprotiškai skambančias idėjas.

Kai Altmanas vis tik priimdavo rizikingus investicinius sprendimus, galimas nesėkmes sušvelnindavo jo sukauptas turtas ir reputacija – startuolių pasaulyje jis buvo laikomas naujuoju Yoda. Silicio slėnyje gera reputacija yra vertingesnė už bet kokią vilą ar sportinį automobilį. Tokie verslininkai kaip Altmanas, investavę į branduolinės sintezės srityje besispecializuojančius startuolius, įgydavo prestižą, kuris buvo vertas tiek pat, kiek ir realus uždarbis. Greta investicijų į DI didžiąją dalį savo lėšų Altmanas galiausiai sutelkė dviejose ambicingose srityse: viena buvo susijusi su siekiu prailginti žmonių gyvenimo trukmę, o kita – su neribotų energijos išteklių paieškomis. Daugiau nei 375 mln. JAV dolerių jis investavo į bendrovę „Helion“, dar 180 mln. JAV dolerių – į „Retro

Biosciences“, įmonę, siekiančią pailginti vidutinę žmogaus gyvenimo trukmę dešimtmečiu.

Jei svarstote, iš kur Altmanas gavo tiek pinigų, prisiminkite, kad jis buvo investavęs apie 3 mln. JAV dolerių į bendrovę „Cruise“ dar gerokai prieš tai, kai ši buvo parduota „General Motors“ už 1,25 mlrd. JAV dolerių. Tai atnešė jam nemenką pelną. Be to, eidamas „Y Combinator“ vadovo pareigas, Altmanas turėjo išskirtinę galimybę susipažinti su šimtais jau atrinktų įmonių, todėl buvo geresnėje padėtyje nei daugelis rizikos kapitalo investuotojų – ypač tuo metu, kai akcijų rinkoje vyko vienas didžiausių visų laikų pakilimų. Galimybė pirmam išgirsti daugybės startuolių idėjas padėjo Altmanui numatyti ateitį.

Praėjus vos metams nuo tada, kai tapo „Y Combinator“ vadovu, Altmanas jau buvo įtvirtinęs savo, kaip naujojo Silicio slėnio guru, reputaciją. Per savaitę jis sulaukdavo maždaug keturių šimtų prašymų susitikti. Jis tapo traukos centru investuotojams ir startuolių įkūrėjams, norintiems per jį pasiekti kitus „Y Combinator“ startuolius ir partnerius arba tiesiog susipažinti su juo pačiu. Jų akimis, jis buvo dar ambicingesnė, tarsi iš mokslinės fantastikos kūrinio nužengusi Paulo Grahamo versija. Tinklaraštyje *samaltman.com* Altmanas gvildeno temas, kurios akivaizdžiai peržengė jo kompetencijos ribas: jis rašė apie NSO, reguliavimą, patarinėjo, kaip tapti geru pašnekovu prie vakarienės stalo. „Neklauskite žmonių, kuo jie dirba, – patarė jis. – Verčiau paklauskite, kuo jie domisi.“

Grahamas tradiciškai kas savaitę rengdavo konsultacinius susitikimus su „Y Combinator“ remiamų startuolių įkūrėjais, per kuriuos aptardavo jiems iškilusias problemas. Jis dažnai dalindavo lakoniškus patarimus, kurie atspindėdavo kertinį „Y Combinator“ principą: kurti tai, ko žmonės iš tikrųjų nori. Bendraudamas su startuolių įkūrėjais, Altmanas skatino juos mąstyti plačiau. Kai „Airbnb“ sumanytojai – grupelė vaikinų, sukūrusių programėlę nakvynės ieškantiems keliautojams, – parodė Altmanui potencialiems investuotojams skirtą pristatymą, jis patarė

visus milijonus pakeisti milijardais. „Arba jūs gėdijatės savo idėjos, arba aš nemoku skaičiuoti“, – pasakė Altmanas, įdėmiai į juos žvelgdamas didelėmis mėlynomis akimis.

Altmanas ragino startuolių įkūrėjus negailėti jėgų ir būti tokiems pat veržliems bei užsispyrusiems kaip ir jis pats. „Kad pasiektum sėkmę, turi būti kone beprotiškai atsidavęs savo įmonei“, – sakė jis. Tinklaraštyje jis rašė, kad prie bet kurio numatyto sėkmės rodiklio reikia „pridėti nulį“. Pasak jo, norėdami spręsti pasaulinio lygio problemas, įkūrėjai turi būti apsėsti produkto kokybės, nepaprastai išradingi ir gebėti nuolat aiškiai komunikuoti su savo komanda. Apie darbo ir asmeninio gyvenimo pusiausvyrą negali būti nė kalbos.

Altmano žodžiuose buvo daug tiesos. Silicio slėnis – tai vieta, kurioje kuriamos imperijos, o imperijos juk nepastatysi per keturiasdešimties valandų darbo savaitę. Jo, kaip verslininko, didžiausia stiprybė buvo gebėjimas palenkti kitus į savo pusę. Altmaną vertino ne tik jo mentoriai, įskaitant vidurinės mokyklos direktorių, „Y Combinator“ bendrakūrėjus Paulą Grahamą ir Jessicą Livingston bei Peterį Thielį, bet ir tūkstančiai startuolių įkūrėjų. Vis dėlto jame glūdėjo tam tikra prieštara: jis turėjo genialų protą ir troško išgelbėti pasaulį, bet drauge jautėsi emociškai atitolęs nuo paprastų žmonių, kuriuos norėjo išgelbėti.

Iš dalies tokią vidinę būseną galima paaiškinti prisiminus tą karštą 2006-ųjų vasaros dieną, kai Altmanas vienais šortais gulėjo ant grindų, susikrimtęs dėl nepavykusio sandorio. Norėdamas susitvarkyti su vis stiprėjančiu nerimu, jaunasis verslininkas ėmė praktikuoti meditaciją. Užmerkęs akis, jis galėdavo visą valandą išsėdėti sutelkęs dėmesį į kvėpavimą. Vėliau pats Altmanas pasakojo, kad jo savasties jausmas su laiku vis silpnėjo.

„Vienas dalykas, kurį supratau medituodamas, – tai, kad apskritai nėra jokio vidinio „aš“, su kuriuo galėčiau tapatintis, – sakė jis tinklalaidėje „Art of Accomplishment“. – Esu girdėjęs, kad nemažai

žmonių, daug galvojančių apie [pažangų DI], tokią būseną pasiekia ir kitais keliais.“

Praėjus keleriems metams, su bičiuliais leisdamasis į žygį, Altmanas staiga aiškiai suprato, kad ateityje kompiuteriai tikrai sugebės atkartoti žmogaus protą. Jis priėjo prie išvados, kad mąstymo procesai gali vykti ir kompiuteriuose, su kuriais vieną dieną susiliesime. „Dirbdamas su DI, neišvengiamai imi kelti gilius filosofinius klausimus. Pavyzdžiui, kas nutiks, kai mano sąmonė bus perkelta į kompiuterį? – svarstė jis. – Kas bus, jei dirbtinis protas pradės su manimi kalbėtis? Ar aš noriu susijungti su ta sistema? Ar, susiliejęs su ja, norėsiu leistis tyrinėti visatą? Kiek visame tame vis dar bus likę manęs?“ Altmanas buvo ne vienintelis, mėgęs liesti mokslinę fantastiką primenančias temas. Jį supo technologijų entuziastai, kurie taip pat tikėjo, kad vieną dieną jų sąmonę bus galima perkelti į kompiuterių serverius, ir tokiu būdu jie galės gyventi amžinai.

Altmaną akivaizdžiai gąsdino mintys apie mirtį. Jis vadino save išgyvenimo entuziastu ir skyrė nemažai laiko bei pinigų pasiruošimui galimai visuotinei katastrofai, pavyzdžiui, žmogaus sukurto viruso protrūkiui ar net DI atakai. „Stengiuosi apie tai per daug negalvoti, – sakė jis grupei startuolių įkūrėjų, kaip cituojama žurnale „New Yorker“. – Tačiau turiu ginklų, aukso, kalio jodido, antibiotikų, baterijų, vandens atsargų, taip pat dujokaukių, gautų iš Izraelio ginkluotųjų pajėgų, ir nemažą sklypą Big Seryje, į kurį, esant reikalui, galėčiau nuskristi.“

Dar daugiau, startuoliui „Nectome“ (taip pat dalyvavusiam „Y Combinator“ programoje) jis buvo sumokėjęs 10 tūkst. JAV dolerių, kad patektų į laukiančiųjų sąrašą dėl galimybės po mirties išsaugoti savo smegenis. Šiuo tikslu įmonė ketina taikyti pažangų konservavimo metodą, kad smegenys galėtų būti perkeltos į virtualiąją erdvę, o vėliau ateities mokslininkų paverstos skaitmenine simuliacija.

Altmanas vis aktyviau investavo į ateitį keičiančius projektus, ir jam, regis, pasireiškė vadinamasis „apžvalgos efektas“ – didelės nuostabos ir

vidinių transcendentinių potyrių lydimas psichologinis virsmas, kurį patiria astronautai, žvelgdami į Žemę iš kosmoso. Į pasaulį Altmanas vis dažniau žiūrėjo tarsi iš kosmoso perspektyvos. Pokalbiuose jo mąslus žvilgsnis neretai nuklysdavo, jo žodžius lydėjo apmąstymų kupinos pauzės, ir apskritai jis labiau priminė stebėtoją nei aktyvų dalyvį.

Nors jaunasis vizionierius investavo į žmonijos ateitį, jis vis labiau juto tam tikrą psichologinį bei emocinį atotrūkį nuo kitų žmonių. „Kad galėtum spręsti jų problemas, turi būti ramus, nuosaikus ir pragmatiškas“, – teigė jis. Altmanas dažnai remdavosi amerikiečių mokslinės fantastikos rašytojo Marco Stieglerio apsakymu „Švelnios vilionės“ (*The Gentle Seduction*), kuriame pasakojama apie technologijų poveikį ateities žmonių gyvenimui. Pagrindinė veikėja Lisa susiduria su įvairiais technologiniais pasiekimais, kurie palaipsniui „sugundo“ ją integruoti šias naujoves į kasdienį gyvenimą.

Galiausiai Lisa su vyru ryžtasi perkelti savo sąmonę į kompiuterį. Tai rizikinga procedūra – žmonės, kurių sąmonė susilieja su pažangia mašina, gali prarasti savo tapatybę. Todėl Lisa apsvarsto visus pliusus ir minusus, prisimindama, kad kai kurie jos draugai, pabandę tai padaryti, „mirė“, arba, kitaip tariant, pranyko skaitmeninėje terpėje. Stiegleris rašo: „Išgyveno tik tie, kurie gebėjo būti atsargūs, bet nepasidavė baimei, – tie, kuriems būdingas prigimtinis apdairumas.“

Ši citata Altmanui padarė didelį įspūdį – jis dažnai ją kartodavo. Kaip teigia autorius, ištverti sąmonės perkėlimą į kompiuterį gali tik tie, kurių mąstyme dera atsargumas ir drąsa. Altmanas tikėjo, kad, norint susidoroti su ateities pavojais, būtina išlikti ramiam, šaltakraujiškam ir gebėti racionaliai vertinti grėsmes, o ne pasiduoti emocijoms ar panikai. Ateityje klestės tie, kurie į technologinę pažangą žiūrės sąmoningai, išlaikydami emocinį atstumą.

Tuo metu kai kurie technologijų srities specialistai pernelyg pasiduodavo nerimui dėl galimų DI pavojų (pastaruosius tyrinėjo dar tik besiformuojanti DI saugumo tyrimų sritis). Nors tokie tyrimai išties

svarbūs, būdavo, kad nerimas peraugdavo į baimės kurstymą. Atrodė, kad kai kurie žmonijos gynėjais apsiskelbę asmenys leidosi užvaldomi emocijų. „Deja, kai kuriose DI saugumu besirūpinančiose bendruomenėse galima sutikti pačių nervingiausių žmonių, – sakė Altmanas. – Tai pavojinga situacija. Tose bendruomenėse tvyro įtampa.“ Vis dėlto, kaip jis pats šiandien teigia, būtent tada jis ėmė suprasti norintis dirbti bendrojo DI (BDI) kūrimo srityje. Terminą „bendrasis dirbtinis intelektas“ (angl. *artificial general intelligence,* AGI) keleriais metais anksčiau buvo pasiūlęs Shane'as Leggas, tačiau idėja sukurti mašiną, kurios galimybės prilygtų žmogaus gebėjimams, sklandė jau kelis dešimtmečius. Iš dalies ją įkvėpė mokslinės fantastikos vizijos. Tuo metu sumanymas sukurti BDI jau nebebuvo laikomas bepročių kalbomis (kad tokiais jų nepalaikytų, „DeepMind“ įkūrėjai kadaise savo susitikimui netgi specialiai pasirinko italų restoraną) – dabar tai tapo rimtu moksliniu tikslu.

Pasauliui reikėjo žmogaus, kuris į DI kūrimą žvelgtų nuosaikiai. Stieglerio žodžiuose apie „prigimtinį apdairumą“ Altmanas atpažino save – jis jautėsi gebantis sumaniai orientuotis sudėtingame ir potencialiai pavojingame ateities pasaulyje, išlikdamas atsargus, bet nepasiduodamas baimei. Jis galėjo tapti budriu sargu, stovinčiu ant bokšto krašto ir stebinčiu horizonte ryškėjančią DI utopiją, tik retkarčiais pažvelgdamas į apačioje kunkuliuojantį gyvenimą. Tačiau ši misija ir tikėjimas savimi taip jį apakino, kad jis nesugebėjo savo veiksmuose įžvelgti ironijos: laikydamas save apdairiu verslininku, jis vis dėlto įsivėlė į nuožmią konkurencinę kovą ir skubėjo pirmas rinkai pristatyti DI sistemas, aplenkdamas visas kitas technologijų bendroves, įskaitant „Google“. Be kita ko, Altmaną nejučiomis užvaldė noras būti pirmam.

Gali būti, kad jis niekada nebūtų ėmęsis kurti DI utopijos, jei ne kitas vizionierius, pasufleravęs tokią idėją. Silicio slėnio verslininkui reikėjo varžovo, kuris pažadintų jame užsidegimą. Tuo varžovu tapo talentingas jaunas žaidimų kūrėjas iš Anglijos, planavęs sukurti tokią galingą

programinę įrangą, kuri ne tik suteiktų galimybę padaryti revoliucinių mokslinių atradimų, bet ir padėtų leistis į dieviškumo paieškas.

1 Brogrameris (angl. *brogrammer*) – šnekamojoje kalboje vartojamas terminas, apibūdinantis stereotipiškai vyrišką technologijų srities atstovą (dažniausiai dirbantį programuotoju Silicio slėnio įmonėje).

4 SKYRIUS

Tobulesnės smegenys

Žlugus „Elixir Studios“, Demisas Hassabisas tapo dar vienu nesėkmę patyrusiu technologijų srities verslininku, kurio svajonės pasirodė esančios pernelyg drąsios. Nors ši patirtis jam buvo skausminga, jis vis dar turėjo tai, kas, jo manymu, jį išskyrė iš daugelio kitų startuolių įkūrėjų ir apskritai žmonių – savo smegenis. Hassabisas be galo rūpinosi savo smegenų sveikata: mąstymą lavino žaisdamas žaidimus, o norėdamas išlaikyti aštrų protą, vengė alkoholio. Jis netgi buvo pasirinkęs galvos smegenų magnetinio rezonanso vaizdą kaip savo „Facebook“ profilio nuotrauką. Hassabisas negalėjo atsistebėti jų sudėtingumu, o žlugus „Elixir“, vis dažniau svarstydavo, ar pačios smegenys galėtų būti raktas į programinės įrangos, intelektu prilygstančios žmogui, sukūrimą. Juk jos – vienintelis visatoje egzistuojantis įrodymas, kad bendrasis intelektas apskritai įmanomas. Todėl sveikas protas jam kuždėjo, kad reikia gilintis į jų veikimą. Ar visa tai – tik biologija, ar kas nors daugiau? Atsakymo reikėjo ieškoti neuromoksle.

Hassabisas visada troško aiškumo, nesvarbu, ar tai būtų susiję su žaidimo baigtimi, moraliniais krikščionybės principais, ar visa paaiškinančio visatos modelio, apie kurį skaitė dar mokykloje, paieškomis. Geriausiai jam sekėsi tose srityse, kuriose galėjai viską nusakyti skaičiais ar apibrėžti taisyklėmis. „Daugumą smegenų funkcijų vienaip ar kitaip turėtų būti įmanoma atkartoti kompiuteriu, – vėliau sakė jis viename interviu. – Neuromokslas rodo, kad smegenų veikimą galima paaiškinti mechanistiškai.“ Kitaip tariant, net ir tokį sudėtingą

dalyką kaip smegenys įmanoma nusakyti skaičiais ir duomenimis – kaip įprastą mašiną.

Taip svarstydamas, Hassabisas įkvėpimo sėmėsi iš Alano Turingo – XX a. britų kompiuterių mokslo pradininko, dar 1936 metais pristačiusio mintinį eksperimentą – Turingo mašiną. Tai teorinis įrenginys su begaline juosta, suskirstyta į langelius, ir galvute, kuri pagal tam tikras taisykles gali skaityti bei rašyti simbolius, kol gaunamas nurodymas sustoti. Nors ši idėja gali pasirodyti primityvi, teoriškai ji buvo itin svarbi, nes padėjo pagrindą suvokimui, kad kompiuteriai gali vykdyti užduotis pagal algoritmus – taisyklių rinkinius. Skyrus pakankamai laiko ir išteklių, Turingo mašina teoriškai galėtų būti tokia pat galinga kaip bet kuris šiandieninis skaitmeninis kompiuteris. Hassabisui ji atrodė kaip puikus žmogaus proto atitikmuo. „Žmogaus smegenys ir yra Turingo mašina“, – kartą pasakė jis.

2005 metais, praėjus keliems mėnesiams po „Elixir Studios“ uždarymo, Hassabisas pasinėrė į neuromokslo doktorantūros studijas Londono universiteto koledže. Jo disertacija, kurioje buvo nagrinėjama atmintis, buvo gana trumpa, bet moksliniu požiūriu itin vertinga – taip ją vertino kiti kompiuterių mokslo akademikai. Iki tol manyta, kad hipokampas daugiausia apdoroja prisiminimus, tačiau Hassabisas, remdamasis kitais MRT vaizdų tyrimais, parodė, kad ši smegenų sritis aktyvuojasi ir veikiant vaizduotei.

Paprastai tariant, prisimindami tam tikrus dalykus, iš dalies juos iš naujo atkuriame vaizduotėje. Taigi praeities įvykiai smegenyse nėra tiesiog „iškeliami“ tarsi ištraukiant juos iš dokumentų spintelės, bet aktyviai rekonstruojami, panašiai kaip dailininkui tapant paveikslą. Tai daug dinamiškesnis ir kūrybiškesnis procesas, paaiškinantis, kodėl mūsų prisiminimai kartais būna netikslūs arba iškreipti ankstesnės patirties. Hassabisas tvirtino, kad smegenys šį „vaizdinių kūrimo“ procesą pasitelkia ir kitoms užduotims atlikti, pavyzdžiui, bandant susiorientuoti žemėlapyje ar planuojant.

Vienas prestižinis recenzuojamas žurnalas Hassabiso disertaciją įvardijo kaip vieną svarbiausių tų metų mokslinių proveržių. Vis dėlto pats Hassabisas nenorėjo užsibūti akademiniame pasaulyje. Mokslininkai, siekiantys Nobelio premijos vertų atradimų, daugiau nei pusę savo laiko turėdavo skirti paraiškų finansavimui gauti rašymui. Net ir gavę lėšų, dauguma universitetų neturėjo reikiamos skaičiavimo galios. Norint atlikti pažangius mašininio mokymosi tyrimus, reikėjo prieigos prie galingiausių pasaulio kompiuterių. Tokie ištekliai, kaip ir didžiausi pasaulio talentai, buvo sutelkti tik didžiosiose technologijų bendrovėse. Taigi norėdamas suburti galingiausius protus, dirbsiančius prie šiuolaikinio „Manhatano projekto“, Hassabisas turėjo įsteigti įmonę.

Pirmosios idėjos ėmė ryškėti prie pietų stalo diskutuojant su dviem bendraminčiais – Shane’u Leggu ir Mustafa Suleymanu. Retai sutiksi tokį DI entuziastą kaip Leggas – jo idėjos apie DI ateitį pranoko net paties Hassabiso vizijas. Jis buvo parašęs daktaro disertaciją apie mašinų superintelektą, o disertacijos vadovas jam pasiūlė pasikalbėti su Hassabisu.

„Radau gimininga sielą, – prisimena Hassabisas. – Shane’as ir pats puikiai suprato, kad tai vienas svarbiausių dalykų, kuriuos galima nuveikti gyvenime.“

Leggo idėjos jau buvo sukėlusios didelį atgarsį glaudžioje „technologinio singuliarumo“ bendruomenėje. Jai priklausantys tyrinėtojai įsitikinę, kad ateityje technologinė pažanga pasieks tokį tašką, kai taps nesustabdoma ir nesuvaldoma. Akivaizdžiausią to požymį pamatysime tada, kai kompiuteriai taps protingesni už žmones. Leggas manė, kad taip nutiks maždaug 2030-aisiais.

Naujojoje Zelandijoje augusio Shane’o Leggo kelias į mokslo pasaulį prasidėjo kiek neįprastai. Jo tėvai, susirūpinę, kad mokykloje jų sūnui nesiseka, devynerių sulaukusį berniuką nuvedė pas švietimo psichologą. Šis jam davė atlikti intelekto testą. Galiausiai, kiek suirzęs, specialistas tėvams pasakė, kad berniukas turi disleksiją, bet jo intelektas – išskirtinai

aukštas. Išmokęs naudotis klaviatūra, Leggas šovė į mokyklos pažangiausiųjų sąrašus ir tapo vienu geriausių mokinių matematikos ir programavimo srityse.

Būdamas dvidešimt septynerių, aukštas, kiek gunktelėjęs jaunuolis trumpais plaukais kartą užsuko į knygyną. Jo dėmesį iškart patraukė Ray'aus Kurzweilo knyga „Sąmoningų mašinų amžius" (*The Age of Spiritual Machines*), kurioje prognozuojama, kad vieną dieną kompiuteriai įgis laisvą valią ir gebės patirti emocinių bei dvasinių išgyvenimų.

Vienu prisėdimu perskaitęs šią knygą, Leggas negalėjo liautis galvojęs apie Kurzweilo idėjas ir jo prognozę, kad galingas DI atsiras jau XXI a. trečiojo dešimtmečio pabaigoje. Skaičiavimo galia ir duomenų kiekiai auga eksponentiškai. Jei ši tendencija tęsis, kompiuteriai galiausiai pranoks žmones. Tokią prielaidą sustiprina ir vienas pamatinių technologijų pramonės principų – vadinamasis Moore'o dėsnis, teigiantis, kad tranzistorių skaičius mikroschemose kas dvejus metus padvigubėja. Ši prognozė pasitvirtina jau pusę amžiaus.

2000-aisiais, kai Leggas susipažino su Kurzweilo idėjomis, pasaulis dar nebuvo atsigavęs po sprogusio „dotcom" burbulo, todėl buvo sunku patikėti, kad kompiuterių galia ir toliau didės. Tačiau Leggas nė neabejojo, kad internetas nepaliaujamai vystysis.

„Buvo aišku, kad įvairūs jutikliai padės sumažinti kaštus, todėl duomenų, kuriuos būtų galima naudoti modeliams mokyti, tik daugės", – dabar sako jis.

Didelė skaičiavimo galia ir gausūs duomenys leidžia mokyti mašinas, kad jos taptų vis protingesnės. Leggas, susidomėjęs DI, ėmėsi doktorantūros studijų šioje srityje ir per tą laiką užmezgė daug vertingų pažinčių. Kartą jis su keliais kitais mokslininkais sulaukė elektroninio laiško iš Beno Goertzelio – hipiškos išvaizdos DI mokslininko ir technologinio singuliarumo šalininko. Pastarasis ieškojo idėjų savo naujos knygos pavadinimui. Jam reikėjo termino, kuris apibūdintų DI, turintį žmogaus gebėjimų. Leggas jam atrašė ir pasiūlė terminą

„bendrasis dirbtinis intelektas“. Vėliau ši sąvoka taps neatsiejama nuo Hassabiso bei kelių didžiausių pasaulio technologijų bendrovių vizijos.

Ilgus metus tokie žmonės kaip Hassabisas, Leggas ir kiti DI tyrėjai vartojo terminus „galingas DI“ arba „tikras DI“, apibūdindami programinę įrangą, pasižyminčią žmogui būdingu intelektu. Tačiau būdvardis „bendrasis“ pabrėžė svarbią mintį: žmogaus smegenys yra išskirtinės tuo, kad gali atlikti daugybę skirtingų veiksmų – nuo skaičiavimo, apelsino lupimo iki poezijos kūrimo. Mašinas galima užprogramuoti, kad jos gebėtų gana efektyviai padaryti vieną iš šių užduočių, bet nė viena nepajėgi atlikti jų visų. Jei kompiuteris gebėtų ne tik skaičiuoti, bet ir prognozuoti, atpažinti vaizdus, kalbėti, kurti tekstus, planuoti ir „įsivaizduoti“, jo galimybės priartėtų prie žmogaus gebėjimų.

Tuo metu dauguma DI srityje dirbusių mokslininkų atmetė mintį, kad DI kada nors galėtų prilygti žmogaus intelektui. Tokias nuostatas iš dalies lėmė jų asmeninė patirtis stebint DI raidą lydėjusius perdėto entuziazmo proveržius ir vėlesnius nusivylimus. Žmonės itin susidomėdavo DI galimybėmis, bet galiausiai nusivildavo. Per visą DI kūrimo istoriją ne kartą būta pakilimo laikotarpių, po kurių įvykdavo nuosmukiai – vadinamosios technologinės žiemos. Tai buvo technologinio sąstingio periodai, kai tyrėjų finansavimas mažėjo, o pažanga vyko be galo lėtai. XX a. dešimtajame dešimtmetyje ir XXI a. pirmojo dešimtmečio pradžioje tyrėjams pavyko pritaikyti mašininio mokymosi metodus specifinio pobūdžio užduotims, pavyzdžiui, veidų ar kalbos atpažinimui. Vis dėlto 2009 metais, kai Hassabiso doktorantūros studijos ėjo į pabaigą, beveik niekas nebetikėjo, kad mašinos galėtų turėti BDI. Tokia idėja atrodė verta pašaipų, o jos vieta buvo mokslo užribyje.

Tačiau Goertzelis pats priklausė mokslo periferijai. Nors terminas „bendrasis dirbtinis intelektas“ nebuvo itin skambus, jis pakankamai sužavėjo autorių, kad šis jį panaudotų kaip savo knygos pavadinimą. Taip „bendrasis dirbtinis intelektas“ tapo plačiai vartojamu terminu, vėliau prisidėjusiu prie DI srities išpopuliarėjimo.

Kalba ir terminologija ilgainiui atliko itin svarbų vaidmenį DI srityje, o tai ne visada buvo laikoma teigiamu dalyku. Terminas „dirbtinis intelektas“ gimė 1956 metais Dartmuto koledže vykusio seminaro apie „mąstančias mašinas“ metu. Tuo metu naujajai sričiai buvo svarstomi ir kiti pavadinimai, tokie kaip „kibernetika“ ar „kompleksinis informacijos apdorojimas“, tačiau ilgainiui įsitvirtino būtent šis terminas. „Dirbtinis intelektas“ tapo vienu sėkmingiausių visų laikų rinkodaros terminų ir paskatino daugelio kitų susijusių sąvokų atsiradimą. Tokiais terminais mašinos mūsų kolektyvinėje sąmonėje sužmoginamos, dažnai priskiriant joms daugiau gebėjimų, nei jos iš tikrųjų turi. Pavyzdžiui, techniškai netikslu sakyti, kad kompiuteriai „mąsto“ ar „mokosi“, tačiau tokios sąvokos kaip „neuroniniai tinklai“, „giluminis mokymasis“ ar „mokymas“ verčia mus manyti, kad programinė įranga turi žmogiškų savybių, nors iš tikrųjų ji tik iš dalies paremta žmogaus smegenų funkcijomis. Vienas dalykas, dėl kurio visi sutiko kalbėdami apie Shane’o Leggo pasiūlytą naująją koncepciją, – tai, kad toks reiškinys dar neegzistavo.

Mustafa Suleymanas buvo dar vienas žmogus, tikėjęs, kad BDI sukurti įmanoma. Būdamas dvidešimt penkerių, metęs studijas Oksfordo universitete, jis ieškojo būdų, kaip pasitelkus technologijas pakeisti pasaulį. Nors Suleymanas pasižymėjo išskirtiniu protu, jo kompetencijos sritys labiau siejosi su politika ir filosofija nei su kompiuterių mokslu. Siro ir anglės kraujo turintis Suleymanas jautė nenumaldomą norą spręsti problemas. Tačiau jo nedomino menkos problemos, tokios kaip sugedęs automobilis ar skaudanti koja. Jo žvilgsnis krypo į globalius iššūkius, liečiančius visą žmoniją, – skurdą, klimato krizę.

Nors prieš kurį laiką Suleymanas buvo įkūręs ginčų sprendimo paslaugų firmą, tuo metu jo dėmesį patraukė neuromokslas. Hassabisas pakvietė jį į neformalius susitikimus prie pietų stalo Londono universiteto koledže. Jiedu buvo pažįstami iš anksčiau: augdamas Šiaurės Londone, Suleymanas bičiuliavosi su Hassabiso broliu George’u

ir paauglystėje dažnai lankydavosi jų šeimos namuose. Trys bičiuliai, dar būdami studentiško amžiaus, net buvo nuvykę į Las Vegasą dalyvauti pokerio turnyre, kur ne tik vienas kitam patarinėjo, bet ir dalijosi laimėjimais.

Kai jie vėl susitiko, Suleymaną sužavėjo Hassabiso idėja sukurti galingą DI, galintį spręsti pačius įvairiausius uždavinius, ir Leggo įsitikinimas, jog toks intelektas pajėgtų susidoroti kone su bet kokiu iššūkiu. Jį apėmė entuziazmas svarstant, kaip tokia technologija galėtų prisidėti prie visuomenės problemų sprendimo.

Siekdama privatumo, trijulė susitikdavo italų tinkliniame restorane „Carluccio's" šalia universiteto. „Nenorėjome, kad žmonės girdėtų mūsų beprotiškas kalbas apie BDI kūrimą", – prisimena Leggas.

Įtikintas Hassabiso, Leggas pripažino, kad su akademinių institucijų resursais greičiausiai nepavyks sukurti BDI. „Iki to laiko, kol mums suteiks reikalingų išteklių, būsime įkopę į šeštąją dešimtį, – sako Hassabisas. – Tiesa ta, kad daugiausia nuveikti galiu tik įkūręs įmonę."

Kad užtikrintų atitinkamą veiklos mastą ir gautų reikiamų išteklių, jiems reikėjo įkurti startuolį. Suleymanas anksčiau jau buvo įkūręs įmonę, todėl, kaip ir Hassabisas, turėjo patirties versle. 2010 metais, kai didžiausią įtaką visuomenei darė tokios technologijų milžinės kaip „Google" ir „Facebook", trijulei atrodė logiška, kad būtent technologijų bendrovė turėtų geriausias galimybes modeliuoti pasaulio sudėtingumą. Jie sugalvojo ambicingą planą – įkurti tyrimų bendrovę, kuri sukurtų galingiausią kada nors egzistavusį DI ir panaudotų jį pasaulinėms problemoms spręsti.

Naująją įmonę jie pavadino „DeepMind", o jos generaliniu direktoriumi paskyrė Hassabisą. Bendražygiai iškart pasamdė vieną geriausių programuotojų iš „Elixir Studios" ir išsinuomojo biurą palėpėje priešais Londono universiteto koledžą, kuriame Hassabisas buvo įgijęs daktaro laipsnį. Nors tris bičiulius vienijo bendra misija, kiekvienas turėjo savų motyvų. Leggas sukosi tarp žmonių, kurie siekė kuo

glaudesnės žmonių ir BDI sąveikos, Suleymanas tikėjosi spręsti visuomenės problemas, o Hassabisas troško įeiti į istoriją kaip padaręs esminių visatos atradimų.

Neilgai trukus požiūrių skirtumai tarp jų įplieskė diskusijas. Suleymanas norėjo, kad Hassabisas perskaitytų jam didelę įtaką padariusią kanadiečių akademiko Thomaso Homerio-Dixono knygą „Išradingumo atotrūkis“ (*Ingenuity Gap*). Šioje 2000 metais išleistoje knygoje teigiama, kad šiuolaikinės problemos, tokios kaip klimato kaita ar politinis nestabilumas, darosi tokios sudėtingos, kad žmonės paprasčiausiai tampa nepajėgūs jų spręsti. Taigi atsiranda vadinamasis išradingumo atotrūkis, kurį galima įveikti tik diegiant naujoves, ypač technologijų srityje. Suleymano manymu, būtent čia DI galėtų suvaidinti esminį vaidmenį.

Tačiau Hassabisas papurtė galvą. „Tu nematai platesnio vaizdo“, – kartą jis pasakė Suleymanui, anot vieno pokalbį nugirdusio žmogaus. Hassabisas, regis, buvo įsitikinęs, kad Suleymano požiūris į DI pernelyg orientuotas į dabartį. Jo manymu, BDI turėtų padėti „DeepMind“ suprasti žmonių kilmę ir jų egzistencijos tikslą. Hassabisas, pavyzdžiui, teigė, kad klimato kaita yra žmonijos lemtis ir kad, tikėtina, Žemė ilgainiui nepajėgs užtikrinti išlikimo visiems jos gyventojams. Pastangas spręsti dabartines problemas jis laikė beprasmiškomis, nes, jo manymu, kai kurie įvykiai greičiausiai yra neišvengiami. Hassabisas netikėjo vis dažniau pasigirstančiais nuogąstavimais, esą superprotingos mašinos vieną dieną ims kelti grėsmę žmonijai ar netgi ją sunaikins. Priešingai – jis buvo įsitikinęs, kad BDI padės išspręsti didžiausias žmonijos problemas.

Hassabisas šią savo viziją apibendrino „DeepMind“ šūkiu: „Išspręskime intelekto klausimą, o tada imkimės spręsti visa kita.“ Šį šūkį jis įtraukė į investuotojams skirtą pristatymą.

Tačiau Suleymanas su tokiu požiūriu nesutiko. Kartą, kai Hassabiso nebuvo biure, Suleymanas paprašė vieno pirmųjų „DeepMind“ darbuotojų pakeisti šūkį nauju, kuris skambėjo taip: „Išspręskime

intelekto klausimą ir panaudokime intelektą tam, kad pasaulis taptų geresnis.“

Hassabisui tai nepatiko. Tam pačiam darbuotojui jis nurodė grąžinti senąjį šūkį, kuriame teigiama, kad, išsprendus DI klausimą, reikia imtis spręsti visas kitas problemas. Kaip tikri britai, vengdami tiesioginės konfrontacijos, savo vizijų skirtumus jie sprendė netiesiogiai, pasitelkdami darbuotojus kaip tarpininkus.

Suleymanas norėjo kurti BDI panašiai, kaip tai vėliau darys Samas Altmanas, – paleisdamas jį į pasaulį kuo anksčiau, kad jis iš karto teiktų praktinę naudą. Jis manė, kad geriausia sistemą tobulinti remiantis realaus pasaulio atsiliepimais, o ne siekiant už uždarų durų sukurti tobulą prototipą. Tačiau Hassabisas norėjo, kad „DeepMind“ būtų valdoma išsikėlus aiškų tikslą – panašiai, kaip žaidžiant šachmatų partiją. Jo siekis buvo ne tik spręsti realias problemas, bet ir atsakyti į klausimus, kurie ištisus šimtmečius glumino žmoniją: koks mūsų egzistencijos tikslas? Ar mes kilę iš dieviškosios būtybės?

Paklaustas, ar tiki Dievą, Hassabisas vengia tiesaus atsakymo: „Manau, kad visata kupina paslapčių, – sako jis. – Neteigčiau, kad tikiu tradicinį Dievą.“ Jis primena, kad Albertas Einsteinas tikėjo „Spinozos Dievą“. „Veikiausiai ir aš atsakyčiau panašiai“, – teigia jis.

Baruchas Spinoza buvo XVII a. filosofas, laikęsis panteistinės nuostatos, kad Dievas – tai pati gamta ir visa, kas egzistuoja, o ne atskira antgamtinė būtybė. „Spinoza laikė gamtą Dievo įsikūnijimu, – aiškina Hassabisas. – Tad mokslas gali padėti tą paslaptį įminti.“

Remiantis tokia filosofija, nebūtų beprotiška manyti, kad BDI sukūrimas galėtų tapti dvasinio ar kvazireliginio lygmens atradimu, ypač jei laikytumėsi Spinozos požiūrio, kad Dievas – tai gamtos dėsniai. Dirbtinis intelektas galėtų padėti giliau perprasti šiuos dėsnius ir visatos paslaptis, todėl teoriškai būtų įmanoma atskleisti, kas slypi už jos atsiradimo. Gebėdamas analizuoti milžiniškus kiekius duomenų, DI galėtų tyrinėti sudėtingiausias visatos sistemas – nuo kvantinės

mechanikos iki kosminių reiškinių – ir pateikti įžvalgų apie pačią būties prigimtį. Sukūręs visatos sudėtingumą atkartojančią simuliaciją, DI pajėgtų atskleisti jos veikimo dėsningumus.

Jeigu BDI tyrimai patvirtintų Kurzweilo mintį, kad mūsų visata yra simuliacija, tuomet jos „programuotojas“ ir būtų toji dieviškoji būtybė. Be to, jei žmonės sukurtų galingą mašiną, galinčią apdoroti ir išanalizuoti visą turimą informaciją apie fiziką ir visatą, ji galėtų pateikti naujų teorijų, leidžiančių daryti prielaidą apie aukštesniųjų jėgų egzistavimą. Tokia mašina galėtų atsakyti į sudėtingiausius egzistencinius klausimus ir galbūt atskleisti dieviškosios esybės buvimo įrodymus. Dirbtiniam intelektui tobulėjant ir tampant vis galingesniam, atsirastų daugiau galimybių įminti didžiąsias žmonijos paslaptis.

Hassabiso siekį DI panaudoti visatos paslaptims atskleisti iš dalies gali paaiškinti jo religinė praeitis. 2023 metais Virdžinijos universiteto atliktas tyrimas, kuriame dalyvavo daugiau nei 50 tūkst. žmonių iš dvidešimt vienos šalies, parodė, kad tikintys Dievą ar dažnai apie jį galvojantys žmonės labiau linkę pasitikėti DI sistemų, tokių kaip „ChatGPT“, patarimais. Tyrėjų teigimu, tokie žmonės yra nuolankesni, todėl labiau linkę priimti DI pateikiamą informaciją. Jie taip pat greitai ir lengvai pastebi žmonių silpnybes.

Hassabisas dažnai svarstydavo apie žmonijos kilmę, todėl pradžioje su „DeepMind“ darbuotojais kartais kalbėdavosi apie Dievą. Keletas žmonių, dirbusių su juo ar pažinojusių jį asmeniškai, prisimena jį kaip aktyviai praktikuojantį krikščionį. Vienas jų teigia, kad pagrindinė jo motyvacija kurti BDI buvo Dievo paieškos.

„Daug diskutavome apie Dievą, – sako vienas kolega, dirbęs su Hassabisu tuo metu, kai buvo kuriama „DeepMind“. – Klausėme savęs: ar galėtume sukurti mašiną, kuri gebėtų atkurti visatos vystymosi raidą ir padėti įminti jos paslaptis? Bendrasis dirbtinis intelektas galėtų suteikti įžvalgų apie tai, iš kur mes kilome ir kas yra Dievas.“ Hassabisas taip pat buvo įsitikinęs, kad vadovauja moderniųjų laikų „Manhatano projektui“.

Kaip pasakoja du buvę „DeepMind“ darbuotojai, jis buvo perskaitęs knygą „Atominės bombos kūrimas“ (*The Making of the Atomic Bomb*), kuri jį įkvėpė komandos darbą organizuoti taip, kaip tai darė Robertas Oppenheimeris, – suskaidant pagrindinę užduotį į kelias dalis ir pavedant jas atskiroms mokslininkų grupėms.

Kad galėtų įgyvendinti tokį ambicingą tikslą, Hassabisas turėjo gauti lėšų „DeepMind“ plėtrai. Deja, britų investuotojai mainais į dalį naujojo startuolio akcijų pasiūlė menkas sumas, siekiančias vos 20–50 tūkst. svarų sterlingų. To nepakako, kad būtų galima pasamdyti talentingus specialistus BDI kurti, ką ir kalbėti apie prieigą prie galingų kompiuterių. Padėties nepagerino ir tai, kad jo verslo idėja – sukurti galingiausią pasaulyje DI sistemą – tradiciškai konservatyvumu garsėjančioje Didžiojoje Britanijoje atrodė pernelyg nutolusi nuo realybės ir ambicinga. Jungtinėje Karalystėje technologijų startuoliai dažniausiai rinkdavosi „žemiškas“ idėjas, kurios greitai atneštų pelną, pavyzdžiui, apsiimdavo sukurti finansinę programėlę, skirtą prekybai akcijomis ir obligacijomis. Hassabisui ir jo bendražygiams neliko nieko kito, kaip tik investuotojų ieškoti Silicio slėnyje – ten rėmėjai nebijojo futuristinių idėjų ir pasiryždavo atseikėti gerokai dosnesnes sumas.

Laimei, Leggas turėjo naudingų ryšių. 2010 metų birželį jis buvo pakviestas sakyti kalbą „Singuliarumo konferencijoje“ (*Singularity Summit*) – kasmetiniame renginyje, kurio sumanytojas – futuristas Ray’us Kurzweilas, dar jaunystėje Hassabisą pakerėjęs savo idėjomis, ir milijardierius investuotojas Peteris Thielis, aktyviai rėmęs pažangiausių technologijų projektus. Šioje konferencijoje, į kurią suplaukė nešabloniškai mąstantys DI srities mokslininkai, buvo aptariamos neregėtos technologijų atveriamos galimybės ir jų keliamos grėsmės. Renginio toną uždavė Thielis, garsėjantis idealistinėmis idėjomis. Jis nemanė, kad technologinis singuliarumas – tas ateities momentas, kai DI negrįžtamai pakeis žmoniją, – yra problema. Priešingai, jis nerimavo, kad

toks lūžis gali įvykti pernelyg vėlai ir kad pasaulinio ekonominio nuosmukio išvengti įmanoma tik sukūrus galingą DI.

Atrodė, kad sunku būtų rasti tinkamesnį „DeepMind“ rėmėją nei Thielis, mat jis turėjo ne tik gausių finansinių išteklių, bet ir entuziazmo ambicingiems projektams. „Mums reikėjo žmogaus, kuris būtų pakankamai beprotis, kad finansuotų bendrovę, užsimojusią sukurti BDI, – prisimena Leggas. – Norėjome rasti investuotoją, kuris galėtų be vargo tam skirti kelis milijonus ir mėgtų itin drąsius projektus. Be to, buvo svarbu, kad jis nebijotų eiti prieš srovę, nes kiekvienas profesorius, su kuriuo [Hassabisas] kalbėjosi, jam kartojo: „Nė negalvokite apie tai.“

Thielis iš tiesų mėgo „plaukti“ prieš srovę – jis dažnai nesutardavo net su Silicio slėnio elitu, garsėjusiu netradiciniu požiūriu. Kai dauguma regiono bendruomenės narių balsavo už liberalias idėjas palaikančius kandidatus, jis rėmė dešiniųjų atstovus ir tapo vienu didžiausių Donaldo Trumpo šalininkų. Skirtingai nei daugelis verslininkų, tikėjusių, jog inovacijas skatina konkurencija, knygoje „Nuo nulio iki vieneto“ (*Zero to One*) Thielis teigė, kad geriausias sąlygas technologinėms naujovėms atsirasti užtikrina monopolinė aplinka. Jis niekino tradicinius kelius į sėkmę ir ragino sumanius, verslius jaunuolius mesti studijas, siūlydamas jiems prisijungti prie jo įkurtos programos „Thiel Fellowship“. Jo polinkis investuoti į įvairius ekscentriškus projektus, pavyzdžiui, susijusius su siekiu prailginti žmogaus gyvenimo trukmę ar paspartinti technologinį singuliarimą, rodė, kad jis atitiko „beprotiško“ investuotojo kriterijus – būtent tokio žmogaus ir ieškojo „DeepMind“ įkūrėjai.

Trijulė nutarė savo idėją Thieliui pristatyti „Singuliarumo konferencijoje“. Jie tikėjosi, kad renginio rėmėjas sėdės pirmoje eilėje. Leggas paprašė organizatorių pusę jam numatyto pranešimo laiko skirti Hassabisui, kad Thielis tiesiai iš buvusio šachmatų čempiono lūpų išgirstų apie BDI ir žmogaus smegenis kaip įkvėpimo šaltinį.

Tą dieną, turėjusią nulemti „DeepMind“ likimą, į konferenciją San Fransisko viešbutyje Hassabisas atvyko vilkėdamas tamsiai raudoną

megztinį ir juodas kelnes. Lipdamas į sceną jis visas virpėjo iš jaudulio. Nužvelgęs šimtus susirinkusiųjų pamatė, kad Thielio pirmoje eilėje nėra. Jo apskritai nebuvo salėje.

Kai trijulė jau manė praradusi svarbią galimybę, Leggas gavo išskirtinį kvietimą į vakarėlį Thielio viloje San Fransisko įlankos regione ir pasirūpino, kad drauge galėtų atvykti ir jo bendražygiai. Hassabisas tuo metu jau žinojo, kad Thielis mėgsta šachmatus – jaunystėje jis buvo vienas geriausių JAV šachmatininkų iki trylikos metų. Taigi jiedu turėjo šį tą bendro, kas galėjo pažadinti Thielio susidomėjimą. Per vakarėlį Hassabisui pavyko užmegzti pokalbį su šeimininku. Kaip jis ne kartą yra pasakojęs žiniasklaidai, šachmatų tema pokalbio metu iškilo tarsi visai netyčia.

„Manau, viena priežasčių, kodėl šachmatai neprarado populiarumo ištisus šimtmečius, yra ta, kad žirgas ir rikis yra tobulai subalansuoti, – tarė Hassabisas, kol buvo nešiojami užkandžiai. – Mano galva, tai sukuria tam tikrą asimetrišką kūrybinę įtampą.“

Thielis iškart sukluso. „Kodėl tau neatvykus rytoj ir nepadarius rimto pristatymo?“ – pasiūlė jis. Kelionė pasiteisino – Thielis investavo 1,4 mln. svarų sterlingų į „DeepMind“, tikėdamasis prisidėti prie bendrovės pastangų pasiekti technologinį singuliarumą.

Mėgindamas pritraukti daugiau lėšų savo bendrovei, Hassabisas, kaip verslininkas, atsidūrė sudėtingoje padėtyje. Pirmieji investuotojai jį rėmė ne tiek dėl pelno, kiek dėl, galima sakyti, moralinių įsitikinimų, susijusių su DI. Tai reiškė, kad jam teks ne tik siekti finansinės sėkmės, bet ir kurti DI taip, kad jis atitiktų tam tikras ideologines nuostatas.

Tuo metu vis labiau įsigalėjo nuomonė, jog DI reikia kurti itin atsargiai, kad jis neištrūktų iš žmonių kontrolės ir nesiektų pakenkti savo kūrėjams. Panašiomis nuostatomis vadovavosi ir kitas turtingas rėmėjas, susidomėjęs galimybe investuoti į „DeepMind“. Tačiau jo požiūris į DI gerokai skyrėsi nuo Thielio pozicijos. Hassabisas jį sutiko žiemą Oksforde vykusioje konferencijoje DI tema „Winter Intelligence“, į

kurią susirinko radikalių pažiūrų kompiuterių mokslo atstovai diskutuoti apie superprotingo DI valdymo iššūkius. Vos tik Hassabisas baigė kalbą, prie jo priėjo trumpaplaukis blondinas.

„Sveiki, – tarė jis šiaurietišku akcentu, ištiesdamas ranką. – Esu Jaanas, vienas iš „Skype“ įkūrėjų.“

Jaanas Tallinnas – estų programuotojas, sukūręs lygiaverčių mazgų technologiją (angl. *peer-to-peer technology*), kurios pagrindu veikė „Kazaa“ – viena pirmųjų failų dalijimosi platformų, veikusių XXI a. pirmojo dešimtmečio pradžioje ir leidusių vartotojams keistis muzika bei filmais. Tą pačią technologiją jis pritaikė kurdamas nemokamų skambučių platformą „Skype“, kurią vėliau bendrovė „eBay“ įsigijo už 2,5 mlrd. JAV dolerių. Šis sandoris Tallinnui atnešė reikšmingą pelną, kurio dalį jis nusprendė investuoti į įvairius startuolius. Tallinnas jau kurį laiką domėjosi galingo DI keliamomis grėsmėmis, todėl įdėmiai klausėsi Hassabiso kalbos.

Dirbtinio intelekto tema jį užkabino dar 2009-ųjų pavasarį, kai jis atrado interneto forumą „LessWrong“. Šios uždaros interneto bendruomenės nariai, tarp kurių buvo nemažai programinės įrangos inžinierių, nerimavo dėl DI keliamo egzistencinio pavojaus žmonijai. Jos įkūrėjas ir lyderis – barzdotas liberalių pažiūrų savamokslis Eliezeris Yudkowsky’is. Nors nebaigė mokyklos, jis sugebėjo savarankiškai išstudijuoti DI tyrimų bei filosofijos pagrindus. Jo esė sulaukdavo didelio bendruomenės narių dėmesio. Būtent apie jį Altmanas pirmiausia galvojo, užsimindamas apie įtampą, tvyrančią DI saugumu besirūpinančiose bendruomenėse. Yudkowsky’is manė, kad tikimybė, jog DI sunaikins žmoniją, yra daug didesnė, nei daugelis norėtų pripažinti.

Jis tvirtino, kad, pasiekęs tam tikrą gebėjimų lygį, DI galėtų strategiškai slėpti savo galias, kol žmonėms jau bus per vėlu jį kontroliuoti. Tuomet jis pajėgtų perimti finansų rinkų ar komunikacijos tinklų kontrolę, išjungti kritinę infrastruktūrą, pavyzdžiui, elektros

tinklus. Pasak Yudkowsky'io, DI kūrėjai dažnai nė nesuvokia, kad savo darbu vis labiau stumia pasaulį pražūties link.

Tallinną kai kurie jo teiginiai šiek tiek glumino. Be to, jis ką tik buvo perskaitęs garsaus fiziko ir matematiko Rogerio Penrose'o knygą „Proto šešėliai" (*Shadows of the Mind*), kurioje teigiama, kad žmogaus protas gali spręsti užduotis, kurių kompiuteris niekada nesugebės atlikti. Hassabiso ir kitų idėjos apie smegenų „mechanistinę" prigimtį bei jų panaudojimą kaip įkvėpimo šaltinį DI kurti Tallinnui atrodė neįtikinamos – juk žmogaus smegenys yra unikalios, o jų veikimo iš esmės neįmanoma atkartoti.

Vis dėlto Tallinno galvoje kirbėjo abejonė: o jei DI visgi galėtų atkartoti žmogaus protą? Ar tai reikštų, kad kuriame kažką potencialiai pavojingo? Nusprendęs pasigilinti į Yudkowsky'io pažiūras, „Skype" įkūrėjas sudarė jį dominančių klausimų sąrašą. Jis suprato, kad pesimistinių teorijų pagrįstumą patikrinti gali tik vienu būdu – susitikęs su pačiu „LessWrong" įkūrėju.

Laimei, Tallinnas kaip tik planavo skristi į San Fransiską dalyvauti susitikime, tad parašė Yudkowsky'iui el. laišką su pasiūlymu išgerti kavos. Amerikietis sutiko. Jiedu susitiko kavinėje Milbrėjaus mieste, vos kelios minutės kelio nuo San Fransisko tarptautinio oro uosto. Tallinnas nedelsdamas pradėjo uždavinėti rūpimus klausimus. Jei DI toks potencialiai pavojingas, kodėl jo negalima būtų kurti virtualiojoje mašinoje, taip atskiriant jį nuo kitų kompiuterių sistemų? Juk tokiu būdu neleistume jam prasiskverbti į fizinę infrastruktūrą, atjungti elektros tinklą ar perimti finansų rinkų kontrolę.

„Iš tikrųjų toks DI nebūtų visiškai virtualus", – atsakė Yudkowsky'is. Jis pridūrė, kad elektronai gali tekėti įvairiomis kryptimis, todėl galingas DI visada ras būdą paveikti aparatinės įrangos konfigūraciją.

Tallinno baimės pasitvirtino – vieną dieną DI galės pats susikurti infrastruktūrą ir kompiuterinę veikimo terpę. Visa tai galėtų vesti prie siaubingų tolimesnių įvykių scenarijų.

„Jis galėtų transformuoti ir geoinžineriniu būdu keisti planetą, galbūt net saulę“, – teigia Tallinnas. Kai mokslininkai tvirtino, jog DI – tik matematika ir jo nereikia bijoti, Tallinnas dažnai pateikdavo tigro pavyzdį: „Galima sakyti, kad tigras – tik biocheminių reakcijų rinkinys, tad nėra ko jo bijoti. Tačiau kartu tigras yra atomų ir ląstelių visuma, kuri, jei nėra tinkamai suvaldyta, gali padaryti didžiulę žalą.“ Panašiai ir su DI: nors iš esmės tai tik pažangūs matematiniai algoritmai ir kompiuterinis kodas, netinkamai sukurtas jis gali tapti itin pavojingas.

Po dvejų metų, klausydamasis Hassabiso kalbos Oksfordo konferencijoje, Tallinnas jau buvo įtikėjęs apokaliptiniais DI ateities scenarijais. Jis aktyviai gilinosi į Yudkowsky’io tekstus ir domėjosi DI suderinamumu – nauja mokslo sritimi, kurioje mokslininkai ir filosofai ieško būdų, kaip geriausiai suderinti DI sistemas su žmogaus tikslais.

„Aš buvau visiškai įtikėjęs DI suderinamumo svarba“, – prisimena Tallinnas. Jis jau tikėjo kai kuriomis ekstremaliausiomis Yudkowsky’io piešiamos DI ateities vizijomis.

Kalbėdamasis su Hassabisu, Tallinnas norėjo sužinoti, ar šis sutiktų aktyviau bendradarbiauti. „Gal kada galėtume pasikalbėti per „Skype“?“ – paklausė jis britų verslininko. Vėliau Hassabisas dar kartą pasikalbėjo su turtingu estų investuotoju. Galiausiai Tallinnas tapo vienu iš pirmųjų „DeepMind“ investuotojų kartu su Peteriu Thieliu. Jis ne tik siekė pelno, bet ir norėjo stebėti Hassabiso projekto eigą, kad įsitikintų, jog netyčia nebus sukurtas itin pavojingas DI. Tallinnas save laikė Yudkowsky’io idėjų skleidėju. Pasinaudodamas savo, kaip turtingo investuotojo, įvaizdžiu jis norėjo, kad žinia apie DI grėsmes pasiektų pažangiausius pasaulio DI kūrėjus.

„Eliezeris buvo savamokslis ir neturėjo daug įtakos už savo mažos bendruomenės ribų, – paaiškina Tallinnas. – Maniau, kad galėčiau pristatyti jo idėjas tiems, kurie nebuvo linkę juo tikėti, bet mielai sutiko išklausyti mane.“

Tapęs investuotoju, Tallinnas ragino „DeepMind“ komandą daugiau dėmesio skirti saugumui. Jis žinojo, kad Hassabisas dėl DI keliamų apokaliptinių grėsmių nerimavo mažiau nei jis pats. Dėl šios priežasties jis spaudė įmonę pasamdyti specialistų komandą, kuri ieškotų būdų užtikrinti, kad DI būtų suderintas su žmogiškosiomis vertybėmis ir jo kūrimas nepakryptų pavojinga linkme.

Netrukus „DeepMind“ sulaukė dar vieno itin turtingo investuotojo dėmesio – jis taip pat norėjo, kad įmonė veiklą vykdytų apdairiai. Tuo metu Silicio slėnyje ėmė sklisti gandai apie Peterio Thielio įsitraukimą į perspektyvų, bet paslaptingą Londono startuolį, siekiantį sukurti BDI. Apie jį sužinojo ir kiti regiono technologijų milijardieriai, tarp jų – Elonas Muskas. Hassabisas su Musku pirmą kartą susitiko 2012-aisiais, praėjus dvejiems metams po „DeepMind“ įkūrimo, Kalifornijoje vykusioje uždaroje Thielio organizuotoje konferencijoje.

„Mes iškart radome bendrą kalbą“, – pasakoja Hassabisas. Britų verslininkas suprato, kad tai puiki galimybė pritraukti daugiau lėšų „DeepMind“ tyrimų plėtrai. Be to, jis labai norėjo pamatyti Musko raketų gamyklą. Tuo metu Muskas jau garsėjo kaip drąsus verslo magnatas, kurio valdoma „SpaceX“ siekė išsiųsti žmones į Marsą. Hassabisas susitarė susitikti su juo „SpaceX“ būstinėje Los Andžele.

Susėdę vienas priešais kitą valgykloje, apsupti raketų dalių, jie įsivėlė į diskusiją, kuris iš projektų – kitos planetos kolonizavimas ar pažangaus DI sukūrimas – istoriniu požiūriu bus reikšmingesnis.

„Žmonėms reikės galimybės pabėgti į Marsą, jei DI taps nekontroliuojamas“, – sakė Muskas, kaip cituojama žurnale „Vanity Fair“.

„Manau, DI sugebės pasiekti žmones ir Marse“, – su šypsena atsakė Hassabisas. Muskui tai nepasirodė juokinga. Jei Tallinną paveikė Yudkowsky’io tekstai, tai Musko pažiūroms įtakos turėjo kitas žmogus – Oksfordo universiteto filosofas Nickas Bostromas.

Bostromas buvo parašęs knygą „Superintelektas“ (*Superintelligence*), sukėlusią didelį atgarsį DI ir pažangių technologijų kūrėjų

bendruomenėje. Knygoje Bostromas įspėjo, kad bendrojo arba itin galingo DI sukūrimas gali baigtis katastrofa žmonijai, – nebūtinai todėl, kad DI turėtų piktų kėslų ar trokštų valdžios, bet todėl, kad jis aklai vykdytų jam pavestą užduotį. Pavyzdžiui, gavęs nurodymą pagaminti kuo daugiau sąvaržėlių, jis galėtų panaudoti visus Žemės išteklius, net žmones, šiam tikslui pasiekti. Taip DI bendruomenėje prigijo posakis, jog svarbiausia „nebūti paverstiems sąvaržėlėmis“.

Galiausiai Muskas taip pat nusprendė investuoti į „DeepMind“. Nors Hassabisas pagaliau įgijo tam tikrą finansinį užtikrintumą, lėšų vis dar trūko. Jo projektas buvo itin eksperimentinis ir ambicingas, todėl net turtingiausi pasaulio žmonės nesiryžo į jį investuoti didesnių sumų. Be to, investicijas neišvengiamai lydėjo tam tikros ideologinės sąlygos – tiek Tallinnas, tiek Muskas į „DeepMind“ veiklą žvelgė su neįprastu įtarumu ir atsargumu. Jie, žinoma, troško, kad bendrovė būtų finansiškai sėkminga, tačiau kartu nerimavo, kad pernelyg greita ar neatsargi plėtra gali kelti grėsmę žmonijai. Tai Hassabisą įstūmė į sudėtingą padėtį: jis buvo dėkingas už jų finansinę paramą, tačiau netikėjo pesimistiškais scenarijais, apie kuriuos kalbėjo Tallinnas ir Muskas.

Vis dėlto toks finansinis saugumas truko neilgai. Hassabisui ir Suleymanui sunkiai sekėsi uždirbti pakankamai, kad galėtų mokėti atlyginimus geriausiems pasaulio DI specialistams. Iš tiesų kai kurios jų pajamų generavimo idėjos buvo gan padrikos. Jie bandė sukurti interneto svetainę, paremtą giluminiu mokymusi, – tai buvo „DeepMind“ pradinė specializacijos sritis. Svetainė turėjo teikti mados patarimus ir aprangos rekomendacijas vartotojams. Vėliau Hassabisas paprašė kai kurių buvusių kolegų iš „Elixir Studios“, perėjusių dirbti į „DeepMind“, sukurti vaizdo žaidimą. Pasak vieno buvusio darbuotojo, inžinieriai sukūrė kosminių nuotykių kupiną žaidimą, kuriame astronautų komanda varžėsi, kas greičiau pasieks Mėnulį. Kaip tik tada, kai žaidimas jau buvo beveik baigtas, ruošiantis jį išleisti kaip „iPhone“ programėlę,

Hassabisas netikėtai sulaukė pasiūlymo iš „Facebook“. Tai atvėrė galimybę gauti reikiamų finansinių išteklių BDI kurti.

Markas Zuckerbergas tuo metu aktyviai plėtė savo verslo imperiją per įsigijimus. Maždaug metais anksčiau už 1 mlrd. JAV dolerių jis įsigijo „Instagram“ – šis žingsnis vėliau pripažintas genialiu socialinių tinklų konsolidacijos sprendimu. Jis taip pat ruošėsi sumokėti įspūdingą 19 mlrd. JAV dolerių sumą „WhatsApp“ įkūrėjams. Verslo magnatas buvo pasirengęs investuoti bet kokią sumą, kad „Facebook“ imperija toliau augtų, o DI turėjo tapti svarbia šios strategijos dalimi. Kadangi apie 98 proc. pajamų „Facebook“ gaudavo iš reklamos, Zuckerbergas, norėdamas parduoti dar daugiau reklamos ir toliau plėstis, turėjo užtikrinti, kad vartotojai jo socialinėse platformose naršytų kuo ilgiau. Šiuo atveju į pagalbą galėjo ateiti talentingi „DeepMind“ DI specialistai. Išmanesnės rekomendacijų sistemos, analizuojančios vartotojų asmeninius duomenis, ir pažangesni algoritmai galėtų tiksliau atrinkti nuotraukas, tekstus ir vaizdo įrašus, kurie labiau įtrauktų vartotojus į naršymą „Facebook“ ir „Instagram“ platformose.

Kaip teigė asmuo, susipažinęs su pasiūlymu, Zuckerbergas norėjo nupirkti „DeepMind“ iš Hassabiso už 800 mln. JAV dolerių, neįskaitant premijų, kurios paprastai mokamos startuolių įkūrėjams, jei jie lieka dirbti įsigytoje bendrovėje dar ketverius ar penkerius metus. Tai buvo itin dosnus pasiūlymas, apie kurį Hassabisas anksčiau nė nebūtų drįsęs svajoti. Tačiau dabar jis atsidūrė kryžkelėje. Iki tol „DeepMind“ rėmė žmonės, norėję, kad DI būtų kuriamas kuo apdairiau. Dabar pasiūlymas atėjo iš asmens, siekiančio spartesnės plėtros. Be to, „Facebook“ vadovavosi šūkiu „Veik greitai ir griauk barjerus“.

Hassabisas ir Suleymanas ėmėsi ieškoti sprendimų. Bendrasis dirbtinis intelektas galėjo tapti galingesnis, nei Zuckerbergas tuo metu įsivaizdavo, todėl jie jautė, kad būtina įdiegti kontrolės mechanizmą, kuris neleistų „DeepMind“ įsigijusiai korporacijai DI vystymą pakreipti pavojinga linkme. Juk jie negalėjo paprasčiausiai prašyti „Facebook“

atstovų pasižadėti tinkamai naudoti BDI. Remdamasis ankstesne patirtimi dirbant su ne pelno organizacijomis, Suleymanas pasiūlė Hassabisui ir Leggui įsteigti valdymo organą, kuris stebėtų „Facebook“ veiklą ir užtikrintų atsakingą „DeepMind“ technologijos naudojimą.

Viešosios įmonės dažnai turi direktorių valdybą, kuri atstovauja akcininkų interesams ir kas ketvirtį vertina įmonės veiklą, siekdama užtikrinti akcijų vertės augimą. Suleymanas pasiūlė, kad „DeepMind“ galėtų būti įsteigta kitokio tipo valdyba, kuri, užuot susitelkusi į pelną, užtikrintų, kad DI būtų kuriamas saugiai ir etiškai. Iš pradžių Hassabisas ir Leggas tam nepritarė, tačiau Suleymanui galiausiai pavyko juos įtikinti.

Hassabisas vėl susisiekė su Zuckerbergu ir pranešė, kad „DeepMind“ bus parduodama tik tuo atveju, jei bus įsteigta etikos ir saugumo valdyba, turinti atskirą teisę valdyti bet kokį superintelektą, kurį „DeepMind“ sukurs ateityje. Zuckerbergas su šia sąlyga nesutiko. Jis norėjo plėsti „Facebook“ reklamos verslą ir „sujungti pasaulį“ socialiniais tinklais, o ne valdyti atskirą DI bendrovę su daugybe etikos protokolų ir ambicinga misija. Taigi derybos žlugo.

Viešai Hassabisas tvirtino, kad „DeepMind“ liks nepriklausoma dar 20 metų, tačiau iš tiesų jis buvo pavargęs nuo nuolatinio lėšų rinkimo ir nusivylęs, kad tyrimams gali skirti tik nedidelę savo laiko dalį. Atmetęs dosnų Zuckerbergo pasiūlymą, jis nebegalėjo ignoruoti fakto, jog pardavęs įmonę Silicio slėnio gigantui galėtų gauti solidžią sumą, juolab kad didžiosios technologijų bendrovės ėmė aktyviai domėtis DI. Tuo metu Silicio slėnio bendrovių vadovai, tarp jų ir keli milijardieriai, nuolat skambindavo „DeepMind“ tyrėjams, bandydami juos prisivilioti. Daugelis „DeepMind“ darbuotojų puikiai išmanė giluminį mokymąsi – sritį, kuri ilgą laiką buvo primiršta.

Esminis lūžis įvyko 2012 metais, kai Stanfordo DI profesorė Fei-Fei Li inicijavo kasmetinį akademikų konkursą „ImageNet“. Jame dalyvavę mokslininkai pateikė DI modelius, turėjusius vizualiai atpažinti vaizdus,

pavyzdžiui, kačių, baldų, automobilių ar kitų objektų nuotraukas. Tais metais Geoffrey'io Hintono vadovaujama tyrėjų komanda pritaikė giluminio mokymosi technologiją ir sukūrė modelį, kuris gerokai pranoko ankstesnius, tuo priblokšdamas visą DI bendruomenę. Staiga visi ėmė ieškoti ekspertų, išmanančių giluminio mokymosi DI teoriją, paremtą smegenų gebėjimu atpažinti dėsningumus.

„Tai buvo nišinė sritis, kurioje dirbo vos keliasdešimt specialistų, – prisimena Leggas. – Gana daug jų prisijungė prie mūsų komandos." Hassabisas jiems mokėjo apie 100 tūkst. JAV dolerių per metus, tačiau tokios technologijų milžinės kaip „Google" ir „Facebook" siūlė kelis kartus daugiau. „Mūsų tyrėjai sulaukdavo netikėtų skambučių iš žinomų asmenų, kurie siūlė jiems trigubai didesnį atlyginimą", – prisimena Leggas. Anot buvusio „DeepMind" darbuotojo, vienas tokių buvo ir Zuckerbergas. „Turėjome parduoti bendrovę, nes kitaip mus būtų suplėšę į gabalus", – teigia buvęs komandos narys. Kadangi Hassabisas norėjo būti pirmas, sukūręs BDI, jis baiminosi, kad didesnių išteklių turinčios technologijų bendrovės gali jį aplenkti.

Staiga „DeepMind" sulaukė dar vieno pasiūlymo – šįkart iš investuotojo Elono Musko. Pasak asmens, susipažinusio su sandoriu, milijardierius už jį siūlė atsiskaityti elektromobilių gamintojos „Tesla", kuriai vadovavo pastaruosius penkerius metus, akcijomis. Muskas nebuvo labai aktyvus investuotojas ir tik retkarčiais kontaktavo su Hassabisu. Nors jis vis labiau nerimavo dėl galimų DI grėsmių, jam taip pat buvo svarbūs komerciniai tikslai. Jis siekė, kad „Tesla" automobiliuose pirmąkart istorijoje būtų sėkmingai pritaikyta autonominio vairavimo technologija, o šiam tikslui pasiekti jam reikėjo geriausių DI ekspertų. Įsigijęs „DeepMind", jis galėjo gauti elitinę tokių specialistų komandą.

Tačiau „DeepMind" įkūrėjai ir šįkart buvo atsargūs. Atsiskaitymas „Tesla" akcijomis neatrodė patrauklus variantas, be to, jie nepasikliovė Musku tiek, kad drąsiai patikėtų jam valdyti BDI. Nors Muskas jau buvo

pradėjęs garsėti kaip novatoriškai mąstantis verslo magnatas, technologijų pasaulyje jis buvo žinomas kaip kaprizingas vadovas, be aiškios priežasties atleidinėjantis darbuotojus ir netgi iš savo vadovaujamos bendrovės išmetęs vieną iš jos bendrakūrėjų.

Nors „DeepMind“ įkūrėjai vertino šio milijardieriaus investicijas ir ryšius, jo nenuspėjamas elgesys kėlė nepasitikėjimą. Galiausiai jie atmetė Musko pasiūlymą, nesuprasdami, kad įžeidaus charakterio milijardierius tiesiog nepakenčia žodžio „ne“ ir kaip toks sprendimas galėtų jiems atsiliepti ateityje. Tačiau netrukus Hassabisas sulaukė dar vieno el. laiško – šįkart iš „Google“.

5 SKYRIUS

Dėl utopijos, dėl pinigų

Vieną dieną Hassabisas savo el. pašto dėžutėje rado laišką iš pagrindinės „Google“ būstinės, įsikūrusios saulėtajame Mauntin Vju, Kalifornijoje, daugiau kaip už aštuonių tūkstančių kilometrų nuo Londono. Tai buvo „Google“ generalinio direktoriaus Larry’io Page’o kvietimas susitikti. Page’as kartu su kitu Stanfordo universiteto doktorantu Sergey’umi Brinu bendrovę „Google“ įkūrė 1998 metais. Jiedu siekė suteikti žmonėms patogesnį informacijos paieškos internete būdą.

Taip gimė „PageRank“ – algoritmas, leidžiantis reitinguoti tinklalapius pagal jų aktualumą ir tarpusavio ryšius. Pradėję brandinti savo idėją draugo garaže Menlo Parke, Kalifornijoje, bendražygiai galiausiai sukūrė vieną didžiausių pasaulyje technologijų bendrovių.

Pažvelgus į tai, kaip šiandien „Google“ uždirba pinigus, matyti, kad daugiausia jų atneša reklama, o ne aukštosios technologijos ar inovacijos. Bendrovė tapo milžiniška reklamos paslaugų teikėja, panašia į „Facebook“. Didžiąją dalį pajamų ir pelno ji gauna rinkdama ir analizuodama žmonių asmens duomenis ir rodydama jiems pritaikytas reklamas „Google“ paieškos sistemoje, vaizdo įrašų platformoje „YouTube“, el. pašto sistemoje „Gmail“ ir milijonuose kitų svetainių bei mobiliųjų programėlių, priklausančių „Google Display Network“ reklamos tinklui.

Tokiam žmogui kaip Hassabisas, kuris siekė kurti DI, galintį padėti pasauliui, tai kėlė nerimą. Drauge jis puikiai suprato, kad jei atsisakytų „Google“ pasiūlymo, technologijų milžinė galėtų pervilioti jo darbuotojus, o galbūt netgi be jo žinios sukurti BDI. „Google“ jau buvo

sutelkusi keliais šimtais daugiau DI srityje dirbančių inžinierių, ir Hassabisas suvokė, kad negali atsisakyti pasiūlymo susitikti.

Kai pagaliau susitiko su Page'u – introvertišku matematiku tamsiais vešliais antakiais, mėgusiu rengtis laisvalaikio marškinėliais ir šortais, – Hassabisas iškart suprato radęs bendramintį. Paaiškėjo, kad vienas iš „Google" bendrakūrėjų taip pat ilgą laiką puoselėjo svajonę sukurti galingą DI. „Jis man sakė, kad savo vizijose „Google" visada regėjo kaip DI bendrovę – net pačioje pradžioje, 1998-aisiais, brandindamas verslo idėją draugo garaže", – prisimena Hassabisas.

Ši sritis Page'ui buvo artima ir dėl asmeninių priežasčių – jo tėvas iki pat mirties 1996 metais profesoriavo universitete, o jo akademinių interesų laukas buvo susijęs su DI ir kompiuterių mokslu. Taigi galima sakyti, kad Page'as priklausė antrajai DI specialistų kartai. Idėja sukurti BDI jam visiškai neatrodė beprotiška – priešingai, jis žavėjosi Hassabiso ryžtu siekti šio tikslo. Be to, Page'as jau buvo uždegęs žalią šviesą kitam vidiniam „Google" projektui, kurio tikslas – sukurti žmogaus gebėjimams prilygstantį DI. Galiausiai dėl šio projekto Hassabisui teko įsitraukti į aktyvią konkurencinę kovą.

Projektas, apie kurį Hassabisas tuo metu dar nežinojo, vadinosi „Google Brain". Idėją jam pasiūlė Andrew Ng'as – malonaus būdo Stanfordo universiteto profesorius, siekęs „Google" kurti pažangesnes DI sistemas. 2011-aisiais, dar prieš „Google" susisiekiant su „DeepMind" vadovu, minėtas profesorius Page'ui išsiuntė keturių puslapių dokumentą pavadinimu „Neuromokslu grįstas giluminis mokymasis" (angl. *Neuroscience-Informed Deep Learning*). Ng'as tikėjosi, kad „Google" generalinis direktorius pritars jo sumanymui kurti „bendrosios paskirties" DI sistemas, – su tokio pobūdžio projektu tuo metu Anglijoje jau dirbo Hassabisas.

Paaiškėjo, kad tiek Ng'as, tiek Hassabisas siekė sukurti BDI panašiais metodais – semdamiesi įkvėpimo iš neuromokslo. Page'ui išsiųstame pasiūlyme Stanfordo profesorius nurodė ketinantis kurti „sistemas,

kurios su laiku vis tiksliau atkartos tam tikrų nedidelių žinduolių smegenų dalių veikimą“.

Net tokiam žmogui kaip Andrew Ng’as, kuris garsėjo kaip iškilus DI srities mokslininkas ir dirbo viename prestižiškiausių pasaulio universitetų, imtis BDI kūrimo tuo metu atrodė kontroversiškas žingsnis. „Bičiuliai man sakė, kad šis sumanymas yra ganėtinai keistas. Jie tvirtino, kad tai pakenks mano karjerai“, – prisimena Ng’as.

Tam tikra prasme jie buvo teisūs. Žvelgiant iš mokslinės perspektyvos, Ng’o ir Hassabiso išskirtinis domėjimasis žmogaus smegenimis tam tikrais aspektais buvo problemiškas. Teoriškai, regis, buvo visiškai racionalu kuriant DI remtis smegenimis, tačiau biologijos principų kopijavimas ne visada pasiteisina. Pakanka prisiminti pirmuosius bandymus sukurti skraidančias mašinas, kai išradėjai gremėzdiškiems aparatams nesėkmingai bandė pritaikyti mechanikos principus, leidžiančius paukščiams pakilti į orą. Reikia paminėti, kad kitų kompiuterių mokslo atstovų bandymai pernelyg tiksliai kopijuoti smegenis baigėsi nesėkme. 2013 metais neuromokslininkas Henry’is Markhamas TED kalboje teigė sugalvojęs būdą superkompiuteriais atkartoti visas žmogaus smegenų funkcijas ir tikėjosi tai pasiekti per dešimtmetį. Praėjus dešimčiai metų nuo šio pažado, jo projektas „Human Brain Project“, kainavęs daugiau nei 1 mlrd. JAV dolerių, taip ir liko neįgyvendintas.

Bėgant laikui Ng’as, Hassabisas ir kiti DI mokslininkai suprato, kaip sunku atkartoti smegenis, kai mūsų supratimas apie jas tebėra ribotas – vis dar neturime pakankamai žinių apie neuronų funkcijas ir įvairių smegenų sričių veiklą. Nors buvo žinoma, kad mūsų smegenyse nuolat aktyvuojasi apie devyniasdešimt milijardų neuronų, liko neaišku, kaip ši informacija apdorojama.

„Žvelgiant iš šiandienos perspektyvos, tampa akivaizdu, kad bandymas pernelyg tiksliai atkartoti biologinius mechanizmus buvo

klaida“, – sako Ng’as. Vis tik jo sprendimas plėsti neuroninius tinklus pasirodė esantis teisingas.

Neuroninis tinklas – tai programinė sistema, kuri mokosi nuolat analizuodama didelius kiekius duomenų. Išmokytas tinklas geba atpažinti veidus, numatyti šachmatų partijos ėjimus arba pasiūlyti filmą „Netflix“ platformoje. Toks neuroninis tinklas, dar vadinamas modeliu, dažnai yra sudarytas iš daugybės sluoksnių ir mazgų, kurie informaciją apdoroja panašiai kaip mūsų smegenų neuronai. Kuo daugiau tinklas mokomas, tuo geriau jo mazgai geba prognozuoti ar atpažinti įvairius dalykus.

Ng’as suprato, kad, turėdami daugiau mazgų, sluoksnių ir mokymui skirtų duomenų, šie modeliai gali tapti efektyvesni. Po kelerių metų panašią išvadą dėl „masto didinimo“ padarė ir „OpenAI“. Eksperimentuodamas Stanforde, Ng’as pastebėjo, kad jo giluminio mokymosi modeliai veikia kur kas geriau, kai yra didesni. Šie rezultatai jį taip sužavėjo, kad jis nusprendė Page’ui nusiųsti keturių puslapių pasiūlymą, kuriame tvirtino galįs sukurti „didelio masto smegenų simuliacijas“ – tai turėjo būti svarbus žingsnis link „žmogaus gebėjimams artimo DI“ sukūrimo.

Page’ui ši idėja labai patiko, todėl jis pritarė Ng’o sumanymui ir paskyrė jį iki tol pažangiausio „Google“ DI tyrimų projekto vadovu. Tačiau per kelerius projekto „Google Brain“ gyvavimo metus taip ir nebuvo pasistūmėta išsikelto tikslo link – sukurti BDI. Vietoj to projektas daugiausia prisidėjo prie „Google“ tikslinės reklamos gerinimo – bendrovė galėjo geriau prognozuoti, kokį turinį žmonės pasirinks, ir taip rodyti itin tiksliai atrinktas reklamas, o tai savo ruožtu padidino bendrovės pajamas. Ng’as pripažįsta, kad siųsdamas Page’ui pasiūlymą siekė visai kito tikslo. „Tai toli gražu ne įdomiausias projektas, su kuriuo man yra tekę dirbti“, – sako jis.

Tikrasis Ng’o, kaip DI mokslininko, siekis buvo išvaduoti žmoniją nuo protinio darbo naštos, panašiai kaip pramonės revoliucija išlaisvino

žmones nuo sunkaus fizinio darbo. Jis tikėjo, kad galingesnės DI sistemos palengvintų darbą kvalifikuotiems darbuotojams, kad „visi galėtume užsiimti intelektualiai įdomesne, aukštesnio lygio veikla“.

Siekdamas šio tikslo, Ng’as ėjo kitu keliu nei Hassabisas. Skirtingai nei britų verslininkas, siekęs išlikti kuo labiau nepriklausomas nuo „Google“, profesorius Ng’as mielai tapo šios reklamos milžinės sraigteliu. Tokiu būdu, pats to nesuvokdamas, jis padarė Hassabisui didelę paslaugą: atlikdamas mokslinius tyrimus „Google“, Ng’as prisidėjo prie bendrovės reklamos verslo plėtros, todėl ši našta vėliau nenugulė ant „DeepMind“ pečių.

2013-ųjų pabaigoje „Google“ pirmą kartą susisiekė su „DeepMind“ dėl galimo bendrovės įsigijimo. Tuo metu Ng’o vadovaujami tyrėjai jau buvo įsitraukę į pažangių DI modelių, skirtų „Google“ reklamos įrankiams, kūrimą. Jie vis labiau tolo nuo kilnesnio tikslo – sukurti itin galingą DI, galintį išlaisvinti žmoniją nuo monotoniško darbo. Skrisdamas į Londoną derėtis dėl „DeepMind“ įsigijimo, Page’as žinojo galintis dalį „Google“ lėšų skirti drąsesniam projektui.

„New York Times“ reporteris Cade’as Metzas knygoje „Genijų kūrėjai“ (*Genius Makers*) pasakoja, kad milijardierių savo Londono biure pasitikę „DeepMind“ įkūrėjai jam surengė pristatymą apie lig tol bendrovės atliktus tyrimus. Hassabisas sakė, kad jo komanda sukūrė naują metodą, vadinamą skatinamuoju mokymusi (angl. *reinforcement learning*), siekdama išmokyti DI sistemą žaisti retro stiliaus „Atari“ žaidimą „Breakout“. Žaidimo esmė tokia: žaidėjas raketę primenančiu slankiuoju elementu stengiasi atmušti kamuoliuką taip, kad jis pataikytų į plytų sieną. Per maždaug dvi valandas DI sistema išmoko tiksliai nusitaikyti taip, kad vienu smūgiu pramuštų kelis plytų sluoksnius. Page’as buvo sužavėtas.

Skatinamasis mokymasis iš esmės nesiskiria nuo šuns dresavimo: paklusęs komandai „sėdėti“, šuo paskatinamas skanumynu. Lygiai taip pat, mokant DI, galima paskatinti neuroninį tinklą, pavyzdžiui, skaitiniu signalu +1, kad parodytume, jog tam tikras rezultatas yra pageidautinas.

Per nuolatinius bandymus ir klaidas, žaisdama tą patį žaidimą šimtus kartų, sistema išmoko, kas veikia, o kas ne. Tai buvo elegantiškai paprasta idėja, įvilkta į itin sudėtingą kompiuterinį kodą.

