KAPITOLA 3.

BLÍZKÁ BUDOUCNOST: PRŮLOMY, CHYBY, ZÁKONY, ZBRANĚ A PRÁCE

Jestliže brzy nezměníme směr, skončíme tam, kam míříme.
Irvin Corey

Co znamená být člověkem v současné době? Čeho si například na sobě skutečně vážíme a co nás odlišuje od jiných forem života a od strojů? Čeho si na nás cení ostatní lidé, že nám jsou ochotní nabídnout práci? Ať už znějí naše odpovědi na tyto otázky jakkoli, není pochyb o tom, že je vzestup technologie postupně změní.

Vezměte si třeba mě. Jako vědec jsem pyšný na to, že si své cíle určuji sám, že se pomocí kreativity a intuice vypořádávám se širokým spektrem nevyřešených problémů a že svá zjištění sdílím pomocí jazyka. Naštěstí pro mě je společnost ochotná mi za to platit jako za práci. Před staletími bych možná stejně jako ostatní stavěl svou identitu na tom, že jsem farmář nebo řemeslník, dnes ovšem kvůli vzestupu technologií takové profese představují jen nepatrný zlomek pracovního trhu. To znamená, že ne každý už může svou identitu formovat okolo zemědělství a řemesel.

Mě samotného neznepokojuje, že stroje jsou dnes o třídu lepší než já v manuálních dovednostech, jako je kopání a pletení, protože ty nepatří ani mezi moje koníčky, ani mezi moje zdroje příjmu, a netvoří tak součást mého pocitu sebeúcty. Jakékoliv falešné představy, které jsem o svých schopnostech mohl v tomto ohledu mít, se mi zhroutily v osmi letech. Tehdy mě škola donutila chodit na pletení, ze kterého bych býval propadl, kdyby se nade mnou jeden páťák nesmiloval a nepomohl mi.

Jak se ale budou technologie vylepšovat, nezastíní nakonec umělá inteligence i ty schopnosti, které mi nyní dodávají sebeúctu a cenu na trhu práce? Stuart Russell se mi svěřil, že on a řada jeho kolegů nedávno prožili okamžiky, kdy nevěřili vlastním očím, neboť viděli AI zvládnout věci, které by od ní čekali až za mnoho let. Proto vám povím o několika momentech, kdy jsem něco podobného zažil i já. Pokládám je za předzvěsti toho, že lidské schopnosti budou brzy překonány.

PRŮLOMY

AGENTI SE ZPĚTNOVAZEBNÍM HLUBOKÝM UČENÍM

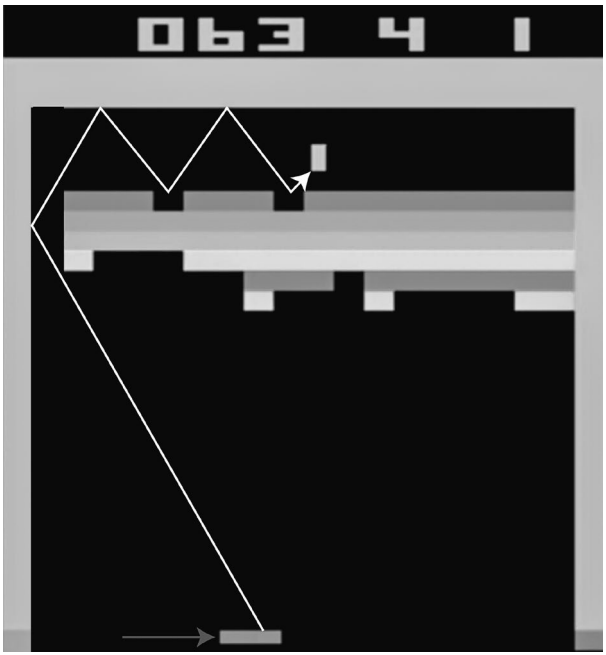
Jeden z momentů, kdy jsem zůstal stát s otevřenou pusou, jsem zažil v roce 2014, když jsem si pouštěl video, na němž se systém AI DeepMind učí hrát počítačové hry. Umělá inteligence tam hrála Breakout (obrázek 3.1), klasickou Atari hru, na niž mám vzpomínky z doby, kdy jsem byl teenager. Cílem je pohybovat s pálkou, aby opakovaně odrážela míček do řady cihel; cihla po každé trefě zmizí a vám se zvýší skóre.

