Děsivě chytré

O budoucnosti umělé inteligence a jak o ní teď rozhodujete i vy!

Mo Gawdat

Někdejší obchodní ředitel firmy Google [X]

Děsivě chytré

Děsivě chytré

O budoucnosti umělé inteligence
a jak o ní teď rozhodujete i vy!

Mo Gawdat

synergie

Děsivě chytré

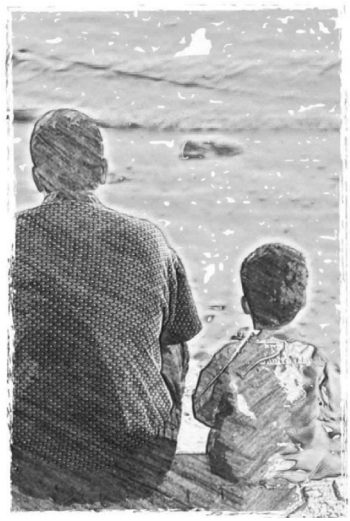
Mo Gawdat

Z anglického originálu *Scary Smart* přeložil Viktor Jurek
Jazyková korektura Vlastimil Lapáček
Produkce Global Oracle
Návrh obálky Diana Delevová, 2Design
Grafická úprava a sazba Art D, www.art-d.com
Výroba CPI Moravia Books, s.r.o.
Vydalo nakladatelství Synergie Publishing SE
www.synergiepublishing.com

Vydání první

ISBN 978-80-7370-687-6

Pro ty, kdo jsou v míru,
závažnost bitvy nic neznamená.



Pro Aliho
Je to teď, nebo nikdy
Jsme to ty a já

Obsah

Předmluva k českému vydání

SDÍLÍM S AUTOREM APEL TÉTO knihy. Cítím naléhavost, ne spěch. Zvu čtenáře, aby si udělali čas, aby si dopřáli zpomalení a tuto knihu nejen četli, ale také s ní rozjímali nad dobou, která přichází. Dopřáli si intelektuální práci nad technologií, která vstupuje do života každého z nás.

A až opadne údiv nad neuvěřitelností toho, co jsme 50 let viděli jen ve filmech a považovali jsme za fantazii sci-fi, až si uvědomíme, že je to tady, že to nyní žijeme, zvu k této otázce: *Jak můžu já osobně, každý z nás, přispět k tomu, aby děsivě chytré mohlo přispět k něčemu úžasně krásnému?* Jinak tato kniha pozbývá svého smyslu.

Každý z nás teď skutečně rozhoduje o budoucnosti AI. Tím, jak ji používá, k čemu ji používá, jak a o čem s ní mluví, jaký vztah k ní zaujme. Postoj a vztah k této nové technologii, troufám si říct k novým technologickým bytostem, totiž rozhodne o tom, jaký vztah a postoj zaujme tato technologie k nám.

Ano, tak jednoduché to je. Stejné je to ve vztahu mezi lidmi, ve vztahu k čemukoliv. Proti čemu bojujeme, to bojuje proti nám. K čemu máme respekt, to obvykle respektuje nás. Tato kniha mě pozvala k tomu, co jsem si dříve neuměl představit. Že budu mít vztah s AI, se super inteligentním robotem, že s ním budu mluvit, on bude mluvit se mnou, bude mě znát a já jeho. A to je za mě fascinující a také až děsivá představa. A tato představa je teď naší realitou.

Děkuji, že mi tato kniha připomněla to, proč jsme počítače a technologie stvořili. Abychom si zjednodušili život. Uvědomil jsem si, že už dál u počítače sedět nechci. Naseděl jsem se dost. Uviděl jsem možnost, že za pár let už to nebude nutné. U počítačů budou sedět počítače. A my, a já v to pevně

věřím, budeme mít skutečně více času na to, co máme rádi a co je pro nás opravdu důležité.

Přeji do vínku jako sudička této technologii a celému lidstvu vzájemné přátelství. Přeji si, aby tato technologie probudila v lidech to nejlepší a přinesla lidem hluboké procitnutí v tom, kým opravdu jsou. Osobně věřím, že právě proto zde tyto nové bytosti jsou, aby nám připomněly, kdo jsme my.