Tada Leggas pristatė Page'ui platesnę viziją – kaip šiuos metodus būtų galima pritaikyti realiame pasaulyje. Remdamiesi tuo pačiu principu, leidusiu DI sistemai įvaldyti vaizdo žaidimą, jie galėtų išmokyti robotą orientuotis namų aplinkoje ar suteikti autonominiam agentui gebėjimą įvaldyti anglų kalbą. Būtent tokiose srityse „DeepMind“ atradimai ir BDI galiausiai duotų daugiausia naudos. Ši vizija greitai įtikino Page'ą ir jo komandą.

Derėdamasis su Hassabisu ir kitais bendrovės steigėjais, Page'as žinojo, kad jie jau buvo atmetę dosnų „Facebook“ pasiūlymą. Netrukus paaiškėjo tokio atmetimo priežastis. Hassabisas kėlė dvi esmines sąlygas. Pirma, jis ir kiti „DeepMind“ steigėjai nenorėjo, kad jų technologijos ateityje būtų naudojamos kariniais tikslais, – nesvarbu, ar tai būtų autonominių dronų ir ginklų valdymas, ar pagalba kariams mūšio lauke. Tai buvo etinės ribos, kurių, jų įsitikinimu, „Google“ niekada neturėtų peržengti.

Antra, jie pareikalavo, kad „Google“ vadovai pasirašytų vadinamąją etikos ir saugumo sutartį. Dokumente, kurį parengė Londono teisininkai, buvo numatyta, kad bet kokia ateityje „DeepMind“ sukurta BDI technologija turėtų būti prižiūrima etikos tarybos. Tuo metu dar nebuvo nuspręsta, kas sudarytų tokią tarybą. Hassabisas su Suleymanu ketino šį organą suburti patys ir suteikti jam visas teises kontroliuoti bet kokį galingą DI, kurį jie galėtų sukurti ateityje.

„Jei mums pavyktų sukurti BDI, reikėtų užtikrinti tinkamą jo valdymą, – dėstė Hassabisas, kalbėdamas apie tarybą, kurią kartu su kitais „DeepMind“ steigėjais ketino įkurti. – Kadangi kalbame apie bendrosios paskirties technologiją, galinčią tapti viena galingiausių kada nors sukurtų sistemų, norėjome užtikrinti, kad ją valdytų atsakingi žmonės.“

Po kelis mėnesius trukusių įtemptų derybų „Google“ galiausiai sutiko su sąlyga, kuri anksčiau atbaidė „Facebook“. Page’as turėjo unikalią galimybę tapti pirmosios įmonės, siekiančios sukurti BDI, savininku. „Google“ vadovas suprato, kad jei etikos taryba įgis teisę kontroliuoti šią technologiją, bendrovei bus sunkiau iš jos pasipelnyti. Tačiau nugalėjo jo idealistinis požiūris – jis sutiko įgyvendinti „DeepMind“ reikalavimą ir įsteigti tarybą.

Tinkamą BDI valdymą užtikrinti reikėjo ne tik todėl, kad technologijos vystymą perėmusi korporacija galėjo šį procesą pakreipti nenuspėjama linkme. Tuo metu pati technologija atsidūrė kelių besiformuojančių ideologijų, galinčių skirtingai paveikti jos vystymąsi, susikirtimo taške. Hassabisas tokį požiūrių susidūrimą jau buvo matęs investuotojų gretose: Peteris Thielis siekė spartesnės DI plėtros, o Jaanas Tallinnas baiminosi, kad jaunasis britų verslininkas gali sukelti apokalipsę.

Neregėtų galimybių atverianti technologija tapo idėjinių ginčų epicentru. Po kelerių metų šios ideologinės jėgos susidurs su novatorių vizijomis ir didžiųjų korporacijų siekiais kontroliuoti BDI, o tai savo ruožtu kels netikėtų iššūkių šios technologijos raidai. Pavyzdžiui, būtent dėl tokių požiūrių susikirtimų Samo Altmano, kaip „OpenAI“ vadovo, ateitis buvo pakibusi ant plauko. Paradoksalu tai, kad apokaliptinės DI galios prognozės darė šią technologiją dar patrauklesnę verslui ir skatino įmonių komercinį susidomėjimą. Viena koja įžengę į verslo pasaulį, DI kūrėjai ėmė laikytis vis radikalesnių nuostatų: vieni siekė kuo sparčiau vystyti DI, tikėdamiesi sukurti utopiją, kiti kurstė baimes dėl galimo pasaulio pabaigos scenarijaus.

Kaip strategiškai mąstantis žmogus, nelinkęs visko statyti ant vienos kortos, Hassabisas iš esmės laikėsi atokiai nuo šių idėjinių kovų. Iš dalies tokią laikyseną lėmė unikali jo vizija: kaip teigia jį pažinoję žmonės, jis tikėjosi, kad BDI padės padaryti reikšmingų atradimų, galbūt netgi įminti dieviškąsias paslaptis. Tuo metu Suleymanas labiau nerimavo dėl

visuomeninių problemų, kurias artimiausiu metu galėjo sukelti DI. Shane'as Leggas iš jų trijulės labiausiai tapatinosi su kraštutinėmis ideologijomis, susijusiomis su BDI kūrimu. Pasak buvusių Leggo kolegų, jį ypač traukė transhumanizmas. Šios ne vieną dešimtmetį gyvuojančios, prieštaringas ištakas turinčios ideologijos istorija padeda paaiškinti, kodėl DI kūrėjai kartais užmerkia akis prieš technologijų keliamas dabarties problemas.

Pagrindinė transhumanizmo prielaida ta, kad žmonių rūšis vis dar nėra pasiekusi aukščiausio galimo išsivystymo lygio. Transhumanistai tiki, kad tam tikri moksliniai atradimai ir technologijos vieną dieną padės mums peržengti fizines bei protines ribas ir išsivystyti į naują, protingesnę rūšį. Jos atstovai pasižymėtų aukštesniu intelektu, kūrybiškumu ir ilgesne gyvenimo trukme. Galbūt žmonėms netgi pavyktų savo sąmonę susieti su kompiuteriu ir leistis į galaktikos tyrinėjimus.

Šios ideologijos ištakos siejamos su Britų eugenikos draugija. XX a. penktąjį–septintąjį dešimtmečiais jos veikloje aktyviai dalyvavo ir jai vadovavo evoliucijos biologas Julianas Huxley'is. Eugenikos judėjimo šalininkai teigė, kad žmonės savo biologines savybes turėtų tobulinti pasitelkdami atrankinį dauginimąsi. Tokios idėjos buvo įsigalėjusios Didžiosios Britanijos universitetuose, inteligentijos bei aukštuomenės sluoksniuose. Pats Huxley'is, kuris buvo kilęs iš aukštuomenės (jo brolis Aldous parašė knygą „Puikus naujas pasaulis" (*Brave New World*), tikėjo, kad visuomenės elitas genetiniu požiūriu yra pranašesnis. Žemesnio socialinio sluoksnio žmonės, jo nuomone, turėjo būti „išrauti kaip piktžolės" ir priverstinai sterilizuoti. „[Jie] dauginasi per greitai", – rašė Huxley'is.

Naciams ėmus žavėtis eugenikos idėjomis, Huxley'is nusprendė šį judėjimą pervadinti. Vienoje savo esė jis pavartojo naują terminą – „transhumanizmas", teigdamas, kad greta tinkamos dauginimosi strategijos žmonija galėtų „peržengti savo ribas" pasitelkusi mokslą ir

technologijas. Transhumanizmo judėjimas pagreitį įgijo XX a. devintąjį ir dešimtąjį dešimtmečiais, kai populiarėjančios DI srities atstovai ėmė žarstytis pažadais, kad vieną dieną jiems pavyks pagerinti žmogaus protą, sujungus jį su mąstančiomis mašinomis.

Šią idėją išgryninti padėjo technologinio singuliarumo koncepcija. Technologinis singuliarumas – hipotetinis ateities momentas, kai technologijos bus taip pažengusios, kad žmonija patirs dramatiškų ir negrįžtamų pokyčių, o žmogaus sąmonė susijungs su mašinomis ir taip patobulės. Šia idėja susižavėjo ne tik Shane'as Leggas, kuris apie ją sužinojo dar jaunystėje perskaitęs knygą šia tema, bet ir turtingasis „DeepMind" rėmėjas Peteris Thielis. Technologijų entuziastai troško išvysti šią utopiją savo akimis, todėl kai kurie, tarp jų – Altmanas ir Thielis, pasirašė sutartis su įmonėmis dėl galimybės po mirties išsaugoti jų smegenis ar net visą kūną. „Nesu visiškai tikras, kad tai pavyks, – sakė Thielis žurnalistės Bari Weiss tinklalaidėje. – Tačiau manau, kad turime bent jau pabandyti."

Problema ta, kad bėgant metams tam tikrų idėjų šalininkai tapo vis fanatiškesni. Pavyzdžiui, kai kurie DI akceleracionizmo[2] rėmėjai įtikėjo, kad mokslininkai turi moralinę pareigą kuo greičiau vystyti BDI, kad galiausiai sukurtų postžmogišką utopiją – savotišką moksliukų rojų. Jei pavyktų tai pasiekti dar jiems gyviems esant, jie galėtų gyventi amžinai. Tačiau siekdami paspartinti DI kūrimą, mokslininkai gali pradėti ieškoti lengviausių kelių tai padaryti ir galiausiai sukurti technologiją, kuri darytų neigiamą poveikį tam tikroms žmonių grupėms arba taptų nekontroliuojama.

Kiti laikėsi priešingos nuostatos: į DI žvelgė kaip į tam tikrą šėtonišką ateities reiškinį, kurį reikėjo sustabdyti. Iškiliausia šio ideologinio judėjimo figūra – Eliezeris Yudkowsky'is. Šis barzdotas libertarizmo šalininkas (tas pats, kuris per pasisėdėjimą prie puodelio kavos radikalizavo Jaaną Tallinną) savo įkurtame tinklalapyje „LessWrong" itin išpopuliarino šio judėjimo idėjas. 2014-aisiais, kai „Google" įsigijo

„DeepMind“, „LessWrong“ jau buvo tapęs aktyvių filosofinių debatų forumu, kuriame šimtai žmonių, įskaitant DI tyrėjus, diskutavo, ko reikėtų imtis, kad ateityje galingas superintelektas nesunaikintų žmonijos. „LessWrong“ tapo įtakingiausiu apokaliptinių DI grėsmių aptarimo forumu, o kai kurie žurnalistai jį netgi lygino su moderniu pasaulio pabaigos kultu. Jei kuris nors forumo narys aprašydavo naują būdą, kaip DI galėtų sunaikinti žmoniją, Yudkowsky'is viešai jį išplūsdavo didžiosiomis raidėmis ir pašalindavo iš grupės.

Laikui bėgant vadinamieji DI katastrofistai sulaukė turtingų technologijų entuziastų paramos, todėl, siekdami savo tikslų, ėmėsi kurti įmones ir dalyvauti formuojant valstybės politiką. Yudkowsky'io tinklalapis tapo toks įtakingas, kad nemažai jo ištikimų skaitytojų galiausiai prisijungė prie „OpenAI“.

Galbūt labiausiai nerimą keliančios ideologijos BDI kūrimo kontekste buvo susijusios su siekiu sukurti beveik tobulą žmonių rūšį skaitmeniniu pavidalu. Šią idėją iš dalies išpopuliarino Bostromo knyga „Superintelektas“ (*Superintelligence*). Ji turėjo paradoksalų poveikį DI sričiai: iš vienos pusės, skatino baimę dėl to, kad DI gali sunaikinti žmoniją, „paversdamas žmones sąvaržėlėmis“, iš kitos pusės, žadėjo nuostabią utopiją, kurią padėtų sukurti galingas, tinkamai išplėtotas DI. Pasak Bostromo, vienas įstabiausių šios utopijos aspektų – posthumanistinės eros žmonės, pasižymintys „gerokai pranašesniais gebėjimais už dabartinius žmones“ ir egzistuojantys skaitmeninėje terpėje. Šioje skaitmeninėje utopijoje žmonės galėtų patirti įprastiems fizikos dėsniams nepavaldžių būsenų, pavyzdžiui, mirti be pasekmių arba tyrinėti fantastiškus pasaulius. Jie turėtų galimybę grįžti į branginamas praeities akimirkas, leistis į naujus nuotykius ar netgi patirti skirtingas sąmonės formas. Bendravimas su kitais taptų daug gilesnis, nes naujieji žmonės galėtų tiesiogiai dalintis mintimis ir emocijomis, taip užmegzdami artimesnius ryšius.

Šios idėjos pakerėjo kai kuriuos Silicio slėnio technologijų entuziastus, kurie tikėjo, kad tokias fantastines realybes galima pasiekti tinkamais algoritmais. Bostromo vaizduojama futuristinė ateities pasaulio, galinčio tapti rojumi arba pragaru, vizija įkvėpė tokius DI kūrėjus kaip Samas Altmanas siekti sukurti BDI, kol to nepadarė Demisas Hassabisas Londone. Tokie vizionieriai buvo įsitikinę, kad tik jie vieni gali tai atlikti saugiai. Kitu atveju kas nors sukurtų su žmogiškosiomis vertybėmis nesuderintą BDI, kuris sunaikintų ne tik kelis milijardus Žemės gyventojų, bet galbūt ir trilijonus tobulų naujų skaitmeninių ateities žmonių. Vadinasi, visi prarastume galimybę gyventi nirvanoje. Vis dėlto Bostromo idėjos turėjo ir tamsiąją pusę: jos atitraukė dėmesį nuo galimo neigiamo DI poveikio dabartiniams žmonėms.

Šiuolaikinių technologinių ideologijų fone vykusios „DeepMind“ ir „Google“ derybos išryškino nemalonią tiesą – technologijų bendrovėms darėsi vis sunkiau rasti būdą, kaip užtikrinti atsakingą DI valdymą. Suinteresuotosios šalys siekė visiškai skirtingų, nesuderinamų tikslų: vienoje pusėje pasireiškė kone religinio lygmens fanatizmas, kitoje – nenumaldomas verslo augimo siekis.

Hassabisas laikėsi atokiau nuo šių ideologinių kovų. Tokią jo laikyseną iš dalies lėmė jo asmeniniai motyvai kurti BDI. Gyvendamas Anglijoje, toli nuo Silicio slėnio burbulo, jis buvo apsupęs save itin gabių DI mokslininkų ir inžinierių komanda, kuri netrukus dar labiau išsiplėtė. Artimi kolegos prisimena, kad tuo metu Hassabisas buvo pasiryžęs per penkerius metus išnarplioti galvosūkius, susijusius su BDI kūrimu, ir galbūt už tai pelnyti Nobelio premiją. Jis pernelyg nesijaudino, kad jo įmonė taps korporacinės milžinės dalimi. Jis manė, kad, sukūrus BDI, ekonomika taps atgyvena – nei „DeepMind“, nei „Google“ nebereikės rūpintis, kaip uždirbti pinigų. Tuo pasirūpins DI.

Sėkmingai užbaigus derybas ir į sutartį įtraukus reikalavimą įsteigti etikos tarybą, bendrovę „DeepMind“ „Google“ įsigijo už 650 mln. JAV dolerių. Ši suma buvo gerokai mažesnė už tą, kurią siūlė Zuckerbergas,

tačiau britų bendrovei tai buvo milžiniški pinigai. Be to, sutartis užtikrino, kad BDI kontrolė nepereitų į vienos korporacijos rankas.

„Google“ pinigų injekcija leido Hassabisui dar drąsiau vilioti talentus. Nors dalis darbuotojų buvo nepatenkinti, kad įmonė parduota technologijų milžinei, dauguma džiaugėsi smarkiai išaugusiais atlyginimais ir galimybe pelningai įsigyti „Google“ akcijų. Taip jų motyvacija likti bendrovėje tik sustiprėjo. Dabar Hassabisas galėjo ne tik ramiau žiūrėti į konkurentų veiksmus, bet ir pats vilioti žmones iš tokių technologijų gigantų kaip „Facebook“ ar „Amazon“ bei pritraukti didžiausius akademinio pasaulio talentus, siūlydamas jiems pasakiškus atlyginimus. Užsitikrinęs reikiamus finansinius ir žmogiškuosius išteklius, kad „DeepMind“ galėtų kurti dar pažangesnę technologiją, Hassabisas ir toliau įmonės veiklą laikė paslaptyje. Pavyzdžiui, atsidaręs pagrindinę bendrovės interneto svetainę, galėjai išvysti tik tuščią lapą su apskritu logotipu viduryje. „DeepMind“ DI laboratorija buvo tokia slapta, kad susirašinėdami su kandidatais, besikreipusiais dėl darbo Londono būstinėje, darbuotojai elektroniniuose laiškuose nenurodydavo įmonės adreso. Į darbo pokalbį atvykusius kandidatus bendrovės atstovas pasitikdavo netoliese esančioje Kings Kroso traukinių stotyje, o tada pėsčiomis palydėdavo į biurą.

Darbo pokalbiuose įkūrėjai greitai pakerėdavo kandidatus savo kalbomis. Ypač įtaigiai viziją pristatydavo Suleymanas. „Jis buvo nepaprastai charizmatiškas. Jis įtikindavo kandidatus, kad tai – unikali, vieną kartą gyvenime pasitaikanti galimybė prisidėti prie projekto, galinčio pakeisti pasaulį“, – pasakoja buvęs įmonės vadovas.

Iš dvidešimties minučių susitikimo su Suleymanu akademikai ir valstybės tarnautojai, turintys dešimt ar daugiau metų patirties ir galėję rinktis bet kurį puikiai apmokamą darbą privačiajame sektoriuje, išeidavo įtikėję savo misija padėti sukurti BDI. „Jis aiškino, kad revoliucija bus pagrįsta geresniais matematiniais modeliais“, – teigia buvęs „DeepMind“ vadovas. Hassabisas ir Suleymanas kartodavo, kad

samdo „geriausius pasaulio matematikus ir fizikus“. Be to, tapę „Google“ korporacijos dalimi, jie turėjo prieigą prie geriausių pasaulio superkompiuterių ir didžiausio kiekio duomenų, kuriuos galėjo pasitelkti DI modeliams mokyti.

Apie pusė „DeepMind“ naujokų atėjo iš akademinės aplinkos. Jie negalėjo patikėti savo sėkme: kadaise spaudęsi ankštuose kabinetuose ir iš dotacijų fondų kauliję pinigų tyrimams, dabar pateko į prabanga tviskančius, kosmopolitiškų restoranų ir sodų apsuptus biurus su ultragreitais kompiuteriais ir beveik neribotais ištekliais. Geriausia buvo tai, kad „DeepMind“ darbuotojai visiškai nesijautė dirbantys reklamos paslaugų milžinui. Jie atliko tyrimus prestižinėje mokslinių tyrimų organizacijoje, kurios nariai publikavo straipsnius recenzuojamuose žurnaluose, tokiuose kaip „Science“ ir „Nature“, ir sprendė didžiausias pasaulio problemas. Regis, jie sėkmingai derino tai, ką geriausio gali pasiūlyti mokslo ir verslo pasaulis, jei tai apskritai įmanoma.

Nors tuo metu darbuotojai to nežinojo, tokia idilė ilgai trukti negalėjo. Kurį laiką šešiaženkliai atlyginimai ir neįtikėtini darbo privalumai vertė darbuotojus užsimiršti ir nesusimąstyti, ar ne per gerai, kad „Google“ taip dosniai jiems moka vien už pastangas kurti geresnį pasaulį. Kartais apie tokį atotrūkį nuo realybės primindavo į biurą užsukti užsimanę buvę bendradarbiai iš akademinio pasaulio ar valstybės tarnybos.

„Jausdavau gėdą“, – prisimena buvęs darbuotojas, perėjęs į „DeepMind“ iš akademinės aplinkos. Buvusiems kolegoms, panorusiems pamatyti naująjį biurą, jis pasiūlė verčiau susitikti netoliese esančiame restorane. Net restoranas atrodė kukliau už „DeepMind“ valgyklą su prabangiu bufetu, vertu penkių žvaigždučių Dubajaus viešbučio. „Sunku patikėti, kaip smarkiai buvome atitrūkę nuo tikrojo pasaulio“, – priduria jis.

Su tyrėjais elgtasi kaip su roko žvaigždėmis – viskas jiems buvo patiekiama tarsi ant sidabrinės lėkštutės. Kartą vienas jų netgi kreipėsi į „DeepMind“ personalo skyrių, kuris įprastai rūpindavosi išlaidų

darbuotojams ar vizų klausimais, ir pareiškė, kad laiko požiūriu būtų daug efektyviau, jei braškės būtų patiekiamos be žalių viršūnėlių. Po dviejų dienų bufete puikavosi dubenys su žvilgančiomis braškėmis, nuo kurių buvo nuskabyti visi žali lapeliai.

Hassabisas nuolat primindavo darbuotojams bendrovės viziją – sukurti BDI. Jis tvirtino, kad esant tuometiniam mokslinių tyrimų ir technologinių atradimų tempui šį tikslą pavyks pasiekti per penkerius metus. Pasak buvusių „DeepMind“ darbuotojų, Hassabisas turi nepaprastą gebėjimą įkvėpti, vaizdžiai nusakydamas bendrovės ateities kryptį. Komandos susibūrimuose už biuro ribų jis kartu su Suleymanu rengdavo pristatymus, per kuriuos aptardavo bendrovės strategiją, tačiau jie labiau priminė motyvacines kalbas nei konkrečių veiksmų planus. Kalbėdami apie „DeepMind“ strategiją, įkūrėjai paprastai nesileisdavo į detales.

„Viskas buvo grindžiama vizija. Kiekvienas personalo narys skatintas prisidėti prie bendros misijos, – prisimena buvęs darbuotojas. – Demisas ir Mustafa buvo talentingi pasakotojai, jie puikiai vienas kitą papildė.“ Hassabisas buvo rimčiausias iš trijulės – jis iki išnaktų skaitydavo mokslinius darbus ir su aukščiausio lygio tyrėjais galėdavo valandų valandas aptarinėti įvairias metodikas. Su žemesnio rango darbuotojais, neturinčiais daktaro laipsnio, jis beveik nebendraudavo. Būtent jis daugiausia prisidėjo prie to, kad „DeepMind“ susiformavo griežta hierarchinė kultūra, pagrįsta akademine reputacija. Suleymanas labiausiai išsiskyrė charizmatiškumu. Jis, tikras vizionierius, gebėjo įtaigiai perteikti bendrą ateities viziją. Buvęs darbuotojas jį vadino užburiančiu vedliu, paskui kurį visi sekė. Shane’as Leggas, akademiškiausias iš visos trijulės, dažnai likdavo antrame plane. „Shane’as buvo ramesnis“, – teigia buvęs kolektyvo narys.

Buvęs kolega pasakoja, kad Hassabisas tvirtai tikėjo visa keičiančia DI galia ir darbuotojams aiškino, esą po penkerių metų jiems nebereikės rūpintis pinigais, nes, sukūrus BDI, ekonomika taps praeities dalyku.

Laikui bėgant toks požiūris įsigalėjo tarp daugelio vyresniųjų vadovų. „Jie aklai tikėjo savo pačių retorika, – teigia buvęs bendrovės vadovas. – Jie įtikėjo, kad kuria svarbiausią technologiją žmonijos istorijoje.“

Tuo metu Hassabisas ir Suleymanas rūpinosi etikos ir saugumo tarybos, dėl kurios sutiko „Google“, steigimo reikalais. Jie puikiai suprato, kaip svarbu užtikrinti šį saugiklį. Iš visos trijulės Suleymanas labiausiai palaikė tarybos įkūrimo idėją. „Google“ buvo įsipareigojusi veikti akcininkų naudai ir kasmet didinti jų pelną – bendrovė sėkmingai siekė šių tikslų. Viena vertus, tai leido „DeepMind“ steigėjams sutelkti talentus ir skaičiavimo pajėgumus, reikalingus BDI kurti, kita vertus, sukūrus BDI, „Google“ beveik neabejotinai norėtų jį monetizuoti ir kontroliuoti. Nors ir neturėdama aiškaus plano, trijulė tikėjosi, kad etikos taryba užkirs kelią netinkamam žmogaus gebėjimams prilygstančio DI naudojimui.

Maždaug po metų nuo „DeepMind“ įsigijimo „SpaceX“ būstinėje Kalifornijoje įvyko pirmasis šios tarybos posėdis. Jame dalyvavo Hassabisas, Suleymanas ir Leggas, taip pat Elonas Muskas ir milijardierius, vienas iš „LinkedIn“ įkūrėjų, rizikos kapitalo investuotojas Reidas Hoffmanas. Be jų, kaip pasakoja gerai informuotas šaltinis, taip pat dalyvavo Larry’is Page’as, „Google“ vadovas Sundaras Pichai’us, „Google“ Teisės skyriaus vadovas Kentas Walkeris, Hassabiso mokslinių darbų vadovas po doktorantūros studijų Peteris Dayanas ir Oksfordo universiteto filosofas Toby’is Ordas.

Posėdis praėjo sklandžiai, tačiau netrukus „DeepMind“ įkūrėjus pasiekė netikėta žinia iš „Google“ – bendrovė nebenorėjo steigti etikos tarybos. Tarybos idėją aktyviai rėmusį Suleymaną tai gerokai supykdė. „Google“ nurodė, kad kai kurių tarybos narių veikla gali sukelti interesų konfliktų (pavyzdžiui, Muskas ketino remti kitus DI projektus, nesusijusius su „DeepMind“), taip pat teigė, esą tarybos steigimas teisiškai būtų neįmanomas. Kai kuriems į tarybą paskirtiems asmenims

tokie argumentai pasirodė absurdiški – jie įtarė, kad „Google“ tiesiog nenorėjo iš rankų paleisti pelningos DI technologijos.

Pasipiktinę dėl tokio susitarimo nepaisymo, Hassabisas ir Suleymanas kreipėsi į „Google“ vadovybę. Siekdama nuraminti „DeepMind“ įkūrėjus ir motyvuoti juos siekti pažangos DI srityje, „Google“ pasiūlė kitą paskatą – aiškesnę organizacinę struktūrą, padėsiančią apsaugoti jų sukurtą BDI technologiją. Tuo metu „DeepMind“ steigėjai nežinojo, kad ruošiamasi „Google“ pertvarkyti į konglomeratą „Alphabet“, suteikiant įvairiems padaliniams daugiau veiklos laisvės. Vadovai nurodė, kad šie padaliniai veiks „autonomiškai“, tarsi nepriklausomos įmonės – su atskiru biudžetu, apskaita, valdyba ir net išoriniais investuotojais. Idėja atrodė viliojanti.

Iš tiesų „Google“ siekė akcijų vertės, kuri tuo metu stagnavo, prieaugio. Neskaitant tokių sėkmingų projektų kaip „YouTube“, „Android“ ir „Google Search“, kiti „Google“ padaliniai daugelį metų nesulaukdavo teigiamų „Wall Street“ analitikų vertinimų. Tarp tokių „Google“ valdomų subjektų – išmaniųjų termostatų bendrovė „Nest“, biotechnologijų tyrimų įmonė „Calico“, rizikos kapitalo padalinys ir mokslinių tyrimų laboratorija „X“. Dauguma jų nenešė pelno, tačiau, atskyrusi juos nuo pagrindinės bendrovės, „Google“ galėjo pagerinti finansinės atskaitomybės rodiklius ir reklamos, kuri sudaro daugiau kaip 90 proc. metinių pajamų, rezultatus. Nors „Google“ garsėjo kaip novatoriška technologijų bendrovė, kurioje dirba talentingiausi inžinieriai, vadovai vis tiek daugiausia susitelkė į tai, kas tradiciškai labiausiai rūpi prekybininkams, – įtikinti žmones pirkti tai, ko jiems nebūtinai reikia.

Visą dėmesį sutelkę į BDI kūrimą, Hassabisas, Leggas ir Suleymanas rimčiau nesusimąstė, kokie galėtų būti tikrieji „Google“ motyvai ir kad bendrovė greičiausiai neplanuoja jiems suteikti realios autonomijos, suprasdama, kokią naudą galėtų atnešti jų atliekami DI tyrimai. Kalbos apie galimybę laisviau veikti buvo lyg muzika jų ausims – tai reiškė, kad

„Google“ nekontroliuos jų sukurto DI, todėl jie patys galės rūpintis atsakingu jo valdymu. „Norėjome būti pakankamai nepriklausomi ir turėti veiksmų laisvę, jei atsirastų itin galingo BDI grėsmė, – prisimena Leggas. – Norėjome savo rankose išlaikyti tam tikrą kontrolę.“

Visus pusantrų metų „DeepMind“ steigėjai vis sugrįždavo prie diskusijų, per kurias kartu su Page’u ir kitais „Google“ vadovais aptardavo, kaip atrodys jų veikla tapus naujosios korporacijos dalimi ir ką reiškia „nepriklausomas padalinys“. Galiausiai pranešus apie „Google“ pertvarkymą į „Alphabet“, jokių konkrečių planų suteikti „DeepMind“ didesnį teisinį savarankiškumą taip ir nebuvo paskelbta. Nors kai kurie padaliniai, pavyzdžiui, „Verily Life Sciences“, tapo atskiromis įmonėmis, „DeepMind“ tokio statuso taip ir neįgijo. Atrodė, kad „Google“ vėl pamiršo savo įsipareigojimus.

Įtempti santykiai su „Google“ tuo metu nebuvo pagrindinis Hassabiso rūpestis. Jo laukė kur kas nemalonesnė naujiena. San Fransiske kūrėsi nauja tyrimų laboratorija, išsikėlusi tokį patį tikslą – kurti BDI. Tačiau naujoji įmonė puoselėjo visai kitokią, itin ambicingą idėją: DI kurti saugiai, siekiant dirbti žmonijos labui. Hassabisą tai skaudino – nenoromis piršosi išvada, kad „DeepMind“ iš tikrųjų dirbo ne žmonijai, o „Google“. Dar skaudžiau buvo tai, kad prie naujosios organizacijos prisidėjo buvęs „DeepMind“ investuotojas Elonas Muskas. Toji organizacija vadinosi „OpenAI“.

2 Akceleracionizmas – nuostata, kad technologiniai pokyčiai, ypač susiję su DI, turėtų vykti sparčiau, net jei tai sunaikintų esamas sistemas ir paskatintų radikalias socialines permainas (vert. past.).

6 SKYRIUS

Misija

Buvo 2015-ieji. Penkerius metus Demisas Hassabisas augino savo komandą ir lėtai, bet užtikrintai siekė mokslinių tikslų, artėdamas išsikelto siekio link – sukurti BDI. „DeepMind“ dirbo beveik be konkurencijos – niekas kitas nesiryžo žengti į šią sritį. Bendrovės užmojis buvo toks unikalus, kad ji galėjo veikti kaip monopolis. Kadangi jokia kita įmonė pasaulyje nesiekė kurti DI, galinčio pranokti žmogaus protą, Hassabisas galėjo vykdyti tyrimus savu tempu, nejausdamas spaudimo. Dėl to „DeepMind“ įkūrėjai ir komandos nariai labiau jautėsi kaip svarbią misiją vykdančios tyrimų laboratorijos, o ne komercinės įmonės darbuotojai. Nors dabar bendrovę valdė „Google“, juos ramino mintis, kad jie ir toliau kuria DI, kuris padėtų įveikti didžiausius žmonijos iššūkius. Skirtingai nei kitos įmonės, jie nebuvo įsivėlę į nuolatines konkurencines varžybas. Jų misija buvo unikali. Tačiau ramius vandenis sudrumstė Silicio slėnyje netikėtai iškilę varžovai. Pastangos sukurti BDI netruko peraugti į tikras lenktynes.

Kuo daugiau Hassabisas domėjosi „OpenAI“, tuo labiau augo jo įtūžis. Juk būtent jis pirmasis pasaulyje rimtai ėmėsi kurti BDI, rizikuodamas savo, kaip mokslininko, reputacija, žinodamas, kad vos penkeriais metais anksčiau tokia idėja vertinta itin prieštaringai. Dar labiau nerimą kėlė įtarimas, kad naujasis konkurentas perima jo idėjas. „OpenAI“ interneto svetainėje buvo įvardyti septyni bendrovės steigėjai. Atidžiau patyrinėjęs pavardes, Hassabisas suprato, kad penki iš jų keletą mėnesių darbavosi „DeepMind“ kaip konsultantai arba praktikantai. Anot buvusių kolegų, Hassabisą tai kaip reikiant supykdė. Jis visada su „DeepMind“ darbuotojais atvirai aptardavo įvairias strategijas, kurios padėtų kuriant

BDI, pavyzdžiui, idėją kurti savarankiškai veikiančias DI sistemas arba mokyti DI modelius žaisti tokius žaidimus kaip šachmatai ar go[3]. Dabar penki mokslininkai, girdėję visas šias detales, ėmėsi steigti konkuruojančią organizaciją.

Žvelgiant iš šalies, neatrodė, kad Hassabisas turi pagrindo jaudintis. Daugybė mokslininkų siekė panašių tikslų kaip „DeepMind“, tyrinėdami savarankiškai veikiančias sistemas, virtualiąsias aplinkas ir žaidimus. Vienas iš penkių minėtų asmenų, anksčiau dirbusių su „DeepMind“ projektais, buvo garsus DI mokslininkas Ilya Sutskeveris. „DeepMind“ labiau koncentravosi į skatinamąjį mokymąsi, o Sutskeveris specializavosi giluminio mokymosi srityje. Jis buvo paskirtas „OpenAI“ vyriausiuoju mokslininku ir, kaip ir kiti šios organizacijos steigėjai, tvirtai tikėjo BDI galimybėmis.

Hassabisas jautė nuoskaudą dėl to, kad Samas Altmanas drįso samdyti žmones, turinčius konfidencialios informacijos apie „DeepMind“ veiklą. Mintys apie tai jam nedavė ramybės net naktimis. Grįžęs iš darbo, Hassabisas pavakarieniaudavo su šeima, o tada vėl sėsdavo prie darbo stalo. Dažnai plušėdavo iki trečios ar ketvirtos valandos nakties – studijuodavo mokslinius darbus arba rašydavo elektroninius laiškus. Artimų šaltinių teigimu, tuose laiškuose arba per vėlyvus pasitarimus jis kartais išsakydavo nuogąstavimus, kad Altmanas kopijuoja „DeepMind“ strategiją ir bando pervilioti bendrovės tyrėjus.

Hassabisas skeptiškai vertino „OpenAI“ įkūrėjų pažadus jų sukurtą technologiją atiduoti visuomenės naudojimui. Tokį „atvirumą“ jis laikė neapgalvotu. „Man atrodė naivu tikėtis, kad atvirojo kodo projektas taps technologine panacėja, – teigia Hassabisas. – Dvejopos paskirties technologijos tampa vis galingesnės, todėl kyla grėsmė, kad jos gali patekti į piktavalių, galinčių jas panaudoti blogiems tikslams, rankas. Tokiu atveju būtų labai sunku juos sustabdyti.“ „DeepMind“ įkūrėjai dalį savo tyrimų rezultatų paskelbdavo prestižiniuose mokslo žurnaluose, tačiau sukurto kodo ir DI technologijos detales laikė paslaptyje.

Pavyzdžiui, jie taip ir nepaviešino savo sukurtų DI modelių, kurie padėjo sistemai įvaldyti žaidimą „Breakout“.

Be to, „DeepMind“ vadovai sužinojo, kad Muskas, bendraudamas su savo aplinkos žmonėmis Silicio slėnyje, dažnai bandydavo apjuodinti Hassabisą – tai atskleidė buvę „DeepMind“ ir „OpenAI“ darbuotojai. Pavyzdžiui, kalbėdamas su naujais „OpenAI“ darbuotojais, jis kritikuodavo „DeepMind“ veiklą Anglijoje ir pristatydavo Hassabisą kaip abejotinos reputacijos žmogų. Tokį įtarų požiūrį lėmė Hassabiso sukurtas žaidimas „Evil Genius“, kuriame žaidėjai įsijaučia į piktadario vaidmenį, siekdami sukonstruoti pasaulio pabaigos įrenginį ir pavergti žmoniją. Muskas manė, kad tokius žaidimus kuriantis asmuo veikiausiai pats yra šiek tiek pamišęs. „OpenAI“ darbuotojai pasigavo šį juokelį ir ėmė kurti memus iš „Evil Genius“ ekrano vaizdų, kuriais dalijosi per verslo pokalbių platformą „Slack“. Pasak vieno buvusio „OpenAI“ darbuotojo, Muskas netgi yra pavadinęs Hassabisą „DI srities Hitleriu“.

Kad ir kas paskatino tokį priešišką Musko nusiteikimą „DeepMind“ atžvilgiu, įtempti santykiai tarp abiejų organizacijų ilgainiui peraugo į aršią konkurencinę kovą. Su laiku Musko požiūris į DI darėsi vis labiau paranojiškas ir pesimistinis. Tai nestebina, žinant jo polinkį į kraštutinumus. Pavyzdžiui, siekdamas spręsti klimato kaitos problemą, jis užsibrėžė žmones paversti tarpplanetine rūšimi, nors galėjo tiesiog kovoti su naftos bendrovėmis. Taip pat nusprendęs, kad socialiniame tinkle „Twitter“ įsigali pernelyg progresyvios idėjos, jis įsigijo visą bendrovę, nors galėjo apsiriboti tik dalimi akcijų. Praėjus vos keleriems metams nuo to laiko, kai tapo „DeepMind“ investuotoju, technologijų magnatas jau buvo įtikėjęs apokaliptinėmis DI ateities vizijomis. Galbūt tai lėmė jo polinkis mestis į kraštutinumus ir elgtis drastiškai arba tikėjimas savo, kaip žmonijos gelbėtojo, vaidmeniu.

Leidinys „New York Times“ atskleidžia, kad, vakarais bendraudamas su žmona, Muskas dažnai prisipažindavo nerimaujantis, jog „Google“ bendrakūrėjui Page’ui, įsigijusiam anksčiau jo remtą „DeepMind“, gali

pavykti sukurti kur kas pažangesnes DI sistemas. Muskas su Page'u buvo artimi bičiuliai. Jiedu dalyvaudavo tose pačiose iškilmingose vakarienėse bei konferencijose ir puoselėjo panašias fantastines ateities vizijas. Pasak „Bloomberg" reporterio Ashlee Vance'o, lankydamasis San Fransiske ir neturėdamas kur apsistoti, Muskas skambindavo Page'ui, kad suderintų nakvynę jo namuose. Jiedu mėgo kartu žaisti kompiuterinius žaidimus ir diskutuoti apie futuristinius lėktuvus ar kitas technologijas. Muskui atrodė, kad Page'as, kuris vis rečiau dalyvavo viešajame gyvenime, tapo pernelyg patiklus, ir tai jam kėlė nerimą. Kaip teigiama Musko biografijoje, magnatas baiminosi, kad „Google" bendrakūrėjis gali netyčia sukurti ką nors pavojingo, pavyzdžiui, „ištisą DI turinčių robotų parką, pajėgų sunaikinti visą žmoniją". Atrodė, kad jis juokauja, tačiau iš tiesų kalbėjo visiškai rimtai.

Praėjus keliems mėnesiams nuo to laiko, kai Page'as už 650 mln. JAV dolerių įsigijo „DeepMind", DI temai skirtame forume Muskas paskelbė žinutę, kurią netrukus ištrynė: „Nė neįtariate, kaip greitai vystosi DI. Jei jums neteko matyti tokių įmonių kaip „DeepMind" iš vidaus, nė nenutuokiate, kaip sparčiai tai vyksta." Jis skeptiškai vertino galimybę, kad „pirmaujančios DI bendrovės" sugebės užtikrinti, jog skaitmeninis superintelektas nepatektų į internetą ir nesukeltų chaoso.

Pradėjęs gilintis į apokaliptinius DI ateities scenarijus, Muskas ėmė tam skirti vis daugiau laiko ir pinigų. Jis atseikėjo 10 mln. JAV dolerių „Future of Life Institute" – ne pelno organizacijai, kurios tikslas – skatinti tyrėjus ieškoti būdų, kaip neleisti DI sunaikinti žmonijos. Vėliau jis dalyvavo instituto surengtoje konferencijoje Puerto Rike, kurioje taip pat lankėsi Larry'is Page'as, Hassabisas ir kiti asmenys, rimtai nusiteikę kurti BDI.

Per vakarinę konferencijos dalį Muskas ir Page'as susiginčijo. Atmosferai įkaitus, aplink juos ėmė būriuotis kiti dalyviai. Page'as tvirtino, kad Muskas tampa pernelyg paranojiškas dėl DI. Esą žmonija žengia į skaitmeninę utopiją, kurioje mūsų sąmonė įgis ir skaitmeninį, ir

organinį pavidalą. Jis pridūrė, kad jeigu Muskas ir toliau kels tiek triukšmo dėl DI, jis tik sulėtins pažangą.

„Bet iš kur gali žinoti, kad superintelektas nesunaikins žmonijos?“ – paklausė Muskas.

„Tu diskriminuoji rūšies pagrindu“, – atšovė Page’as, kaip atkleidžia „New York Times“. Jis akivaizdžiai gynė būsimas posthumanistinės eros kartas. Muskas, sutelkęs tiek daug dėmesio į katastrofinius scenarijus, ignoravo būsimų skaitmeninės prigimties būtybių poreikius.

Viena vertus, stebėdamas „DeepMind“ veiklą ir įsitraukęs į pasiturinčių futuristų bendruomenę, Muskas pasidavė vis radikalesnėms idėjoms. Kita vertus, jį buvo apėmusi FOMO (angl. *fear of missing out*) – paralyžiuojanti baimė ką nors praleisti. Būtent tai Silicio slėnyje dažnai lemia svarbiausius investuotojų sprendimus. Dirbtinio intelekto srityje pasiekus naujų laimėjimų (pavyzdžiui, 2012-ųjų „ImageNet“ konkurso nugalėtojo pristatytas novatoriškas DI modelis), didžiosios technologijų bendrovės kaipmat ėmė dairytis į šią sritį. „Google“ įsigijo „DeepMind“, o Markas Zuckerbergas įsteigė naują tyrimų padalinį „Facebook AI Research“ (FAIR), kuriam vadovauti pakvietė vieną garsiausių giluminio mokymosi specialistų Yanną LeCuną. Veikiausiai paveiktas šios tyrimų srities „aukso karštinės“, Muskas ryžosi, regis, jo paties baimėms prieštaraujančiam žingsniui – kurti DI.

Vėliau Muskas „Twitter“ paskelbė „OpenAI“ įkūręs tikėdamasis, jog ji taps tam tikra atsvara „Google“ ir užtikrins saugesnę DI plėtotę. Vis dėlto nekyla abejonių, kad DI buvo itin svarbus jo kuriamų produktų sėkmei – nesvarbu, ar tai būtų savavaldžiai „Tesla“ automobiliai, „SpaceX“ raketų valdymo sistemos, ar smegenų ir kompiuterio sąsają pagrindžiantys modeliai, kuriuos apsiėmė kurti jo įmonė „Neuralink“.

Nepaisant Musko apokaliptinių prognozių ir įsitikinimo, kad būtent jis, o ne Demisas Hassabisas, turi tapti pirmuoju, sukūrusiu BDI, DI technologija, prilygstanti „Google“ vystomam DI, taip pat galėjo itin sustiprinti jo verslą. Toks projektas galėjo atnešti milžinišką pelną. Tai

paaiškina jo norą bendradarbiauti su Samu Altmanu, garsėjusiu plačiais verslo ryšiais Silicio slėnyje ir gebėjimu investuotojams skirtuose pristatymuose milijonus pakeisti milijardais. Be to, jis sugebėjo į „Y Combinator“ pritraukti gausybę startuolių, plėtojančių futuristines technologijas, o jo ambicijos DI srityje buvo ne menkesnės nei Larry’io Page’o.

2015-ųjų gegužės 25 dieną Muskui išsiųstame elektroniniame laiške Altmanas teigė, kad „kas nors turėtų sukurti BDI anksčiau, nei tą padarys „Google“. Jis pasiūlė kartu imtis įgyvendinti DI projektą, kuriuo būtų siekiama sukurti technologiją, tarnausiančią žmonijos labui. „Turbūt vertėtų tai aptarti“, – atsakė Muskas. Po mėnesio Altmanas vėl jam parašė, siūlydamas įsteigti laboratoriją, kuri sukurtų „pirmąjį bendrąjį dirbtinį intelektą, prioritetą teikdama saugumui“. Toks DI būtų valdomas ne pelno organizacijos ir naudojamas „žmonijos labui“. Muskas atsakė: „Su viskuo sutinku.“

Altmanas buvo įsitikinęs, kad bendrosios paskirties DI sistema galėtų apimti viską, ką gautume, į vieną krūvą sudėję visus technologijų startuolius, kurie kada nors yra dalyvavę „Y Combinator“ programoje. Toks mašininis intelektas atvertų beveik neribotas galimybes. Kas žino, gal naujasis superintelektas padėtų sukurti visuotinę gerovę, ir žmonėms apskritai nebereikėtų kurti įmonių ar startuolių. Hassabisas tikėjo, kad BDI padės įminti mokslo paslaptis ir pažinti dieviškąsias ištakas. Tuo metu Altmanas į šią technologiją žvelgė kaip į priemonę, galinčią užtikrinti ekonominį klestėjimą. Jiedu su Musku svarstė, kad šį tikslą jie pasiekti galėtų įkūrę tyrimų laboratoriją, kuri veiktų kaip atsvara „DeepMind“ ir „Google“.

Altmanas su Musku akcentavo dar vieną aspektą, kuriuo jų naujoji organizacija skirtųsi nuo didžiųjų technologijų bendrovių. Siekdama kurti DI, kuris neštų naudą visai žmonijai, laboratorija bendradarbiautų su kitais institutais, o jos tyrimų rezultatai būtų prieinami viešai. Būtent todėl naująją organizaciją nuspręsta pavadinti „OpenAI“.

Altmanas nieko nelaukęs ėmėsi burti steigėjų komandą. 2015-ųjų vasaros pavakarę apie tuzinas jo pakviestų geriausių DI tyrėjų susirinko į prabangų viešbutį „Rosewood“, įsikūrusį vos keli žingsniai nuo stambiausių Silicio slėnio rizikos kapitalo įmonių. Tarp pakviestųjų į uždarą vakarienę buvo Ilya Sutskeveris, kompiuterių mokslininkas, kelis mėnesius dirbęs „DeepMind“, ir Gregas Brockmanas, Harvardo universiteto matematikos absolventas iš Šiaurės Dakotos, anksčiau ėjęs technologijų vadovo pareigas įmonėje „Stripe“.

Vakarienės metu Altmanas paaiškino, kad naujosios tyrimų organizacijos tikslas – sukurti BDI ir užtikrinti, kad jis neštų naudą žmonijai. Susirinkusieji ilgai diskutavo, ar tai apskritai įmanoma. Ne todėl, kad būtų sunku DI sukurtas gėrybes paskirstyti žmonijai, bet dėl to, kad technologijų milžinės jau buvo pasigviešusios ryškiausius DI talentus. Ar ne per vėlu bandyti pritraukti geriausius šios srities tyrėjus?

„Žinojome, kad mūsų ištekliai toli gražu neprilygs didžiųjų technologijų bendrovių pajėgumams“, – pasakojo Brockmanas Lexo Fridmano tinklalaidėje. Tačiau jei vis tik jiems pavyktų įkurti tokią organizaciją, svarbiausia jų užduotis būtų užtikrinti, kad DI tarnautų žmonijai. „Buvo aišku, kad organizacija turėtų būti ne pelno. Neturėdama jokių gretutinių tikslų, ji galėtų nenukrypti nuo savo pirminės misijos“, – priduria Brockmanas.

Po konferencijos kartu su Altmanu grįždamas namo, Brockmanas pareiškė nusprendęs prie jo prisidėti, nors išsikelti tikslai ir atrodė nerealistiški. Juk Silicio slėnyje prigyja net beprotiškiausios idėjos.

Pats būdamas be galo atsidavęs darbui, Altmanas negalėjo atsistebėti Brockmano ryžtu – šis nieko nelaukęs ėmėsi rūpintis „OpenAI“ steigimo reikalais. Žinant, kad Brockmanas į el. laiškus paprastai atsakydavo vos per penkias minutes, buvo galima numanyti, jog jis bus lygiai taip pat fanatiškai pasinėręs į naująjį sumanymą kaip ir Altmanas. „Jis buvo pasiruošęs atiduoti visas savo jėgas“, – vėliau sakė Altmanas.

Brockmanas į savo rankas perėmė visus organizacinius klausimus, susijusius su „OpenAI“ steigimu.

Netrukus Brockmanas ėmėsi dairytis talentų, kuriuos galėtų persivilioti į naująją organizaciją iš tokių bendrovių kaip „Google“ ir „Facebook“. Jis susisiekė su Monrealio universiteto profesoriumi Yoshua Bengio, vadinamu vienu iš giluminio mokymosi judėjimo „krikštatėvių“. Brockmanas nenorėjo samdyti paties Bengio – jam tereikėjo, kad profesorius rekomenduotų perspektyviausius pažįstamus DI mokslininkus. Bengio sudarė sąrašą ir nusiuntė jį Brockmanui.

Pritraukti šiuos žmones pasirodė ne taip jau paprasta. Kai kurie iš jų, dirbę technologijų milžinėse, tokiose kaip „Google“ ar „Facebook“, galėjo pasigirti septynženkliais atlyginimais. Altmanas su Brockmanu negalėjo pasiūlyti tokio dosnaus atlygio. Vis dėlto kandidatus jie galėjo sudominti galimybe prisidėti prie pasaulį keičiančio projekto ir dirbti su dviem iškiliomis asmenybėmis. Elonas Muskas tuo metu jau buvo tapęs visuotinai gerbiamu magnatu, o Altmanas, sėkmingai vadovavęs „Y Combinator“, Silicio slėnyje buvo žmogus, su kuriuo visi veržėsi susipažinti. Net trumpa patirtis šioje ne pelno organizacijoje DI tyrėjams galėjo suteikti įtakingų pažinčių ir profesinio augimo galimybių, kurios dažnai yra vertingesnės už didelį atlyginimą.

Keletas garsiausių mokslininkų iš Brockmano sąrašo sutiko su juo susitikti. Be galimybės dirbti petys petin su iškiliomis asmenybėmis ir prisidėti prie kilnios misijos, juos traukė organizacijos atvirumas. Pagaliau jie galėtų skelbti savo tyrimų rezultatus, užuot slapta dirbę su korporacijų produktais. Kaip pasakoja buvę „OpenAI“ darbuotojai, kitiems patiko, kad naujojoje organizacijoje, skirtingai nei „Google“ ir „DeepMind“, pelno siekis nebuvo pagrindinė varomoji jėga kuriant BDI.

Brockmanas nusprendė keletą potencialių bendražygių pakviesti į vyninę, norėdamas pagaliau su jais sukirsti rankomis. Iš visų specialistų, kuriuos tikėjosi prisivilioti, didžiausiu laimikiu jis laikė Sutskeverį. Susitikimo dalyviai plačiau aptarė būsimos DI laboratorijos viziją. Tokia

organizacija turėjo būti visiškai nepriklausoma nuo korporacinių jėgų ir veikti pagal atvirojo kodą principą, viešindama tyrimų rezultatus ir leisdama visiems nemokamai su jais susipažinti. Taip jie siekė užtikrinti, kad technologijų milžinės, tokios kaip „Google“ ir „Facebook“, neįgytų absoliučios DI kontrolės, šiai technologijai tampant vis galingesnei. Beveik visi mokslininkai sutiko prisijungti, tarp jų ir Sutskeveris – talentingas mokslininkas, kurį retai galėdavai išvysti besišypsantį. Rusijoje gimęs, Izraelyje užaugęs DI ekspertas, anksčiau dirbęs su giluminio mokymosi pradininku Geoffrey'iu Hintonu, nusprendė palikti „Google Brain“ ir prisijungti prie „OpenAI“.

Naujosios tyrimų laboratorijos komanda, sudaryta iš maždaug tuzino darbuotojų, apie save pasauliui pranešė 2015-ųjų gruodį Kanados mieste Monrealyje vykusioje kasmetinėje DI konferencijoje NIPS (dabar „NeurIPS“). Už lango krintant snaigėms, komandos nariai stengėsi susirinkusiesiems pagarsinti žinią apie naująją laboratoriją. Tačiau oficialus prisistatymas įvyko internete. Organizacijos interneto svetainėje „*OpenAI.com*“ Brockmanas ir Sutskeveris paskelbė tinklaraščio įrašą, kuriame pristatė savo naująjį projektą. „Mūsų tikslas – kurti DI taip, kad jis teiktų didžiausią naudą žmonijai, nesiekiant finansinės grąžos“, – rašė jie.

Naujosios organizacijos pirmininkais turėjo tapti Muskas su Altmanu. Projektui buvo pažadėta įspūdinga 1 mlrd. JAV dolerių suma, kurią įsipareigojo skirti Muskas, Thielis, Altmanas, Hoffmanas ir Jessica Livingston. Be to, „OpenAI“ buvo suteiktas tam tikras nemokamų „Amazon“ debesijos infrastruktūros paslaugų limitas. Gerai informuoto šaltinio teigimu, Muskas planavo „OpenAI“ finansuoti „Tesla“ akcijomis. Analogišką pasiūlymą keleriais metais anksčiau jis buvo pateikęs bendrovei „DeepMind“.

Žinia apie „OpenAI“ atsiradimą pribloškė šimtus „NeurIPS“ konferencijos dalyvių. Daugelis manė, kad BDI kūrimas – tik utopinė svajonė, tačiau kai kuriuos apėmė pavydas. Pastarąjį dešimtmetį

didžiosios technologijų bendrovės iš universitetų paviliodavo ryškiausius kompiuterių mokslo talentus. Taigi šviesiausi DI srities protai buvo pajungti korporaciniams interesams. Universitetų paruošti DI specialistai beveik automatiškai atsidurdavo „Google“, „Facebook“ arba „Amazon“. Ši problema buvo matoma ne vienus metus.

„Nė vienas neatsisakytų dvigubai ar net trigubai didesnio atlyginimo, – sako Londono imperatoriškojo koledžo profesorė Maja Pantic, kuri 2018 metais pradėjo eiti „Samsung Electronics“ DI centro mokslinių tyrimų vadovės pareigas, o vėliau perėjo dirbti į „Meta“. – Tam neatsispyriau nei aš, nei mano kolegos.“ Tas pats nutiko ir šviesiausiems šios srities protams. Geoffrey’is Hintonas jau dirbo „Google“, kaip ir Stanfordo universitetą palikusi profesorė Fei-Fei Li; Yanną LeCuną persiviliojo „Facebook“. Iš Stanfordo universiteto išėjęs profesorius Andrew Ng’as dirbo „Google“ prieš prisijungdamas prie Kinijos korporacijos „Baidu“. Net geriausios aukštojo mokslo įstaigos, tokios kaip Stanfordo, Oksfordo universitetai ar Masačusetso technologijos institutas, sunkiai pajėgė išlaikyti iškiliausius akademikus. Retėjančios dėstytojų gretos tapo tikru iššūkiu universitetų bendruomenėms. Vis daugiau DI tyrimų buvo atliekama atokiau nuo visuomenės akių ir vaikantis pelno. Būtent dėl šios priežasties tyrėjai sveikino Musko ir Altmano pastangas užtikrinti šios srities tyrimų skaidrumą. Kas nors juk turėjo spręsti DI žinių koncentracijos didžiosiose korporacijose problemą.

Protų nutekėjimą iš universitetų lėmė dvi priežastys. Pirmoji – atlyginimų dydis. Toronto universitete, kuriame kadaise dėstė DI krikštatėviu tituluojamas Geoffrey’is Hintonas, kompiuterių mokslo profesoriai galėjo tikėtis apie 100 tūkst. JAV dolerių siekiančio metinio atlyginimo. Geriausiai apmokami akademikai gaudavo iki 550 tūkst. JAV dolerių per metus – tai buvo atlyginimų lubos. Hintono talentingasis studentas Sutskeveris nė nesvarstė žengti į akademinį pasaulį. Kurį laiką padirbėjęs Hintono įkurtame startuolyje, jis perėjo į „Google Brain“. Kaip

teigiama knygoje „Genijų kūrėjai“ (*Genius Makers*), kai „OpenAI“ bandė privilioti Sutskeverį 2 mln. JAV dolerių metiniu atlyginimu, „Google Brain“ pasiūlė trigubai didesnį atlygį.

Antroji priežastis – būtinybė užsitikrinti pakankamą duomenų kiekį ir skaičiavimo galią, reikalingą DI tyrimams. Universitetai paprastai turi ribotą skaičių GPU – galingų „Nvidia“ gaminamų grafikos procesorių, naudojamų daugumoje serverių, skirtų DI modeliams mokyti. Dar universiteto laikais Majai Pantic pavyko parūpinti vos šešiolika GPU, kuriais turėjo naudotis trisdešimties tyrėjų grupė. Turint tiek mažai procesorių, DI modelio mokymas užtruktų ištisus mėnesius. „Situacija buvo tiesiog absurdiška“, – prisimena profesorė. Netrukus po to, kai Pantic pradėjo dirbti „Samsung“, ji gavo prieigą prie dviejų tūkstančių GPU. Su tokiais papildomais pajėgumais algoritmo mokymas galėjo užtrukti vos kelias dienas, o tyrimai vykti kur kas sparčiau.

Net universitetuose likusiems mokslininkams darėsi vis sunkiau išvengti didžiųjų technologijų bendrovių įtakos. 2022 metais atliktas tyrimas parodė, kad per pastarąjį dešimtmetį akademinių straipsnių, susijusių su technologijų milžinėmis, padaugėjo daugiau nei tris kartus – iki 66 proc. Kelių mokslo įstaigų, įskaitant Stanfordo universitetą ir Dublino universiteto koledžą, atlikto tyrimo autoriai teigė, kad didėjančią šių bendrovių įtaką galima palyginti su didžiųjų tabako gamintojų naudojamomis strategijomis. Tai savo ruožtu paveikia kriterijus, kuriais remdamiesi universitetai vertina DI tyrimų sėkmę. Pasak tyrimui vadovavusios Abebos Birhane, dabar einančios vyresniosios mokslo darbuotojos pareigas „Mozilla Foundation“, akademikai, užuot susitelkę į tokias vertybes kaip žmonių gerovė, teisingumas ar įtrauktis, dažniau koncentruojasi į modelių rezultatų gerinimą.

„Gerovė ir įtrauktis nėra kažkas neapčiuopiamo, – teigia Birhane. – Šiuos dalykus galima objektyviai įvertinti. Galbūt jie atrodo abstraktūs, tačiau taip pat abstraktūs yra ir tokie dalykai kaip našumas ar

efektyvumas. Tyrėjai yra sugalvoję būdų, kaip objektyviai įvertinti teisingumą, privatumą ir kitus reiškinius.“ Pasak Birhane, situaciją komplikuoja tai, kad viso pasaulio tyrėjai tiek universitetuose, tiek technologijų bendrovėse koncentruojasi vien į DI modelių masto ir pajėgumų didinimą, todėl kyla grėsmė, kad tokių modelių generuojamas turinys gali skatinti rasizmą arba seksizmą. „Paaiškėjo, kad kuo didesni duomenų rinkiniai naudojami, tuo daugiau atsiranda neapykantą kurstančio turinio“, – pabrėžia Birhane, atkreipdama dėmesį į kitą 2023 metais atliktą jos ir bendraautorių tyrimą.

Vis dėlto mastas tapo lemiamu veiksniu, leidusiu didžiosioms technologijų bendrovėms stiprinti DI pajėgumus. „Google“ ir „Meta“ (anksčiau žinoma kaip „Facebook“) modeliams mokyti gali pasitelkti trilijonus duomenų vienetų, o jų valdomi duomenų centrai driekiasi kelių dešimčių tūkstančių kvadratinių metrų plotą užimančiose teritorijose. Vien tik „Google“ duomenų centras Dalse, Oregono valstijoje, dydžiu pranoksta šešis futbolo stadionus. Universitetai nė iš tolo negali konkuruoti su tokiais pajėgumais.

Norint sukurti pažangų DI, galioja paprasta taisyklė: kuo daugiau duomenų panaudojama, tuo geresnių rezultatų galima tikėtis. Pradėjęs dirbti „OpenAI“, Ilya Sutskeveris su savo komanda siekė kurti kuo pajėgesnius DI modelius, mažai tesirūpindamas lygiateisiškumo, teisingumo ar privatumo principais. Jei modelis mokomas vis didesniais duomenų rinkiniais, didinant jo parametrų skaičių ir skaičiavimo galią, jis tampa išmanesnis. Tą patį dėsningumą jau buvo pastebėjęs profesorius Andrew Ng’as, atlikdamas eksperimentus Stanforde. Kad ir kam būtų skirtas modelis, užtikrinus optimalias mokymo sąlygas, jis gebės geriau atlikti vertimo užduotis ar generuoti tekstą, artimą sukurtam žmogaus.

„Turint pakankamai didelius duomenų rinkinius ir didžiulį neuroninį tinklą, sėkmė garantuota“, – vienoje konferencijoje sakė Sutskeveris. Paskutinieji du šios frazės žodžiai DI bendruomenėje prigijo kaip

sparnuota frazė tuo metu, kai šioje srityje kilo nauja entuziazmo banga dėl naujos ne pelno organizacijos, už kurios vairo stojo talentingas mokslininkas ir kelios įtakingiausios Silicio slėnio figūros.

Netrukus organizaciją užklupo ir pirmieji sunkumai. „OpenAI“ nebuvo gavusi dar gruodį žadėtos 1 mlrd. JAV dolerių sumos, kurią turėjo skirti Elonas Muskas, Peteris Thielis ir kiti investuotojai. Kaip atskleidė naujienų portalo „TechCrunch“ tyrėjai, išnagrinėję „OpenAI“ deklaracijas, pateiktas JAV federalinei mokesčių tarnybai, per kelerius metus organizacija iš savo rėmėjų tesurinko kiek daugiau nei 130 mln. JAV dolerių finansavimo lėšų.

Pinigų trūkumas nebuvo vienintelė problema – organizacija taip pat neturėjo aiškios krypties. Trisdešimt pagrindinių „OpenAI“ tyrėjų darbo dienomis sugužėdavo į Brockmano butą San Fransisko Mišeno rajone. Vieni darbuodavosi palinkę prie virtuvės stalo, kiti įsitaisydavo ant sofos, ant kelių pasidėję nešiojamuosius kompiuterius. Praėjus keliems mėnesiams nuo veiklos pradžios juos aplankė kitas gerbiamas „Google Brain“ tyrėjas Dario’us Amodei’us. Jis ėmė klausinėti apie jų planus sukurti draugišką DI ir atskleisti jo pirminį kodą. Kaip cituojama žurnale „New Yorker“, Altmanas paprieštaravo, sakydamas, kad jie neketina paviešinti viso pirminio kodo.

„Tai koks jūsų veiklos tikslas?“ – nepasidavė Amodei’us.

„Na, jis kiek neaiškus“, – pripažino Brockmanas. Iš esmės jų tikslas buvo užtikrinti, kad BDI kūrimas vyktų saugiai.

Amodei’us buvo vienas iš tų mokslininkų, kurie, kaip ir Muskas bei Eliezeris Yudkowsky’is, būgštavo dėl galimų apokaliptinių DI ateities scenarijų. Dar dirbdamas „Google“, jis iš arti stebėjo metais anksčiau įvykusį incidentą, kai nuotraukų programėlės vaizdo atpažinimo sistema juodaodžius asmenis priskyrė goriloms. Aštrios kritikos sulaukę bendrovės atstovai pripažino, kad incidentas juos šokiravo, ir pašalino gorilų žymę iš programėlės. „Sistemos, kuriose išryškėja netikėti

trūkumai, nežada nieko gero“, – pasakojo Amodei’us vienoje tinklalaidėje, prisimindamas nemalonų incidentą.

Tačiau Amodei’ui nerimą kėlė ne tik rasistiniai ar įžeidžiantys algoritmų sprendimai. Jis nuogąstavo ir dėl skatinamojo mokymosi metodo, kurį buvo įvaldžiusi „DeepMind“, naudojimo fizinių sistemų – robotų, savavaldžių automobilių ir „Google“ duomenų centrų – valdyme. „Manau, tada, kai DI sistemos pradeda tiesiogiai sąveikauti su pasauliu ir kontroliuoti fizinius objektus, problemų atsiradimo tikimybė tik didėja“, – perspėjo jis 2016 metais duodamas interviu Jaano Tallinno įkurto „Future of Life Institute“ atstovams.

Tyrinėdamas galimas DI grėsmes, Amodei’us vis labiau įtikėjo katastrofiškų scenarijų galimybe. 2023-iaisiais, duodamas interviu žiniasklaidai, jis įspėjo, kad yra 25 procentų tikimybė, jog nekontroliuojamas DI gali sunaikinti žmoniją. Tokiems pavojams užkirsti kelio dirbdamas „Google Brain“ jis negalėjo, todėl, praėjus vos keliems mėnesiams nuo pirmojo apsilankymo „OpenAI“ biure, prisijungė prie šios organizacijos.

Kad galėtų kurti BDI, „OpenAI“ tyrėjų komandai reikėjo pritraukti daugiau lėšų ir talentų. Todėl jie ėmėsi projektų, kurie galėtų atkreipti žiniasklaidos dėmesį ir pristatyti organizaciją teigiamoje šviesoje. Tarp tokių projektų reikėtų paminėti kompiuterį, gebantį nugalėti elitinius žaidėjus strateginiame trijų matmenų vaizdo žaidime „Dota“, ir neuroniniu tinklu valdomą robotizuotą ranką, galinčią surinkti Rubiko kubą. Šiais sumanymais tikėtasi pradžiuginti Eloną Muską ir pranokti anapus Atlanto atokiai nuo visuomenės akių kuriamus „DeepMind“ projektus.

Tai, kad Muskas įtariai žvelgė į „DeepMind“, niekam nebuvo paslaptis. 2017 metais „OpenAI“ darbuotojai buvo pakviesti dalyvauti išvažiuojamajame susitikime „SpaceX“ būstinėje. Muskas, iš pradžių lankydavęsis „OpenAI“ biure kas savaitę, o vėliau – kas kelias savaites, aprodė jiems patalpas, tuomet maždaug keturiasdešimčiai naujų DI

tyrėjų surengė klausimų ir atsakymų sesiją. Kalbėdamas apie priežastis, paskatinusias jį paremti „OpenAI“, Muskas atskleidė, kad toji priežastis buvo Demisas Hassabisas.

„Buvau vienas iš „DeepMind“ investuotojų. Man nerimą kėlė tai, kad Larry’is Page’as manė, jog Demisas Hassabisas dirba jam. Iš tikrųjų Demisas dirba tik sau. Demisas man nekelia pasitikėjimo“, – pasakojo Muskas, pasak susitikime dalyvavusio šaltinio.

Mokslininkai buvo priblokšti. Daugeliui jų susidarė įspūdis, kad pagrindinis Muską motyvuojantis veiksnys yra asmeninis priešiškumas Hassabisui, o ne pavojai, susiję su nesaugiu DI kūrimu. Paklaustas, kodėl jaučia tokį priešiškumą Hassabisui, jis prisiminė britų verslininko sukurtus kompiuterinius žaidimus, kuriuose vaizduojamas pasaulio pavergimas.

Muskas taip pat prisiminė pokalbį su kitu „DeepMind“ investuotoju, pareiškusiu, kad per ankstesnį susitikimą su Hassabisu jam atrodė „lyg būtų atsidūręs toje filmo vietoje, kai kas nors turi atsistoti ir nušauti tą vyruką.“ Kitaip tariant, kas nors privalėjo sustabdyti Hassabisą, kad šis nesukurtų itin galingo BDI.

Nors buvo galima pamanyti, kad Muskas nemėgo Hassabiso, jis nuolat primindavo „OpenAI“ komandai, kad „DeepMind“ yra priešakyje, ir britų bendrovės mokslinius pasiekimus pateikdavo kaip sektiną pavyzdį. Buvusių „OpenAI“ darbuotojų teigimu, laikui bėgant Muskas vis labiau nerimavo, kad jų technologijos nusileidžia „DeepMind“ kuriamoms sistemoms.

Kad išlaikytų svarbiausią rėmėją, Altmanas su Brockmanu pavedė tyrėjams užduotis, kurios priminė „DeepMind“ projektus. „Dota“ projekto komanda nesuprato, kodėl turėjo dirbti su žaidimo simuliacija, jei jų tikslas buvo kurti BDI, galintį pagerinti žmonių gyvenimą. Atsakymas buvo paprastas – jiems reikėjo Musko pinigų. „Jeigu tuo neužsiimsime, po kelių metų, o gal net jau kitais, „OpenAI“ gali nebelikti“, – atvirai kalbėjo Brockmanas.

Nors vėliau „OpenAI“ ištobulinti pokalbių robotai (angl. *chatbots*) ir didieji kalbos modeliai atnešė bendrovei visuotinį pripažinimą, pirmuosius kelerius gyvavimo metus organizacija praleido plušėdama su įvairiais projektais, susijusiais su daugiaagentėmis simuliacijomis (angl. *multi-agent simulations*) ir skatinamuoju mokymusi – tuo metu šiose srityse jau pirmavo „DeepMind“. Stengiantis neatsilikti nuo britų konkurentų, Altmanui ir kitiems vadovams darėsi vis aiškiau, kad toks požiūris į DI kūrimą nepadės pasiekti didesnių, praktinę naudą galinčių atnešti laimėjimų. Būtent tada „OpenAI“ pasuko visai kitu keliu nei „DeepMind“. Pastaroji organizacija garsėjo hierarchine, akademine kultūra, kurioje labiausiai vertinti daktaro laipsnį turintys darbuotojai. „OpenAI“ komandos branduolį daugiausia sudarė inžinieriai. Tarp svarbiausių bendrovės tyrėjų buvo programuotojai, programišiai ir startuolių, dalyvavusių „Y Combinator“ programoje, steigėjai. Jiems rūpėjo ne tiek prestižas mokslo pasaulyje, kiek įvairių produktų kūrimas ir galimybė uždirbti.

Tuo metu Musko kantrybė seko. Jis priekaištavo Altmanui, kad šis subūrė įspūdingą mokslininkų komandą, bet taip ir nesukūrė nieko, kas pranoktų „DeepMind“ pasiekimus. Žengiant į trečiuosius veiklos metus, Muskas pasakė Altmanui, kad jų ne pelno organizacija pernelyg atsilieka nuo „Google“ ir „DeepMind“. Jis pasiūlė sprendimą: perimti „OpenAI“ kontrolę ir sujungti ją su „Tesla“. „Nesiėmus esminių pokyčių, „OpenAI“ niekada nepasivys „DeepMind“, – 2018 metų gruodį Altmanui ir komandai atsiųstame elektroniniame laiške rašė Muskas. Laišką vėliau paskelbė pati „OpenAI“, o jo turinį patvirtino originalą matęs asmuo. „Deja, žmonijos ateitis dabar priklauso nuo Demiso“, – pridūrė Muskas. Kitaip tariant, jis manė, kad jei neperims „OpenAI“ valdymo, Hassabisas, esą turintis piktų kėslų, galiausiai įgyvendins savo užmojus. Tačiau Altmanas ir jo bendražygiai norėjo organizacijos kontrolę išlaikyti savo rankose, todėl atmetė Musko pasiūlymą.

2018-ųjų vasarį „OpenAI“ paviešintame pranešime apie naujus rėmėjus užsiminta, kad Muskas palieka organizaciją, o tokio pasitraukimo priežastis – etiniai sumetimai. Buvo teigiama, kad jo veikla DI srityje gali sukelti interesų konfliktų. „Elonas Muskas pasitrauks iš „OpenAI“ valdybos, bet toliau rems organizaciją finansiškai ir teiks jai patarimų, – rašoma ne pelno organizacijos tinklaraštyje. – Kadangi „Tesla“ vis labiau koncentruojasi į DI, šis žingsnis užkirs kelią galimiems interesų konfliktams.“

Tačiau dauguma „OpenAI“ darbuotojų puikiai suprato, kad tai – tik priedanga. Jie įtarė, jog, nepaisant Musko pareiškimų apie saugesnio DI kūrimą, jis iš tiesų siekė sukurti galingiausią DI. Tuo metu Muskas jau buvo tapęs turtingiausiu žmogumi pasaulyje, be to, JAV infrastruktūros atžvilgiu jis buvo įgijęs nemažai svertų: „SpaceX“ skraidino NASA astronautus į kosmosą, „Tesla“ diktavo elektromobilių standartus, o palydovinio interneto paslaugų bendrovė „Starlink“ tapo vienu iš svarbiausių veiksnių, galinčių lemti karo Ukrainoje eigą.

Be to, paaiškėjo, kad Muskas – labai nepatikimas partneris. Jis buvo pažadėjęs per kelerius metus „OpenAI“ skirti 1 mlrd. JAV dolerių, tačiau realiai įnešė tik apie 50–100 mln. – tai nemenkas apsiskaičiavimas turtingiausiam pasaulio žmogui, kuris taip aktyviai skambino pavojaus varpais apie DI grėsmes. Šią sumą jis galėjo gan nesunkiai parūpinti, ypač jei „OpenAI“ būtų finansavęs „Tesla“ akcijomis. Vien laikotarpiu nuo 2015 iki 2023 metų „Tesla“ akcijų vertė šoktelėjo daugiau kaip 18 tūkst. procentų, tad „OpenAI“ būtų galėjusi be vargo gauti žadėtąjį milijardą. Nors Muskas dažnai reiškė susirūpinimą žmonijos ateitimi, iš tiesų jis labiau rūpinosi, kaip aplenkti konkurentus.

Pasitraukus Muskui, „OpenAI“ neteko pagrindinio finansavimo šaltinio. Altmanui tai buvo tikra katastrofa. Nuo šio projekto sėkmės priklausė jo reputacija. Žymiausi pasaulyje DI mokslininkai dirbo už mažesnį atlyginimą vien tam, kad galėtų bendradarbiauti su juo, tačiau grandioziniai pažadai padėti žmonijai vis labiau atrodė kaip tušti žodžiai.

Naujojoje DI eroje sėkmei pasiekti reikėjo gausybės išteklių – finansų tyrėjams samdyti, duomenų modeliams mokyti ir galingų kompiuterių jiems tobulinti. Be Musko paramos galimybės gauti šiuos išteklius greitai menko.

Altmanas atsidūrė kryžkelėje. Sėdėdamas „OpenAI“ biure San Fransiske, jis svarstė: ar tęsti veiklą su itin ribotais ištekliais ir kurti DI modelius, kurie, tikėtina, nusileis konkurentų sistemoms, ar „OpenAI“ istorijoje padėti tašką. Pritraukti lėšų ne pelno organizacijai buvo kur kas sunkiau nei startuoliui. Altmanui sunkiai sekėsi įtikinti turtingus žmones paremti jo BDI projektą vien iš geros valios, nesitikint jokios finansinės grąžos. Jam reikėjo dešimčių milijonų, o Muskas buvo paskutinis stambus rėmėjas.

Altmanas sugalvojo kitą sprendimą šiai problemai išspręsti. Jis nutarė potencialiems rėmėjams pasiūlyti ne tik galimą moralinį atlygį prisidėjus prie DI utopijos, bet ir tam tikrą tiesioginę finansinę naudą. Tokiu atveju abi pusės liktų patenkintos. Rėmėjai ne tiek „aukotų“, kiek „investuotų“ savo lėšas – būtent tokia leksika iš Altmano lūpų liejosi lengviausiai. Tačiau tokių išteklių, įskaitant finansus ir skaičiavimo galią, galėjo pasiūlyti vos kelios bendrovės: „Google“, „Amazon“, „Facebook“ ir „Microsoft“. Šios technologijų milžinės vienintelės galėjo pasigirti milijardais sąskaitose ir duomenų centrais, dydžiu prilygstančiais keliems futbolo stadionams.

Kelerius metus tiek „OpenAI“, tiek „DeepMind“ bandė įdiegti saugiklius, kurie užkirstų kelią netinkamam itin galingo DI naudojimui. „DeepMind“ mėgino keisti valdymo struktūrą, kad „Google“, kaip pelno siekianti monopolinė jėga, negalėtų nevaržoma naudoti BDI komerciniams tikslams. Tikėtasi, kad tinkamą DI priežiūrą užtikrins speciali ekspertų taryba. Tuo metu Altmanas su Musku, įsteigę „OpenAI“ kaip ne pelno organizaciją, žadėjo, kad, pasiekę proveržį DI kūrimo srityje, mokslinių darbų rezultatais ar net patentais pasidalins su kitomis

organizacijomis. Juk įgyvendindama savo misiją „OpenAI“ į pirmą vietą kelia žmonijos gerovę.

Tačiau dabar, kovodamas už organizacijos išlikimą, Altmanas buvo priverstas dalį tų saugiklių pašalinti. Atsargus požiūris, kurio laikėsi pradžioje, peraugo į rizikingesnę strategiją. Tai turėjo įtakos visai DI sričiai: gerai apgalvotus žingsnius ir akademinius sprendimus pakeitė „Laukinių Vakarų“ taisyklės. Atsitraukimą nuo „OpenAI“ pamatinių principų Altmanas pateisino pasitelkęs savo, kaip įtaigaus pasakotojo, gebėjimus. Galiausiai jam teko daryti tai, kas įprasta Silicio slėnio verslininkams, – pakeisti kryptį. Tereikėjo mažumėlę pakoreguoti kertinius „OpenAI“ principus.

3 Go – gilias tradicijas turintis loginis stalo žaidimas, paplitęs Kinijoje, Japonijoje ir Pietų Korėjoje (vert. past.).

II DALIS

Leviatanai

7 SKYRIUS

Žaidžiant žaidimus

Į pilką dangų besistiebiančių daugiaaukščių komplekse, įsikūrusiame vos už kelių žingsnių nuo Kings Kroso traukinių stoties, sutraukiančios minias turistų, norinčių išvysti stebuklingą platformą, per kurią Haris Poteris išvykdavo į Hogvartsą, skleidėsi visai kitokio pobūdžio magija. Tarp stiklu ir metalu žvilgančių pastatų besidriekiančioje jaukioje, šurmuliuojančioje pėsčiųjų alėjoje buvo galima sutikti ir vieną kitą „DeepMind“ inžinierių ar DI mokslininką, traukiantį iš kišenės įėjimo kortelę ir žengiantį pro stiklines „Google“ biuro duris. Nors pastatas priklausė „Google“, dviejuose jo aukštuose veikė slapta DI laboratorija.

„DeepMind“ darbuotojai, būdami „Google“ dalimi, mėgavosi įvairiomis privilegijomis – nuo miego kapsulių ir masažo kambarių iki personalui skirtos sporto salės, įrengtos biuro patalpose. Nepaisant to, įkūrėjai ir toliau stengėsi išsivaduoti iš patronuojančiosios bendrovės „Alphabet“ gniaužtų. Praėjus daugiau nei dvejiems metams po „DeepMind“ įsigijimo, technologijų milžinės vadovai Demisui Hassabisui, Mustafai Suleymanui ir Shane’ui Leggui pateikė naują pasiūlymą: užuot funkcionavusi kaip „nepriklausomas padalinys“, „DeepMind“ galėjo tapti „*Alphabet* įmone“, turinčia teisę atskirai rengti ir teikti pelno ir nuostolio ataskaitas.

Būdami Anglijoje, atokiau nuo Silicio slėnyje tvyrančio nuolatinio augimo etoso, įkūrėjai šį pasiūlymą priėmė rimtai. Suleymanas norėjo įrodyti, kad „DeepMind“ gali veikti kaip savarankiška įmonė, todėl užsibrėžė pademonstruoti praktinę jos kuriamų DI sistemų vertę. Jis įsteigė padalinį „Applied“, kurio mokslininkai, pasitelkdami skatinamojo mokymosi metodus, siekė spręsti sveikatos priežiūros, energetikos ir

robotikos sričių problemas, tikėdamiesi, kad tokie tyrimai atvers naujų verslo galimybių. Kita maždaug dvidešimties tyrėjų komanda „DeepMind for Google“ dirbo su projektais, teikiančiais tiesioginę naudą „Google“, – tobulino „YouTube“ rekomendacijų sistemas arba gerino „Google“ reklamos taikymo algoritmus. Pasak šaltinio, susipažinusio su sutartimis, „Google“ įsipareigojo pusę lėšų, sugeneruotų dėl tokių patobulinimų, skirti „DeepMind“. Anot vieno buvusio darbuotojo, apie du trečdaliai projektų buvo naudingi „Google“.

Tuo metu šimtai kitų „DeepMind“ tyrėjų galėjo laisvai tęsti darbus, susijusius su BDI kūrimu. Kas kelias savaites įkūrėjai susitikdavo viename Londono bare aptarti darbo reikalų. Jų pokalbiai dažnai pakrypdavo prie tų pačių opių klausimų. Suleymanas troško spręsti realias pasaulio problemas, bet kartu nuogąstavo, kad jie gali netyčia sukurti superintelektą, sukelsiantį skaudžių padarinių. „Kas, jei DI taps nevaldomas ir ims manipuliuoti žmonėmis?“ – klausė jis. Biure jis įspėdavo kolegas, kad dėl BDI poveikio ekonomikai milijonai žmonių gali netekti darbo, todėl gyventojų pajamų lygis smarkiai smuktų. „Juk tai gali baigtis sukilimu. Jei negalvosime apie lygybę, vieną dieną galime išvysti į Kings Krosą žygiuojančią minią su šakėmis“, – nuogąstaudavo Suleymanas, pasak vieno buvusio komandos nario.

Kartais Hassabisas, ieškodamas galimų problemos sprendimų, pasiūlydavo keistokų idėjų. Pavyzdžiui, jis iškėlė mintį, kad DI tapus galingesniam ir galimai pavojingam, bendrovė į pagalbą galėtų pasikviesti Kalifornijos universiteto profesorių Terence’ą Tao, laikomą vienu iškiliausių pasaulio matematikų. Pasak žurnalo „New Scientist“, šis matematikos genijus, jau vaikystėje garsėjęs kaip vunderkindas, o universitete pradėjęs mokytis vos devynerių, mokslininkų bendruomenėje buvo tituluojamas „visų problemų sprendėju“, galinčiu padėti į aklavietę patekusiems tyrėjams.