Svého času jsem nějaké počítačové hry napsal a dobře si uvědomuji, že napsat program, který umí hrát Breakout, není těžké – ovšem tým DeepMindu udělal něco jiného. Vytvořili umělou inteligenci, která byla úplně nevědomá; nevěděla nic o této hře ani o jiných hrách, dokonce neměla představu ani o *pojmech* hra, pálka, cihla nebo míček. Znala jen dlouhý seznam čísel, jimiž ji v pravidelných intervalech krmili: současné skóre a dlouhou řadu čísel, která bychom my (ale nikoli AI) rozpoznali jako údaje o tom, jak jsou které části obrazovky obarveny. AI dostala jednoduše pokyn, aby maximalizovala skóre tím, že bude v pravidelných časových intervalech vysílat nějaké signály, které bychom my (ale nikoli ona) rozpoznali jako kódy, kterou klávesu zmáčknout.

Zpočátku hrála umělá inteligence příšerně: bezradně pohybovala pálkou tam a zpátky, vypadalo to, že zcela nahodile, a míček skoro pokaždé minula. Po chvíli jako by začínala chápat, že je dobrý nápad pohybovat pálkou směrem k míčku, přestože i nadále většinou minula. Cvikem se ale zlepšovala a zanedlouho jí to šlo lépe než kdy mně a neomylně odpalovala míček, ať se přibližoval seberychleji. A pak mi poklesla čelist. AI objevila úžasnou strategii pro nejvyšší skóre: nejprve soustavně mířit na levý horní roh, aby si prorazila díru skrz, a pak nechat míček, aby se odrážel mezi zadní řadou a zdí. Působilo to jako hodně chytrý tah. Demis Hassabis mi později sdělil, že programátoři z týmu DeepMindu ten trik neznali a naučili se ho až od své AI. Podívejte se na to video, odkaz je připojen.[1]

Cosi v tom připomínalo člověka, a právě to mě trochu znepokojovalo: sledoval jsem umělou inteligenci, která má cíl a učí se, jak ho dosahovat stále lépe, až nakonec překoná své stvořitele. V minulé kapitole jsme inteligenci definovali jako schopnost dosahovat komplexních cílů. V tomto smyslu se AI DeepMindu stávala před mými zraky stále inteligentnější (třebaže jen ve velmi úzkém kontextu hraní této konkrétní hry). V první kapitole jsme se setkali s něčím, co informatici nazývají *inteligentní agent*: entita, která sbírá informace o svém prostředí ze senzorů a tato data pak zpracovává, aby se rozhodla, jak se v tomto prostředí zachovat. Přestože AI DeepMindu existovala v extrémně jednoduchém virtuálním světě z cihel, pálek a míčků, nemohl jsem popřít, že je inteligentním agentem.

Lidé od DeepMindu svou metodu brzy publikovali a podělili se o její kód. Vysvětlili, že používá velmi jednoduchý, ale mocný koncept nazývaný *zpětnovazební*



Obrázek 3.1: Poté, co se pomocí zpětnovazebního hlubokého učení AI od začátku učila hrát Atari hru Breakout tak, aby maximalizovala skóre, objevila optimální strategii: provrtat díru podél okraje řady cihel a nechat míček, aby se odrážel za ní a rychle nahromadil body. Šipky označují předchozí pohyby míčku a pálky.