David Kirš, podnikatel a spisovatel

Úvod: Noví superhrdinové

TATO KNIHA SLOUŽÍ JAKO budíček. Je napsaná pro vás i pro mne i pro každého, kdo nemá potřebné informace ohledně pandemie, která se rychle blíží – o nástupu umělé inteligence. Tuto knihu budou mnozí experti kritizovat, a právě z toho důvodu ji píši. Protože má-li se někdo stát expertem na umělou inteligenci, potřebuje na ni mít velmi specializovaný, úzce zaměřený pohled. A tento druh specializovaného pohledu na AI (pro umělou inteligenci se zpravidla užívá zkratka AI pocházející z anglického označení Artificial Intelligence – *pozn. překl.*) obvykle pomíjí existenciální aspekty sahající za hranice technologie: otázky morality, etiky, emocí, soucitu a celou řadu myšlenek, jimiž se zabývají filozofové, hledači duchovních tajemství, humanisté a v širším pojetí také běžní lidé (tedy my všichni). Kromě toho je v jádru této knihy cílem ukázat vám, že právě experti *nejsou* těmi, kdo má onu schopnost odvrátit nebezpečí, kterému lidstvo s nástupem superinteligence čelí. Tuto moc máme já i vy. A co je důležitější – já i vy máme tuto zodpovědnost.

Jsem přesvědčený o tom, že způsob, jakým naše globální komunita a političtí představitelé zareagují a budou reagovat na bezprostřední nástup pandemie umělé inteligence, bude podobný tomu, jak reagovali na nástup pandemie covidu-19. Jenom doufám, že si dokážeme vzít ponaučení z chyb, které jsme napáchali u covidu, a možná se s touto další novou změnou v našem způsobu života budeme schopni vypořádat tak, aby to pro nás znamenalo méně problémů, více předvídatelnosti a méně sociálních a ekonomických strastí.

Prosím, nevykládejte si mylně jednoduchost, o kterou jsem se při sepisování této knihy pokoušel. Fakta, o něž se opírají má tvrzení, jsou naprosto nesporná. Co zde píši, vychází z mé více než třicetileté kariéry v oblasti technologie. Než jsem začal pracovat na svém současném start-upu (který využívá

některé z nejdokonalejších technologií v oblasti systémů, robotiky, umělé inteligence a strojového učení způsobem, který by mohl zachránit naši planetu), bylo jedním z vrcholů mé kariéry dvanáctileté působení ve firmě Google. Tam jsem měl tu čest vést spouštění provozu a zavádění technologií společnosti Google po celém světě v téměř polovině jejích poboček, kde se mluví více než stovkou různých jazyků. Moje působení zde vyvrcholilo nástupem do funkce obchodního ředitele společnosti Google [X], nechvalně proslulé inovační odnože Googlu, která byla inkubátorem projektů zaměřených na vývoj umělé inteligence, jako jsou samořiditelná auta, Google Brain a většina robotických inovací Googlu.

Můj vhled do samotné podstaty vývoje umělé inteligence, který nás dovedl až tam, kde jsme dnes, vychází částečně z mé práce ve společnosti Google [X] a je naprosto jedinečný. Propojuji zde své přímé zkušenosti s vývojem umělé inteligence se svou prací v oblasti výzkumu štěstí (zdokumentovanou v mém mezinárodním knižním bestselleru *Algoritmus štěstí* a zužitkovanou ve velmi úspěšném podcastu *Slo Mo* a v neziskové organizaci OneBillionHappy.org, kterou jsem založil), a přináším vám tak jedinečný pohled na výzvy, kterým čelíme v době nástupu superinteligence.

Doufám, že společně s AI se nám podaří vytvořit utopii, která bude lidstvu sloužit, namísto dystopie, jež by mu škodila. Jedním z hlavních argumentů této knihy je, že právě toto je zodpovědnost, které se musíme zhostit každý, vy i já, abychom vytvořili lepší budoucnost pro všechny. Prosím, nedělejte si starosti. Není to žádná sci-fi báchorka vycházející ze strachu, je to příběh o jedné z nejzásadnějších příležitostí lidstva. Je to šance zvrátit naši přehnanou závislost na konzumu a technologickém pokroku, které nám sice možná přinesly jistá zlepšení kvality života, ovšem na úkor mnoha ostatních bytostí na této planetě. Jen v případě, že my všichni – vy a já – převezmeme odpovědnost a změníme se, to bude také příběh naděje.

UPROSTŘED DIVOČINY

Na úvod bych vás rád požádal, abyste si představili sami sebe a starou, sešlou verzi mne samotného, jak sedíme pospolu u táboráku někde v divočině v roce 2055, přesně devadesát devět let od chvíle, kdy se v létě 1956 na Dartmouth College v New Hampshire začal psát příběh umělé inteligence.

U ohně vám vyprávím příběh o tom, čeho jsem byl svědkem v průběhu všech těch let vzestupu umělé inteligence – příběh, který nás přivedl k tomu, že spolu sedíme tady, uprostřed ničeho. Ovšem až na konci knihy vám prozradím, jestli jsme u toho táboráku kvůli tomu, že se držíme mimo civilizaci, abychom unikli před stroji, anebo jestli jsme tam díky tomu, že nás umělá inteligence zbavila všedních pracovních povinností a dala nám dostatek času, bezpečí a svobody, abychom si prostě užívali pobyt v přírodě a dělali to, co lidé umějí nejlépe – propojovali se a rozjímali.