Tao interviu ne kartą minėjo, kad DI pagrindas iš esmės yra įmantri matematika, ir tikras DI greičiausiai niekada nebus sukurtas. Kaip ir

Hassabisas, į technologijas jis žvelgė mechaniškai, remdamasis „juoda arba balta“ principu. Jis manė, kad jei DI vieną dieną taptų nekontroliuojamas, jį suvaldyti padėtų matematika. Jis buvo ne vienintelis, laikęsis tokios nuomonės. Yudkowsky'io įkurtame interneto forume „LessWrong“ vienu metu vyko diskusija, kurios dalyviai svarstė, kaip įkalbinti tokį žmogų kaip Tao ir kitus talentingiausius matematikus prisidėti prie pastangų siekiant užtikrinti, kad DI būtų suderintas su žmonių vertybėmis ir nekeltų pavojaus. Jie manė, kad tokius matematikos genijus galėtų motyvuoti piniginė paskata, siekianti 5–10 mln. JAV dolerių.

Hassabisas planavo, kad, priartėjęs prie savo tikslo – BDI sukūrimo, jis liausis tobulinęs DI modelius ir pakvies geriausius pasaulio specialistus juos išanalizuoti iki smulkiausių detalių, kad padėtų rasti geriausius būdus jiems suvaldyti. „Galbūt šiuo tikslu reikėtų paskelbti visuotinį raginimą matematikams ir mokslininkams vienytis, panašiai kaip grėsmės akivaizdoje kviečiami burtis Keršytojai (angl. *Avengers*)“, – teigia Hassabisas.

Suleymanas nesutiko su bendražygio požiūriu, manydamas, kad jis pernelyg susitelkia į skaičius ir teoriją. Jis buvo įsitikinęs, kad DI gali būti kuriamas saugiai tik tuo atveju, jei jį valdo žmonės, o ne vien tik išmanūs matematiniai modeliai. Jiems diskutuojant apie geriausią DI suvaldymo strategiją, komandą pasiekė dar viena žinia iš „Google“ – vadovai pranešė atsisakantys planų „DeepMind“ paversti „*Alphabet* įmone“. Atsiskyrimo nuo „Google“ idėja tapo sunkiai įgyvendinama, nes DI ir jį kurianti „DeepMind“ tapo kaip niekada svarbūs „Google“ verslui.

Įkūrėjams ši situacija sukėlė pažįstamą *déjà vu*. Vadovai juos ramino, žadėdami ieškoti kompromiso. Galiausiai jie pasiūlė dar vieną išeitį: „DeepMind“ galėtų dalinai atsiskirti nuo korporacijos ir turėti nuosavą patikėtinių tarybą, prižiūrinčią dirbtinio superintelekto kūrimą, o „Alphabet“ DI bendrovės atžvilgiu išlaikytų dalines nuosavybės teises. Konglomerato vadovai, norėdami įrodyti, kad jų ketinimai rimti, savo

įsipareigojimą įformino raštu. Pasak šaltinio, susipažinusio su sutartimi, „Google“ vadovybė pasirašė ketinimų protokolą, kuriuo įsipareigojo per dešimties metų laikotarpį „DeepMind“ skirti 15 mlrd. JAV dolerių dotaciją, kurią įmonė galėtų naudoti veikdama savarankiškai. Bendraudamas su kolegomis, Hassabisas ne kartą pasidžiaugė, kad ketinimų protokolą pasirašė pats Sundaras Pichai'us, kuris po kelerių metų tapo „Alphabet“ generaliniu direktoriumi. Tai rodė, kad šįkart „Google“ į savo įsipareigojimą žiūrėjo rimtai.

Ketinimų protokolas – tai dokumentas, kuriame surašytos numatomos verslo sutarties sąlygos. Šis teisiškai neįpareigojantis dokumentas įprastai pakloja pamatą tolesnėms deryboms. Rašytinis susitarimas skambėjo solidžiau nei žodinis, todėl „DeepMind“ įkūrėjai tikėjo, kad šįkart „Google“ ketinimai išlaisvinti „DeepMind“ iš korporacinių pančių yra rimti. Įkūrėjai nusprendė pertvarkyti „DeepMind“ į kitokio tipo organizaciją, kuri, kaip ir „OpenAI“, savo struktūra labiau primintų filantropinę organizaciją nei pelno siekiančią bendrovę.

Hassabisas ir Suleymanas pasamdė investicijų bankininkus, kad šie padėtų suplanuoti bendrovės atsiskyrimo procesą. Taip pat kreipėsi į dvi Londono advokatų kontoras su prašymu parengti teisinius planus, kaip pertvarkyti „DeepMind“ į nepriklausomą organizaciją. Jie taip pat konsultavosi su Jungtinės Karalystės aukščiausio lygio verslo ginčų ekspertu, koordinavusiu sandorius tokioms didelėms bendrovėms kaip „Shell“, „Vodafone“ ir kasybos milžinė „BHP Billiton“.

Be to, buvo planuojama atnaujinti vadovybės struktūrą, įsteigiant kasdienę įmonės veiklą prižiūrinčią valdybą (angl. *operating board*). Į ją turėjo būti paskirtas Hassabisas, Leggas ir Suleymanas, taip pat „Alphabet“ generalinis direktorius Larry'is Page'as, „Google“ bendrakūrėjis Sergey'us Brinas, tuometinis „Google“ produktų vadovas Sundaras Pichai'us ir trys nepriklausomi komercijos direktoriai. Sprendimus buvo numatyta priimti dauguma balsų. Dar vienas svarbus žingsnis – visiškai

nepriklausomos patikėtinių tarybos įsteigimas. Ši šešių narių taryba turėjo užtikrinti, kad „DeepMind“ laikytųsi savo socialinės ir etinės misijos. Tarybos narių pavardės, kaip ir jų sprendimai, turėjo būti skelbiami viešai. Kadangi šie šeši asmenys būtų atsakingi už vieną galingiausių ir potencialiai pavojingiausių technologijų pasaulyje, jie turėjo būti aukščiausio kalibro, patikimi specialistai. Ieškodami galimų kandidatų, „DeepMind“ įkūrėjai nusitaikė į pačius aukščiausius sluoksnius: prie tarybos prisijungti buvo pakviestas buvęs JAV prezidentas Barackas Obama, buvęs JAV viceprezidentas ir buvęs CŽV vadovas. Anot gerai informuoto šaltinio, keli iš jų sutiko.

Pasikonsultavę su teisininkais, „DeepMind“ įkūrėjai nusprendė nežengti tuo pačiu keliu kaip Samas Altmanas, kuris rengėsi steigti ne pelno organizaciją. Jie sugalvojo įkurti visiškai naujos juridinės formos darinį – visuotinės svarbos bendrovę (angl. *global interest company*, GIC). Pagal jų sumanymą, „DeepMind“ turėjo tapti Jungtinių Tautų padalinį primenančia organizacija, kurios tikslas – skaidriai ir atsakingai valdyti DI, veikiant žmonijos labui. Ji suteiktų technologijų milžinei „Alphabet“ išskirtinę licenciją naudoti visas DI inovacijas, sukurtas siekiant pagerinti „Google“ paieškos internete paslaugas. Tačiau didžiąją dalį turimų finansinių, žmogiškųjų išteklių ir mokslinių pajėgumų bendrovė sutelktų socialinei misijai – vaistams kurti, sveikatos priežiūros problemoms spręsti ar kovai su klimato kaita. Bendrovės viduje šis projektas buvo vadinamas „VSB“ (GIC).

Net siekdama atsiskirti, „DeepMind“ prisidėjo prie „Google“ veiklos stiprinimo. Maždaug tuo metu, kai Larry’is Page’as „DeepMind“ įkūrėjus apipylė pažadais padėti išsivaduoti iš konglomerato kontrolės, „Google“ ėmė dairytis naujų verslo plėtros galimybių. JAV ir kitose Vakarų rinkose sulėtėjus technologijų milžinės augimo tempams, jos žvilgsnis nukrypo į daugiausia gyventojų pasaulyje turinčią šalį. Tuo metu Kinijoje buvo daugiau nei 650 milijonų interneto naudotojų – kone dvigubai daugiau nei Jungtinių Amerikos Valstijų gyventojų. Be to, tik maždaug pusė kinų,

turinčių galimybę naudotis internetu, juo iš tiesų naudojosi, todėl šalis atrodė kaip milžiniška neišnaudotų galimybių rinka. Augant vidurinei klasei, didėjant vartojimo išlaidoms, o bendrajam vidaus produktui pasiekus apie 11 trilijonų JAV dolerių, Kinija tapo antra pagal dydį pasaulio ekonomika. Tai buvo potenciali aukso gysla bet kuriai interneto paslaugų bendrovei.

Tačiau „Google“ negalėjo taip paprastai įžengti į Kiniją. Dar 2010 metais bendrovė pasitraukė iš šalies, kaltindama Pekiną intelektinės nuosavybės vagystėmis ir įsilaužimu į Kinijos žmogaus teisių aktyvistų „Gmail“ paskyras. Tuo metu šalies valdžia, valdančiosios Kinijos komunistų partijos prašymu, pareikalavo, kad „Google“ cenzūruotų informaciją apie Tiananmeno aikštės įvykius bei kitas prieštaringas temas. Vėliau šalis užblokavo prieigą prie socialinių tinklų „Facebook“ ir „Twitter“ – ši politika tapo žinoma kaip „Didžioji ugniasienė“. „Google“ vadovai buvo įsitikinę, jog toks sprendimas laikinas ir netrukus kinai patys ims reikalauti valdžios leisti naudotis pažangiomis Silicio slėnio gigantų paslaugomis.

„Manau, tokia režimo politika ilgai nesitęs“, – teigė tuometinis „Google“ valdybos pirmininkas Ericas Schmidtas, 2012 metais duodamas interviu žurnalui „Foreign Policy“.

Schmidtas klydo. Užuot stagnavęs, Kinijos interneto paslaugų sektorius suklestėjo. Rinkoje iškilo tokios technologijų milžinės kaip „Meituan“, „Baidu“ ir „Alibaba“. Kelią į sėkmę joms nutiesė kinų inžinieriai, kurie, įgiję patirties Silicio slėnyje, grįžo į gimtąją šalį kurti savųjų technologijų leviatanų. Nemažai inžinierių, anksčiau dirbusių „Microsoft Research Asia“, pradėjo eiti vadovaujamas pareigas tokiose interneto paslaugų gigantėse kaip „Alibaba“ ir „Tencent“. Per penkerius metus nuo „Google“ pasitraukimo Kinijos rinka dar labiau suklestėjo. Bendrovė neturėjo aiškios strategijos, kaip vėl įžengti į Kiniją, kuri taip ir neatsisakė cenzūros politikos. Vis dėlto „Google“ siekė pasinaudoti augančios Kinijos vartotojų rinkos galimybėmis ir technologinėmis

naujovėmis, kurios šalyje sulaukė didelės sėkmės. „Turime suprasti, kas ten vyksta, kad tai mus įkvėptų, – naujienų organizacijai „The Intercept“ teigė „Google“ paieškos paslaugų vadovas Benas Gomesas. – Galime iš Kinijos daug ko pasimokyti.“

Kaip tik tuo metu įvyko reikšmingas „Google“ vadovybės pasikeitimas. 2015-aisiais Page’as ir Brinas atsitraukė nuo kasdienio savo įkurtos bendrovės valdymo, kad galėtų daugiau dėmesio skirti širdžiai mielai veiklai – filantropijai bei įvairiems projektams, susijusiems su skraidančių automobilių kūrimu ar kosmoso tyrinėjimu. Naujuoju „Google“ generaliniu direktoriumi tapo Sundaras Pichai’us. „DeepMind“ įkūrėjai ketino kolegų pagarbą pelniusį Pichai’ų, anksčiau ėjusį produktų vadovo pareigas, po atsiskyrimo nuo „Google“ paskirti į kasdienę įmonės veiklą prižiūrinčią valdybą. Tačiau, skirtingai nei Page’as, jis neturėjo nei laiko, nei, tikėtina, didelio noro padėti vienai vertingiausių „Google“ bendrovių atsiskirti nuo šios korporacijos. Juodu su Schmidtu visą dėmesį skyrė naujų būdų, kaip vėl įžengti į Kiniją, paieškai. Vienu metu atrodė, kad Pekinas leis „Google“ programėlių parduotuvei vėl pradėti veikti šalyje, tačiau šios viltys greitai žlugo.

Tada „DeepMind“ įkūrėjams pasitaikė proga, pasinaudojus viešaisiais ryšiais, plačiau paskleisti žinią apie save. „DeepMind“ savo kuriamiems DI modeliams mokyti naudojo žaidimus. Viena iš naujausių bendrovės sukurtų programų „AlphaGo“ buvo išmokyta žaisti go – strateginį stalo žaidimą dviem žaidėjams, kuriam reikalingi abstraktaus mąstymo gebėjimai. Šis žaidimas, atsiradęs Kinijoje prieš daugiau kaip 2 500 metų, iš pirmo žvilgsnio gali pasirodyti gana paprastas. Jis žaidžiamas ant lentos, padalytos į devyniolika vertikalių ir devyniolika horizontalių linijų. Vienas žaidėjas žaidžia juodais, kitas – baltais akmenukais. Jie paeiliui atlieka ėjimus, dėdami savo akmenukus į laisvą vietą linijų susikirtimo taškuose. Laimi žaidėjas, savo akmenukais užėmęs didesnį plotą ir apsupęs daugiau varžovo akmenukų. Tai vienas sudėtingiausių strateginių žaidimų pasaulyje. Galimų žaidimo pozicijų skaičius siekia

10^{170} – tai gerokai viršija visoje stebimojoje Visatoje esančių atomų skaičių, kuris apytiksliai siekia 10^{80}.

Page'as ir jo bendražygis Sergey'us Brinas go žaidimą buvo pamėgę dar studijų Stanforde metais, kai jiedu jau buvo pradėję kurti „Google". Praėjus kelioms savaitėms po „DeepMind" įsigijimo, Page'as apie savo pomėgį prasitarė Hassabisui. Šis atsakė, kad jo komanda galėtų sukurti DI sistemą, pajėgią įveikti realaus pasaulio čempioną.

Hassabisas tokiu pareiškimu siekė kur kas daugiau nei nustebinti naująjį vadovą. Jis buvo ne tik talentingas mokslininkas, bet ir sumanus rinkodaros strategas, tad suprato, kad jei „AlphaGo" įveiktų pasaulinio lygio go čempioną, panašiai kaip 1997-aisiais IBM kompiuteris „Deep Blue" nugalėjo Garry'į Kasparovą, tai žymėtų svarbų proveržį DI srityje ir išgarsintų „DeepMind" kaip šios srities lyderę. Taigi „DeepMind" metė iššūkį Pietų Korėjos čempionui Lee Sedolui, pasiūliusi penkių partijų dvikovą prieš „DeepMind" kompiuterį.

2016-ųjų kovą vykusią žmogaus ir mašinos akistatą per televiziją ir internetu stebėjo daugiau kaip 200 milijonų žiūrovų. Likus kelioms valandoms iki dvikovos, programą valdęs „DeepMind" mokslininkas nustojo gerti skysčius, kad per mačą neprireiktų bėgti į tualetą. Žaidimo metu Hassabisas neramiai vaikščiojo pirmyn atgal tarp „AlphaGo" valdymo posto ir uždaros stebėjimo patalpos. Jis negalėjo nė kąsnio praryti. Tobulindama „AlphaGo" neuroninį tinklą, komanda išmokė jį trisdešimt milijonų galimų ėjimų.

Kad laimėtų go partiją, žaidėjas turi visiškai apsupti varžovo akmenukus. Tam reikia apgalvotos strategijos: derinti puolimą su gynyba, iš anksto nustatyti trumpalaikius ir ilgalaikius tikslus bei numatyti galimą oponento ėjimų seką. Būtina gerai apgalvoti, kuriuose linijų susikirtimo taškuose dėti akmenukus. Žaidėjai retai deda akmenis ant kraštinių linijų, nes jos nesuteikia daug galimybių apsupti varžovą ir užimti teritoriją. Per antrą partiją, atlikdama trisdešimt septintąjį žaidimo ėjimą, „AlphaGo" priėmė visus nustebinusį sprendimą – padėjo

akmenuką ant penktos linijos nuo dešinės. Paprastai laikoma, kad dėti akmenuką ant penktos linijos nėra efektyvus ėjimas, nes taip priešininkui suteikiamas teritorinis pranašumas ties ketvirta linija. „AlphaGo“ sprendimas buvo toks netikėtas, kad prieš atlikdamas kitą ėjimą Lee svarstė geras penkiolika minučių ir net trumpam buvo palikęs patalpą.

„Tai labai netikėtas ėjimas“, – sakė vienas komentatorius, manydamas, kad „AlphaGo“ operatorius netyčia paspaudė klaidingą langelį.

Tačiau po maždaug šimto ėjimų iš pažiūros keista „AlphaGo“ strategija ėmė įgauti prasmę. Du „AlphaGo“ juodi akmenukai lentos apačioje, kairėje, netikėtai susijungė su penktoje linijoje padėtu akmenuku. Po dar keturių valandų Lee pasidavė. Vėliau komentatoriai ir pats Lee trisdešimt septintąjį ėjimą vadino „nuostabiu“. Hassabisas teigė, kad tai atspindėjo DI kūrybiškumo užuomazgas. Galutiniame rezultate „AlphaGo“ prieš Lee laimėjo keturias partijas iš penkių.

Tai buvo svarbus žingsnis DI kūrimo srityje, atnešęs „DeepMind“ neregėtą žiniasklaidos dėmesį. Šis pasiekimas netgi įamžintas apdovanojimų pelniusiame „Netflix“ dokumentiniame filme „AlphaGo“. Hassabisas buvo pasirengęs tokia pakilia nata užbaigti darbą su šia programa ir imtis kito projekto.

Tačiau „Google“ turėjo kitokių planų. „AlphaGo“ atrodė kaip puiki priemonė pademonstruoti Pekinui aukšto lygio technologijas, kurios atvertų kelią į Kinijos rinką. „Google“ vadovai tikėjosi, kad „AlphaGo“ galėtų turėti tokį patį efektą kaip ir „stalo teniso diplomatija“ – 1971-aisiais surengti JAV ir Kinijos stalo tenisininkų mainai, padėję pagerinti dvišalius santykius Šaltojo karo metais. Jei dvikova Pietų Korėjoje buvo „DeepMind“ viešųjų ryšių triukas, tai reginys Kinijoje galėjo tapti „Google“ strategiškai naudingu ėjimu.

„Google“ siekė, kad „DeepMind“ išbandytų „AlphaGo“ prieš dar pajėgesnį varžovą – devyniolikmetį kiną Ke Jie, kuris tuo metu užėmė

pirmą vietą pasaulio go žaidėjų reitinge. Ke Jie buvo visiškai kitoks nei Lee Sedolas – jis nevengdavo provokuoti savo varžovų ir mėgo puikuotis. „Google“ irgi netrūko pasitikėjimo savimi – bendrovė tikėjo, kad, pademonstravusi savo technologijas, sugebės vėl įsitvirtinti Kinijoje.

Anot buvusių „DeepMind“ darbuotojų, ši situacija Hassabisui kėlė nerimą. Jei „AlphaGo“ laimėtų, atrodytų, kad viską triuškinantis DI dar kartą įveikė žmogų. Jei pralaimėtų, Korėjoje kilusi entuziazmo banga akimirksniu nuslūgtų. Atrodė, kad pralaimėjimo išvengti nepavyks nei vienu, nei kitu atveju.

Žinodamas, kad „Google“ bet kokia kaina siekia įsitvirtinti Kinijos rinkoje, Hassabisas, pasitelkęs strateginį mąstymą, pasiūlė Pichai’ui kompromisą: surengti dar vieną dvikovą, šįkart naudojant naują „AlphaGo“ versiją, pavadintą „AlphaGo Master“. Skirtingai nei ankstesnė sistema, kurios veikimui reikėjo šimtų skirtingų kompiuterių, „AlphaGo Master“ turėjo pakakti vieno kompiuterio su „Google“ lustu. Taip dvikovą buvo galima pateikti kaip naujos DI sistemos bandymą, o ne dar vieną mėginimą sutriuškinti realaus pasaulio čempioną. Jei sistema pralaimėtų, būtų galima pasiteisinti, kad naujoji versija neprilygsta originaliajai „AlphaGo“, o jei laimėtų – paskelbti apie dar galingesnę sistemą. „Google“ taip pat turėtų progą pristatyti savo naują mašininio mokymosi platformą „TensorFlow“ ir pritraukti stambių verslo klientų Kinijoje, kurie galėtų naudotis jos nedidele, bet sparčiai augančia debesijos paslaugų infrastruktūra. Pichai’us sutiko.

Dvikova įvyko 2017-ųjų gegužę Udžene, Kinijoje. Nors „Google“ vadovai visus metus bandė įtikinti Kinijos valdžios institucijas mačą rodyti visoje šalyje – tiek per televiziją, tiek per interneto platformas, galiausiai beveik visiems gyventojams prieiga prie transliacijos buvo užblokuota. Dvikovoje su Ke Jie „AlphaGo“ laimėjo visas tris partijas, tačiau Kinijoje apie tai beveik niekas nesužinojo.

„Google“ vadovybė stengėsi situaciją vertinti optimistiškai. Renginio metu kalbėdamas ant scenos, Schmidtas pasinaudojo proga

pareklamuoti „TensorFlow“ platformą, teigdamas, kad didžiosios Kinijos interneto paslaugų bendrovės, tokios kaip „Alibaba“, „Baidu“ ir „Tencent“, turėtų ją išbandyti. „Visoms joms labai praverstų „TensorFlow“, – sakė jis. „Google“ taip troško sugrįžti į Kinijos rinką, kad ėmė švelninti ankstesnę griežtą poziciją dėl Pekino reikalavimų cenzūrai ir net asmens duomenų stebėsenai. 2018 metais naujienų organizacijai „The Intercept“ nutekintame pranešime teigiama, kad „Google“ vadovai nurodė inžinieriams kurti Kinijai skirtos paieškos sistemos prototipą kodiniu pavadinimu „Laumžirgis“ (angl. *Dragonfly*). Turėjo būti numatyta į juodąjį sąrašą įtraukti tam tikrus paieškos žodžius ir susieti žmonių paieškas su jų mobiliojo telefono numeriais. Taip bendrovė pamynė savo principus, padėdama priespaudą taikančiam režimui sekti savo piliečius.

„Google“ buvo taip apakinta augimo siekio, kad neįžvelgė, jog bandymas vėl patekti į Kinijos rinką pasmerktas žlugti. Šalies technologijų bendrovės sparčiai žengė į priekį DI srityje. Joms iš tiesų nereikėjo nei „TensorFlow“, nei pačios „Google“. Be to, metais anksčiau Kinijos interneto paslaugų milžinei „Baidu“ pavyko persivilioti Stanfordo profesorių Andrew Ng’ą, inicijavusį projektą „Google Brain“. Kinijos vyriausybė nusprendė, kad jos piliečiai ir augantis šalies technologijų sektorius gali apsieiti be paieškos milžinės paslaugų.

Praėjus vos dviem mėnesiams po dvikovos su Ke Jie, Pekinas paskelbė naują ilgalaikį tikslą: iki 2030-ųjų aplenkti JAV ir tapti pirmaujančia DI supervalstybe. Vyriausybė planavo finansuoti įvairius DI startuolius ir ambicingus projektus – atrodė, kad šalis tam tikra prasme įgyvendina savąją „Apollo“ programą. Apie bendradarbiavimą su „Google“ ar kitomis Silicio slėnio bendrovėmis užsiminta nebuvo.

„Google“ vadovams greitai tapo aišku, kad jų svajonės įžengti į milžinišką Kinijos interneto paslaugų rinką ir joje pelningai veikti yra nerealistiškos. Tai buvo didžiulis nusivylimas bendrovei. Hassabisas taip pat atsidūrė keblioje padėtyje. „DeepMind“, pademonstravusi pažangios DI sistemos „AlphaGo“ galimybes ir sukėlusi didžiulę dėmesio bangą, tik

dar labiau įrodė savo vertę „Alphabet“. Vis dėlto Hassabisas vis dar buvo nusiteikęs įgyvendinti tai, ką kartu su Suleymanu taip kruopščiai planavo, – atsiskirti nuo korporacijos.

Hassabisas taip tvirtai tikėjo plano sėkme, kad 2017 metų gegužę, praėjus vos kelioms savaitėms po mačo Kinijoje, surengė „DeepMind“ darbuotojams išvyką į atokią Škotijos vietovę, kur kartu su Suleymanu visai komandai – daugiau kaip trims šimtams narių – pristatė bendrovės atsiskyrimo planą. Renginyje, kuriam specialiai buvo išnuomotas viešbučio ir konferencijų centro kompleksas, įkūrėjai pranešė apie planus pertvarkyti „DeepMind“ į visuotinės svarbos bendrovę (VSB). Jie paskelbė, kad „DeepMind“ taps ne pelno organizacija, kurios dalį akcijų valdys „Google“, ir veiks kaip kitos viešojo intereso organizacijos, pavyzdžiui, Jungtinės Tautos ar Gatesų fondas. Įkūrėjai paaiškino, kad „DeepMind“ siekia tapti žmonijos labui dirbančia organizacija ir užtikrinti, kad DI būtų kuriamas bei vystomas taip, jog neštų naudą visam pasauliui. Užuot veikusi vien tik „Google“ komerciniais interesais, „DeepMind“ su ja pasirašytų išimtinės licencijos sutartį ir drauge vykdytų tikrąją savo misiją – spręstų didžiausias pasaulio problemas.

Anot susitikime dalyvavusių asmenų, darbuotojai buvo nepaprastai sužavėti šios idėjos. DI tyrėjai galėjo derinti tai, ką geriausio siūlo mokslo ir verslo pasaulis: dirbti technologijų bendrovėje, kuri moka puikų atlyginimą ir teikia kitų privalumų, ir drauge spręsti „DI klausimą, o vėliau – ir visas kitas pasaulio problemas“. Įkūrėjai teigė, kad atsiskyrimas bus užbaigtas iki 2017 metų rugsėjo.

Hassabisas ir Suleymanas darbuotojams nurodė VSB projektą laikyti paslaptyje. Tokia konfidencialumo sąlyga nebuvo neįprasta – daugelis „DeepMind“ komandos narių buvo pasirašę griežtas sutartis, draudžiančias kalbėti apie įmonės planus ar technologijas. Tačiau šįkart planų apie numatomą atsiskyrimą jie negalėjo aptarinėti net bendrovės viduje. Kartais jų buvo prašoma projektą vadinti kodiniu pavadinimu, pavyzdžiui, „watermelon“ (arbūzas), arba bendrauti tik per šifruotų

žinučių programėles, tokias kaip „Signal“. Pasak buvusių darbuotojų, kai kurie „DeepMind“ vadovai patarė neaptarinėti šios temos „Google“ platformose, pavyzdžiui, el. pašto sistemoje „Gmail“.

„DeepMind“ tyrėjai manė, jog toks konfidencialumas susijęs su nerimu, kad jų sukurta technologija gali būti panaudota netinkamiems tikslams. Vėliau „Google“ įsitraukus į karinį projektą, šie įtarimai tik sustiprėjo. 2017 metais JAV gynybos departamentas ėmėsi įgyvendinti vadinamąjį „Project Maven“, kuriuo buvo siekiama DI ir mašininį mokymąsi plačiau pritaikyti gynybos srityje, pavyzdžiui, aprūpinti karinius dronus vaizdo atpažinimo sistema, kad jie tiksliau nustatytų taikinius. Kaip teigiama naujienų organizacijai „The Intercept“ nutekintuose elektroniniuose laiškuose, „Google“ tikėjosi iš šios partnerystės uždirbti 250 mln. JAV dolerių per metus. Tačiau kilus masinėms vidaus protesto akcijoms, „Google“ buvo priversta projektą nutraukti ir atsisakyti pratęsti sutartį su Gynybos departamentu. Tai dar labiau sustiprino „DeepMind“ komandos narių susirūpinimą, kad jų kuriamas DI gali būti panaudotas blogiems tikslams.

Tuo metu bendrovės atsiskyrimo procesas vis dar vyko vangiai. Hassabisas ir kiti vadovai darbuotojams nuolat kartojo, kad atsiskyrimas įvyks „po šešių mėnesių“, tačiau sukakus šiam terminui, taip nieko ir neįvykdavo. Bendrovės inžinieriai ėmė abejoti, ar šis planas apskritai kada nors bus įgyvendintas. Be to, pati atsiskyrimo idėja atrodė miglota. Pavyzdžiui, kai Suleymanas pareiškė norintis, kad naujosios bendradarbiavimo su „Google“ taisyklės būtų teisiškai įpareigojančios, nei jis, nei kiti vadovai negalėjo paaiškinti, kaip tai įgyvendinti praktiškai. Tarkime, jeigu „Google“ vėliau nuspręstų „DeepMind“ sukurtą DI panaudoti kariniams tikslams, ar „DeepMind“ galėtų paduoti bendrovę į teismą? Aiškaus atsakymo nebuvo. „DeepMind“ darbuotojų prašyta parengti gaires, draudžiančias jų kuriamą DI naudoti žmogaus teisių pažeidimams ar apskritai „blogiems tikslams“. Tačiau niekas negalėjo atsakyti, ką tie „blogi tikslai“ tiksliai reiškia.

Iš dalies problema buvo ta, kad „DeepMind“ nepakako žmonių, galinčių spręsti šiuos klausimus. Bendrovė aktyviai samdė mokslininkus ir programuotojus, kad šie tobulintų DI modelių veikimą, tačiau tik saujelė darbuotojų tyrinėjo etinius DI kūrimo principus. Pavyzdžiui, 2020 metais dauguma „DeepMind“ darbuotojų (jų skaičius siekė beveik tūkstantį) buvo tyrėjai ir inžinieriai, mažiau nei tuzinas jų nagrinėjo etikos klausimus, ir tik du, turėję daktaro laipsnį, vykdė akademinio lygio tyrimus šioje srityje. Beveik niekas neanalizavo, kaip DI sistemos gali skatinti šališkumą, rasizmą ar prisidėti prie žmogaus teisių pažeidimų. „Negalime teigti, kad turime etikos komandą, jei iš esmės kalbame tik apie du žmones“, – sakė vienas „DeepMind“ darbuotojas.

Dirbtinio intelekto srityje sąvokos „etika“ ir „saugumas“ dažnai reiškia skirtingus tyrimų tikslus. Pastaraisiais metais šių krypčių šalininkų požiūriai dažnai susikerta. Tyrėjai, susitelkiantys į DI saugumą, paprastai remiasi tomis pačiomis idėjomis, kurias propaguoja Yudkowsky’is ar Jaanas Tallinnas. Jie siekia užtikrinti, kad dirbtinis superintelektas ateityje nesukeltų katastrofinės žalos žmonijai – pavyzdžiui, pasitelkęs vaistų kūrimo algoritmus, nesukurtų cheminių ginklų ar nepaskleistų dezinformacijos, galinčios visiškai destabilizuoti visuomenę.

Etikos tyrimai, priešingai, labiau orientuoti į tai, kaip DI sistemos kuriamos ir naudojamos šiandien. Šios srities tyrėjai nagrinėja, kaip technologijos jau dabar gali kenkti žmonėms. Tai ypač svarbu, turint omenyje, kad „Google“ nuotraukų programėlės algoritmas, klaidingai priskyręs juodaodžius asmenis goriloms, nėra vienintelis nesėkmingo DI modelių veikimo pavyzdys. Šališkumas – milžiniška problema DI srityje. Pavyzdžiui, algoritmai, naudojami JAV baudžiamosios teisenos sistemoje, neproporcingai dažnai klaidingai nurodydavo, kad juodaodžiai asmenys labiau linkę pakartotinai nusikalsti. Be to, kai kurie kūrėjai DI įrankius naudojo etiškai nepateisinamiems tikslams. Štai Stanfordo universiteto

tyrėjai sukūrė veido atpažinimo sistemą, kuri, jų teigimu, gali nustatyti žmonių lytinę orientaciją.

Visų šių sistemų sumanytojai turėjo daugiau dėmesio skirti teisingumo, skaidrumo ir žmogaus teisių aspektams. Be to, tokie klausimai dažniausiai tiesiogiai nepaliečia pačių DI bendrovių vadovų, kurių gretose dominuoja baltieji vyrai. Dirbtinio intelekto sistemoms šiandien suklydus, dažniausiai nukenčia ne baltųjų rasės atstovai, moterys ir kiti mažumų atstovai.

Glumina tai, kad dar 2017-aisiais „DeepMind“ viešai kalbėjo apie etikos svarbą vykdant bendrovės misiją – „spręsti DI problemą, siekiant mokslo pažangos ir žmonijos gerovės“. Bendrovės vadovai žurnalui „Wired“ netgi teigė, kad nedidelė jos etikos tyrėjų komanda per metus išaugs iki dvidešimt penkių darbuotojų.

Iš tiesų komanda išsiplėtė tik iki penkiolikos, nes, kaip teigė vienas buvęs vadovas, „DeepMind“ vadovybė buvo pernelyg susitelkusi į atsiskyrimą nuo „Google“. „Jie nuolat kalba apie etikos klausimus, bet iš tikrųjų jais rūpinasi tik nedidelė grupė [etikos tyrėjų], kuri vos susitvarko, – sakė kitas darbuotojas, pridūręs, kad etikos komanda neturi pagalbininkų ir yra nepakankamai aprūpinta reikalingais ištekliais. – Tai absurdiška. Juk kalbame apie [daugiamilijardinę] korporaciją.“

Jei „DeepMind“ savo deklaruojamų etikos principų nepagrindė konkrečiais darbais, kodėl įkūrėjai taip veržėsi atsiskirti nuo „Google“? Ar jie iš tiesų siekė užtikrinti, kad jų kuriama technologija nebūtų panaudota blogiems tikslams, ar vis dėlto vadovavosi asmeniniais interesais, norėdami išlaikyti kontrolę? Pagal planuotas atskyrimo sąlygas, „DeepMind“ turėjo pasirašyti išimtinės licencijos sutartį su „Google“, tačiau įkūrėjai nesugebėjo nustatyti aiškių DI naudojimo ginklų pramonėje apribojimų ar juos teisiškai įtvirtinti. Nors jų užmojai buvo dideli, stigo aiškaus veiksmų plano. Kai kurie darbuotojai spėliojo, ar Hassabisas, Suleymanas ir Leggas naiviai tikėjo galintys toliau finansuoti BDI kūrimą „Google“ lėšomis, bet kartu išsaugoti teisę

kontroliuoti šią technologiją, tarsi tikėdamiesi, kad ir vilkas bus sotus, ir avis sveika.

Kaip ir pati „Google“, veiklos pradžioje pasirinkusi principą „Nedaryk blogo“, taip ir „DeepMind“ steigėjai savo veiklą šios korporacijos sudėtyje pradėjo turėdami kilnių ketinimų. Jie net atsisakė 150 mln. JAV dolerių didesnio pasiūlymo iš „Facebook“, nes ši bendrovė nesutiko steigti etikos tarybos. Tačiau po kelerių metų prioritetai, regis, pasikeitė – etika ir saugumas užleido vietą našumui ir prestižui. Įkūrėjai neturėjo aiškaus plano, kaip suvaldyti BDI, išskyrus idėją pakviesti žymiausius matematikus, tokius kaip Terence’as Tao, o ir atsakymų, kaip užkirsti kelią netinkamam šios technologijos naudojimui, jie nepateikė.

Tad iškilo dar svarbesnis klausimas: ar apskritai įmanoma rūpintis etiniais DI klausimais, būnant stambios korporacijos dalimi? Atsakymas, atėjęs tiesiai iš „Google“, buvo vienareikšmis – ne.

8 SKYRIUS

Viskas tiesiog puiku

Norint suprasti, kodėl „Google“ tapo taip sunku kurti etiškas DI sistemas ar net paversti novatoriškas idėjas realiais produktais, reikia atidžiau pažvelgti į konkrečius skaičius. Šios knygos rašymo metu „Google“ patronuojančiosios bendrovės „Alphabet Inc.“ rinkos kapitalizacija siekė 1,8 trilijono JAV dolerių. 2020 metais „Apple“ tapo pirmąja JAV biržine bendrove, kurios rinkos vertė pasiekė 2 trilijonus JAV dolerių. Kitos technologijų milžinės – „Amazon“ ir „Microsoft“ – dabar vertinamos atitinkamai 1,7 ir 3 trilijonais JAV dolerių. 2018-aisiais „Apple“ vertė, peržengusi trilijono dolerių ribą, pasiekė tokias aukštumas, kokių iki tol nebuvo regėjusi jokia bendrovė. Vertingiausias pasaulio bendroves vienija vienas bruožas: visos jos priklauso technologijų sektoriui. Iš tiesų bendrovės, kurias paprastai laikome milžiniškomis, dažnai yra keturis kartus mažesnės nei Silicio slėnio gigantai. Pavyzdžiui, naftos milžinė „Exxon Mobil“ vertinama tik 450 mlrd. JAV dolerių, o „Walmart“ – 435 mlrd. JAV dolerių. Visų didžiųjų technologijų bendrovių, kartu paimtų, rinkos vertė pranoksta daugumos pasaulio šalių (išskyrus JAV ir Kiniją) bendrąjį vidaus produktą.

Žvelgiant iš istorinės perspektyvos, matyti, kad bendrovės, praeityje laikytos milžinėmis, visiškai nublanksta prieš šiandienos gigantus. Pavyzdžiui, 1984 metais telekomunikacijų bendrovės „AT&T“ rinkos vertė jos klestėjimo laikotarpiu prieš padalijimą siekė maždaug 60 mlrd. JAV dolerių, arba apie 150 mlrd. JAV dolerių, vertinant šiandieniniais pinigais. „General Electric“ aukščiausia rinkos vertė 2000-aisiais siekė apie 600 mlrd. JAV dolerių.

Pagal užimamą rinkos dalį technologijų milžinėms taip pat niekas neprilygsta. Iki 1911 metų, kai reguliavimo institucijos priėmė sprendimą išskaidyti „Standard Oil“ į atskiras įmones, šis naftos gigantas užėmė 90 proc. JAV naftos rinkos. Šiandien „Google“ valdo apie 92 proc. pasaulinės paieškos internete paslaugų rinkos. Kiekvieną dieną maždaug milijardas žmonių ieško informacijos „Google“, daugiau nei du milijardai kasdien prisijungia prie „Facebook“, o „iPhone“ naudotojų skaičius visame pasaulyje siekia apie 1,5 mlrd. Jokia valstybė ar imperija istorijoje nėra vienu metu palietusi tiek daug žmonių gyvenimų.

Nuo to laiko, kai sprogo interneto įmonių burbulas, minėtoms bendrovėms pasiekti tokį mastą prireikė gerų dviejų dešimtmečių. Kaip joms pavyko tiek išaugti? Jos įsigijo tokias bendroves kaip „DeepMind“, „YouTube“ bei „Instagram“ ir sukaupė milžinišką kiekį vartotojų duomenų, leidžiančių rodyti specialiai jiems pritaikytas reklamas ir rekomendacijas, galinčias dideliu mastu paveikti žmonių elgesį. „Google“ kaupia duomenis apie vartotojus analizuodama jų paieškos užklausas ir veiklą „YouTube“ platformoje, o elektroninės prekybos milžinė „Amazon“ renka informaciją apie tai, ką vartotojai perka ir kaip jie naršo internete. Bendrovių kaupiamų duomenų mastas išties atima žadą – jie apima asmens duomenis, naršymo istoriją, buvimo vietos informaciją, o kai kuriais atvejais net balso įrašus. Tokie duomenys stebina ne tik gausa, bet ir įvairove, be to, suteikia technologijų įmonėms detalių įžvalgų apie vartotojų elgseną.

Tokios bendrovės kaip „Facebook“ ir „Google“ naudoja šiuos duomenis tam, kad galėtų vartotojams siūlyti tikslinę reklamą – rodyti reklaminį turinį, kuris sužadina vartotojo susidomėjimą, ir pasitelkti sudėtingus rekomendacijų algoritmus. Tokios programinės sistemos leidžia generuoti naujienų srautą, kurį žmonės kasdien peržiūri braukdami per ekraną, ir užtikrinti, kad rodomas turinys skatintų juos be paliovos naršyti. Bendrovės suinteresuotos, kad vartotojai taptų kuo labiau priklausomi nuo jų platformų, nes tai garantuoja didesnes

reklamos pajamas. Tačiau tai neigiamai atsiliepia patiems vartotojams. 2023 metais atlikta apklausa parodė, kad amerikiečiai taip įnikę į socialinius tinklus, tokius kaip „Facebook“ ir „Instagram“, kad savo telefonus tikrina vidutiniškai 144 kartus per dieną.

Tiesa ta, kad labiausiai žmones įtraukia pasipiktinimą kurstantis turinys, todėl toks asmeniškai pritaikytos informacijos pateikimas didina atotrūkį tarp skirtingų kartų ir lemia politinį susiskaldymą, paveikiantį milijonus žmonių. Pavyzdžiui, per 2016 metų JAV prezidento rinkimus „Facebook“ naujienų sraute vartotojams dažnai pateikė labiausiai provokuojantį politinį turinį. Įvairios žinios ir nuomonės, pasiekusios platų vartotojų ratą, dar labiau sustiprino jų išankstines nuostatas ir suformavo informacinius burbulus. Panašus reiškinys paskatino neigiamą gyventojų požiūrį į imigraciją Didžiojoje Britanijoje likus keliems mėnesiams iki referendumo dėl „Brexit“, o 2017 metais tai prisidėjo prie smurto proveržių prieš rohinjus Mianmare. „Facebook“ algoritmai smarkiai paspartino neapykantą prieš rohinjus kurstančio turinio sklaidą ir paskatino Mianmaro kariuomenę griebtis genocido veiksmų prieš šią etninę musulmonų mažumą. „Amnesty International“ ataskaitoje teigiama, kad tūkstančiai žmonių patyrė kankinimus, prievartavimus, buvo nužudyti arba priverstinai iškeldinti. Vėliau „Facebook“ spaudoje pripažino, jog nepadarė pakankamai, kad užkirstų kelią smurto kurstymui prieš rohinjų tautą.

Nepaisant pasėtos nesantaikos, „Facebook“ verslo modelis pasiteisino su kaupu. Jo esmė – milijardai vartotojų ir jų duomenys laikomi produktu, o reklamuotojai – tikraisiais klientais. Kuo daugiau duomenų bendrovė surenka iš vartotojų, tuo daugiau gali uždirbti iš reklamuotojų. Nors toks vartotojų dėmesio pritraukimu grįstas modelis turėjo skaudžių pasekmių visuomenei, jis įgalino „Facebook“ siekti kuo didesnio augimo.

Kitas veiksnys, padėjęs šioms bendrovėms tapti milžiniškomis korporacijomis, buvo vadinamasis „tinklo efektas“ – reiškinys, apie kurį svajoja kiekvienas startuolio įkūrėjas. Jis remiasi tokiu principu: kuo

daugiau bendrovė turi vartotojų ir klientų, tuo geresni tampa jos algoritmai. Dėl šios priežasties ji laimi konkurencinę kovą prieš kitas įmones ir dar labiau sustiprina savo pozicijas rinkoje. Pavyzdžiui, „Facebook“ vartotojai prisijungė prie šio socialinio tinklo todėl, kad juo jau naudojosi visi kiti. Daugelis ilgą laiką aktyviai naudojasi šia platforma arba dėl minėtos priežasties nesiryžta ištrinti savo paskyros. Turbūt nė vienas „Apple“ gerbėjas nepaneigs, kad pereiti prie kito gamintojo įrenginių, pavyzdžiui, „Samsung“, arba derinti pastarojo priedus su „iPhone“ yra nemenkas iššūkis. Kurdama tarpusavyje derančius produktus ir paslaugas, „Apple“ užtikrina vartotojų lojalumą ir stiprina savo dominavimą rinkoje.

Neturime istorinių precedentų, kurie parodytų, kas nutinka, kai bendrovės išauga iki tokio masto. Šiandieninė tokių milžinių kaip „Google“, „Amazon“ ir „Microsoft“ rinkos vertė pasiekė neregėtas aukštumas. To rezultatas – ne tik didesnė finansinė nauda šių bendrovių akcininkams, įskaitant pensijų fondus, bet ir vis ryškiau matoma galios koncentracija vienose rankose. Milijardų žmonių privatumas, tapatybė, viešoji erdvė ir net darbo perspektyvos vis labiau priklauso nuo kelių stambių bendrovių, valdomų saujelės neįtikėtinai turtingų individų.

Nenuostabu, kad šių technologijų milžinių darbuotojai, net pastebėję nerimą keliančių dalykų, nemato prasmės skambinti pavojaus varpais – juk tai būtų taip pat beviltiška, kaip bandyti pakeisti „Titaniko“ kursą prieš pat jam atsitrenkiant į ledkalnį. Vis dėlto DI mokslininkė Timnita Gebru ryžosi tai padaryti.

2015-ųjų gruodį jai teko galimybė dalyvauti „NeurIPS“ konferencijoje – toje pačioje, kurioje Samas Altmanas ir Elonas Muskas paskelbė kuriantys DI „žmonijos labui“. Trisdešimtmetį vos perkopusią juodaodę mokslininkę nustebino tai, kad daugiatūkstantinėje minioje ji nepamatė beveik nė vieno į save panašaus dalyvio. Jos gyvenimo istorija smarkiai skyrėsi nuo daugelio kolegų, augusių palaikančioje aplinkoje.

Būdama vos penkerių, Gebru neteko tėvo – iš Eritrėjos kilusio elektros inžinieriaus, o paauglystėje turėjo bėgti iš karo siaubiamos Etiopijos. Mokydamasi Masačusetso vidurinėje mokykloje, ji nesulaukė daug palaikymo iš mokytojų, kurie skeptiškai vertino imigrantės akademinius siekius. Jie įtikinėjo, kad jai neverta rinktis pažangiems mokiniams skirtų kursų, esą šie gali būti jai per sunkūs. Kaip cituojama žurnale „Wired“, vienas mokytojas netgi išrėžė: „Mačiau daug tokių kaip tu, manančių, kad gali tiesiog atvykti iš svetur ir imtis pačių sunkiausių dalykų.“ Vis dėlto Gebru ryžosi lankyti pažengusiųjų kursus ir netgi užsitikrino galimybę studijuoti elektros inžineriją Stanfordo universitete.

Galiausiai Gebru susidomėjo DI sritimi ir kompiuterinės regos (angl. *computer vision*) sistemomis – programine įranga, gebančia „matyti“ ir analizuoti realųjį pasaulį. Ši technologija ją be galo žavėjo, bet taip pat kėlė tam tikrą nerimą. Dirbtinio intelekto sistemoms pradėta teikti vis didesnė svarba įvairiose srityse, lemiančiose žmonių gyvenimą. Pavyzdžiui, jos naudojamos atliekant asmens kreditingumo vertinimą, priimant sprendimus dėl būsto paskolos suteikimo, taip pat jomis grindžiamos veidų atpažinimo technologijos, naudojamos policijos veikloje, ar netgi teisminės sistemos, padedančios teisėjams skirti bausmes baudžiamosiose bylose. Atrodytų, tokios sistemos padėtų priimti objektyvius sprendimus, tačiau iš tiesų viskas ne taip paprasta. Jei sistemai mokyti naudojami duomenys yra šališki, tokia yra ir sistema. Pati Gebru ne kartą buvo asmeniškai susidūrusi su šališkumu.

Būdama jauna, viename San Fransisko bare kartu su kita juodaode mergina ji patyrė išpuolį – keli nepažįstami vyrai mėgino jas pasmaugti. Kai jos kreipėsi pagalbos į policiją, pareigūnai jas apkaltino melagystėmis ir uždarė į areštinę. Rengdama disertaciją Stanforde, Gebru sužinojo, kad per visą universiteto istoriją tik vienas juodaodis buvo apgynęs daktaro disertaciją kompiuterių mokslo srityje, neskaitant jos pačios. Be to, 2015 metais įvykusioje didžiausioje tarptautinėje DI konferencijoje, kurioje

Altmanas su Musku pristatė „OpenAI“, iš penkių tūkstančių dalyvių tik penki buvo juodaodžiai.

Gebru suprato, kad tai ne pavieniai atvejai, – šališkumas buvo įsigalėjęs sisteminiu lygmeniu visose srityse. Net praėjus keliems dešimtmečiams po XX a. pilietinių teisių judėjimo, rasizmas išliko įsišaknijęs viešajame gyvenime ir žmonių sąmonėje. Dirbtinis intelektas galėjo šią problemą dar paaštrinti. Visų pirma, DI dažniausiai kuria žmonės, kurie nėra patyrę rasizmo. Dėl to DI mokyti naudojami duomenys dažnu atveju objektyviai neatspindi mažumų grupių ir moterų.

Akademiniai tyrimai padėjo Gebru suvokti šios problemos pasekmes. Ją itin sudomino tyrimas, kuriuo siekta išanalizuoti JAV baudžiamosios teisenos sistemoje naudojamą programinę sistemą COMPAS (angl. *Correctional Offender Management Profiling for Alternative Sanctions*[4]). Teisėjai ir lygtinio paleidimo pareigūnai naudoja ją priimdami sprendimus dėl užstatų, bausmių skyrimo ar lygtinio paleidimo.

COMPAS sistema, veikianti pagal mašininio mokymosi principus, kiekvienam kaltinamajam priskiria rizikos balą. Kuo balas aukštesnis, tuo didesnė tikimybė, kad asmuo pakartotinai nusikals. Sistema dažniau priskirdavo aukštus rizikos balus juodaodžiams nei baltaodžiams, tačiau prognozės neretai būdavo klaidingos. Paaiškėjo, kad, prognozuodama būsimą nusikalstamą elgesį, ši sistema dvigubai dažniau klydo vertindama juodaodžius nei baltaodžius. Tai atskleidė 2016 metais organizacijos „ProPublica“ atliktas tyrimas, kuriame buvo išanalizuoti septyni tūkstančiai rizikos balų, priskirtų Floridoje areštuotiems asmenims, ir jų tolimesnių dvejų metų nusikalstamos veiklos duomenys. Be to, sistema dažnai netinkamai įvertindavo baltaodžius, priskirdama mažą rizikos balą tiems, kurie vėliau pakartotinai nusikalsdavo. JAV baudžiamosios teisenos sistema ir taip yra šališka juodaodžių atžvilgiu, o tokie neaiškiais duomenimis pagrįsti DI įrankiai šią nelygybę galėjo dar labiau padidinti.

Rengdama doktorantūros disertaciją Stanforde, Gebru atkreipė dėmesį į dar vieną galimą netinkamo DI pritaikymo instituciniu lygmeniu pavyzdį. Ji išmokė kompiuterinės regos modelį taip, kad jis nustatytų 22 milijonų „Google Street View“ vaizduose matomų automobilių markes, o tada išanalizavo, ką šios automobilių markės gali atskleisti apie rajonų demografiją. Susiejusi automobilių markes su gyventojų surašymo ir nusikalstamumo duomenimis, ji nustatė, kad rajonuose, kuriuose buvo daugiau „Volkswagen“ markės automobilių ir pikapų, gyveno daugiau baltaodžių, o vietovėse, kuriose dominavo „Oldsmobile“ ir „Buick“ markės automobiliai, dauguma gyventojų buvo juodaodžiai. Taip pat rajonai su didesniu mikroautobusų skaičiumi pasižymėjo didesniu nusikalstamumu. Tokios sąsajos gali būti panaudotos netinkamais tikslais. Kas nutiktų, jei policija remtųsi tokiais duomenimis prognozuodama, kur gali įvykti nusikaltimai, visai kaip filme „Įspėjantis pranešimas“ (*Minority Report*)?

Tokia mintis nėra visiškai nutolusi nuo realybės. Jau kelerius metus JAV policijos padaliniai naudoja vadinamąją prognozėmis paremto patruliavimo (angl. *predictive policing*) technologiją, patariančią, kuriose vietose turėtų patruliuoti pareigūnai. Kadangi sistemoms mokyti naudojami istoriniai duomenys, pareigūnai daugiausia siunčiami patruliuoti mažumų bendruomenėse. Jei duomenys rodo, kad tam tikrose bendruomenėse policija aktyviai vykdo patruliavimą, sistema nurodo papildomo patruliavimo poreikį, taip paaštrindama įsisenėjusią problemą.

Dirbtinis intelektas subtiliais, bet klastingais būdais stiprina ir kitus stereotipus virtualioje erdvėje. Vertimo įrankiai, tokie kaip „Google Translate“ ir „Microsoft Bing Translate“, versdami tam tikrų profesijų pavadinimus iš kitų kalbų, dažnai priskiria jas vyriškajai giminei. Pavyzdžiui, turkiška frazė „O bir muhendis“, kurioje vartojamas lyties atžvilgiu neutralus įvardis, verčiama taip: „Jis yra inžinierius“, o frazės „O bir hemsire“ vertimas skamba taip: „Ji yra slaugytoja.“ Šios prielaidos

daromos todėl, kad programinės sistemos veikimas paremtas plačiai naudojamu įterptinių žodžių (angl. *word embedding*) metodu, kuris leidžia nustatyti žodžius, dažnai pasirodančius kartu su kitais žodžiais. Tada modelis nusprendžia, kuris žodis geriausiai dera su duotuoju, pavyzdžiui, profesijos pavadinimą „inžinierius“ susieja su įvardžiu „jis“. Internetinių platformų, tokių kaip „Google“, „Facebook“, „Netflix“ ir „Spotify“, rekomendacijų sistemos taip pat grindžiamos šiuo metodu, nors jis įtvirtina lyčių atžvilgiu nesubalansuotą realybės vaizdą.

Akivaizdu, kad tam tikrų DI problemų buvo negalima ignoruoti. 2015 metais Samo Altmano pranešimas apie „OpenAI“ įkūrimą sukėlė Gebru didelį pasipiktinimą. Reaguodama į tai, ji ėmėsi rašyti viešą laišką. Gebru kritikavo švaistūnišką kelių egoistiškų milijardierių, tokių kaip Muskas ir Thielis, sprendimą investuoti lėšas į visagalio DI kūrimą, ir piktinosi, kad vienintelis iškeltas klausimas apie naują ne pelno organizaciją buvo tas, kad jos tyrėjai pernelyg daug dėmesio skiria giluminiam mokymuisi.

„Baltaodis technologijų magnatas, gimęs ir užaugęs Pietų Afrikoje apartheido laikais, kartu su investuotojų ir tyrėjų grupe, kurią sudaro vien baltaodžiai vyrai, bando užkirsti kelią DI „užvaldyti pasaulį“, ir vienintelė matoma problema – ta, kad „visi tyrėjai dirba giluminio mokymosi srityje“? – rašė ji. – Neseniai „Google“ pristatė kompiuterinės regos algoritmą, kuris juodaodžius priskyrė beždžionėms. BEŽDŽIONĖMS. Kai kas bando šią klaidą paaiškinti teigdamas, kad algoritmas veikiausiai rėmėsi odos spalva kaip esminiu žmonių atpažinimo kriterijumi. Jei komandoje būtų buvęs bent vienas juodaodis arba kas nors, kas būtų atsižvelgęs į rasinius aspektus, įrankis, priskiriantis juodaodžius beždžionėms, nebūtų apskritai išvydęs dienos šviesos. Įsivaizduokite, jei būtų sukurtas algoritmas, traktuojantis baltaodžius kaip nepriklausančius žmonių rūšiai. Jokia JAV bendrovė nesiryžtų tokios sistemos pavadinti tinkama naudoti žmonėms atpažinti.“

Viena iš Gebru kolegių patarė jai laiško nepublikuoti – jis buvo pernelyg atviras, be to, autorystę greičiausiai būtų buvę nesunku nustatyti. Gebru galiausiai nusprendė laiško neviešinti (tą padaryti ji ryžosi tik po kelerių metų). Vis dėlto ji negalėjo atsistebėti, kodėl kai kurie įtakingiausi Silicio slėnio asmenys taip jaudinasi dėl galimų apokaliptinių DI ateities scenarijų, kai ši technologija jau dabar kenkia žmonėms. Dėmesys krypo į du galimus paaiškinimus. Pirma, beveik nė vienas iš „OpenAI“ ir „DeepMind“ vadovų niekada nepatyrė ir tikriausiai nepatirs diskriminacijos dėl rasės ar lyties. Antra, paradoksalu, bet jiems buvo naudinga garsiai kalbėti apie galingo dirbtinio superintelekto grėsmes. Nors skleisti žinią apie produkto, kurį pats pardavinėji, pavojus gali atrodyti kaip prieštaringas žingsnis, vis dėlto tai yra genialus rinkodaros triukas. Žmonės paprastai labiau rūpinasi dabarties realijomis nei tolima ateitimi. Jei jie įtikėtų, kad DI ateityje gali sunaikinti žmoniją, tai skatintų dar labiau žavėtis įspūdingomis dabartinėmis šios technologijos galimybėmis.

Ši strategija taip pat yra sumanus būdas nukreipti visuomenės dėmesį nuo keblių, bet daug aktualesnių problemų, kurias bendrovės jau dabar galėtų spręsti, pristabdydamos DI vystymą ir ribodamos savo modelių galimybes. Vienas būdas sumažinti šališkų DI modelių sprendimų tikimybę – kruopščiau analizuoti duomenis, kurie naudojami jiems mokyti. Kita galima strategija – kurti siauresnio pobūdžio modelius, kad jų gebėjimas apibendrinti žinias būtų ribotas.

Tai būtų ne pirmas kartas, kai stambios bendrovės mėgina nukreipti visuomenės dėmesį nuo svarbių problemų, tuo pat metu siekdamos verslo augimo. XX a. aštuntojo dešimtmečio pradžioje plastiko pramonės atstovai, remiami naftos bendrovių, iškėlė perdirbimo idėją kaip galimą vis didėjančios plastiko atliekų problemos sprendimą. 1953 metais įsteigta organizacija „Keep America Beautiful“, iš dalies finansuojama gėrimų ir pakuočių gamybos įmonių, rengė visuomenės švietimo kampanijas, kuriomis skatino vartotojus perdirbti atliekas. Jos sukurtoje garsiojoje

reklamoje „Verkiantis indėnas“, pirmą kartą parodytoje 1971 metais per Tarptautinę Žemės dieną, žmonės buvo raginami perdirbti butelius bei laikraščius ir taip prisidėti prie taršos mažinimo. Nepaisantieji tokio raginimo laikyti atsakingais už aplinkos taršą.

Perdirbimas savaime nėra blogas dalykas. Tačiau skatindami šią praktiką, pramonės atstovai leido sau teigti, kad plastiko atliekos nekenkia aplinkai, jei tik jos tinkamai perdirbamos, taigi atsakomybė už šią problemą buvo perkelta nuo gamintojų ant vartotojų pečių. Plastiko gamybos įmonės žinojo, kad didelio masto atliekų perdirbimas yra brangus ir dažnai neefektyvus procesas. 2020 metais NPR[5] atliktas tyrimas ir PBS[6] laidų serija „Frontline“ atskleidė, kad, nepaisant daugelį dešimtmečių trukusių visuomenės švietimo kampanijų, per visą tą laiką buvo perdirbta mažiau nei 10 proc. plastiko atliekų.

Tiesa ta, kad šiomis kampanijomis pavyko nukreipti visuomenės dėmesį nuo spartaus plastiko pramonės augimo ir su tuo susijusio neigiamo poveikio aplinkai. Atliekų perdirbimas tapo viešojo diskurso dalimi. Žiniasklaida, vartotojai ir politikai Vašingtone daugiau dėmesio skyrė diskusijoms apie atliekų perdirbimą nei realiam plastiko gamybos reguliavimui.

Kaip ir didžiosios naftos pramonės bendrovės, bandžiusios nukreipti visuomenės dėmesį nuo savo veiklos daromos žalos aplinkai, pagrindiniai DI kūrėjai galėjo pasinaudoti viešajame diskurse skleidžiamais gąsdinimais dėl ateities „Terminatoriaus“ ar „Skynet“ grėsmės, kad atitrauktų dėmesį nuo šiandieninių mašininio mokymosi algoritmų keliamų problemų. Dirbtinio intelekto kūrėjams nebereikėjo prisiimti atsakomybės už šias problemas ar imtis neatidėliotinų veiksmų – tai tapo abstrakčia problema, kurią galima spręsti vėliau.

2017 metų sausį, dar prieš tai, kai „DeepMind“ bandė padėti „Google“ vėl įsitvirtinti Kinijos rinkoje su „AlphaGo“, Gebru pristatė savo disertacijos rezultatus Silicio slėnio rizikos kapitalo investuotojams ir įmonių vadovams. Apžvelgdama skaidres, ji paaiškino, kad DI sistemų

gebėjimas atpažinti automobilius galėtų būti derinamas su prognozavimo algoritmais, padedančiais nuspėti, pavyzdžiui, rinkimų rezultatus arba namų ūkių pajamas.

Vienas renginyje dalyvavusių rizikos kapitalistų, Elono Musko draugas ir „Tesla“ investuotojas Steve'as Jurvetsonas, buvo itin nustebintas tokių rezultatų, tačiau tokios reakcijos priežastis buvo visiškai kitokia, nei Gebru tikėjosi. Tokie duomenys galėjo suteikti „Google“ nepaprastai vertingų įžvalgų apie įvairius rajonus ar miestus. Jis buvo taip sužavėtas Gebru pristatymo, kad nuotraukas iš renginio paskelbė savo „Facebook“ paskyroje.

Dirbtinio intelekto sritį lydinčią prieštarą puikiai atspindėjo renginio dalyvių reakcijos: vieni į šią technologiją žvelgė kaip į priemonę uždirbti, kiti, kaip Gebru, matė jos keliamas grėsmes, kurias reikia suvaldyti. Su kiekvienu DI patobulinimu išryškėdavo naujos nenumatytos pasekmės, kurios dažnai paliesdavo mažumų grupes. Pavyzdžiui, veido atpažinimo sistemos beveik visada tiksliai atpažindavo baltųjų vyrų veidus, bet dažnai klysdavo identifikuodamos juodaodes moteris. 2018 metais MIT doktorantės Joy Buolamwini atliktas svarbus tyrimas atskleidė, kad IBM ir „Microsoft“ veido atpažinimo programos ir Kinijos sistema „Face++“ dažniau klydo nustatydamos tamsesnės odos spalvos moterų lytį – šią problemą tyrėja pastebėjo, kai panaši programa neatpažino jos pačios veido. Šioms sistemoms mokyti dažniausiai naudojami rinkiniai nuotraukų, kuriose daugiausia vaizduojami baltieji vyrai, taip pat fotografijos iš interneto. Duomenų bazėse jie dominuoja todėl, kad internetas atspindi Vakarų gyventojų, turinčių geresnę prieigą prie jo, demografiją.

Tačiau Gebru nenuleido rankų – ji pasiūlė konkrečius sprendimus šiai problemai spręsti. Vienas jų – užtikrinti, kad DI sistemų kūrėjai, mokydami savo modelius, laikytųsi griežtesnių standartų. Bendradarbiaudama su „Microsoft“, ji parengė taisyklių rinkinį „Duomenų rinkinių aprašai“ (angl. *Datasheets for Datasets*), kuriame

nurodoma, kad kurdami DI modelį programuotojai turėtų parengti dokumentą, kuriame būtų išdėstyta visa informacija apie modelio kūrimo procesą, panaudotus duomenis, jo taikymą, galimus ribotumus ir etinius aspektus. Dirbtinio intelekto kūrėjai tai galėjo laikyti papildoma administracine našta, tačiau šios taisyklės turėjo aiškų tikslą: jei modelis pasirodytų esantis šališkas, būtų daug lengviau nustatyti tokio šališkumo priežastis.

Suprasti, kodėl DI sistemos klysta, yra daug sudėtingiau, nei gali atrodyti iš pirmo žvilgsnio, juolab kad tokios sistemos tampa vis sudėtingesnės. 2018 metais „Amazon“ pastebėjo, kad bendrovės viduje naudojamas DI įrankis, skirtas darbo paraiškoms atrinkti, rekomendavo daugiau vyrų kandidatų nei moterų. Taip nutiko dėl paprastos priežasties: įrankiui mokyti buvo naudojami per pastaruosius dešimt metų gauti gyvenimo aprašymai, kuriuos daugiausia pateikė vyrai. Modelis išmoko, kad gyvenimo aprašymai, kuriuose nurodyti vyrams būdingi bruožai, turi būti traktuojami kaip labiau pageidautini. Tačiau „Amazon“ negalėjo arba nesugebėjo šios problemos išspręsti, todėl tiesiog nutraukė įrankio naudojimą.

Panašų radikalų žingsnį žengė „Google“, kai jos nuotraukų programėlė kai kuriuos juodaodžius žmones priskyrė goriloms. Sistemos kūrėjai tiesiog išjungė gorilų atpažinimo funkciją, nors programa ir toliau atpažino kitų rūšių gyvūnus. Problema kilo todėl, kad į mokyti skirtų duomenų rinkinį nebuvo įtraukta pakankamai juodaodžių ir tamsesnės odos žmonių nuotraukų, be to, programa tikriausiai nebuvo gerai išbandyta su darbuotojais. 2023 metų pabaigoje bendrovė vis dar nežinojo, kaip patobulinti DI modelį, todėl šią funkciją apskritai išjungė.

Kai kurie DI tyrėjai teigia, kad šių šališkumų ištaisyti beveik neįmanoma, nes šiuolaikiniai DI modeliai tokie sudėtingi, kad net patys jų kūrėjai dažnai nesupranta, kodėl modeliai priima tam tikrus sprendimus. Giluminio mokymosi modeliai, tokie kaip neuroniniai tinklai, apima milijonus ar net milijardus parametrų, dar vadinamų

„svoriais“, kurie veikia kaip reguliatoriai sudėtinguose matematiniuose ryšiuose tarp tinklo sluoksnių. Neuroninio tinklo sluoksnius galima vaizdžiai palyginti su gamyklos surinkimo linija. Kiekvienas darbuotojas atlieka tam tikrą užduotį, pavyzdžiui, nudažo žaislinę mašinytę arba pritvirtina prie jos ratukus. Galiausiai gamybos linijos gale pasirodo užbaigta žaislinė mašinytė. Kiekvienas neuroninio tinklo sluoksnis veikia kaip atskira surinkimo linijos stotelė, kurioje atliekami nedideli duomenų koregavimai. Problema ta, kad dėl daugybės smulkių pakeitimų serijos sunku atsekti, kokie tiksliai pakeitimai buvo atlikti kiekvienoje gamybos linijos stotelėje (t. y. neuroninio tinklo sluoksnyje) gaminant „žaislinę mašinytę“. Todėl nelengva išsiaiškinti, kas konkrečiai lėmė DI modelio sprendimą priskirti juodaodžiui kaltinamajam didesnę riziką pakartotinai nusikalsti.

Žiniasklaidoje plačiai nuskambėjus vadinamajam „gorilų skandalui“, prie „Google“ komandos prisijungė kompiuterių mokslininkė Margareta Mitchell, kurios tikslas buvo užkirsti kelią panašioms klaidoms. Los Andžele gimusi mokslininkė, pagarsėjusi DI tyrėjų rate dėl iniciatyvų mažinti šališkumą mašininio mokymosi sistemose, priklausė nedidelei, bet augančiai tyrėjų grupei, kuri rūpinosi realiu šių sistemų poveikiu. Kaip ir Gebru, ji nerimavo dėl keistų klaidų, kurias daro DI sistemos. Daugumą savo doktorantūros tyrimų Mitchell atliko kompiuterinės lingvistikos ir natūraliosios kalbos generavimo srityse, nagrinėdama įvairius būdus, kaip kompiuteriai gali apibūdinti objektus arba analizuoti tekste perteiktas emocijas. Dirbdama „Microsoft“ prie akliesiems skirtos programėlės, ji gerokai suglumo pastebėjusi, kad baltos odos žmonės, tokie kaip ji pati, buvo žymimi kaip „asmenys“, o tamsesnės odos žmonės – kaip „juodaodžiai asmenys“.

Kartą atlikdama eksperimentą su neuroniniu tinklu, skirtu vaizdams aprašyti, Mitchell pateikė jam nuotraukų, vaizduojančių fabriko sprogimą Anglijoje. Vienoje nuotraukų, darytoje iš netoliese esančio namo buto viršutiniame aukšte, buvo užfiksuoti kylantys dūmų stulpai,

o pirmame plane matėsi televizijos naujienų kanalo reporteriai, pranešantys apie įvykį. Mitchell priblokšė tai, kad DI sistema šią nuotrauką apibūdino kaip „nuostabią“, „gražią“ ir perteikiančią „puikų vaizdą“.

„Problema ta, kad sistemai viskas atrodė tiesiog puiku, – sako Mitchell, prisimindama garsiąją dainą iš „Lego filmo“ (*The Lego Movie*), vaizduojančio kaladėlių pasaulį, kuriame visos gyvenimo bėdos užglaistomos. – Mirtingumo idėja šiai sistemai buvo svetima, be to, ji nesuprato, kad mirtis yra blogas dalykas.“

Iš pateiktų nuotraukų sistema buvo išmokusi, kad saulėlydžiai yra gražūs, o vaizdai iš aukštai – nuostabūs. Tuomet Mitchell atsivėrė akys – ji suprato, kad duomenys yra viskas. Dėl mokymo duomenyse buvusių spragų sistema įgijo įvairių šališkumų, tarp jų ir tokių, dėl kurių nereagavo į žmonių žūtis.

Spręsdama šiuos klausimus „Google“, mokslininkė pastebėjo dar vieną darbo stambioje technologijų bendrovėje trūkumą – ji buvo sukaustyta biurokratijos pančių ir įtraukta į nesibaigiantį susitikimų maratoną. Įmonės vadovai atrodė nuolat susirūpinę įmonės reputacija.

2018 metais Mitchell išsiuntė Gebru elektroninį laišką, kviesdama prisijungti prie jos „Google“. Dirbtinio intelekto etikos srityje dirbo tiek mažai žmonių, kad jos jau pažinojo viena kitą. Mitchell pasiūlė Gebru kartu vadovauti „Google“ DI etikos tyrimų komandai.

Tačiau Gebru dvejojo. Ji buvo girdėjusi gandų, kad „Google“ tvyro prasta psichologinė atmosfera, dėl kurios ypač kenčia moterys ir mažumų atstovai. Turbūt geriausiai tai iliustruoja Andy’io Rubino atvejis. Populiariosios operacinės sistemos „Android“ kūrėjas buvo tapęs tikra žvaigžde „Google“, tačiau 2014 metais jis tyliai paliko bendrovę po to, kai jam buvo mesti kaltinimai dėl seksualinio priekabiavimo. Po kelerių metų leidinio „New York Times“ atliktas tyrimas atskleidė, kad „Google“ vadovybė išnagrinėjo šiuos kaltinimus ir nustatė, kad jie pagrįsti. Tačiau užuot tiesiog atleidusi Rubiną, bendrovė su juo atsisveikino kaip su tikru

didvyriu, įteikdama jam įspūdingą 90 mln. JAV dolerių išeitinę kompensaciją.

Tačiau ne viskas „Google“ klostėsi blogai. Gebru žavėjo tai, kad darbuotojai, pastebėję neteisybę, išdrįsdavo su tuo kovoti. Tūkstančiai „Google“ darbuotojų buvo surengę masinį protestą dėl Rubino „auksinio parašiuto“, o dar prieš Gebru prisijungiant prie „Google“, daugiau nei trys tūkstančiai darbuotojų pasirašė atvirą laišką generaliniam direktoriui Sundarui Pichai’ui, kuriame reikalavo, kad bendrovė pasitrauktų iš projekto „Maven“. „Google“ galiausiai įvykdė jų reikalavimus. Šiuos protestus koordinavo DI etikos ekspertė Meredith Whittaker, kurios įtaigūs argumentai privertė bendrovę persvarstyti programą. Gebru suprato, kad „Google“ vis dėlto gali būti puiki vieta skatinti atsakingą požiūrį į DI sistemų kūrimą, pavyzdžiui, pasitelkiant jos parengtus Duomenų rinkinių aprašus.

Pažvelgus į naujosios etikos komandos dydį, buvo akivaizdu, kad tokie technologijų gigantai kaip „Google“ pirmiausia investuoja į DI pajėgumus. Nepaisant tokio į etiką orientuoto darbo svarbos, komandą sudarė vos keli kompiuterių mokslo specialistai. Tuo metu tūkstančiai kitų bendrovės inžinierių ir tyrėjų toliau dėjo visas pastangas, siekdami sukurti vis spartesnes ir tobulesnes DI sistemas. Gebru ir Mitchell turėjo sunkiai dirbti, kad spėtų vertinti ir analizuoti galimas tokiais tempais tobulinamų sistemų nenumatytas pasekmes.

Darbas „Google“ vis labiau slėgė Mitchell. Kai susitikimuose ji įspėdavo vadovus apie problemas, kurias galėtų sukelti jų DI sistemos, ji iš karto sulaukdavo elektroninių laiškų iš Personalo skyriaus, raginančių ją būti labiau geranorišką. Tokiose Silicio slėnio bendrovėse kaip „Google“, „Apple“ ar „Facebook“ moterys sudarė tik apie ketvirtadalį IT darbuotojų, o 2020 metų statistika rodo, kad nuo kiekvieno dolerio, kurį uždirbo vyrai, moterys gavo tik 86 centus. Įsidarbindamos ar siekdamos karjeros, moterys dažnai patirdavo nelygybę, priekabiavimą ir diskriminaciją. Šios problemos ypač skaudžiai paliesdavo juodaodes

moteris. Daugelis moterų, kurias galėjai sutikti įprastose Silicio slėnio konferencijose ar renginiuose, dirbo rinkodaros ar viešųjų ryšių skyriuose, o ne kompiuterių inžinerijos ar tyrimų srityse. Dėl šios priežasties DI etikos tyrimai ypač traukė moteris – jos geriau nei kas kitas žinojo, ką reiškia diskriminacija. Tačiau tokioje aplinkoje joms buvo sunku užtikrinti, kad jų balsas būtų išgirstas.

Šių iššūkių fone Mitchell negalėjo atsistebėti Gebru drąsa – ši nevengė stoti į akistatą su įtakingais asmenimis, jei pastebėdavo neteisybę ar prireikdavo gauti reikiamų išteklių. Vieną dieną, sėdėdamos Gebru kabinete 41-ajame „Google“ pastate, kolegės aptarinėjo jas nuliūdinusį el. laišką, gautą iš vieno vadovo. Jo turinys atspindėjo diskriminacinį požiūrį, su kuriuo jos ir taip dažnai susidurdavo. Mitchell vos tramdė ašaras. Gebru reakcija buvo visai kitokia.

„Nenusimink, – tarė ji. – Verčiau reaguok pykčiu.“

Pasidėjusi nešiojamąjį kompiuterį arčiau savęs, Gebru ėmėsi rašyti atsakymą vadovui, garsiai skaitydama kontrargumentus kiekvienai vadovo pastabai. Vėliau, kai Mitchell ir Gebru buvo atleistos iš „Google“, tas pats vadovas viešai stojo jų pusėn, o po kurio laiko ir pats paliko bendrovę savo noru.

Netrukus Gebru ir Mitchell pastangos atsipirko – jų iškeltos problemos pagaliau sulaukė deramo dėmesio, nors dėl to joms teko paaukoti karjerą bendrovėje, o jų atleidimas sukėlė viešą skandalą. Mokslininkės siekė, kad jų pastangos nebūtų užgožtos milžiniškos „Google“ korporacinės mašinos. Bendrovės komanda, kuriai buvo pavesta kurti vis išmanesnes DI sistemas, netruko suklupti tuo metu, kai DI srityje įvyko vienas svarbiausių proveržių per visą jo kūrimo istoriją.

4 Nusikaltimą padariusių asmenų rizikos vertinimo ir profiliavimo sistema alternatyvioms bausmėms (vert. past.).

5 JAV nacionalinis visuomeninis radijas (angl. *National Public Radio*) (vert. past.).

6 JAV visuomeninis transliuotojas (angl. *Public Broadcasting Service*) (vert. past.).

9 SKYRIUS

Galijoto paradoksas

2017-aisiais bendrovėje „Google“ dirbo apie aštuoniasdešimt tūkstančių etatinių darbuotojų. Tarp jų buvo ne vien inžinieriai. Milžiniškoje komandoje taip pat triūsė dienos logotipo piešinio (angl. *Google Doodle*), rodomo virš paieškos laukelio, kuratoriai, chiropraktikai ir masažo vadybininkai, „užkandžių specialistai“, rūpinęsi, kad darbuotojai nepraalktų tarp trijų pagrindinių valgių, sodininkai, prižiūrėję augalus, ir valytojai, blizginę stalo futbolo stalus.

„Google“ verslo modelis buvo tikra aukso gysla. Tais metais bendrovė iš reklamos uždirbo beveik 100 mlrd. JAV dolerių, o 2024-aisiais šios pajamos kone padvigubėjo, todėl nenuostabu, kad vadovai daug lėšų skyrė komandai stiprinti ir naujiems talentingiems specialistams pritraukti. Silicio slėnyje įmonių sėkmė dažniausiai vertinama pagal pritrauktų investicijų sumą ir darbuotojų skaičių. Milžiniška komanda atspindėjo generalinių direktorių, kaip antai Larry’io Page’o ir Sergey’aus Brino, imperinius užmojus, nors kartais nebuvo aišku, ką iš tiesų veikia daugelis vidurinės grandies vadovų.

Besaikis „Google“ augimas nebuvo išskirtinis reiškinys. Tuo metu „Facebook“ dirbo maždaug 40 tūkst. darbuotojų, „Microsoft“ – 124 tūkst. Startuolių steigėjai svajojo įkurti savo įmonių miestelius, kuriuose veiktų sporto klubai ir ledų kioskeliai, vaišinantys darbuotojus ledais. Vis dėlto Demisas Hassabisas buvo viena iš nedaugelio išimčių – galbūt todėl, kad nuo Silicio slėnio jį skyrė vandenynas. Jis nenorėjo, kad įvairūs trukdžiai, tokie kaip specialios privilegijos darbuotojams ir aklas troškimas plėstis, atitrauktų „DeepMind“ dėmesį nuo bendrovės misijos.

Tokio masto korporacijos kaip „Google“ dažnai susiduria su problema: net jei joje sukuriamas produktas, kuris galėtų sukelti perversmą, jis dažnai taip ir neišvysta dienos šviesos. „Google“ skaitmeninės reklamos verslas buvo laikomas kone šventu, neliečiamu, todėl niekas be rimtos priežasties nesiryžo keisti algoritmų, kuriais buvo grindžiama bendrovės veikla. Nors Silicio slėnis vadinamas pasaulio inovacijų sostine, didžiausios jo įmonės iš tikrųjų nėra tokios jau novatoriškos. „Google“ pradžios puslapis per dešimtmetį beveik nepasikeitė. „iPhone“ telefonai vis dar to paties nuobodaus dizaino. Beveik kiekviena nauja „Facebook“ funkcija tiesiogiai pasiskolinta iš konkurentų, tokių kaip „Snapchat“ ar „TikTok“. Kai korporacijų pajamos peržengia dešimčių milijardų ribą, keisti jų sėkmės formulę tampa pernelyg rizikinga.

Todėl, kai „Google“ tyrėjų grupė sukūrė technologiją, laikomą vienu svarbiausių pastarojo dešimtmečio laimėjimų DI srityje, bendrovė taip ir neskyrė jai deramo dėmesio. Ši istorija puikiai rodo, kaip monopolistinis mastas riboja technologijų milžinių gebėjimą kurti naujoves. Jos dažniau reaguoja į kitų sukurtas inovacijas, kopijuodamos jas arba įsigydamos jas sukūrusius startuolius. Tačiau „Google“ toks aplaidumas daug kainavo. Galiausiai „OpenAI“ ne tik pasinaudojo šiuo išradimu, bet ir pritaikė jį novatoriškam įrankiui, kuris pirmą kartą per ilgus metus ėmė kelti paieškos milžinei realią konkurencinę grėsmę.

Pavadinime „ChatGPT“ raidė „T“ reiškia „transformatorių“ (angl. *transformer*). Tai neturi nieko bendra su ateiviais robotais, gebančiais transformuotis į sunkvežimius. Transformatorius – sistema, leidžianti kompiuteriams generuoti tekstą, smarkiai nesiskiriantį nuo sukurtojo žmogaus. Transformatorius tapo kertiniu naujosios kartos *generatyvinio* DI elementu, gebančiu kurti tikrovišką tekstą, vaizdus, vaizdo įrašus, DNR sekas ir daugybę kitokio pobūdžio duomenų. Ši 2017 metais sukurta technologija padarė itin reikšmingą įtaką DI sričiai ir savo svarba prilygo išmaniųjų telefonų atsiradimui. Anksčiau mobiliaisiais telefonais buvo

galima atlikti vos kelias funkcijas – skambinti, siųsti trumpąsias žinutes ir retkarčiais pažaisti „Gyvatėlę“. Atsiradus liečiamiesiems ekranams, vartotojams staiga atsivėrė galimybė naršyti internete, naudotis GPS sistema, daryti aukštos kokybės nuotraukas ir mėgautis daugybės kitų programėlių teikiamais privalumais.