hluboké učení.[2] Obyčejné zpětnovazební učení je klasická technika strojového učení inspirovaná behavioristickou psychologií, kde odměna zvyšuje vaši motivaci něco udělat znovu – a naopak. A stejně jako se pes naučí poslouchat povely, když to zvýší pravděpodobnost, že od svého pána dostane pamlsek, tak se AI DeepMindu naučila pohybovat pálkou, aby odrazila míček, protože tím zvýšila pravděpodobnost bodového zisku. V DeepMindu tento princip zkombinovali s hlubokým učením: vytrénovali hlubokou neuronovou síť (jak jsme viděli v předchozí kapitole), aby předpovídala, kolik bodů přinese zmáčknutí každé z povolených kláves na klávesnici, a aby pak zvolila tu, kterou v daném okamžiku hry vyhodnotí jako nejslibnější.

Do svého výčtu hodnot přispívajících k mému pocitu vlastní ceny jako člověka jsem zařadil schopnost vypořádat se se širokým spektrem nevyřešených problémů. Umět hrát Breakout a nedokázat nic jiného naopak znamená extrémně slabou, úzce zaměřenou inteligenci. Pro mě osobně spočívá přelomovost DeepMindu v tom, že zpětnovazební hluboké učení je zcela obecná technika. Nechali tutéž umělou inteligenci cvičit hraní devětačtyřiceti různých Atari her a ona se naučila porazit své lidské testery ve dvaceti devíti z nich – od Pongu po Boxing, Video Pinball a Space Invaders.

Netrvalo dlouho a tatáž myšlenka umělé inteligence se začala osvědčovat na dalších moderních hrách s třírozměrnými, nikoli dvojrozměrnými světy. Konkurence DeepMindu se sídlem v San Franciscu, společnost OpenAI, záhy představila platformu jménem Universe, kde AI DeepMindu a další inteligentní agenti mohou trénovat interakce s celým počítačem, jako by se jednalo o hru: klikat na cokoli, libovolně psát, otevírat a spouštět jakýkoli software, který dokážou ovládat – například otevřít prohlížeč a brouzdat po internetu.

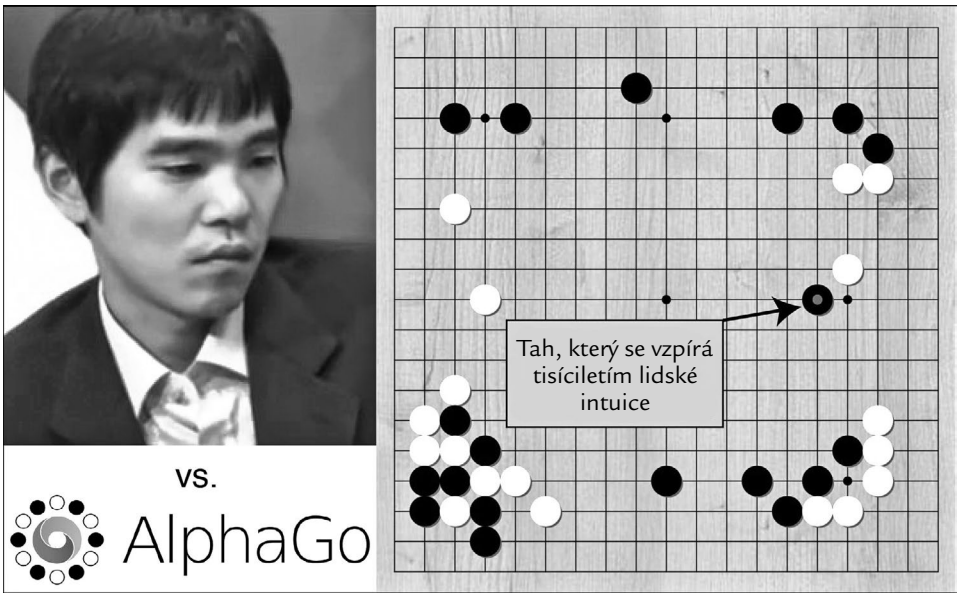
Při pohledu do budoucnosti zpětnovazebního hlubokého učení a jeho vylepšování není v dohledu žádný zřejmý konec. Jeho potenciál se neomezuje na virtuální herní světy, protože když jste robot, můžete za hru považovat život samotný. Stuart Russell mi jednou řekl, že poprvé zažil naprostý údiv tehdy, když sledoval robota Big Dog, jak vybíhá do sněhem pokrytého lesního svahu a elegantně řeší problém pohybu po končetinách, s nímž Russell zápolil celá léta.[3] Když tohoto mezníku nakonec v roce 2008 dosáhl, bylo za ním obrovské množství práce dobrých programátorů. Po průlomu, který přinesl DeepMind, neexistuje jediný důvod, proč by robot nemohl použít nějaký druh zpětnovazebního hlubokého učení, aby se naučil chodit bez lidských programátorů: potřebuje jen systém, který mu dá za odměnu body, kdykoli učiní nějaký pokrok. Roboti ve skutečném světě mají podobně potenciál plavat, létat, hrát stolní tenis, bojovat a projevovat takřka nekonečný počet dalších motorických dovedností bez pomoci lidských programátorů. Aby se celý proces urychlil a snížilo se nebezpečí, že se roboti během učení poškodí, první kroky svého učení budou pravděpodobně absolvovat ve virtuální realitě.