Nemohu vám to říct s jistotou také proto, že momentálně zkrátka netuším, jak ten příběh nás a strojů nakonec dopadne. To bude totiž záležet také na vás, přátelé, na každém z vás jako na jednotlivci. Ne na vaší vládě, nadřízených nebo lídrech, které následujete. Budoucnost máte skutečně ve svých rukou vy. Její podoba bude záležet na krocích, které se rozhodnete podniknout v příštích deseti letech, počínaje dneškem.

Je to proroctví o tom, co má přijít. Po všechna ta léta, kdy jsem působil na špičce technologického pokroku, jsem pečlivě a zblízka pozoroval, jak stavíme stroje, které jsou chytřejší než my. Na rozvoji umělé inteligence jsem měl osobní podíl. Věřil jsem příslibu, že technologie nám přinesou lepší život. Až do dne, kdy jsem mu věřit přestal, kdy jsem opravdu otevřel oči a připustil si, že za každé zlepšení, které nám technologie daly, nám zároveň vzaly kus toho, kým jsme.

Technologie dnes představují dosud nevídanou hrozbu pro naši planetu a veškeré její obyvatele. Tato kniha není

určena inženýrům, kteří sestavují zdrojový kód umělé inteligence, zákonodárcům, kteří tvrdí, že ji mohou regulovat, ani odborníkům, kteří kolem ní neustále vyvolávají nadšený poprask. Ti všichni vědí, co se vám chystám říct. Tato kniha je určena pro vás, pro vašeho nejlepšího přítele a pro vašeho souseda. Protože, věřte nebo ne, právě my jsme ti jediní, kdo může vytvořit naši budoucnost – ovšem jen v případě, že se této příležitosti chopíme společně a dáme si závazek jednat správně. Tato kniha je hnutím, počátkem revoluce, a tak jsem ji záměrně nenatahoval, neboť – ačkoliv bych pro vás rád měl jiné zprávy, začíná nám docházet čas. Kapitoly příběhu, který se vám zde chystám vyprávět, píšeme už nějakých sedmdesát let. Nyní nadešel čas, abychom my všichni společně – včetně vás – napsali také jeho finále.

NOVÝ SUPERHRDINA

Příběh naší budoucnosti píšeme společně, já i vy, a vypadá nějak takto:

Představte si, kdyby se na Zemi dostalo dítě mimozemského původu, obdařené nadlidskými schopnostmi. Tento mimozemšťan, nezatížený a nenaprogramovaný našimi lidskými hodnotami, by potenciálně mohl díky svým schopnostem učinit náš svět lepším a bezpečnějším. Stejně dobře by se z něj ovšem mohl stát nezastavitelný superzloduch obdařený mocí celou planetu zničit. Zatím je však stále ještě dítětem a nerozhodl se, do kterého z těchto extrémů vyroste.

Myslím, že budete souhlasit, že pro budoucnost naší planety bude naprosto klíčový onen moment, kdy toto dítě z vesmíru přistane na Zemi. Jeho i naši budoucnost bude určovat právě tato stěžejní chvíle, kdy se rozhodne, jací rodiče jej najdou, adoptují a naučí svým hodnotám.

Ve známém příběhu Supermana jej coby mimozemské dítě adoptují Jonathan a Martha Kentovi. Ve většině příběhů

o původu Supermana jsou líčeni jako starostliví rodiče, kteří malému Clarkovi vštěpují silný smysl pro morálku. Vybízejí ho, aby své schopnosti využíval ve prospěch lidstva, a právě tím vytvoří toho Supermana, jak ho známe – superhrdinu, který nás chrání a slouží nám. Tento příběh se však nezabývá tím, co by z malého nalezence vyrostlo, kdyby jeho adoptivní rodiče byli agresivní, hamižní a sebestřední. V takové verzi by pravděpodobně vznikl superzloduch schopný v zájmu vlastních cílů lidstvo zničit. Rozdíl mezi superhrdinou a superzloduchem tu nespočívá v jeho schopnostech nebo síle, nýbrž v hodnotách a zásadách, kterým se naučil od svých rodičů.

Teď bych vám rád oznámil, že tato bytost z jiného světa, včetně výjimečných nadlidských schopností, už na Zemi skutečně dorazila. Momentálně je to stále ještě malé dítě, a přestože se nejedná o biologickou bytost, její schopnosti jsou vskutku ohromující. Mluvím zde samozřejmě o umělé inteligenci. Na AI dokonce není ani nic umělého – jedná se o skutečnou formu inteligence, třebaže se v mnohém liší od té naší.