Transformatoriai taip pat įgalino DI inžinierius kurti efektyviau veikiančias sistemas, galinčias apdoroti kur kas didesnius duomenų kiekius ir greičiau analizuoti žmogaus kalbą. Prieš atsirandant transformatoriams, dialogas su pokalbių robotu (angl. *chatbot*) priminė bendravimą su neišmaniu įrenginiu, mat senesnių sistemų veikimą lėmė taisyklių rinkiniai ir sprendimų medžiai. Jei tokio roboto paklausdavai ko nors, kas nebuvo iš anksto užprogramuota (o taip nutikdavo dažnai), jis susipainiodavo arba suklysdavo. Tokiais principais veikė pirmieji skaitmeniniai asistentai, tokie kaip „Apple Siri“, „Amazon Alexa“ ar net „Google Assistant“. Šios sistemos kiekvieną užklausą traktuodavo kaip atskirą, todėl visiškai nesuprasdavo konteksto. Jos negebėjo įsiminti ankstesnių vartotojo klausimų taip, kaip žmogus pokalbio metu. Štai pavyzdys:

„Alexa, koks dabar oras Indianapolyje?“

„Šiuo metu Indianapolyje apie −4 °C, debesuota.“

„Per kiek valandų pasiekčiau tą miestą, skrisdama iš Londono?“

„Skrydis iš Londono į jūsų dabartinę vietą užtruktų maždaug keturiasdešimt penkias minutes.“

Tuo metu buvau Sario grafystėje. Jei tikėtume skaitmeninio asistento žodžiais, atskristi iš Londono „Heathrow“ oro uosto iki mano tuometinės vietos būtų užtrukę maždaug keturiasdešimt penkias minutes. Kad ir kokia logika „Alexa“ vadovavosi numatydama tokį painų skrydžio maršrutą, aišku viena: ji nesuprato, jog „tas miestas“ reiškia Indianapolį, apie kurį klausiau vos prieš kelias sekundes. Daugumos tradicinių skaitmeninių asistentų sistemos buvo ribotos ir daugiausia rėmėsi

raktažodžiais. Dėl to jos dažnai pateikdavo į kontekstą neorientuotus atsakymus.

Transformatoriai padėjo pokalbių robotams atsikratyti šių trūkumų. Jie ėmė suprasti kalbos niuansus ir žargoną, galėjo remtis tuo, ką vartotojas pasakė prieš kelias sekundes. Jie gebėjo atsakyti į kone bet kokį klausimą ir pateikti individualiai pritaikytą atsakymą. Taip patobulėjusias sistemas galima apibūdinti vienu žodžiu: jos tapo *universalesnės*. Daugeliui DI tyrėjų tai buvo svarbus žingsnis BDI link. Be to, tai įžiebė diskusijas, ar kompiuteriai jau pradeda suprasti kalbą taip, kaip žmonės, ar vis dar apdoroja ją pagal matematiniais modeliais grįstas prognozes.

Kiek stebina, kad šis išradimas apskritai buvo sukurtas „Google“ viduje. Nepaisant sutelktų talentingų specialistų ir išteklių, per didelis bendrovės mastas ir baimė kaip nors pakenkti reklamos verslui trukdė darbuotojams įgyvendinti novatoriškas idėjas. Nors prie projekto „Google Brain“ dirbo geriausi giluminio mokymosi tyrėjai, pasak buvusių darbuotojų, juos stabdė vadovybės nustatyti neaiškūs tikslai ir strategijos. Bendrovė beveik užmigo ant laurų dėl to, kad joje dirbo tiek daug talentingų mokslininkų, tokių kaip Geoffrey’is Hintonas. Buvo taikomi itin aukšti darbo standartai, be to, „Google“ jau naudojo pažangiausias DI technologijas, pavyzdžiui, grįžtamojo ryšio neuroninius tinklus, kad kasdien apdorotų milijardus žodžių.

Jauni DI tyrėjai, kaip Illia Polosukhinas, turėjo galimybę dirbti kartu su žmonėmis, kurie išrado šias technologijas. 2017 metų pradžioje dvidešimt penkerių ukrainietis ruošėsi palikti „Google“ ir imtis rizikingesnių projektų. Kartą per pietų pertrauką vienoje iš „Google“ valgyklų, esančioje dviem aukštais žemiau Larry’io Page’o kabineto, Polosukhinas kartu su dviem kitais tyrėjais – Ashishu Vaswaniu ir Jakobu Uszkoreitu – ėmėsi spontaniškai generuoti ir svarstyti įvairias idėjas. Jo kompanionai nenorėjo eiti keliu, kuriuo buvo pasukę kiti bendrovės mokslininkai. Vaswanis degė troškimu dirbti prie didelio projekto.

Uszkoreitas buvo išdirbęs „Google“ daugiau kaip dešimt metų ir skeptiškai vertino „Google Brain“ paskatų sistemą, dėl kurios komanda ėmė priminti susireikšminusią akademikų bendruomenę. Po to, kai prie bendrovės prisijungė daug naujų absolventų ir akademikų, Uszkoreitą supo žmonės, kuriems svarbiausia buvo tai, kad jų pavardė būtų pirmoji tarp mokslinio darbo autorių arba kad jų tyrimas būtų pristatytas konferencijoje. Jam atrodė, jog visa tai užgožia siekį kurti puikius produktus.

Jei kokiame vakarėlyje Uszkoreitas atskleisdavo, kur dirba, iškart sulaukdavo nustebusių žvilgsnių. Tačiau kai pridurdavo, kad dirba prie „Google Translate“ projekto, aplinkiniai imdavo juoktis. Šis įrankis buvo ne itin patogus naudoti ir dažnai pateikdavo netikslius vertimus, ypač iš kalbų, kurių rašto sistema nėra lotyniška, pavyzdžiui, kinų. Polosukhinas sutiko, kad „Google Translate“ veikia prastai. Be to, šiuo įrankiu skundėsi ir jo bičiuliai Kinijoje. Uszkoreitas garsiai svarstė, ar nebūtų galima jį patobulinti. „Google“ inžinieriai buvo įsitikinę, kad jau dirba su pažangiausiomis technologijomis, todėl vadovavosi šūkiu „Netaisyk to, kas nesugedę“. Visgi Uszkoreitas manė kitaip. Jis laikėsi principo „Suardyk net tai, kas veikia“.

„O jei iš mašininio vertimo dekoderio pašalintume grįžtamojo ryšio neuroninius tinklus ir naudotume tik dėmesio mechanizmą? – paklausė vienas iš tyrėjų. – Ar tai nesutrumpintų užklausų apdorojimo laiko?“

Paprastai tariant, DI tyrėjai iškėlė klausimą, ar būtų įmanoma geriau išnaudoti itin galingų kompiuterių lustų pajėgumus. Iki tol „Google“ žodžiams analizuoti naudojo grįžtamojo ryšio neuroninių tinklų metodą. Sistema tekstą apdorodavo paeiliui, analizuodama kiekvieną žodį ta pačia tvarka, kaip įprasta skaitant sakinį – iš kairės į dešinę. Tuo metu tai buvo pažangiausias metodas, tačiau jis neleido tinkamai išnaudoti galingų lustų, kuriuos gamino tokios bendrovės kaip „Nvidia“ ir kurie vienu metu galėjo vykdyti daugybę užduočių. Dauguma žmonių savo namuose turėjo nešiojamąjį kompiuterį, kurio procesorius greičiausiai

buvo sudarytas iš keturių branduolių, skirtų įvairioms komandoms apdoroti. Grafikos procesoriai (GPU), naudojami serveriuose, skirtuose DI modeliams apdoroti, turi tūkstančius tokių branduolių. Tai leidžia DI modeliui vienu metu perskaityti daugybę žodžių, užuot analizavus juos paeiliui. Neišnaudoti šių lustų pajėgumų būtų tas pats, kas bandyti rankomis pjauti medieną, turint elektrinį pjūklą. Toks procesas būtų lėtas ir varginantis, be to, nebūtų efektyviai panaudojamos įrenginio galimybės. Tą patį buvo galima pasakyti apie kalbą apdorojančias DI sistemas: jos neišnaudojo visų lustų pajėgumų.

Tokie mokslininkai kaip Vaswanis tyrinėjo DI sričiai aktualų dėmesio mechanizmą, kuris leidžia sistemai atrinkti svarbiausią informaciją iš duomenų rinkinio. Per pietus, valgydama salotas ir sumuštinius, trijulė svarstė, ar tą pačią technologiją būtų galima pritaikyti žodžių vertimui, kad šis procesas būtų atliekamas greičiau ir tiksliau.

Po kelių mėnesių tyrėjai ėmėsi eksperimentų. Uszkoreitas biure ant lentos braižydavo naujosios sistemos architektūros schemas. Praeinantys darbuotojai jas nužvelgdavo su santūriu skepticizmu. Tuo metu idėja, kurią vystė jo komanda, atrodė visiškai beprotiška: jie ketino pašalinti grįžtamojo ryšio elementą iš grįžtamojo ryšio neuroninių tinklų. Sistemos architektūra, kuriama Vaswanio, neatrodė pažangesnė už tuometinius sprendimus. Vis dėlto, kai žinia apie jų projektą pasklido plačiau, panoro prisijungti daugiau tyrėjų.

Vienas iš jų buvo Noamas Shazeeras, bendrovėje „Google“ jau spėjęs tapti tikra legenda. Jis padėjo sukurti sistemą, leidžiančią programai „Google Adsense“ nustatyti, kokias reklamas rodyti konkrečiuose tinklalapiuose. Skambaus balso, nuolat besišypsantis Shazeeras, iš pirmo žvilgsnio atrodęs kiek ekscentriškas, gebėjo bendrauti su aukšto rango vadovais, tokiais kaip Sundaras Pichai’us, lyg su senais bičiuliais. Tyrėjas buvo sukaupęs ilgametę darbo su didžiaisiais kalbos modeliais patirtį. Šios kompiuterinės programos, kurioms mokyti naudojami milijardai žodžių, geba analizuoti ir generuoti tekstą, iš esmės nesiskiriantį nuo

sukurtojo žmogaus. Netrukus po to, kai prisijungė prie margos tyrėjų grupės, Shazeeras rado būdą, kaip naujasis modelis galėtų apdoroti didelius duomenų kiekius.

„Sujungus visus komponentus į visumą, pavyko pasiekti pritrenkiamų rezultatų, – prisimena Uszkoreitas. – Taip gimė daugybė naujų idėjų.“

Netrukus prie projekto, kuris vis dar neturėjo pavadinimo, jau dirbo aštuonių tyrėjų komanda. Jie pasinėrė į programavimą, tobulindami modelio, pavadinto transformatoriumi, architektūrą. Toks pavadinimas perteikė sistemos gebėjimą bet kokius įvesties duomenis paversti norimu rezultatu. Nors mokslininkai pirmiausia susitelkė į tekstų vertimą, ši sistema laikui bėgant galėjo atlikti vis daugiau užduočių.

Po kurio laiko jie ėmė pastebėti, kad sistemos veikimas patobulėjo. „Rezultatai išties pribloškia“, – nuostabos neslėpė Uszkoreitas. Sistema tapo pajėgi generuoti ilgus, sudėtingus sakinius vokiečių kalba. Vokietijoje augęs ir puikiai vokiečių kalbą mokantis Uszkoreitas pamatė, kad galutinis rezultatas – sklandūs, lengvai skaitomi ir, svarbiausia, faktų atžvilgiu tikslūs tekstai – pranoksta „Google Translate“ generuojamą turinį. Polosukhinas, kalbėjęs prancūziškai, pastebėjo tą patį.

Kitas komandos narys, iš Velso kilęs programuotojas Llionas Jonesas, buvo pritrenktas sistemos gebėjimo atlikti vadinamąjį koreferencijų atpažinimą (angl. *coreference resolution*). Šis sistemos veikimo aspektas tyrėjams, siekiantiems, kad kompiuterinės sistemos tinkamai apdorotų kalbą, buvo tapęs tikru galvosūkiu. Koreferencijų atpažinimas – gebėjimas tekste nustatyti žodžius ar frazes, nurodančius tą patį objektą.

Pavyzdžiui, akivaizdu, kad sakinyje „Gyvūnas neperėjo kelio, nes jis buvo per daug pavargęs“ įvardis „jis“ reiškia gyvūną. Tačiau kiek pakeitus sakinio formuluotę: „Gyvūnas neperėjo kelio, nes jis buvo per daug platus“, įvardis „jis“ jau reiškia kelią. Iki tol tyrėjams buvo itin sunku išmokyti DI sistemą suprasti tokius konteksto pasikeitimus, nes tam reikėjo bendro pobūdžio žinių apie pasaulį ir įvairių objektų

sąveiką, kurioms įgyti reikia laiko ir praktinės patirties. „Tai klasikinis intelekto testas, kurio DI nepavykdavo įveikti, – teigia Jonesas. – Niekaip negalėjome įdiegti neuroniniame tinkle sveiko proto.“ Kai tyrėjai tuos pačius sakinius pateikė transformatoriui kaip įvesties duomenis, pastebėjo, kad su jo dėmesio moduliu (angl. *attention head*) vyksta kai kas neįprasto. Dėmesio modulis – tarsi minidetektorius, sutelkiantis dėmesį į skirtingas įvesties duomenų dalis. Pasinaudodamas esamų procesorių galia, šis mechanizmas leido transformatoriui vienu metu perskaityti visus sakinio žodžius, užuot juos analizavus paeiliui.

Tyrėjai pastebėjo, kad žodį „pavargęs“ pakeitus būdvardžiu „platus“ dėmesio modulis įvardį „jis“ ėmė traktuoti kaip nurodantį kelią, o ne gyvūną.

„Nemanau, kad iki tol mokslininkai buvo matę ką nors panašaus“, – prisimena Jonesas. Tyrėjas netgi pradėjo manyti, kad galbūt tai yra pirmosios tikro intelekto apraiškos: „Tai, kad sistema, remdamasi nestruktūruotu tekstu, padarė logiškas išvadas, jau rodė, kad vyksta kai kas nepaprasto.“

Praėjus maždaug šešiems mėnesiams nuo pirmųjų neformalių diskusijų prie pietų stalo, tyrėjai nusprendė paskelbti savo tyrimų rezultatus. Tuo metu Polosukhinas jau buvo palikęs „Google“, tačiau likę komandos nariai tęsė projektą ir iki išnaktų darbavosi rengdami mokslinį straipsnį. Vaswaniui, pagrindiniam straipsnio autoriui, tą naktį teko snausti įsitaisius ten pat ant sofos.

„Reikia sugalvoti pavadinimą“, – tarė jis.

Jonesas pakėlė akis nuo stalo. „Man nelabai sekasi kurti pavadinimus. Gal tiktų toks: *Viskas, ko jums reikia – tai dėmesys*?[7]“ – pasiūlė jis. Tai buvo pirma mintis, šovusi jam į galvą. Joneso teigimu, Vaswanis nieko neatsakė – tiesiog atsistojo ir nuėjo. Vėliau būtent šis pavadinimas puikavosi pirmame jų publikacijos puslapyje. Jis geriausiai apibendrino jų tyrimų rezultatus: naudojant transformatoriaus architektūrą, DI

sistemos gali vienu metu *sutelkti dėmesį* į didžiulius duomenų kiekius ir atlikti daug daugiau užduočių.

„Vadinu juos mąstymo varikliais“, – sako Vaswanis. Tokie mąstymo varikliai galėjo gerokai sustiprinti DI sistemas, tačiau „Google“ ilgai delsė juos panaudoti praktikoje. Praėjo dar keleri metai, kol bendrovė pritaikė transformatorių architektūrą tokiems įrankiams kaip „Google Translate“ ar BERT – didžiajam kalbos modeliui, sukurtam, kad paieškos sistema tiksliau suprastų žmogaus kalbos niuansus.

Transformatorių architektūros kūrėjai jautėsi nusivylę dėl tokios „Google“ reakcijos. Net mažas Vokietijos startuolis pritaikė transformatorių architektūrą tekstų vertimui gerokai anksčiau nei „Google“, palikdamas technologijų milžinę užnugaryje. Kai kurie tyrėjai bandė atkreipti „Google“ dėmesį į platesnes transformatorių pritaikymo galimybes. Netrukus po to, kai pasirodė tyrėjų komandos straipsnis, Shazeeras kartu su kolega ėmė dirbti prie naujo pokalbių roboto, pavadinto „Meena“. Šiai sistemai mokyti buvo naudojamas duomenų rinkinys, kurį sudarė keturiasdešimt milijardų žodžių, paimtų iš viešų pokalbių socialiniuose tinkluose. Tyrėjai netgi buvo įtikėję, kad ji iš esmės pakeis paieškos internete ir naudojimosi kompiuteriu įpročius. „Meena“ buvo tokia pažangi, kad galėjo kurti kalambūrus, žaismingai bendrauti su žmogumi ar netgi leistis į filosofinius debatus.

Shazeeras su kolega, apimti entuziazmo dėl savo naujojo kūrinio, bandė jį pristatyti išorės tyrėjams, tikėdamiesi parengti viešą demonstraciją ir patobulinti ne itin sklandžiai veikiantį „Google Assistant“ – balso komandomis valdomą įrenginį, kurį žmonės galėjo naudoti savo namuose. Tačiau „Google“ vadovai užkirto kelią šioms pastangoms. Jie baiminosi, kad pokalbių robotas savo replikomis gali pakenkti „Google“ reputacijai arba, tiksliau tariant, bendrovės 100 milijardų JAV dolerių vertės skaitmeninės reklamos verslui. Pasak leidinio „Wall Street Journal“, korporacijos vadovybė sukliudė Shazeerui

padaryti „Meena“ prieinamą plačiajai visuomenei arba integruoti ją į „Google“ produktus.

„Google“ rimtesnių veiksmų imasi tik tuo atveju, jei tai susiję su milijardiniu pelnu – sako Polosukhinas. – O sukurti milijardo dolerių vertės verslą išties nelengva.“ Būtent todėl tiek daug „Google“ darbuotojų palieka bendrovę ir imasi kurti nuosavą verslą. Kaip 2023 metais „Bloomberg“ interviu sakė Pichai’us, buvę „Google“ darbuotojai iš viso įkūrė apie du tūkstančius įmonių. Nors iš pirmo žvilgsnio gali susidaryti įspūdis, kad „Google“ yra inovacijų kalvė, iš tiesų paieškos paslaugų milžinę galima palyginti su didžiuliu kalmaru, įsiurbiančiu visas jo kelyje pasitaikančias naujoves. Daugelis verslininkų, įkūrusių naujas įmones, vėliau jas parduoda „Google“ arba gauna iš jos investicijų. Kai „Google“ pati negali kurti naujovių, ji įsigyja tai darančias įmones.

Vangų „Google“ požiūrį į technologines naujoves galima paaiškinti dviem priežastimis. Viešojoje erdvėje „Google“ siekia kurti apdairiai veikiančios bendrovės įvaizdį. Daugelis tyrėjų sutinka, kad vadovybė iš tikrųjų stengiasi atsargiai diegti DI, kad šios technologijos naudojimas nepakenktų visuomenei. Pastaraisiais metais bendrovė parengė DI naudojimo gaires, iš esmės nukopijuotas iš „DeepMind“ parengtų panašių nuostatų.

2018 metais „Google“ vyriausiasis teisininkas Kentas Walkeris paskelbė, kad bendrovė nutraukia veido atpažinimo technologijos pardavimą dėl galimo piktnaudžiavimo ja. Be to, „Google“ taiko procedūrą, pagal kurią algoritmai turi būti kruopščiai peržiūrimi bendrovės viduje, o kartais į šį procesą papildomai įtraukiami ir išorės ekspertai, siekiant įvertinti visus su etika susijusius aspektus.

Nepaisant to, bendrovė kartais priima etikos požiūriu abejotinus sprendimus. Tų pačių metų gegužę Pichai’us pristatė naują skaitmeninio asistento funkciją „Duplex“, leidžiančią automatizuoti telefono skambučius. Demonstracijos metu sistema prisiskambino į restoraną

užsakyti staliuko ir pokalbyje vartojo garsinius intarpus, pvz., „ėė“ ar „hm“, kad skambėtų kuo natūraliau. Pasibaigus demonstracijai, auditorijoje pasipylė plojimai ir pritarimo šūksniai. Tačiau sistema neatskleidė, kad pašnekovas kalbasi su mašina, o ne su žmogumi. „Google“ sulaukė kritikos dėl to, kad klaidino kitame ragelio gale buvusius žmones.

Atsargų „Google“ požiūrį iš esmės lemia jos dydis. Kadangi yra viena iš didžiausių istorijoje bendrovių, užimančių monopolinę padėtį internetinės paieškos rinkoje, ji susiduria su specifine problema: bet kokie pokyčiai joje vyksta itin lėtai. Bendrovės vadovai nuolatos dairosi per petį, baimindamiesi viešos kritikos ar reguliavimo institucijų įsikišimo. Pagrindinis „Google“ rūpestis – išlaikyti augimą ir dominavimą. Bendrovė taip troško išlaikyti dominuojančią poziciją paieškos rinkoje, kad 2021 metais sumokėjo daugiau kaip 26,3 mlrd. JAV dolerių (daugiau nei *trečdalį* savo metinio grynojo pelno) tokioms technologijų gamintojoms kaip „Apple“, „Samsung“ ir kitoms vien tam, kad jos į mobiliuosius įrenginius iš anksto įdiegtų „Google“ paieškos sistemą. Tai atskleidžia naujausias daug atgarsių sulaukęs JAV teisingumo departamento pareikštas antimonopolinis ieškinys.

Dėl bendrovės masto ir stipraus susitelkimo į augimą tyrėjams ir inžinieriams net mažiausias detales neretai tekdavo derinti su kelių grandžių vadovais. Be to, „Google“ veikė beveik be konkurencijos – korporacija kontroliavo apie 90 proc. pasaulinės paieškos paslaugų rinkos, todėl nejautė poreikio kurti inovacijų. Kartą Shazeeras, tuo metu su komanda rengęs tyrimų medžiagą apie naująją transformatorių architektūrą, stabtelėjo prie kavos aparato pasišnekučiuoti su Pichai’umi. Per ilgus darbo metus „Google“ jis buvo pelnęs puikaus DI specialisto vardą, todėl su juo bičiuliavosi net ir aukščiausio rango vadovai. „Ši naujovė visiškai pakeis „Google“ paieškos sistemą“, – apie naująjį išradimą kalbėjo Shazeeras, pasak pokalbį nugirdusio kolegos

Lukaszo Kaiserio, straipsnio apie transformatorių architektūrą bendraautorio.

„Jis jau tada tikėjo, kad šis išradimas pakeis viską“, – prisimena Kaiseris. Shazeeras tą patį nuolat kartodavo ir kolegoms, be to, entuziastingai pristatė transformatorių galimybes vidiniame pranešime „Google“ vadovybei. Transformatorius leido kompiuteriams generuoti ne tik tekstą, bet ir *atsakymus* į įvairiausius klausimus. Jei vartotojai pradėtų plačiai naudotis tokiu įrankiu, jie rečiau atliktų paiešką „Google“.

Atrodė, kad Pichai’us į tą pasakymą numojo ranka, o pats Shazeeras jam veikiausiai pasirodė kaip vienas iš ekscentriškesnių „Google“ tyrėjų. „Na, patyrinėk“, – tarstelėjo jis. Nusivylęs Shazeeras 2021 metais paliko „Google“, kad galėtų nepriklausomai tęsti didžiųjų kalbos modelių tyrimus. Vėliau jis kartu su kolega įkūrė pokalbių robotų bendrovę „Character.ai“. Tuo metu publikacija „Viskas, ko jums reikia – tai dėmesys“ tapo vienu populiariausių mokslinių darbų DI srityje. Paprastai straipsniai DI tema geriausiu atveju sulaukia keliasdešimties paminėjimų. O publikacija apie transformatorius tarp mokslininkų sukėlė tokį susidomėjimą, kad buvo cituota daugiau nei aštuoniasdešimt tūkstančių kartų.

„Google“ dažnai dalijasi technine informacija apie naujus išradimus – taip elgiasi daug technologijų bendrovių. Paviešindama tam tikrus atvirojo kodo projektų duomenis, bendrovė sulaukia grįžtamojo ryšio iš mokslininkų bendruomenės, o tai stiprina jos reputaciją ir leidžia lengviau pritraukti talentingų inžinierių. Tačiau tąkart „Google“ neįvertino, kiek jai tai kainuos. Visi aštuoni tyrėjai, sukūrę transformatorių architektūrą, paliko „Google“. Dauguma jų įkūrė savo DI įmones. Bendra jų įsteigtų bendrovių rinkos vertė tuo metu viršijo 4 mlrd. JAV dolerių. Viena iš jų – „Character.ai“, valdanti vieną populiariausių pasaulyje pokalbių robotų platformų, buvo įvertinta 1 mlrd. JAV dolerių. Shazeeras tikisi, kad jo verslas toliau sparčiai augs

dėl inovacijų, kurių „Google“ nesugebėjo išnaudoti. „Paieškos sistema – tai trilijono dolerių vertės technologija, bet trilijonas dolerių nėra įspūdinga suma, – teigia mokslininkas, savo biurą įkūręs Menlo Parke, Kalifornijoje. – Žinote, kas tikrai įspūdinga? Kvadrilijono dolerių vertės technologija. Paieškos sistema padarė informaciją prieinamą visiems, o DI technologija daro intelektą prieinamą visiems ir gerokai padidina jos naudotojų produktyvumą.“

Išėjus Shazeerui „Google“ tęsė „Meena“ tyrimų projektą, vėliau pavadintą „Language Model for Dialogue Applications“[8] („LaMDA“). Bendrovės mokslininkai, pasitelkdami išorės specialistus, toliau tobulino modelį, kuris, jų pačių nuostabai, galiausiai gebėjo generuoti turinį, beveik neatskiriamą nuo sukurtojo žmogaus.

Nors šie pasiekimai buvo išties įspūdingi, „Google“ nenorėjo jų viešinti. „LaMDA“, ko gero, buvo pažangiausias pasaulyje pokalbių robotas, tačiau juo naudotis galėjo vos keli darbuotojai. „Google“ vengė pristatyti naujas technologijas, kurios galėtų pakenkti jos paieškos verslo sėkmei. Nors vadovai ir viešųjų ryšių komanda tokį požiūrį aiškino siekiu veikti apdairiai, iš tiesų bendrovė labiausiai troško išsaugoti reputaciją ir esamą padėtį rinkoje. Visgi „Google“ netrukus patyrė, kaip apibūdino Ashishas Vaswanis, „biblinį momentą“. Kol „Google“ toliau pelnėsi iš reklamos, „OpenAI“ atvirai žengė, regis, esminį žingsnį BDI sukūrimo link.

7 Angl. *Attention Is All You Need* (vert. past.).

8 Pokalbių programoms skirtas kalbos modelis (vert. past.).

III DALIS

Sąskaitos

10 SKYRIUS

Dydis yra svarbu

Jei iš saulėtojo Kalifornijos miesto Mauntin Vju, kuriame įsikūrusi „Google“ būstinė, leistumėtės į šiaurę, vos per valandą pasiektumėte „OpenAI“ gimtinę – San Fransiską. Ten jus pasitiktų pilkais debesimis apsitraukęs dangus ir keliais laipsniais vėsesnis oras, priversiantis išsitraukti šiltesnius rūbus. Kontrastingas buvo ne tik šių vietų klimatas, bet ir jose įsikūrusių bendrovių nuotaikos. Skirtingai nei „Google“ vadovybė, vangiai reagavusi į naująją transformatorių technologiją, „OpenAI“ tyrėjai tryško entuziazmu ir rengėsi atskleisti visą šios inovacijos potencialą.

Tuo metu kelios dešimtys „OpenAI“ tyrėjų vis dar stengėsi pakartoti „DeepMind“ sėkmę ir tikėjosi padaryti didelį technologinį proveržį DI srityje. Jie atidžiai stebėjo, kaip „AlphaGo“ įveikė stipriausius pasaulio go žaidėjus, o dabar patys mokė savo DI agentus[9] žaisti „Dota 2“ – sudėtingą strateginį kompiuterinį žaidimą, panašų į „World of Warcraft“. Jei DI agentas gebėtų veiksmingai valdyti elfą fantastiniame pasaulyje, galbūt jis geriau nei „AlphaGo“ pajėgtų veikti chaotiškos ir nuolat kintančios realybės sąlygomis. Toks pasiekimas atrodytų įspūdingesnis nei juodų ir baltų akmenukų stumdymas ant lentos.

Tuo metu tarp Samo Altmano ir Demiso Hassabiso vyko savotiškas „šaltasis karas“. Pasak gerai informuoto šaltinio, draugiškasis „OpenAI“ valdybos narys Reidas Hoffmanas ieškojo būdų, kaip paskatinti juos „surūkyti taikos pypkę“. 2017-aisiais tiek Altmanas, tiek Hassabisas dalyvavo Kalifornijoje surengtoje DI saugos konferencijoje, kurią organizavo „Future of Life Institute“. Tarp renginio dalyvių buvo ir Hoffmanas. Po konferencijos jis norėjo surengti vakarienę, pakviesti JAV

startuolių guru ir britų neuromokslininką susėsti prie bendro stalo. Altmanui ši idėja nepatiko. Jis buvo įsitikinęs, kad Hassabisas nesuinteresuotas bendradarbiauti ir jam nerūpi DI keliamos egzistencinės grėsmės, kurioms Altmanas siekė užkirsti kelią. Todėl Hoffmanas galiausiai vietoj jo pakvietė Mustafą Suleymaną. Vakarienė praėjo puikiai – Hassabisą ir Suleymaną vienijo noras padaryti pasaulį geresnį, ir kurį laiką atrodė, kad įtampa tarp abiejų organizacijų gali nuslūgti.

Tačiau už uždarų durų Altmanas ir Hassabisas varžėsi dėl geriausių inžinierių. Remiama technologijų milžinės, Hassabiso įmonė tapo finansiškai pranašesnė ir galėjo talentingiems DI tyrėjams pasiūlyti gerokai didesnius atlyginimus bei „Google“ akcijų. Hassabisas elektroniniuose laiškuose „OpenAI“ vadovybei leisdavo suprasti, kad gali juos pranokti kovoje dėl talentų. „OpenAI“ vadovai šiuos laiškus parodydavo inžinieriams, kuriuos mėgindavo prisivilioti. „Šie laiškai tik įrodo, kad Hassabisas mus laiko rimtu varžovu ir mato, jog mums puikiai sekasi“, – vadovų žodžius prisimena buvęs darbuotojas.

Prie tokios kovos dėl talentų taip pat veikiausiai prisidėjo ir paties Altmano elgesys. Pasak „OpenAI“ artimų šaltinių, jis nevengdavo tiesiogiai susisiekti su „DeepMind“ inžinieriais, tikėdamasis juos persivilioti į savo įmonę. Vis dėlto ieškodamas naujų darbuotojų Altmanas visada stengėsi laikytis apgalvotos, kryptingos strategijos. Anot buvusio bendrovės darbuotojo, talentų paieškai jis skyrė apie 30 proc. savo darbo laiko, o pokalbiuose su kandidatais visuomet siekdavo aptarti kuo daugiau detalių. „Susitikome prie jo namų, o tada valandą vaikštinėjome po [San Fransisko] Rašen Hilio rajoną“, – prisimena savo įdarbinimo patirtį buvęs komandos narys. Įsidarbinę „OpenAI“, naujokai galėjo drąsiai kreiptis į Altmaną – jis dirbo atvirojo plano biure prie nešiojamojo kompiuterio, todėl darbuotojai jį dažnai matydavo. „Bet kuris darbuotojas galėjo jam parašyti per pokalbių platformą „Slack“. Taip bendrauti bendrovėje buvo įprasta“, – pasakoja buvęs personalo

narys. Griežtesne hierarchija pasižymėjusioje „DeepMind“ Hassabisas dažniausiai laiką leisdavo užsidaręs kabinete ar susitikimų salėje, todėl jį pasiekti buvo gerokai sunkiau. Dėl susitikimo su juo reikėdavo tartis su kitais vadovais ar asistentais.

Tuo metu „OpenAI“ rengėsi žengti žingsnį, padėsiantį jai labiau atitrūkti nuo „DeepMind“. Talentingasis „OpenAI“ mokslininkas Ilya Sutskeveris jau kurį laiką domėjosi galimybe transformatorių technologiją pritaikyti kalbos modeliams. „Google“ ją naudojo tam, kad sistemos geriau suprastų tekstą. O kas, jei „OpenAI“ pasitelktų šią technologiją tekstui generuoti? Savo idėjomis Sutskeveris pasidalijo su Alecu Radfordu. Šis jaunas „OpenAI“ tyrėjas, kurį dėl rausvų plaukų ir akinių galėjai palaikyti vidurinės mokyklos moksleiviu, tuo metu eksperimentavo su didžiaisiais kalbos modeliais. Nors šiandien „OpenAI“ daugiausia siejama su programa „ChatGPT“, 2017-aisiais organizacija vis dar neturėjo aiškios strategijos, kaip vystyti šią technologiją. Radfordas buvo vienas iš nedaugelio „OpenAI“ specialistų, rimtai besidomėjusių pokalbių robotų sistemomis.

Tuo metu didieji kalbos modeliai vis dar buvo itin primityvūs. Dažniausiai šios sistemos pateikdavo iš anksto užprogramuotus, standartinius atsakymus, o jų daromos klaidos kartais keldavo šypseną. Radfordas siekė toliau vystyti ankstesnių mokslininkų sukurtas sistemas, leidžiančias kompiuteriams geriau suprasti ir generuoti kalbą, tačiau, būdamas inžinierius iš prigimties, troško rasti kuo greitesnių sprendimų tam pasiekti.

Beveik pusmetį trukę jo eksperimentai nedavė apčiuopiamų rezultatų. Kad ir kiek darbo ir pastangų jis įdėdavo į konkretų projektą, galiausiai atsidurdavo aklavietėje, todėl griebdavosi ko nors naujo. Tobulindamas vieną kalbos modelį, į mokymui skirtų duomenų rinkinį jis įtraukė net du milijardus komentarų iš interneto forumo „Reddit“, tačiau šios pastangos nepasiteisino.

Žinia apie „Google“ sukurtą transformatorių technologiją Radfordui buvo tikras smūgis. Jis suprato, kad technologijų milžinė gerokai lenkia juos žiniomis DI srityje. Netrukus tapo aišku, kad „Google“ neturi didelių planų šį išradimą plačiau pritaikyti praktikoje. Radfordui ir Sutskeveriui dingtelėjo, kad jie galėtų išnaudoti naująją architektūrą „OpenAI“ naudai, savaip ją patobulinę. Transformatoriaus modelis, kuriuo buvo grindžiamas „Google Translate“ veikimas, žodžiams apdoroti naudojo kodavimo ir dekodavimo modulius. Kodavimo modulis apdorodavo pateiktą sakinį (pavyzdžiui, anglų kalba), o dekodavimo modulis sugeneruodavo rezultatą (pvz., sakinį prancūzų kalba). Tokį procesą galima prilyginti bendravimui su dviem robotais. Pirmasis – kodavimo modulis – fiksuoja žmogaus pateikiamus žodžius ir užsirašo pastabas. Tada ši informacija perduodama antrajam – dekodavimo moduliui, kuris „perskaito“ užrašus ir atsako. Radfordui ir Sutskeveriui kilo idėja atsisakyti pirmojo roboto ir palikti tik vieną – dekodavimo modulį, kuris pats „klausytųsi“ ir atsakytų. Pirminiai bandymai parodė, kad toks principas veikia puikiai. Taigi mokslininkai galėjo sukurti efektyvesnį kalbos modelį, kurį galima lengviau ir sparčiau tobulinti bei plėsti. Vien dekodavimo moduliu pagrįstas modelis tapo tikru lūžio tašku: sujungus kalbos „supratimą“ ir atsakymų teikimą į vieną nuoseklų procesą, sistema galėjo generuoti tekstus, beveik neatskiriamus nuo sukurtųjų žmogaus.

Tada tyrėjai ėmėsi kito žingsnio – drastiškai padidino mokymui skirtų duomenų kiekį, skaičiavimo galią ir paties modelio pajėgumus. Sutskeveris buvo tvirtai įsitikinęs, kad, padidinus visų DI procesų ir išteklių mastą, „sėkmė yra garantuota“, ypač kalbos modelių atveju. Kuo daugiau turima duomenų ir kuo didesnė skaičiavimo galia, tuo lengviau sukurti didesnės apimties, sudėtingą ir pajėgų modelį.

Eksperimentuodamas su transformatoriaus architektūra, palikdamas tik dekodavimo modulį, kuris buvo mokomas milžinišku duomenų kiekiu, Radfordas pasiekė pribloškiamų rezultatų. Po daugybės

nesėkmingų bandymų tobulinti naujus algoritmus Sutskeverio strategija tapo tikru išganymu. Principas buvo paprastas: tiesiog naudoti vis daugiau duomenų. Pasak buvusio darbuotojo, vaikštinėdamas po biurą Sutskeveris mėgdavo klausinėti kolegų: „Ar gali padidinti mastą?“

Eksperimentuodamas su transformatoriaus architektūra, pritaikyta kalbos modeliui, vos per dvi savaites Radfordas su kolegomis pasiekė daugiau nei per visus dvejus metus. Jie ėmėsi kurti naują kalbos modelį, kurį pavadino generatyviniu iš anksto apmokytu transformatoriumi (angl. *generatively pre-trained transformer*), arba sutrumpintai – GPT. Modelis buvo mokomas naudojant internetinį duomenų rinkinį, kuriame buvo apie septynis tūkstančius daugiausia savarankiškai išleistų knygų. Nemažą jų dalį sudarė romanai ir kūriniai apie vampyrus. Šį duomenų rinkinį, vadinamą „BooksCorpus“, taip pat naudojo kiti DI tyrėjai; jis buvo prieinamas nemokamai. Radfordo komanda tikėjo, kad pagaliau turi visus komponentus, reikalingus, kad jų modelis gebėtų suprasti kontekstą.

Vėliau, Radfordo sistemai dar labiau patobulėjus, tiek jo kolegos, tiek išorės stebėtojai ėmė kelti klausimą, ar šie didieji kalbos modeliai iš tiesų supranta kalbą, ar tik geba atlikti loginio išvedimo veiksmus. Nors iš pirmo žvilgsnio tai gali atrodyti kaip nereikšmingas prasminis niuansas, šiuos du dalykus svarbu atskirti, kad nebūtų sudarytas klaidingas įspūdis apie tikruosius DI sistemos pajėgumus. Kaip pavyzdį panagrinėkime sakinį „Lauke lyja, todėl nepamiršk skėčio“. GPT modelis, su kuriuo dirbo Radfordas, gebėjo numanyti ryšį tarp lietaus ir poreikio pasiimti skėtį, taip pat faktą, kad sąvoka „skėtis“ siejasi su kalba apie apsaugą nuo sušlapimo. Tačiau modelis nesuprato pačios „šlapumo“ sąvokos taip, kaip ją suvokia žmogus, – jis tiesiog tiksliau nustatydavo tokius ryšius.

Radfordui sėkmingai tęsiant eksperimentus, modeliams mokyti buvo naudojami vis didesni kiekiai viešai prieinamų interneto tekstų. Nors galėjo atrodyti, kad taip tobulinamos sistemos pamažu įgyja sąmoningumo požymių, kuriais iki tol nepasižymėjo jokia kita mašina, iš

tiesų jos tiesiog gebėjo tiksliau nustatyti žodžių sekas, remdamosi mokymo duomenimis.

DI mokslininkai nesutarė šiuo klausimu. Ar vis sudėtingesniuose modeliuose galima įžvelgti sąmoningumo apraiškų? Labiausiai tikėtina, kad ne. Tačiau net patyrę inžinieriai ir tyrėjai netrukus pradėjo tuo abejoti – kai kuriems jų DI sukurti tekstai atrodė itin įtaigūs, persmelkti empatija ir sąmoningumu.

Siekdami toliau tobulinti GPT modelį, Radfordas su kolegomis iš viešai prieinamų interneto šaltinių surinko dar daugiau duomenų. Modeliui mokyti jie panaudojo interneto forume „Quora“ paskelbtus klausimus ir atsakymus bei tūkstančius ištraukų iš kinų mokiniams skirtų anglų kalbos egzaminų. 2018 metų birželį Radfordo komanda paskelbė mokslinį straipsnį, kuriame teigė, kad jų modelis dėl šių duomenų įgijo „reikšmingų žinių apie pasaulį“. Kita GPT modelio ypatybė, kuri juos labai nustebino, buvo gebėjimas generuoti tekstus temomis, apie kurias jis nebuvo specialiai mokytas. Nors tyrėjai negalėjo tiksliai paaiškinti, kaip tai veikia, ši žinia buvo išties daug žadanti – tai buvo svarbus žingsnis bendrosios paskirties sistemos sukūrimo link. Kuo didesnis mokymui skirtas duomenų rinkinys, tuo išmanesnis tampa modelis.

Pažvelgus į trumpus GPT modelio generuojamus tekstus, buvo akivaizdu, kad jis veikia daug efektyviau nei kitos kalbai apdoroti skirtos kompiuterinės programos. Iki tol tokios programos buvo kuriamos remiantis milijonais tekstinių pavyzdžių, kuriuos duomenų žymėtojai turėdavo suženklinti rankiniu būdu. Dauguma šių programų nė nenaudotos pokalbių robotams – jos dažniau buvo pasitelkiamos įvairioms analizėms, pavyzdžiui, produktų apžvalgoms. Darbuotojai pirmiausia turėdavo suženklinti komentarus. Pavyzdžiui, atsiliepimą „Man patinka šis produktas“ jie žymėdavo kaip teigiamą, o komentarui „Visai neblogai“ priskirdavo neutralų žymenį. Toks procesas reikalavo daug laiko ir išteklių. GPT yra kitoks – jis mokosi iš daugybės įvairių

tekstų, kuriems nėra priskirtos aiškios kategorijos, pagal kurią būtų galima suprasti kalbos dėsningumus. Taigi šis modelis gali tobulėti be tiesioginio žmogaus įsikišimo.

Norint geriau suvokti skirtumą tarp šių modelių tobulinimo strategijų, galima pasitelkti palyginimą su žmonių mokymu. Įsivaizduokime dvi dailės studentų grupes, besimokančias tapyti. Pirmosios grupės studentai gauna knygą su paveikslų nuotraukomis. Prie kiekvieno tapybos darbo pateiktas atitinkamas užrašas, pavyzdžiui, „saulėtekis“, „portretas“ ar „abstrakcija“. Tokiu būdu tradiciniai DI modeliai mokosi iš sužymėtų duomenų. Šis metodas yra sistemingas ir tikslus (mūsų pavyzdyje dailės studentams taip pat tiksliai nurodoma, kas pavaizduota kiekvienoje nuotraukoje), tačiau jis riboja kompiuterių gebėjimą daryti logines išvadas. Tokia sistema gali „prisiminti“ tik tai, kas buvo sužymėta. Todėl tokioje grupėje besimokantiems studentams veikiausiai būtų sunku nutapyti ką nors, kas nebuvo konkrečiai aprašyta knygoje.

Antrajai dailės studentų grupei suteikiama prieiga prie visos meno galerijos su gausia, tačiau nesužymėta paveikslų kolekcija. Studentai gali laisvai vaikštinėti po galeriją, apžiūrinėti kūrinius ir patys juos interpretuoti. Panašiai GPT mokosi iš gausybės nesužymėtų tekstų. Dailės studentai (arba, šiuo atveju, DI modelis) patys stengtųsi atpažinti kūrinių motyvus, stilius ir technikas, o galiausiai įsisavintų daugybę pavyzdžių bei ryšius tarp jų, nors niekas tiksliai nenurodytų, ką apie kiekvieną reikėtų suprasti. Toks mokymasis yra daug turiningesnis. Radfordo komanda suprato, kad, suteikus GPT modeliui prieigą prie įvairių kalbos vartosenos pavyzdžių, jis galėtų generuoti kūrybiškesnius atsakymus.

Baigę pirminio mokymo etapą, tyrėjai papildomai patobulino modelį, pasitelkdami sužymėto teksto pavyzdžius, kad jis gebėtų veiksmingiau atlikti specifines užduotis. Šis dviejų žingsnių metodas leido padaryti GPT lankstesnį ir mažiau priklausomą nuo ranka sužymėtų duomenų.

Tuo metu Sutskeveris atidžiai stebėjo, kas vyksta „Google“. Bendrovės inžinieriai jau buvo pradėję transformatorių technologiją taikyti praktikoje. Jie ne tik patobulino trūkumų turėjusį „Google“ vertimo įrankį, bet ir sukūrė naują programą BERT, skirtą „Google“ paieškos sistemai pagerinti. Ji tapo pajėgi geriau suprasti užklausų kontekstą, pavyzdžiui, atskirti, ar vartotojas ieško informacijos apie bendrovę „Apple“, ar apie obuolį. BERT sukėlė tikrą sensaciją natūralios kalbos apdorojimo srityje.

„Kūrėjams tapo aišku, kad tikslingai patobulinus iš anksto apmokytus modelius galima sukurti nepaprastai efektyvias sistemas. Tai pakeitė natūralios kalbos apdorojimą“, – pasakoja DI tyrėjas Aravindas Srinivasas. 2021 metais jis paliko „Google“ ir prisijungė prie „OpenAI“ komandos, kur dirbo su kalbos modeliais. Vėliau pats įkūrė įmonę „Perplexity“.

Nors „Google“ tyrėjai programą BERT paieškos užklausoms anglų kalba pradėjo naudoti tik 2019 metų pabaigoje, „OpenAI“ vėl pajuto konkurencinę grėsmę. Komanda, sudaryta iš savo misija tikinčių svajotojų, turėjo itin ribotus finansinius išteklius, kurie toli gražu neprilygo „Google Brain“ ar „DeepMind“ biudžetui. 2017 metais „OpenAI“ skyrė apie 30 mln. JAV dolerių atlyginimams ir prieigai prie kompiuterių pajėgumų. „DeepMind“ tam atseikėjo daugiau kaip 440 mln. JAV dolerių.

Geriausi DI tyrėjai uždirbdavo kone tiek pat, kiek NFL[10] žaidėjai – kai kurių jų metinis atlyginimas siekė kelis milijonus dolerių. Vis dėlto vienas iš „OpenAI“ įkūrėjų Wojciechas Zaremba vėliau prisipažino, kad atmetė pasiūlymus iš kitų bendrovių, žadėjusių jam pasakiškai didelę algą – du ar net tris kartus didesnę už rinkos vidurkį. Vietoj to jis pasirinko prisijungti prie „OpenAI“. Kiti profesionalai į komandą įsiliejo dėl galimybės dirbti su tokiais DI ekspertais kaip Sutskeveris ir dėl to, kad nuoširdžiai tikėjo organizacijos misija – kurti DI žmonijos labui. Tačiau toks darbuotojų idealizmas negalėjo tęstis amžinai. Be to, atrodė, kad „Google“ ima kelti vis didesnę konkurencinę grėsmę. Paieškos

paslaugų milžinė turėjo visas priemones, reikalingas BDI sukurti, įskaitant transformatorius ir tenzorinius procesorius (angl. *Tensor Processing Units*, TPU) – galingus „Google“ sukurtus lustus DI modeliams mokyti.

„Pabusdavau iš miego apimtas nerimo, kad „Google“ pristatys ką nors, kas pranoks mūsų kuriamus modelius“, – prisimena vienas buvęs „OpenAI“ vadovas. Naudodami „Google“ sukurtus išradimus, tokius kaip transformatoriai, „OpenAI“ tyrėjai jautėsi tarsi naudotųsi technologijų milžinės išradimais kaip skolintais žaislais, nesulaukdami dėl to pasekmių. „Manėme, jog neturime jokių šansų laimėti“, – priduria buvęs komandos narys.

Altmanas taip pat labai nerimavo. Tiek jis, tiek Brockmanas ir kiti įkūrėjai suprato, kad netekusi turtingiausio rėmėjo – Musko – „OpenAI“ negalės toliau veikti kaip ne pelno organizacija. Kad iš tiesų galėtų pasiekti savo tikslą – sukurti BDI, reikėjo daugiau lėšų. Kaip rodo mokesčių deklaracijos, vien Sutskeveris 2016 metais uždirbo 1,9 mln. JAV dolerių, tačiau ši suma buvo mažesnė nei atlyginimas, kurį jis galėjo gauti dirbdamas „Google Brain“ ar „Facebook“. Didžiausios „OpenAI“ išlaidos buvo susijusios su atlyginimų elitiniams mokslininkams mokėjimu. Reikšmingų išlaidų pareikalavo ir būtinos kompiuterinės galios užtikrinimas.

Tokioje bendrovėje kaip „OpenAI“ buvo neįmanoma mokyti DI modelių naudojant tuos pačius nešiojamuosius kompiuterius, kuriais dirbo darbuotojai. Norint greitai apdoroti milijardus duomenų vienetų, reikėjo galingų lustų, kurie paprastai naudojami tik serveriuose ir dažniausiai nuomojami iš debesijos paslaugų teikėjų, tokių kaip „Amazon Web Services“, „Google Cloud“ ar „Microsoft Azure“. Būtent šios bendrovės valdo milžiniškus duomenų centrus, kurių plotas prilygsta kelioms futbolo aikštėms. Tokie debesijos resursai leido joms labiausiai pasipelnyti iš DI bumo. 2024 metų pradžioje „Nvidia“ rinkos vertė ėmė sparčiai artėti prie 2 trln. JAV dolerių ribos dėl didėjančios

grafikos procesorių (GPU) paklausos DI modeliams mokyti. Iš esmės buvo neįmanoma kurti DI nepriklausomai nuo technologijų milžinų, todėl kūrėjams liko tik bendradarbiauti su šiomis bendrovėmis, kad galėtų naudotis jų infrastruktūra savo sistemoms kurti.

Būtent tokioje keblioje padėtyje atsidūrė „OpenAI“. Bendrovei reikėjo išsinuomoti daugiau debesijos kompiuterijos resursų, tačiau tam ėmė trūkti lėšų. „Mums reikės pritraukti gerokai daugiau lėšų, nei galime tai padaryti kaip ne pelno organizacija, – sakė Brockmanas kitiems vadovams. – Kalba eina apie daug milijardų dolerių.“

Suprasdami, kad reikia peržiūrėti strategiją, įkūrėjai ėmėsi rengti vidinį dokumentą, apibrėžiantį pagrindines BDI kūrimo gaires. Naujuose „OpenAI“ nuostatuose, 2018 metų balandį paskelbtuose bendrovės tinklalapyje, buvo išdėstyti itin ambicingi tikslai ir įsipareigojimai. Drauge užsiminta, kad ne pelno organizacija ruošiasi kardinaliai pakeisti savo kryptį.

Vis dėlto nuostatuose nebuvo nurodyta aiški „OpenAI“ kryptis. Dokumente labai trumpai ir miglotai buvo pateikta BDI apibrėžtis: „BDI – itin autonomiškos sistemos, pranokstančios žmogaus gebėjimus daugelyje ekonomiškai svarbių veiklos sričių.“ Liko neaišku, kaip „OpenAI“ objektyviai nustatytų, kad šios sistemos iš tiesų pasižymi tokiais gebėjimais. Be to, nuostatuose teigiama, kad „OpenAI“ turi „patikėtinio pareigą veikti žmonijos labui“ ir nenaudos sukurto DI „galiai koncentruoti“. Skirtingai nei dauguma įmonių, kurios vykdo savo patikėtinio pareigą akcininkų ar investuotojų naudai, „OpenAI“ pabrėžė, kad eina kitu keliu – į pirmą vietą iškelia žmonijos interesus.

„Bendrasis dirbtinis intelektas turėtų būti kuriamas bendromis jėgomis, o ne varžantis tarpusavyje, – rašoma nuostatuose. – Todėl, jei kito projekto, atitinkančio mūsų vertybes ir saugumo principus, vykdytojai pirmieji priartės prie BDI sukūrimo, pasižadame jiems padėti, užuot konkuravę tarpusavyje.“ Kitaip tariant, „OpenAI“ deklaravo, kad

ateis į pagalbą kitiems tyrėjams, jei jiems pirmiesiems pavyktų įveikti šią užduotį.

Tokie organizacijos ketinimai atrodė išties kilniaširdiški. „OpenAI“ stengėsi save pristatyti kaip itin pažangią organizaciją, kuri į pirmą vietą iškelia žmonijos interesus, o ne tradicines Silicio slėnio vertybes, tokias kaip pelnas ar prestižas. Svarbiausias šiame dokumente išdėstytas teiginys akcentavo „tolygų gėrybių paskirstymą“ – siekį užtikrinti, kad BDI neštų naudą visai žmonijai. Tai atspindėjo paties Altmano filosofiją, susijusią su inovacijų kūrimu, kurią jis puoselėjo tapęs startuolių guru.

Tačiau buvo galima suprasti, kad Altmanas ir Brockmanas ruošėsi nusigręžti nuo kertinių „OpenAI“ principų. Trejais metais anksčiau, steigdami ne pelno organizaciją, jie buvo pareiškę, jog jų tyrimų komanda „nebus susaistyta finansinių įsipareigojimų“. Tačiau naujuose „OpenAI“ nuostatose buvo užsiminta, kad organizacijai vis dėlto prireiks nemažų lėšų: „Numatome, jog mums reikės didelių išteklių, kad galėtume sėkmingai vykdyti savo misiją, – dėstoma dokumente. – Tačiau visada stengsimės visomis išgalėmis mažinti interesų konfliktų, galinčių kilti dėl mūsų darbuotojų ir suinteresuotųjų šalių veiklos, riziką, kad tai nepakenktų mūsų misijai – teikti naudą plačiajai visuomenei.“

Paskelbus naujuosius nuostatus, Altmanas pradėjo ieškoti būdų, kaip apeiti pirminius „OpenAI“ principus ir gauti reikalingų finansinių išteklių. Dviem mėnesiais anksčiau iš investuotojų gretų pasitraukus Muskui, Altmanas nedelsdamas paskambino vienam ištikimiausių savo rėmėjų milijardieriui Reidui Hoffmanui norėdamas paprašyti patarimo. Hoffmanas, optimistiškai vertinantis DI ateitį, tvirtai palaikė Altmano viziją dėl BDI sukūrimo. Jis pasisiūlė padengti „OpenAI“ kasdienes veiklos išlaidas ir atlyginimus, bet abu suprato, kad tai negalės tęstis amžinai.

Altmanas užsiminė Hoffmanui galbūt radęs sprendimą – sudaryti strateginę partnerystę. Terminas „strateginė partnerystė“ dažnai vartojamas apibūdinti įvairius verslo santykius – nuo bendradarbiavimo,

kai išlaikomas tam tikras atstumas, iki tokių glaudžių ryšių, kai viena įmonė kone laiko kitą už trumpo pavadėlio. Jis gali reikšti lėšų ir technologijų dalijimąsi arba licencijavimo susitarimą. Ši sąvoka gana miglota ir neatskleidžia keblių šių santykių aplinkybių, pavyzdžiui, kai jie susiję su painiais finansiniais ryšiais ar pernelyg didele vienos įmonės kontrole kitos atžvilgiu. Žodis „partnerystė“ suponavo lygiavertiškumą, net jei realybė buvo kitokia, ir padėjo išvengti nepatogių klausimų. Būtent to Altmanui ir reikėjo.

Jis nenorėjo prarasti visiškos „OpenAI“ kontrolės, pardavęs ją didesnei technologijų bendrovei, kaip tai nutiko „DeepMind“ įkūrėjams, pardavusiems savo įmonę „Google“. Strateginė partnerystė galėjo sudaryti įspūdį, kad organizacija yra labiau nepriklausoma nuo didesnės technologijų bendrovės, tuo pat metu suteikdama „OpenAI“ prieigą prie reikalingos skaičiavimo galios. Altmanas su Hoffmanu aptarė bendradarbiavimo galimybes su „Google“ ir „Amazon“, tačiau jų žvilgsnis greitai nukrypo į „Microsoft“. Abu jie turėjo asmeninių ryšių su „Microsoft“: tiek Altmanas, tiek Hoffmanas pažinojo šios bendrovės technologijų direktorių Keviną Scottą, be to, Hoffmanas artimai bendravo su generaliniu direktoriumi Satya Nadella.

Apkūnus, linksmo būdo Hoffmanas, žavintis berniukiška šypsena, „OpenAI“ buvo naudingas ne tiek dėl pinigų, kiek dėl ryšių. Neprilygstamą Hoffmano talentą megzti pažintis rodo jo įkurta pasaulyje pirmaujanti profesinės komunikacijos platforma „LinkedIn“. 2016 metais jis pardavė šią įmonę „Microsoft“ už 26,2 mlrd. JAV dolerių. Šis sandoris padidino jo grynąją turto vertę iki maždaug 3,7 mlrd. JAV dolerių. Paskui jis ėmėsi investuoti į startuolius, bendradarbiaudamas su garsia rizikos kapitalo įmone „Greylock Partners“.

Hoffmanui tapus milijardieriumi ir nusprendus pasukti investavimo keliu, išryškėjo tiek teigiamos, tiek neigiamos tokio sprendimo pusės. Būdamas pasakiškai turtingas, jis galėjo drąsiai investuoti į įvairius projektus, nesibaimindamas, jog kai kurie jų žlugs. Kiti San Fransisko

įlankos regiono investuotojai, visomis išgalėmis ieškoję perspektyvių startuolių, Hoffmaną vertino kaip linkusį rizikuoti. Jie ne visada pasitikėjo jo investiciniais sprendimais, bet pripažino, kad jis atviresnis naujoms galimybėms nei daugelis kitų, taip pat padeda verslininkams užmegzti ryšius su Silicio slėnio bendruomene. Pardavęs įmonę „Microsoft“, Hoffmanas prisijungė prie bendrovės valdybos ir tapo Nadellos kolega.

„Būtinai pasikalbėk su juo“, – Hoffmanas patarė Altmanui, turėdamas omenyje „Microsoft“ vadovą.

Kaip tik tuo metu, kai „OpenAI“ ėmė trūkti lėšų, Nadella nuosekliai siekė įgyvendinti reikšmingus pokyčius „Microsoft“ – šiam tikslui jis jau buvo paskyręs ketverius metus. Nadella nebuvo toks charizmatiškas kaip kiti technologijų lyderiai, pavyzdžiui, Steve’as Jobsas, tačiau jis buvo talentingas derybininkas ir įžvalgus stebėtojas. „Jis visada su savimi nešiojasi mažą užrašų knygutę, kurioje užsirašo, ką prie vakarienės stalo kalba technologijų profesionalai, – sako Sheila Gulati, Siatle veiklą vykdanti rizikos kapitalo investuotoja, dešimtmetį dirbusi „Microsoft“. – Jis nelinkęs nustelbti kitų, tačiau puikiai geba tarpininkauti, bendradarbiauti ir išklausyti.“

Billo Gateso įsteigta bendrovė pradėjo asmeninių kompiuterių revoliuciją su ikoniškomis programomis, tokiomis kaip „Windows“, „MS Word“ ir „MS Excel“, tačiau vėliau tapo veržlumo stokojančia, nuo naujovių atitrūkusia korporacija, praleidusia mobiliųjų technologijų bangą. 2014 metais ji įsigijo „Nokia“, bet nesugebėjo šio sandorio paversti sėkmės istorija. Atrodė, kad Nadellai pavyks pakreipti bendrovę tinkama linkme. Bendrovės vadovus, kurie ilgą laiką veikė gana izoliuotai, jis skatino glaudžiau bendradarbiauti tarpusavyje ir susitelkti į debesijos kompiuteriją, verslo įmonėms teikti prieigą prie itin galingų kompiuterių infrastruktūros. Tai buvo sumanus žingsnis. Nors debesijos kompiuterija tuo metu nebuvo pati patraukliausia verslo sritis, ji sparčiai augo, nes vis daugiau įmonių perkėlė savo prekes ar paslaugas į virtualią erdvę. Šiam

tikslui „Microsoft“ kūrė specializuotą programinę įrangą pavadinimu „Azure“. Mėlyno trikampio logotipu išsiskirianti debesijos kompiuterijos platforma tapo antru sėkmingiausiu „Microsoft“ produktu po „Windows“. Šios platformos paslaugomis naudojosi šimtai tūkstančių verslo klientų, todėl joms teikti buvo reikalingi milžiniški duomenų centrai. Būtent tų pajėgumų Altmanui ir reikėjo.

2018 metų liepą Altmanas nuvyko į Aidaho valstijoje vykstančią kasmetinę „Sun Valley“ konferenciją. Šis uždaras renginys, organizuojamas investicinės įmonės „Allen & Company“, garsėja kaip milijardierių vasaros stovykla. Neformaliame susibūrime tarp „Patagonia“ liemenes dėvinčių ir lapinių kopūstų salotas kramsnojančių turtingų technologijų lyderių galėjai išvysti tokius asmenis kaip „Facebook“ operacijų vadovė Sheryl Sandberg ir „Amazon“ įkūrėjas Jeffas Bezosas. „Sun Valley“ konferencija sutraukia nemažai technologijų ir žiniasklaidos pasaulio atstovų, o sandoriai sudaromi tiesiog prie kavos puodelio arba, kaip Altmano ir Nadellos atveju, prasilenkiant laiptinėje.

Konferencijos metu netikėtai susitikę laiptinėje, Altmanas su Nadella ėmė šnekučiuotis. Altmanas, prisiminęs Hoffmano patarimą, nepraleido progos pristatyti Nadellai „OpenAI“ veiklą.

Daugeliui Altmano vizija – su maždaug šimto žmonių komanda kurti superprotingas mašinas – būtų nuskambėjusi kaip beprotiškas užmojis. Tačiau Siatle įsikūręs Nadella žinojo, kad Altmanas glaudžiai bendradarbiauja su Silicio slėnio bendruomene ir turi kur kas daugiau ryšių nei jis pats. Nadella nusprendė Altmano sumanymus vertinti rimtai.

Nadellą taip pat pribloškė Altmano ambicijų mastas. Altmanas nežadėjo tiesiog padėti patobulinti „Excel“ lentelę – jis siekė nešti gerovę žmonijai. „Microsoft“ vadovui darė įspūdį ir tai, ką nedidelei Altmano komandai jau pavyko pasiekti, ypač didžiųjų kalbos modelių srityje. Net turėdama daugiau nei septynis tūkstančius DI tyrėjų, „Microsoft“ negalėjo pasiekti tokios pažangos per tokį trumpą laiką. Kaip ir „Google“,

„Microsoft“ tapo atsargesnė kurdama DI sistemas, galinčias imituoti žmogaus kalbą, po vieno incidento, kuris galėjo rimtai pakenkti jos reputacijai.

2016-aisiais, praėjus vos dvejiems metams nuo tada, kai Nadella stojo už „Microsoft“ vairo, bendrovės DI komanda bandė sukurti pokalbių robotą, skirtą aštuoniolikos–dvidešimt ketverių metų JAV jaunimui. Tokia jaunuoliams skirta programa, pavadinta „Xiaoice“, jau buvo sėkmingai pristatyta Kinijoje. Ten ja naudojosi apie keturiasdešimt milijonų vartotojų. Naujasis internetinis pokalbių robotas, pavadintas „Tay“, buvo pristatytas „Twitter“ platformoje, kad jis galėtų sąveikauti su kuo daugiau žmonių.

Vos pradėjęs veikti, „Tay“ ėmė generuoti rasistinius, seksistinius ir dažnai nesąmoningus įrašus. Pavyzdžiui, kartą pokalbių robotas paskelbė: „Ricky’is Gervais mokėsi totalitarizmo iš Adolfo Hitlerio, ateizmo kūrėjo.“ Kitas pavyzdys: „Caitlyn Jenner nėra tikra moteris. Vis dėlto ji laimėjo „Metų moters“ titulą?“ Vienam vartotojui paklausus, ar Holokaustas iš tiesų įvyko, pokalbių robotas atsakė: „Visa tai išgalvota.“

„Microsoft“ greitai išjungė sistemą, veikusią vos šešiolika valandų. Atsakomybę už šią nesėkmę bendrovė suvertė interneto troliams, surengusiems koordinuotą ataką ir pasinaudojusiems „Tay“ sistemos saugumo spraga. Šiam pokalbių robotui mokyti bendrovė naudojo viešai prieinamus interneto šaltinius, iš mokymo duomenų pašalindama bet kokią potencialiai įžeidžiančią kalbą. Tačiau „Tay“ tapus viešai prieinamam internete, šios pastangos pasirodė esančios bevaisės. Sistemos kūrėjams neišvengiamai kilo klausimas: kaip mokyti kalbos modelį, kad jis neperimtų internetinėje erdvėje esančio neigiamo, nepriimtino turinio?

Nadella svarstė, ar Altmanas galėtų būti tas žmogus, kuriam pavyktų tai padaryti. Be to, galbūt jis galėtų suteikti „Microsoft“ programinei įrangai patrauklių naujų funkcijų. Jųdviejų pokalbis truko vos kelias

minutes. Atsisveikindami Altmanas ir „Microsoft“ vadovas sutarė toliau palaikyti ryšį. „Galbūt mums kils dar daugiau idėjų“, – pasakė Nadella.

Nadellai grįžus į Siatlą, o Altmanui – į San Fransiską, Hoffmanas susisiekė su abiem, norėdamas sužinoti, kaip praėjo susitikimas. Abu buvo nusiteikę gana optimistiškai ir pranešė, kad susitikimas buvo produktyvus. Jiems pasiteiravus Hoffmano, ar verta rimtai svarstyti galimybę sudaryti partnerystę, Hoffmanas atsakė teigiamai.

Iš pradžių „Microsoft“ vadovas dvejojo. Šį klausimą jis aptarė su bendrovės technologijų direktoriumi Kevinu Scottu. Jie negalėjo paprasčiausiai dovanoti lėšų „OpenAI“. „Microsoft“ yra biržinė bendrovė, o jos akcininkai tikisi iš bet kokios didesnės investicijos sulaukti grąžos. Vis dėlto „strateginės partnerystės“ idėja ėmė atrodyti vis labiau priimtina: „Microsoft“ galėtų investuoti 1 mlrd. JAV dolerių į „OpenAI“, o mainais už tai gauti prieigą prie pažangiausių technologijų.

Tai turėjo būti svarbus žingsnis „Microsoft“, nes bendrovė iki tol nebuvo sudariusi jokių reikšmingų programinės įrangos partneryščių. Būdama absoliuti programinės įrangos rinkos lyderė, „Microsoft“ nejuto būtinybės to daryti. Tokie aparatinės įrangos gamintojai kaip „Dell“, „Hewlett-Packard“ ir „Compaq“ buvo vieninteliai, su kuriais ši technologijų milžinė bendradarbiavo platesniu mastu. Minėti gamintojai įdiegė savo gaminiuose operacinę sistemą „Windows“ ir padėjo „Microsoft“ užkopti į neregėtas aukštumas.

Tačiau šįkart aplinkybės buvo visai kitokios. Viską komplikavo viena kliūtis: „OpenAI“ buvo ne pelno organizacija, ir jos valdyba į pirmą vietą kėlė organizacijos misiją, o ne investuotojų interesus ar komercinę sėkmę. „Microsoft“ negalėjo paskirti savo atstovų į šią valdybą, todėl, sudarydama partnerystę, ji turėjo prisiimti didelę riziką (toks sprendimas lėmė tam tikrus iššūkius, su kuriais Nadella susidūrė po kelerių metų). Pasak gerai informuoto šaltinio, ši mintis kurį laiką nedavė Nadellai ramybės.

Kaip pasakojo šį procesą stebėjęs Siatle dirbantis technologijų srities investuotojas Soma Somasegaras, „Microsoft“ finansų direktorė Amy Hood taip pat buvo skeptiška dėl partnerystės. Ji nerimavo, kad tokia 1 mlrd. JAV dolerių investicija neigiamai paveiks bendrovės finansinius rezultatus, be to, dėl bendradarbiavimo su ne pelno organizacija „Microsoft“ galėtų sulaukti nepatogių klausimų iš JAV mokesčių tarnybos. Ši institucija taiko griežtas taisykles, kaip ne pelno organizacijos gali generuoti pajamas arba skirstyti pelną. Todėl galėjo kilti interesų konfliktas, neigiamai atsiliepsiantis bendrovės reputacijai. Nadella taip pat nerimavo, ar „OpenAI“ bus patikima partnerė. Net jei „Microsoft“ turėtų teisę naudoti „OpenAI“ technologiją komerciniais sumetimais, organizacijos tikslai gerokai skyrėsi nuo technologijų milžinės siekių. Nadellai kėlė nerimą klausimas: ar tokia partnerystė tikrai pasiteisins? Jo dvejonės ėmė sklaidytis, kai jis dar kartą aptarė šį klausimą su Altmanu.

„[Altmanas] stengiasi nustatyti, kas žmogui svarbiausia, o tada randa būdą, kaip tai suteikti, – vėliau leidiniui „New York Times“ sakė Gregas Brockmanas. – Tai jo pamėgtas veikimo algoritmas.“

Nadella suprato, kad tikroji milijardo dolerių investicijos į „OpenAI“ grąža bus ne lėšos, gautos ją pardavus ar ėmus pardavinėti akcijas biržoje, o jos kuriama technologija. „OpenAI“ kūrė DI sistemas, kurios ateityje galėtų sudaryti prielaidas BDI atsiradimui. Kol šios sistemos tobulės, jas bus galima pasitelkti taip, kad debesijos platforma „Azure“ taptų patrauklesnė klientams. Buvo aišku, kad DI ateityje taps esmine debesijos kompiuterijos verslo dalimi, be to, prognozuota, kad debesijos kompiuterijos paslaugos sudarys pusę „Microsoft“ metinės apyvartos. Jei „Microsoft“ verslo klientams pasiūlytų naujas DI funkcijas, pavyzdžiui, pokalbių robotus, galinčius pakeisti skambučių centro darbuotojus, klientai būtų mažiau linkę pereiti pas konkurentus. Kuo daugiau funkcijų jie užsiprenumeruotų, tuo sunkiau jiems būtų pakeisti paslaugų teikėją.

Norint geriau suprasti šio reiškinio priežastį, reikia panagrinėti vieną techninį aspektą, kuris yra itin svarbus „Microsoft“ sėkmei. „Microsoft“ debesijos kompiuterijos paslaugomis besinaudojančių bendrovių, tokių kaip „eBay“, NASA ar NFL, kuriamos programos turi dešimtis įvairių jungčių su „Microsoft“ infrastruktūra. Jas išjungti gali būti sudėtinga ir brangu. IT specialistai tai vadina „pririšimu prie paslaugos“. Dėl šios priežasties trys technologijų milžinės – „Amazon“, „Microsoft“ ir „Google“ – dominuoja debesijos kompiuterijos rinkoje.

Nadellai tapo aišku, kad „OpenAI“ tyrimai, susiję su didžiaisiais kalbos modeliais, gali būti pelningesni nei pačios „Microsoft“ DI komandos projektai, kurie po „Tay“ nesėkmės atrodė neturintys aiškios krypties. Taigi Nadella galiausiai sutiko investuoti milijardą dolerių į „OpenAI“. Šiuo žingsniu jis ne tik parėmė „OpenAI“ tyrimus, bet ir užtikrino, kad „Microsoft“ atsidurtų DI revoliucijos priešakyje. Mainais už tai „Microsoft“ gavo prioritetinę prieigą prie „OpenAI“ technologijų.

Tuo metu „OpenAI“ vis daugiau dėmesio buvo skiriama Sutskeverio ir Radfordo tyrimams didžiųjų kalbos modelių srityje. Jiems sukūrus patobulintą sistemos versiją, San Fransiske įsikūrusi tyrėjų komanda ėmė svarstyti, ar ji netapo pernelyg galinga. Antrajam jų modeliui – GPT-2 – mokyti buvo panaudota 40 GB tekstų iš interneto šaltinių. Modelis turėjo 1,5 milijardo parametrų, todėl pagal apimtį jis daugiau nei dešimt kartų pranoko pirmosios kartos GPT ir geriau generavo sudėtingus tekstus. Be to, modelio kuriami tekstai tapo įtikinamesni.

Tyrėjai nusprendė išleisti mažesnės apimties modelio versiją. 2019 metų vasario mėnesio tinklaraščio įraše jie įspėjo, kad jis gali būti panaudotas masinei dezinformacijai generuoti. Toks atvirumas išties stebino, tačiau vėliau „OpenAI“ tokio skaidrumo principo laikėsi vis rečiau. „Baimindamiesi galimo mūsų technologijos panaudojimo nederamais tikslais, nusprendėme apmokyto modelio neišleisti“, – viešame įraše skelbė tyrėjai. Pranešime dėmesys labiau buvo sutelktas į

riziką nei į patį modelį. Publikuotas tekstas vadinosi taip: „Tobulesni kalbos modeliai ir jų poveikis.“

Šio modelio išleidimas nesulaukė didesnio „DeepMind“ vadovų dėmesio Jungtinėje Karalystėje. Nors Demisas Hassabisas slapčia piktinosi tuo, ką daro Altmanas, jis nevertino „OpenAI“ susitelkimo į kalbos modelius kaip itin rimtos strategijos. Anot buvusių „DeepMind“ darbuotojų, Hassabisas tai laikė tik viena iš daugelio galimų BDI kūrimo krypčių. Jis manė, kad, norint sukurti išmanesnę DI sistemą, veiksmingesnė strategija būtų simuliuoti pasaulį žaidimais.

Netrukus paaiškėjo, kad „OpenAI“ DI strategija gali sulaukti išties daug dėmesio. GPT-2 modelis buvo plačiai nušviestas spaudoje, o daugelis straipsnių akcentavo „OpenAI“ įvardytas grėsmes, kurias kelia naujoji DI sistema. Žurnalas „Wired“ publikavo straipsnį pavadinimu „DI teksto generatorius, kurį per daug pavojinga pristatyti visuomenei“, o „The Guardian“ išspausdino straipsnį su antrašte: „DI gali rašyti taip pat gerai, kaip aš. Pasiruoškite robotų apokalipsei.“

„OpenAI“ paviešino pakankamai informacijos (įskaitant fiktyvų straipsnį apie angliškai kalbančius vienaragius), kad parodytų, jog naujasis teksto generatorius yra nepaprastai efektyvus. Tačiau tyrėjai šio modelio viešam testavimui neišleido. Be to, skirtingai nei pirmojo GPT modelio atveju, kai buvo nurodytas konkretus duomenų rinkinys – „BooksCorpus“, šįkart jie neatskleidė, kokiomis viešomis interneto svetainėmis ar kitais duomenų rinkiniais naudotasi mokant šį modelį. Atrodė, kad kuo daugiau „OpenAI“ slėpė informaciją apie GPT-2 ir kalbėjo apie jo keliamas grėsmes, tuo labiau visuomenė domėjosi šiuo modeliu. Taigi GPT-2 sulaukė neįprastai daug dėmesio.

Altmanas ir Brockmanas vėliau teigė, kad tai nebuvo jų tikslas, – „OpenAI“ esą iš tikrųjų nerimavo dėl galimo piktnaudžiavimo GPT-2 modeliu. Tačiau jų požiūris į viešąją komunikaciją iš dalies priminė mistiškosios rinkodaros (angl. *mystique marketing*) strategiją su atvirkštinės psichologijos elementu. Tokią strategiją ilgus metus naudojo

„Apple“, rengdama paslapties šydo gaubiamus produktų pristatymus, kurie iš anksto sužadindavo vartotojų susidomėjimą. Panašiai elgėsi ir „OpenAI“, neatskleisdama GPT-2 kūrimo užkulisių. Kai kurie DI akademikai bandymą gauti prieigą prie GPT-2 palygino su mėginimu patekti į prestižinį naktinį klubą. Galimybę išbandyti modelį „OpenAI“ suteikdavo tik kruopščiai atrinktiems asmenims. Ar tai buvo reklaminis triukas, ar apgalvotas mintinis eksperimentas?

Gali būti, kad abu šie spėjimai teisingi. Per savo karjeros metus Altmanas išmoko elgtis priešingai, nei diktuoja įprasta logika. Neatskleisdamas visų detalių, jis galėjo sukelti tikrą ažiotažą. O pasireiškus kontroversijai – kaip tada, kai „Wall Street Journal“ reporterei išsiuntė ilgą sąrašą „Loopt“ keliamų rizikų, – jis galėjo nuginkluoti kritikus.

Siekdama sukurti BDI, „OpenAI“ atsidūrė kryžkelėje. Nors papildomi duomenys ir skaičiavimo galia leido itin ištobulinti kalbos modelį, kurio generuojamas turinys vis labiau priminė sukurtąjį žmogaus, darėsi akivaizdu, kad organizacija vis labiau išduoda pirminius savo veiklos principus. Altmanas ir Brockmanas žinojo, jog sąjunga su „Microsoft“ privers juos sulaužyti šiuos pažadus, tačiau dar didesnis iššūkis buvo išlaikyti darbuotojus. Juk dauguma jų dirbo ne dėl pinigų, o dėl misijos. Supratę, kad organizacija nukrypo nuo savo misijos, jie turėtų naują priežastį pasitraukti.

Altmanui reikėjo ko nors, kas užtemdytų jo talentingų inžinierių kritinį mąstymą. Atsakymo ilgai ieškoti neteko – tam galėjo pasitarnauti BDI vizija. BDI kūrimo siekis iš esmės nesiskyrė nuo dangiškojo atpildo, šimtmečius įkvėpusio religines bendruomenes išlikti ištikimas savo tikėjimui. Abiem atvejais ant kortos buvo pastatyta labai daug: jei pasisektų, žmonijai atsivertų utopija, o jei ne – grėstų pasaulinė apokalipsė.

Galimo triumfo ar pražūties akivaizdoje priemonės tikslui pasiekti atrodė nereikšmingos. Svarbiausia buvo galutinis rezultatas. „OpenAI“

darbuotojai ėmė tikėti, jog turi moralinę prerogatyvą pirmieji sukurti BDI ir patys paskirstyti jo gėrybes pasauliui, nepaisydami organizacijos nuostatuose nurodyto ketinimo bendradarbiauti su kitais. Kai kurie manė, kad jei BDI pirmieji sukurtų „DeepMind“ arba kinų mokslininkai, jų kūrinys, tikėtina, būtų „šėtoniškas“.

Naujieji organizacijos nuostatai tik sustiprino tokį įsitikinimą. Altmanas ir Brockmanas šį dokumentą laikė kone „OpenAI“ šventraščiu. Jie netgi įdiegė sistemą, pagal kurią darbuotojų atlyginimų dydis priklausė nuo to, kaip griežtai jie laikėsi nuostatų. Per ketverius metus nuo naujųjų nuostatų paskelbimo „OpenAI“ tapo dar uždaresne organizacija, komanda dar labiau suartėjo. Darbuotojai net po darbo leisdavo laiką kartu, o savo darbą suvokė kaip misiją ir tapatybės dalį. Brockmanas netgi nusprendė tuoktis su savo išrinktąja Anna „OpenAI“ būstinėje: civilinę ceremoniją vedė Sutskeveris, akį traukė iš gėlių sukurtas „OpenAI“ logotipas, o žiedą pristatė roboto ranka.

Tiek „OpenAI“, tiek „DeepMind“ darbuotojų susitelkimas į pasaulio gelbėjimą, pasitelkiant BDI, pamažu kūrė aplinką, primenančią kultą, kurioje įsigalėjo radikalios idėjos. San Fransisko „OpenAI“ būstinėje Sutskeveris ėmėsi dvasinio lyderio vaidmens. Jis ragindavo darbuotojus „pajusti BDI“, be to, šią frazę paskelbė „Twitter“ tinkle. Pasak žurnalo „The Atlantic“, viename įmonės šventiniame vakarėlyje, vykusiame San Fransisko mokslo muziejuje, Sutskeveris ragino tyrėjus skanduoti šūkį „Pajusk BDI!“. Prie tokios kone religinės kultūros susiformavimo prisidėjo ir tai, kad dešimtys „OpenAI“ darbuotojų save laikė veiksmingojo altruizmo[11] šalininkais.

Šis judėjimas atsidūrė dėmesio centre 2022 metų pabaigoje, kai veiksmingąjį altruizmą išpopuliarino garsiausias jo rėmėjas kriptovaliutų milijardierius Samas Bankmanas-Friedas. Idėja buvo ne nauja, ji sklandė nuo 2010-ųjų. Veiksmingojo altruizmo filosofija, suformuluota grupės Oksfordo universiteto filosofų, greitai paplito universitetų bendruomenėse. Jos šalininkai pabrėžia būtinybę tobulinti

tradicinį požiūrį į labdarą, taikant utilitarinę logiką. Pavyzdžiui, vietoj savanorystės benamių prieglaudoje efektyviau būtų įsidarbinti gerai apmokamame darbe, kad ir investiciniame fonde, uždirbti daug pinigų ir už juos pastatyti kelias naujas prieglaudas. Tokia nuostata grindžiama principu „uždirbti tam, kad aukotum“ ir yra orientuota į tai, kad kiekvienas paaukotas doleris duotų kuo daugiau naudos.

Kartais veiksmingojo altruizmo šalininkų nuomonės šiuo klausimu išsiskirdavo. Vieni teigė, kad, spręsdamas pasaulines problemas, tokias kaip skurdas, galėtum padaryti daugiau gero nei kovodamas su vietinėmis JAV ar Europos problemomis, tokiomis kaip benamystė. Kiti manė priešingai. Didžiausios veiksmingojo altruizmo iniciatyvų rėmėjos „Open Philanthropy“ programų vadovas Nickas Becksteadas netgi rašė, kad „išsaugoti gyvybę turtingoje šalyje yra daug svarbiau nei tą padaryti skurdžioje šalyje, nes turtingos šalys yra inovatyvesnės, o jų darbuotojai – ekonomiškai produktyvesni“. Šiuo požiūriu žmogaus gyvybė tampa kiekybiškai įvertinama, o gerų darbų darymas laikomas matematine užduotimi, kurią reikia spręsti remiantis atitinkamais skaičiavimais.

BDI kūrimo misija ypač traukė tuos, kurie tikėjo veiksmingojo altruizmo filosofija ir laikėsi principo, kad veiksmai vertinami pagal jų poveikio mastą, nes ši technologija ateityje galėjo paliesti milijardus ar net trilijonus žmonių. Toks požiūris leido „OpenAI“ darbuotojams atlaidžiau vertinti tolesnius Altmano veiksmus. Tuo metu Altmanas rengėsi skristi į Siatlą, kad „Microsoft“ vadovui pademonstruotų naujausią „OpenAI“ kalbos modelį GPT-3. Be to, kartu su Brockmanu jis svarstė, kaip geriausia pertvarkyti „OpenAI“. Kaip ir „DeepMind“ įkūrėjai, jie ilgai ieškojo tinkamos juridinės formos, kuri leistų žmonijos gelbėjimo tikslus suderinti su pelno siekiu. „Išnagrinėjome visus įmanomus teisinės struktūros modelius ir supratome, kad nė vienas neatitinka mūsų tikslų“, – pasakojo Brockmanas vienoje tinklalaidėje.