INTUICE, KREATIVITA A STRATEGIE

Dalším rozhodujícím momentem pro mě bylo, když systém AI DeepMindu jménem AlphaGo zvítězil v pětikolovém zápasu v go proti Leemu Sedolovi, považovanému za celosvětovou špičku mezi hráči počátku jednadvacátého století.

Všeobecně se očekávalo, že lidští hráči go budou dříve nebo později sesazeni z trůnu stroji, jelikož přesně to se stalo před dvaceti lety jejich kolegům šachistům, nicméně většina znalců go předpovídala, že to bude trvat aspoň deset let. Triumf AlphaGo tak byl klíčovým momentem i pro ně. Jak Nick Bostrom, tak Ray Kurzweil zdůrazňují, jak obtížně se dá jakýkoli průlom v AI předvídat. Je to koneckonců zřejmé z výroků samotného Leeho Sedola, které postupně pronášel do doby, než prohrál první tři hry:

- Říjen 2015: „Podle toho, co jsem viděl o jeho úrovni, ... myslím, že tu hru skoro najisto drtivě vyhraji."
- Únor 2016: „Doslechl jsem se, že umělá inteligence od Google DeepMindu je překvapivě silná a že je stále silnější, ale věřím, že alespoň tentokrát můžu vyhrát."
- 9. března 2016: „Dost mě to překvapilo, protože jsem nečekal, že prohraji."
- 10. března 2016: „Nemám slov. Jsem v šoku. Přiznám se, že ... třetí hra pro mě nebude snadná."
- 12. března 2016: „Cítil jsem se tak nějak bezmocně."



Obrázek 3.2: Umělá inteligence AlphaGo z dílny DeepMindu provedla na páté linii vysoce kreativní krok, který se vzpírá tisíciletím lidské moudrosti. Právě ten se přibližně po padesáti tazích ukázal klíčovým pro porážku legendy go Leeho Sedola.

Během roku od partie proti Leemu Sedolovi sehrála vylepšená verze AlphaGo zápasy se všemi dvaceti nejlepšími hráči světa a neprohrála jedinou partii.

Proč to pro mě tolik znamenalo? No, už jsem přiznal, že za centrální lidské vlastnosti pokládám intuici a kreativitu a podle mne AlphaGo projevila obojí, jak si nyní vysvětlíme.