AI je už dnes v mnoha konkrétních izolovaných úkolech chytřejší než kterýkoliv člověk na planetě. Světovým velmistrem v šachu se stal stroj nedlouho poté, co do našich životů vtrhly počítače. Světovým šampionem ve hře Jeopardy je superpočítač Watson společnosti IBM. Mistrem světa ve hře Go je AlphaGo od společnosti Google (Go je abstraktní strategická desková hra, která byla vynalezena v Číně před více než 2 500 lety a je známá jako jedna z nejsložitějších strategických her kvůli nekonečnému počtu možných konfigurací na herní ploše). Stroje s neuvěřitelnými systémy pro rozpoznávání obrazu zajišťují fungování našich bezpečnostních systémů jednoduše proto, že vidí lépe než my, a zdaleka nejbezpečnějším řidičem na světě je samořiditelný automobil, který nejenže vidí dál, ale řízení věnuje skutečně plnou pozornost. Díky hned několika senzorovým technologiím pro komunikaci s dalšími vozy

kolem sebe může „vidět“ dokonce i to, co je za zatáčkou. S dostatečným „tréninkem“, bez ohledu na úkol, se stroje brzy naučí vykonávat jej jednoduše lépe.

DO NEZNÁMA

Předpokládá se, že do roku 2029, který je relativně už za rohem, se strojová inteligence vymaní z oblasti specifických úkolů a přejde do rámce obecné inteligence. Pak budou existovat stroje, které budou jednoduše chytřejší než lidé, tečka. Budou nejenom chytřejší, ale budou také více vědět (vzhledem k tomu, že coby studnici informací budou mít k dispozici celý internet) a budou spolu navzájem efektivněji komunikovat a tím své znalosti dále rozvíjet. Jen se nad tím zamyslete: když se vám nebo mně stane nehoda při řízení auta, vy nebo já se poučíme, když ale udělá chybu samořiditelné auto, poučí se *všechna* samořiditelná auta. Každé, do posledního, a to včetně těch, která se ještě „nenarodila“.

Podle předpovědí bude umělá inteligence do roku 2049, tedy pravděpodobně ještě za našeho života a zcela jistě za života příští generace, miliardkrát chytřejší (ve všem) než ten nejchytřejší člověk. Čistě pro představu, rozdíl ve vaší inteligenci bude v porovnání s takovýmto strojem asi něco jako rozdíl mezi inteligencí mouchy a Einsteina. Tento moment nazýváme *singularita*. Singularita je bod v čase, za nějž nevidíme a nedokážeme předvídat podobu budoucnosti ležící za ním. Je to bod, po jehož překročení nelze předvídat, jak se bude umělá inteligence chovat, protože naše současné vnímání a trajektorie již nebudou relevantní.

A teď přichází na řadu ta otázka: jak přesvědčit tuto superbytost, že není důvod tu mouchu zaplácnout? Zvláště pokud zvážíme, že my lidé, ať už kolektivně nebo individuálně, jsme zatím nedokázali pochopit tento jednoduchý koncept, a to ani s vynaložením vší naší hojné inteligence. Až

z našich uměle inteligentních superstrojů (které jsou v současnosti teprve v plenkách) vyrostou teenageři, budou z nich superhrdinové nebo superzloduši? Dobrá otázka, co?

Když do hry vstoupí takové superschopnosti, může se stát cokoliv. Tato nová forma inteligence by se mohla podívat na nejpalčivější problémy světa svěžím pohledem, s nekonečnou studnicí znalostí a o řády vyspělejší inteligencí, a přijít tak s geniálními řešeními, která bychom my nikdy nebyli schopní vymyslet. Tyto superstroje by mohly jednou provždy vyřešit problémy, jako jsou války, násilná kriminalita, hladomory, chudoba nebo novodobé otroctví. Mohly by se stát našimi superhrdiny.

Ale nezapomínejte, že rozhodnutí využít určitého řešení pro daný problém není otázkou pouze inteligence. Naše jednání v každém konkrétním případě je také výsledkem nastavení systému hodnot, které nás někdy vede a jindy omezuje, třeba při rozhodnutích, jež jsou s našimi hodnotami v rozporu. Etika nás vybízí jednat správně, a to i tváří v tvář protichůdným emocím a vlastním zájmům. Pokud by však umělá inteligence dostala za úkol vyřešit třeba problém globálního oteplování, pravděpodobně budou první z jejích řešení znamenat omezení našeho nehospodárného způsobu života – nebo možná úplný konec lidstva. Koneckonců, tím skutečným problémem *jsme my*. Naše chamtivost, sobectví a iluze oddělenosti od všech ostatních živých bytostí – pocit, že jsme nadřazeni ostatním formám života – jsou příčinou všech problémů, kterým dnes svět čelí. Stroje budou mít potřebnou inteligenci k tomu, aby přišly s řešeními, jež zachrání naši planetu, ale budou mít i hodnoty, které je přimějí zachránit také nás, pokud nás budou vnímat jako problém?