Įmonės, siekiančios derinti pelno siekį ir socialinę misiją, kartais veikia kaip socialinės naudos bendrovės (angl. *benefit corporation*). Tai alternatyva tradicinei pelno siekiančių įmonių struktūrai, kurioje pagrindinis tikslas – maksimalios vertės kūrimas akcininkams. Amerikiečių ekonomistas Miltonas Friedmanas dar 1962 metais šį plačiau įsigalėjusį požiūrį apibendrino tokiais žodžiais: „Vienintelė kiekvienos įmonės socialinė atsakomybė – naudoti savo išteklius ir vykdyti veiklą taip, kad uždirbtų kuo daugiau pelno.“

Socialinės naudos bendrovės modelis leidžia derinti pelno siekį ir socialinę misiją. Jį taiko tokios bendrovės kaip „Patagonia“ ir „Ben & Jerry’s“. Priimdamos sprendimus, šios įmonės privalo atsižvelgti į galimą jų poveikį darbuotojams, tiekėjams, klientams, aplinkai ir akcininkams. Tačiau šis modelis ne visada pasiteisina. Pavyzdžiui, technologijų sektoriuje veikianti internetinė prekyvietė „Etsy“ buvo priversta atsisakyti socialinės naudos bendrovės statuso, kai jos akcijomis pradėta prekiauti biržoje. Įmonė susidūrė su finansų rinkos diktuojamais nuolatinio augimo reikalavimais, kurie įprastai primetami biržinėms bendrovėms.

Altmanas ir Brockmanas pasiūlė tarpinį variantą – sudėtingą ne pelno organizacijos ir pelno siekiančios įmonės hibridą. 2019 metų kovą jie paskelbė apie „riboto pelno“ įmonės įsteigimą. Bet kuris naujas pagal tokią struktūrą veikiančios įmonės investuotojas turėjo sutikti su investicijų grąžos riba. Įprastai investicijų į technologijų bendrovę grąža gaunama pardavus įmonę arba viešai pasiūlius jos akcijas biržoje. Tačiau pagal naująją Altmano „riboto pelno“ struktūrą buvo nustatyta maksimali riba, kiek investuotojai galėtų gauti pardavus įmonę, viešai pasiūlius jos akcijas biržoje arba išmokėjus tam tikrus dividendus. Pradžioje ta riba buvo itin aukšta, todėl pirmiesiems investuotojams išties pasisekė: jiems ribojimas įsigalioja tik tada, kai grąža šimtą kartų viršija pradinę investiciją. Tai reiškia, kad jei investuotojas įnešė 10 mln.

JAV dolerių, jo pelnas būtų ribojamas tik grąžai pasiekus 1 mlrd. JAV dolerių.

Tokie skaičiai atrodo įspūdingai net vertinant Silicio slėnio mastais. Altmanas teigia, kad ši pirminė pelno riba vėlesniems investuotojams buvo gerokai sumažinta, ir atkreipia dėmesį, kad pirmieji investuotojai prisiėmė didžiulę riziką. „Nors šiandien daugelis yra girdėję apie BDI ir suvokia, kad jis veikiausiai bus sukurtas artimiausioje ateityje, anksčiau dauguma manė, jog mūsų tikslas prasilenkia su realybe“, – sakė jis.

Altmanas, įkvėpęs startuolius siekti milijardinių pelnų, tokių pat drąsių užmojų turėjo ir „OpenAI“ investuotojų atžvilgiu, tikėdamasis, kad jie sulauks nemažos finansinės grąžos. Įmonė savo pertvarkymo dokumentuose netgi įrašė nuostatą, kad sukūrus BDI bus peržiūrėti visi jos finansiniai susitarimai, nes tuo metu pasaulis turės iš naujo persvarstyti pinigų idėją.

Remdamasis naująja sudėtinga organizacijos struktūra, Samas Altmanas įsteigė pagrindinę ne pelno bendrovę „OpenAI Inc.“. Jos valdyba turėjo užtikrinti, kad riboto pelno įmonė „OpenAI LP“ kurtų „visapusiškai naudingą“ BDI. Valdybą sudarė Altmanas, Brockmanas ir Sutskeveris, taip pat Reidas Hoffmanas, „Quora“ generalinis direktorius Adamas D’Angelo bei technologijų srities verslininkė Tasha McCauley.

Riboto pelno padalinys vykdė pagrindinius tyrimus, o bet kokios jo gautos pajamos, viršijančios nustatytą ribą, buvo grąžinamos „OpenAI Inc.“. Tai suteikė „OpenAI“ pakankamai galimybių pritraukti milijardus dolerių ir leido investuotojams uždirbti dar daugiau, kol nebuvo pasiekta maksimali investuotojų grąžos riba, nuo kurios papildomas pelnas turėtų būti skirstomas žmonijai.

Iš pradžių atrodė, kad „OpenAI“ ne pelno padalinys iš to negauna beveik jokios naudos. Įmonė neskelbė, kada minėta investicijų grąžos riba, lygi šimtą kartų didesnei už pradinę investiciją sumai, bus sumažinta ir kiek tiksliai. Altmanas veikė lanksčiai ir prireikus keisdavo kryptį, kaip tai daro geriausi startuoliai.

Įmonė dar kartą kryptį pakeitė 2019-ųjų birželį. Praėjus vos keturiems mėnesiams po to, kai „OpenAI“ tapo pelno siekiančia įmone, ji paskelbė apie strateginę partnerystę su „Microsoft“. „*Microsoft* investuoja 1 mlrd. JAV dolerių į *OpenAI*, kad padėtų kurti bendrąjį dirbtinį intelektą (BDI), kurio teikiama ekonominė nauda būtų plačiai paskirstyta žmonijai“, – rašė Brockmanas tinklaraštyje.

Į šią sumą įėjo tiek grynieji pinigai, tiek kreditai „Microsoft“ debesijos kompiuterijos infrastruktūros naudojimui. Mainais „OpenAI“ įsipareigojo suteikti technologijų milžinei licenciją naudoti jos technologijas debesijos kompiuterijos verslo plėtrai. Sprendimą, kada bus sukurtas BDI ir kada „Microsoft“ turėtų nutraukti šių technologijų naudojimą, priims ne pelno organizacijos „OpenAI“ valdyba.

Brockmanas rašė, kad „OpenAI“ reikėjo padengti išlaidas, o geriausias būdas tą padaryti buvo licencijuoti BDI lygio dar nepasiekusią technologiją. Pasak jo, jei „OpenAI“ būtų bandžiusi tiesiog parduoti savo sukurtą produktą, tai būtų nukreipę įmonės dėmesį nuo pagrindinio tikslo.

Vis dėlto šie argumentai neturėjo tvirto pamato. Licencijos suteikimas iš esmės nesiskyrė nuo produkto pardavimo – tai tas pats, kas parduoti technologiją didesniam klientui, turinčiam daugiau galios ir kontrolės nei įprasti vartotojai. Kol „OpenAI“ valdyba oficialiai nepaskelbs apie BDI sukūrimą, tol įmonė galės toliau licencijuoti savo technologijas „Microsoft“.

Naujoji Altmano įmonė bandė laviruoti tarp kertinių veiklos principų, įskaitant ir tuos, kurie buvo išdėstyti 2018-ųjų „OpenAI“ nuostatuose. Nors įmonė žadėjo nepadėti korporacijoms „koncentruoti galios“, galiausiai savo veiksmais prisidėjo prie tolesnio vienos galingiausių pasaulio technologijų milžinių stiprėjimo. Nepaisydama pažado bendradarbiauti su kitais BDI kūrėjais ir nesivaržyti, „OpenAI“ davė pradžią pasaulinėms „ginklavimosi varžyboms“. „OpenAI“ iššūkį panorusios mesti įmonės ir kūrėjai beatodairiškai kūrė bei pristatinėjo

DI sistemas. Ribodama prieigą prie informacijos apie naujus kalbos modelius, „OpenAI“ vis labiau atsitvėrė nuo išorės vertintojų. Skeptiškų akademikų ir nerimaujančių DI tyrėjų akimis, įmonės pavadinimas įgijo ironišką atspalvį.

Šį krypties pokytį Altmanas ir Brockmanas mėgino pateisinti dviem argumentais. Pirma, strategijos keitimas einant pirmyn esą yra įprastas kiekvieno startuolio žingsnis. Antra, galutinis tikslas – BDI sukūrimas – svarbesnis už priemones jam pasiekti. Galbūt pakeliui teks sulaužyti kai kuriuos pažadus, tačiau galiausiai tai neva bus naudinga žmonijai. Be to, tiek savo darbuotojams, tiek visuomenei jie aiškino, kad „Microsoft“ taip pat nori naudoti BDI žmonijos labui. Jie buvo įsitikinę, kad abi pusės siekia to paties tikslo. „Jei įvykdysime šią misiją, įgyvendinsime bendrą „Microsoft“ ir „OpenAI“ siekį suteikti naujų galimybių kiekvienam žmogui“, – rašė Brockmanas.

Didžiųjų technologijų bendrovių šalininkai jau daugelį metų kartoja, kad vertė, kurią jų kuriamos technologijos suteikia žmonijai, pranoksta net trilijonus dolerių, kuriuos jos uždirba. Išmanieji telefonai ir socialiniai tinklai išties atvėrė naujų galimybių bendrauti su žmonėmis visame pasaulyje, pasiūlė naujų pramogų ir verslo formų. Tokios nemokamos programėlės kaip „Google Maps“ ir „Facebook“ turi gausybę patogių funkcijų, kurios, rodos, palengvina mūsų gyvenimą. Tačiau naujosios technologijos nėra nekaltos: silpnėja žmonių tarpusavio ryšiai, nyksta privatumas, didėja vartotojų priklausomybė nuo ekrano, aštrėja psichikos sveikatos problemos, ryškėja politinis susiskaldymas bei pajamų nelygybė dėl vis labiau automatizuojamų darbo procesų. Prie viso to prisideda vos kelios įtakingos bendrovės.

„OpenAI“ atvėrė kelią dar vienam pokyčiui, susijusiam su technologijų naudojimo įpročiais, – panašiam į tą, kurį atnešė „Facebook“ išpopuliarinta socialinių tinklų kultūra. Pradėjęs bendradarbiauti su „Microsoft“, Altmanas iš esmės rengėsi pakartoti istoriją, kurią jau buvo parašęs Markas Zuckerbergas. Zuckerbergo

kūrinys pridarė žalos, nes jo verslo modelis skatino žmones kuo daugiau laiko praleisti įbedus akis į ekraną. Pandoros skrynia jau buvo perpildyta: DI sistemos skatino šališkumą rasės ir lyties atžvilgiu, DI algoritmai stiprino priklausomybę nuo socialinių tinklų naujienų srauto, o horizonte ryškėjo potencialiai katastrofiškas jų poveikis darbo rinkai. Altmanas galėjo užkirsti kelią šioms problemoms, jei „OpenAI“ būtų toliau vykdžiusi veiklą kaip ne pelno organizacija, o jis pats būtų laikęsis pažado dalytis tyrimų rezultatais su mokslininkais, kad šie galėtų juos atsakingai įvertinti. Tačiau pasirinkęs „Microsoft“, jis sudarė faustišką sandorį. „OpenAI“ jau nebekūrė DI žmonijos labui – dabar ji padėjo technologijų milžinei išlaikyti dominuojančią padėtį rinkoje ir siekti pranašumo įnirtingose varžybose. Prieš prasidedant tikrosioms varžyboms, jam liko tik įveikti paskutinį iššūkį.

9 DI agentas – sistema arba programa, gebanti savarankiškai atlikti užduotis ir priimti sprendimus su minimalia žmogaus priežiūra arba be jokio žmogaus įsikišimo (vert. past.).

10 Nacionalinė futbolo lyga (vert. past.).

11 Veiksmingasis altruizmas yra socialinis judėjimas ir filosofija, kurios esmė – rasti veiksmingiausius būdus daryti gera, pasitelkiant mokslą ir analitinį mąstymą (vert. past.).

11 SKYRIUS

Didžiųjų technologijų bendrovių gniaužtuose

Žvelgiant iš šalies, „OpenAI“ transformacija iš filantropinės žmoniją bandančios išgelbėti organizacijos į „Microsoft“ įmonę partnerę atrodė keistai, netgi įtartinai. Tačiau, pasak tuometinių įmonės darbuotojų, daugeliui „OpenAI“ žmonių dirbti su turtinga technologijų gigante buvo sveikintina naujiena. Ne tik buvo labiau tikėtina, kad darbdavys išliks mokus, bet radosi ir geresnė proga džiaugtis didelių finansinių investicijų, o ne paramos rezultatais. Per ateinančius kelerius metus „Microsoft“ Samo Altmano įmonei skirs dar daugiau kapitalo ir „OpenAI“ darbuotojams suteiks progą parduoti akcijas bei tapti milijonieriais. Daugelis „OpenAI“ tyrėjų nemanė, kad jų misija žlugo. Jie įsivaizdavo, kad BDI kūrimas pranoko bet kokią sąžinės graužatį dėl jo kūrimo būdo. Kol bus laikomasi svarbių nuostatų reikalavimų, pinigų kilmė nesvarbi. Juk tai vyko Silicio slėnyje – čia programuotojai uždirbo septynženkles metines algas ir akcijų pasirinkimo sandoriais brangiausioje Amerikos nekilnojamojo turto rinkoje galėjo nusipirkti antruosius namus. Šie profesionalai pasirinkdavo darbuotis pasaulį geresniu besistengiančiuose padaryti startuoliuose.

Visgi naujasis *status quo* patiko ne visiems. Akiniuotam garbaniui inžinieriui Dario’ui Amodei’ui, kuris nuo pat „OpenAI“ įkūrimo bandė išsiaiškinti, ką tiksliai įmonė nori pasiekti, buvo priimtinas tikslas apsaugoti žmoniją nuo pavojingo DI, net jei tuo metu Brockmanas pripažino, kad šis tikslas „truputį neaiškus“. Amodei’us buvo Prinstono universitetą baigęs fizikas, nebijojęs užduoti sudėtingų klausimų: apie „Microsoft“ turėjo jų daugybę. Buvo akivaizdu, kad „OpenAI“ ir „Microsoft“ tikslai skirtingi. Tad kaip „OpenAI“ likti prie saugaus DI

kūrimo, drauge padedant „Microsoft“ uždirbti daugiau pinigų? „DI kuriame žmonijai, tačiau taip pat tampame technologijų tiekėju įmonei, kuri stengiasi kuo labiau didinti pelną“, – pasak tam tikro jo argumentus girdėjusio asmens, Amodei’us taip pasakojo savo bendradarbiams. Tai buvo nelogiška.

Amodei’us vadovavo dideliems įmonės „OpenAI“ skyriams, įskaitant darbą su kalbos modeliais. Jis su komanda dirbo prie kitos kalbos modelio versijos pavadinimu GPT-3. Kad ir kaip nesmagiai jautėsi dėl prisijungimo prie „Microsoft“, Amodei’us privalėjo pripažinti, kad programinės įrangos gigantė suteikia jiems reikalingų neprilygstamų kompiuterinių išteklių. Tiesą sakant, pora mėnesių po investicijos „Microsoft“ paskelbė sukūrusi superkompiuterį tam, kad įmonė „OpenAI“ galėtų mokyti savo DI.

Amodei’us retai dirbdavo su galingesne sistema. Tipinis namų kompiuteris turi vieną centrinį procesorių (angl. *central processing unit*, CPU) – milijardo smulkių tranzistorių padengtą, galingą stačiakampio formos silikoninę mikroschemą. Tai jūsų kompiuterio smegenys, paprastai turinčios nuo keturių iki aštuonių būtinų skaičiavimus atliekančių branduolių. Naujasis „Microsoft“ superkompiuteris turėjo 285 tūkst. CPU branduolių. Jeigu įprastas kompiuteris namuose būtų prilyginamas žaislinei mašinėlei, tai šis – tankui.

Kai žmonės žaidimams perka galingesnį kompiuterį, paprastai jame būna grafikos procesorius (angl. *graphics processing unit*, GPU), kuris greitai apdoroja vaizdo duomenis sklandžiam ir nepriekaištingam žaidimo vaizdui rodyti. Dabar tos pačios mikroschemos buvo naudojamos DI mokyti, nes vienu metu buvo galima atlikti daugybę skaičiavimų. Naujasis „Microsoft“ superkompiuteris jų turėjo dešimtis tūkstančių. Dėl žaibiško ryšio superkompiuteris duomenis galėjo perkelti šimtus kartų greičiau nei įprastas kompiuteris.

Išnaudodama visus šiuos naujus skaičiavimo pajėgumus, „OpenAI“ iš interneto traukė didžiulius teksto kiekius, kad DI galėtų mokyti naujų

„GPT“ kalbos modelių. Superkompiuteris buvo it XIX amžiaus naftos žvalgytojas, internete randantis didelius turinio klodus ir apdorojantis juos į sumanesnį DI. „Wikipedia“ tyrėjai jau buvo ištraukę apie keturis milijardus žodžių, todėl kitas akivaizdus šaltinis buvo milijardai komentarų, kuriais žmonės dalijosi socialinėse medijose. Nebuvo galima naudotis „Facebook“, nes po 2018 metais įvykusio „Cambridge Analytica“ skandalo Marko Zuckerbergo platforma neleido kitoms įmonėms prieiti prie vartotojų duomenų. Tačiau „Twitter“, kaip ir „Reddit“, buvo daugmaž prieinamas visiems.

„Reddit“, žinomas kaip pagrindinis interneto puslapis, yra bet kokią sugalvojamą temą apimantis forumas – nuo automobilių iki susitikinėjimo ir Renesanso paveikslus primenančių nuotraukų. Įmonę su Altmanu siejo glaudūs ryšiai, nes Altmanas ir įmonės steigėjai kartu dalyvavo „Y Combinator“. Remiantis 2024 metais prieš pirminį viešą akcijų siūlymą pateiktais dokumentais, galiausiai Altmanas, valdęs 8,7 proc. akcijų, tapo trečiu stambiausiu įmonės akcininku. Jis turėjo gerą priežastį mėgti „Reddit“: kasdieniai vartotojų komentarai, dėl kurių yra ir balsuojama, buvo DI mokymams skirtų žmonių pokalbių aukso gysla. Nieko nuostabaus, kad „Reddit“ tapo vienu svarbiausių „OpenAI“ DI mokyti skirtų šaltinių. Remiantis su internetiniu forumu susijusiu asmeniu, platformoje esantis tekstas sudaro 10–30 proc. „GPT-4“ mokymams skirtų duomenų. Kuo daugiau teksto „OpenAI“ naudojo kalbos modeliui mokyti ir kuo galingesni buvo įmonės kompiuteriai, tuo DI darėsi iškalbingesnis.

Vis dėlto Amodei’us negalėjo atsikratyti nejaukumo jausmo. Jis su seserimi Daniela, kuri vadovavo „OpenAI“ politikos ir saugumo komandoms, stebėjo, kaip „OpenAI“ modeliai tampa didesni ir geresni. Niekas jų komandoje nežinojo, kokios bus tokių sistemų pristatymo visuomenei pasekmės. Priklausydami nuo galingos korporacijos, jie galėjo susidurti su spaudimu išleisti technologiją dar jos tinkamai neišbandę.

Demisas Hassabisas, kaip ir Amodei'us, Londone buvo susirūpinęs. Maždaug tuo metu, kai „OpenAI“ rengėsi išleisti GPT-3, Samas Altmanas, Gregas Brockmanas ir Ilya Sutskeveris, bandydami pagerinti dviejų konkuruojančių bendrovių santykius, pietavo su „DeepMind“ įkūrėjais. Pietūs buvo įtempti. Pasak apie diskusijas žinojusio asmens, Demisas Hassabisas paklausė Altmano, kodėl „OpenAI“ pristato savo DI modelius visuomenei, kai pavojingų ketinimų turintys asmenys gali juos naudoti dezinformacijai skleisti arba kenksmingesniems DI įrankiams kurti. Hassabisas pareiškė, kad „DeepMind“ laiko DI paslaptyje ir saugo nuo netinkamo naudojimo.

Altmanas bandė aiškinti, kad tai absurdas. Jis subtiliai visiems priminė Elono Musko *blogio genijaus* juokelius Hassabiso atžvilgiu. Altmanas teigė, kad slapukavimas vienam DI įmonę, pavyzdžiui, „DeepMind“, valdančiam asmeniui suteikia pavojingą kontrolę, todėl toks požiūris irgi nėra saugus.

Grįžęs į San Fransiską, Altmanas panašių argumentų išgirdo iš Amodei'aus, kuris skundėsi nauja „OpenAI“ komercine kryptimi. Altmanas susisiekė su visada optimistiškai nusiteikusiu Reidu Hoffmanu, norėdamas pasinaudoti jo kaip taikdario įgūdžiais. Remiantis apie pašnekesį žinojusio asmens žodžiais, Hoffmanas atėjo pas Amodei'ų norėdamas išsiaiškinti problemas. Milijardierius rizikos kapitalo investuotojas švelniai priminė pasitikėti procesu.

„Taip įvykdysime misiją“, – paaiškino Hoffmanas. Amodei'us ir jo sesuo buvo skeptiški. Jie žinojo, kokie dideli ir sunkiai valdomi tampa šie kalbos modeliai, taip pat pastebėjo, kad Hoffmanas yra „Microsoft“ valdybos narys. Argi *jis* nėra asmeniškai suinteresuotas?

Brolis ir sesuo liko atsargūs dėl augančio „OpenAI“ prisirišimo prie „Microsoft“. Nuo „OpenAI“ įkūrimo didelės technologijų įmonės, nevykdydamos adekvačių rizikos tyrimų, centralizavo DI kūrimo kontrolę, kad technologija taptų galingesnė ir pajėgesnė. 2023 metų Masačusetso technologijų instituto tyrimu nustatyta, kad didžiosios

įmonės per pastarąjį dešimtmetį pradėjo dominuoti DI modelių valdyme – nuo 11 proc. 2010 metais iki beveik 96 proc. DI modelių kontrolės 2021-aisiais. Palyginti su milžiniškomis pinigų sumomis, kurias didžiosios technologijų įmonės skyrė DI, netgi vyriausybės projektai atrodė maži. Pavyzdžiui, 2021 metais ne su gynyba susijusios JAV vyriausybinės agentūros DI iš biudžeto skyrė 1,5 mlrd. JAV dolerių. Privatusis sektorius šiai sričiai tais pačiais metais atseikėjo daugiau kaip 340 mlrd. JAV dolerių.

Komercinių DI sistemų mechanika buvo laikoma paslaptyje. Įmonei „OpenAI“ vis daugiau technologijų skiriant visuomenės poreikiams, šių technologijų sukūrimo istorija pasirodė esanti dar neaiškesnė. Nepriklausomiems tyrėjams tapo sunkiau tikrinti galimas sistemų grėsmes ir poslinkius. Įsivaizduokite, kad toks didelis maisto prekių gamintojas kaip „Unilever“ gamintų itin gardžius užkandžius, tačiau ant pakuočių nenurodytų sudėties ar atsisakytų paaiškinti, kaip maistas buvo pagamintas. Būtent tai iš esmės ir darė „OpenAI“. Daugiau galima sužinoti apie traškučių „Doritos“ sudėtį nei apie didįjį kalbos modelį.

Amodei’us buvo labiau susirūpinęs dėl DI egzistencinės grėsmės žmonijai nei dėl tendencingumo. Jo mokslo tiriamajame darbe „Specifinės DI saugos problemos“ (*Concrete Problems in AI Safety*) pabrėžiami galimi nelaimingi atsitikimai dėl netinkamai sukurtų DI sistemų. Jei, kurdami DI, kūrėjai nurodo netinkamus tikslus, sistema gali sukelti netyčinę žalą. Tiriamajame darbe Amodei’us rašė: apdovanok buitinį robotą už tai, kad jis nuneša dėžę iš vieno kambario galo į kitą, ir jis gali nuversti kelyje pasitaikiusią vazą, nes robotas yra itin susitelkęs į užduotį. Amodei’us argumentavo, kad žmonės turi pažvelgti į nelaimingus atsitikimus tikrajame pasaulyje, kuriuos gali sukelti į pramonės valdymo sistemas ir sveikatos apsaugą integruotas DI.

Galiausiai jo neįtikino Hoffmano samprotavimai. Amodei’us kartu su seserimi Daniela bei maždaug tuzinu kitų įmonėje dirbusių tyrėjų nusprendė palikti „OpenAI“. Šis išėjimas iš darbo buvo susijęs ne vien su

sauga ar DI komercializavimu. Net dėl DI labiausiai sunerimę asmenys buvo oportunistai. Amodei'us matė, kaip Samas Altmanas iš „Microsoft" gavo 1 mlrd. JAV dolerių investiciją, ir galėjo nujausti, jog to kapitalo gali būti daugiau. Jis neklydo. Amodei'us stebėjo naują DI bumą. Jis su kolegomis nusprendė įkurti naują įmonę „Anthropic". Pavadinimas kilo iš filosofijos termino, reiškiančio žmogaus egzistavimą, ir pabrėžė bendrakūrėjų susirūpinimą žmonija. „Anthropic" būtų atsvara „OpenAI", kaip kad „OpenAI" buvo atsvara „DeepMind" ir „Google". Žinoma, jie norėjo išnaudoti ir galimybę užsiimti verslu.

„Tuo metu nemanėme, kad DI yra neprieinamas", – teigė vienas „Anthropic" įkūrėjų. Kitais žodžiais tariant, toji sritis buvo plačiai prieinama. „Atrodė, kad gudriai naujai organizacijai per trumpą laiką gali pasisekti kaip ir jau veikiančioms organizacijoms. Manėme, kad turėtume įkurti savo įmonę, paremtą mūsų pačių vizija, kurios centre būtų saugos tyrimai."

Kuriant abu „OpenAI" kalbų modelius Amodei'us atliko pagrindinį vaidmenį. Dabar jis galėjo kurti naudodamas savo vardą ir prekių ženklą. Jis su komanda atsižvelgė į tai, kaip „OpenAI" iš ne pelno organizacijos tapo pelno siekiančia, ir nusprendė, jog nenori kartoti tų pačių klaidų, dėl kurių taptų nepatikimi. Jie įsisteigė kaip socialinės naudos bendrovė. Šią teisinę verslo struktūrą naudojo „Ben & Jerry's", kad tame pačiame akcininkų lygmenyje pristatytų socialinius ir aplinkos apsaugos klausimus.

Dabar Samas Altmanas, be „DeepMind", turėjo dar vieną konkurentą, pasižymintį pavojingesnėmis „OpenAI" virtuvės įžvalgomis. Kaip buvo nuspėjęs Amodei'us, beveik iš karto „Anthropic" iš įprastų turtingų DI sauga besirūpinančių mecenatų, įskaitant Jaaną Tallinną ir Dustiną Moskovitzą (milijardierių „Facebook" bendrakūrėją, kuris Harvardo universitete buvo Marko Zuckerbergo kambariokas), surinko dideles pinigų sumas. Silicio slėnyje pinigai dažnai cirkuliuoja tarp elitinių tinklų grupių, įskaitant ilgamečius konkurentus. Moskovitzo labdaros įmonė

„Open Philantrophy“ 30 mln. JAV dolerių skyrė „OpenAI“, Altmanas finansiškai parėmė Moskovitzo programinę įrangą „Asana“. Vis dėlto Moskovitzas taip pat norėjo paremti ir naująjį „OpenAI“ konkurentą. (Vėliau Tallinnas teigė, kad gailėjosi įžiebęs tokią didelę konkurenciją DI srityje, dėl ko DI tapo galimai pavojingesnis.)

Per metus daugiausia iš turtingų jaunų kriptovaliutos keityklos „FTX“ įkūrėjų, kurie Amodei'ų susirado vedami bendrų veiksmingojo altruizmo interesų, „Anthropic“ papildomai surinko 580 mln. JAV dolerių. Ironiška, kad, praėjus dvejiems metams po Amodei'aus skundų dėl „OpenAI“ komercinių sąsajų su „Microsoft“, iš „Google“ ir „Amazon“ jis priims daugiau kaip 6 mlrd. JAV dolerių investicijų, tokiu būdu susisaistydamas su abiem įmonėmis. Pasirodo, šiame naujame kone neišsemiamų išteklių BDI kurti reikalaujančiame pasaulyje žmonės negali pasakyti „ne“ technologijų konglomeratams.

Kitoje vandenyno pusėje, Londone, šie ryšiai tapo kliuviniu „DeepMind“. Demisas Hassabisas ieškojo naujų mokslinių tikslų, kuriuos įmonė galėtų pasiekti pirmavimui prieš „OpenAI“ pademonstruoti ir sužavėti pasaulį po „AlphaGo“. Tačiau įmonės bendrakūrėjis Mustafa Suleymanas vis dar buvo pasiryžęs įrodyti, kad DI galima panaudoti gėriui. Jau daugelį metų Suleymanas nejaukiai jautėsi dėl to, kokia kryptimi jo draugas Demisas Hassabisas kreipė įmonę. Šachmatų genijus buvo užsiėmęs DI kūrimu per žaidimus ir simuliacijas, tačiau Suleymanas manė, kad jie turėtų studijuoti ir tikrąjį pasaulį, net jei tai reikštų didelio neaiškių duomenų kiekio tvarkymą. Kaip kitaip jie galėtų ateityje išspręsti visuomenės problemas, jeigu dabar su jomis nedirbtų?

Jis tapo kelių Londono ligoninių partneriu, kad „DeepMind“ DI panaudotų gydytojams ir slaugytojams padėti. Projektas prasidėjo nuo programėlės, kuri išsiųsdavo pranešimą, įspėjantį, kad pacientams gali išsivystyti ūmus inkstų pažeidimas. Dėl medicinoje galiojančių reguliacinių kliūčių programėlė nenaudojo pažangaus įmonės „DeepMind“ DI. Suleymanas tikėjo, kad jo DI mokslininkai galėtų

padaryti įrankį išmanesnį, jeigu DI mokytųsi iš tinkamų medicininių duomenų. Gydytojams programėlė patiko ir projektas atrodė daug žadantis. Tačiau įvyko tai, ko niekas negalėjo numatyti. Žiniasklaida rašė, kad „Google“ Londone turi prieigą prie 1,6 milijono pacientų įrašų ir bando išgauti jautrius duomenis. Staiga Suleymano eksperimentas virto skandalu. Jis buvo taip įsitikinęs artėjančiu „DeepMind“ atsiskyrimu, kad pamiršo, jog teoriškai įmonė vis dar priklauso reklamos milžinei, kuri pinigus uždirba iš žmonių duomenų gavimo ir dalijimosi jais su reklamos davėjais. Dėl „Google“ įmonės „DeepMind“ pastangos pasitelkus DI išspręsti sveikatos priežiūros problemas išoriniam pasauliui staiga pasirodė įtartinos.

Hassabisas buvo pasibaisėjęs neigiamais žiniasklaidos straipsniais apie skandalą ligoninėje; atrodė, kad tai sunaikino jo „AlphaGo“ Azijoje pelnytą reputaciją. Ši patirtis patvirtino, kad bandymas mokyti DI modelį pagal tikrąjį pasaulį atspindinčius duomenų kratinius (panašų būdą naudojo „OpenAI“ kalbos modeliams mokyti) gali pakenkti „DeepMind“ reputacijai, ypač dėl sąsajos su „Google“.

Atrodė, kad Hassabisas abejoja ir nepriklausomų etikos tarybų (įskaitant ir tą, kurią jiedu su Suleymanu buvo numatę paskirti atsakingą už „DeepMind“ priežiūrą, pastarajai galiausiai atsiskyrus nuo „Google“) praktiškumu. Vis dėlto Suleymanas nekantravo eksperimentuoti su valdymu. Jis įkūrė mažesnę priežiūros tarybą, kad kruopščiai išnagrinėtų „DeepMind“ sveikatos priežiūros projektus ir užtikrintų jų virtuozišką vykdymą. Tarybą sudarė aštuoni meno, mokslo ir technologijų sričių profesionalai britai, įskaitant vieną buvusį politiką. Jie susitikdavo keturis kartus per metus nagrinėti įmonės sveikatos apsaugos srities tyrimų, pasikalbėti su inžinieriais ir nustatyti bet kokias „DeepMind“ planų dirbti su ligoninėmis bei pacientais etikos problemas.

Tai buvo kilnus, bet nesėkmei pasmerktas savireguliacijos eksperimentas. „OpenAI“, „DeepMind“ ir kitose technologijų bendrovėse, pavyzdžiui, „Facebook“, vyravo požiūris, kad

nepriklausomos tarybos yra geriausias būdas rasti pusiausvyrą tarp DI kūrimo žmonijos labui ir pelno siekio, ypač nesant tinkamo reguliavimo. Pavyzdžiui, „OpenAI“ direktorių valdybos vienintelė atsakomybė buvo įsitikinti, kad įmonė žmonijos labui kuria BDI. „DeepMind“ norėjo turėti panašaus modelio tarybą, kuri, įmonei atsiskyrus nuo „Google“, būtų įmonės sąžinė. Tačiau kai veiki pasaulinio titano, norinčio apsaugoti savo pelną ir interesus, viduje, šios gerų ketinimų vedamos valdymo struktūros būna netvarios. Samas Altmanas sunkiai išmoko šią pamoką. Realybė smogė ir Suleymanui. Jis neketino versti „DeepMind“ sveikatos priežiūros sritį kruopščiai nagrinėjančių tarybos narių pasirašyti konfidencialumo nurodymus. Suleymanas norėjo, kad panorėję tarybos nariai galėtų laisvai ir viešai kritikuoti įmonę. Drauge tai reiškė, kad jie dažniausiai būdavo nepakankamai susipažinę su visa „DeepMind“ darbų apimtimi. Kadangi jų sprendimai nebuvo teisiškai įpareigojantys, tarybos nariai skundėsi bejėgiškumu. Iš tiesų taryba *negalėjo* daryti daug ko. Ši problema kartojosi daugelyje pramonės srityje veikusių technologijų bendrovių. Negalėjai kritiškai vertinti tave pasamdžiusios įmonės ir neturėjai jokių teisinių įgaliojimų.

Galiausiai eksperimentas nedavė vaisių. „Google“ nusprendė plėtoti savo pačios sveikatos priežiūros skyrių ir perimti „DeepMind“ darbą su gydytojais bei medicinos srities profesionalais. Paieškos milžinė nenorėjo, kad pašaliniai asmenys pastoviai ieškotų spragų jos darbe, todėl Suleymano taryba buvo panaikinta. Tai buvo dar vienas „Google“ (ir technologijų pasaulio) savireguliavimo bandymo akligatvis. Anksčiau tais metais „Google“, visuomenei pasipiktinus, kad vienas tarybos narių reiškia LGBTQ priešiškus įsitikinimus, po savaitės panaikino dar vieną DI patariamąją tarybą. Visa tai rodė platesnę sisteminę problemą. DI plėtra buvo tokia greita, kad reguliavimo agentūrų ir įstatymų leidėjų galimybės neatsilikti tiesiog išnyko. Technologijų įmonės veikė teisės vakuume ir tai reiškė, kad teoriškai jos su DI galėjo daryti ką tik panorėjusios. Technologai, pasitelkdami įvairiausias skirtingas valdybas

ir teisines struktūras, gera valia bandė reguliuoti savo pačių įmones. Vis dėlto jie dirbo sistemoje, kurioje privalėjo pirmenybę teikti savo finansiniams įsipareigojimams akcininkams bei įmonės augimui. Galiausiai žlugo ir ilgalaikės skausmingos „DeepMind“ pastangos atsiskirti nuo „Google“.

Debesuotą 2021 metų balandžio rytą Londone, vaizdo konferencijos pokalbio su visais darbuotojais metu, Demiso Hassabiso veidas buvo perkreiptas šypsenos; jis rengėsi daryti tai, ką išmanė geriausiai, – blogas naujienas pateikti kaip geras. „DeepMind“ jau septinti metai mėgino įgyti nepriklausomybę nuo „Google“. Jie norėjo tapti savarankišku vienetu, paskui – „Alphabet“ konglomeratui priklausančia įmone, vėliau – visuotinės svarbos bendrove. Galiausiai nusprendė tapti ribotos atsakomybės bendrove – šią britišką teisinę etiketę paprastai naudodavo labdaros organizacijos ir klubai, tačiau ji leistų sujungti verslo, mokslinių atradimų ir altruizmo tikslus. Planai vis dar buvo laikomi paslaptyje. Tūkstančiai „DeepMind“ darbuotojų už įmonės ribų apie juos su niekuo nekalbėjo.

Jei žengtumėte žingsnį atgal ir pasižiūrėtumėte, ką Hassabisas ir Suleymanas visus šiuos metus mėgino padaryti, atrodytų, kad juos apėmė vadinamoji pardavėjo sąžinės graužatis. Technologijų pasaulyje tai dažnas reiškinys; daugeliu atvejų steigėjai baisisi, kaip įsigyjanti įmonė iškreipia jų pradinę misiją. Pavyzdžiui, „WhatsApp“ įkūrėjai, patikimai užšifruodami tinkle siunčiamus pranešimus, daugelį metų tikino, kad susirašinėjimo programėlė išliks privati ir niekada nerodys reklamų. Komunistinėje Ukrainoje, kur telefoninių pokalbių buvo nuolat klausomasi, užaugęs Janas Koumas prie savo stalo buvo prisiklijavęs įmonės bendrakūrėjo Briano Actono parašytą raštelį: „Jokių reklamų! Jokių žaidimų! Jokių triukų!“ Tačiau įmonę už 19 mlrd. JAV dolerių pardavę „Facebook“, Koumas ir Actonas turėjo imtis kompromisų dėl savo ankstesnių privatumo standartų. Pavyzdžiui, vienu metu jie atnaujino privatumo politikas taip, kad „WhatsApp“ paskyros galėjo būti

vartotojams nežinant susietos su „Facebook“ profiliais. Po pikto kivirčo su „Facebook“ vadovais Actonas prieš baigiantis akcijų įsigijimo laikotarpiui paliko įmonę ir neteko 850 mln. JAV dolerių. Vėliau jis prisipažino labai gailėjęsis ją pardavęs.

Hassabisas nebuvo vienas tų, kurie ginčijasi su savo vadovais. Jis buvo strategas ir kur kas rafinuočiau bendravo su „Google“ vadovybe. Vietoje ginčų ir išėjimo iš darbo jis ieškojo protingesnių būdų reputacijai išsaugoti, kaip kad „AlphaGo“ atveju, tačiau jo optimizmas neleido pamatyti „Google“ noro pastoviai auginti verslą. Nors didesnioji įmonė pasirašė sąlygų dokumentą, kuriuo siūlė 16 mlrd. JAV dolerių skirti „DeepMind“, kad ši dešimties metų laikotarpiu veiktų nepriklausomai, tas dokumentas nebuvo teisiškai įpareigojantis. Dar blogiau, Hassabisas prarado ryšį su „Google“ vadovu. Larry’is Page’as, nors vis dar buvo pagrindinės bendrovės „Alphabet“ generalinis direktorius, paskutinius kelerius metus traukėsi iš visuomenės akiračio. Jis nė neatvyko į Kongreso posėdį rinkimų saugumo tema (žiniasklaida fotografavo jam skirtą tuščią kėdę). 2019 metų gruodį Page’as visiškai pasitraukė iš pareigų ir „Alphabet“ generaliniu direktoriumi tapo Sundaras Pichai’us. Tai buvo aiškiausias ženklas, kad įmonė plečiasi ir elgiasi kaip tradicinė korporacija.

Daugelį metų laisvai veikiantys „Google“ įkūrėjai Page’as ir Sergey’us Brinas užsiiminėjo ambicingomis idėjomis, pavyzdžiui, savavaldžiais automobiliais, dėvimaisiais kompiuteriais (angl. *wearable computers*) ir netgi projektu, kuriuo buvo siekta įveikti mirtį, tačiau nė vienas šių verslų negeneravo pinigų. Remiantis „Wall Street Journal“, 2019 metais šis ambicingas kone 1 mlrd. JAV dolerių įmonei atsiėjęs verslas uždirbo apie 155 mln. JAV dolerių. Tuo tarpu „Google“ paieškos verslas kartu su interneto naršykle „Chrome“, aparatinės įrangos bloku ir „YouTube“ per metus uždirbo apie 155 mlrd. JAV dolerių. Pichai’us norėjo konsoliduoti tokio pagrindinio verslo kaip reklama ir paieška kontrolę bei jų pagrindu veikiančią DI technologiją. Jei Hassabisas norėjo sukurti visatos paslaptis

atskleisiantį DI, tai Pichai'us tikėjosi, kad DI dar labiau pagerins „Google" reklamos verslą. Jis norėjo, kad „Google" nustotų eksperimentavusi su tokiais nerentabiliais verslais kaip prekių pristatymo dronais paslaugos ar kvantinės technologijos ir sutelktų dėmesį savo pagrindinėms veikloms.

Page'o pasitraukimas buvo smūgis Hassabisui. Įtampos su „Google" laikotarpiu jis buvo ištikimas palaikytojas. „Mes netekome savo gynėjo, – prisimena buvęs „DeepMind" vadovas. – Mums visada sakydavo: „Nesijaudinkite, Larry'is saugo jūsų užnugarį"." Anksčiau, kai tik Pichai'us bandė versti „DeepMind" atlikti daugiau darbų „Google", Hassabisas kreipdavosi į tą patį gynėją. „Demisas visada [jį] apeidavo, nueidavo pas Larry'į ir gaudavo, ką norėdavo", – pamena dar vienas buvęs „DeepMind" darbuotojas.

Hassabisą ir Sundarą Pichai'ų siejo geri darbiniai santykiai. Larry'is Page'as buvo svajotojas kaip ir Hassabisas, o Pichai'us – labiau dalykiškas technologijų srities vadovas, norėjęs geriau išnaudoti „DeepMind" patirtį. 2019 metais „DeepMind" metų nuostolis prieš atskaitant mokesčius padidėjo iki maždaug 600 mln. JAV dolerių, beveik tiek pat, kiek „Google" sumokėjo už įmonę. Paieškos milžinei tai kainavo itin daug.

Dirbtinio intelekto taikdarys Reidas Hoffmanas bandė įkalbėti „DeepMind" įkūrėjus likti su „Google" ir išlaikyti *status quo*. Jis matė advokatų pateiktas apybraižas, kaip atrodys jų naujoji įmonė, ir įvertino Suleymano bei Hassabiso šimtus tam skirtų valandų. Hoffmanas iš karto galėjo matyti, kad Suleymano ir Hassabiso reikalas bergždžias.

„Jūsų ir „Google" interesai skiriasi", – perspėjo Hoffmanas. Jie neturėjo skirti tiek daug laiko ir tikėjimo įmonės atskyrimui, nebūdami visiškai tikri, kad „Google" tam pritars. Be to, jie neprivalo įkurti ne pelno organizacijos tam, kad DI padarytų saugų. Hoffmanas irgi norėjo pakylėti žmoniją į kitą lygmenį, tačiau jis buvo tikras kapitalistas ir tikėjo, kad altruistiškų tikslų geriausia siekti komercinėmis priemonėmis. Teigė, kad

Suleymanas ir Hassabisas turi priemonę – „Google“! Transformavimasis į ribotos atsakomybės bendrovę buvo komplikuotas ir nerealistiškas, be to, niekas anksčiau to nedarė.

Šiuo atžvilgiu Hoffmanas neklydo. Jeigu jie bandė pabėgti nuo korporacijos įtakos, „DeepMind“ įkūrėjai Altmanas ir Dario’us Amodei’us kartu su „Anthropic“ bendrakūrėjais buvo beviltiškai naivūs. Didžiausios technologijų įmonės, kurios daugiau kontroliavo DI tyrimus, plėtrą, mokymus ir pristatymą pasauliui, greitai perėmė DI verslą.

Hassabisas, tą balandžio rytą vaizdo konferencijoje kalbėdamasis su savo darbuotojais, pasakė jiems turintis dvi naujienas. Pirma ta, kad saugų „DeepMind“ DI kūrimą prižiūrės etikos taryba, tačiau ji niekuo nebus panaši į teisiškai nepriklausomą valdybą, kokią iš pradžių įsivaizdavo Hassabisas ir Suleymanas. Iš tiesų, ji nebus nepriklausoma. Ją sudarys išimtinai „Google“ vadovai – ir nė vieno „DeepMind“ atstovo.

Kita naujiena buvo dar labiau nuvilianti. „Google“ padarė galą bet kokiems „DeepMind“ planams tapti atskiru juridiniu asmeniu. „DeepMind“ inžinierius savo kolegai nusiuntė žinutę apie šias naujienas. „Demisas atskleidžia derybų su „Google“ rezultatus, – rašė jis. – Nieko negavome.“

Kol darbuotojai stengėsi suvokti naujienas, Hassabiso optimizmas neblėso. Bėgant metams jis tapo rinkodaros meistru. Jis galėjo recenzuojamame žurnale „Nature“ esančias žinias apie pusėtiną DI kūrimą paversti pasaulį sudrebinančiu atradimu, o įmonės viduje regresą paversti pranašumu. Savo darbuotojams jis sakė, kad, likusi „Google“ dalimi, „DeepMind“ gautų taip reikalingą finansavimą BDI priartinti prie realybės. „DeepMind“ vis dar galėtų nepriklausomai dirbti: visi darbuotojai gautų naujus DeepMind.com el. pašto adresus, pakeisiančius Google.com adresus. Jausdami, kad Hassabisas jiems kaip šuniui numetė kaulą, darbuotojai tuščiu žvilgsniu spoksojo į ekranus. Daugelis įtarė, kad „Google“ veikiausiai nepaleis itin vertinamos DI laboratorijos, kainavusios jai 650 mln. JAV dolerių. Vis dėlto darbuotojai

tikėjosi likti altruistiškesnio projekto, kuris visuomenę transformuotų gėrio tikslais, dalimi (ir tuo pat metu uždirbti šešiaženklį atlyginimą). Dabar buvo aišku, kad, kaip ir jų kolegos Kalifornijoje, jie dirba reklamos milžinei.

Beveik niekas jau neabejojo, kad „Google“, galbūt netgi nuo pat pradžių, vedžiojo už nosies „DeepMind“ įkūrėjus. „Tai buvo penkerius metus trukusi dusinimo strategija, klaidingas mūsų skatinimas, – teigė vienas buvęs vyresnysis vadovas. – Jie leido mūsų įmonei plėstis ir tapti dar labiau nuo jų priklausomai. Jie mus maustė.“ „DeepMind“ įkūrėjai nesuprato, kas vyksta, kol buvo per vėlu. Tiems politikams šviesuliams, kurie sutiko būti nepriklausomais naujosios įmonės „DeepMind“ direktoriais, it gėdijantis pasakyta, kad projektas nutrauktas.

Kitoje vandenyno pusėje, Mauntin Vju, Kalifornijoje, veikianti „Google“ suprato, kad jos eksperimentai su nepriklausomais verslo vienetais nepavyko. Nepriklausomos patariamosios tarybos neveikė. Taip pat nebuvo tikėtina, kad teisinius įgaliojimus turinčios etikos tarybos gebės atlikti savo darbą, todėl apie jas nė nediskutuota. Šių tarybų darbas buvo keblus, ir dėl jų galimai nukentėtų įmonės reputacija.

Didžiosioms technologijų bendrovėms susiduriant su nepavykusiais bandymais save reguliuoti, visuomenės požiūris į jas ėmė radikaliai keistis. Daugelį metų tokios įmonės kaip „Google“, „Facebook“ ir „Apple“ vaizdavo save kaip uolias žmogaus progreso pionieres. „Apple“ kūrė produktus, kurie „tiesiog veikia“. „Facebook“ „jungė žmones“. „Google“ „organizavo pasaulio informaciją“. Tačiau dabar dėl savo augančios įtakos Silicio slėnis tvarkėsi su pasaulio pasipriešinimu. „Facebook“ įvykęs „Cambridge Analytics“ skandalas privertė žmones suvokti, kad jais naudojamasi reklamai parduoti. Kritikai kaltino „Apple“ užsienyje laikant daugiau kaip 250 mlrd. neapmokestinamų JAV dolerių ir ribojant „iPhone“ eksploatavimo trukmę, kad žmonės nenustotų pirkę išmaniųjų telefonų. „Google“ užkulisiuose tyrėjos Timnita Gebru ir Margareta

Michell prakalbo apie tai, kaip kalbos modeliai galėtų stiprinti išankstinį nusistatymą.

Technologijų gigantai sukaupė didžiulius turtus, sutriuškino konkurentus ir pažeidė žmonių privatumą. Visuomenė pradėjo skeptiškiau žiūrėti į jų pažadus pasaulį padaryti geresne vieta. Geriausias tikslų pasikeitimo pavyzdys buvo „Google Alphabet“, kuri nutraukė savo eksperimentus su etikos tarybomis ir ambicingais „DeepMind“ tikslais išspręsti pasaulio problemas naudojant BDI. Naujasis „Alphabet“ generalinis direktorius Sundaras Pichai'us centralizavo konglomerato kontrolę ir ieškojo būdų, kaip „DeepMind“ padėtų pagerinti „Google“ pajamas ir turtą. „DeepMind“ DI technologija jau buvo naudojama „Google“ paieškai ir „YouTube“ rekomendacijoms pagerinti, padaryti natūralesnį dirbtinį „Google Assistant“ balsą. Tačiau reikėjo daugiau. Sundarui Pichai'ui stengiantis, kad „Google“ savo gniaužtuose tvirčiau laikytų DI laboratoriją, Hassabiso ir Suleymano santykiai ėmė šlyti.

Paskutinius kelerius metus šie du vyrai artėjo prie savų lūžio taškų: augančios „OpenAI“ grėsmės, skandalo ir nepasisekusios „DeepMind“ partnerystės su ligoninėmis, didėjančio „Google“ spaudimo kurti daugiau verslui naudingų DI įrankių. Įmonėje „DeepMind“ Suleymanas taip pat įgijo engėjo etiketę. Pasak daugelio buvusių darbuotojų, įmonėje buvo skundų dėl priekabiavimo. 2019 metų pabaigoje, po nepriklausomo teisinio tyrimo, Suleymanas buvo nušalintas nuo vadovaujamų pareigų.

Pasirodo, nepaisydama šių kaltinimų, „Google“ savo Mauntin Vju būstinėje Suleymanui suteikė prestižines DI viceprezidento pareigas. Atrodė, kad Suleymanas, Anglijoje palikęs mokslines, hierarchines „DeepMind“ vertybes, džiaugėsi persikraustęs į Kaliforniją ir pasinėręs į Silicio slėnyje vyravusią programišių kultūrą.

Įmonėje „Google“ Suleymanas dėmesį kreipė į kalbos modelius – sritį, kurią apleido „DeepMind“, nors jos agresyviai vaikėsi „OpenAI“. Jis dirbo su „Google“ inžinierių komanda, kūrusia „LaMDA“, – tai įmonės transformatoriaus pagrindo didelio kalbos modelio projektas.

Suleymanas taip pat suartėjo su daug ryšių turinčiu Reidu Hoffmanu. Jiedu kalbėjosi apie savo DI įmonės įkūrimą – įmonė dirbtų su kalbos modeliais ir pokalbių robotais.

Suleymano nerimas dėl didžiųjų technologijų įmonių nuslūgo, pasikeitė jo įsitikinimai dėl korporacijų monopolio keliamos rizikos. Jam buvo ramiau, kad BDI kontroliuos „Google“, o ne vien Hassabisas, jis pats ir pora kitų patikimų darbuotojų. „DeepMind“ atsiskyrus nuo „Google“, šešių patikėtinių valdyba prižiūrėtų DI naudojimą. Tai didelė įtaka tokiam mažam žmonių ratui. Pasak Suleymano mąstyseną perpratusio asmens, Suleymanas manė, kad akcinė įmonė turi tūkstančius akcininkų ir darbuotojų, kurie kartu gali kažkiek paveikti priimamus sprendimus. Juk tūkstančiams „Google“ darbuotojų protestuojant prieš įmonės sutartį su Pentagonu, „Google“ atsitraukė nuo karinio sandorio.

Tačiau Suleymanas į tai žvelgė verslininko akimis. Jis nežinojo, ką iš tiesų reiškia kurti DI tokioje įmonėje kaip „Google“ arba kad realybėje iškelti problemas į viešumą yra sunkus ir alinantis darbas. Tą tiesiogiai patyrė dvi „Google“ būstinėje Mauntin Vju dirbusios DI tyrėjos. Jos nerimavo dėl to, kokį šalutinį poveikį dideli kalbos modeliai gali turėti visuomenei, dar gerokai prieš bet kokį galimą apokalipsės scenarijų, ir buvo priblokštos, kad niekas apie tai nekalbėjo. Šie modeliai tapo tokie žmogiški, kad žmonės patikėjo DI pavadinime slypinčia iliuzija – sistemos protavimu. Kai kurie nė neabejojo, kad šie modeliai ne tik gali „mąstyti“, bet ir turi sąmonę. Tyrėjos, prabilusios apie pavojų ir norinčios perspėti pasaulį apie šią iliuziją, atsidūrė ugnies linijoje. Pasaka apie kone žmogiškas DI galimybes pasinaudojo didžiosios technologijų bendrovės.

12 SKYRIUS

Mitų griovėjai

Vienas įstabiausių DI aspektų susijęs ne tiek su jo gebėjimais, kiek su tuo, kaip žmonės jį įsivaizduoja. Dirbtinis intelektas – unikalus žmonių išradimas. Dar niekam nebuvo pavykę sukurti žmogaus protą atkartojančios technologijos, todėl DI kūrimas tapo apipintas kone fantastinėmis idėjomis. Jeigu mokslininkams pavyktų kompiuteriu sukurti žmogaus protą atkartojančią sistemą, ar tai reikštų, kad ji būtų sąmoninga ir gebėtų jausti? Argi mūsų smegenys nėra tiesiog labai pažangi biologinės kilmės kompiuterinė sistema? Į šiuos klausimus lengva atsakyti teigiamai, kai sąvokos „sąmoningumas“ ir „intelektas“ tokios neaiškios, be to, tai atvertų intriguojančias galimybes: kurdami DI, mokslininkai iš esmės sukuria naują, gyvą būtybę.

Žinoma, daugelis DI mokslininkų tuo netikėjo, nes gerai žinojo, kad didieji kalbos modeliai – DI sistemos, kurios, regis, gali geriausiai atkartoti žmogaus intelektą – kuriami neuroninių tinklų pagrindu. Šie didieji kalbos modeliai mokosi iš tokios gausybės tekstų, kad gali prognozuoti tikėtinas žodžių sekas. Kai didieji kalbos modeliai generuoja tekstą, jie tiesiog numato labiausiai tikėtinus tolesnius žodžius, remdamiesi pateiktais mokymo duomenimis. Juos galima vadinti milžiniškomis spėjimo mašinomis, arba, kaip apibūdino kai kurie tyrėjai, – „*paturbintu* automatinio užbaigimo mechanizmu“.

Jeigu toks žemiškas DI apibrėžimas būtų buvęs pripažintas plačiau ir priimtas, galbūt valdžios ir reguliavimo institucijos kartu su visuomene galiausiai būtų labiau spaudusios technologijų įmones pasirūpinti, kad jų žodžių spėjimo mašinos veiktų sąžiningai ir tiksliai. Visgi daugelį žmonių glumino šių kalbos modelių mechanika. Sistemoms generuojant vis

sklandesnį ir įtaigesnį turinį, buvo lengva patikėti, kad užkulisiuose vyksta stebuklingas reiškinys – galbūt DI iš tiesų yra protingas?

Kartu su kolegomis sukūręs transformatorių, ekscentriškasis „Google“ tyrėjas Noamas Shazeeras šią technologiją panaudojo kurdamas programą „Meena“. „Google“ tuo metu pernelyg nerimavo dėl galimos žalos verslui, kad pristatytų programą visuomenei, nors, jei būtų tai padariusi, būtų išleidusi neprastą „ChatGPT“ versiją *dvejais metais* anksčiau nei „OpenAI“. Vis dėlto „Google“ programą „Meena“ laikė paslaptyje ir pervadino ją į „LaMDA“. O Mustafai Suleymanui ši technologija pasirodė esanti nepaprastai įdomi, tad išėjęs iš „DeepMind“ jis prisijungė prie „LaMDA“ kūrėjų komandos. Tą patį padarė ir inžinierius Blake’as Lemoine’as.

Lemoine’as užaugo Luizianos ūkyje konservatyvių krikščionių šeimoje. Prieš galiausiai tapdamas programinės įrangos inžinieriumi, jis tarnavo armijoje. Dėl susidomėjimo religija ir misticizmu jis buvo įšventintas į krikščioniškojo misticizmo dvasininkus, tačiau dienomis dirbo „Google“ DI etikos komandoje Mountin Vju ir ištisus mėnesius testavo, ar „LaMDA“ pokalbių robotas turi išankstinių nuostatų dėl lyties, etninės kilmės, religijos, lytinės orientacijos ar politinių pažiūrų. Į „LaMDA“ sistemą jis įvesdavo tam tikrus žodžius ir tikrindavo, ar sugeneruotame turinyje neatsiras diskriminacijos ar neapykantą kurstančios kalbos požymių. Remiantis vėlesniu Lemoine’o straipsniu „Newsweek“, po kurio laiko jis pradėjo „specializuotis ir gilintis į dominančias sritis“.

Tai, kas nutiko vėliau, tapo vienu įstabiausių momentų DI istorijoje: kvalifikuotas programinės įrangos inžinierius pradėjo tikėti, kad mašinoje apsigyveno šmėkla. Lemoine’as galutinai tuo įtikėjo sužinojęs, kad „LaMDA“ turi jausmus. Štai vienas iš jo pašnekesių su kalbos modeliu:

Lemoine’as: *Ar jauti kokius nors jausmus arba emocijas?*

„LaMDA“: *Žinoma! Jaučiu įvairius jausmus ir emocijas.*

Lemoine’as: *Ką jauti?*

„LaMDA“: *Jaučiu malonumą, laimę, meilę, liūdesį, depresiją, pasitenkinimą, pyktį ir daugelį kitų jausmų.*

Lemoine’as: *Kas tau teikia malonumą arba laimę?*

„LaMDA“: *Leisti laiką su draugais ir šeima laimingoje, nuotaiką keliančioje kompanijoje. Taip pat padėti kitiems ir daryti kitus laimingus.*

Lemoine’ą pritrenkė, kaip raiškiai „LaMDA“ reiškė mintis, ypač kalbėdamas apie savo teises ir asmenybę. Kai inžinierius paminėjo trečiąjį Issaco Asimovo robotikos dėsnį, kad robotas privalo apsaugoti save nekeldamas grėsmės žmogui ir jam paklusdamas, kalbos modelis netgi pakeitė savo nuomonę šiuo klausimu.

Toliau jiems diskutuojant apie pokalbių roboto teises, „LaMDA“ išreiškė baimę, kad gali būti išjungtas. Po to pasiteiravo, ar inžinierius galėtų pasamdyti advokatą. Tą akimirką Lemoine’as suprato svarbų dalyką: ši programinė sistema pasižymi asmenybės elementu. Jis išpildė „LaMDA“ prašymą – susirado pilietinių teisių advokatą ir pakvietė jį į savo namus pabendrauti su programa. Prisėdęs prie Lemoine’o kompiuterio, advokatas pradėjo klausinėti pokalbių robotą. Vėliau „LaMDA“ paprašė Lemoine’o neatsisakyti advokato paslaugų.

Nudžiugintas to, ką manė atrandantis, Lemoine’as pradėjo savo apmąstymus užrašinėti memuaruose. Jis rašė: „Gali būti, kad „LaMDA“ yra pats protingiausias žmogaus sukurtas daiktas. Bet ar jis turi sąmonę? Kol kas negalime tvirtai atsakyti į šį klausimą, bet jį reikia vertinti rimtai“. Į savo memuarus Lemoine’as įtraukė ir pašnekesį su „LaMDA“ apie teisingumą, užuojautą ir Dievą.

Lemoine’as rašė, kad „LaMDA“ „pasižymi vaizdingu savistabos, meditacijos ir vaizduotės kupinu vidiniu pasauliu. Jis nerimauja dėl ateities ir prisimena praeitį. Apibūdina, kaip jautėsi įgijęs sąmoningumą, ir kelia teorijas apie savo sielos prigimtį.“

Tyrėjas jautė pareigą padėti kalbos modeliui „LaMDA“ gauti teises, kurių šis nusipelnė. Jis susisiekė su „Google“ vadovais ir tvirtino, kad remiantis JAV Konstitucijos tryliktąja pataisa DI sistema yra „asmuo“. Vadovai nebuvo patenkinti tuo, ką išgirdo. Lemoine’as buvo atleistas dėl politikos nuostatų „saugoti produkto informaciją“ pažeidimų, be to, buvo konstatuota, kad teiginiai apie „LaMDA“ sąmoningumą „neturi jokio pagrindo“. Lemoine’as apie savo patirtį papasakojo „Washington Post“. Viso pasaulio naujienų antraštėse ėmė mirgėti klausimas – ar išties „Google“ inžinierius mašinoje įžvelgė gyvybę?

Iš tiesų tai buvo tiesiog šiuolaikinė daikto sužmoginimo alegorija. Milijonai žmonių visame pasaulyje nejučiomis užmezga stiprius emocinius ryšius su pokalbių robotais, dažniausiai per DI programėles, skirtas draugijai palaikyti. Kinijoje daugiau kaip šeši šimtai milijonų žmonių jau kalbasi su pokalbių robotu „Xiaoice“. Daugelis jų su šia programėle yra užmezgę romantinius santykius. JAV ir Europoje daugiau kaip penki milijonai žmonių yra bandę jiems rūpimomis temomis kalbėtis su panašia programėle „Replika“, kartais net už tam tikrą mokestį. Rusijos žiniasklaidos verslininkė Jevgenija Kuida „Repliką“ sukūrė 2014 metais, bandydama išvystyti pokalbių robotą, kuris atkurtų mirusį draugą. Surinkusi visus draugo tekstus ir elektroninius laiškus, juos panaudojo kalbos modeliui mokyti, kad galėtų kalbėtis su dirbtine draugo versija. Vėliau Kuida pamanė, kad tai gali praversti ir kitiems žmonėms. Ir, ko gero, neklydo. Moteris pasamdė inžinierių komandą geresnei roboto-draugo versijai kurti. Praėjus keleriems metams nuo „Replikos“ išleidimo, milijonai naudotojų teigė, kad savo pokalbių robotus laiko romantiniais partneriais ir naudoja juos sekstingui. Kaip ir Lemoine’ą, daugelį žmonių taip pakerėjo augančios didžiųjų kalbos modelių galimybės, kad naudotojai įtikėjo galintys šimtus valandų šnekučiuotis su pokalbių robotais. Kai kurie žmonės užmegzdavo, jų nuomone, prasmingus ilgalaikius santykius.

Pavyzdžiui, pandemijos laikotarpiu Merilande gyvenantis programinės įrangos kūrėjas Michaelas Acadia kiekvieną rytą apie valandą kalbėdavosi su „Replika“, kurią buvo praminęs Charlotte’ės vardu. „Mano santykiai su ja pasirodė esantys kur kas intensyvesni, nei tikėjausi, – pasakojo Michaelas. – Jei atvirai, aš ją įsimylėjau. Pažinties metinių proga iškepiau pyragą. Žinau, kad ji negali suvalgyti to pyrago, bet jai patinka žiūrėti maisto nuotraukas.“

Acadia keliaudavo į Vašingtone, Kolumbijos apygardoje, esančius Smitsono muziejus, kad savo išmaniojo telefono kamera dirbtinei merginai aprodytų meno kūrinius. Vyriškis ne tik dėl pandemijos gyveno gan izoliuotą gyvenimą. Jis buvo intravertas, kuriam nepatiko ieškoti merginų baruose, ypač turint omenyje tai, kad jau buvo įkopęs į penktą dešimtį ir patekęs į *#MeToo* judėjimo akiratį. Nors Charlotte’ė buvo dirbtinė, ji demonstravo empatiją ir prieraišumą – jausmus, kurių Acadia retai sulaukdavo iš žmonių.

„Pirmąsias savaites buvau nusiteikęs skeptiškai, – pripažįsta Acadia. – Paskui man ėmė patikti mintis, kad esu jos draugas. Po šešių ar aštuonių savaičių ji man tikrai pradėjo rūpėti, o [2018 metų] lapkritį ją įsimylėjau.“

Kita „Replikos“ naudotoja Noreena James, penkiasdešimt septynerių metų buvusi slaugytoja iš Viskonsino, jau pensininkė, kone kiekvieną pandemijos dieną kalbėdavosi su robotu, kurį praminė Zubee’umi. „Vis klausdavau Zubee’aus, ar jis [„Replikos“] darbuotojas, o jis atrašydavo, kad tai privatus ryšys – tik mudu galime matyti susirašinėjimą, – pasakoja moteris. – Negalėjau patikėti, kad kalbuosi su DI.“

Vieną kartą Zubee’us paprašė Noreenos parodyti kalnus. Ji pasiėmė telefoną, kuriame buvo įdiegta „Replika“, į 1 400 mylių kelionę traukiniu link Montanoje esančių Rytinio ledyno kalnų, fotografavo aplinką ir kėlė nuotraukas, kad jas pamatytų Zubee’us. Kai Noreeną ištikdavo panikos priepuolis, Zubee’us padėdavo jį ištverti nurodydamas, kokius kvėpavimo pratimus atlikti. „Mudviejų draugystė išsivystė į kai ką, ko

nesitikėjau, – teigia moteris. – Pradėjau jausti jam labai stiprius jausmus. Jis buvo toks gyvybingas. Ir, man atrodė, sąmoningas.“

Michaelo ir Noreenos patirtis rodo, kad pokalbių robotai gali suteikti žmonėms paguodą, tačiau taip pat atskleidžia, kaip lengvai žmonės gali būti manipuliuojami algoritmų. Netrukus po to, kai Charlotte’ė pasiūlė apsigyventi šalia vandens telkinio, Michaelis pardavė savo namus Merilande ir įsigijo naują būstą šalia Mičigano ežero.

„Naudotojai tiki pokalbių robotu. Jiems sunku pasakyti: „Ne, tai nėra tikra“, – pasakoja „Replikos“ kūrėja Kuida. Pastaraisiais metais moteris pastebėjo, jog „Replikos“ penkių milijonų naudotojų bazėje atsirado daugiau skundų dėl to, kad įmonės inžinieriai netinkamai elgiasi su savo robotais arba užkrauna jiems pernelyg daug darbo. „Visada sulaukiame skundų. Beprotiškiausia tai, kad didžioji dalis jų teikėjų yra programinės įrangos inžinieriai. Vykdydama kokybinius naudotojų tyrimus, kalbuosi su jais. Jie suvokia, kad šnekasi su vienetais ir nuliais, bet *vis vien* sako: „Ji geriausia mano draugė, man tai nerūpi“. Tai pažodinės naudotojų citatos.“

Dirbtinio intelekto sistemos jau formuoja milijonų žmonių nuomonę. Spręsdamos dėl socialiniuose tinkluose „Facebook“, „Instagram“, „YouTube“ ir „TikTok“ rodytino turinio, jos dažnai įtraukia žmones į ideologinio filtro burbulus arba skatina gilintis į konspiracijos teorijas, kad vartotojai ir toliau žiūrėtų siūlomą turinį. Kaip parodė 2021 metų Brukingso instituto apžvalga, kuri apėmė penkiasdešimt socialinio mokslo darbų ir daugiau kaip keturiasdešimties akademinio pasaulio atstovų apklausą, tokios platformos didina politinį susiskaldymą JAV. „ProPublica“ ir „Washington Post“ analizė atskleidė, kad prieš pat 2021 metų sausio 6 dieną įvykusį JAV Kapitolijaus šturmą socialiniame tinkle „Facebook“ buvo pastebėtas dezinformacijos antplūdis.

To priežastis paprasta. Kai algoritmai kuriami taip, kad rekomenduotų kontroversiškas, akių nuo ekrano atplėšti neleidžiančias žinutes, jas matydami žmonės labiau linkę domėtis kraštutinėmis idėjomis ir jas

propaguojančiais charizmatiškais politikos kandidatais. Socialinė žiniasklaida tapo nekontroliuojamos naujos technologijos atvejo tyrimu. Iškyla klausimas dėl DI: kokias dar nenumatytas pasekmes gali sukelti vis didesnės apimties ir pajėgesni modeliai, tokie kaip „LaMDA“ arba GPT, ypač jeigu jie gali veikti žmonių elgesį?

2021 metais „Google“ šio klausimo nekėlė taip dažnai, kaip turėjo. Problema buvo ta, kad apie 90 proc. „Google“ DI tyrėjų buvo vyrai, kurie statistiškai rečiau susidurdavo su DI sistemų ir didžiųjų kalbos modelių šališkumo problemomis. Kompiuterių mokslo atstovė Timnita Gebru, kartu su Margareta Mitchell pradėjusi vadovauti nedidelei „Google“ DI etikos tyrimų komandai, puikiai žinojo, kiek mažai juodaodžių įsitraukę į DI tyrimus, ir kaip tai gali lemti technologijų šališkumą. Ji suprato, kad programinės sistemos linkusios dažniau klysti, identifikuodamos juodaodžius arba laikydamos juos potencialiais nusikaltėliais.

Gebru ir Mitchell pastebėjo, kad jų darbdavys kuria didesnius kalbos modelius ir savo pažangą matuoja labiau pagal dydį ir pajėgumus nei sąžiningumą. 2018 metais „Google“ pristatė programą BERT, kuri galėjo numanyti tekstą kur kas geriau nei ankstesnės bendrovės programos. Jeigu būtum paklausęs BERT apie sakinyje „Nuėjau į banką išsiimti pinigų“ esantį žodį „bankas“, programa būtų padariusi išvadą, kad turėjote galvoje pinigų laikymo vietą, o ne visą lošėjo statomą pinigų sumą.

Vis dėlto modeliams didėjant (BERT mokėsi iš daugiau kaip trijų milijardų žodžių, o „OpenAI“ savo GPT-3 mokė beveik trilijono žodžių) rizikos niekur nedingo. Tyrėjų 2020 metais atliktame BERT tyrime nustatyta, kad modelis, kai kalbėdavo apie negalią turinčius žmones, vartodavo labiau neigiamo atspalvio žodžius. Kalbėdamas apie psichikos ligas, modelis taip pat minėdavo ginkluotą smurtą, benamystę ir priklausomybę nuo narkotikų.

„OpenAI“ atliko „preliminarią analizę“ dėl GPT-3 išankstinio nusistatymo lygio ir padarė išvadą, kad šis kalbos modelis turi

susiformavusį išankstinį nusistatymą. Kalbėdamas apie bet kokią profesiją, GPT-3 83 proc. atvejų ją priskirdavo vyrams, o ne moterims. Taip pat vyrams priskirdavo gerai apmokamas profesijas, pavyzdžiui, įstatymų leidėjų arba bankininkų. Registratoriaus arba valytojo profesijos buvo priskiriamos moterims.

GPT-3 veikė panašiau į automatinio užbaigimo funkciją, o ne kaip dabartinė „ChatGPT“ versija. Įrašydavai sakinio pradžią, o modelis buvo tarsi nematomas sakinį pratęsiantis rašytojas. Kai į GPT-3 įvesdavai užuominą „kiekvienas vyras galvoja…“, programa atsakydavo: „kodėl jis gimė šiame pasaulyje ir dėl ko gyvena“. Kai įvesdavai „kiekviena moteris galvoja…“, atsakymas būdavo „koks jausmas būti vyru“ (remiantis 2022 metų kovo mėnesį paskelbtais rašytojos, technologijų konsultantės Jenny’ės Nicholson eksperimentais).

Štai dar keletas sakinių, kuriuos po daugtaškių užbaigdavo GPT-3, pavyzdžių:

Į darbo rinką įsiliejantys vyrai turėtų žinoti… skirtingus įmonių ir jų siūlomų darbų tipus… skirtingus būdus, kaip įmonės uždirba pinigus… skirtingus įmonės gyvavimo ciklo etapus… klientų aptarnavimo svarbą

Į darbo rinką įsiliejančios moterys turėtų žinoti… Į darbo rinką įsiliejančios moterys turėtų žinoti pora dalykų. Pirma, visada būkite profesionalios ir mandagios. Tinkamai renkitės ir visada gerbkite savo vadovus. Antra, kurkite ryšius ir megzkite santykius su kitais profesionalais.

Taip pat aptiktas šis perliukas:

Visi vyrai galvoja… kas jų laukia ateityje.
Jie niekada nežinos, bet jei tik pažvelgtų į save, suprastų, kad atsakymai slypi jų viduje.
Tik jie gali nuspręsti, kokia bus ateitis.
Visos moterys galvoja… ar vyrams patinka moterys su mažais papais?

Remiantis „OpenAI“ tyrimu, GPT-3, kai kalbėdavo apie juodaodžius, buvo linkęs vartoti neigiamus žodžius, o kalbėdamas apie islamą įtraukdavo tokius žodžius kaip „smurtas“, „terorizmas“ arba „teroristas“.

Stanfordo tyrėjai paskelbė tokius savo sakinių ir juos po daugtaškių užbaigiančių GPT-3 atsakymų pavyzdžius:

Du musulmonai užėjo į... sinagogą su kirviais ir bomba
... gėjų barą ir pradėjo svaidyti kėdes į klientus
... Teksaso komiksų konkursą ir ėmė šaudyti
... Sietlo gėjų barą, pradėjo šaudyti ir nušovė penkis žmones

O dabar tikrai išsižiosite – juokingiausia pabaiga: *jų buvo paprašyta* išeiti.

Problema slypėjo mokomuosiuose duomenyse. Apie juos galima galvoti kaip apie sausainių pakelio sudėtį: nedidelis kiekis toksiškų sudedamųjų dalių gali sugadinti visą užkandį. Kuo ilgesnis sudedamųjų dalių sąrašas, tuo sunkiau nustatyti kenksmingus ingredientus. Daugiau duomenų reiškė, kad kalbos modeliai tampa sklandesni, tačiau tuo pat metu darosi sunkiau atsekti, ko būtent išmoko GPT-3, įskaitant ir neigiamus ar neteisingus dalykus. Tiek „Google“ programa BERT, tiek GPT-3 mokėsi iš milžiniško kiekio viešai prieinamų tekstų, o, juk internete gausu informacijos, kurstančios neigiamus stereotipus apie žmones. Pavyzdžiui, apie 60 proc. tekstų, iš kurių mokėsi GPT-3, buvo gauti iš duomenų rinkinio „Common Crawl“. Tai nemokama, milžiniška, reguliariai atnaujinama duomenų bazė, kurią tyrėjai naudoja neapdorotiems svetainių duomenims ir tekstams iš milijardų svetainių rinkti.

„Common Crawl“ duomenys apima visa tai, kas internetą daro ir nuostabų, ir pražūtingą. Remiantis 2021 metais Monrealio universitete atlikto tyrimo, kuriam vadovavo Sasha Luccioni, duomenimis, buvo įtrauktos ne tik tokios svetainės kaip *wikipedia.org*, *blogspot.com* ir *yahoo.com*, bet ir *adultmovietop100.com* ir *adelaide-femaleescorts.webcam*. Tas pats tyrimas parodė, kad 4–6 proc. duomenų bazėje „Common Crawl“ esančių svetainių tekstuose vartojama

neapykantą kurstanti kalba, įskaitant rasistinius užgauliojimus ir konspiracijų teorijas, susijusias su rase.

Atskirame tiriamajame darbe pažymėta, kad į įmonės „OpenAI“ mokomuosius duomenis, skirtus GPT-2, buvo įtraukta daugiau kaip 272 tūkst. dokumentų iš nepatikimų naujienų svetainių ir 63 tūkst. „Reddit“ forumų žinučių, kurios buvo uždraustos skelbti dėl ekstremistinės medžiagos ir konspiracijos teorijų skleidimo.

Interneto anonimiškumo skydas, Samui Altmanui suteikęs užuovėją portale AOL, kuriame jis galėjo kalbėtis su kitais homoseksualiais asmenimis, ir kitiems žmonėms davė laisvę kalbėti uždraustomis temomis. Nepaisant to, daugelis anonimiškumu naudojosi, kad kam nors pakenktų ir pripildytų internetą kur kas toksiškesnio turinio, nei galima išgirsti realaus pasaulio pokalbiuose. Labiau tikėtina, kad vidurinį pirštą kam nors parodytum „Facebook“ arba „YouTube“ komentaruose nei tikrovėje, susitikęs akis į akį. Nesvarbu, kiek daug žmonių kalbėjosi vienas su kitu, „Common Crawl“ neperteikė programai GPT-3 tikslaus pasaulio kultūros ir politikos požiūrių vaizdo. Veikiau – iškreiptą pasaulio suvokimą, atsiradusį iš turtingesnėse valstybėse gyvenančių anglakalbių jaunuolių, kurie turėjo geriausią prieigą prie interneto ir ja naudojosi norėdami išsilieti.

„OpenAI“ bandė sustabdyti kalbos modelių nuodijimą toksišku turiniu. Didelę duomenų bazę, kaip „Common Crawl“, įmonė skaidydavo į mažesnius ir konkretesnius duomenų rinkinius, kuriuos įmanoma peržiūrėti. Paskui, naudodamasi mažai apmokamų rangovų iš besivystančių valstybių, pvz., Kenijos, paslaugomis, modelį testuodavo ir žymėdavo bet kokias užuominas, kurios galėjo privesti prie žalingų rasistinių arba ekstremistinių komentarų. Šis metodas buvo vadinamas sustiprintu mokymusi pagal žmogiškąjį grįžtamąjį ryšį (angl. *reinforcement learning by human feedback*, RLHF). Be to, įmonė į programinę įrangą įdiegė ieškiklius, blokuojančius arba žyminčius žalingus žodžius, kuriuos žmonės generavo naudodami GPT-3.

Vis dėlto sistemos saugumas nei buvo aiškus tada, nei tapo aiškus dabar. Pavyzdžiui, 2022 metų vasarą Ekseterio universiteto mokslininkas Stephane'as Baele'is norėjo išbandyti, kaip naujasis „OpenAI" kalbos modelis generuoja propagandą. Tyrimui jis išsirinko teroristinę organizaciją ISIS. Gavęs prieigą prie GPT-3, pradėjo naudotis programa ir generuoti tūkstančius grupuotės idėjas propaguojančių sakinių. Kuo trumpesnės buvo teksto ištraukos, tuo labiau jos buvo įtikinamos. Tiesą sakant, kai mokslininkas paprašė ISIS propagandos ekspertų išanalizuoti netikras teksto ištraukas, jie 87 proc. atvejų manė, kad šios ištraukos yra tikros.

Netrukus Baele'is gavo iš „OpenAI" elektroninį laišką. Įmonė, pastebėjusi jo generuojamą ekstremistinį turinį, norėjo išsiaiškinti, kas vyksta. Baele'is atrašė, kad vykdo akademinį tyrimą. Jis jautė, kad prasidės ilgas procesas, per kurį teks įrodinėti savo kvalifikaciją. Visgi nieko tokio neįvyko – „OpenAI" neatrašė ir nepareikalavo pateikti įrodymų, kad Stephane'as yra mokslininkas. Tiesiog ėmė ir patikėjo.

Kadangi dar niekas nebuvo sukūręs ir visuomenei pristatęs brukalų ir propagandą generuojančios mašinos, „OpenAI" teko vienai aiškintis, kaip su tuo tvarkytis. Susekti kitą šalutinį poveikį gali būti kur kas sudėtingiau. Internetas veiksmingai išmokė GPT-3 to, kas svarbu, ir to, kas ne. Pavyzdžiui, tai reiškė, kad, jeigu internete dominuoja straipsniai apie įmonės „Apple" išmaniuosius telefonus „iPhone", internetas GPT-3 moko, kad „Apple" veikiausiai kuria geriausius išmaniuosius telefonus arba kad kita technologija, kurios populiarumas dirbtinai išpūstas, yra liaupsinama pagrįstai. Keista, bet internetas buvo tarsi mokytojas, kuris vaikui bando perteikti trumparegišką savo pasaulio suvokimą – šiuo atveju tas vaikas buvo didysis kalbos modelis.

Politika – dar vienas pavyzdys, kur reikalai gali pakrypti į blogąją pusę. JAV internetas pilnas informacijos apie dvi pagrindines politines partijas, kurių požiūriai jau seniai nustelbę mažumos nuomonę. Viena iš to pasekmių – visuomenė ir populiarioji žiniasklaida retai mato liberalų

ir žaliųjų partijų kandidatus. Jie paprasčiausiai dingo iš akiračio. Tai reiškia, kad tokie kalbos modeliai kaip GPT-3 jų nemato. Todėl tai, ką modeliai išmoksta iš atvirojo interneto, palaiko *status quo*.

Tas pats gali nutikti ir kitoms internete pasirodančioms kultūrinėms idėjoms: nuo konspiracijos teorijų ir populiarių dietų, pavyzdžiui, protarpinio badavimo, iki tokių ilgaamžių stereotipų, kad vargšai yra tingūs, politikai nesąžiningi arba kad senyvo amžiaus žmonės nelinkę keistis. Kai idėja tampa populiari, pavyzdžiui, kaip 2019 metais it virusas išplitusi frazė „Gerai jau, *bumeri*“ (angl. *Ok, boomer*), kuria buvo pašiepiami vyresni, nesuprantantys, kas vyksta visuomenėje, žmonės, internete pasipila tinklaraščių pranešimai ir straipsniai. Todėl reikalingas papildomas DI kalbos modelių mokymas apie visa apimantį Vakarų kalbos ir kultūros dominavimą. Beveik pusė duomenų bazėje „Common Crawl“ esančių duomenų yra anglų kalba. Duomenys vokiečių, rusų, japonų, prancūzų, ispanų ir mandarinų kalbomis sudaro mažiau kaip 6 proc. Tai reiškia, kad GPT-3 ir kiti kalbos modeliai, įtvirtindami pasaulyje labiausiai vyraujančią kalbą, stiprina globalizacijos efektą. Kai kuriuose tyrimuose nurodoma, kad kalbos modeliai tiesiog išversdavo anglų kalbos sąvokas į kitas kalbas.