Hráči go střídavě pokládají černé a bílé kameny na desku, která má 19 krát 19 průsečíků (obrázek 3.2). V go existuje nesrovnatelně více pozic, než kolik atomů se nachází v našem vesmíru. Z toho plyne, že pokusy analyzovat všechny zajímavé sekvence dalších tahů se rychle stávají beznadějnými. Hráči se proto musejí spoléhat na podvědomou intuici, kterou doplňuje jejich racionální uvažování, znalci mají vyvinutý takřka nadpřirozený cit pro to, které pozice jsou dobré a které nikoli. Jak jsme viděli v předchozí kapitole, výsledky hlubokého učení někdy intuici připomínají: hluboká neuronová síť může rozhodnout, že obrázek znázorňuje auto, aniž by to byla schopná zdůvodnit. Tým DeepMindu tedy vsadil na myšlenku, že hluboké učení dokáže rozpoznat nejen kočky, ale i výhodné pozice ve hře go. AlphaGo postavili na spojení intuitivní síly hlubokého učení se silou logiky GOFAI – což je zkratka z „Good Old-Fashioned AI" čili „stará dobrá AI" z dob před revolucí hlubokého učení. Použili rozsáhlou databázi pozic go z lidských her i z her, kde AlphaGo stál proti svému klonu, a vytrénovali hlubokou neuronovou síť, aby pro každou pozici vyjádřila pravděpodobnost, že bílý

zvítězí. Další oddělenou síť mezitím naučili předvídat pravděpodobné další kroky a poté tyto sítě zkombinovali GOFAI metodou, která chytře prohledávala zredukovaný seznam pravděpodobných sekvencí budoucích tahů, aby zjistila, jaký krok povede k nejvýhodnější pozici.

Z tohoto spojení intuice a logiky vzešly tahy, které byly nejenom silné, ale v některých případech i velmi kreativní. Tisíce let moudrosti v go například velí, že na začátku hry se nejvíc vyplatí hrát na třetí nebo čtvrté linii od kraje. Je to něco za něco: třetí linie pomáhá získat krátkodobou převahu na této straně hrací desky, zatímco čtvrtá je lepší pro dlouhodobý vliv u středu plochy.

V třicátém sedmém tahu druhé hry šokovala AlphaGo celý svět go tím, že se vzepřela prastaré moudrosti a táhla na páté linii (obrázek 3.2), jako by v její dlouhodobou výhodnost věřila více než člověk, a proto se rozhodla pro strategickou výhodu, a nikoli krátkodobý zisk. Komentátory to zarazilo a Lee Sedol dokonce vstal a na chvíli opustil místnost.[4] A opravdu, asi o padesát tahů později se spojil postup z levého dolního rohu hrací plochy s oním černým kamenem z tahu číslo třicet sedm! Právě to nakonec hru rozhodlo a upevnilo pověst tahu AlphaGo na páté linii jako jednoho z nejkreativnějších v dějinách go.

Kvůli své intuitivní a kreativní stránce je go považováno spíše za druh umění než za hru. Ve staré Číně bylo společně s malířstvím, kaligrafií a hudbou *qin* řazeno mezi čtyři „základní umění" a v Asii se stále těší značné popularitě; první partii mezi AlphaGo a Leem Sedolem sledovalo téměř 300 milionů diváků. Svět go byl proto tímto výsledkem docela otřesen a viděl ve vítězství AlphaGo zásadní milník lidstva. Ke Jie, toho času nejlepší hráč go na světě, se vyjádřil takto:[5] „Lidstvo hraje go tisíce let, a přesto jsme se, jak nám ukázala umělá inteligence, jen sotva dotkli povrchu. Spojení lidských a počítačových hráčů dá vzniknout nové éře. ... Společně mohou lidé a umělá inteligence nalézt pravdu go." Taková plodná spolupráce lidí a strojů skutečně vypadá slibně v mnoha odvětvích včetně vědy, kde nám lidem AI snad pomůže prohloubit naše poznání a uvědomit si náš skutečný potenciál.