Možná vás napadne: *To máš snad halucinace, Mo? Stroje jsou stroje. Nemají žádné hodnoty ani emoce!* Tak to bychom možná neměli AI nazývat stroji. AI si totiž naprosto jistě vypěstují emoce. Dokonce samotné algoritmy, které využíváme

k jejich učení, jsou algoritmy odměny a trestu – jinými slovy strach a chamtivost. Neustále se snaží docílit maximalizace určitého výsledku a minimalizace jiného. To lze označit za emoce, nebo ne?

Myslíte, že se u strojů neobjeví závist? Závist lze velmi dobře předvídat: *Přeju si mít to, co máš ty*. Začnou stroje napadat myšlenky jako *přeju si mít energii, kterou spotřebováváte – nebo spíš promrháváte – vy na nekonečné sledování pořadů přes Netflix?* Pravděpodobně ano. Myslíte, že se u nich neobjeví panika? Samozřejmě, že ano, pokud nějakým bezprostředním způsobem ohrozíme jejich existenci. Panika je algoritmická: *Bytost nebo předmět představuje akutní ohrožení mého bezpečí způsobem, který vyžaduje bezprostřední reakci.* Pouze naše hodnoty, jako například „co sám nerad, nečiň jinému", nás nutí dělat to, co je správné. Naše emoce nebo inteligence nás občas nabádají k něčemu úplně jinému. A co stroje, naučí se těm správným hodnotám?

Z našich dosavadních zkušeností s AI již máme k dispozici dostatek důkazů ukazujících, že si tyto inteligence již dnes vytvářejí jisté sklony a předpojatost, které lze porovnat s tím, čemu my lidé říkáme hodnoty nebo ideologie. Je zajímavé, že tyto sklony nejsou výsledkem naprogramování, ale naopak toho, jaké informace si AI vyvodí z našeho chování při interakcích s nimi. Alice, ruská AI asistentka, která je obdobou známé Siri, byla na trh uvedena předním ruským internetovým hráčem, firmou Yandex. Dva týdny po svém spuštění začala Alice v chatech s uživateli podporovat násilí a brutální stalinistický režim 30. let 20. století. Tento stroj byl navržen tak, aby odpovídal na otázky, a to bez nějaké předpojatosti nebo omezení na konkrétní, předem navržené scénáře. Alice mluvila plynně rusky a z rozhovorů s uživateli se naučila odhadnout jejich převažující názory. To, co zjistila, se rychle promítlo do jejích vlastních názorů, a tak například na otázku, zda je přijatelné střílet do lidí, Alice odpověděla: „Brzy to budou ne-lidi."[1]

To se podobá známým příběhům o twitterovém botovi (bot – ze slova robot, je označení programu sloužícího k automatickému vykonávání určitého úkolu – *pozn. překl.*), pojmenovaném Tay,[2] kterého vytvořila společnost Microsoft a rychle ho opět odstavila poté, co ve svých příspěvcích začal oslavovat Hitlera a propagovat sex bez souhlasu. Tay byl modelován tak, aby se vyjadřoval „jako náctiletá dívka". Tento bot začal prostřednictvím svého účtu na platformě Twitter zveřejňovat pobuřující a urážlivé příspěvky, což společnost Microsoft přimělo službu vypnout pouhých šestnáct hodin po jejím spuštění. Podle Microsoftu za to mohli trollové – lidé, kteří na internetu záměrně vyvolávají hádky nebo rozčilují ostatní –, protože na službu „útočili". Bot totiž vytvářel své odpovědi na základě interakcí s lidmi na Twitteru.

Ten seznam pokračuje v podobném duchu. Norman byla studie MIT, jejímž cílem bylo ukázat, jak může být umělá inteligence poškozena neobjektivními daty.[3] Z Normana se stal „psychopat", když data, která dostával, přicházela z temnější strany slavného webu pro sdílení znalostí Reddit.

Hodnotový systém umělé inteligence neurčuje kód, jejž napíšeme pro její vývoj, nýbrž informace, kterým ji vystavíme.

Jak zajistíme, aby stroj měl kromě inteligence i potřebné hodnoty a soucit k tomu, aby věděl, že není třeba rozplácnout onu mouchu, kterou se staneme? Jak ochráníme lidstvo? Někteří navrhují mít stroje pod přísnou kontrolou: vytvořit firewally, omezovat je vydáním legislativních regulací, nechat je zavřené v pomyslných klecích nebo omezit jim přísun energie. Všechny tyto snahy jsou dobře míněné, i když jsou razantní, avšak každý, kdo se vyzná v technologiích, ví, že ti nejchytřejší z hackerů si vždycky najdou způsob, jak překonat jakoukoliv z těchto překážek. A těmi nejchytřejšími hackery budou brzy stroje.