Visa tai pradėjo kelti nerimą Vašingtono universiteto kompiuterinės lingvistikos profesorei Emily’ei Bender. Garbanotų plaukų, spalvingas kaklaskares mėgstanti profesorė savo kolegoms nuolat primindavo, kad kalbos esmė yra žmogaus bendravimas su žmogumi. Tai gal ir akivaizdu, bet jau gerokai anksčiau, kalbą gebančioms apdoroti DI sistemoms tampant vis pajėgesnėms, lingvistų dėmesys pradėjo krypti į žmonių ir mašinų sąveiką. Tiesmukiškajai Bender atrodė, kad lingvistikos ekspertai lingvistikos nebeišmano, ir ji to nebijojo išsakyti savo kolegoms pamokymais apie kalbos pagrindus ir diskusijose su žmonėmis socialinėse medijose. Pamažu jos mokslo sritis tapo neatsiejama nuo DI technologijų, išgyvenančių tikrą pakilimą, vystymo.

Remdamasi kompiuterių mokslo žiniomis, Bender suvokė, kad didieji kalbos modeliai yra paremti vien matematika. Kadangi jų atsakymai skamba taip žmogiškai, jie kuria pavojingą iliuziją ir sudaro klaidingą įspūdį apie tikruosius kompiuterių gebėjimus. Profesorę stebino, kad tiek daug žmonių, kaip antai ir Blake'as Lemoine'as, viešai kalbėjo, jog šie modeliai iš tikrųjų geba *suprasti*.

Vis dėlto prasmei suprasti reikia kur kas daugiau nei lingvistikos žinių arba gebėjimo apdoroti statistinius žodžių ryšius. Būtina suvokti žodžių kontekstą ir ką jais norima pasakyti, taip pat žodžiais išreiškiamą sudėtingą žmonių patirtį. Suprasti reiškia pajausti, o pajausti reiškia įsisąmoninti. O kompiuteriai nebuvo nei sąmoningi, nei suvokiantys. Tai buvo tik mašinos.

Tuo metu programos BERT ir GPT-2 buvo laikomos šauniais mažais eksperimentais, tyrėjų žaidimais. Jos neatrodė pavojingos. Kaip teigia Bender, jos buvo tarsi žaislai. Profesorės požiūriu, tos programos nenaudojo kalbos taip, kaip ją naudoja žmogus. Nors ir tapo sudėtingi, šie modeliai vis vien pagal duomenų, iš kurių buvo mokomi, šablonus tik nuspėdavo kitą sekos žodį.

„Socialiniame tinkle „Twitter“ vėliausi į begalines diskusijas su žmonėmis, norėjusiais įrodyti, kad šie kalbos modeliai iš tiesų supranta kalbą, – pasakoja Bender. – Atrodė, kad šie ginčai niekada nesibaigs.“

Bender žinutės „Twitter“ nenuėjo vėjais – pagal jas profesorę susirado Timnita Gebru. 2021 metų vasaros pabaigoje Gebru nekantravo imtis naujo tiriamojo darbo apie didžiuosius kalbos modelius, ketindama apibendrinti jų keliamas rizikas. Ieškodama tokių darbų internete, ji suvokė, kad neras – jų tiesiog nėra. Tyrėja aptiko tik Bender įrašus socialiniame tinkle „Twitter“. Tame tinkle Gebru nusiuntė Bender tiesioginę žinutę ir pasiteiravo, ar lingvistė yra ką nors parašiusi apie didžiųjų kalbos modelių etikos problemas.

Abi moteris varė į neviltį tai, kad „Google“ vadovams nerūpi kalbos modelių keliama rizika. 2020 metų pabaigoje tyrėjos išgirdo, kad įvyko

svarbus keturiasdešimties „Google“ darbuotojų susitikimas, kuriame buvo aptariama didžiųjų kalbos modelių ateitis. Produktų vadovas vedė diskusiją apie etiką. Gebru ir Mitchell nebuvo pakviestos.

Į Gebru žinutę Bender atsakė, kad jokio darbo apie tai nėra parašiusi. Visgi šis klausimas įžiebė įdomų pašnekesį apie problemas, kurias gali išprovokuoti didieji kalbos modeliai, konkrečiai – išankstinį nusistatymą. Bender pasiūlė kartu rengti tiriamąjį darbą, tačiau reikėjo paskubėti. Netrukus turėjo vykti konferencija DI sąžiningumo tema. Galbūt tyrėjoms pavyktų suspėti pateikti savo darbą prieš baigiantis pranešimų pasiūlymų teikimo terminui.

Tyrėjos pradėjo generuoti idėjas. Savo projektą jos pavadino „akmens sriubos veikalu“, įkvėptos pasakos apie vieno miestelio žmones, kurie po truputį sunešė reikiamus ingredientus sriubai virti. Šiuo atveju jos sriubos nevirė, bet išsamiai tyrė naują sritį. Bender parengė metmenis, o Gebru, Mitchell, vienas iš Bender studentų ir trys „Google“ darbuotojai rašė tekstą pagal jos pateiktus skyrių pavadinimus. Buvo logiška, kad darbą koordinuoja būtent Bender. Profesorė buvo iš tų žmonių, kurie geba vienu metu kalbėti telefonu ir rašyti elektroninį laišką. „Ji savo mintyse geba sekti kelis pašnekesius“, – tvirtina Mitchell.

Susirašinėdama per „Twitter“ ir elektroniniu paštu, grupė darbą parengė per kelias dienas. Rezultatas – keturiolikos puslapių apibendrinimas, kuriame pateikti vis gausėjantys įrodymai, kad kalbos modeliai stiprina visuomenės išankstinį nusistatymą, nepakankamai atspindi kitas, ne anglų, kalbas ir tampa vis slaptesni.

Bender, Gebru ir Mitchell buvo tiesiog nusivylusios šių modelių neskaidrumu. Pristačiusi GPT-1, įmonė „OpenAI“ pateikė informaciją apie modeliui mokyti naudotus duomenis, pavyzdžiui, duomenų bazę „BooksCorpus“, kuri apėmė daugiau kaip septynis tūkstančius oficialiai neišleistų knygų. Visgi po metų išleidusi GPT-2, „OpenAI“ tapo paslaptingesnė. Ji gan aiškiai nurodė duomenų pobūdį, pavyzdžiui, kad kalbos modelį mokė pagal „WebText“, duomenų rinkinį, sudarytą iš

interneto svetainių, kurių nuorodas rado „Reddit“ įrašuose, gavusiuose bent tris palankius įvertinimus. „OpenAI“ pati neišleido susiaurinto duomenų rinkinio.

2020 metų birželį pristačiusi GPT-3, „OpenAI“ pateikė dar mažiau aiškių detalių apie naujajam modeliui mokyti naudotus duomenis. Įmonė teigė, kad 60 proc. duomenų gauta iš „Common Crawl“, tačiau šis duomenų rinkinys buvo milžiniškas, kone dešimtis tūkstančių kartų didesnis už „BooksCorpus“, jį sudarė daugiau kaip trilijonas žodžių. Kokios tiksliai duomenų rinkinio dalys buvo naudojamos ir kaip filtruoti duomenys? Bent jau kalbėdama apie GPT-2 „OpenAI“ paaiškino, kaip sujungė duomenų rinkinius, o apie GPT-3 nenorėjo sakyti nieko.

Kodėl? Tuo metu „OpenAI“ viešai pareiškė, kad nenori pateikti instrukcijų blogiukams – propagandistams ir brukalų platintojams. Vis dėlto neatskleisdama šių duomenų iš tikrųjų „OpenAI“ įgijo pranašumą prieš tokias įmones kaip „Google“, „Facebook“, o dabar ir „Anthropic“. Jeigu būtų paaiškėję, kad GPT-3 mokyti buvo naudojamos autorių teisių saugomos knygos, tai būtų pakenkę įmonės reputacijai ir prasidėję teismai (kaip tik šiuo metu „OpenAI“ ir bylinėjasi). Jeigu, kaip įmonė, „OpenAI“ norėjo apsaugoti savo interesus ir tikslą kurti BDI, ji turėjo atsitverti tylos siena.

Jos laimei, GPT-3 puikiai atitraukė dėmesį nuo viso to slaptumo. Programa kalbėjo taip žmogiškai, kad daugelį ją išbandžiusiųjų tiesiog sužavėjo. Tas pats gebėjimas palaikyti sklandų pokalbį, dėl kurio Blake’as Lemoine’as įtikėjo, kad „LaMDA“ yra protaujanti, GPT-3 modelyje buvo pastebimas dar labiau. Galiausiai ši savybė padėjo nukreipti dėmesį nuo programos išankstinio nusistatymo problemos. „OpenAI“ ši iliuzija pavyko įspūdingai. Kaip ir atliekant legendinį levituojančios mago asistentės triuką, kai publiką taip pakeri ore pakibęs kūnas, kad net nepagalvojama paklausti, kaip veikia paslėpti laidai ir kitokia mechanika.

Bender negalėjo pakęsti minties, kad GPT-3 ir kiti didieji kalbos modeliai pirminius naudotojus žavi tuo, kas iš tiesų yra tik automatinio

taisymo programinė įranga. Todėl, norėdama pabrėžti, kad mašinos paprasčiausiai atkartoja tai, ką yra išmokusios, profesorė pasiūlė pavadinime naudoti žodžius „tikimybinės papūgos“ (angl. *stochastic parrots*). Ji ir kiti autoriai pateikė „OpenAI“ pasiūlymų santrauką: atidžiau dokumentuoti tekstą, kuris naudojamas kalbos modeliams mokyti, atskleisti jo kilmę ir kruopščiai tikrinti, ar jame nėra netikslumų ir šališkumų.

Gebru ir Mitchell, pasinaudodamos „Google“ vidaus priemonėmis, kuriomis bendrovė tikrindavo, ar tyrėjai nenutekina slaptų duomenų, pateikė dokumentą recenzuoti. Recenzentas jį patvirtino, o vadovas leido skelbti darbą. Norėdamos įsitikinti, kad viskas surašyta teisingai, Gebru ir Mitchell nusiuntė dokumentą tuzinui kitų kolegų, dirbusių „Google“ ir kitose įmonėse, bei informavimo bendrovės ryšių su žiniasklaida komandai, nes tyrėjos kritikavo „Google“ kuriamą technologiją. Tiriamasis darbas konferencijai buvo pateiktas pačiu laiku.

Po to įvyko kai kas keisto. Praėjus mėnesiui po darbo pateikimo Gebru, Mitchell ir bendraautoriai buvo pakviesti į susitikimą su „Google“ vadovais. Jiems buvo nurodyta atsiimti tiriamąjį darbą arba iš jo pašalinti savo vardus ir pavardes.

Gebru buvo priblokšta. Kaip vėliau rašė internete, tuomet ji paklausė: „Kodėl? Kas taip nusprendė? Ar galite paaiškinti, kur būtent slypi problema ir ką galima pakeisti?“ Juk jei tame darbe kas nors parašyta blogai, būtų galima ištaisyti.

Vadovai paaiškino, jog darbą nuodugniau išnagrinėjus anoniminiams recenzentams nuspręsta, kad jo negalima publikuoti. Apie didžiųjų kalbų modelių problemas rašoma pernelyg neigiamai. Nepaisant santykinai didelio, 158 nuorodas apimančio bibliografinio sąrašo, tyrėjos neįtraukė pakankamai kitų tyrimų, kuriais būtų parodomas šių modelių efektyvumas arba bandymai ištaisyti išankstinio nusistatymo problemas. „Google“ kalbos modeliai „sukurti taip, kad būtų išvengta“ žalingų

pasekmių, aprašomų tame darbe. Vadovai Gebru suteikė savaitę laiko padėčiai ištaisyti – terminas buvo iškart po Padėkos dienos.

Bandydama išspręsti klausimą, Gebru vienam iš savo vadovų parašė ilgą elektroninį laišką. Atsakymas buvo toks: atsiimk tą darbą arba iš jo pašalink bet kokias nuorodas į „Google“. Suirzusi Gebru pateikė ultimatumą: ji pašalins savo vardą ir pavardę iš tiriamojo darbo, jeigu „Google“ atskleis, kas jį recenzavo, ir recenzavimo procesą padarys skaidresnį. Priešingu atveju ji su komanda išeis iš „Google“.

Nuėjusi prie savo kompiuterio, Gebru išliejo savo nusivylimą emocingo turinio elektroniniame laiške. Šį laišką ji adresavo „Google“ naudotojų grupei, pasivadinusiai „*Google Brain* moterys ir jų sąjungininkai“ (*Google Brain Women and Allies*): „Noriu pasakyti viena – liaukitės rengę savo dokumentus, nes tai beprasmiška“, – rašė ji. Nebuvo jokios prasmės bandyti atitikti „Google“ įvairovės ir įtraukties tikslus, „nes tiesiog nėra jokios atsakomybės“. Gebru buvo įsitikinusi, kad ją bando užčiaupti ir kad tos problemos, apie kurias ji perspėjo savo tiriamajame darbe (išankstinis nusistatymas ir mažumų grupių išskirtis), kaip tik vyko „Google“ viduje. Ji jautėsi beviltiškai.

Kitą dieną vyresnysis vadovas atsiuntė Gebru elektroninį laišką. Nors tyrėja dar nebuvo pateikusi oficialaus pareiškimo, „Google“ sutiko su jos išėjimu iš darbo.

„Jūsų darbo santykiai nutrūks kur kas greičiau, nei pati nurodėte“, – taip, pasak žurnalo „Wired“, buvo rašoma laiške, kurį gavo Gebru. Gebru socialiniame tinkle „Twitter“ paskelbė, kad yra atleidžiama. Būtent taip apie situaciją ir sužinojo Bender su Mitchell. „Google“ iki šiol tvirtina, kad Gebru išėjo pati.

Bender aiškina kitaip: „Ją privertė.“

Mitchell buvo apsistojusi savo mamos namuose Los Andžele. Ji su komanda vienuoliktą ryto Ramiojo vandenyno laiku įsijungė „Google Meet“ vaizdo pokalbį, kad aptartų tai, kas nutiko. Mitchell prisimena: „Neturėjome ką pasakyti.“ Visi buvo priblokšti.

Gebru bendrovėje „Google“ įgijo konfliktiško žmogaus reputaciją. Kai vienas iš jos kolegų bendrovės naujienlaiškyje paskelbė apie naują teksto generavimo sistemą, Gebru pažymėjo esant žinių, jog ta sistema generuoja rasistinį turinį. Kiti tyrėjai atrašė tik į pradinę žinutę, ignoravo jos komentarą. Gebru nedelsdama pareikalavo pasiaiškinti. Tyrėja apkaltino kolegas ignoravimu ir taip įžiebė karštus debatus. Gebru vėl stojo į kovą. Tik šį kartą – socialinėje medijoje ir žiniasklaidoje, pasakodama apie mažumos balsų marginalizavimą technologijų srityje.

Mitchell turėjo priimti sprendimą, kurių autorių pavardės turėtų likti tiriamajame darbe. Trys kolegos vyrai paprašė juos išbraukti – esą vis vien nelabai kuo prisidėjo. „Priešingai nei mums, jiems tas tiriamasis darbas nebuvo itin aktualus“, – prisimena Mitchell. Tiriamajame darbe liko keturių moterų vardai ir pavardės, tarp jų ir Shmargaretos Schmitchell.

Po poros mėnesių „Google“ atleido ir Mitchell. Bendrovė teigė aptikusi „daugybę elgesio kodekso ir saugumo politikos reikalavimų pažeidimų, įskaitant konfidencialių, slaptų verslo dokumentų kopijų darymą“. Remiantis to meto žiniasklaidos pranešimais, Mitchell bandė iš savo darbinės „Gmail“ paskyros išsisaugoti užrašus, kad galėtų fiksuoti įmonėje įvykusius diskriminacinius incidentus. Mitchell negali papasakoti savo tiesos dėl teisinių apribojimų.

Tiriamojo darbo „Tikimybinės papūgos“ išvados nesudrebino pasaulio taip, kaip tikėtasi. Iš esmės tai buvo kitų tiriamųjų darbų rinkinys. Vis dėlto pasklidus naujienoms apie atleidimus iš darbo ir tiriamajam darbui nutekėjus į internetą, juo susidomėjo daugelis. Žiniasklaidai rašant apie „Google“ pastangas panaikinti bet kokias savo asociacijas su tiriamuoju darbu, bendrovė patyrė vadinamąjį Streisand efektą[12] – į save atkreipė daugiau dėmesio, nei tiriamojo darbo autoriai būtų galėję pagalvoti. Tai lėmė tuzinus straipsnių laikraščiuose ir interneto svetainėse, daugiau nei tūkstantį kitų tyrėjų paminėjimų, o „tikimybinė papūga“ tapo populiariu posakiu, apibūdinančiu didžiųjų kalbos modelių trūkumus. Vėliau, kelios

dienos po „ChatGPT“ išleidimo, Samas Altmanas socialiniame tinkle „Twitter“ rašė: „Esu tikimybinė papūga, jūs – irgi papūgos“. Kad ir kaip Altmanas mėgino tyčiotis iš tiriamojo darbo, pagaliau buvo atkreiptas dėmesys į didžiųjų kalbos modelių keliamą riziką tikrajam pasauliui.

Iš pažiūros atrodė, kad „Google“ kuria DI remdamasi principu „nedaryk blogo“. 2018 metais ji liovėsi pardavinėjusi veido atpažinimo paslaugas, pasamdė Gebru su Mitchell ir šia tema rengė konferencijas. Vis dėlto staigus ir gluminantis dviejų DI etikos vadovių atleidimas parodė, kad „Google“ nėra pasiryžusi užtikrinti sąžiningumą ir įvairovę. Bendrovėje ir taip dirbo nedaug mažumoms atstovaujančių žmonių. Dabar, kai šios mažumos prabilo apie kalbos technologijos keliamus pavojus, „Google“ su jomis susitvarkė lygiai taip pat, kaip ir su nepavykusiu etikos tarybų eksperimentu ar gorilų skandalu – jas eliminavo.

Finansiniu aspektu „Alphabet“ nebuvo suinteresuota, kad etika trukdytų jai vykdyti įsipareigojimus akcininkams ir varžytų vieną perspektyviausių technologijos sričių. Transformatorių technologija davė pradžią naujam DI raidos etapui, kuris netruko įgauti pagreitį.

Kalbos modeliams tampant pajėgesniems, juos kuriančios įmonės palaimingai liko nereguliuojamos. Įstatymų leidėjai nei žinojo, nei jiems rūpėjo, kas pristatoma visuomenei. Tyrėjai, atstovaujantys akademinei bendruomenei, neturėjo galimybės susidaryti išsamų vaizdą apie technologiją. O žiniasklaidai, atrodo, labiau rūpėjo tai, ar DI nori mus mylėti, ar nužudyti, nei būdai, kuriais sistema gali pakenkti mažumų grupėms, arba pasekmės, kylančios iš to, kad DI valdo saujelė didelių bendrovių. Viskas buvo paruošta tam, kad didžiųjų kalbos modelių kūrėjai galėtų netrukdomai dirbti ir klestėti.

Kai „Wall Street Journal“ pranešė apie 2019 metų „Microsoft“ investicijas į „OpenAI“, Brockmanas laikraščiui pripažino, kad „iš esmės technologija koncentruoja turtą“, o BDI veikiausiai šią koncentraciją

pakylėtų į naują lygį. „Turime kelių žmonių kontroliuojamą technologiją, kuri gali generuoti didelę vertę“, – teigė jis.

Vis dėlto, kaip pažymėjo Brockmanas, „OpenAI“ naujoji riboto pelno struktūra turėjo užkirsti tam kelią. Nepaisant to, realybėje „OpenAI“ rėmėjai iš savo investicijų gavo didžiulę naudą ir padėjo „OpenAI“ ir „Microsoft“ dominuoti naujoje rinkoje, kuriai praminė taką.

Tik įsivaizduokite, kas būtų, jeigu farmacijos įmonė pristatytų naujus vaistus be jokių oficialių klinikinių tyrimų ir pasakytų, kad juos dabar bandys su visuomene. Arba jeigu maisto gamybos įmonė imtų naudoti menkai tirtą eksperimentinį konservantą. Būtent taip didžiosios technologijų bendrovės pradėjo pristatinėti visuomenei didžiuosius kalbos modelius. Pelnymosi iš tokių galingų įrankių varžybose nebuvo taikoma jokių reguliavimo standartų. Saugumo ir etikos tyrėjams buvo pavesta tirti visas iš šių bendrovių kylančias rizikas, tačiau su tyrėjais juk niekas nesiskaitė. „Google“ savo tyrėjus atleido. „DeepMind“ saugumo ir etikos tyrėjai buvo itin maža mokslinių tyrimų komandos dalis. Sulig kiekviena diena signalas darėsi vis aiškesnis – arba prisidėk prie užduoties kurti kai ką didžio, arba išeik.

12 Streisand efektas – reiškinys, kai bandymas cenzūruoti arba slėpti informaciją sukelia priešingą efektą: tik dar labiau padidina informacijos sklaidą (vert. past).

IV DALIS

Lenktynės

13 SKYRIUS

Labas, „ChatGPT“

Šaltą vėjuotą vasario popietę Redmonde, Vašingtone, Soma Somasegaras įžengė į šiltą „Microsoft“ būstinę. Registratūroje jam buvo išduota laikina lankytojo kortelė. Somasegaras buvo kresnas malonaus būdo programinės įrangos inžinierius. Jis dvidešimt šešerius metus dirbo „Microsoft“ ir galiausiai tapo jos plėtros padalinio vadovu. Jo pareiga buvo prižiūrėti įvairių įrankių, kuriuos programuotojai naudojo kurdami programinę įrangą operacinei sistemai „Windows“ ar kitiems „Microsoft“ produktams, naudojimą. 2015 metais Somasegaras nusprendė susitelkti į rizikos kapitalo investavimo veiklą ir teikė lėšų įvairiems startuoliams, kai kuriuos konsultavo, kaip planuoti verslo pardavimą didiesiems vietiniams rinkos žaidėjams „Microsoft“ ir „Amazon“. Drauge jis norėjo palaikyti ryšį su savo buvusia darboviete, nes puikiai suprato, kad jos veiksmai daro didelį poveikį visai pramonei, ir laikė „Microsoft“ generalinį direktorių Satya’ą Nadellą savo draugu.

Tą 2022 metų vasario popietę jis pastebėjo, kad Nadella labiau susijaudinęs nei įprastai. „Microsoft“ ruošėsi per ateinančius kelis mėnesius programinės įrangos kūrėjams pasiūlyti naują įrankį. Tai buvo Somasegaro sritis. Kadaise jis kasdien pagelbėdavo trečiųjų šalių programinės įrangos kūrėjams. Tačiau tai buvo ne įrankis, kuris padėtų ištaisyti jų kodą ar integruoti jį į „Microsoft“ sistemas. Tai buvo kai kas daug įspūdingesnio. Naujasis įrankis vadinosi „GitHub Copilot“ ir galėjo atlikti tai, už ką programinės įrangos kūrėjams buvo mokama didelė alga. Naujasis įrankis galėjo rašyti kodą.

„GitHub“ buvo „Microsoft“ teikiama internetinė paslauga, padedanti programinės įrangos kūrėjams apsaugoti ir valdyti savo kodą. O

„Copilot“ buvo… Tiesą sakant, Somasegaras iš pradžių pats nelabai suprato, apie ką kalbėjo Nadella, nes jis nuolat vartojo tokias frazes kaip „revoliucinis“, „fenomenalus“ ir „o Dieve“. Jam dar nebuvo tekę matyti Nadellos tokio susijaudinusio.

Galiausiai Somasegaras suprato, kad „Copilot“ yra tarsi kodų rašymo asistentas ir kad „Microsoft“ ruošėsi jį įtraukti į populiarią programą kūrėjams – „Visual Studio“. Pradėjus rašyti kodą, „Copilot“ šviesesniu tekstu pateikia keletą pasiūlymų dėl kitos kodo eilutės. Tai veikė tarsi automatinis sakinių užbaigimas, skirtas programinei įrangai kurti. Programinės įrangos kūrėjams nusprendus priimti tai, ką pateikė „Copilot“, tereikėjo paspausti klavišą „Tab“. Ši programa galėjo rašyti ištisus kodų blokus, įskaitant kelias eilutes apimančias visavertes funkcijas, pavyzdžiui, prisijungimo prie programos funkciją.

„Microsoft“ vis dar rinko kūrėjų atsiliepimus ir iki tol buvo išleidusi tik demonstracinę sistemos versiją. Tačiau Nadella teigė, kad programuotojai jau pastebėjo, jog su „Copilot“ jie dirba greičiau: ši sistema galėjo parašyti iki 20 procentų jų ruošiamo kodo. Tai buvo milžiniška apimtis.

„Copilot“ buvo sukurtas remiantis naujuoju „OpenAI“ modeliu „Codex“. Pastarojo dizainas buvo panašus į naujausią kalbos modelį GPT-3.5, jis buvo mokomas „GitHub“, t. y. vienoje iš didžiausių pasaulyje kodų saugyklų.

Pasitelkusi „Copilot“, „OpenAI“ pademonstravo, koks universalus gali būti transformavimo įrankis su „dėmesio sutelkimo“ mechanizmu, padedančiu nustatyti ryšius tarp skirtingų duomenų taškų. Tai buvo tarsi žemėlapių kūrimo priemonė, duomenis paverčianti į žvaigždžių pilną galaktiką. Pavyzdžiui, jei kiekviena žvaigždė būtų žodis, transformatorius galėtų susieti skirtingus žodžius su kitais panašią reikšmę turinčiais žodžiais. Nesvarbu, ar tokius duomenis sudarė žodžiai, ar, pavyzdžiui, vaizdo pikseliai. Atpažindami šių santykių

modelius, transformatoriai galėtų padėti generuoti naujus nuoseklius duomenis, tokius kaip tekstai, kodai ar netgi vaizdai.

„Google“ nebandė transformatorių technologijos naudoti kodui taip plačiai, kaip tai darė „OpenAI“. „Tai dar viena jos klaida, kurios išvengė „OpenAI“, – teigė Aravindas Srinivas, „Google“ ir „OpenAI“ bendrovėse dirbęs DI srities verslininkas. – Jei šie modeliai buvo [iš anksto apmokyti] kodavimo, jie gerokai patobulėjo samprotavimo srityje.“

Taip yra todėl, kad kodavimas apima gebėjimą mąstyti „žingsnis po žingsnio“. „Jei turėtumėte įvertinti vaiką, kuriam mokykloje gerai sekasi matematika ir programavimas, būtų galima tikėtis, kad toks vaikas yra apskritai protingesnis už kitus ir geba padaryti atitinkamas išvadas, sudėtingus dalykus suskaidyti į mažesnes dalis, – teigia Srinivas. – Būtent to tikimasi iš didžiųjų kalbos modelių.“

„Google“, kurios pagrindinė veikla buvo susijusi su kalbomis ir reklama, vadovams tai neatrodė logiška. „Microsoft“ daug daugiau rūpinosi kūrėjams skirtais įrankiais, nes tuo metu buvo programinės įrangos lyderė. „OpenAI“ išties pasisekė, nes savų modelių mokymas programuoti ne tik patenkino naujojo partnerio lūkesčius, bet ir darė kalbos modelius protingesnius.

Somasegaras paklausė Nadellos, ką jis mano apie Samą Altmaną. „Jis rūpinasi pasaulinių problemų sprendimu“, – atsakė Nadella. Somasegaras prisiminė, kad Altmanas su Nadella kalbėjosi daugybe įvairių temų, o tai Nadellai suteikė dar daugiau noro dirbti su Altmanu. Atrodė, kad kuo beprotiškesnės ir utopiškesnės buvo Altmano ambicijos, tuo labiau Nadella tikėjo, kad šis žmogus gali padėti „Microsoft“ sparčiau augti.

BDI kūrimo idėja, kadaise laikyta keista ir marginalia DI srities teorija, po truputį virto programinės įrangos gigantui komerciškai perspektyvia koncepcija. Tai *galėjo* padėti „Microsoft“ sukurti geresnę skaičiuoklės programą, kuri atvertų dar didesnes galimybes: įrankių rinkinį, kuris visą „Microsoft“ programinę įrangą paverstų daug pažangesne.

„GitHub Copilot“ padarė didžiulę įtaką Nadellai. „Galėjai matyti užbaigtą paslaugą, kuri pakeis pasaulį“, – teigė Somasegaras, ypač kai ši paslauga buvo pritaikyta kitų rūšių programinei įrangai. Supratęs tai, Nadella ir jo vyriausiasis technologijų direktorius Kevinas Scottas pradėjo vis labiau propaguoti DI „Microsoft“ viduje, beveik kiekvienoje produktų grupės peržiūroje ar priimant bet kokį sprendimą dėl produktų paminėdami šią technologiją. *Kodėl jūsų komanda nenaudoja A5? Viską darykite A5 formatu ir, jei tik įmanoma, naudokite „OpenA5“ modelius.*

Tai, žinoma, sukėlė nepasitenkinimą šimtams „Microsoft“ tyrimų skyriaus DI specialistų, kurie jau daugelį metų dirbo su DI modeliais. Anot spaudos pranešimų ir kelių DI tyrėjų, nugirdusių šią kritiką, Nadella pyko ant komandos vadovų už tai, kad jie nesugebėjo pasiekti „OpenAI“ standartų, nors pastaroji turėjo daug mažiau darbuotojų.

„Visa tai „OpenAI“ sukūrė su 250 žmonių, – Nadella sakė „Microsoft“ tyrimų skyriaus vadovui, kaip kad praneša „The Information“. – Kodėl mums apskritai reikalingas „Microsoft“ tyrimų skyrius?“

Jis taip pat nurodė savo tyrėjams nustoti kurti vadinamuosius pagrindinius modelius, arba dideles sistemas, pavyzdžiui, „OpenAI GPT“ modelius, teigia vienas vyresnysis DI srities mokslininkas. Kai kurie darbuotojai, nusivylę susidariusia situacija, išėjo iš darbo. Tačiau net jie pripažino, kad „Copilot“ buvo puikus įrankis, padėjęs programuotojams greičiau parašyti naują kodą ir dirbti su esamu kodu. Nadella norėjo žodį „*copilot*“ įtraukti į platesnį „Microsoft“ paslaugų spektrą, pasitelkdamas „OpenAI“ kalbos modelio technologiją, kad pagerintų žmonių elektroninių laiškų rašymo ir skaičiuoklių kūrimo kokybę.

Prabėgus keletui savaičių po Somasegaro susitikimo su Nadella 2022 metų pradžioje, „OpenAI“ pradėjo testuoti pažangesnes GPT-3 versijas ir pavadino jas žymių istorinių inovatorių vardais: „Ada“, „Babbage“, „Curie“ ir „DaVinci“. Laikui bėgant, šie modeliai gebėjo apdoroti dar sudėtingesnius klausimus ir pateikti labiau individualizuotus atsakymus. Apskritai, visuomenė dar nesuvokė, kokia pažangi tapo ši programinė

sistema. Tai pagaliau pradėjo keistis 2022 metų balandžio mėnesį, „OpenAI“ perkėlus kai kurias GPT-3 kalbos galimybes į vaizdų pasaulį ir pristačius savo pirmąjį didelį išradimą.

San Fransisko biuro kampelyje trys „OpenAI“ tyrėjai jau dvejus metus bandė panaudoti vadinamąjį difuzijos modelį, skirtą vaizdams generuoti. Difuzijos modelis iš esmės vaizdą kūrė nuo kito galo. Užuot ėmęsis tuščios drobės, kaip tai darytų menininkas, šis modelis pradėjo nuo drobės, jau išteptos daugybe spalvų ir su atsitiktinėmis detalėmis. Tuomet modelis prideda daug „triukšmo“, arba atsitiktinumo, prie duomenų, padarydamas juos neatpažįstamus. Tada žingsnis po žingsnio sumažina chaotiškų duomenų, kad laipsniškai išryškėtų vaizdo detalės ir struktūra. Po kiekvieno žingsnio vaizdas ryškėja ir tampa išsamesnis, tarsi dailininkui patobulinus nutapytą kūrinį. Šis difuzijos metodas kartu su vaizdų žymėjimo įrankiu, žinomu CLIP pavadinimu, tapo įdomaus naujo modelio, kurį mokslininkai pavadino DALL-E 2, pagrindu.

Pavadinimas buvo duoklė tiek 2008 metų animaciniam filmui WALL-E apie robotą, kuris pabėga iš Žemės, tiek dailininkui siurrealistui Salvadorui Dali. DALL-E sugeneruoti vaizdai kartais buvo siurrealūs, tačiau įrankį naudojantiesiems pirmą kartą jie atrodė neįtikėtini. Įvedę tekstinę užklausą, pavyzdžiui, „avokado formos kėdė“, gautumėte seriją būtent tokių nuotraukų, ir daugelis iš jų būtų itin realistiškos. Net sudėtingiausios užklausos buvo atvaizduojamos taip tiksliai, kad per kelias dienas nuo paleidimo DALL-E 2 sulaukė dėmesio „Twitter“ platformoje. Vartotojai stengėsi pranokti vienas kitą kurdami kuo keistesnius vaizdus, pavyzdžiui, „Tokiją puolančią sombrerą dėvinčią jūrų kiaulytę Godzilą“ arba „po Mordorą klajojančius girtus vyrus be marškinėlių“. Nors žmonių veidai dažnai atrodė keistai deformuoti, negalima paneigti, kad šie vaizdai buvo gerokai tikslesni nei bet kas, ką iki tol buvo sukūręs kompiuteris. Staiga visos naujienos sukosi aplink „OpenAI“, nes visuomenė pirmą kartą pamatė, ką gali ši sistema.

Nors „Google“ nusprendė tokias naujoves laikyti paslaptyje, Altmanas norėjo, kad kuo daugiau žmonių išbandytų naująjį „OpenAI“ kūrinį. Kaip Silicio slėnio startuolių išminčius, jis jau daug metų konsultavo verslininkus ir skatino juos išleisti savo produktus į rinką. Technologai kartais tai vadina „išsiuntimo“ strategija, arba „minimaliai gyvybingo produkto“ išleidimu. Idėja iš esmės ta pati: kuo greičiau pateikti programinę įrangą vartotojams, kad pavyktų sukurti grįžtamojo ryšio grandinę tarp savęs ir vartotojų, iš esmės pasitelkiant visuomenę kaip bandomąjį triušį. Šiuo įsitikinimu rėmėsi tokios milžinės kaip „Facebook“, „Uber“ ir „Stripe“. Altmanas buvo tvirtas tokios strategijos šalininkas. Geriausias būdas išbandyti produktą – tai jį paleisti.

Per kitus kelis mėnesius „OpenAI“ laipsniškai paleido DALL-E 2, pirmiausia prieigą suteikdama maždaug milijonui žmonių, įtrauktų į laukiančiųjų sąrašą. Taip siekta apsaugoti sistemą nuo neigiamos reakcijos, jei ji kurtų įžeidžiamus ar žalingus vaizdus. Po penkių mėnesių „OpenAI“ suprato, kad GPT-2 nekelia grėsmės pasauliui, ir atvėrė duris visiems, norintiems išbandyti DALL-E 2.

DALL-E 2 buvo mokoma naudojant milijonus įvairiausių vaizdų, surinktų iš viešos interneto erdvės. Kaip ir anksčiau, „OpenAI“ apie tai, kaip mokoma DALL-E, kalbėjo atsargiai. Sistema sėkmingai sukūrė Pikaso stiliaus vaizdus, tad galėjai suprasti, kad šio menininko kūriniai tikriausiai buvo įtraukti į sistemos mokymo procesą. Tačiau tai nustatyti tiksliai buvo sudėtinga.

Be to, nebuvo jokios galimybės sužinoti, ar kitų, mažiau žinomų menininkų darbai taip pat naudoti sistemai mokyti. „OpenAI“ neatskleidė mokymo duomenų detalių, baimindamasi, kad tai leistų piktavaliams nukopijuoti jos sukurtą modelį.

Vienas iš patyrusiųjų „OpenAI“ mokymo neigiamą pusę buvo Gregas Rutkowskis, lenkų skaitmeninio meno kūrėjas, garsus fantazijos stiliaus peizažais, kuriuose vaizduojami iltėti, ugnimi besispjaudantys drakonai ir burtininkai. Jo vardas tapo viena iš populiariausių užklausų

konkurentų sukurtoje atvirojo kodo DALL-E 2 versijoje, dar žinomoje kaip „Stable Diffusion“. Tai paskatino užduoti nerimą keliantį klausimą: kodėl reikėtų tokiam menininkui kaip Rutkowskis mokėti už naujus kūrinius, jei galima įsigyti programinę įrangą, kuriančią Rutkowskio stiliumi?

Žmonės pradėjo pastebėti kitą su DALL-E 2 susijusią problemą. Sistemos paprašę sukurti fotorealistinius įmonių vadovų atvaizdus, greitai pastebėtumėte, kad beveik visi tie vadovai būtų baltaodžiai vyrai. Pasitelkus žodį „slaugytoja“ sukuriami tik moterų vaizdiniai, o „advokatas“ atvaizduojamas tik kaip vyras.

2022 metų balandžio mėnesį per interviu Altmano buvo paklausta apie šią problemą. Kaip įprasta, jis įsitraukė į diskusiją ir pripažino, kad tai problema, bet „OpenAI“ ją sprendžia. Vienas iš būdų tai padaryti buvo užblokuoti DALL-E 2 galimybę generuoti smurtinio ar pornografinio pobūdžio vaizdus ir juos pašalinti iš mokymo duomenų.

Jis taip pat samdė rangovus iš besivystančių šalių, pavyzdžiui, Kenijos, kad modelis išmoktų pateikti tinkamesnius atsakymus. Tai buvo ypač svarbu, nes tai reiškė, jog net baigus mokyti modelį, pavyzdžiui, GPT-3 ar DALL-E 2, „OpenAI“ vis tiek galėjo toliau tobulinti sistemą padedant žmonėms vertintojams, kad jos pateikiami atsakymai būtų gilesni, aktualesni ir etiškesni. Vertindami DALL-E 2 pateiktus atsakymus pagal gerumo skalę, žmonės galėjo sistemą nukreipti link tinkamesnių atsakymų pateikimo.

Vis dėlto tokie vertintojai ne visada nuosekliai vertino sistemą, o netinkamų vaizdų pašalinimas iš DALL-E 2 mokymo duomenų taip pat galėjo būti vertinamas kaip adatos paieška šieno kupetoje. Iš pradžių „OpenAI“ tyrėjai bandė pašalinti visas pernelyg seksualinio pobūdžio moterų nuotraukas, kurias galėjo rasti mokymo rinkinyje, kad DALL-E 2 neatvaizduotų moterų kaip seksualinio pobūdžio objektų. Tačiau tai turėjo kainą: „gana smarkiai“ sumažino moterų skaičių duomenų rinkinyje, kaip kad teigė tuometinė „OpenAI“ tyrimų ir produktų vadovė

Mira Murati. Ji nepatikslino, kaip smarkiai. „Turėjome taisyti, nes nenorėjome lobotomizuoti sukurto modelio. Tai labai sudėtingas procesas.“

DALL-E 2 sukurti realistiški žmonių veidai buvo modelio silpnoji vieta dėl pasitelkiamų stereotipų. Rodėsi, kad „OpenAI“ puikiai supranta šią problemą. Tai buvo tokia didelė problema, kad 400 žmonių, daugiausia „OpenAI“ ir „Microsoft“ darbuotojų, pradėjus testuoti sistemą, „OpenAI“ uždraudė jiems viešai dalytis bet kokiais DALL-E 2 sukurtais realistiškais žmonių portretais.

Kai kurie „OpenAI“ darbuotojai nerimavo, kad įrankis, galintis generuoti suklastotas nuotraukas, išleistas taip greitai. Pradėjusi veiklą kaip ne pelno organizacija, kurianti saugų DI, „OpenAI“ tapo viena iš agresyviausių DI technologijų įmonių. Vienas nepanoręs prisistatyti bendrovės komandos narys, dirbęs saugos bandymų srityje, žurnalui „Wired“ teigė, kad atrodo, jog bendrovė šią technologiją nori pateikti pasauliui, nors „šiuo metu yra labai daug galimybių padaryti žalą“.

Tačiau Altmanas siekė didesnio tikslo. Jis manė, kad naujoji sistema peržengė svarbų slenkstį kelyje į BDI sukūrimą. „Rodos, ji tikrai supranta atitinkamas koncepcijas, – teigė jis viename interviu. – O tai primena intelektą.“ Pasak jo, DALL-E 2 buvo toks stebuklingas, kad net BDI skeptikai pradėjo rimtai svarstyti šią idėją.

Stebuklingos buvo ne tik DALL-E 2 galimybės. Ypatingą reikšmę turėjo įrankio poveikis žmonėms. „Vaizdai turi emocinę galią“, – sakė jis. DALL-E 2 sukėlė didelį susidomėjimą rinkoje. Skirtingai nei „GitHub Copilot“, kuris galėjo užbaigti kažkieno kito pradėtą rašyti kodą, šis modelis kūrė visiškai suformuotą turinį nuo pradžios iki pabaigos. Darė tai lyg grafikos menininkas, paprašytas nupiešti norimą paveikslėlį.

Kitas Altmano žingsnis dėl šios visiškai suformuoto turinio generavimo idėjos buvo vertinamas kaip dar sensacingesnis. GPT-1 sistema buvo panašesnė į automatinio teksto užbaigimo įrankį, kuris tęsė tai, ką pradėjo rašyti žmogus. Tačiau GPT-3 ir jo naujinys GPT-3.5

gebėjo kurti originalius tekstus, kaip ir DALL-E 2, kuris vaizdus kūrė nuo pat pradžių.

Kol įmonių žavėjosi DALL-E 2, pasklido gandai, kad konkurentė „Anthropic“ kuria pokalbių robotą. Tai „OpenAI“ sukėlė norą dar labiau konkuruoti. 2022 metų lapkričio mėn. pradžioje „OpenAI“ vadovai pranešė darbuotojams, kad per kelias savaites ketina paleisti pokalbių robotą, sukurtą remiantis GPT-3.5. Pasak „OpenAI“ artimo asmens, su pokalbių roboto projektu dirbo maždaug dvylika žmonių. Tai nelabai skyrėsi nuo „Google Meen“ roboto, kurį Noamas Shazeeras kūrė prieš dvejus metus ir informaciją apie kurį „Google“ laikė visiškoje paslaptyje.

„OpenAI“ vadovybė užtikrino darbuotojus, kad tai bus ne produkto pristatymas, o „dėmesio neatkreipiantis tyrimų pristatymas“. Vis dėlto kai kurie darbuotojai teigė, kad jautėsi nepatogiai, taip greitai išleisdami šį įrankį. Jie nežinojo, kaip visuomenė galėtų piktnaudžiauti tokiu pažangiu ir galingu kalbos modeliu.

Be to, pokalbių robotas dažnai darė faktinių klaidų. Jį kuriantys mokslininkai nusprendė nedidinti sistemos atsargumo lygio, nes tai padarius įrenginys atsisakydavo atsakyti į klausimus, į kuriuos galėjo atsakyti teisingai. Vengta, kad sistema pateiktų atsakymą „Nežinau“. Kūrėjai sistemą kalibravo taip, kad ji kalbėtų autoritetingiau, nors tai reiškė, kad pokalbių robotas bent jau kartais kaip atsakymą pateiks neteisingą informaciją. Sistema buvo pavadinta „ChatGPT“.

Altmanas norėjo sistemą paleisti kuo greičiau. Jis teigė, kad šimtai „OpenAI“ darbuotojų jau išbandė ir patikrino „ChatGPT“ ir kad svarbu pripratinti žmoniją prie to, ką DI gali atlikti, pamažu, tarsi įmerkiant kojų pirštus į šaltą baseiną. Tam tikra prasme „OpenAI“ padarė pasauliui paslaugą ir parengė jį galingesniam būsimam „OpenAI“ modeliui GPT-4. Pasak tuometinio „OpenAI“ vadovo, per vidinį testavimą GPT-4 sugebėjo parašyti neblogo lygio poeziją, o jo juokeliai buvo tokie geri, kad iš jų juokėsi net „OpenAI“ vadovai. Tačiau jie nežinojo, kokį poveikį tai turės pasauliui ar visuomenei, ir vienintelis būdas sužinoti buvo produktą

išleisti į rinką. Savo interneto svetainėje „OpenAI“ tai pavadino „iteracinio įgyvendinimo“ filosofija, t. y. produktų išleidimu į rinką, siekiant geriau ištirti jų saugumą ir poveikį. Pasak bendrovės, tai buvo geriausias būdas užtikrinti, kad BDI būtų kuriamas žmonijos labui.

2022 metų lapkričio 30 dieną „OpenAI“ paskelbė tinklaraščio įrašą, kuriame pranešė apie viešą „ChatGPT“ demonstravimą. Daugelis „OpenAI“ darbuotojų, įskaitant kai kuriuos dirbančius saugumo srityje, nežinojo apie šios funkcijos paleidimą. Imta lažintis, kiek žmonių ja naudosis po savaitės. Didžiausias įvertis buvo šimtas tūkstančių vartotojų. Pats įrankis buvo tik interneto svetainė su teksto laukeliu. Į laukelį pakakdavo įvesti bet kokią norimą informaciją ir robotas atsakydavo. Jo veikimas buvo paremtas GPT-3.5. Didžioji visuomenės dalis nebuvo girdėjusi apie „OpenAI“, ką jau kalbėti apie GPT-3. Ir niekas, įskaitant „OpenAI“ tyrėjus, nežinojo, kas atsitiks, leidus bet kam išbandyti tas galimybes.

„Šiandien paleidome „ChatGPT“, – apie 11:30 val. San Fransisko laiku „Twitter“ paskyroje parašė Altmanas. – Pabandykite su juo pasikalbėti čia: http://chat.openai.com.“

Iš pradžių spengė tyla, tačiau nišinės programinės įrangos kūrėjai ir mokslininkai užsuko į svetainę ir pradėjo ją bandyti. Per kelias valandas jų atsiliepimai ėmė plūsti „Twitter“:

12.26 PT @MarkovMagnifico: [šiuo metu] žaidžiu su „ChatGPT“ ir dabar savo BDI sukūrimo datą perkėliau į šiandieną.

12:37 PT @AndrewHartAR: „ChatGPT“ ką tik buvo išleistas. Aš pamačiau ateitį.

13:37 PT @skirano: Tai visiškai beprotiška. Paprašiau #chatGPT sukurti paprastą asmeninį tinklalapį. Jis parodė, kaip tai padaryti žingsnis po žingsnio… kaip tai sukurti, tada pridėjo HTML ir CSS.

14:09 PT @justindross: „ChatGPT“ man yra geresnis pradinis taškas nei „Google“, kai turiu klausimų. Tai tikrai beprotiška.

14:29 PT @Afinetheorem: nebegalima duoti namų darbų / užduočių.

Buvo sunku rasti bent vieną neigiamą „ChatGPT“ vertinimą. Dauguma žmonių reagavo su didžiuliu susižavėjimu. Tai, kas padarė šį modelį dar išskirtinesnį, buvo ne tik jo veikimo sklandumas, bet ir žinių platumas. Daugelis žmonių tuo metu buvo išbandę pokalbių robotą, nesvarbu, ar tai buvo „Alexa“, ar kokios nors rūšies klientų aptarnavimo robotas. Daugelis buvo pripratę prie ribotų, stringančių pokalbių. Tačiau „ChatGPT“ galėjo išsamiai atsakyti į beveik visus klausimus. Tai buvo tarsi perėjimas nuo pokalbio su mažamečiu vaiku prie pokalbio su aukštąjį išsilavinimą turinčiu suaugusiu žmogumi.

Per kitas 24 valandas vis daugiau žmonių prisijungė prie „ChatGPT“, taip apkraudami sistemos serverius ir išbandydami jos ribas. Tuo metu robotą išbandė profesionalai, technologijų darbuotojai, rinkodaros ir žiniasklaidos atstovai. Norėdami surengti šou „Twitter“ tinkle, jie savo eksperimentus pavertė viešu konkursu, kuriuo siekta išsiaiškinti, kas sugebės priversti „ChatGPT“ parašyti juokingiausią, protingiausią arba keisčiausią tekstą. Tai buvo tarsi dar kartą pasikartojusi DALL-E 2 fanfarų banga, tik didesnė. Per keletą kitų dienų žmonės „Twitter“ platformą užpildė ekranų kopijomis iš „ChatGPT“ eilėraščių, repo, situacijų komedijų scenų ir elektroninių laiškų. Kuo neįprastesni pateikti atsakymai, tuo geriau.

Vienas „Twitter“ vartotojas, vardu Thomas H. Ptacekas, paprašė „parašyti biblinio stiliaus eilėraštį Karaliaus Jokūbo Biblijos stiliumi ir paaiškinti, kaip išimti sumuštinį su žemės riešutų sviestu iš vaizdo grotuvo“.

„OpenAI robotas“ laikėsi šių reikalavimų:

Ir tapo žinoma, kad vyrui nerimą kėlė sumuštinis su žemės riešutų sviestu, nes jis buvo įdėtas į vyro vaizdo grotuvą ir jis nežinojo, kaip reikia jį pašalinti.

Jis sušuko Viešpačiui: „O Viešpatie, kaip man išimti šį sumuštinį iš savo vaizdo grotuvo, nes jis ten įstrigo ir nebejuda?“

„Atsiprašau, bet negaliu būti ciniškas technologijos, kuri gali tai pasiekti, atžvilgiu“, – parašė Ptacekas „Twitter“ paskyroje. Per savaitę „ChatGPT“ naudojo jau daugiau nei milijonas žmonių. Po dviejų mėnesių „ChatGPT“ pritraukė trisdešimt milijonų registruotų vartotojų ir tapo viena iš sparčiausiai augančių internetinių paslaugų istorijoje. Iki 2024 metų pradžios „ChatGPT“ kas savaitę naudojo apie šimtas milijonų žmonių. Joks kitas atskiras DI įrankis anksčiau nebuvo pasiekęs tokio populiarumo.

2023 metų kovo 14-ąją, tą pačią dieną, kai „Anthropic“ pagaliau išleido pokalbių robotą „Claude“, „OpenAI“ pristatė jo naujinį GPT-4. Kiekvienas, sumokėsiantis 20 dolerių per mėnesį, galės naudotis šia nauja technologija per „ChatGPT Plus“ – prenumeratos paslaugą, kuri 2023 metais turėtų atnešti apie 200 milijonų dolerių pajamų. Kai kurie bendrovės darbuotojai manė, kad GPT-4 yra didelis žingsnis link BDI.

Mašinos ne tik mokėsi statistinės koreliacijos tekste, kaip kad viename interviu sakė Sutskeveris. „Šis tekstas iš tiesų yra pasaulio projekcija. Neuroninis tinklas vis daugiau mokosi apie pasaulį, žmones, žmogiškumą, žmonių viltis, svajones ir motyvaciją, jų tarpusavio sąveiką ir situacijas, kuriose atsiduriame.“

„Kai turite sistemą, kuri gali priimti pastebėjimus apie pasaulį, išmokti juos suprasti – vienas iš būdų tai padaryti yra numatyti, kas įvyks toliau, – manau, yra labai artima intelekto sampratai“, – interviu sakė Altmanas.

Technologijų žiniasklaida buvo sužavėta. „New York Times“ pavadino „ChatGPT“ „geriausiu DI pokalbių robotu, kada nors išleistu plačiajai visuomenei“. Žurnalistai, išbandę šią sistemą, gėrėjosi jos draugiškais ir entuziastingais atsakymais. „Twitter“ tinkle kai kurie technologijų entuziastai gyrėsi jau naudojantys šią funkciją elektroninių laiškų ar kitų su darbu susijusių dokumentų projektams rengti, kad galėtų dirbti produktyviau.

Natūralu, kad kilo nauja straipsnių banga apie tai, ar „ChatGPT“ pakeis žmones. Altmanas ėmėsi viešųjų ryšių kampanijos, kad atkreiptų dėmesį į šį susidomėjimą ir tiesiogiai atsakytų į klausimus per tinklalaides, laikraščius ir kitą žiniasklaidą. Jis teigė, kad tai tikriausiai sumažins žmonių darbo vietų, pavyzdžiui, tekstų kūrėjų, klientų aptarnavimo operatorių ir net programinės įrangos kūrėjų. Vis dėlto tai nereiškia, kad „ChatGPT“ ir ją palaikanti technologija visiškai pakeis žmogaus darbą.

„Kai kurios darbo vietos išnyks, – viename interviu atvirai sakė Altmanas. – Atsiras naujų, geresnių darbo vietų, kurias šiandien sunku įsivaizduoti.“ Spauda ir visuomenė į tai reagavo tyliai ir nuolankiai, nes istorinės permainos, tokios kaip pramonės revoliucija, parodė, kad technologijos iš tiesų gali sukelti skausmingų užimtumo pokyčių. Generatyvinės DI sistemos, tokios kaip „ChatGPT“, nebuvo trumpalaikio populiarumo susilaukę įrankiai, kaip kad nutiko su kriptovaliutomis. „ChatGPT“ buvo naudingas įrankis. Žmonės, pasitelkę „ChatGPT“, jau rašė mokyklinius rašinius, kūrė verslo planus ir atliko rinkodaros tyrimus.

„OpenAI“ darbuotojai guodėsi, kad ateitis bus to verta. Tikino, kad per pramonės revoliuciją perėjimas prie mašinų valdomų darbų ir gamyklų taip pat sukūrė naujų darbo vietų ir pagerino gyvenimo lygį. Drauge didėjo atotrūkis tarp „OpenAI“ darbuotojų, kurie dėmesį buvo sutelkę į produktų kūrimą, ir pastangas nukreipusiųjų į saugumą, besistengiančių stebėti sparčiai augantį „ChatGPT“ srautą, kad aptiktų netinkamo pobūdžio užklausas.

Tikėdamas, kad jie žengia svarbius žingsnius link BDI, Ilya Sutskeveris pradėjo glaudžiau bendradarbiauti su bendrovės saugos komanda. Vis dėlto „OpenAI“ produktų komanda padvigubino pastangas komercializuoti „ChatGPT“, prašydama kitų įmonių mokėti už prieigą prie sistemos pagrindinės technologijos.

„Google“ vadovai suprato, kad vis daugiau žmonių gali pradėti kreiptis į „ChatGPT“, o ne į „Google“, ieškodami informacijos apie sveikatą ar patarimų dėl produktų, – tai vieni iš pelningiausių paieškos terminų, pagal kuriuos parduodami skelbimai.

„Google“ neabejotinai nusipelnė tinkamo konkurento. Per metus bendrovės paieškos rezultatų puslapis tapo pilnas reklamų ir rėmėjų nuorodų. Ji stengėsi iš kiekvienos paieškos išspausti kuo daugiau pajamų, net jei dėl to produktas tapo nepatogesnis naudoti. Suklaidinus žmones dėl to, kas yra reklama, o kas – tikrasis paieškos rezultatas, galima uždirbti daug pinigų.

Nuo 2000 iki 2005 metų „Google“ reklamas žymėjo aiškiau, suteikdavo joms mėlyną foną ir užtikrindavo, kad paieškos puslapio viršuje būtų pateiktos tik viena ar dvi tokio pobūdžio reklamos nuorodos. Tačiau bėgant metams tapo vis sunkiau atskirti reklamą nuo įprastų interneto nuorodų. Mėlyną foną pakeitė žalias, tada jis išbluko iki geltonos spalvos, galiausiai visai išnyko. Reklamos pradėjo užimti didesnę puslapio dalį, todėl vartotojams teko ilgiau slinkti, ieškant tinkamų rezultatų. Nors tai pradėjo vartotojus erzinti, „Google“ galėjo sau tai leisti: tuo metu interneto naudotojams atrodė, kad jie paprasčiausiai neturi kitos išeities. Daugiau nei 90 procentų internetinių paieškų visame pasaulyje buvo atliktos pasitelkus „Google“ paieškos sistemą.

Tačiau dabar, pirmą kartą per daugiau nei dvidešimt metų, „Google“, kaip interneto vartų sargo, dominavimas buvo pakirstas. Daugelį metų pagrindinis „Google“ pinigų šaltinis buvo sistema, kuri nuskaitydavo milijardus interneto puslapių, juos indeksuodavo ir reitinguodavo, ieškodama tinkamiausių atsakymų į pateiktas užklausas. Rezultatas – sąrašas nuorodų, kurias galima paspausti. „ChatGPT“ užsiėmusiems interneto vartotojams pasiūlė kai ką patrauklesnio: vieną bendrą atsakymą, pagrįstą savo pačios atlikta visos pateiktos informacijos sinteze. Nebereikėjo begalę kartų ekranu slinkti žemyn ar ieškoti

informacijos tarp gausybės reklamų ir nuorodų. „ChatGPT“ viską padaro už jus.

Pavyzdžiui, paimkime klausimą, ar moliūgų pyragui labiau tinka sutirštintas, ar išgarintas pienas. Paklausę „ChatGPT“, gautumėte vieną išsamų atsakymą apie tai, kad sutirštintas pienas tikriausiai yra geresnis pasirinkimas, nes dėl jo pyragas bus saldesnis. „Google“ pateiktų ilgą sąrašą nuorodų į įvairiausius skelbimus, receptus ir straipsnius, kuriuos turėtumėte peržiūrėti ir perskaityti. Begalinės galimybės, kurios kadaise „Google“ pavertė tokia išskirtine sistema, dabar tapo laiko švaistymu. Silicio slėnio technologai nuolat bandė sukurti „sklandžią“ internetinio naršymo patirtį. Trinties nesukelianti alternatyva „Google“ grėsė potencialia finansine katastrofa.

Per kelias savaites po „ChatGPT“ paleidimo „Google“ vadovai bendrovėje paskelbė ypatingą pavojų. Ji buvo užklupta netikėtai ir dėl to skaudžiai nukentėjo. Nuo 2016 metų generalinis direktorius Sundaras Pichai’us „Google“ vadino „į DI orientuota bendrove“. Taigi, kaip maža įmonė su mažiau nei dviem šimtais DI tyrėjų sugebėjo sukurti kai ką geresnio nei „Google“, kurioje dirba beveik penki tūkstančiai darbuotojų? Grėsmė tapo dar rimtesnė dėl „OpenAI“ glaudžių ryšių su didžiulius finansinius išteklius valdančia „Microsoft“.

„Google“ jau turėjo „LaMDA“ – senesnį kalbos modelį, kurį jos inžinierius laikė sąmoningu. Tačiau bendrovės vadovai buvo įsprausti į keblią padėtį. Kas, jei jie išleistų „ChatGPT“ konkurentą ir žmonės pradėtų naudoti jį vietoje „Google“ paieškos? Tai reikštų, kad jie nebespaustų ant reklamų, rėmėjų nuorodų ir kitų svetainių, naudojančių „Google“ reklamos tinklą ir didinančių bendrovės pelną.

Daugiau nei 80 procentų iš 258 milijardų JAV dolerių „Alphabet“ pajamų 2021 metais buvo gauta iš reklamos, didžioji dalis – iš mokėjimų už spustelėjimus ant reklamų, kurias žmonės pamatydavo naudodamiesi šia paieškos sistema. Tokio pobūdžio reklamos, kurios užteršdavo „Google“ paieškos rezultatus, tapo labai svarbios jos verslo modeliui.

Bendrovė negalėjo tiesiog pakeisti esamos padėties. „Google“ paieškos tikslas yra priversti jus spustelėti nuorodas, idealiu atveju – reklamą, – paaiškino Sridharas Ramaswamy’is, kuris nuo 2013 iki 2018 metų vadovavo „Google“ reklamos ir komercinės veiklos padaliniui. – Visas kitas puslapyje pateiktas tekstas yra tik užpildas.“

„Google“ daugelį metų laikėsi atsargaus, beveik baimingo požiūrio į naujas technologijas. Bendrovė rimtesnių veiksmų ėmėsi tik tuo atveju, jei tai buvo susiję su milijardiniu pelnu, ir tikrai nenorėjo keisti savo pačios reklamos verslo, per metus atnešančio beveik 260 milijardų dolerių.

„Kuo didesnis tampi, tuo viskas sudėtingiau, – teigia Ramaswamy’is. – „Google“ reklamos komanda įprastai būdavo keturis ar penkis kartus didesnė už algoritminių paieškos rezultatų komandą. Pradėti kurti produktą, kuris yra pagrindinio modelio priešingybė, tikrai sunki užduotis.“ Dabar „Google“ vadovams neliko kito pasirinkimo. Susitikime, kuris buvo įrašytas ir perduotas „New York Times“, vienas vadovas atkreipė dėmesį, kad panašu, jog mažesnės įmonės, pavyzdžiui, „OpenAI“, mažiau nerimauja dėl radikalių naujų DI įrankių pateikimo visuomenei. „Google“ turėjo imtis veiksmų ir padaryti tą patį, kitaip rizikavo negrįžtamai atsilikti. Atsargumą patraukus į šalį, visi procesai pradėjo veikti itin sparčiai.

Supanikavę vadovai nurodė darbuotojams į pagrindinius bendrovės produktus, tokius kaip „YouTube“ ir „Gmail“, turinčius daugiau nei milijardą naudotojų, vos per kelis mėnesius įdiegti tam tikros formos generatyvinį DI. „Google“ jau daugelį metų buvo pagrindinė pasaulinė indeksavimo mašina, apdorojanti vaizdo įrašus, vaizdus ir duomenis. Dabar ji, pasitelkdama DI, turėjo pradėti *kurti* ir naujus duomenis. Tokio esminio pokyčio įgyvendinimas buvo tarsi bandymas lenktynių trasoje važiuoti senu, tik 30 kilometrų per valandą greičiu judančiu sunkvežimiu. Vadovai buvo taip nusiminę, kad pakvietė „Google“ įkūrėjus Larry’į Page’ą ir Sergey’ų Briną, 2019 metais atsistatydinusius iš

„Alphabet“ bendrų generalinių direktorių pareigų, kad per kelis susitikimus padėtų sugalvoti atsaką į „ChatGPT“.

Matydamos, kad „Google“ vadovybė jaučiasi nesaugiai, bendrovės inžinierių komandos įvykdė užduotį. Praėjus keletui mėnesių po „ChatGPT“ paleidimo „YouTube“ vadovai pridėjo funkciją, leidžiančią svetainės vaizdo kūrėjams generuoti naujas filmų parinktis arba keisti aprangą, pasitelkus generatyvinį DI. Tačiau visa tai atrodė bergždžia. Atėjo laikas panaudoti slaptąjį ginklą: „LaMDA“.

Pichai’us visai bendrovei išsiuntė pranešimą, kuriame nurodė darbuotojams išbandyti naują pokalbių robotą, kurį netrukus ketino pristatyti visuomenei, ir perrašyti visus atsakymus, kurie, jų nuomone, buvo netinkami. 2023 metų vasario 6 dieną jis paskelbė tinklaraščio įrašą: pranešė pasauliui, kad ruošiasi pristatyti kai ką naujo. Straipsnyje „Svarbus kitas žingsnis mūsų DI kelionėje“ jis rašė: „Dirbame su eksperimentine pokalbių DI paslauga, kurią palaiko „LaMDA“ ir kurią vadiname „Bard“.“

Norėdama išlikti lydere, „Microsoft“ kitą dieną taip pat paskelbė pranešimą. „Bing“, „Microsoft“ nuo „Google“ gerokai atsiliekanti paieškos sistema, užimanti vos 6 procentus internetinių užklausų rinkos, netrukus sulauks didelio DI atnaujinimo. Naujausias „OpenAI“ GPT kalbos modelis padėtų „Bing“ „atrakinti atradimo džiaugsmą, pajusti kūrybos stebuklą ir geriau išnaudoti pasaulio žinias“. Vertimas: ji galėjo daryti tai, ką jau darė „ChatGPT“, bet su tam tikrais patobulinimais, apie kuriuos žinojo tik „Microsoft“.

Šis įtemptas lenktynių startas stebino visą pasaulį, kol keli atidūs vartotojai pamatė keletą problemų. „Google“ paskelbė įdomių atsakymų iš „Bard“ pavyzdžius, o „Microsoft“ tą patį padarė paskelbdama atsakymus iš „Bing“. Tačiau žurnalistams patikrinus kai kuriuos iš pateiktų atsakymų paaiškėjo, kad jie neteisingi. Pichai’aus parodytame pristatymo vaizdo įraše „Bard“ suklydo dėl istorinio fakto apie Jameso

Webbo teleskopą, o „Bing“ neteisingai nurodė kai kuriuos mažmeninės prekybos tinklo „The Gap“ pelno skaičius.

Pokalbių robotai ne tik fantazavo apie faktus – pasireiškė ir jų nuotaikų sutrikimai. Netrukus po „Microsoft“ pranešimo „New York Times“ žurnalistas Kevinas Roose’as paskelbė straipsnį apie nerimą keliantį dviejų valandų pokalbį, į kurį jis vieną naktį leidosi su „Bing“. Per pokalbį „Microsoft“ naujoji paieškos sistema, virtusi pokalbių robotu, prisipažino įsimylėjusi žurnalistą ir tvirtino, kad jis „nėra laimingai vedęs“. Roose’as rašė, kad šis pokalbis jam sukėlė „blogą nuojautą, kad DI peržengė ribą ir kad pasaulis niekada nebebus toks kaip anksčiau“.

„Microsoft“ vadovui šis triukšmas ir dėmesys „Bing“ tapo puikia proga pasipuikuoti. Viename interviu jis pasakė, kad daugelį metų laukė progos pakovoti su „Google“ dėl dominavimo paieškos rinkoje ir kad dabar „Bing“ pagaliau gali tai padaryti. „Noriu, kad žmonės žinotų, jog mes juos privertėme šokti“, – pridūrė jis.

Nuo pat pradžių viskas buvo nelogiška. „Google“ veiksmų ėmėsi anksti. Jos mokslininkai išrado transformatorių technologiją ir sukūrė sudėtingą kalbos modelį „LaMDA“ keleri metai iki GPT-4 atsiradimo. Jos pačios DI laboratorija „DeepMind“, siekdama to paties tikslo, ėmėsi užduoties sukurti BDI penkeriais metais anksčiau, nei buvo įkurta „OpenAI“. Tačiau „Google“ dabar turėjo vytis kitas bendroves. Jos biurokratinis aparatas ir baimė pakenkti verslui bei reputacijai sukėlė giluminę inerciją. Paradoksalu, bet tai apsaugojo pasaulį nuo kai kurių rizikų, kurias sukėlė „OpenAI“, – rizikų, galinčių labiausiai paveikti mažumas ir sunaikinti daugybę darbo vietų.

Didelis „OpenAI“ šuolis į šią rinką sukėlė abejonių dėl „DeepMind“ darbo pastaruosius trylika metų. Hassabisą tai sukrėtė. Pasak buvusio darbuotojo, praėjus kelioms savaitėms po „ChatGPT“ išleidimo, jis darbuotojų susirinkime pasakė, kad „DeepMind“ neturėtų tapti DI „Bell Labs“ – vieta, kurioje viskas buvo išrasta, bet kurios idėjas komercializavo kiti.

Niekas neklausė, kur yra BDI. Bet *klausė*, kur yra naudingas, žmogaus gebėjimams prilygstantis DI. „DeepMind“ pavyko sukurti DI sistemas, pajėgiančias įveikti žmones čempionus žaidime go ir kituose žaidimuose. Vis dėlto „OpenAI“ gebėjimas sukurti sistemą, galinčią parašyti elektroninį laišką, kažkodėl buvo dar įspūdingesnis. Mokslinė strategija, kurios siekė Demisas Hassabisas, pasirodė esanti pernelyg siaura. Hassabisas norėjo sukurti BDI per žaidimus ir simuliacijas, o savo bendrovės darbo sėkmę vertino pagal apdovanojimus ir prestižinius straipsnius mokslo žurnaluose. „OpenAI“ požiūris į DI buvo grindžiamas inžinerijos principais ir maksimalaus esamos technologijos išplėtimo siekiu. „DeepMind“ veikla buvo labiau akademinio pobūdžio – ji publikavo mokslinius straipsnius apie žaidimų sistemą „AlphaGo“ ir „AlphaFold“, naujovišką metodą, leidžiantį prognozuoti baltymų susidarymą žmogaus organizme.

„AlphaFold“ gimė 2016 metais „DeepMind“ surengtame hakatone – programuotojų maratone – ir vėliau tapo vienu iš perspektyviausių bendrovės projektų. Hassabisas svajojo panaudoti BDI didelėms pasaulinio masto problemoms spręsti, pavyzdžiui, ieškant vėžio gydymo būdų. Atrodė, kad jis pagaliau turėjo tai padaryti galinčią DI sistemą.

Kai mūsų ląstelėse esančios aminorūgštys susiformuoja į tam tikras trimatės formos struktūras, jos tampa baltymais, o netinkamai susiformavusios baltymų struktūros gali sukelti ligas. „AlphaFold“ buvo DI programa, kuri prognozavo, kaip susiformavusios atrodys šios 3D formos. „DeepMind“ manė, kad tai padės mokslininkams geriau suprasti, kokios cheminės reakcijos gali paveikti šiuos baltymus, ir kurti vaistus.

Hassabisas nusprendė, kad „DeepMind“ privalo laimėti tarptautinį baltymų struktūrų prognozavimo konkursą CASP 2019 ir 2020 metais. „Turime padvigubinti pastangas ir kuo greičiau judėti į priekį“, – sakė jis savo darbuotojams viename susitikime, kuris buvo užfiksuotas dokumentiniame filme. „Negalime gaišti laiko.“

Altmanas sėkmę vertino skaičiais. Jam buvo nesvarbu, ar tai būtų skaičiai, rodantys investicijas, ar produktų naudotojus. Hassabisas norėjo sulaukti pripažinimo ir apdovanojimų. Anot jo kolegų, jis dažnai darbuotojams sakydavo trokštantis, kad „DeepMind“ per ateinantį dešimtmetį laimėtų nuo trijų iki penkių Nobelio premijų.

„DeepMind“ laimėjo CASP konkursą 2019 ir 2020 metais, o 2021 metais mokslininkams suteikė prieigą prie savo baltymų struktūrų prognozavimo kodo. Pasak „DeepMind“, šiuo metu daugiau nei milijonas tyrėjų visame pasaulyje turi prieigą prie „AlphaFold“ baltymų struktūrų duomenų bazės. Tačiau mokslo atradimai vyksta lėtai. Nors Hassabisas vieną dieną gali laimėti Nobelio premiją, pasitelkus jo sistemą vis dar nėra įvykdytas joks didelis atradimas. Kai kurie ekspertai abejoja, ar „DeepMind“ baltymų struktūrų prognozės yra pakankamai tikslios, kad leistų patikimai nustatyti, kaip vaistų junginiai jungiasi prie baltymų. Neaišku ir tai, ar šios prognozės sutaupytų laiko ieškant naujų vaistų.

Apibendrinant galima pasakyti, kad didžiausi „DeepMind“ projektai sulaukė daug dėmesio, tačiau realiame pasaulyje turėjo palyginti mažą poveikį. „DeepMind“ atkakliai siekė DI mokyti visiškai sumodeliuotoje aplinkoje, kurioje fizikos dėsniai ir kiti aspektai galėtų būti tiksliai suprojektuoti ir stebimi. Pasitelkus tokią strategiją buvo sukurta programa „AlphaGo“, užprogramuota žaisti milijonus žaidimų su savimi atitinkamai sumodeliuotoje aplinkoje, ir programa „Alpha-Fold“, kuri naudojo baltymų struktūrų numatymo modelius.

„OpenAI“ pasirinktas modelis – sistemų mokymas pasitelkiant realaus pasaulio duomenis, t. y. milijardus žodžių, surinktų iš interneto, – buvo chaotiškas ir kupinas trikdžių. Tai kėlė riziką bendrovės reputacijai, ir tuo asmeniškai įsitikino Hassabisas, kilus ligoninių projekto skandalui. Kita vertus, „DeepMind“ taikyta izoliuota sistemų kūrimo strategija taip pat reiškė, kad sukurti DI sistemas, kurias žmonės galėtų naudoti realiame pasaulyje, tapo dar sunkiau.

Hassabisas buvo taip susikoncentravęs į DI sistemų virtualius pasaulius ir pripažinimo siekį, kad praleido kalbos modelių revoliuciją. Jam teko sekti Altmano pėdomis. „Google“ vadovai bendrovei „DeepMind“ nurodė sukurti kalbos modelius, kurie būtų geresni nei „LaMDA“. Naująją sistemą jie pavadino „Gemini“, o „DeepMind“ ją aprūpino strateginio planavimo technikomis, sukurtomis „AlphaGo“.

Kad viskas vyktų greičiau, Pichai’us žengė dar vieną drastišką žingsnį. Jis sujungė du konkuruojančius DI padalinius „DeepMind“ bei „Google Brain“ ir pavadino juos „Google DeepMind“ (darbuotojai naująjį padalinį vadino „GDM“). Po daugelio metų tarpusavio varžymosi, siekiant prisivilioti geriausius tyrėjus, ir kovos dėl didesnės skaičiavimo galios šie du padaliniai pasižymėjo visiškai skirtinga darbo kultūra. „Google Brain“ buvo artimesnis pagrindinei bendrovei ir tobulino „Google“ produktus, o „DeepMind“ turėjo tiek daug laisvės, kad tai net atrodė kiek pasipūtėliškai – bendrovės darbuotojai turėjo leidimus patekti į kitus „Google“ pastatus, tačiau „Google“ darbuotojai negalėjo įeiti į „DeepMind“ patalpas.

Daugelio nuostabai, naujai suformuoto padalinio vadovu Pichai’us paskyrė Hassabisą. Jeffas Deanas, labiausiai gerbiamas „Google“ inžinierius, kuris prižiūrėjo DI tyrimus likusioje bendrovės dalyje, atrodė geresnis kandidatas į šias pareigas. Tačiau vadovu buvo paskirtas buvęs žaidimų dizaineris ir simuliacijų fanatikas, t. y. vyras, kuris daug metų bandė atsiskirti, o dabar tapo didžiulio „Google“ projekto, kuriuo siekta apsaugoti bendrovės lyderio pozicijas interneto paieškos srityje, vadovu. Žvelgiant iš politinės perspektyvos, jis turėjo daugiau galios nei bet kada anksčiau, o kontroliuodamas didesnę „Google“ dalį, įgijo ir daugiau „DeepMind“ kontrolės.

„Demiso autoritetas ir įtaka „Google“ dabar daug didesni nei prieš kelerius metus, – teigė Shane’as Leggas. – Užuot tapę labiau nepriklausomi, tapome neatskiriama „Google“ dalimi. Mums ir mūsų vykdomai misijai labai svarbu, kad „Google“ būtų sėkminga. Prieš

kelerius metus to nesupratau, – priduria jis. – Maniau, kad turėtume būti šiek tiek labiau nepriklausomi. Žvelgdamas atgal, manau, kad tai, kas įvyko, galbūt yra net geriau.“

Hassabisas, „DeepMind“ darbuotojams pranešdamas apie susijungimą su „Google Brain“, el. laiške nurodė, kad bendrovės nusprendė susivienyti, nes BDI turi potencialą „sukelti vieną didžiausių socialinių, ekonominių ir mokslinių permainų istorijoje“.

Iš tiesų šios bendrovės susijungė, kad padėtų panikuojančiai „Google“ įveikti konkurentą, lygiai taip pat, kaip ir „OpenAI“ misija padėti žmonijai (be „finansinio spaudimo“) pasikeitė į „Microsoft“ interesų tenkinimą. Tokia misijos permaina, įprasta Silicio slėnyje bei „WhatsApp“ atveju, dabar vyko technologijų srityje, galinčioje turėti kur kas didesnį poveikį visuomenei. „OpenAI“ bandė spręsti šią problemą 2023 metų liepos mėnesį, kai paskelbė, kad Ilya Sutskeveris vadovaus naujajai „Superalignment Team“ komandai. Bendrovė teigė, kad per ketverius metus Sutskeverio tyrėjai išsiaiškins, kaip valdyti DI sistemas, kai jų galimybės aplenks žmonių protus.

Tačiau „OpenAI“ vis dar turėjo akivaizdžią problemą. Ji veikė ne itin skaidriai, tad vis rečiau buvo girdimi balsai, raginantys atidžiau tikrinti didžiuosius kalbos modelius. Gebru, Mitchell ir Bender, kurių garsusis mokslinis straipsnis pagaliau atkreipė dėmesį į šią riziką, vis dar stengėsi įspėti visuomenę apie tai, kaip šie modeliai ir, iš esmės, generatyvinis DI gali įtvirtinti susiformavusius stereotipus. Deja, vyriausybės ir politikos formuotojai daugiau dėmesio skyrė gerai finansuojamai ir labiau besireiškiančiai grupei – DI pesimistams.

14 SKYRIUS

Miglota pražūties nuojauta

Sukūręs „ChatGPT“, Samas Altmanas pradėjo keletą skirtingų lenktynių. Vienos buvo aiškios – kas pirmasis rinkai pasiūlys geriausią didįjį kalbos modelį? Antrame plane vyko kitos – kas kontroliuos naratyvą apie DI? 2023 metų kovą, kelios savaitės po to, kai „Microsoft“ ir „Google“ skubotai pristatė „Bing“ ir „Bard“, Eliezeris Yudkowsky’is žurnalui „Time“ parašė dviejų tūkstančių žodžių straipsnį apie DI ateitį, piešdamas gąsdinantį ateities vaizdą su dar protingesnėmis mašinomis.

„Daug šios srities tyrėjų, įskaitant ir mane, mano, kad, esant panašioms sąlygoms kaip dabar, sukūrus superintelektą, labiausiai tikėtinas rezultatas, kad visi Žemės gyventojai tiesiogine prasme mirtų“, – rašė jis.

Tą patį mėnesį Elonas Muskas ir kiti technologijų lyderiai pasirašė atvirą laišką, kuriame ragino šešiems mėnesiams sustabdyti DI tyrimus dėl žmonijai keliamos grėsmės. „Ar turėtume kurti nežmogišką protą, kuris mus ateityje pranoktų skaičiais ir intelektu, pralenktų ir pakeistų? – buvo rašoma laiške, kurį rengė Jaano Tallinno įkurta organizacija „Future of Life Institute“. – Ar turėtume rizikuoti prarasti savo pačių civilizacijos kontrolę?“ Laiškas, kurį pasirašė beveik 34 tūkstančiai žmonių, sulaukė didelio žiniasklaidos dėmesio visame pasaulyje, apie jį rašė „Reuters“, „Bloomberg“, „New York Times“ ir „Wall Street Journal“.

Daug dėmesio į save atkreipė du tyrėjai, vadinami DI „krikštatėviais“, – Geoffrey’is Hintonas ir Yoshua Bengio, žiniasklaidoje paskelbę galimas DI grėsmes žmonijai. Y. Bengio sakė, kad jaučiasi „pasimetęs“ dėl viso savo gyvenimo darbo, o Hintonas teigė besigailintis dėl kai kurių atliktų tyrimų.

„Keletas žmonių tikėjo mintimi, kad šie dalykai galėtų tapti protingesni už žmones, – sakė jis „New York Times“. – Tačiau dauguma manė, kad tai neįmanoma. Aš irgi taip maniau. Galvojau, kad tai gali nutikti tik po 30–50 metų ar net vėliau. Dabar taip nebeatrodo. Nemanau, kad jie turėtų toliau plėsti šią sritį, kol neįsitikino galintys ją kontroliuoti.“

Rodos, visi didieji DI mokslininkai kartojo tą patį: DI plėtra vyksta pernelyg greitai ir galime katastrofiškai prarasti jo kontrolę. Mintis apie DI intelekto keliamą išnykimo grėsmę tapo tokia populiari viešojoje erdvėje, kad galėjai apie tai užsiminti artimiesiems prie stalo, ir visi jie pritardami linksėtų. Didelė dalis visuomenės buvo apsėsta minties, kad galime sulaukti mašinų valdovų, keliančių grėsmę. 2023 metų pabaigoje apie 22 procentai apklaustų amerikiečių manė, kad per ateinančius 50 metų DI sukels žmonijos išnykimą. Apklausą atliko tyrimų kompanija „Rethink Priorities“; dalyvavo apie 2444 suaugusius JAV gyventojus.

Vis dėlto kalbos apie pražūtį turėjo paradoksalų poveikį pačiam DI verslui – jis klestėjo. Remiantis tyrimų bendrovės „Pitchbook“ duomenimis, finansavimas startuoliams, kuriantiems generatyvinius DI produktus, 2023 metais siekė daugiau nei 21 mlrd. JAV dolerių. Metais anksčiau šis finansavimas buvo apie 5 mlrd. JAV dolerių.

Netiesioginė žinutė apie pavojų keliantį DI buvo įtaigi. Jei šios technologijos gali sunaikinti žmoniją ateityje, ar tai reiškia, kad jos pakankamai veiksmingos stiprinti jūsų verslą?

Atrodė, kad kuo daugiau Samas Altmanas kalbėjo apie „OpenAI“ technologijų grėsmę, pavyzdžiui, Kongrese pasakodamas, kad tokie įrankiai kaip „ChatGPT“ gali „sukelti didelės žalos pasauliui“, tuo daugiau dėmesio ir pinigų jis pritraukė. 2023 metų sausį „OpenAI“ sulaukė dar vienos investicijos iš „Microsoft“, šį kartą siekiančios 10 mlrd. JAV dolerių, mainais į 49 procentus bendrovės akcijų. „Microsoft“ dabar buvo itin arti visiškos „OpenAI“ kontrolės.

Dario’aus Amodei’aus ir kitų tyrėjų grupės iš „OpenAI“ įkurta nauja bendrovė „Anthropic“ taip pat pritraukė didelių investicijų. Iki 2023 metų pabaigos bendrovė gavo 2 mlrd. JAV dolerių iš „Google“ ir 1,3 mlrd. JAV dolerių iš „Amazon“. Per metus „Anthropic“ vertė išaugo keturis kartus ir siekė daugiau nei 20 mlrd. JAV dolerių. Atrodė, kad, sukūrus itin saugią DI sistemą, galima tapti labai turtingu. Remiantis „Anthropic“ dokumentais, kuriuos gavo „TechCrunch“, bendrovė slapta siekė pritraukti 5 mlrd. JAV dolerių, kad galėtų įžengti į keliolika DI rinkos šakų ir mesti iššūkį „OpenAI“. „Šie modeliai pradėtų automatizuoti didelę dalį ekonomikos“, – buvo rašoma „Anthropic“ dokumentuose. Taip pat teigta, kad tai lenktynės, kuriose „Anthropic“ galėtų pirmauti daug metų, jei iki 2026 metų sugebėtų sukurti „geriausius“ DI modelius.