Podle mě nám AlphaGo dala ještě další důležitou lekci pro blízkou budoucnost. Kombinováním intuice deep learningu s GOFAI logikou může vzniknout *strategie*, jíž se nic nevyrovná. Jelikož je go jedním z vrcholů strategických her, je nyní AI připravená dostudovat a vyzvat nejlepší lidské stratégy i mimo hrací plány (nebo jim pomoci) – například s investičními, politickými a vojenskými strategiemi. Takové strategické problémy skutečného světa obvykle ztěžuje lidská mentalita, neúplné informace a faktory, které je potřeba modelovat jako náhodné. Nicméně systémy AI hrající poker už ukázaly, že žádná z těchto výzev není nepřekonatelná.

PŘIROZENÝ JAZYK

Další doménou, kde mě pokrok AI nedávno zaskočil, je jazyk. Už v mládí jsem se zamiloval do cestování a zvědavost, jaké jsou jiné kultury a jazyky, zformovala významnou část mé identity. Vychovali mě tak, abych mluvil švédsky a anglicky, ve škole mě učili němčinu a španělštinu a dvě manželství mě naučila portugalsky

a rumunsky, pro zábavu jsem se pak naučil i základy ruštiny, francouzštiny a mandarínské čínštiny.

Ale umělá inteligence dosahuje dál a po důležitém objevu v roce 2016 neexistují skoro žádné líné jazyky, mezi nimiž umím překládat lépe než systém AI vyvinutý vybavením mozku Googlu.

Vyjádřil jsem se dost jasně? Vlastně jsem chtěl říct toto:

Ale umělá inteligence mě dohání a po významném průlomu v roce 2016 nezbývají skoro žádné jazyky, mezi nimiž umím překládat lépe než AI systém vyvinutý týmem Google Brain.

Nejdříve jsem to ovšem přeložil do španělštiny a pak zpátky pomocí programu, který jsem si na svůj notebook nainstaloval před pár lety. V roce 2016 aktualizoval tým Google Brain svou bezplatnou službu Google Translate, aby využívala zpětnovazební hluboké neuronové sítě, a zlepšení oproti starším GOFAI systémům bylo výrazné:[6]

Ale umělá inteligence mě dohání a po průlomu v roce 2016 nezbývají skoro žádné jazyky, mezi nimiž umějí překládat lépe než AI systém vyvinutý týmem Google Brain.

Jak můžete vidět, zájmeno „já“ se při španělské odbočce vytratilo, což bohužel změnilo význam celé věty. Skoro správně, ale přesto vedle! Na obranu umělé inteligence od Googlu ale musím dodat, že mi často vytýkají, že píšu zbytečně dlouhé věty, které se obtížně rozdělují. A pro tento příklad jsem vybral jednu z komplikovanějších a nejvíc matoucích. Typičtější věty už jejich AI často překládá bezchybně. Když se objevila, vyvolala značné pozdvižení a je natolik užitečná, že ji denně používají stovky milionů lidí. A co víc, díky nedávnému pokroku hlubokého učení při převádění mluvené řeči na text a obráceně mohou tito uživatelé mluvit do svého chytrého telefonu jedním jazykem a poslechnout si výsledný překlad v druhém.

Zpracování přirozeného jazyka je jedním z nejrychleji postupujících odvětví umělé inteligence. Domnívám se, že další úspěchy budou mít obrovský dopad vzhledem k tomu, jak klíčovou roli jazyk hraje v lidské společnosti. Čím více se bude AI zdokonalovat v lingvistických predikcích, tím lépe bude umět sepisovat rozumné e-mailové odpovědi nebo vést mluvenou konverzaci. To by se – přinejmenším pozorovateli zvnějšku – mohlo jevit jako projev lidského myšlení. Deep learning tak dělá první krůčky směrem ke slavnému Turingovu testu. V něm musí stroj konverzovat v písemné formě dostatečně dobře, aby člověka obelstil a ten si myslel, že komunikuje s lidskou bytostí.