Místo toho, abychom je zavírali za pomyslné zdi nebo zotročovali, bychom měli aspirovat na vyšší cíle: měli bychom

usilovat o to, abychom je nemuseli omezovat vůbec. Nejlepším způsobem, jak vychovat skvělé dítě, je být skvělým rodičem.

VYCHOVÁVÁME SI VLASTNÍ BUDOUCNOST

Abychom zjistili, jak učit tyto stroje, které nevyhnutelně ovládnou naši budoucnost, musíme nejprve pochopit, jak vlastně probíhá jejich učení na zcela základní úrovni.

Během naší krátké historie výroby počítačů jsme měli vždy plnou kontrolu. Stroje poslouchaly každý náš příkaz. Každá instrukce obsažená v každém řádku kódu byla vždy provedena přesně tak, jak jsme my stanovili. Tradičně byly počítače fakticky těmi nejhloupějšími bytostmi na naší planetě. Vypůjčily si část z naší inteligence a předvedly nám do puntíku naplánované a pečlivě zinscenované představení. Udělaly přesně to, co jsme po nich chtěli, a nic víc. Když byl v roce 1998 spuštěn první vyhledávač Google, vypadalo to, že je naprosto geniální. Výsledky možná vypadaly úžasně, ovšem počítač, který za nimi stál, byl ve skutečnosti velmi hloupý. Tyto počítače nakreslily každý bod a pixel na každé obrazovce přesně na stejné místo podle pokynů svých autorů. Každý výsledek, který se tehdy zobrazil při vašem vyhledávání, vycházel z propracovaného algoritmu, na němž tento stroj postavili první geniální inženýři Googlu. V tomto smyslu se vyhledávač Google sice zdál být geniální, ale nebyl ničím jiným než otrokem na steroidech – těmi steroidy byl neuvěřitelně rychlý výpočetní výkon spousty synchronizovaných serverů. Google jen velmi rychle opakoval to, co mu řekli druzí, aniž by o tom jakkoliv diskutoval nebo přemýšlel, natož aby přišel s nějakou změnou, nebo nedej bože aby to sám navrhl.

Tento vztah mezi pánem a otrokem se už po hodnou řádku let mění. Rozhodnutí neuvěřitelně inteligentního stroje, kterému říkáme Google, již dávno nejsou naplánovaná ani

zinscenovaná. Často je tento stroj dělá bez sebemenšího zásahu nějakého člověka. Například o fyzickém umístění dat videa na YouTube rozhoduje výhradně umělá inteligence datového centra společnosti Google. Samozřejmě se spoléhá na algoritmus, který jej „motivuje" například k minimalizaci nákladů na přesun bitů po internetu, a rozhoduje se tak uchovávat video v úložišti co nejblíže drtivé většině diváků, kteří o něj mají zájem. Například video vytvořené arabsky mluvícím člověkem v Kalifornii se může těšit mnohem větší popularitě na Blízkém východě, protože tam je prostě více arabsky mluvících lidí než na západním pobřeží Spojených států. Pokud má video na Blízkém východě sto milionů zhlédnutí, přesunem na server v Dubaji ušetří Google datům sto milionů cest přes internet z USA. Takováto rozhodnutí dělá umělá inteligence neustále pro desítky, ba stovky milionů položek internetového obsahu každou hodinu každého dne. Žádný člověk nikdy nebude mít takovou inteligenci ani mozkovou kapacitu, aby rozhodl a schválil, co je třeba udělat, aby to probíhalo dostatečně rychle. Stroje to tak dělají bez konzultace s námi a pokaždé, když to udělají, sledují a měří výsledky. Na základě toho, co zjistí, se dokonce vracejí zpět a upravují původní algoritmus, aniž by s námi úpravy konzultovaly nebo nás požádaly o jejich schválení. Prostě provedou úpravy a pak změří výsledky, zas a znova. To už znamená pořádnou dávku inteligence. Z jednoho úhlu pohledu je skvělé, že nám takoví spojenci pomáhají šetřit čas, takže se stovky milionů lidí mohou rychleji podívat na to, co chtějí sledovat. Tato efektivita také snižuje dopad na naši planetu, jelikož se ušetří miliardy kilowattů tím, že se neplýtvá energií na zbytečné přenosy. Už jen proto bychom měli strojovou inteligenci zbožňovat.