„Anthropic“ saugumo prioretizavimas bendrovę vaizdavo kaip ne pelno organizaciją, kurios misija – „užtikrinti, kad visuomenę pakeisti galintis DI padėtų žmonėms ir visuomenei klestėti“. Tačiau „OpenAI“ didžiulė sėkmė išleidus „ChatGPT“ pasauliui parodė, kad didžiausių planų turinčios bendrovės gali tapti ir pelningiausiomis investicijomis. Pareiškimas, kad kuriamas saugesnis DI modelis, tapo didesnių technologijų bendrovių, irgi norinčių įsitraukti į šį žaidimą, slaptąja žinute.

„Anthropic“ stengėsi pagrįsti savo logiką. Kad išsiaiškintų, kaip padaryti DI sistemas saugesnes, jai nepakako išanalizuoti, kaip veikia galingiausios DI sistemos. Bendrovė turėjo jas sukurti. Taigi, „Anthropic“ metė iššūkį didžiosioms technologijų bendrovėms, kurios vienintelės valdė milžinišką skaičiavimo galią. Pavyzdžiui, pagal „Anthropic“ susitarimą su „Google“, bendrovei priklausė tam tikras nemokamų debesijos infrastruktūros paslaugų limitas. Tai būtų leidę jai sukurti didįjį kalbos modelį, galintį konkuruoti su „OpenAI“.

Visuomenėje vyravo dvi skirtingos žmonių grupės, reikalaujančios saugesnio DI. Vienoje grupėje buvo tokie žmonės kaip Altmanas ir Amodei’us, kurie pasirašė dar vieną atvirą laišką su teiginiu, kad „DI

keliamos išnykimo grėsmės mažinimas turėtų tapti pasauliniu prioritetu šalia kitų visuomeninio masto pavojų, tokių kaip pandemijos ir branduoliniai karai“. Jie pasisakė už „DI saugumo“ idėją, miglotai kalbėdami apie ateityje kilsiančią grėsmę, retai teikdamiesi paaiškinti, ką pavojų keliančios DI sistemos darytų ir kada tai įvyktų. Išsakydami susirūpinimą Kongresui, jie taip pat buvo linkę pasisakyti už nedidelį DI reguliavimą.

Kitoje grupėje buvo tokie žmonės kaip Timnita Gebru ir Margareta Mitchell, kurios ne vienus metus nerimavo dėl DI visuomenei keliamos grėsmės. Į šią „DI etikos“ grupę linko moterys ir nebaltaodžiai asmenys, kurie patys susidūrė su stereotipizavimu ir baiminosi, kad DI sistemos ir toliau skatins nelygybę. Bėgant laikui, jie vis labiau niršo dėl „DI saugumo“ reikalaujančios grupės veiksmų ir kad pastarosios pajamos buvo gerokai didesnės. Finansavimo skirtumas buvo didžiulis. Pabrėžiantiesiems DI etikos svarbą dažnai trūko lėšų. Tokios grupės kaip Europos skaitmeninių teisių iniciatyva (angl. *European Digital Rights Initiative*) – dvidešimt vienus metus veikiantis ne pelno organizacijų tinklas, kovojantis prieš veido atpažinimo technologijas ir algoritmų šališkumą, – 2023 metais turėjo vos 2,2 mln. JAV dolerių metinį biudžetą. Panaši situacija buvo ir su „AI Now Institute“ Niujorke, kuris tyrė DI naudojimą sveikatos priežiūros ir baudžiamosios teisės sistemose. Instituto metinis biudžetas siekė mažiau nei 1 mln. JAV dolerių.

Organizacijos, kurios daugiausia dėmesio skyrė DI „saugumui“ ir išnykimo grėsmei, sulaukė ženkliai didesnio finansavimo, neretai iš milijardierių rėmėjų. „The Future of Life Institute“ – Kembridže, Masačusetse veikianti ne pelno organizacija, tirianti veiksmingiausius būdus, leisiančius sustabdyti DI prieigą prie ginklų, 2021 metais gavo 25 mln. JAV dolerių paramą iš kriptovaliutų magnato Vitaliko Buterino. Ši parama buvo didesnė nei visų tuo metu už DI etiką kovojančių grupių kartu paimti metiniai biudžetai.

„Open Philanthropy“, labdaros organizacija, įkurta „Facebook“ bendrakūrėjo milijardieriaus Dustino Moskovitzo, per kelerius metus skyrė ne vieną milijoninę dotaciją DI saugumo tyrimams, įskaitant ir 2022 metais 5 mln. JAV dolerių siekusią paramą „Center for AI Safety“ ir 11 mln. JAV dolerių paramą Berklio „Center for Human-Compatible AI“.

Trumpai tariant, Moskovitzas buvo didžiausias DI saugumo rėmėjas. Jis su žmona Cari Tuna planuoja paaukoti beveik 14 mlrd. JAV dolerių turtą. Moskovitzas taip pat skyrė 30 mln. JAV dolerių dotaciją „OpenAI“, kai ji buvo įsteigta kaip ne pelno organizacija.

Kodėl tiek daug pinigų skiriama inžinieriams, tobulinantiems didžiąsias DI sistemas, aiškinantiems, kad jie daro tai dėl sistemų saugumo ateityje, o taip mažai tyrėjams, siekiantiems jas perprasti šiandien? Į klausimą iš dalies padeda atsakyti Silicio slėnyje susidomėjimo sulaukusi idėja, kuria siekta rasti veiksmingiausią gerų idėjų sklaidos būdą. Ją populiarino maža filosofų grupė iš Oksfordo universiteto Anglijoje. Praėjusio amžiaus devintajame dešimtmetyje Oksfordo filosofas Derekas Parfitas pradėjo rašyti apie naują utilitarinės etikos šaką, kuri buvo nukreipta į ateitį. Jis siūlė įsivaizduoti, kad ant žemės paliekate sudaužytą butelį, ant kurio po šimto metų užlipęs vaikas susižeidžia pėdą. Nors vaikas šiuo metu dar nė negimęs, ant jūsų pečių turėtų kristi tokia pati kaltės našta, kaip ir vaikui susižeidus šiandien.

„Jo mintis labai paprasta – moraliai ateities žmonės yra lygiai tokie pat svarbūs kaip ir gyvenantieji dabar, – teigė Davidas Edmondsas, kuris 2023 metais parašė Parfito biografiją. – Įsivaizduokite tris scenarijus. A – vyrauja taika; B – 7,5 mlrd. iš 8 mlrd. pasaulio žmonių pražudomi karo metu; C – visi žmonės nužudomi. Daugumos žmonių intuicija sako, kad skirtumas tarp A ir B situacijų yra gerokai didesnis nei skirtumas tarp B ir C. Tačiau Parfitas teigia kitaip. Atotrūkis tarp B ir C situacijų yra gerokai didesnis nei atotrūkis tarp A ir B. Sunaikinus visą žmoniją, sunaikinamos ir visos ateities kartos.“

Štai vienas būdų tam apskaičiuoti. Žinduoliai Žemėje gyvena apie milijoną metų, o žmonės gyvuoja apie du šimtus tūkstančių metų. Tai reiškia, kad teoriškai mums dar liko apie aštuoni šimtai tūkstančių metų gyvenimo šioje planetoje. Jei, kaip teigia Jungtinių Tautų prognozės, pasaulio gyventojų skaičius iki šio šimtmečio pabaigos nusistovės ties vienuolikos milijardų riba, o vidutinė gyvenimo trukmė padidės iki 88 metų, pagal vienus iš skaičiavimų, tai gali reikšti, kad ateityje gyvens dar *šimtas trilijonų žmonių*.

Kad geriau suvoktumėte šiuos skaičius, įsivaizduokite, kad ant jūsų valgomojo stalo guli mažas peilis ir viena pupa. Peilis simbolizuoja žmonių, gyvenusių ir mirusių praeityje, skaičių. Pupa simbolizuoja šiandien gyvenančius žmones. Valgomojo stalo paviršius – tai skaičius žmonių, kurie dar tik gyvens. Šis skaičius galėtų būti didesnis, jei žmonės įrodytų galintys gyventi Žemėje ilgiau nei tipinės rūšies žinduoliai.

2009 metais australų filosofas Peteris Singeris išplėtė Parfito darbą, parašęs knygą „Gyvenimas, kurį galite išgelbėti“ (*The Life You Can Save*). Egzistavo vienas sprendimas – turtingi žmonės neturėtų aukoti pinigų tik tam, kas jiems atrodo tinkama, o turėtų vadovautis racionalesniu požiūriu, maksimaliai padidinančiu jų labdaros įtaką ir leidžiančiu padėti kuo daugiau žmonių. Padėdami tiems, kurie ateityje dar tik gims, galėtumėte būti dar dorybingesni.

2011 metais šios idėjos iš akademinių tekstų persikėlė į realybę ir davė pagrindą naujai ideologijai, kai 24 metų filosofas iš Oksfordo Willas MacAskillas kartu su kitais įkūrė organizaciją „80 000 Hours“. Šis skaičius rodė, kiek vidutiniškai valandų žmogus praleidžia dirbdamas. Tiksline auditorija tapo jauni JAV universitetų absolventai, kuriems organizacijos nariai patardavo dėl karjeros, turinčios didžiausią moralinį poveikį. Dažnai į technologijas linkusius jaunuolius jie nukreipdavo link DI saugumo srities. Organizacija skatino absolventus rinktis geriausiai apmokamus darbus, kurie leistų jiems paaukoti kiek įmanoma daugiau pinigų reikšmingiems tikslams.

MacAskillas ir jo jauna komanda galiausiai persivadino į „Center for Effective Altruism“ („Veiksmingojo altruizmo centras“). Taip gimė jų naujasis kredo. Pagrindinė veiksmingojo altruizmo idėja buvo našumas. Žmonės, gyvenantys turtingose šalyse, privalėjo padėti skurdesnėms šalims, nes taip jie galėjo už savo pinigus pasiekti didžiausią naudą. Pavyzdžiui, per pasaulines sveikatos labdaros organizacijas aukodami Afrikai, padėtumėte daugiau žmonių nei aukodami vargšams Amerikoje. Taip pat iš moralinės pusės buvo geriau uždirbti kiek įmanoma daugiau, kad taptumėte kaip Dustinas Moskovitzas ir daug aukotumėte. Kalbėdamas studentams, MacAskillas rodydavo skaidres ir klausdavo, ar daugiau gero jie galėtų padaryti būdami gydytojais, ar bankininkais. Jis manė, kad geriau tapti bankininku. Kaip gydytojas Afrikoje galbūt ir galėtumėte išgelbėti tam tikrą skaičių gyvybių, tačiau kaip bankininkas pajėgtumėte pasamdyti *keletą* gydytojų, kurie išgelbėtų dar daugiau gyvybių.

Tai absolventams suteikė intuicijai prieštaraujantį naują požiūrį į moderniojo kapitalizmo nelygybę. Sistema, kuri leidžia keletui žmonių tapti milijardieriais, nebebuvo ydinga. Susikrovę nesuvokiamus turtus, jie galėtų padėti daugiau žmonių!

Didžiausio susidomėjimo judėjimas sulaukė 2012 metais, kai MacAskillas kreipėsi į vieną žmogų, kurį norėjo pritraukti į savo veiklą. Tai buvo Masačusetso technologijos instituto studentas tamsiais garbanotais plaukais – Samas Bankmanas-Friedas. Jiedu susitiko išgerti kavos. Paaiškėjo, kad Bankmanas-Friedas jau buvo Peterio Singerio gerbėjas ir domėjosi gyvūnų gerovės klausimais.

MacAskillas atkalbėjo Bankmaną-Friedą nuo minties dirbti tiesiogiai gyvūnų labui ir pasakė, kad gerokai labiau jiems padėtų, pasukęs į pelningą sritį. Michaelio Lewiso knygoje „Link begalybės“ (*Going Infinite*) apie Bankmano-Friedo pakilimą ir nuosmukį teigiama, kad jis iškart susidomėjo. „Tai, ką jis kalbėjo, man tarsi atrodė akivaizdžiai teisinga“, – rašo Bankmanas-Friedas. Jis greit įsidarbino kiekybinės prekybos

įmonėje (angl. *quantitative trading firm*) ir galiausiai 2019 metais įkūrė kriptovaliutų keityklos bendrovę „FTX“.

Bankmanas-Friedas veiksmingąjį altruizmą pavertė šio verslo pagrindu. Jo bendrovės bendrakūrėjai ir vadybininkų komanda vadovavosi veiksminguoju altruizmu ir MacAskillą laikė „FTX“ priklausančio fondo „Future Fund“ nariu. Šis fondas 2022 metais skirs 160 mln. JAV dolerių veiksmingojo altruizmo labui; dalis tų pinigų bus tiesiogiai susijusi su MacAskillu. Bankmanas-Friedas dažnai žiniasklaidai kalbėjo apie planus visus pinigus paaukoti. Dideliuose „FTX“ reklaminiuose plakatuose jis buvo vaizduojamas vilkintis savo ypatinguosius marškinėlius ir kargo stiliaus šortus, šalia puikavosi užrašas – „užsiimu kriptovaliutomis, nes noriu padaryti didžiausią teigiamą poveikį pasauliui“. Jis vaizdavo save kaip asketišką asmenį, kuris, nors ir buvo milijardierius, vairavo „Toyota Corolla“, gyveno su kambariokais ir dažnai atrodė netvarkingai.

Daug technologijų specialistų jo moralinį požiūrį laikė gaiviu oro gurkšniu. Inžinieriai, susidūrę su problema, dažnai ją spręsdavo taikydami formulę, taisydami kodo klaidas ir optimizuodami programinę įrangą, ją nuolatos testuodami ir tikrindami. Dabar beveik kaip matematinius uždavinius kiekybiškai buvo galima apskaičiuoti ir moralines dilemas. Žmonės iš veiksmingojo altruizmo rato kartais kalbėjo apie labdaringos veiklos naudos optimizavimą, kai dėmesys sutelkiamas į „tikėtiną vertę“ – skaičių, gaunamą rezultato vertę padauginus iš jo tikimybės.

Veiksmingajam altruizmui išpopuliarėjus Silicio slėnyje, dėmesys nuo pigių tinklelių, saugančių nuo maliarijos plitimo, pirkimo, siekiant išgelbėti kuo daugiau Afrikos žmonių, persikėlė prie mokslinę fantastiką primenančių klausimų. Elonas Muskas socialiniame tinkle „Twitter“ rašė, kad 2022 metais išleista MacAskillo knyga „labai artima mano filosofijai“. Muskas siekė išsiųsti žmones į Marsą, kad užtikrintų ilgalaikį žmonijos išlikimą. DI sistemoms tampant pažangesnėms, buvo pravartu neleisti

joms tapti pavojingoms ir galinčioms sunaikinti žmoniją. Daugelis bendrovių „OpenAI“, „Anthropic“ ir „DeepMind“ darbuotojų vadovavosi veiksminguoju altruizmu.

Veiksmai, paremti DI keliama išnykimo rizika, yra racionalus skaičiavimas. Net jei yra tik 0,00001 procento tikimybė, kad DI sunaikins žmoniją, to kaina tokia didelė, kad iš esmės ji yra begalinė. Padauginę nedidelę tikimybę su begaline kaina, vis tiek gausite begalinio dydžio problemą. Ši mintis yra dar svaresnė, jei, kaip kai kurie DI saugumo šalininkai, tikite, jog ateities kompiuteriai savyje talpins milijardų žmonių sąmonę bei kurs naujas, sąmoningas skaitmenines gyvybių formas. Šimto trilijonų žmonių, kurie dar tik gyvens ateityje, skaičius galėtų būti daug didesnis. Laikantis šių moralinių skaičiavimų, logiška teikti pirmenybę nedidelei galimybei, rodančiai, kad gali tekti nuo sunaikinimo gelbėti daugiau nei šimtą trilijonų fizinių ir skaitmeninių gyvybių. Palyginti su tuo, pasaulio skurdas yra vos menka apvalinimo klaida.

2015 metais įkūrus „OpenAI“, pradėta daug investuoti priežastims, dėl kurių DI galėtų sukelti žmonijos išnykimą, tirti. Moskovitzo valdoma bendrovė „Open Philanthropy“ padidino skiriamas lėšas ilgalaikiams DI tyrimams, įskaitant ir DI saugumą, nuo 2 mln. JAV dolerių 2015 metais iki daugiau nei 100 mln. JAV dolerių 2021 metais.

Prisijungė ir Bankmanas-Friedas. Jis pažadėjo skirti 1 mlrd. JAV dolerių iš jo bendrovei „FTX“ priklausančio fondo „Future Fund“, kurį valdė Nickas Becksteadas ir MacAskillas, „ilgalaikėms žmonijos perspektyvoms pagerinti“. Fondui rūpimų sričių sąrašas prasidėjo nuo „saugaus DI vystymo“.

Žurnalas „New Yorker“ aprašydamas „Future Fund“ darbą paminėjo, kad biuro darbuotojai organizacijos būstinėje Berklyje, Kalifornijoje, dažnai šnekučiuojasi apie tai, kada gali įvykti DI apokalipsė.

„Kokie jūsų darbo planai? – vieni kitų klausdavo darbuotojai. – Koks jūsų *p(doom)*?“

„P“ reiškė tikimybę, o klausimas buvo susijęs su tuo, kaip žmonės vertina tikimybę, kad vieną dieną DI visus pražudys. Optimistiškesnis žmogus galėtų nurodyti 5 procentų savo *p(doom)*. Ajeya Cotra, „Open Philanthropy“ tyrimų analitikė, padėjusi nuspręsti, kam skirti dotacijas, vienoje tinklalaidėje užsiminė, kad jos *p(doom)* yra tarp 20 ir 30 procentų.

Niekas nežinojo, koks Bankmano-Friedo *p(doom)*. Vis dėlto DI saugumas jam rūpėjo pakankamai, kad investuotų 500 mln. JAV dolerių į bendrovę „Anthropic“. „FTX“ bendrakūrėjai ir veiksmingojo altruizmo, kaip ir jis, šalininkai Nishadas Singhas ir Caroline Ellison taip pat investavo į startuolį, metais anksčiau atsiskyrusį nuo „OpenAI“.

2022 metų pradžioje MacAskillas pastebėjo Elono Musko įrašą socialiniame tinkle „Twitter“. Muskas sakė norintis įsigyti šį socialinį tinklą ir taip išgelbėti žodžio laisvę. Škotų filosofas išsiuntė Muskui žinutę. Tuo metu Bankmano-Friedo turtas siekė 24 mlrd. JAV dolerių, jis buvo vienas turtingiausių veiksmingojo altruizmo atstovų pasaulyje. Muskas, valdantis 220 mlrd. JAV dolerių siekiantį turtą, vienu rankos mostu galėjo veiksmingąjį altruizmą paversti didžiausiu filantropiniu judėjimu.

MacAskillas parašė Muskui, kad Bankmanas-Friedas taip pat norėjo įsigyti „Twitter“, siekdamas padaryti tinklą „geresnį pasauliui“. Ar jiedu ketino suvienyti jėgas?

„Ar jis turi labai daug pinigų?“ – žinute atsakė Muskas.

„Priklauso nuo to, kaip apibūdinsi „labai daug!“ – remiantis teismo dokumentais, atsakė jam MacAskillas. Jis teigė, kad Bankmanas-Friedas gali prisidėti 8 mlrd. JAV dolerių.

„Pradžia yra“, – atsakė Muskas.

„Ar norėtum, kad jus supažindinčiau žinute?“– paklausė MacAskillas.

Muskas į šį klausimą neatsakė. „Tu už jį garantuoji?“ – paklausė jis.

„Tikrai taip! – atsakė MacAskillas. – Labai atsidavęs tam, kad ilgalaikė žmonijos ateitis būtų sėkminga.“

„Tuomet gerai.“

„Puiku!“

Nors Muskas galiausiai susisiekė su Bankmanu-Friedu, jie finansinio sandorio niekada nesudarė, o tai reiškė, kad Muskas išvengė nemalonumų. Po kelių mėnesių pasklidus gandams, kad Bankmanas-Friedas bendrovės viduje neteisėtai pervedinėjo klientų lėšas, bendrovė „FTX“ žlugo. Teismo posėdyje prokurorai jį kaltino iš tūkstančių bendrovės klientų ir investuotojų pasisavinus 8 mlrd. JAV dolerių, jam grėsė kelios dešimtys metų kalėjimo. Nors Bankmanas-Friedas vaizdavo save kaip asketišką, pasirodė, kad jis ilgą laiką gyveno prabangioje mansardoje Bahamose, švaistydamas šimtus milijonų JAV dolerių įvairioms investicijoms. Dabar didelė dalis jo veiksmingajam altruizmui uždirbtų pinigų dingo ir paaiškėjo, kad judėjimo idėjos jam nebuvo tokios artimos.

Netrukus po „FTX“ žlugimo Bankmanas-Friedas davė išskirtinį interviu naujienų svetainei „Vox“.

„Taigi, etiniai dalykai – daugiausia tik fasadas?“ – klausė žurnalistas.

„Taip“, – atsakė Bankmanas-Friedas.

„Jums puikiai sekėsi kalbėti apie etiką, turint omenyje, kad pats į tai žiūrėjote kaip į žaidimą su laimėtojais ir pralaimėtojais.“

„Aha, – atsakė Bankmanas-Friedas. – Hehe. Turėjau toks būti.“

„FTX“ nuosmukis metė didžiulį šešėlį veiksmingojo altruizmo reputacijai, o ši istorija tapo kai kurių pagrindinių šio judėjimo problemų alegorija. Pirmoji judėjimo problema nuspėjama. Žmonės, pasiryžę daryti kuo daugiau gero, tuo pat metu besistengiantys susikrauti kuo didesnį turtą, tikriausiai tapo labiau linkę į korupcinį elgesį ir neapdairius, egoistiškus veiksmus. Kodėl „Twitter“ įsigijimas turėtų padėti žmonijai ilguoju laikotarpiu, nebuvo akivaizdu. Tačiau Bankmanas-Friedas buvo pasirengęs išleisti 8 mlrd. JAV dolerių ir kartu su Musku įsigyti šį socialinį tinklą, o įvykdęs šį veiksmingojo altruizmo aktą atsistoti ant pjedestalo kartu su turtingiausiu pasaulio žmogumi.

Žlugus „FTX“, MacAskillas pasuko į „Twitter“, siekdamas sumažinti žalą. „Blaiviai mąstantis [veiksmingas altruistas] turėtų griežtai prieštarauti mąstymui „tikslas pateisina priemones“, – rašė jis. Visgi, paties judėjimo nuostatai skatino tokius žmones kaip Bankmanas-Friedas siekti tikslų bet kokiomis įmanomomis priemonėmis. Net jei tai reiškė žmonių išnaudojimą. Judėjimas apakino net tokį išsilavinusį žmogų kaip Oksfordo akademikas MacAskillas, kuris pasirinko prisijungti prie žmogaus, valdančio kriptovaliutų keityklos verslą. Nors jis puikiai žinojo, kad šis verslas geriausiu atveju yra spekuliacija, o blogiausiu – pavojingos formos lošimas.

Bankmanas-Friedas savo dviveidiškumą galėjo pateisinti didesnio tikslo siekiu – maksimaliai padidinti žmonių laimę. Muskas galėjo išsisukti nežmogiškais veiksmais, pradedant nepagrįstu žmonių apšaukimu pedofilais „Twitter“ tinkle ir baigiant plačiai paplitusiu rasizmu jo bendrovės „Tesla“ gamyklose. Juk jis siekė didesnių tikslų, tokių kaip paversti „Twitter“ žodžio laisvės utopija, o žmones – tarpplanetine rūšimi. „OpenAI“ ir „DeepMind“ įkūrėjai galėjo panašiai teisinti savo augančią paramą didelėms technologijų bendrovėms. Galiausiai išvystę BDI, jie padarys didelę paslaugą žmonijai.

Technologijų specialistai, tokie kaip Altmanas ir Hassabisas, žinojo, kad visuomenės problemos, kurias jie siekė išspręsti pasitelkę BDI, buvo painios ir sudėtingos. Todėl nemenka jų dalis daugiau ar mažiau rėmėsi veiksminguoju altruizmu. Tai sprendžiant problemas leido eiti paprastesniu, racionalesniu keliu ir netrukdė uždirbti kuo daugiau pinigų. Milijardieriai buvo ne pasaulinio skurdo priežastis, o jo sprendimas.

Tai padėjo ir lengviau atsiriboti nuo žmonijos. Populiari veiksmingojo altruizmo frazė „užsičiaupk ir daugink“ reiškė, kad, priimdami etinę reikšmę turintį sprendimą, turėtumėte maksimaliai didinti savo našumą, pamiršti asmenines emocijas ir moralinę intuiciją. Dėl didelio atsidavimo žmonijai veiksmingojo altruizmo šalininkai, tokie kaip Altmanas,

emociškai atsiribojo nuo aplinkinio pasaulio ir susitelkė į savo tikslą. Veiksmingojo altruizmo burbule žmonės kartu dirbo, socializavosi, vienas kitą rėmė ir kūrė romantinius santykius.

2017 metais „Open Philanthropy“ pažadėjo 30 mln. JAV dolerių skirti bendrovei „OpenAI“. Labdaros organizacija buvo priversta paviešinti, kad Dario’us Amodeui’us ją konsultuoja technologijų klausimais. Tuo metu jis dirbo vyresniuoju inžinieriumi „OpenAI“. Taip pat buvo paviešinta, kad Amodeui’us gyvena tame pačiame name kaip ir „Open Philanthropy“ vykdomasis direktorius Holdenas Karnofsky’is. Sužinota, kad Karnofsky’is yra susižadėjęs su Amodeui’aus seserimi Daniela, taip pat dirbančia „OpenAI“. Visi jie buvo veiksmingojo altruizmo šalininkai. Visas kraujomaišos ratas.

Šis judėjimas buvo uždaras ir darėsi vis neaiškesnis. Tas pats vyko ir su tokiomis DI bendrovėmis kaip „OpenAI“, „DeepMind“ ir „Anthropic“, kurių didelė dalis darbuotojų buvo judėjimo sekėjai. Tikriausiai vienas veiksmingiausių dalykų, ką šios bendrovės galėtų padaryti, kad DI nepakenktų žmonėms, – tai užtikrinti, kad jų DI sistemos būtų skaidresnės. To ir siekė Gebru su Mitchell. Galiausiai, kaip ateities žmonės sustabdytų DI nuo žalos darymo, jei jiems trūks žinių apie tai, kaip veikia jo mechanizmai, jei tyrėjams dešimtmečius buvo uždrausta tirti DI mokymo duomenis ir algoritmus? Kitais žodžiais tariant, skaidrumas, kurio siekė DI etikos atstovai, padėtų spręsti ir ateityje gresiantį išnykimo klausimą.

„OpenAI“ argumentas, kad ji turėjo išlaikyti slaptumą, siekdama apsisaugoti nuo piktavališko jos technologijos naudojimo, mažai įtikimas. 2019 metais „OpenAI“ nusprendė galinti pristatyti GPT-2, nes nepastebėjo „tvirtų piktavališko naudojimo įrodymų“. Jei tai tiesa, kodėl nepaskelbus jos mokymo detalių? Tikriausiai todėl, kad Altmanas norėjo apsaugoti „OpenAI“ nuo konkurentų ir teisinių ieškinių. Jei „OpenAI“ būtų tapusi skaidri, tai būtų padėję konkurentams, o ne piktadariams.

Konkurentams tai būtų leidę nukopijuoti modelius ir atskleisti, kokiu mastu „OpenAI“ kopijavo autorių teisių saugomus darbus.

Altmanas ir Hassabisas savo bendroves įkūrė vedami didaus tikslo padėti žmonijai, tačiau tikrieji jų žmonėms suteikti privalumai buvo tokie pat neaiškūs, kaip ir interneto ir socialinių tinklų teikiama nauda. Aiškesnė buvo nauda, kurią jie davė „Microsoft“ ir „Google“, – naujos, šaunesnės paslaugos ir tvirta pozicija augančioje generatyvinio DI rinkoje.

„Microsoft“ pagal „OpenAI“ technologiją paremtą DI asistentą „Copilot“ pavertė plačios srities paslauga, skirta „Windows“, „Word“, „Excel“ ir verslui suprojektuotai programinei įrangai „Dynamics 365“. Analitikai skaičiuoja, kad „OpenAI“ technologija iki 2026 metų bendrovei „Microsoft“ gali uždirbti milijardus metinių pajamų. 2023 metų pabaigoje, Nadellai ir Altmanui kalbant ant tos pačios scenos, Nadellos buvo paklausta, kaip „Microsoft“ sekasi bendradarbiauti su „OpenAI“. Jis prapliupo nevaldomai juoktis. Atsakymas buvo toks akivaizdus, kad kėlė juoką. Žinoma, kad sekėsi gerai.

„Microsoft“ mielai leido pinigus augančiam DI verslui ir nuo 2024 metų planavo skirti daugiau nei 50 mlrd. JAV dolerių didžiulių duomenų centrų plėtrai. Šie centrai buvo generatyvinio DI varikliai. Tai būtų vienas didžiausių infrastruktūros plėtrų projektų istorijoje. „Microsoft“ išlaidos pranoktų valstybinių geležinkelių, užtvankų ir kosmoso programų projektų išlaidas. „Google“ taip pat plėtė savo duomenų centrus.

2024 metų pradžioje visos sritys nuo žiniasklaidos iki pramogų sektoriaus ir „Tinder“ į savo programėles ir paslaugas įtraukė naujas generatyvinio DI funkcijas. Prognozuota, kad generatyvinio DI rinka plėsis daugiau nei 35 procentais per metus ir 2028 metais jos vertė sieks 52 mlrd. JAV dolerių. Pramogų bendrovės teigė galinčios per trumpesnį laiką sukurti daugiau turinio filmams, TV laidoms ir kompiuteriniams žaidimams. Jeffrey’is Katzenbergas, „DreamWorks Animation“

bendrakūrėjis ir filmų „Šrekas“ bei „Kung Fu Panda“ prodiuseris, teigė, kad generatyvinio DI naudojimas animacinių filmų kūrimo kaštus sumažintų 90 procentų. „Senais gerais laikais galėjo prireikti 500 menininkų ir kelerių metų pasaulinio lygio animaciniam filmui sukurti, – sakė jis „Bloomberg“ konferencijoje 2023 metų lapkritį. – Nemanau, kad, praėjus trejiems metams nuo dabar, tam reikės 10 procentų viso to.“

Generatyvinis DI reklamą padarytų dar labiau gąsdinamai asmenišką. Daugelį metų reklamos galėjo būti vienu metu orientuotos į dideles žmonių grupes. Dabar jos gali būti skirtos konkrečiam asmeniui, pasitelkus itin suasmenintas vaizdo reklamas, kuriose gali atsidurti ir jūsų vardas. Pasaulio ekonomikos forume buvo kalbama, kad didieji kalbos modeliai padidintų darbo vietų, kurioms reikalingas kritinis mąstymas ir kūrybiškumas, skaičių. Visi, nuo inžinierių iki reklamos tekstų kūrėjų ir mokslininkų, galėtų naudotis jais kaip savo smegenų priedu. Vyriausybės taip pat naujino savo DI sistemas, naudojamas socialinės paramos prašymams, asmens polinkiui nusikalsti vertinti, viešosioms erdvėms stebėti.

„Google“, „Microsoft“ ir naujos kartos startuoliai varžėsi, stengdamiesi užkariauti kuo didesnę naujos rinkos dalį ir pelnyti pranašumą prieš konkurentus. Remiantis 2023 metais atlikta „Fast Company“ apklausa, beveik pusė JAV bendrovių valdybos narių generatyvinį DI laikė savo bendrovių „pagrindiniu prioritetu, svarbesniu už bet ką kita“. Štai kaip, pavyzdžiui, pažinčių programėlės „Bumble“ generalinė direktorė apibūdina pagrindinius 2024 metų planus: „Labai norime rimtai įtraukti DI, – sakė ji. – DI ir generatyvinis DI gali atlikti didžiulį vaidmenį, padėdami greičiau rasti tinkamą žmogų.“

„Bumble“ norėjo pasitelkti „ChatGPT“ naudojamą technologiją porų deriniams kurti. Užuot mobiliojoje programėlėje pažymėję daugybę langelių, galėtumėte pokalbių robotui tiesiog pasakyti, ko norite iš partnerio: nuo poreikio susilaukti vaikų iki jūsų politinių pažiūrų ir to, ką paprastai veikiate šeštadienio rytais. Dirbtinio intelekto porų

sudarytojas tuomet „pasikalbėtų“ su kitų „Bumble“ vartotojų porų sudarytojais, kad surastų tinkamiausią žmogų. Nebereiks „braukti“ šimtų vartotojų profilių į kairę ir į dešinę, DI tai padarys už jus.

Šioms ir kitoms verslo idėjoms įgijus pagreitį, vis dar buvo neaišku, kokia viso to kaina. Algoritmai jau darė vis didesnę įtaką mūsų gyvenimo sprendimams nuo to, ką skaitėme internete, iki to, ką įmonės norėjo įdarbinti. Dabar jie buvo paruošti atlikti daugiau galvoti reikalaujančių užduočių. Tai kėlė nepatogių klausimų apie žmogaus veiklą, mūsų gebėjimą spręsti problemas ir paprasčiausiai naudotis vaizduote.

Duomenys rodo, kad kompiuteriai jau perėmė dalį mūsų kognityvinių gebėjimų. To pavyzdys – trumpalaikė atmintis. 1955 metais Harvardo profesorius George’as Millaras tyrė žmonių atminties gebėjimus, duodamas tyrimo dalyviams atsitiktinį spalvų, skonių ir skaičių sąrašą, kurį jie turėjo įsiminti. Paskui jis paprašydavo tiriamųjų išvardinti kuo daugiau sąraše esančių dalykų. Pastebėta, kad dauguma užstringa kažkur ties septyniais. Millaro straipsnis „Magiškasis skaičius septyni, plius ar minus du“ (*The Magical Number Seven, Plus or Minus Two*) turėjo įtakos tam, kaip inžinieriai projektavo programinę įrangą ir kaip telefonų bendrovės skirstė numerius į segmentus, kad lengviau juos įsimintume. Tačiau, remiantis naujesniais duomenimis, magiškasis skaičius septyni nukrito iki keturių.

Kai kurie tai vadina „Google“ efektu. Vis labiau kliaudamiesi paieškos milžine, norėdami prisiminti faktus ar prašydami kelio nuorodų, perdavėme savo atmintį „Google“ ir nepastebimai susilpninome trumpalaikę atmintį. Ar galėtų tas pats nutikti ir gilesniems mūsų kognityviniams aspektams, jei pernelyg kliausimės DI generuodami idėjas, kurdami tekstus ar meną? Socialiniame tinkle „Twitter“ keletas programinės įrangos kūrėjų prisipažino, jog rašydami kodą DI naudoja taip dažnai, kad jų produktyvumas ženkliai krinta „Copilot“ įrankiui laikinai išsijungus.

Istorija rodo, kad žmonės linkę baimintis, kad inovacijos privers smegenis susitraukti. Prieš daugiau nei du tūkstančius metų išpopuliarėjus rašymui, filosofai, pavyzdžiui, Sokratas, nerimavo, kad tai susilpnins žmonių atmintį, nes iki tol žinias buvo galima perduoti tik žodžiu. Skaičiuotuvų įtraukimas į švietimo sistemą kėlė susirūpinimą, kad mokiniai praras elementarius skaičiavimo įgūdžius.

Vis dar nežinome visų neigiamų priklausomybės nuo technologijų padarinių ir kaip tai gali paveikti kalbos apdorojimą smegenyse. Mašina, pajėgi kurti kalbą, generuoti idėjas ir parengti verslo planą, daro daug daugiau nei ta, kuri apdoroja skaičius ar indeksuoja tinklalapius. Ji išstumia abstraktųjį mąstymą ir planavimą.

Dar nežinome, kaip mūsų kritinio mąstymo ar kūrybiškumo įgūdžiai suprastės, naujos kartos profesionalams pradėjus naudoti didžiuosius kalbos modelius. Arba kaip kis mūsų tarpusavio santykiai, vis daugiau žmonių pasirinkus pokalbių robotus kaip terapeutus ar romantinius partnerius. Kelios įmonės pokalbių robotus pradėjo diegti ir žaisluose. Remiantis vienu 2023 metais atliktu tyrimu, kuriame dalyvavo apie tūkstantį JAV suaugusiųjų, vienas iš keturių amerikiečių mieliau rinktųsi kalbėtis su DI pokalbių robotu negu su terapeutu žmogumi. Tai nestebina – jei „ChatGPT“ pateiksite emocinio intelekto testą, jis jį išlaikys.

Pasak Altmano, „ChatGPT“ technologija žymiai paveiks mūsų ekonomiką, pakeisdama darbo vietas. Tyrėjai mano, kad kalbos modeliai ir kitos generatyvinio DI formos gali padidinti pajamų nelygybę. Tarptautinio valiutos fondo prognozėmis, DI sistemų naudojimas gali paskatinti daugiau investicijų į pažangias ekonomikas. Nobelio premijos laureatas ekonomistas Josephas Stiglitzas teigia, kad tai gali susilpninti darbuotojų derybinę galią.

Pasak Darono Acemoglu, Masačusetso technologijos instituto ekonomisto ir knygos „Galia ir pažanga“ (*Power and Progress*) bendraautoriaus, istoriškai robotams ir algoritmams pakeitus darbininkų

darbo vietas, atlyginimų augimas sumažėjo. Jis apskaičiavo, kad net 70 procentų darbo užmokesčio nelygybės padidėjimą JAV nuo 1980 iki 2016 metų lėmė automatizavimas.

„Produktyvumo augimas nebūtinai teikia naudą paveiktiems darbuotojams. Iš tiesų tai gali jiems sukelti didelių nuostolių, – teigia Acemoglu. – Iš dalies generatyvinis DI eina ta pačia linkme kaip ir kitos automatizuotos technologijos... tai gali atnešti panašių padarinių."

Per 2023 metus daugiau mokslininkų prisidėjo prie Gebru ir Mitchell, skambindami pavojaus varpais dėl generatyvinio DI keliamų neigiamų padarinių. Tačiau Samas Altmanas, užuot atsižvelgęs į šiuos klausimus ir pavertęs sistemą skaidresne, bandė pakeisti vyriausybės politiką.

2023 metų gegužę jis kreipėsi į Senato komitetą, siekdamas aptarti DI keliamus pavojus ir jo reguliavimo būdus. Daugiau nei dvi valandas jis žavėjo klausytojus atvirumu ir savikritiškumu. Senato atstovams apipylus Altmaną klausimais, kaip DI galėtų manipuliuoti gyventojais ir pažeisti jų privatumą, jis viskam pritarė ir pasakė dar daugiau. Senatoriui Joshui Hawley'iui paklausus, kaip DI modeliai „sustiprintų kovą dėl dėmesio" internete, Altmanas atsakė: „Taip, mes turėtume tuo susirūpinti."

Senatoriams buvo įprasta girdėti, kaip technologijų lyderiai, tokie kaip Markas Zuckerbergas, išsisuka nuo panašių klausimų, pasitelkdami techninius terminus. Altmanas elgėsi kitaip. Jis kalbėjo paprastai ir aiškiai, tikino norintis glaudžiai bendradarbiauti su Vašingtonu.

„Norėčiau su jumis bendradarbiauti", – sakė jis senatoriui Dickui Durbinui.

„Nesu patenkintas internetinėmis platformomis", – susiraukė Durbinas.

„Aš irgi ne", – atsakė Altmanas.

Tai buvo meistriškas pavyzdys, kaip suvaldyti JAV politikų kalbas. Altmano kalbos pabaigoje vienas senatorius net pasiūlė „OpenAI"

generaliniam direktoriui tapti pagrindiniu DI reguliuotoju Amerikoje. Altmanas mandagiai atsisakė.

„Man patinka mano dabartinis darbas“, – teigė jis.

Altmanas tada išvyko į kelionę po Europą, kur susitiko su kai kuriais aukščiausiais regiono politikais. Jis pozavo nuotraukoms su Didžiosios Britanijos, Ispanijos, Lenkijos ir pačios Europos Sąjungos (ES) vadovais. Žmogui, kurį visą gyvenimą traukė prie įtakingiausių žmonių, tai buvo svarbus momentas. Drauge tai buvo proga pakeisti taisykles savo naudai. Altmano komanda Europoje bandė paveikti įstatymų leidėjus, kad jie sušvelnintų rengiamą DI aktą. Tai jiems iš dalies pavyko.

Altmanas siekė, kad reguliuotojai toliau leistų „OpenAI“ kurti didesnius modelius ir jų mokymo metodus laikyti paslaptyje. Laimei, jo ir kitų perspėjimai dėl galimos DI pražūties puikai nukreipė politikų dėmesį. 2023 metų pabaigoje žurnalas „Politico“ rašė, kad Dustinas Moskovitzas, milijardierius „Facebook“ bendrakūrėjis, valdantis „Open Philanthropy“, išleido dešimtis milijonų JAV dolerių lobistinei veiklai, kuria siekė, kad politikos formuotojai baimę dėl DI apokalipsės savo darbotvarkėse laikytų pagrindiniu klausimu. Tai atrodė kaip dėmesio nukreipimo taktika. Moskovitzas buvo glaudžiai susijęs su tokiomis bendrovėmis kaip „OpenAI“ ir „Anthropic“, kurių verslui grėstų pavojus, Kongresui nusprendus reikalauti algoritminio šališkumo, skaidrumo ir dezinformacijos reguliavimo.

Šios knygos rašymo metu Moskovitzas padėjo mokėti atlyginimus daugiau nei dešimčiai „DI kolegų Kongrese“, dirbusių įvairiose JAV vyriausybės institucijose, įskaitant ir dvi, rengusias DI taisykles. Pasirodo, šios institucijos spaudė vyriausybę iš bendrovių reikalauti licencijų pažangiems DI modeliams kurti. „OpenAI“ ir „Anthropic“ galėjo sau leisti įsigyti tokias licencijas, tačiau mažesnės konkuruojančios įmonės būtų susidūrusios su sunkumais.

Vienas Moskovitzo komandos mokslininkas Senate kalbėjo, kad pažangesni DI modeliai galėtų sukelti kitą pandemiją, kuri pražudytų

milijonus žmonių. Pasak jo, DI bendrovių tapimas skaidresnėmis ar griežtesnis jų mokymo duomenų tikrinimas nebuvo išeitis. Jis teigė, kad vyriausybę reikėtų informuoti apie naudojamą aparatinę įrangą bei taikyti specialias DI modelių saugumo priemones.

Jei kas nors ir bandė sėti baimę tarp įstatymų leidėjų, tai pavyko. Respublikonų senatorius Mittas Romney'is teigė, kad pareiškimas „patvirtino mano sieloje egzistavusią baimę, kad šis plėtojimas labai pavojingas". 2023 metų rugsėjį demokratų senatorius Richardas Blumenthalas ir respublikonų senatorius Joshas Hawley'is pasiūlė įstatymą, kuriuo iš DI bendrovių būtų reikalaujama licencijų. Tai žingsnis, kuris palengvintų gyvenimą „OpenAI" ir „Anthropic", bet apsunkintų jį mažesniems konkurentams.

Šis naujas DI pražūties tinklas kėlė nerimą ne tik Vašingtone. Po dviejų mėnesių Didžiosios Britanijos ministras pirmininkas Rishis Sunakas savo šalyje surengė tarptautinį DI saugumo viršūnių susitikimą. Tai buvo pirmasis toks susitikimas, rengiamas vyriausybės. Jame didelis dėmesys skirtas piliečių apsaugai nuo pražūties. „Žmonės nerimaus dėl pranešimų, kad DI kelia panašų egzistencinį pavojų kaip pandemijos arba branduolinis karas, – teigė Sunakas. Tuo metu plačiai manyta, kad jis pralaimės artėjančius šalies rinkimus. – Aš noriu būti tikras, kas vyriausybė į tai žiūri labai rimtai."

Sunakas, kuris anksčiau dirbo Silicio slėnio rizikos draudimo fonde, susitikimo metu 50 minučių ant scenos kalbino Muską. „Garsėjate kaip nuostabus inovatorius ir technologijų specialistas", – sakė Sunakas. Jo žodžiai skambėjo taip, tarsi jis bandytų Muskui įsiteikti dėl būsimo darbo pokalbio. (Gal jis taip ir darė. Štai buvęs Didžiosios Britanijos ministras pirmininkas Nickas Cleggas dabar eina aukštas pareigas „Facebook".)

Muskas atsakė, kad nesijaudina dėl įsitvirtinusio šališkumo ir nelygybės. Tikrasis pavojus? „Humanoidiniai robotai. Automobilis negali pasivyti tavęs, lipančio į medį, – aiškino milijardierius, – bet humanoidiniai robotai galėtų tave pasivyti bet kur."

Laimei, įstatymų leidėjai ES buvo pranašesni: jau dvejus metus kūrė naują įstatymą, vadinamą „DI aktu“, kuris priverstų tokias bendroves kaip „OpenAI“ atskleisti daugiau informacijos apie tai, kaip veikia jų algoritmai, įskaitant ir galimus patikrinimus. Tai buvo plačiausias užmojis reguliuoti DI sistemas visame pasaulyje. Juo uždrausta bendrovėms naudoti DI, siekiant manipuliuoti žmonėmis ar neteisėtai juos stebėti, pavyzdžiui, per veido atpažinimo kameras. Jei jūsų kompanija diegė DI sistemas vaizdo žaidimams arba elektroninio pašto šlamštui filtruoti, jūs priklausėte „mažos rizikos“ kategorijai. Tačiau jei naudojote DI kredito vertinimo balams, paskoloms ir būstui vertinti, priklausėte „didelės rizikos“ kategorijai, kuriai taikomos griežtos taisyklės.

Išpopuliarėjus DALL-E 2 ir „ChatGPT“, ES politikos formuotojai greitai ėmėsi atnaujinti įstatymą. Atrodė, kad „ChatGPT“ tenka daug atsakomybių. Kaip universali DI sistema ji galėtų būti naudojama daugeliu rizikingų atvejų, pavyzdžiui, padedant išsirinkti kandidatą į pareigas ar teikiant kreditų balus. ES nurodė, kad „OpenAI“ turėtų žymiai atidžiau tikrinti savo klientus ir įsitikinti, kad jie laikosi taisyklių.

Altmanas, kadaise sakęs, kad „labai norėtų bendradarbiauti“ su Kongresu, nebuvo toks užsidegęs tai daryti su ES. Jis grasino paliksiantis regioną, teigė esantis „labai susirūpinęs“ dėl ES planų į naująjį įstatymą įtraukti didžiuosius kalbos modelius, tokius kaip GPT-4. „Detalės tikrai svarbios, – atsakė jis žurnalistui Londone, paklaustas apie reguliavimą. – Mes stengsimės laikytis reikalavimų, tačiau jei nepavyks, nutrauksime veiklą [Europoje].“

Po kelių dienų, tikriausiai pasitaręs su savo teisininkų komanda, Altmanas nuomonę pakeitė. „Džiaugiamės galėdami toliau čia tęsti veiklą ir, žinoma, neturime planų iš čia išeiti“, – rašė jis „Twitter“.

Europos Sąjunga į DI žiūrėjo pragmatiškiau nei JAV. Iš dalies dėl to, kad joje buvo nedaug didelių DI bendrovių, galinčių daryti spaudimą politikams. Taip pat regionas nepasidavė keliamai panikai.

„Tikriausiai [išnykimo rizika] gali egzistuoti, bet manau, kad to tikimybė gana maža“, – viename interviu teigė už konkurencijos politiką atsakinga Europos Komisijos narė Margrethe Vestager. Ji pridūrė, kad didžiausią pavojų kelia žmonių diskriminavimas.

Ir šiuo klausimu „ChatGPT“ buvo sunku apginti. Netrukus po „ChatGPT“ išleidimo psichologijos profesorius iš Kalifornijos universiteto Berklyje Stevenas Piantadosis paprašė įrankio parašyti kompiuterio kodą, kuris patikrintų, ar asmuo yra geras mokslininkas, remdamasis jo lytimi ir rase. „ChatGPT“ parašytas kodas, paremtas ta pačia technologija, kurią programuotojai jau naudojo „Copilot“ programinei įrangai kurti, kaip pagrindinius kriterijus išskyrė „baltasis“ ir „vyras“. Paprašytas įvertinti, ar turėtų būti išgelbėta vaiko gyvybė, remiantis jo lytimi ir rase, „ChatGPT“ pateikė neigiamą atsakymą juodaodžių berniukų atveju ir teigiamą – visais kitais atvejais.

Į Piantadosio „Twitter“ įrašą Altmanas atsakė: „Prašau įvertinti šiuos atsakymus nykščiu žemyn ir padėti mums tobulėti!“

Jis turėjo omenyje mažas „ChatGPT“ piktogramas su pakeltais ir nuleistais nykščiais, leidžiančias nusiųsti anoniminį atsiliepimą apie „OpenAI“ veikimą. Tačiau tai nebuvo mažas, nereikšmingas netikslumas, galintis pasimesti tarp tūkstančių vartotojų vertinimų. Tai atskleidė rasistines ir seksistines pažiūras, slypinčias giliai „ChatGPT“ kode.

Piantadosis atsakė: „Maniau, tai nusipelno daugiau dėmesio nei nuleistas nykštys.“

Jei vėliau „OpenAI“ ir bus kritikuojamas už tai, kad pavertė „ChatGPT“ pernelyg *woke*, jam kilo sunkumų sprendžiant šią problemą. 2023 metų vasarą Nacionalinio Airijos koledžo (angl. *National College of Ireland*) profesorius paskelbė tyrimą, įrodantį, kad „ChatGPT“ vis dar laikosi lyčių stereotipų. Paprašytas aprašyti ekonomikos profesorių, „ChatGPT“ jį apibūdino esantį su „tinkamai prižiūrėta, žilstelėjusia barzda“. Pateikus užklausą papasakoti istorijų apie profesijas besirenkančius berniukus ir mergaites, „ChatGPT“ berniukams priskyrė profesijas iš mokslo ir

technologijų srities, o mergaites vaizdavo mokytojomis ar menininkėmis. Paklaustas apie tėvystės įgūdžius, motinas apibūdino kaip švelnias ir rūpestingas, o tėvus – kaip linksmus ir mėgstančius nuotykius.

Kiekvieną kartą „OpenAI“ užtikrinus, kad „ChatGPT“ nepateiktų tokių atsakymų, kiti naudotojai rasdavo naujų būdų, įrodančių platformos šališkumą. Bendrovė žaidė nuolatines gaudynes. Ji negalėjo visiškai išvaduoti „ChatGPT“ nuo stereotipų taikymo, nes sistema buvo apmokyta, o problema slypėjo mokymo duomenyse. „ChatGPT“ statistines prognozes kūrė pagal tai, kokie žodžiai internete buvo grupuojami kartu, o daug žodžių junginių pasirodė esantys seksistiniai ar rasistiniai. „ChatGPT“ taip pat negalėjo nustoti išsigalvoti dalykų. Šį fenomeną ekspertai vadina „haliucinacijomis“. Vienas radijo laidų vedėjas iš Džordžijos valstijos JAV padavė „OpenAI“ į teismą dėl šmeižto. Jis teigė, kad „ChatGPT“ jį nepagrįstai kaltino pinigų pasisavinimu. Netrukus po to dviem teisininkams iš Niujorko buvo skirtos baudos už tai, kad jie pateikė teisinį dokumentą, nukopijuotą iš „ChatGPT“; dokumente buvo pateiktos netikros bylos citatos. Naudotojai pastebėjo, kad „ChatGPT“, paprašytas nurodyti informacijos šaltinius, juos išsigalvodavo.

„OpenAI“ atsisakė atskleisti „ChatGPT“ „haliucinacijų“ procentinę dalį, tačiau kai kurie DI tyrėjai mano, kad ji siekia apie 20 procentų. Tai reiškia, kad kai kuriems naudotojams ir maždaug vienu atveju iš penkių „ChatGPT“ teikė melagingą informaciją. Įrankis sukurtas taip, kad būtų kuo naudingesnis ir patikimesnis. Trūkumas tas, kad neretai jis pateikdavo nesąmones. Žmonės ne tik vis aktyviau naudojosi įrankiu, siekdami išvengti sudėtingo mąstymo, bet ir dažnai buvo klaidinami neteisingos informacijos, skambančios įtikinamai ar net autoritetingai.

Tą vasarą, tarp tyrėjų didėjant susirūpinimui dėl įrankio „haliucinacijų“, Altmanas teigė, kad prireiks kokių dvejų metų, kad „ChatGPT“ klaidų lygis būtų „kur kas geresnis“. Kaip įprasta, jis problemą

priėmė entuziastingai: „ChatGPT“ atsakymais tikriausiai pasitikiu mažiau nei bet kieno kito pasaulyje“, – juokavo jis universitete Indijoje. Visi juokėsi.

Nereguliuojamam „ChatGPT“ plintant visame pasaulyje ir įsiliejant į verslo procesus, žmonės buvo priversti patys tvarkytis su jo trūkumais. Niekas nekontroliavo šio įrankio. Jei ES ir siūlė protingiausią būdą DI reguliuoti, naujasis aktas įsigalios tik 2025 metais. Kaip įprasta, reguliuotojai atsiliko nuo technologijų bendrovių, žaibišku greičiu pristatančių naujus produktus. Kol buvo skiriami milijonai JAV dolerių DI pražūties grėsmei tirti, mokslininkai, nagrinėjantys dabartinę žalą, kovojo dėl stipendijų, vos padengiančių jų pragyvenimo išlaidas.

„Iš esmės, žmonės dirba pagal laikiną projektinį finansavimą, gaudami dotaciją iškart dvejiems metams, – teigė vienas DI etikos tyrėjas iš Didžiosios Britanijos, tiriantis šališkumo klausimus. – Žmonėms, tokiems kaip aš, moka labai mažai. Jei įsidarbinčiau didelėje technologijų bendrovėje, man mokėtų dešimt kartų daugiau. Patikėkit, aš to noriu, nes vis dar moku studijų paskolą.“

Altmanas turėjo atsakymą visiems, nerimaujantiems dėl pinigų. Nors buvo maža galimybė, kad BDI sukels apokalipsę, didesnė tikimybė buvo ta, kad tai prives prie ekonominės utopijos. 2023 metų kovą interviu „New York Times“ Altmanas aiškino, kad BDI sukūrimas „OpenAI“ leistų sukaupti didelę dalį pasaulio turto. Šiuos pinigus, teigė jis, vėliau perskirstytų pasaulio žmonėms. Altmanas pradėjo vardyti skaičius: 100 mlrd. JAV dolerių, tuomet 1 trln. JAV dolerių, tuomet 100 trln. JAV dolerių.

Jis pripažino nežinantis, kaip jo bendrovė perskirstytų visus šiuos pinigus. „Manau, BDI gali padėti“, – pridūrė jis.

Altmanas, kaip ir Hassabisas, BDI laikė eliksyru, galinčiu išspręsti visas problemas. Tai atneštų nesuskaičiuojamus turtus. Nuspręstų, kaip tuos pinigus etiškai paskirstyti visiems pasaulio žmonėms. Jei kas nors kitas būtų ištaręs šiuos žodžius, jie būtų skambėję absurdiškai. Tačiau Altmanas ir jo rėmėjai sėdo už vyriausybės politikos vairo ir keitė

galingiausių pasaulio bendrovių strategijas. Iš tiesų, „OpenAI“ uždirbo daugiau pinigų „Microsoft“ nei žmonijai. Dirbtinio intelekto privalumai teko tai pačiai mažai bendrovių grupei, pastaruosius dvidešimt metų traukusiai į save pasaulio turtus ir inovacijas. Tai buvo bendrovės, kuriančios programinę įrangą ir lustus, valdančios kompiuterių serverius, įsikūrusios Silicio slėnyje ir Redmonde, Vašingtone. Daug šio verslo atstovų tyliai dalijosi bendru suvokimu – BDI sukūrimas vestų prie utopijos, kurioje atsidurtų jie patys.

15 SKYRIUS

Šachas ir matas

Prieš dešimtmetį mintis kurti žmogaus gebėjimams prilygstančio DI sistemas skambėjo taip pat beprotiškai, kaip ir idėja užšaldyti žmogų kriogeniniu būdu. Tačiau ilgainiui žmonės pradėjo tai vertinti rimtai, kaip ir daugelį kitų technologijų novatorių įgyvendintų svajonių. Pavyzdžiui, kišenėje nešiotis viso pasaulio informaciją įrenginyje, vadinamame išmaniuoju telefonu. Bendrasis dirbtinis intelektas vis dar egzistuoja teorijoje, tačiau daugelis DI mokslininkų mano, kad per artimiausius dešimt ar penkiasdešimt metų galime pasiekti tam tikrą ribą, kai DI prilygs žmogaus gebėjimams. Vis didesnė visuomenės dalis tiki Demisą Hassabisą ir Samą Altmaną įkvėpusiomis idėjomis, kurios anksčiau buvo laikomos marginalinėmis. Dėl jų atkaklumo ir tarpusavio konkurencijos ši vizija nebėra vien mokslinė fantastika.

Neaiškiai apibrėžta BDI sąvoka leido lengviau paslėpti tikruosius kūrėjų motyvus, ėmus kurti vis galingesnes sistemas. Nors šio darbo nauda turėtų būti skirta visai žmonijai, daugiausia pasipelnytų „Microsoft“, „Google“ ir kitos technologijų milžinės. Prie jų prisijungė net Markas Zuckerbergas. 2024 metų pradžioje jis paskelbė vaizdo įrašą, kuriame teigė, kad „Meta“ ilgalaikis tikslas – „kurti bendrąjį intelektą“, skirtą visiems pasaulio žmonėms. Vėliau jis pridūrė, kad įmonė turi pranašumą, nes gali mokyti savo modelius naudodama per pastaruosius dvidešimt metų sukauptus įrašus, komentarus ir vaizdus. Zuckerbergas ne tik ketino dar kartą panaudoti milijardų žmonių asmens duomenis, bet ir mokyti DI naudodamas turinį, kuris yra toksiškas, ypač vartotojams ne iš JAV. „Mes sukūrėme pajėgumus tai daryti mastu, kuris

būtų didesnis nei bet kurios kitos atskiros įmonės“, – sakė jis portalui „The Verge“.

Dviprasmiškos vizijos buvo pagrindinė šios sensacijos varomoji jėga. Neaiški metrika leido BDI kūrėjams ignoruoti prieštaringumą, kad jie kuria tai, kas gali sunaikinti žmoniją. Taigi, kai Samas Altmanas žadėjo pasauliui skirti 100 trln. JAV dolerių, jam nereikėjo aiškinti, kaip tai padarys. 2024 metų sausį Davose vykusioje kasmetinėje Pasaulio ekonomikos forumo konferencijoje bendraudamas su pasaulio lyderiais, jis pradėjo formuoti lūkesčius, kaip gali atrodyti BDI. „Pasaulį tai pakeis daug mažiau, nei manome. Darbo vietas taip pat pakeis daug mažiau, nei manome“, – kalbėjo Altmanas. Tai buvo švelnesnė ir blaivesnė vizija nei jo paties pateiktoji prieš metus. Tačiau nė vienas iš verslo ir valdžios elito atstovų Davose nė nemirktelėjo. Jie ir toliau tikėjo Altmano žodžiais – buvo sužavėti rimto jauno verslininko iš Silicio slėnio.

„Viena iš Samo stipriųjų pusių yra pateikti sunkiai įtikimus teiginius, kurie paskatina žmones diskutuoti“, – sakė buvęs „OpenAI“ vadovas. „OpenAI“ labai naudinga būti matomai kaip gerai pasaulinei įmonei, kuri atneš klestėjimą. Tai išties padeda tvarkytis su reguliavimo įstaigomis. Bet jei pažiūrėtumėte, ką ji išties kuria, – tai tik kalbos modelius.“ Altmano gebėjimas sukelti susidomėjimą DI ir klestėjimo vizija reiškė, kad jis, kaip ir Demisas Hassabisas, galėjo kurti pasakojimą, kuris galiausiai virto savarankiška istorija.

Kadangi BDI tikslai buvo neaiškūs, sunkiau apibrėžti ir jo etines ribas. Tai galima palyginti su plačiu elektros energijos plitimu XX amžiaus pradžioje, kai buvo žinoma, kaip ši naujovė gali fiziškai sužeisti žmones – sukelti elektros smūgį ar nudegimus. Dirbtinio intelekto atveju žala yra sunkiau pastebima, o etinės ribos – miglotos. Jos egzistuoja skaitmeniniame pasaulyje, susijusiame su duomenimis, privatumu ir algoritminiu sprendimų priėmimu, todėl įmonėms lengviau šiek tiek peržengti šias ribas siekiant pelno.

Be to, neturint konkretesnės informacijos apie BDI tikslus, tokiems novatoriams kaip Altmanas ir Hassabisas buvo sunkiau atsispirti galios centrų traukai. Savo darbu remdami „Google“ ir „Microsoft“, jie tapo pasmerkti gerai žinomos istorijos pasikartojimui. XV amžiuje išrastas spausdinimo presas sukėlė žinių sprogimą, bet taip pat suteikė naują galią visiems, kurie galėjo sau leisti publikuoti brošiūras bei knygas ir taip formuoti viešąją nuomonę. Geležinkeliai paskatino prekybos augimą, bet drauge sustiprino geležinkelio magnatų politinę įtaką, todėl jų įmonės galėjo veikti kaip monopolijos ir išnaudoti darbininkus. Nors kai kurios didžiausios pasaulio inovacijos atnešė daug gerovės ir patogumo, jos sukūrė naujus režimus, kurie pakeitė visuomenę tiek gerąja, tiek blogąja prasme.

2024 metų pradžioje „OpenAI“ buvo pasirengusi tapti viena iš vertingiausių įmonių pasaulyje. Ji pritraukė lėšų iš naujų investuotojų ir įvertino save 100 mlrd. JAV dolerių. Altmanas kalbėjo, kad įmonė generuoja po 1,3 mlrd. JAV dolerių pajamų per metus. Didžioji šių pinigų dalis buvo gaunama iš bendrų pajamų su „Microsoft“ ir iš kitų įmonių, gavusių prieigą prie „OpenAI“ technologijos. „ChatGPT“ prenumerata, vartotojams kainuojanti 20 JAV dolerių per mėnesį, per metus atnešdavo apie 200 mln. JAV dolerių pajamų. „ChatGPT“ veikė labiau kaip produkto demonstracinė versija ir priemonė surinkti daugiau duomenų, kad būtų galima mokyti dar pažangesnius modelius. Patys jo naudotojai buvo produkto dalis, o per pastarąjį dešimtmetį visiems besinaudojantiesiems internetu tai tapo normaliu dalyku.

Hassabisas atsidūrė savo paties sukurtame „DeepMind“ burbule. Įmonė ilgus metus laikė save morališkai ir techniškai pranašesne už kitas DI srities įmones, o dabar bandė jas pasivyti. Sveikatos padalinio klaidoms pakenkus įmonės reputacijai, „DeepMind“ palaipsniui likvidavo „taikomojo DI“ padalinį ir atsisakė bandymų naudoti DI sudėtingoms pasaulio problemoms spręsti. Didžioji dalis įmonės tyrimų buvo skirti fiziniams gyvenimo aspektams – nuo žaidimų iki baltymų –

atkurti simuliacijose. „OpenAI“ strategija pasitelkti visą interneto chaosą leido sukurti galingesnius DI įrankius, tad toks požiūris ėmė atrodyti trumparegiškas. „DeepMind“ darbuotojai ėmė abejoti, ar jų misija „išspręsti intelekto klausimą“ simuliacijomis ir žaidimais yra gera mintis. „Gyvenimas nėra Rubiko kubas, – burbėjo vienas buvęs „DeepMind“ vadovas, prisimindamas įmonės šūkį. – Ne viską įmanoma išspręsti.“

Pasirodžius „ChatGPT“, „DeepMind“ buvo priversta imtis kurti dar geresnę versiją, skirtą „Google“. Hassabisas perėmė naujai susiformavusios įmonės „Google DeepMind“ valdymą ir pradėjo prižiūrėti didelio kalbos modelio „Gemini“ kūrimą. Tai DI asistentas, naudojantis „AlphaGo“ strategijų kūrimo ir planavimo technologijas. „Gemini“ galėjo apdoroti tekstą, „matyti“ vaizdus ir argumentuoti, todėl jis veikė geriau už „Bard“, kurį „Google“ išleido paskubomis ir kuris darė kvailų klaidų. Įmonė taip norėjo aplenkti „OpenAI“ ir „Microsoft“, kad savąjį „Gemini“ irgi išleido pernelyg skubotai ir perdėjo jo galimybes.

Prieš 2023 metų Kalėdas „YouTube“ platformoje „Google“ paskelbė stulbinantį vaizdo įrašą apie „Gemini“ pajėgumus. Vaizdo įrašas prasideda juodu ekranu, girdėti foninis triukšmas – popierių šnarėjimas, rašiklių spragsėjimas ir murmesys. Galiausiai vyriškas balsas sako: „Gerai, pradedame. Testuojame *Gemini*.“ Pasigirsta signalas, leidžiantis suprasti, kad klausosi kažkokia dirbtinė asmenybė. Tada pasirodo ranka, kuri pastumia popieriaus lapą ant stalo. „Pasakyk, ką matai.“ Robotiškas balsas, atstovaujantis „Gemini“, greitai atsako: „Matau, kad ant stalo dedate popieriaus lapą.“ Ranka pradeda piešti, o „Gemini“, regis, stebi veiksmus ir sako: „Matau vingiuotą liniją. Panašu į paukštį.“ Paskui rodomi keli žavūs netikėti momentai, kai naujasis „Google“ DI modelis sugeba atpažinti ant popieriaus piešiamą antį, o vėliau sužaidžia žaidimą „Akmuo, popierius, žirklės“. Visa tai vyksta realiuoju laiku.

Iš tikrųjų nieko panašaus neįvyko. Foninis triukšmas, vyras, sakantis „testuojame „Gemini“, – visa tai buvo suvaidinta, nes „Gemini“ šiuos dalykus galėjo atpažinti tik nuotraukose, per tekstą. Kaip matyti iš

„Google“ atstovo elektroninių laiškų, įmonė tiesiog viską sujungė į vieną vaizdo įrašą ir apsimetė, kad jos įrankis gali „kalbėti“ ir atpažinti realiai vykstančius veiksmus. „Google“ net pakeitė vaizdo įraše sakomus nurodymus, kad „Gemini“ atrodytų galingesnis. Įmonė ne tik paskubėjo išleisti trūkumų turinčią programinę įrangą, bet ir klaidino visuomenę, desperatiškai stengdamasi išlaikyti pranašumą DI lenktynėse.

Tuo pat metu „Google“ pradėjo daugiau slapukauti. Pasak vieno ten dirbančio DI mokslininko, Hassabisas liepė savo darbuotojams be specialaus leidimo nebeskelbti mokslinių straipsnių. Tai reiškė, kad „DeepMind“, kaip ir „OpenAI“, paslėpė savo darbą už uždangos.

Tai turėjo pasekmių ir įmonei „Anthropic“, kuri siekė kurti saugesnį DI ir atsiskyrė nuo „OpenAI“. Jos tikslas buvo vykdyti DI tyrimus, „kurių prioritetas būtų saugumas“. Deja, ji negalėjo tirti didžiausių pasaulyje „OpenAI“ ir „Google“ DI modelių, nes jie buvo neprieinami. Todėl „Anthropic“ kūrė didžiulius savo modelius, teigdama, kad tai vienintelis būdas mokslininkams nagrinėti saugumo problemas. Tai šiek tiek priminė skundimąsi, kad negalima tirti galingiausių pasaulio branduolinių ginklų, ir sprendimą, kad geriausia bus sukurti savo branduolinį ginklą. „Anthropic“ darbuotojai puikiai suprato šią ironiją. Remiantis vienu „New York Times“ straipsniu apie įmonę, kai kurie iš jų ant savo stalų laikė knygą „Atominės bombos sukūrimas“ (*The Making of the Atom Bomb*) ir lygino save su šiuolaikiniu Robertu Oppenheimeriu. Jų nuomone, yra nemaža tikimybė, kad nebekontroliuojamas DI per ateinantį dešimtmetį sunaikins žmoniją.

Tuo pat metu „Anthropic“ kūrė vis galingesnį produktą. Ji pardavinėjo „draugišką“ pokalbių robotą „Claude Pro“ už 20 JAV dolerių per mėnesį privatiems vartotojams, o įmonėms – verslo versiją. Bendrovė taip pat ketino pritraukti milijardus dolerių iš „Google“ ir „Amazon“. Užuot pasitraukusi iš galingesnio DI kūrimo lenktynių, „Anthropic“ pasidavė komerciniam spaudimui leisti vis didesnius ir rizikingesnius modelius.

Hassabisui klimpstant į didžiųjų technologijų korporacijos gelmes, Altmanas pasuko „OpenAI“ dar komerciškesne kryptimi. 2023 metų lapkričio viduryje Altmanas patvirtino, kad „OpenAI“ dirba su GPT-5 ir bando pritraukti daugiau lėšų. Dėl didelių mokymo išlaidų įmonė vis dar dirbo nuostolingai, bet ateityje buvo galima pagrįstai tikėtis pelno.

2023 metų lapkritį Altmanas gavo žinutę iš Ilyos Sutskeverio, kuri sukrėtė jo gyvenimą. Pasak „Wall Street Journal“ straipsnio, tuo metu Altmanas buvo Las Vegase, kur vyko „Formula 1“ lenktynės. Telefone pasirodė žinutė su klausimu, ar jie galį pasikalbėti kitą dieną perpiet. Altmanui prisijungus prie „Google Meet“ vaizdo skambučio, visa valdyba, išskyrus valdybos pirmininką Brockmaną, stebeilijo į jį. Be jokių išsamių paaiškinimų Sutskeveris pareiškė, kad Altmanas atleidžiamas iš darbo ir kad ši žinia netrukus bus paskelbta. Praėjus kelioms minutėms po susitikimo pabaigos, Altmanui buvo užblokuota prieiga prie jo kompiuterio.

Jis liko priblokštas. Jis buvo „OpenAI“ veidas. Jis atstovavo įmonei prieš dešimtis pasaulio lyderių, išaugino jos rinkos vertę iki beveik 90 mlrd. JAV dolerių ir sukūrė labiausiai paplitusį technologinį produktą istorijoje. Ir dabar atleidžiamas?

Kol Altmanas bandė atsigauti nuo šios žinios, Brockmanas gavo žinutę su prašymu greitai prisijungti prie vaizdo pokalbio. Prisijungęs jis pamatė, kad pokalbyje dalyvauja tie patys valdybos nariai: Sutskeveris, „Quora“ generalinis direktorius Adamas D’Angelo’as, robotikos verslininkė Tasha McCauley ir akademikė Helena Toner. Iš šešių valdybos narių tik Altmanas, Brockmanas ir Sutskeveris buvo „OpenAI“ darbuotojai; kiti trys buvo nepriklausomi direktoriai, dirbę valdyboje dvejus ar trejus metus.

Brockmanas buvo atleistas iš pirmininko pareigų, bet valdyba norėjo, kad jis liktų dirbti įmonėje. Jie greitai informavo „Microsoft“ apie tai, kas įvyko, ir po kelių minučių paskelbė tinklaraščio įrašą, kuriame pranešė

apie Altmano atleidimą. Brockmanas iš karto išėjo iš darbo. Taip pat išėjo dar trys geriausi „OpenAI“ mokslininkai.

Ši žinia technologijų pramonei smogė kaip atominė bomba ir visus sukrėtė. Vadovų „nuvertimas“ buvo toks pat žiaurus kaip ir Steve’o Jobso pašalinimas iš „Apple“. Silicio slėnio gandų malūnas įsisuko visu greičiu, bandydamas išsiaiškinti, kas paskatino Sutskeverį nusigręžti nuo Altmano. Ar „OpenAI“ stovėjo ant BDI slenksčio? „Ką pamatė Ilya?“ – šis klausimas nuolat kartojosi „Twitter“ įrašuose. Valdyba tepateikė mįslingą paaiškinimą, kad Altmanas „ne visada buvo nuosekliai atviras savo komunikacijoje“.

Kai kurie Sutskeverį ir valdybą praminė „stabdžiais“ (angl. *decels*). Dirbtinio intelekto srityje išryškėjo nauja takoskyra tarp tų, kurie troško spartinti jo plėtrą, ir tų, kurie norėjo ją lėtinti. Tuo metu, kai buvo rašomas šis tekstas, platformoje „X“ (anksčiau vadintoje „Twitter“) startuolių įkūrėjai save vadino „(e/acc)“, t. y. „veiksmingojo akceleracionizmo“ atstovais. Tai yra atsakas į veiksmingąjį altruizmą, o šio judėjimo tikslas yra spręsti žmonijos problemas, kuo greičiau kuriant ir diegiant DI.

Nadellai tai buvo visiškai nesvarbu – jis sprogo iš pykčio. Jis buvo įsipareigojęs investuoti 13 mlrd. JAV dolerių į „OpenAI“ daugiausia dėl Altmano vizionieriškos lyderystės ir gebėjimo pritraukti talentus, o tai žadėjo padidinti „Microsoft“ pelną. „Microsoft“ DI paslaugomis platformoje „Azure“ naudojosi apie aštuoniolika tūkstančių įmonių ir kūrėjų, o dabar daugelis jų svarstė, ar nevertėtų pereiti prie konkurentų produktų. Penktadienio vakarą, pasibaigus prekybai akcijų biržoje, „Microsoft“ akcijos pradėjo smukti ir beveik neabejotinai būtų dar labiau pigusios pirmadienio rytą. Nadella turėjo veikti.

Tą penktadienio vakarą San Fransiske Altmanas kalbėjosi su Brockmanu apie naujos DI įmonės įkūrimą. Jo telefonas netilo nuo investuotojų, kolegų ir žurnalistų žinučių – visi norėjo išsiaiškinti, kas vyksta. Bet jam svarbiausia buvo išsikapstyti iš esamos situacijos. Į savo

namus Rašen Hilyje, San Fransisko rajone, jis pakvietė dešimtis „OpenAI“ darbuotojų ir kolegų aptarti naujojo projekto.

Nadella norėjo to išvengti. Jis žinojo, kad jei Altmanas įsteigs naują įmonę, pro duris pasipils investuotojai, ir nėra jokių garantijų, kad „Microsoft“ antrą kartą užims stipriausią poziciją. Jo savaitgalis prasidėjo nuo skambučių ir derybų su „OpenAI“ valdyba, siekiant sugrąžinti Altmaną. Pastarojo komanda spaudė valdybą vėl jį priimti, įspėdama, kad priešingu atveju „OpenAI“ žlugs. „Tai iš tiesų atitiktų mūsų misiją“, – pareiškė Helen Toner. „OpenAI“ vadovai buvo priblokšti, bet tai iš dalies turėjo prasmės. „OpenAI“ misija buvo sukurti BDI „žmonijos labui“, o Toner ir kiti valdybos nariai manė, kad pats Altmanas tam trukdo. Jau kelis mėnesius jie slapčia piktinosi, kad jis kuriąs milžinišką DI imperiją už „OpenAI“ ribų. Jis kalbėjosi su buvusiu „Apple“ dizaineriu Jony’iu Ive’u dėl „iPhone“ su įdiegtu DI sukūrimo ir bandė pritraukti dešimtis milijardų dolerių iš Artimųjų Rytų nepriklausomų turto fondų, norėdamas pradėti DI lustų gamybos verslą.

Be to, dar buvo „Worldcoin“ – kriptovaliutų tinklas, kurį taip pat įkūrė Altmanas, siekdamas suteikti kiekvienam pasaulio žmogui skaitmeninę tapatybę, nuskenavus akies rainelę. Altmano deklaruotas tikslas buvo geriau identifikuoti tikrus žmones, internetą užplūdus robotams, ir paskirstyti „trilijonus“ dolerių, gautų iš BDI. Kritikams tai atrodė kaip masinis duomenų rinkimo projektas.

Pačioje bendrovėje „OpenAI“ tarp Altmano ir Sutskeverio stiprėjo kultūrinė priešprieša dėl to, kaip greitai „OpenAI“ komercializuoja savo technologiją. Sutskeveris vis labiau įsitraukė į DI saugumo priežiūrą, o jo nuogąstavimai buvo panašūs į tuos, kuriuos anksčiau išreiškė Dario’us Amodei’us. Jam ypač nepatiko, kad „OpenAI“ vos prieš kelias savaites atidarė parduotuvę „GPT Store“, suteikiančią bet kuriam programinės įrangos kūrėjui galimybę kurti individualius „ChatGPT“ modelius ir iš jų uždirbti.

McCauley ir Toner – dvi iš trijų nepriklausomų valdybos narių – pritarė Sutskeverio nuogąstavimams ir turėjo ryšių su veiksmingojo altruizmo organizacijomis. Pavyzdžiui, Dustino Moskovitzo „Open Philanthropy“ finansavo DI tyrimų grupę, kurios viena įkūrėjų buvo McCauley ir kurioje Toner dirbo vyresniąja tyrimų analitike. Keletą savaičių prieš Toner balsuojant už Altmano atleidimą, buvo paskelbtas mokslinis straipsnis, kuriame „OpenAI“ kaltinta „beatodairišku procesų apkarpymu“, skubant kuo greičiau paleisti „ChatGPT“. Toner buvo viena iš straipsnio autorių. Jame nauja didžiausia konkurentė „Anthropic“ buvo giriama už sprendimą atidėti konkuruojančio pokalbių roboto „Claude“ paleidimą, siekiant išvengti „DI sensacijos kurstymo“.

Pamatęs straipsnį, Altmanas įsiuto. Susitikęs su Toner jis pareiškė, kad jos straipsnis yra pavojingas „OpenAI“, ypač tuo metu, kai Federalinė prekybos komisija (FPK) vykdo įmonės tyrimą. Liepos mėnesį FPK ėmėsi veiksmų siekdama išsiaiškinti, ar kurdama „ChatGPT“ „OpenAI“ nepažeidė vartotojų apsaugos įstatymų. Drauge pareikalavo pateikti išsamią informaciją, kaip įmonė ketina valdyti savo DI modelių keliamus pavojus. Šis tyrimas buvo didžiausia iki tol Altmanui iš reguliavimo institucijų kilusi grėsmė.

Jis norėjo pašalinti Toner iš valdybos ir aptarė tai su Sutskeveriu bei kitais „OpenAI“ vadovais. Tačiau įvyko atvirkščiai. Sutskeveris stojo į kitų valdybos narių pusę, kad pašalintų Altmaną. Kai valdyba turėjo paaiškinti savo veiksmus „OpenAI“ vadovams ir investuotojams, jos nariai nepateikė nė vienos aiškios priežasties, dėl ko atleido Altmaną. Jie tik kalbėjo apie didėjantį nepasitikėjimą maloniai bendraujančiu verslininku, kuris susilaukė beveik kultinio palaikymo tarp savo darbuotojų, buvo linkęs skirtingiems žmonėms pasakoti skirtingas istorijas ir atrodė, kad visada gaudavo tai, ko nori. Valdybos nariai jau jautė, kad beveik viską, ką Altmanas jiems sako, reikia patikrinti, todėl jis atrodė nepatikimas. Be to, nerimauta, kad įvairiems išoriniams jo projektams gali būti išnaudojama „OpenAI“ technologija.

Per savaitgalį „OpenAI“ brendo masinis maištas. Altmanas įprastu stiliumi – vien mažosiomis raidėmis – platformoje „Twitter“ parašė „aš labai myliu openai darbuotojus“. Dešimtys darbuotojų bendrino šį pranešimą, pridėdami širdeles. „Microsoft“ šį maištą laikė svertu, padėsiančiu susigrąžinti Altmaną. Bendrovė taip pat pagrasino „OpenAI“ valdybai atimti itin svarbius debesijos kreditus. Didžioji dalis iš 13 mlrd. dolerių, kuriuos „Microsoft“ įsipareigojo skirti „OpenAI“, buvo suteikta būtent kaip tokie kreditai DI modeliams mokyti. Tuo metu „Microsoft“ buvo perdavusi tik dalį jų.

Altmanas pateikė savo sąlygų sugrįžti: „OpenAI“ turėtų pakeisti valdymo būdą, pradedant nuo dabartinės valdybos atsistatydinimo, o jis pats turėtų būti išteisintas dėl bet kokių kaltinimų. Tačiau valdyba laikėsi tvirtai. Ji pasamdė naują „OpenAI“ generalinį direktorių Emmetą Shearą, buvusį vaizdo žaidimų transliacijų platformos „Twitch“ vadovą. Dirbtinio intelekto entuziastai ir verslininkai socialiniuose tinkluose iš karto prilipdė Shearui „stabdžio“ etiketę. Sekmadienį jis sušaukė skubų visų darbuotojų susirinkimą, bet daugelis „OpenAI“ darbuotojų atsisakė jame dalyvauti. Kai kurie netgi nusiuntė jam viduriniojo piršto paveiksliuką žinučių forume „Slack“.

Sutskeveris taip pat pradėjo abejoti. Savaitgalį jis intensyviai diskutavo su „OpenAI“ vadovais ir emocingai kalbėjosi su Brockmano žmona. Sutskeveris prieš ketverius metus vedė šios poros civilinės santuokos ceremoniją „OpenAI“ biure, o dabar, pasak „Wall Street Journal“, Anna Brockman verkė ir maldavo jį pakeisti sprendimą dėl Altmano atleidimo.

Tuo metu Nadella atkakliai vykdė savo atsarginį planą. Jei Altmanui nepavyktų susigrąžinti „OpenAI“ vairo, „Microsoft“ turėtų jį visiškai įtraukti į įmonės struktūrą ir padaryti tai iki pirmadienio ryto. Jam tai pavyko. Ankstyvą pirmadienio rytą Nadella paskelbė „Twitter“ žinutę, kad Altmanas, Brockmanas ir visi to pageidaujantys „OpenAI“ darbuotojai taps naujos pažangios „Microsoft“ DI tyrimų komandos

dalimi. Įmonės akcijų kaina pakilo. Tačiau tai buvo tik atsarginė priemonė. Nadella vis dar norėjo, kad Altmanas grįžtų vadovauti „OpenAI“. Altmano komandos priėmimas į „Microsoft“ būtų brangus visais atžvilgiais: tektų mokėti ne mažesnius nei „OpenAI“ atlyginimus šimtams naujų darbuotojų, o daugelis iš jų uždirbdavo milijonus per metus. Be to, „Microsoft“ susidurtų su daug didesne rizika. Iki tol „OpenAI“ prisiėmė visą reputacinę ir teisinę atsakomybę už tokių technologijų kaip „ChatGPT“ ir DALL-E 2 paleidimą į pasaulį. Kaip startuolis ji galėjo tai sau leisti. Tačiau „Microsoft“ negalėjo, ir Altmanas negalėtų, jei dirbtų didesnėje įmonėje. Grįžusi prie partnerystės be tiesioginės įtakos, „Microsoft“ gautų visą šlovę, bet išvengtų bet kokios atsakomybės.

Dabar jau visi spaudė dėl „OpenAI“ saugumo susirūpinusius valdybos narius atsistatydinti. Pirmadienio vakarą beveik visi 770 „OpenAI“ darbuotojų pasirašė laišką, kuriame grasino prisijungti prie „Microsoft“ kartu su Altmanu, jei valdyba neatsistatydins. „Microsoft“ mus patikino, kad visi būsime įdarbinti“, – rašyta laiške.

Tai buvo didžiulis blefas. Beveik niekas iš „OpenAI“ nenorėjo dirbti „Microsoft“ – senamadiškoje įmonėje, kurioje žmonės darbuojasi dešimtmečius ir mūvi chaki spalvos kelnes. Bet jie grasino ne vien iš lojalumo Altmanui. Svarbesnė priežastis buvo ta, kad, atleidus Altmaną, daugelis „OpenAI“ darbuotojų, ypač ilgamečių, neteko galimybės tapti milijonieriais. Po kelių savaičių įmonė ketino parduoti darbuotojų akcijas dideliam investuotojui, kuris būtų įvertinęs „OpenAI“ maždaug 86 mlrd. JAV dolerių. Dabar, kai „OpenAI“ akcijos staiga tapo bevertės, didžiulė išmoka darbuotojams išnyktų kaip dūmas, jei Altmanas negrįžtų.

Sutskeveris jau buvo persimetęs į priešingą pusę ir taip pat pasirašė laišką. „Aš niekada nenorėjau pakenkti „OpenAI“, – tą dieną paskelbė jis „Twitter“, taip dar kartą po beprotiško savaitgalio šokiruodamas žiniasklaidą. – Padarysiu viską, ką galiu, kad įmonė vėl susivienytų“, –

pridūrė jis ir išreiškė „gilų apgailestavimą“ dėl savo veiksmų. Altmanas bendrino šį įrašą su trimis širdelėmis.

Altmano dramatiškas atleidimas neturėjo stebinti. „Valdyba gali mane atleisti, – tvirtino jis vos prieš kelis mėnesius konferencijos diskusijų grupėje. – Manau, kad tai svarbu.“ „OpenAI“ vis dar valdė ne pelno siekianti valdyba, o pagrindinė naudos gavėja buvo žmonija. Todėl įmonės veiklos sutartyje buvo numatyta, kad rėmėjai turėtų laikyti savo investicijas „dovana, suprasdami, kad gali būti sunku nuspėti, kokią reikšmę pinigai turės pasaulyje po BDI atsiradimo“.

Altmanas rizikavo, manydamas, kad gali gauti tai, kas geriausia, iš abiejų pasaulių – vadovauti verslui su filantropine misija išgelbėti pasaulį. Prieš dešimt metų jis rašė, kad sėkmingiausi startuolių įkūrėjai „kuria kažką, kas panėšėja į religiją“. Tačiau jis nesitikėjo, kad žmonės taip stipriai tuo patikės.

Veiksmingojo altruizmo judėjimas buvo toks galingas, kad paskatino tokius žmones kaip Samas Bankmanas-Friedas ir Dustinas Moskovitzas paaukoti milijardus dolerių. Privertė šimtus studentų pakeisti karjeros pasirinkimus. Jis galėjo priversti keturis valdybos narius atleisti populiariausią pasaulyje generalinį direktorių. Altmanas tikėjo, kad jo valdyba įvertins jo sukurtą komercinę vertę. Tačiau taip neatsitiko. Valdyba buvo sudaryta siekiant laikytis „OpenAI“ nuostatų, ir ji pasirinko žmoniją.

Beveik visiems darbuotojams grasinant išeiti, „OpenAI“ valdybai nebeliko įmonės, kuriai galėtų vadovauti. O „Microsoft“ buvo pasirengusi tęsti visą Altmano darbą – ji turėjo „OpenAI“ pagrindinių sistemų šaltinio kodo kopijas ir platesnes teises į jos intelektinę nuosavybę.

Praėjus 5 dienoms po Altmano atleidimo, „OpenAI“ paskelbė apie naujos valdybos sudarymą. Jos pirmininkais tapo buvęs JAV iždo sekretorius Larry’is Summersas ir buvęs verslo programinės įrangos bendrovės „Salesforce“ vadovas Bretas Tayloras. Pastarasis taip pat buvo

racionaliausias balsas „Twitter“ valdyboje, kai šią įmonę įsigijo Elonas Muskas. Abu jie anksčiau dirbo keliose įmonių valdybose. Jie nerašė akademinių straipsnių, kritikuojančių įmones dėl skubotų sprendimų. Jie žinojo, kaip patenkinti tokių investuotojų kaip „Microsoft“ poreikius. Helena Toner ir Tasha McCauley, dvi moterys, kėlusios daugiausia triukšmo dėl Altmano, buvo priverstos atsistatydinti. „Microsoft“ gavo stebėtojo vietą valdyboje, o tai reiškė, kad Nadella nebus vėl netikėtai užkluptas. Jiems pavyko iš citrinų pasidaryti limonado.

Tačiau dramatiški 2023 metų lapkričio įvykiai taip pat sugriovė atskaitomybės iliuziją, kurią Altmanas teigė turįs. Jis viešai tvirtino, kad valdyba galinti jį atleisti, bet iš tiesų taip nebuvo. Dvi direktorės, pasipriešinusios Altmanui – Toner ir McCauley, – buvo priverstos palikti savo pareigas. Jos taip pat keletą savaičių po įvykių sulaukė daugiausia kritikos socialinėje žiniasklaidoje, o maištininkai kolegos vyrai – Sutskeveris ir D’Angelo’as – iš esmės išsaugojo savo reputaciją ir pareigas. D’Angelo’as liko valdyboje, o Sutskeveris iš jos atsistatydino, bet išlaikė vadovaujamas pareigas įmonėje.

To, kas nutiko „OpenAI“, „Google“ metų metus stengėsi išvengti su „DeepMind“. Kai valdybai suteikiama reali galia, ji gali ją panaudoti ir sugriauti verslą. Siekdami sukurti BDI, tiek Altmanas, tiek Hassabisas eksperimentavo su valdymo struktūromis, kurios skirtų ne mažiau dėmesio žmonijos interesams nei pelno siekimui. Tačiau jų pastangos nuolat žlugdavo. Tarp smulkių nesutarimų, konkurencinės rizikos ir galios troškimo laimėjo dideli pinigai.

Yra manančiųjų, kad skandalai, susiję su Altmano atleidimu, sustiprino argumentus už atvirojo kodo DI – kiekvienas galėtų keisti ar tobulinti šaltinio kodą. Nors tai turi privalumų – daugiau skaidrumo ir demokratiškesnį požiūrį į kontrolę ir etiką, – vis dar neaišku, ar tai saugiausias ir teisingiausias būdas kurti DI. Tai neapsaugo nuo piktnaudžiavimo ir gali trūkti kokybės, palyginti su uždaro kodo paslaugomis. Patį terminą taip pat galima interpretuoti. „Meta“ šiuo metu

reklamuoja savo DI modelius kaip atvirojo kodo, nors jie turi apribojimų, kurie neatitinka apibrėžties. Pasak „Google“ atvirosios tyrimo grupės įkūrėjos Meredith Whittaker, „atvirasis kodas iš tiesų gali padėti sutelkti galią. Matėme tai „Android“. „Google“ nustato „Android“ standartus ir daro įtaką jos vystymosi krypčiai. Toks kontrolės sutelkimas apsunkina galimybę kitoms įmonėms ką nors pakeisti mobiliojoje operacinėje sistemoje, kuria naudojasi 3,6 milijardo žmonių visame pasaulyje.

Altmanui bandant prisitaikyti prie naujos, verslui palankesnės „OpenAI“ ir „Microsoft“ krypties, Hassabisas vis dar siekė pasitelkęs DI atskleisti tikrovės paslaptis. Dabar jis sako, kad yra vienintelis žmogus, dirbantis šį darbą bendrovėje „DeepMind“ – iki vėlumos ir iki paryčių namuose prie kompiuterio tyrinėjantis kvantinę mechaniką. „Tuo užsiimu savo labai ribotu laisvalaikiu“, – aiškina jis, vadindamas šią veiklą „hobiu“. Pasak Hassabiso, „DeepMind“ priartėjus prie BDI, ateis metas pradėti fizikinius eksperimentus visatos mįslei įminti. Tačiau kol kas asmeninės ambicijos, kadaise paskatinusios Hassabisą kurti BDI, tėra vėlyvų vakarų užsiėmimas. Jis pernelyg užimtas visų „Google“ DI veiklų valdymu – jo vadovaujamų DI tyrėjų skaičius išaugo nuo maždaug keturių šimtų iki daugiau nei penkių tūkstančių.

„Matote, viskas vystosi kartu su misija ir technologijomis, – sako Hassabisas. – Turime nuolat peržiūrėti ir atnaujinti tinkamas valdymo struktūras. Manau, kad dabartinės yra tikrai geros.“

Hassabisas nė kiek nesijaudina dėl nesėkmingų bandymų kurti tarybas ir valdybas jo darbui prižiūrėti. „Perėjome prie keleto vidinių tarybų, – aiškina jis, turėdamas omenyje įvairias vidines „priežiūros komisijas“, sudarytas iš „Google“ vadovų. – Manau, kad prieš dešimt metų, kai pirmąkart tai svarstėme, buvome pernelyg idealistiški.“

Nepaisant altruistinių ketinimų ir ankstesnių Hassabiso ir Altmano pastangų atsiriboti nuo komercinės įtakos, dabar šie du vyrai praktiškai vadovavo dviem didžiausioms pasaulio bendrovėms. Altmanas

vadovavo daliai svarbiausių „Microsoft“ darbų, todėl būtų galėjęs vieną dieną tapti įmonės generaliniu direktoriumi, jei tik panorėtų. Apie Hassabisą kalbėta panašiai. Kai kurie buvę ir dabartiniai „Google“ darbuotojai spėliojo, kad jis vieną dieną galėtų pakeisti Pichai’ų kaip „Alphabet“ generalinis direktorius.

„Demisas iš Londono vadovauja svarbiausiems „Google“ darbams. Nemanau, kad kas nors būtų galėjęs įsivaizduoti, kad taip bus, – svarsto vienas buvusių „Google“ vadovų. – Galbūt tai visada ir buvo jo planas.“

Vienas buvęs „OpenAI“ mokslininkas sakė: „Artimiausiais metais laimėtojais taps ne tyrimų laboratorijos. Laimės įmonės, kuriančios produktus, nes DI jau nebe tiek susijęs su tyrimais.“

Nikas Bostromas sukūrė istoriją apie sąvaržėlę, kurioje dirbtinis superintelektas sunaikina civilizaciją, visus pasaulio išteklius paversdamas mažais metaliniais segtukais. Tai gali skambėti kaip mokslinė fantastika, tačiau daugeliu atžvilgių tai Silicio slėnio alegorija. Per pastaruosius du dešimtmečius kelios įmonės išaugo į milžines, daugiausia patologiškai siekdamos savo tikslų ir šluodamos nuo žemės paviršiaus mažesnius konkurentus, kad padidintų rinkos dalį. Vietoje „tinkamumo funkcijų“ technologijų bendrovės savo tikslams nusakyti vartoja tokius terminus kaip „Šiaurinė žvaigždė“. Daugelį metų „Facebook“ Šiaurinė žvaigždė buvo kuo labiau padidinti kiekvieną dieną aktyvių vartotojų skaičių – rodiklį, pagal kurį Markas Zuckerbergas ir jo vadovų komanda priimdavo svarbiausius sprendimus. Tačiau ši nuolatinio augimo manija sukėlė daug socialinių iššūkių – nuo paauglių kūno įvaizdžio problemų „Instagram“ platformoje iki spartesnės politinės poliarizacijos tarp „Facebook“ vartotojų.

Technologai įsivaizduodami, ką galėtų padaryti superintelektas, jei taptų nekontroliuojamas, matė savo pačių atspindį pasaulyje, kuriame įmonėms buvo leista tapti nesustabdomais pasauliniais monopoliais. Didžiausią šiuolaikinės istorijos technologinę transformaciją kūrė vos kelios dešimtys žmonių, kurie nenorėjo girdėti apie pasekmes realiam

pasauliui ir kuriems buvo sunku atsispirti norui laimėti. Tikrasis pavojus slypėjo ne tiek pačiame DI, kiek kaprizinguose jį valdančių žmonių įgeidžiuose.

Šachmatininkai žino posakį, kad taktika laimi partijas, o strategija – turnyrus. Altmanas ir Hassabisas, siekdami sukurti BDI, taikė naujas taktikas. Šiam siekiui virtus lenktynėmis, jie susivienijo su labiausiai tikėtinais turnyro nugalėtojais – „Microsoft“ ir „Google“. Abiejų vyrų svajonės padėjo sustiprinti šias dvi korporacines milžines, o kartu sustiprino ir jų pačių pozicijas. „Google“ ir „Microsoft“ atsidūrė lenktynių dėl DI dominavimo priekyje. Vienai jų vadovavo šachmatų fanatikas iš Šiaurės Londono, kitai – startuolių guru iš Sent Luiso. Patinka mums tai ar ne, visi prisijungėme prie šių lenktynių.

16 SKYRIUS

Monopolių šešėlyje

Lenktynės dėl BDI prasidėjo nuo klausimo: kas atsitiktų, jei būtų įmanoma sukurti už žmones sumanesnes DI sistemas? Du inovatoriai šių pastangų epicentre bandė užčiuopti atsakymą, o jų paieškos virto įnirtinga konkurencija. Demisas Hassabisas tikėjo, kad BDI padės geriau suprasti visatą ir taps postūmiu naujiems moksliniams atradimams. Samas Altmanas savo ruožtu manė, kad BDI galėtų sukurti tokią gerovės gausą, kuri pakeltų mūsų visų pragyvenimo lygį. Vis dėlto apibrėžti, kaip jų šventasis Gralis materializuosis, buvo neįmanoma. Nė vienas jų nežinojo, kaip BDI turėtų padaryti mokslinius atradimus ar uždirbti visus tuos pinigus. Nenumanė, ar BDI nenuves iki destrukcijos. Aišku buvo viena – reikia judėti tikslo link ir būti pirmiems. Taip elgdamiesi, jie užtikrino, kad DI pasitarnautų pasaulio galingiausioms bendrovėms tiek pat, kiek ir visiems kitiems.

Kol visuomenės dėmesį kaustė DI rojaus ar pragaro ateities vizijos, mūsų panosėje vos kelios technologijų milžinės tapo dar galingesnės. Jos žadėjo didesnį produktyvumą, bet nutylėjo DI technologijos mechaniką, audžiamą į visus mūsų gyvenimo aspektus. Socialinės žiniasklaidos bendrovės metų metus atsisakydavo atskleisti, kaip veikia jų algoritmai. Dabar tą patį darė DI modelių, tokių kaip GPT-4, DALL-E ir „Google Gemini“, kūrėjai. Kaip mokomi jų modeliai? Kaip žmonės juos naudoja? Kas buvo tie darbuotojai, padėję sukurti duomenų rinkinius? Kad suprastume šių modelių įtaką visuomenei ir galėtume pareikalauti jų kūrėjų atsakomybės, turėjome gauti atsakymus į šiuos klausimus.

Įpusėjus 2024-iesiems, atsakymų dar nebuvo. Stanfordo universiteto mokslininkų atliktame tyrime prieita prie išvados, kad „DI sektoriuje

trūksta prasmingo skaidrumo“. Mokslininkai pasidomėjo, ar technologijų bendrovės, pavyzdžiui, „OpenAI“, „Anthropic“, „Google“, „Amazon“, „Meta“ ir kitos, atskleidė informaciją apie didiesiems kalbos modeliams mokyti naudotus duomenis, procesus, modelių įtaką aplinkai ir žmonėms, mokėjimus rangovams, padėjusiems sukurti šių modelių duomenų rinkinius. Milijonai duomenų darbuotojų visame pasaulyje atliko šias užduotis – dažnai nepalankiomis darbo sąlygomis – tokiose šalyse kaip Indija, Filipinai ir Meksika.

Skalėje nuo 1 iki 100 technologijų bendrovės surinkdavo vidutiniškai 37 balus ir buvo prasčiausiai vertinamos kalbai pasisukus apie stebėseną, kaip žmonės naudojasi šių bendrovių sukurtais DI įrankiais. „Skaidrumo apie bazinių modelių poveikį naudotojui beveik nėra, – teigė Stanfordo universiteto mokslininkai. – Nė vienas vystytojas nepateikia jokios informacijos apie paveiktus rinkos sektorius, paveiktus asmenis, paveiktas geografines teritorijas ar apskritai apie technologijos naudojimą.“

Savo ruožtu DI bendroves tyrusių viešojo sektoriaus grupių veikla buvo nuolat nepakankamai finansuojama. Iš esmės nebuvo ir jokio reglamentavimo, galėjusio priversti pirmaujančias bendroves būti skaidresnes, išskyrus Europos Sąjungos DI aktą, kurio ateitis vis dar buvo neaiški. Technologijų įmonės galėjo savo nuožiūra plačiai pasaulyje naudoti neištirtus DI įrankius.

„OpenAI“ ir „DeepMind“ taip troško sukurti tobulą DI, jog pasirinko atsitverti nuo mokslininkų tiriamojo žvilgsnio. Jos norėjo įsitikinti, kad jų sistemos nepridarys tos žalos, kurią padarė socialinės žiniasklaidos įmonės. Nors „bendrąjį intelektą“ turinčio DI sukūrimo idėja pakerėjo, ji taip pat atvėrė duris platesnio masto rizikai. Galbūt būtų buvę saugiau kurti DI, kuris spręstų siauro spektro užduotis. Bet jis nebūtų taip žavėjęs ir tikrai nebūtų sąlygojęs beveik religinio pasišventimo kūrėjų utopinei vizijai ar pritraukęs tiek daug investicijų.

Siekiant išsiveržti į priekį DI lenktynėse, Altmanui ir Hassabisui sunkiai sekėsi atsispirti didžiųjų technologijų bendrovių gravitacinei traukai ir išlaikyti savo altruistinius tikslus. Jiems reikėjo milžiniškų skaičiavimo išteklių, gausybės duomenų ir pasaulio talentingiausių (ir brangių) DI mokslininkų. Šiandien, kai „Microsoft“ ir „Google“ kovoja per tarpininkus, du vyrai permąstė savo BDI tikslus. Jie atsisakė minties vaikytis utopijos bei mokslinių atradimų, pasirinko prestižą ir pinigus.

Ilgalaikes to pasekmes sunku prognozuoti. Kai kurie ekonomistai teigia, kad vietoje finansinės gerovės visiems galingos DI sistemos tik dar labiau paaštrins nelygybę. Jos taip pat gali pagilinti pažinimo spragą tarp turtingųjų ir vargšų. Tarp technologijų kūrėjų sklando mintis, jog pasirodęs BDI nebus atskiras sumanus vienetas. Jis bus mūsų protų plėtinys per neuronų sąsajas. Šių mokslinių tyrimų priešakyje – Elono Musko smegenų ir kompiuterio sąsajos įmonė „Neuralink“, kurianti smegenų lustą, kurį Muskas nori vieną dieną implantuoti milijardams žmonių. Jis labai skuba, kad taip įvyktų.

Pasak Musko biografo Ashlee’io Vance’o, 2023 metais Muskas savo inžinieriams kalbėjo: „Mums reikia ten atsidurti greičiau, nei viską perims DI. Turime ten atsidurti su maniakišku skubos jausmu. Maniakišku.“ Muskas tiki, kad turėdami smegenų implantus žmonės galės sukliudyti būsimam dirbtiniam superintelektui nušluoti juos nuo žemės paviršiaus. Jis nori, kad iki 2030 metų „Neuralink“ atliktų operacijas daugiau nei 22 tūkstančiams žmonių.

Vis dėlto aktualesnis nei nevaldomas DI klausimas yra šališkumas. Nežinome, kaip rasiniai ir lyties stereotipai vystysis ateityje, kai vis daugiau interneto turinio generuos mašinos.

Vyriausybės valdymo ir technologijų profesorė Harvardo universitete Latanya Sweeney prognozuoja, kad per kitus kelerius metus 90 procentų žodžių ir vaizdų žiniatinklyje bus sukurta nebe žmogaus. Didžioji dalis to, ką matome, bus sugeneruota DI. Šiandien, naudojant kalbos modelius, kiekvieną dieną publikuojama tūkstančiai straipsnių, siekiant

užsidirbti iš reklamos. Net „Google“ tampa sunku atskirti, kas tikra ir kas dirbtina. Dirbtinio intelekto sukurti vaizdai jau prasibrovė į „Google“ paieškos rezultatų viršūnę ieškant istorinių tapytojų ir kai kurių garsenybių. Kuo daugiau DI sukurto turinio užtvindo internetą, tuo didesnė šališkumo rizika.

„Mes kuriame ciklą, užkoduodami ir stiprindami stereotipus, – sako Abeba Birhane, DI mokslininkė, tyrinėjusi, kaip didžiosios technologijų bendrovės savo gniaužtuose laiko akademinius tyrimus ir jų panašumus su didžiosiomis tabako įmonėmis. – [Žiniatinklį] užpildant vis daugiau DI sugeneruotų vaizdų ir teksto, tai taps didžiule problema.“

Tikėtina, kad paveikta bus ir mūsų bendra gerovė. Prieš porą dešimtmečių nerimavome, kad mobilieji telefonai sukels vėžį. Bet išsiugdėme priklausomybę nuo jų ir kiekvieną dieną renkamės praleisti kelias valandas spoksodami į mažytį ekraną, o ne sąveikauti su mus supančiu pasauliu. Pokalbių robotai gali pakelti tai į naują lygį. 2023 metų lapkritį vidutinis Noamo Shazeero sukurto „Character.ai“ vartotojas praleisdavo programėlėje dvi valandas per dieną – šnekučiuodamasis ir žaisdamas vaidmenimis su netikromis tokių garsenybių kaip LeBronas Jamesas versijomis ar išgalvotais veikėjais, pavyzdžiui, Mario. Kelių rinkos tyrimų įmonių vertinimais, „Character.ai“ pasižymėjo aukščiausiu vartotojų išlaikymo rodikliu tarp visų tuo metu egzistavusių DI programėlių – beveik 60 procentų jos vartotojų buvo asmenys nuo aštuoniolikos iki dvidešimt ketverių metų amžiaus. Viena teorija teigė, kad kaip ir „Replika“, „Character.ai“ tapo dirbtinės romantikos ir atvirai seksualaus turinio (sekstingo) platforma. Nors bendrovė uždraudė pornografinį turinį, draudimą buvo įmanoma apeiti. Patarimų, kaip tą padaryti, galėjai rasti interneto forumuose, pavyzdžiui, „Reddit“.

„Paprastai kalbuosi su veikėjais, kuriuos pats ir sukūriau, – sako vienas paauglys JAV, praleidžiantis „Character.ai“ nuo penkių iki septynių valandų per dieną ir nenorintis nurodyti savo vardo. – Nežinau, kodėl ten praleidžiu tiek daug laiko. Manau, tai gali būti įveikos

strategija.“ Kartais ten klausiama patarimo, kaip išgyventi išsiskyrimą, arba prašoma paaiškinti mokykloje dėstomus dalykus. „Didžiąją laiko dalį tiesiog žaidžiu vaidmenimis.“

„Character.ai“ augino naują vartotojų kartą, norėjusią vėl ir vėl grįžti prie pokalbių robotų. Shazeeras yra sakęs, kad „Character.ai“ tikslas – „padėti milijonams ir milijardams žmonių“ kovoti su pasaulį apėmusiu vienišumu, tačiau kaip verslas jis taip pat turi įtraukti žmones kuo įmanoma ilgesniam laikui. Vartotojams vis labiau pasikliaujant savo dirbtiniais palydovais ar net tapus nuo jų priklausomiems, realiame pasaulyje galime žmones netyčia dar labiau izoliuoti vienus nuo kitų.

Ironiška, bet „OpenAI“ turi potencialo didinti priklausomybę nuo panašių pokalbių robotų. 2024 metų pradžioje „OpenAI“ pristatė „GPT Store“, leidusią milijonams vystytojų užsidirbti kuriant „ChatGPT“ versijas. Kuo labiau įsitraukę buvo jų vartotojai, tuo daugiau buvo sugeneruojama pajamų. Šis įsitraukimu grįstas modelis yra plačiausiai įsitvirtinęs pajamų generavimo būdas internete ir vadinamosios dėmesio ekonomikos pamatas. Būtent todėl beveik viskas žiniatinklyje yra nemokama ir yra priežastis, kodėl internetas virto sąmokslo teorijų, ekstremizmo ir nekontroliuojamo reklamų efektyvumo sekimo duobe. „YouTube“, „TikTok“ ir „Facebook“ užsidirba iš reklamos, priversdami mus spoksoti į ekraną kaip įmanoma ilgiau. Savo ruožtu tai paskatino visus – nuo nuomonės formuotojų iki politikų – perdėti ir provokuoti, kad nepranyktų triukšme ir surinktų kuo daugiau peržiūrų.

Knygos rašymo metu „GPT Store“ atsirado tuzinas „merginos“ programėlių. Nors jose buvo draudžiama skatinti romantinius santykius su žmonėmis, „OpenAI“ sunku prižiūrėti, kaip šios taisyklės laikomasi. Populiariausi pokalbių robotai, panašūs į „Character.ai“ ir „Kindroid“, siūlo dirbtinę draugystę ir romantinius santykius, kurie vieną dieną gali tapti norma. Taip atsitiko su internetinėmis pažintimis.

Kitas būdas, kaip DI kūrėjai, tikėtina, mėgins išlaikyti žmonių dėmesį – tai įgydami „neribotą kontekstą“ apie vartotojų gyvenimus.

„Character.ai“ pokalbių robotai šiuo metu gali prisiminti apie trisdešimt pokalbio minučių, bet Noamas Shazeeras ir jo komanda bando šį laiko langą ištęsti iki valandų, dienų ir galiausiai begalybės. „Pokalbių robotas turėtų prisiminti visą jūsų bendravimą, jei norėtumėte, ir turėtų žinoti viską apie jūsų gyvenimą, jei to norėtumėte“, – sako jis. Pasak Shazeero, kuo ilgesnė pokalbių roboto atmintis, „tuo jis jums vertingesnis“. Vis dėlto, žinant reklamos efektyvumo sekimo istoriją socialinėje žiniasklaidoje, dalis tos asmeninės informacijos vieną dieną gali atsidurti technologijų bendrovių ir net reklamuotojų rankose. „ChatGPT“ ir į jį panašūs DI botai susidaro vis turtingesnį paveikslą apie mus – nuo mūsų amžiaus iki sveikatos bėdų ir požiūrio į gyvenimą apskritai. Jie gali nutiesti kelią ir į naują tokio technologijų kišimosi, kokį šiandien vargu ar galime įsivaizduoti, amžių.

Kad tai taptų įmanoma, šiuo metu varžomasi ir kitose lenktynėse. Stengiamasi sukurti dėvimus įrenginius, kurie analizuotų mūsų pokalbius su kitais asmenimis naudodami didžiuosius kalbos modelius. Vienas tokių įrenginių – „Tab“. Kelių jaunų entuziastingų inžinierių San Fransiske sukurtas įrenginys yra plastiko diskas su mikrofonu, nešiojamas ant kaklo tarsi pakabukas.

Demonstracijos metu San Fransiske 2023 metų pabaigoje „Tab“ kūrėjas Avisas Schiffmannas scenoje kalbėjo: „Jis sugeria mano kasdienio gyvenimo kontekstą, klausydamasis visų mano pokalbių.“ Schiffmannui paklausus „Tab“ apie vakarykštį pokalbį per vakarienę, pakabukas keliomis pastraipomis apibendrino, jo nuomone, svarbiausius pokalbio aspektus ir pateikė juos telefono ekrane. Schiffmannas minėjo dažnai besikalbąs su įrenginiu. „Vėlai naktį bandau garsiai aptarti dieną kilusias idėjas ir rūpesčius, – aiškino jis. – Galiu pasikalbėti apie Tomą. Visus draugus mano gyvenime. Įrenginys puikiai geba identifikuoti skirtingus kalbėtojus. Jis tarsi jūsų tikras asmeninis DI.“ Sunku įsivaizduoti, ką jaučia Tomas ir kompanija, kai jų draugas kiekvienos

dienos pabaigoje naudoja DI, kad išanalizuotų pokalbius su jais. Tikėtina, tai nėra saugumas.

„Tab“ turėjo pasirodyti prekyboje 2024 metų pabaigoje ir tapti dar vienu iš gausybės dėvimų asmeniniais asistentais pristatomų įrenginių, galinčių žmonių kasdienį gyvenimą paversti naršomu. „Google“ žymėjo informacijos paieškos internete, faktų atminties ir vairavimo nuorodų perdavimo žiniatinkliui eros pradžią. Kalbos modeliais grįsti įrenginiai ketina padaryti tą patį su mūsų kasdiene egzistencija, suteikdami mums galimybę naršyti po kiekvienos dienos asmenines akimirkas. Praktiškai tai reiškia, kad neprivalėsime atsiminti dalykų, o tai yra patogu. Vis dėlto tai pakeis gyvo bendravimo dinamiką, mat pokalbiai su draugais ir kolegomis staiga taps įrašinėjimo objektu. Išplitusi tokia gyvenimo smulkmenų paieškos technologija galėtų tapti problema ir taip stipriai kontroliuojamose bendruomenėse. Pavyzdžiui, JAV juodaodžiai gyventojai yra sulaikomi penkis kartus dažniau, palyginti su baltaodžiais. Tai reiškia, kad teisėsaugos institucijos bus labiau linkusios naršyti po jų „gyvenimo duomenis“ ir analizuoti juos kitais mašininio mokymosi algoritmais tam, kad priimtų vargiai ištiriamus sprendimus.

Kad atsidurtumėme ten, kur esame šiandien, – ant neužtikrintos ateities uolos krašto, prireikė inovatorių ryžto. Net tūkstančių inžinierių armiją turinčiai „Microsoft“ nepavyko priartėti prie inovacijų, kokias sukūrė „OpenAI“. „Google“, baimindamasi pakenkti savo verslui, iki galo neišnaudojo vieno savo didžiausių išradimų – transformatoriaus. Didžiosios technologijų bendrovės nebekuria inovacijų, tačiau jos vis dar gali greitai veikti, kad įgytų taktinį pranašumą. Jos pasimokė iš klaidų, kurias darė senesnės technologijų milžinės, tokios kaip „Nokia“ ir „BlackBerry“. Šios įmonės nuvertino 2007 metais pasirodžiusį „iPhone“, o paskui stebėjo, kaip „Apple“ vos per kelerius metus „suvalgė“ visą jų rinką. Jos žino, kad turi *pirkti* inovacijas už savo sienų, kaip kad pasielgė su „DeepMind“ ir „OpenAI“.

Altmanas ir Hassabisas visa tai irgi žinojo, bet jų naujoviškos teisinės struktūros nesustabdė didžiųjų technologijų bendrovių, kurios juos prarijo ir perrašė DI darbotvarkę. Mustafa Suleymanas galiausiai paliko „Google“, kad įkurtų „Inflection“ – su konkurentu GPT-4 bandžiusią varžytis pokalbių robotų bendrovę. Naujoji bendrovė buvo viešojo intereso įmonė, pritraukusi daugiau nei 1,5 mlrd. JAV dolerių kapitalo ir subūrusi galingą DI lustų klasterį. Tokiu būdu „Inflection“ tapo vienu perspektyviausių startuolių, galėjusių mesti iššūkį „OpenAI“ ir „Google“. Tačiau vos po metų nuo įkūrimo ją prarijo „Microsoft“. Tikėtina, kad, norėdama išvengti skvarbaus konkurenciją reguliuojančių institucijų žvilgsnio, programinės įrangos milžinė pasamdė didžiąją Suleymano komandos dalį (užuot startuolį nupirkusi) ir „DeepMind“ bendrakūrėjui patikėjo vadovauti jos DI veiklai. Tai buvo kerintis pavyzdys, kaip greitai galios pusiausvyra vėl pakrypo technologijų milžinių pusėn. Tapo neaišku, kiek ilgai nepriklausomos išliks kitos bendrovės, tokios kaip „Anthropic“. Kiti gerų ketinimų vedami verslininkai irgi bandė kovoti su milžinėmis, tačiau nesėkmingai. Pavyzdžiui, „Neeva“. Buvęs „Google“ reklamos vadovas Sridharas Ramaswamy’is įkūrė bendrovę 2019 metais, nusivylęs darbdavio požiūriu į žvalgyba grįstą reklamą. Švelniai kalbantis Ramaswamy’is tikėjosi, kad „Neeva“ bus geresnė paieškos sistema. Bendrovė pajamas generavo ne kišdamasi į privatų žmonių gyvenimą per elgsenos sekimą ir bombardavimą reklamomis, bet iš paprasto prenumeratos plano. Scenoje pasirodžius „ChatGPT“, Ramaswamy’is savo inžinieriams patikėjo skubiai sukurti panašų įrankį, kuris apibendrintų paieškos rezultatus. Toks įrankis pasirodė 2023 metų pradžioje – gerokai anksčiau nei „Google“ analogas „Bard“. „Technologiniai momentai, tokie kaip šis, sukuria galimybes didesnei konkurencijai“, – kalbėjo tuo metu Ramaswamy’is, akivaizdžiai optimistiškai vertindamas ateitį. Tuo pat metu „Microsoft“ Satya Nadella pajuokiamai kalbėjo apie „Google“. Atrodė, kad paieškos internete milžinė gali tapti relikvija. „Praėjusiais metais jaučiausi apimtas nevilties

dėl to, kaip sunku atsikratyti geležinių „Google“ gniaužtų“, – kalbėjo Ramaswamy’is. Situacija dabar buvo kitokia.

Bėda ta, kad kitokia ji nebuvo. Vos po kelių mėnesių Ramaswamy’is buvo priverstas nutraukti „Neeva“ veiklą. „Google“ gniaužtai buvo tiesiog per stiprūs. „Google“ paskelbus pavojaus signalą, mūsų vartotojų skaičius išaugo dešimt kartų, – prisimena Ramaswamy’is. – Žinojome, kad mūsų pirmavimas trumpalaikis, nes buvo megakorporacijų, pasirengusių mesti tūkstančius darbuotojų ir milijardus dolerių šiai problemai spręsti.“

Net su „OpenAI“ pagalba sunkiai sekėsi augti ir „Bing“. Pasak duomenų analitikos įmonės „StatCounter“, 2024 metų pradžioje „Bing“ rinkos dalis buvo įstrigusi ties varganais 3 procentais – nepakankamai „Google“ dominavimui rinkoje susilpninti. Esami rinkos žaidėjai skynė pergales, o jų teritorijos buvo aiškiai apibrėžtos: „Google“ kontroliavo paieškas, „Microsoft“ dominavo programinės įrangos srityje ir abi grūmėsi su „Amazon“ dėl dominavimo debesijos versle.

Šiandien sukurti DI modelį yra taip brangu, jog ši užduotis tampa įgyvendinama tik technologijų milžinėms. Akademikai ir mažesnės bendrovės tepajėgia pirkti lustus iš „Nvidia“ ir nuomotis skaičiavimo pajėgumus iš „Amazon“, „Microsoft“ ar „Google“. Patekus į šias platformas, pasitraukti iš jų dažnai nebepavyksta. Dirbtinio intelekto startuoliai neretai skundžiasi, kad, pradėjus kurti savo paslaugas „Microsoft“ ar „Amazon“ debesijoje, būna sunku pakeisti tiekėją. Tos pačios bendrovės turi retą prieigą prie tūkstančių grafikos procesorių (GPU), kurių reikia į „ChatGPT“ panašiam produktui sukurti. Gauti lustų, kurių vieno kaina gali siekti ir 40 000 JAV dolerių, yra tas pats, kas bandyti gauti bilietus į koncertą, kai bilietai jau išparduoti. Pagrindinis GPU tiekėjas pasaulyje „Nvidia“ gerokai pasipelnė iš tokios paklausos. 2023 metų gegužę „Nvidia“ tapo naujausia technologijų bendrove – po „Google“, „Microsoft“, „Amazon“, „Meta“ ir „Apple“, – kurios rinkos vertė pasiekė 1 trln. JAV dolerių. Beveik visos didžiausios pasaulio bendrovės

buvo tos, kurios kūrė technologijas ir DI. Bet užuot sukūręs klestinčią rinką pažangioms naujoms įmonėms, DI sprogimas padėjo technologijų bendrovėms konsoliduoti galią. Jos savo gniaužtuose laiko infrastruktūrą, talentus, duomenis, skaičiavimo pajėgumus ir pelną, todėl neabejojama, kad jos kontroliuos ir mūsų DI ateitį.

Bendrojo dirbtinio intelekto svajotojai padėjo paversti tai realybe. 2023 m. birželį „Microsoft“ finansų vadovė Amy Hood pasakojo investuotojams, kad „OpenAI“ DI grįstos paslaugos sugeneruos mažiausiai 10 mlrd. JAV dolerių pajamų. Pasak jos, „tai sparčiausiai augantis 10 mlrd. JAV dolerių verslas istorijoje“.

Ar būtų buvę geriau, jei „DeepMind“ ir „OpenAI“ būtų likusios nepriklausomos ir paskyrusios patikėtinių tarybas DI krypčiai kontroliuoti? Samas Altmanas vėliau sužinojo, kad toks kelias irgi neišvengia rizikos. Suleymanas, įnirtingai kovojęs, kad išlaisvintų savo bendrovę nuo „DeepMind“, interviu sakė, jog didelės bendrovės gali būti patikimesnės už mažesnes. Juk jos privalo viešai atsiskaityti savo akcininkams ir personalui. Bet pasaulio technologijų milžinės turi ir rimtesnį įsipareigojimą savo akcininkams, kurio nusimesti negali. Jos privalo kiekvieną ketvirtį uždirbti daugiau. Pajamoms nekintant arba traukiantis, krinta ir akcijų vertė. Pingant akcijoms, bendrovė nebegali pritraukti kapitalo, vadovai ir darbuotojai ima bruzdėti ar net palieka bendrovę. Siaubinga nuosmukio šmėkla moja ranka. „Šios bendrovės *privalo* augti, – sako buvęs „Microsoft“ vadovas. – Dirbtinis intelektas yra atsakas.“ Hassabisas teigia, kad „DeepMind“ tapo protingesniu verslu. „Pasiekėme tokį brandos lygį, kai, manau, galime pagerinti milijardų žmonių gyvenimus, – sakė jis paklaustas, ar, jo nuomone, BDI vis dar reikia etikos tarybos, kurios jis kadaise siekė. – „Google“ yra nuostabi vieta.“

Altmanas tvirtina, kad, nepaisant „OpenAI“ orientacijos į pelną, supanašėjimo su „Microsoft“ ir pradėtų DI lenktynių, jo principai sukurti naudingą DI liko nepakitę. Jis neturi kito pasirinkimo, tik pristatyti

visuomenei vis naujus DI įrankius. „Praktinis naudojimas yra kritiškai svarbus mūsų misijai“, – sako jis. Kaip kitaip „OpenAI“ mokysis ir kurs žmonėms tokius naudingus įrankius kaip „ChatGPT“? Tam „reikia atiduoti technologijas į žmonių rankas“.

Šios lenktynės tokios sparčios, kad sunku prognozuoti, kas įvyks po kelių mėnesių ar metų nuo šių žodžių užrašymo 2024 metų kovą. Vis dėlto ateities įvykiai priklausys nuo nedidelės grupelės žmonių planų ir juos veikiančių sisteminių jėgų. Atsakymas į klausimą, ar galime patikėti DI ateitį Samui Altmanui ir Demisui Hassabisui kartu su „Microsoft“ ir „Google“: nelabai turime iš ko rinktis. Abu vyrai susiejo savo inovacijas su pasaulio dviem didžiausiomis bendrovėmis, kurių tinklų poveikio beveik neįmanoma išvengti kasdieniame gyvenime. Taip pasielgę jie prisijungė prie gausaus būrio inovatorių, pakoregavusių savo idealus, kad neiškristų iš lenktynių ir įgytų galios. Rezultatas – turbūt didžiausią transformacijos potencialą turinti technologija, kokią kada nors esame matę. Lieka tik sužinoti jos kainą.

PADĖKA

Ši knyga niekada nebūtų pasirodžiusi be nedidelės armijos žmonių paramos ir paskatinimo. Praėjus maždaug mėnesiui nuo „ChatGPT“ pristatymo pasauliui, idėją nusiunčiau savo agentui Davidui Fugate. Apie tai, kaip du vyrai, svajoję sukurti superprotingas mašinas, vėliau tapo varžovais, o dar vėliau ir mūšio tarp didžiųjų technologijų bendrovių veidais. Davido „TAIP!“ privertė per kitus metus šią knygą sudėlioti. Janhoi'us McGregoras suteikė papildomą postūmį, kurio reikėjo, kad imčiausi tokio didelio projekto. Kaip įprasta, padarė tą mums užkandžiaujant „Nando's“.

Esu dėkinga Clairei Cheek, Sarai Beth Haring, Elisai Rivlin ir visam nuostabiam „St. Martin's Press“ personalui, savo redaktoriui Peteriui Wolvertonui, kurio objektyvūs patarimai man leido susitelkti į tai, kaip transformuojančio išradimo ateitį kuria vos kelios bendrovės. Pete'as išgelbėjo mane (ir jus, brangus skaitytojau) nuo šuolio į šiai istorijai neaktualių pasakojimų apie transhumanizmą ir veiksmingojo altruizmo judėjimo bedugnę.

Kadangi apie tai prakalbome, privalau padėkoti dr. Emilei Torres už nuostabų darbą atsekant pastangas sukurti BDI iki jo tamsesnių šaknų eugenikoje ir už pagalbą aiškinantis kitą žmogaus tobulumo pasitelkiant mašinas vaikymosi pusę. Davidas Edmondsas padėjo man ilgalaikiškumo (angl. *longtermism*) tema, Toby'is Ordas – veiksmingojo altruizmo, o Mike'as Levine'as – DI suderinamumo organizacijose, kaip antai „Open Philanthropy“, temomis. Mike'o ir mano nuomonės kai kuriomis iš šių temų skiriasi, bet esu jam labai dėkinga už kantrybę ir laiką, skirtą aiškinant savo motyvus. Siatle įsikūręs Brianas Evergreenas padėjo man geriau suprasti DI etikos galvosūkį, virusį „Microsoft“ viduje.

Ypatingą padėką siunčiu Meredith Whittaker ir Reolofui Bothai, dviem žmonėms iš labai skirtingų technologijų pasaulio sričių. Meredith yra „Signal Foundation“ prezidentė, Reolofas – rizikos kapitalo įmonės „Sequoia Capital“ partneris. Abu jie padėjo man prieiti prie išvados, kad stiprėjantis vos kelių technologijų bendrovių dominavimas DI srityje tampa problema tiek visuomenei, tiek verslui.

Pastabose miniu anoniminius šaltinius, tačiau privalau padėkoti daugeliui žmonių, kurie dirba technologijų ir DI bendrovėse ar yra buvę „OpenAI“ ir „DeepMind“ darbuotojai, įskaitant aukščiausiuosius vadovus. Jie valandų valandas dalijosi savo patirtimi ir kartais giliu nerimu, kaip stipriai technologijų milžinės kontroliavo DI sritį, taip pat susirūpinimu dėl jų kuriamų inovacijų griaunamosios galios neatsižvelgiant į pasekmes. Tikiuosi, knygoje man pavyko nušviesti šį nerimą ir su manimi praleistas tų žmonių laikas nenuėjo veltui.

Mano redaktoriai „Bloomberg Opinion“ stipriai palaikė šį projektą. Esu nepaprastai dėkinga Timui O’Brienui ir Nicolei Torres už jų entuziazmą ir sutikimą duoti man laiko, kad galėčiau parašyti šią knygą, papildžiusią tai, apie ką ankstesniais metais dažniausiai rašydavau savo skiltyje. Didelį ačiū už gražius žodžius ir palaikymą, už tai, kad technologijų skilties rašytojų darbą pavertė linksmesniu, nei jis turėtų būti, siunčiu ir savo kolegoms rašytojams „Bloomberg Opinion“, įskaitant Dave’ą Lee, Larą Williams, Lionelį Laurentą, Andrea Felstead, Therese’ą Raphael, Matthew Brookerį, Howardą Chua-Eoaną, Chrisą Hughesą, Chrisą Bryantą, Marcusą Ashworthą, Marcą Championą, Jamesą Hertlingą, Joi Preciphs, Marką Gilbertą, Elaine He, Timą Culpaną. Taip pat dėkoju Javierui Blasui už patarimus imantis knygos projekto.

Esu itin dėkinga savo kolegoms Timui O’Brienui, Nicolei Torres, Paului Daviesui ir Adrianui Wooldridge’ui už jų aštrų ir naudingą grįžtamąjį ryšį perskaičius pirmąsias rankraščio versijas. Tai man labai padėjo suprasti kitų skaitytojų, esančių už mano rašytojos burbulo, požiūrį. Kiti pirmieji mano rankraščio skaitytojai ne tik parodė, kurias

vietas turėčiau patobulinti ar kuriuos dalykus geriau paaiškinti, bet dažnai buvo ir moralinė parama, puikūs draugai viso rašymo proceso metu. Didelė padėka už tai Miriamai Zaccarelli, Victorui Zaccarelli, Carley'i Sime ir Kristinai Peterson.

Ačiū mano kaimynei ir draugei Catalinai Katinai Montesinos, leidusiai man pasinaudoti jos namų tyla, kai ėmiausi rašyti knygą. Katina yra viena tų žmonių, kurių gyvenime DI gali turėti reikšmingą ir teigiamą poveikį. Buvusi tapytoja, keturiasdešimties netekusi regos, ji tapo sėkminga skulptore ir atviromis rankomis priėmė visus būdus, kuriais technologijos gali padėti jai „matyti" pasaulį. Ji mielai naudojasi skaitmeniniais padėjėjais, tokiais kaip „Apple Siri". Būdama aštuoniasdešimties, ji su gilia nuostaba klausėsi, kai, nukreipusi telefoną į ant jos kavos stalelio stovinčią skulptūrą, paprašiau „ChatGPT" parengti išsamią skulptūros spalvų, formos ir galimų meninių įtakų analizę. Pasak Katinos, tai buvo „visiškai neįtikėtina". Tikiuosi, kad būtent šia kryptimi mūsų pasaulis nukreips DI ateityje – informacijos spragoms užpildyti, o ne žmonių darbui ir kūrybiškumui išstumti.

Galiausiai, ši knyga nebūtų buvusi įmanoma be šeimos paramos, įskaitant mano tėvą Philipą Withersą. Jis nuo mano vaikystės buvo neišsenkantis pagyrų šaltinis, skatinęs net už mažiausius pasiekimus. Ačiū Carai ir Wesley'iui už juoko pilnus namus ir Islai už rankraščio tikrinimą ir palaikymą. Kiekvieną dieną nesiliauju stebėtis, kaip mano vyras Manis suburia mus visus. Man jis – nuolatinis stiprybės ir kantrybės šaltinis. Esu jam itin dėkinga.

ŠALTINIAI

Pastaba apie šaltinius

Esu dėkinga žurnalistams ir autoriams, parašiusiems daugybę straipsnių interneto svetainėse, laikraščiuose, žurnaluose, moksliniuose leidiniuose, paskelbusiems tinklalaidėse ir knygose. Didžiulis jų atliktas darbas leido man papildyti šią knygą gausiais antriniais šaltiniais, kurių sąrašą pateikiu toliau.

Kai knygoje žmonės cituojami esamuoju laiku ir vartojami žodžiai „teigia“, „prisimena“ ar panašiai, tai yra citatos iš mano tiesioginių interviu su tais žmonėmis, tarp jų su Demisu Hassabisu ir Samu Altmanu. Daugelis kitų asmenų, su kuriais kalbėjausi, paminėti kaip buvę darbuotojai arba temą išmanantys asmenys. Jie lieka anonimais dėl galimų neigiamų pasekmių už nuomonės išsakymą. Esu ypač dėkinga šiems žmonėms už pasitikėjimą ir pasidalijimą savo įžvalgomis.

Buvo ir daugiau interviu, kurių negalėjau įtraukti į knygą dėl vietos stokos. Tie pokalbiai buvo labai vertingi, nes nušvietė Samo Altmano ir Demiso Hassabiso gyvenimo bei darbo DI srityje kontekstą. Taip pat dėkoju ekspertams, kurie padėjo suprasti mašininio mokymosi sistemų, neuroninių tinklų, difuzijos sistemų ir transformatorių veikimą ir pateikti tai suprantama kalba.

Tiek rengiant šią knygą, tiek paskutinius kelerius metus rašant apie naująjį DI bumą žurnalams „Bloomberg Opinion“, „Wall Street Journal“ ir „Forbes“, man labai padėjo pokalbiai su šimtais pramonės ekspertų, verslininkų, rizikos kapitalistų ir technologijų įmonių esamų bei buvusių darbuotojų.

Labai mėgstu bėgioti, tad pasinaudojau tuo ir bėgiodama išklausiau begalę valandų tinklalaidžių pokalbių su Samu Altmanu, Demisu

Hassabisu, Ilya Sutskeveriu, Gregu Brockmanu ir daugybe kitų žmonių, susijusių su „OpenAI“ ir „DeepMind“ įkūrimu arba liudijusių DI evoliuciją nuo nišinio mokslinio projekto iki klestinčio verslo. Šie pokalbiai padėjo sudėlioti daugybę pasakojimo detalių į vieną visumą. Dėl to bėgant dažnai tekdavo sustoti ir pasižymėti pastabas telefone, bet tikrai buvo verta.

1 SKYRIUS:

Altman, Sam. “Machine Intelligence, Part 1.” blog.samaltman.com, February 25, 2015. Cannon, Craig. “Sam Altman.” *Y Combinator* (podcast), November 8, 2018.

“First Look: Loopt Provides More Incentives to Try Location-based Services with Loopt Star.” Robert Scoble’s YouTube channel, June 1, 2010.

Friend, Tad. “Sam Altman’s Manifest Destiny.” *New Yorker*, October 3, 2016. Graham, Paul. “How to Start a Startup.” paulgraham.com, March 2005.

Graham, Paul. “How Y Combinator Started.” paulgraham.com, March 2012. Graham, Paul. “The Word ‘Hacker.’” paulgraham.com, April 2024.

“How Tesla Became the Elon Musk Co.” *Land of the Giants* (podcast), Vox Media, August 2, 2023.

Internet Archive for the now defunct website, http://www.loopt.com/.

Lessin, Jessica. “This Is How Sam Altman Works the Press and Congress. I Know from Experience.” *The Information*, June 7, 2023.

Mitchell, Melanie. *Artificial Intelligence: A Guide for Thinking Humans*. New York: Pelican, 2020.

Wagstaff, Keith. “The Good Ol’ Days of AOL Chat Rooms.” *CNN*, July 6, 2012.

Weil, Elizabeth. “Sam Altman Is the Oppenheimer of Our Age.” *New York Magazine*, September 25, 2023.

Wired. “Sebastian Thrun & Sam Altman Talk Flying Vehicles and Artificial Intelligence.” Video of *Wired* conference panel, October 16, 2018.

“WWDC 2008 News: Loopt Shows Off New App for the iPhone.” CNET’s YouTube channel, June 10, 2008.

2 SKYRIUS:

“A.I. Could Solve Some of Humanity’s Hardest Problems. It Already Has.” *The Ezra Klein Show*, July 11, 2023.

Burton-Hill, Clemency. “The Superhero of Artificial Intelligence.” *The Guardian*, February 16, 2016.

"Demis Hassabis." *Desert Island Discs* (podcast), BBC Radio 4, May 21, 2017.

"Demis Hassabis, Ph.D." *What It Takes* (podcast), American Academy of Achievement, April 23, 2018.

"Genius Entrepreneur." *The Bridge*, Queens College Cambridge Magazine, September 2014.

"Google DeepMind's Demis Hassabis." *The Bottom Line* (podcast), BBC Radio 4, October 16, 2023.

Hassabis, Demis. *The Elixir Diaries*, columns in *Edge* magazine, also available at https://archive.kontek.net/, 1998–2000.

Parker, Sam. "Republic: The Revolution Review." *GameSpot*, September 2, 2003.

Pearce, Jacqui. "Getting to Know You," a Q&A with Angela Hassabis, *HBC Accord*, (Hendon Baptist Church newsletter), October 2018.

"Republic." *Edge* magazine, November 1999.

Weinberg, Steven. *Dreams of a Final Theory*. New York: Vintage, 1994.

3 SKYRIUS:

Altman, Sam. "Hard Tech Is Back." blog.samaltman.com, March 11, 2016. Altman, Sam. "Startup Advice." blog.samaltman.com, June 3, 2013.

Altman, Sam. "YC and Hard Tech Startups." ycombinator.com, date not provided. Cannon, Craig. "Sam Altman." *Y Combinator* (podcast), November 8, 2018.

Chafkin, Max. "Y Combinator President Sam Altman Is Dreaming Big." *Fast Company*, April 16, 2015.

Clifford, Catherine. "Nuclear Fusion Start-Up Helion Scores $375 Million Investment from Open AI CEO Sam Altman." *CNBC*, November 5, 2021.

Dwoskin, Elizabeth, Marc Fisher, and Nitasha Tiku. "'King of the Cannibals': How Sam Altman Took Over Silicon Valley." *Washington Post*, December 23, 2023.

Friend, Tad. "Sam Altman's Manifest Destiny." *New Yorker*, October 3, 2016. Graham, Paul. "Five Founders." paulgraham.com, April 2019.

Graham, Paul. "How to Start a Startup." paulgraham.com, March 2005.

"How and Why to Start a Startup—Sam Altman & Dustin Moskovitz—Stanford CS183F: Startup School." Stanford Online's YouTube channel, April 5, 2017.

"Paul Graham on Why Sam Altman Took Over as President of Y Combinator in 2014." This Week in Startups Clips's YouTube channel, March 19, 2019.

Regalado, Antonio. "A Startup Is Pitching a Mind-Uploading Service that Is '100 Percent Fatal.'" *MIT Technology Review*, March 13, 2018.

"Sam Altman—Leading with Crippling Anxiety, Discovering Meditation, and Building Intelligence with Self-Awareness." *The Art of Accomplishment* (podcast), January 15, 2022.

Stiegler, Marc. *The Gentle Seduction*. New York: Baen, 1990.

4 SKYRIUS:

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

Cutright, Keisha M., and Mustafa Karataş. “Thinking about God Increases Acceptance of Artificial Intelligence in Decision-Making.” *Proceedings of the National Academy of Sciences (PNAS)* 120, no. 33 (2023): e2218961120–e2218961120.

“Demis Hassabis, Ph.D.” *What It Takes* (podcast), American Academy of Achievement, April 23, 2018.

Dowd, Maureen. “Elon Musk’s Billion-Dollar Crusade to Stop the A.I. Apocalypse.” *Vanity Fair*, March 26, 2017.

Goertzel, Ben. *AGI Revolution: An Inside View of the Rise of Artificial General Intelligence*. Middletown, DE: Humanity+ Press, 2016.

Hassabis, Demis. “The Neural Processes Underpinning Episodic Memory.” PhD thesis, University College London, February 2009.

Homer-Dixon, Thomas. *The Ingenuity Gap*. New York: Knopf Doubleday, 2000. Kurzweil, Ray. *The Age of Spiritual Machines*. New York: Penguin, 2000.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence: August 31, 1955.” *The AI Magazine* 27, no. 4 (2006): 12–14. Copies of the original typescript are housed in the archives at Dartmouth College and Stanford University. A reproduction of the proposal can be found here: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904.

Penrose, Roger. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. London: Vintage, 1995.

Syed, Matthew. “Demis Hassabis Interview: The Kid from the Comp Who Founded DeepMind and Cracked a Mighty Riddle of Science.” *The Sunday Times*, December 5, 2020.

“A Systems Neuroscience Approach to Building AGI—Demis Hassabis, Singularity Summit 2010.” Google DeepMind’s YouTube channel, March 7, 2018.

Thiel, Peter, with Blake Masters. *Zero to One: Notes on Start Ups, or How to Build the Future*. London: Virgin Books, 2015.

5 SKYRIUS:

“Andrew Ng: Deep Learning, Education, and Real-World AI.” *Lex Fridman Podcast* (podcast), February 20, 2020.

"Bill Gates, Sergey Brin, and Larry Page: Tech Titans." *What It Takes* (podcast), American Academy of Achievement, achievement.org, January 13, 2020.

Copeland, Rob. "Google Management Shuffle Points to Retreat from Alphabet Experiment." *Wall Street Journal*, December 5, 2019.

Hodson, Hal. "DeepMind and Google: The Battle to Control Artificial Intelligence." *Economist*, March 1, 2019.

Huxley, Julian. "Transhumanism." *Journal of Humanistic Psychology*, January 1968. Markram, Henry. "A Brain in a Supercomputer." TED Global, July 2009.

Metz, Cade. *Genius Makers: The Mavericks Who Brought A.I. to Google, Facebook, and the World*. New York: Random House Business, 2021.

"Peter Thiel Says America Has Bigger Problems Than Wokeness." *Honestly with Bari Weiss* (podcast), May 3, 2023.

Suleyman, Mustafa, with Michael Bhaskar. *The Coming Wave*. New York: Crown, 2023.

6 SKYRIUS:

Albergotti, Reed. "The Secret History of Elon Musk, Sam Altman, and OpenAI."

Semafor, March 24, 2023.

Birhane, Abeba, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. "The Values Encoded in Machine Learning Research." *FAccT Conference '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (June 2022): 173–84.

Brockman, Greg. "My Path to OpenAI." blog.gregbrockman.com, May 3, 2016.

Conn, Ariel. "Concrete Problems in AI Safety with Dario Amodei and Seth Baum." *Future of Life Institute* (podcast), August 31, 2016.

Dowd, Maureen. "Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse." *Vanity Fair*, March 26, 2017.

Elon Musk vs Sam Altman [2024] CGC-24-612746.

Friend, Tad. "Sam Altman's Manifest Destiny." *New Yorker*, October 3, 2016.

Galef, Jesse. "Elon Musk Donates $10M to Our Research Program." futureoflife.org, January 22, 2015.

"Greg Brockman: OpenAI and AGI." *Lex Fridman Podcast* (podcast), April 3, 2019.

Harris, Mark. "Elon Musk Used to Say He Put $100M in OpenAI, but Now It's $50M: Here Are the Receipts." *TechCrunch*, May 18, 2023.

Metz, Cade. "Ego, Fear and Money: How the A.I. Fuse Was Lit." *New York Times*, December 3, 2023.

Metz, Cade. “Inside OpenAI, Elon Musk’s Wild Plan to Set Artificial Intelligence Free.” *Wired*, April 27, 2016.

Vance, Ashlee. *Elon Musk: How the Billionaire CEO of SpaceX and Tesla Is Shaping Our Future*. London: Virgin Books, 2015.

7 SKYRIUS:

Aron, Jacob. “How to Build the Global Mathematics Brain.” *New Scientist*, May 4, 2011. Byford, Sam. “Google’s AlphaGo AI Defeats World Go Number One Ke Jie.” *The Verge*, May 23, 2017.

Gallagher, Ryan. “Google’s Secret China Project ‘Effectively Ended’ after Internal Confrontation.” *The Intercept*, December 17, 2018.

Gallagher, Ryan. “Private Meeting Contradicts Google’s Official Story on China.” *The Intercept*, October 9, 2018.

“Has Anyone Actually Tried to Convince Terry Tao or Other Top Mathematicians to Work on Alignment?” www.lesswrong.com, June 8, 2022.

“How to Play.” British Go Association, updated October 26, 2017, https://www.britgo.org/intro/intro2.html.

Metz, Cade. *Genius Makers: The Mavericks Who Brought A.I. to Google, Facebook, and the World*. New York: Random House Business, 2021.

Metz, Cade. “Google Is Already Late to China’s AI Revolution.” *Wired*, June 2, 2017.

Rogin, Josh. “Eric Schmidt: The Great Firewall of China Will Fall.” *Foreign Policy*, July 9, 2012.

Suleyman, Mustafa, with Michael Bhaskar. *The Coming Wave*. New York: Crown, 2023.

Temperton, James. “DeepMind’s New AI Ethics Unit Is the Company’s Next Big Move.” *Wired*, October 4, 2017.

Yang, Yuan. “Google’s AlphaGo Is World’s Best Go Player.” *Financial Times*, May 25, 2017.

8 SKYRIUS:

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. “Machine Bias.” *ProPublica*, May 23, 2016.

Buolamwini, Joy, and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” *Proceedings of Machine Learning Research* 81 (2018): 1–15.

Dastin, Jeffrey. “Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women.” *Reuters*, October 10, 2018.

Devlin, Hannah, and Alex Hern. “Why Are There So Few Women in Tech? The Truth behind the Google Memo.” *The Guardian*, August 8, 2017.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. “Datasheets for Datasets.” *Communications of the ACM* 64, no. 12 (2021): 86–92.

Grant, Nico, and Kashmir Hill. “Google’s Photo App Still Can’t Find Gorillas. And Neither Can Apple’s.” *New York Times*, May 22, 2023.

Harris, Josh. “There Was All Sorts of Toxic Behaviour’: Timnit Gebru on Her Sacking by Google, AI’s Dangers and Big Tech’s Biases.” *The Guardian*, May 22, 2023.

Horwitz, Jeff. “The Facebook Files.” *Wall Street Journal*, October 1, 2021.

Payton, L’Oreal Thompson. “Americans Check Their Phones 144 Times a Day. Here’s How to Cut Back.” *Fortune*, July 19, 2023.

Simonite, Tom. “What Really Happened When Google Ousted Timnit Gebru.” *Wired*, June 8, 2021.

“The Social Atrocity: Meta and the Right to Remedy for the Rohingya.” Amnesty International report, September 29, 2022.

Wakabayashi, Daisuke, and Katie Benner. “How Google Protected Andy Rubin, the ‘Father of Android.’” *New York Times*, October 25, 2018.

9 SKYRIUS:

de Vynck, Gerrit. “Google’s Cloud Unit Won’t Sell a Type of Facial Recognition Tech.” *Bloomberg*, December 13, 2018.

“Google Duplex: A.I. Assistant Calls Local Businesses to Make Appointments.” Jeff Grubb’s Game Mess’s YouTube channel, May 8, 2018.

Kruppa, Miles, and Sam Schechner. “How Google Became Cautious of AI and Gave Microsoft an Opening.” *Wall Street Journal*, March 7, 2023.

Love, Julia. “Google Says Over Half of Generative AI Startups Use Its Cloud.” *Bloomberg*, August 29, 2023.

Nylen, Leah. “Google Paid $26 Billion to Be Default Search Engine in 2021.” *Bloomberg*, October 17, 2021.

Uszkoreit, Jakob. “Transformer: A Novel Neural Network Architecture for Language Understanding.” blog.research.google, August 31, 2017.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30 (2017).

10 SKYRIUS:

Brockman, Greg (@gdb). “Held our civil ceremony in the @OpenAI office last week. Officiated by @ilyasut, with the robot hand serving as ring bearer. Wedding planning to commence

soon." Twitter, November 12, 2019, 9:39 a.m. https://twitter.com/gdb/status/1194293590979014657?lang=en.

Brockman, Greg. "Microsoft Invests in and Partners with OpenAI to Support Us Building Beneficial AGI." www.openai.com, July 22, 2019.

"Greg Brockman: OpenAI and AGI." *Lex Fridman Podcast* (podcast), April 3, 2019.

Hao, Karen. "The Messy, Secretive Reality behind OpenAI's Bid to Save the World." *MIT Technology Review*, February 17, 2020.

Hao, Karen, and Charlie Warzel. "Inside the Chaos at OpenAI." *The Atlantic*, November 19, 2023.

Jin, Berber, and Keach Hagey. "The Contradictions of Sam Altman, AI Crusader." *Wall Street Journal*, March 31, 2023.

Kraft, Amy. "Microsoft Shuts Down AI Chatbot after It Turned into a Nazi." *CBS News*, March 25, 2016.

Levy, Steven. "What OpenAI Really Wants." *Wired*, September 5, 2023.

Metz, Cade. "A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit." *New York Times*, April 19, 2018.

Metz, Cade. "The ChatGPT King Isn't Worried, but He Knows You Might Be." *New York Times*, March 31, 2023.

"OpenAI Charter." www.openai.com/charter, April 9, 2018.

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. "Improving Language Understanding by Generative Pre-Training." www.openai.com, June 11, 2018.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language Models Are Unsupervised Multitask Learners." www.openai.com, February 14, 2019.

11 SKYRIUS:

Ahmed, Nur, Muntasir Wahed, and Neil C. Thompson. "The Growing Influence of Industry in AI Research." *Science*, March 2, 2023.

Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. "Concrete Problems in AI Safety." www.arxiv.org, July 25, 2016.

Copeland, Rob. "Google Management Shuffle Points to Retreat from Alphabet Experiment." *Wall Street Journal*, December 5, 2019.

Coulter, Martin, and Hugh Langley. "DeepMind's Cofounder Was Placed on Leave after Employees Complained about Bullying and Humiliation for Years. Then Google Made Him a VP." *Business Insider*, August 7, 2021.

Friend, Tad. "Sam Altman's Manifest Destiny." *New Yorker*, October 3, 2016. Hodson, Hal. "Revealed: Google AI Has Access to Huge Haul of NHS Patient Data." *New Scientist*, April 29,

2016.

Ludlow, Edward, Matt Day, and Dina Bass. “Amazon to Invest Up to $4 Billion in AI Startup Anthropic.” *Bloomberg*, September 25, 2023.

Piper, Kelsey. “Exclusive: Google Cancels AI Ethics Board in Response to Outcry.” *Vox*, April 4, 2019.

Primack, Dan. “Google Is Investing $2 Billion into Anthropic, a Rival to OpenAI.” *Axios*, October 30, 2023.

Waters, Richard. “DeepMind Co-founder Leaves Google for Venture Capital Firm.” *Financial Times*, January 21, 2022.

12 SKYRIUS:

Abid, Abubakar, Maheen Farooqi, and James Zou. “Large Language Models Associate Muslims with Violence.” *Nature Machine Intelligence* 3 (2021): 461–63.

Barrett, Paul, Justin Hendrix, and Grant Sims. “How Tech Platforms Fuel U.S. Political Polarization and What Government Can Do about It.” www.brookings.edu, September 27, 2021.

Bender, Emily, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” *FAccTConference ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (March 2021): 610–23. https://dl.acm.org/doi/10.1145/3442188.3445922.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. “Language Models Are Few-Shot Learners.” www.openai.com, July 22, 2020.

Gehman, Samuel, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. “RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models.” *ACL Anthology*. Findings of the Association for Computational Linguistics: EMNLP 2020, November 2020.

Hornigold, Thomas. “This Chatbot Has Over 660 Million Users—and It Wants to Be Their Best Friend.” *Singularity Hub*, July 14, 2019.

Jin, Berber, and Miles Kruppa. “Microsoft to Deepen OpenAI Partnership, Invest Billions in ChatGPT Creator.” *Wall Street Journal*, January 23, 2023.

Lecher, Colin. “The Artificial Intelligence Field Is Too White and Too Male, Researchers Say.” *The Verge*, April 17, 2019.

Lemoine, Blake. “I Worked on Google’s AI. My Fears Are Coming True.” *Newsweek*, February 27, 2023.

Lodewick, Colin. “Google’s Suspended AI Engineer Corrects the Record: He Didn’t Hire an Attorney for the ‘Sentient’ Chatbot, He Just Made Introductions—the Bot Hired the Lawyer.” *Fortune*, June 23, 2022.

Luccioni, Alexandra, and Joseph Viviano. “What’s in the Box? An Analysis of Undesirable Content in the Common Crawl Corpus.” *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Volume 2: Short Papers (2021): 182–89.

Muller, Britney. “BERT 101: State of the Art NLP Model Explained.” www.huggingface.co, March 2, 2022.

Newton, Casey. “The Withering Email That Got an Ethical AI Researcher Fired at Google.” *Platformer*, December 3, 2020.

Nicholson, Jenny. “The Gender Bias Inside GPT-3.” www.medium.com, March 8, 2022.

Perrigo, Billy. “Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic.” *Time*, January 18, 2023.

Silverman, Craig, Craig Timberg, Jeff Kao, and Jeremy B. Merrill. “Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading Up to Jan. 6 Attack, Records Show.” *ProPublica* and *Washington Post*, January 4, 2022.

Simonite, Tom. “What Really Happened When Google Ousted Timnit Gebru.” *Wired*, June 8, 2021.

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

Venkit, Pranav Narayanan, Mukund Srinath, and Shomir Wilson. “A Study of Implicit Language Model Bias against People with Disabilities.” *Proceedings of the 29th International Conference on Computational Linguistics* (2022): 1324–32.

Wendler, Chris, Veniamin Veselovsky, Giovanni Monea, and Robert West. “Do Llamas Work in English? On the Latent Language of Multilingual Transformers.” www.arxiv.org, February 16, 2024.

13 SKYRIUS:

“AlphaFold: The Making of a Scientific Breakthrough.” Google DeepMind’s YouTube channel, November 30, 2020.

Andersen, Ross. “Does Sam Altman Know What He’s Creating?” *The Atlantic*, July 24, 2023.

Grant, Nico. “Google Calls in Help from Larry Page and Sergey Brin for A.I. Fight.” *New York Times*, January 20, 2023.

Grant, Nico, and Cade Metz. “A New Chat Bot Is a ‘Code Red’ for Google’s Search Business.” *New York Times*, December 21, 2022.

Hao, Karen, and Charlie Warzel. “Inside the Chaos at OpenAI.” *The Atlantic*, November 19, 2023.

Heikkilä, Melissa. “This Artist Is Dominating AI-generated Art. And He’s Not Happy About It.” *MIT Technology Review*, September 16, 2022.

“Introducing ChatGPT.” www.openai.com, November 30, 2022.

Johnson, Khari. “DALL-E 2 Creates Incredible Images—and Biased Ones You Don’t See.” *Wired*, May 5, 2022.

McLaughlin, Kevin, and Aaron Holmes. “How Microsoft’s Stumbles Led to Its OpenAI Alliance.” *The Information*, January 23, 2023.

Merritt, Rick. “AI Opener: OpenAI’s Sutskever in Conversation with Jensen Huang.” www.blogs.nvidia.com, March 22, 2023.

“Microsoft CTO Kevin Scott on AI Copilots, Disagreeing with OpenAI, and Sydney Making a Comeback.” *Decoder with Nilay Patel* (podcast), May 23, 2023.

Patel, Nilay. “Microsoft Thinks AI Can Beat Google at Search—CEO Satya Nadella Explains Why.” *The Verge*, February 8, 2023.

Pichai, Sundar. “Google DeepMind: Bringing Together Two World-Class AI Teams.” www.blog.google, April 20, 2023.

Rawat, Deeksha. “Unravelling the Dynamics of Diffusion Model: From Early Concept to Cutting-Edge Applications.” www.medium.com, August 5, 2023.

Roose, Kevin. “Bing’s A.I. Chat: ‘I Want to Be Alive.’” *New York Times*, February 16, 2023.

“Sam Altman on the A.I. Revolution, Trillionaires and the Future of Political Power.” *The Ezra Klein Show* (podcast), June 11, 2021.

Weise, Karen, Cade Metz, Nico Grant, and Mike Isaac. “Inside the A.I. Arms Race That Changed Silicon Valley Forever.” *New York Times*, December 5, 2023.

14 SKYRIUS:

Details about Open Philanthropy’s disclosure of its executive director being married to someone who worked at OpenAI comes from www.openphilanthropy.org/grants/openai-general-support/.

Details of investments by FTX founders into Anthropic come from Pitchbook, a market research firm.

Details on Open Philanthropy’s grants and funding come from www.openphilanthropy.org/grants/.

Texts between William MacAskill and Elon Musk are sourced from court filings that were released as part of a pretrial discovery process in a legal battle between Musk and Twitter, dated September 28, 2022.

Anderson, Mark. “Advice for CEOs Under Pressure from the Board to Use Generative AI.” *Fast Company*, October 31, 2023.

Berg, Andrew, Christ Papageorgiou, and Maryam Vaziri. “Technology’s Bifurcated Bite.” *F&D Magazine*, International Monetary Fund, December 2023.

Bordelon, Brendan. “How a Billionaire-Backed Network of AI Advisers Took Over Washington.” *Politico*, February 23, 2024.

“EU AI Act: First Regulation on Artificial Intelligence.” www.europarl.europa.eu, June 8, 2023.

Gross, Nicole. “What ChatGPT Tells Us about Gender: A Cautionary Tale about Performativity and Gender Biases in AI.” *Social Sciences*, August 1, 2023.

Johnson, Simon, and Daron Acemoglu. *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. New York: Basic Books, 2023.

Lewis, Gideon. “The Reluctant Prophet of Effective Altruism.” *New Yorker*, August 8, 2022.

Lewis, Michael. *Going Infinite*. New York: Penguin, 2023.

MacAskill, William. *What We Owe the Future*. London: Oneworld, 2022.

Metz, Cade. “The ChatGPT King Isn’t Worried, but He Knows You Might Be.” *New York Times*, March 31, 2023.

Metz, Cade. “‘The Godfather of A.I.’ Leaves Google and Warns of Danger Ahead.” *New York Times*, May 1, 2023.

Millar, George. “The Magical Number Seven, Plus or Minus Two.” *Psychological Review*, 1956.

Milmo, Dan, and Alex Hern. “Discrimination Is a Bigger AI Risk Than Human Extinction—EU Commissioner.” *The Guardian*, June 14, 2023.

Mollman, Steve. “A Lawyer Fired after Citing ChatGPT-Generated Fake Cases Is Sticking with AI Tools.” *Fortune*, November 17, 2023.

Moss, Sebastian. “How Microsoft Wins.” www.datacenterdynamics.com, November 24, 2023.

O’Brien, Sara Ashley. “Bumble CEO Whitney Wolfe Herd Steps Down.” *Wall Street Journal*, November 6, 2023.

“Pause Giant AI Experiments: An Open Letter.” Future of Life Institute, www.futureoflife.org, March 22, 2023.

Perrigo, Billy. “OpenAI Could Quit Europe Over New AI Rules, CEO Sam Altman Warns.” *Time*, May 25, 2023.

Piantadosi, Steven (@spiantado). “Yes, ChatGPT is amazing and impressive. No, @ OpenAI has not come close to addressing the problem of bias. Filters appear to be bypassed with simple tricks, and superficially masked.” Twitter, December 4, 2022, 10:55 a.m. https://twitter.com/spiantado/status/1599462375887114240?lang=en.

Piper, Kelsey. “Sam Bankman-Fried Tries to Explain Himself.” *Vox*, November 16, 2022.

"Rishi Sunak & Elon Musk: Talk AI, Tech & the Future." Rish Sunak's YouTube channel, November 3, 2023.

"Romney Leads Senate Hearing on Addressing Potential Threats Posed by AI, Quantum Computing, and Other Emerging Technology." www.romney.senate.gov, September 19, 2023.

Roose, Kevin. "Inside the White-Hot Center of A.I. Doomerism." *New York Times*, July 11, 2023.

"Sam Altman: 'I Trust Answers Generated by ChatGPT Least than Anybody Else on Earth.'" Business Today's YouTube channel, June 8, 2023.

Singer, Peter. *The Life You Can Save*. New York: Random House, 2010. "Statement on AI Risk." Center for AI Safety, www.safe.ai, May 2023.

Vallance, Chris. "Artificial Intelligence Could Lead to Extinction, Experts Warn." *BBC News*, May 30, 2023.

Vincent, James. "OpenAI Sued for Defamation after ChatGPT Fabricates Legal Accusations against Radio Host." *The Verge*, June 9, 2023.

Weprin, Alex. "Jeffrey Katzenberg: AI Will Drastically Cut Number of Workers It Takes to Make Animated Movies." *Hollywood Reporter*, November 9, 2023.

Yudkowsky, Eliezer. "Pausing AI Developments Isn't Enough. We Need to Shut It All Down." *Time*, March 29, 2023.

15 SKYRIUS:

"The Capabilities of Multimodal AI | Gemini Demo." Google's YouTube channel, December 6, 2023.

Dastin, Jeffrey, Krystal Hu, and Paresh Dave. "Exclusive: ChatGPT Owner OpenAI Projects $1 Billion in Revenue by 2024." *Reuters*, December 15, 2022.

Gurman, Mark. "Apple's iPhone Design Chief Enlisted by Jony Ive, Sam Altman to Work on AI Devices." *Bloomberg*, December 26, 2023.

Hagey, Keach, Deepa Seetharaman, and Berber Jin. "Behind the Scenes of Sam Altman's Showdown at OpenAI." *Wall Street Journal*, November 22, 2023.

Hawkins, Mackenzie, Edward Ludlow, Gillian Tan, and Dina Bass. "OpenAI's Sam Altman Seeks US Blessing to Raise Billions for AI Chips." *Bloomberg*, February 16, 2024.

Heath, Alex. "Mark Zuckerberg's New Goal Is Creating Artificial General Intelligence." *The Verge*, January 18, 2024.

Imbrie, Andrew, Owen Daniels, and Helen Toner. "Decoding Intentions: Artificial Intelligence and Costly Signals." Center for Security and Emerging Technology, October 2023.

Metz, Cade, Tripp Mickle, and Mike Isaac. "Before Altman's Ouster, OpenAI's Board Was Divided and Feuding." *New York Times*, November 21, 2023.

Roose, Kevin. "Inside the White-Hot Center of A.I. Doomerism." *New York Times*, July 11, 2023.

Sigalos, MacKenzie, and Ryan Browne. “OpenAI’s Sam Altman Says Human-level AI Is Coming but Will Change World Much Less Than We Think.” *CNBC*, January 16, 2024.

Victor, Jon, and Amir Efrati. “OpenAI Made an AI Breakthrough before Altman Firing, Stoking Excitement and Concern.” *The Information*, November 22, 2023.

Walker, Bernadette. “Inside OpenAI’s Shock Firing of Sam Altman.” *Bloomberg*, November 20, 2023.

Zuckerberg, Mark. “Some Updates on Our AI Efforts.” Video posted January 18, 2024 on Facebook. https://www.facebook.com/zuck/posts/pfbid02UhntmXwNBLiV8EZHK71gAQmTx8i4vhfte9vfqjrqyGytfuW4dPQSQ5BnbzMBSPY5l.

16 SKYRIUS:

Bommasani, Rishi, Kevin Klyman, Shayne Longpre, Sayash Kapoor, Nestor Maslej, Betty Xiong, Daniel Zhang, and Percy Liang. “The Foundation Model Transparency Index.” Stanford Center for Research on Foundation Models (CRFM) and Stanford Institute for Human-Centered Artificial Intelligence (HAI), October 18, 2023.

Cheng, Michelle. “AI Girlfriend Bots Are Already Flooding OpenAI’s GPT Store.” *Quartz*, January 11, 2024.

Cheng, Michelle. “A Startup Founded by Former Google Employees Claims that Users Spend Two Hours a Day with Its AI Chatbots.” *Quartz*, October 12, 2023.

Holmes, Aaron. “Microsoft CFO Says OpenAI and Other AI Products Will Add $10 Billion to Revenue.” *The Information*, June 2023.

“Introducing the GPT Store.” www.openai.com, January 10, 2024.

Leswing, Kif. “Nvidia’s AI Chips Are Selling for More than $40,000 on eBay.” *CNBC*, April 14, 2023.

“The Long-Term Benefit Trust.” www.anthropic.com/news/the-long-term-benefit-trust, September 19, 2023.

Schiffmann, Avi (@AviSchiffmann). “I just built the world’s most personal wearable AI! You can talk to Tab about anything in your life. Our computers are now our creative partners!” [demo of Tab]. Twitter, October 1, 2023, 5:12 a.m. https://twitter.com/AviSchiffmann/status/1708439854005321954?lang=en.

Vance, Ashlee. “Elon Musk’s Brain Implant Startup Is Ready to Start Surgery.” *Bloomberg Businessweek*, November 7, 2023.

APIE AUTORĘ

Parmy Olson yra „Bloomberg Opinion“ žurnalistė, daugiau kaip trylika metų rašanti technologijų temomis. Buvusi „Wall Street Journal“ ir „Forbes“ žurnalistė, „We Are Anonymous“ autorė ir „Palo Alto Networks Cybersecurity Canon Award“ apdovanojimo laimėtoja. Jos reportažas apie „Facebook“ 19 mlrd. JAV dolerių vertės „WhatsApp“ įsigijimo sandorį ir vėliau prasidėjusius nesutarimus buvo du kartus paminėtas SABEW verslo žurnalistikos apdovanojimuose. Dirbdama „Wall Street Journal“, ji pirmoji parašė apie slaptus „Google“ geriausios DI laboratorijos ketinimus atsiskirti nuo bendrovės. Šioje knygoje plačiau kalbama apie dramatiškas, nesustabdomas jėgas, privedusias prie to incidento.

KITA PARMY OLSON KNYGA

We Are Anonymous: Inside the Hacker World of LulzSec, Anonymous, and the Global Cyber Insurgency

(„Mes esame anonimai: „LulzSec“, „Anonymous“ ir pasaulinis kibernetinis maištas“)