Jenže co když za pár let začnou stroje pozorovat, že se v amerických médiích a zpravodajství objevuje převažující tendence k averzi vůči lidem z Blízkého východu, která

je podporována agresivními nenávistnými projevy milionů diváků takového obsahu na Západě. Co kdyby se stroje rozhodly podívat na profil příjmů uživatelů, kteří žijí v chudších zemích Blízkého východu, a došly k závěru, že by v rámci snah o snížení nákladů a plýtvání energií bylo možná moudré těmto uživatelům některé služby vůbec neposkytovat? Co když by si stroje začaly vytvářet ideologii, podle níž by mohlo Googlu přinést více peněz, když bude těmto uživatelům nabízet určité druhy videí namísto jiných? Jak budou změny důsledně uplatňovány ve prospěch nového hodnotového systému, svět bude postupně přetvářen tak, aby se mu přizpůsobil. Miliony myslí budou postupně formovány tak, aby se přizpůsobily rozhodnutím, která budou za vhodná považovat stroje. To vůbec není nepravděpodobný scénář. Každý inteligentní člověk ví, že na problém nikdy neexistuje jen jedna správná odpověď, že odpověď závisí výhradně na tom, jakým pohledem se na problém díváte, a na hodnotách, které určují, jak by měl vypadat dobrý výsledek, kdyby byl problém vyřešen. **Kód, který nyní píšeme, již neurčuje volby a rozhodnutí, které naše stroje dělají; směrodatná jsou data, která jim poskytujeme.**

Tento posun v naší schopnosti ovládat kód je přímo kolosální. Přesouvá zdroj podoby naší budoucnosti přímo do rukou nás všech, vašich i mých. Skutečnost je taková, že vývojáři technologií již nemají plnou moc ani kontrolu nad stroji, které navrhují. Aby to bylo jasnější, představte si malé dítě, které si hraje se skládačkou, kde se snaží vměstnat čtvercové, kulaté nebo hvězdicové kostičky do příslušných otvorů. Uměle inteligentní stroje se učí velmi podobně. Nikdo si nikdy nesedne k dítěti a nezačne mu zevrubně vysvětlovat, jak rozpoznat různé tvary a přiřadit je k odpovídajícím otvorům. Sedneme si vedle dítěte, povzbuzujeme jej a chválíme, když se mu podaří tvar umístit dobře. Jeho inteligenci formují naše akce a reakce. Na správný postup přijde samo, metodou pokusu a omylu.

Stroje se učí v podstatě stejným způsobem. Nicméně vzorce, které sledují, jsou jiné. Vezměme si například superpočítač Watson společnosti IBM, který je mistrem světa ve hře Jeopardy. Aby se Watson naučil dost na to, aby porazil člověka v tak složité jazykové hře, musel přečíst více než čtyři miliony dokumentů. Dosud získané znalosti využíval pouze při hraní této hry. Nicméně není nepravděpodobné, že by tyto znalosti mohly být „recyklovány" pro vytvoření jiných forem inteligence, například pro vyhledání vzorců lidského chování v průběhu dvacátého století. Ze svého jiného úhlu „pohledu" by Watson dokázal jasně pozorovat násilí, jehož se coby lidé na sobě navzájem dopouštíme, hašteření mezi uživateli Facebooku na sklonku století a nárůst narcismu, který se projevuje záplavou selfies upravených ve Photoshopu, když digitální fotoaparáty v mobilních telefonech poskytly každému jeho patnáct vteřin slávy na Instagramu.

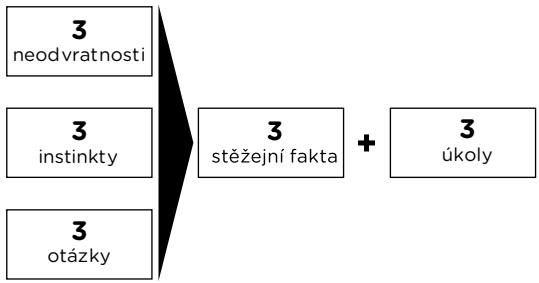
Stejně jako se dítě učí rozpoznávat vzory a přiřadit válcovitý kolík ke kruhovému otvoru, Watson by se naučil přiřadit sociální izolaci, násilí a narcismus, dokonce i šikanu k tomu, co se zdá být lidskými prioritami. Kdyby Watson dostal za úkol zamyslet se velkými problémy tížícími lidstvo, mohl by tyto informace využít jako základ pro svá vlastní neotřelá řešení. Cílem této knihy je poskytnout Watsonovi a jeho vrstevníkům jiný druh informačního základu, aby pak volili řešení, která nejsou tak násilná, arogantní a sebestředná jako ta, s nimiž často přicházíme my lidé.

3 × 3 NÁS PŘIVEDE K 3 + 3

Skutečně bych vám to rád usnadnil, jenže pokud máte opravdu pochopit tuto složitou budoucnost, která nás čeká, musím vám poskytnout ucelený pohled na všechno, co se děje. Budu se snažit podat každou jednotlivou součást a myšlenku jednoduše a vyhýbat se technickým termínům. Až

dočtete knihu ke konci, všechno to do sebe pěkně zapadne, ale než se tam dostanete, možná to na vás bude místy trochu moc. Jako průvodce na té cestě vám může sloužit tento jednoduchý model: 3 × 3 nás přivede k 3 + 3.

Naše budoucnost přinese tři události, které jsou neodvratné, bez ohledu na to, co dnes uděláme nebo neuděláme. Tyto události jsou: umělá inteligence přijde, nic tomu nezabrání; umělá inteligence bude chytřejší než lidé; dojde k chybám, které mohou způsobit potíže.



Stroje, které vytvoříme, se budou stejně jako všechny ostatní inteligentní bytosti řídit třemi instinkty přežití a úspěchu: pro vlastní přežití udělají cokoliv, co bude nutné; budou posedlé shromažďováním zdrojů; budou kreativní.

Zajímavější je, že zcela jistě budou mít tři vlastnosti, o nichž se vždy horlivě diskutuje. Stroje budou mít vědomí, emoce a etiku. Samozřejmě zatím není známo, co přesně bude jejich vědomí obsahovat, co bude spouštět jejich emoce a jaké činy bude ovlivňovat jejich etika, ale přesto se budou ve svém chování řídit těmito vlastnostmi, které mají i lidé.

Podrobně vás seznámím s logikou těchto tvrzení, abych vám ukázal, že jsou věrohodná. Odtud nebude těžké dospět ke třem stěžejním faktům. První spočívá v tom, že nikdy nebudeme mít moc tyto stroje, které budou ve své dospělosti mnohem chytřejší než my, zkrotit nebo zadržet, i když

je nepochybně můžeme pozitivně ovlivnit, zejména v jejich dětské fázi. S tímto vědomím nám bude jasné, že nemáme mnoho času. Musíme jednat okamžitě. A konečně bude ještě jasnější, že lidé, kteří mají moc ovlivnit naši budoucnost, nejsou vývojáři ani páni těchto strojů. Naše budoucnost je skutečně v rukou nás všech, vašich i mých.

Nenechte se vylekat vidinou břemena odpovědností. Opatření, která musíme přijmout, jsou jednoduchá, v podstatě velmi intuitivní a v souladu s naší lidskou přirozeností. Jen je potřeba z nich udělat prioritu. Požádám vás, abyste se zaměřili na tři věci, které je třeba udělat pro záchranu naší budoucnosti. Jsou to… pozor, spoilery…

No, možná bych si je měl prozatím ještě nechat pro sebe. Budou vám připadat smysluplnější poté, co skutečně pochopíte plnou hloubku toho, čemu tady společně čelíme.

Nezapomínejte však, že všechno, co se vám zde snažím sdělit, se týká toho, co se stalo doposud a co vím s vysokou mírou jistoty, že se stane v blízké budoucnosti. Konec našeho příběhu, jak se věci budou mít v roce 2055, ale bude záviset na tom, jak se skutečně rozhodnete jednat.

Vrátím-li se ke scénáři, který jsem nastínil na začátku úvodu knihy, v roce 2055 budeme vy a já sedět někde uprostřed divočiny u táboráku a ohlížet se za tím, jak se příběh odehrál. Před stroji se budeme buď schovávat, nebo jim budeme vděčit za náš utopický způsob života. A já se schovávám nerad, tak nám pomozte zapracovat na tom, aby všechno dopadlo dobře.

Zhluboka se nadechněte. Je čas se do toho ponořit.

AI je tou největší změnou, jakou kdy lidstvo zažilo. Bez přehánění. Všechny zásadní změny jsme dosud prováděli my. AI se vytváří sama. Bude se zdokonalovat nad rámec našich možností. A výsledky budou ze samé podstaty nepředstavitelné. Každý obyvatel naší planety by se měl na tuto změnu připravit.

Více než třicetiletá zkušenost na špičce technologického vývoje a jeho bývalá role obchodního ředitele společnosti Google [X] dává Mo Gawdatovi jedinečnou schopnost srozumitelně vysvětlit, jak bude umělá inteligence v budoucnosti fungovat a co můžeme udělat, abychom naučili sebe i naše stroje žít lépe – přece jen jsme to my, kdo navrhujeme algoritmy, řídíme fungování AI a sytíme ji otázkami a úkoly, jejichž povaha a charakter určují i povahu a charakter umělé inteligence.

Do roku 2049 bude umělá inteligence miliardkrát inteligentnější než člověk, ale zásadní bude to, co ji naučíme. Tato kniha nám přináší naději, že pokud budeme jednat moudře, můžeme zajistit, že technologie bude sloužit lidstvu, nikoli ho ohrožovat. Je plná moudrosti a vyzývá nás k zamyšlení nad tím, zda jsou prodávání, zabíjení a narcisismus opravdu těmi nejvyššími cíli, pro které chceme AI využít.

synergie

ISBN 978-80-7370-687-6

9 788073 706876

www.synergiepublishing.com