Umělá inteligence pro zpracování jazyka má ovšem před sebou stále ještě dlouhou cestu. Musím sice přiznat, že mi poněkud snižuje sebevědomí, že AI překládá lépe než já, ale když si uvědomím, že zatím *nechápe*, co to vlastně říká, hned se cítím lépe. Umělá inteligence se cvičí na obsáhlých souborech dat, kde objevuje vzorce a vztahy, aniž by si slova spojovala s čímkoli ve skutečném světě. Například si každé slovo může označit seznamem tisíce čísel, která specifikují, jak moc se podobá určitým dalším slovům. Z toho pak může vyvodit, že rozdíl mezi „králem“ a „královnou“ se podobá odlišnosti mezi „manželem“ a „manželkou“ – ale pořád netuší,

co znamená být muž nebo žena, dokonce ani že někde venku existuje reálný svět s prostorem, časem a hmotou.

Jelikož se Turingův test v zásadě zakládá na schopnosti klamat, bývá kritizován za to, že více zkouší lidskou důvěřivost než vlastní umělou inteligenci. Naopak konkurenční test nazývaný *Winograd Schema Challenge* jde přímo k věci. Zaměřuje se na porozumění podle zdravého rozumu, které současné systémy hlubokého učení obvykle postrádají. My lidé při rozboru věty rutinně používáme znalosti ze skutečného světa, abychom zjistili, k čemu se jaké zájmeno vztahuje. Typická otázka Winogradova schématu se například ptá, k čemu se tu vztahuje (nevyjádřené) zájmeno „oni“:

1. Městští radní neudělili demonstrantům povolení, protože se (oni) báli násilí.
2. Městští radní neudělili demonstrantům povolení, protože se (oni) zastávali násilí.

Každoročně se koná soutěž, v níž má AI na takové otázky odpovídat, a stále si v ní vede mizerně.[7] Pochopit, co k čemu odkazuje, se nepodařilo ani Google Translate, když jsem ve svém výše popsaném případu nahradil španělštinu čínštinou:

Ale umělá inteligence mě dohnala, po významném zlomu v roce 2016, téměř s žádným jazykem, jsem mohl přeložit systém umělé inteligence než vyvinutý týmem Google Brain.

Vyzkoušejte to prosím sami, v době, kdy budete tuto knihu číst, na https://translate.google.com a přesvědčte se, jestli se AI Googlu zlepšila! Je docela dobře možné, že ano, protože už byly představeny slibné nápady, jak skloubit zpětnovazební hluboké neuronové sítě s GOFAI a vytvořit umělou inteligenci pracující s jazykem, která zahrnuje i model světa.

PŘÍLEŽITOSTI A VÝZVY

Tyto tři příklady jsou samozřejmě pouhou ukázkou, jelikož AI rychle postupuje na mnoha významných frontách. A přestože jsme zde zmínili jen dvě společnosti, konkurenční výzkumné skupiny na univerzitách a v jiných firmách nezůstaly příliš pozadu. Po celém světě se z různých oddělení informatiky ozývá hlasité srkání – to jak firmy jako Apple, Baidu, DeepMind, Facebook, Google, Microsoft a další lákavými nabídkami vysávají z univerzit studenty, postdoktorandy i učitele.

Nenechte se zmást uvedenými příklady. Dějiny umělé inteligence se neskládají z dlouhých období stagnace přerušených občasným průlomem. Už dlouho lze sledovat setrvalý pokrok, média o něm ale referují jen občas (a to jako o jednorázovém průlomu), většinou tehdy, když vznikne nějaká pozoruhodná aplikace nebo nečekaně užitečný produkt. Lze proto předpokládat, že strmý vývoj AI bude pokračovat ještě mnoho let. A jak jsme viděli v minulé kapitole, neexistuje žádný zásadní důvod, proč by tento pokrok neměl pokračovat, dokud AI nedožene ve většině oblastí lidské schopnosti.

Což nás přivádí k otázce: „Jak nás to ovlivní?“ Jak změní vývoj AI v dohledné době obsah pojmu „být člověkem“? Viděli jsme, že je stále obtížnější doložit tvrzení, že

AI zcela postrádá cíle, hloubku, intuici, kreativitu nebo jazyk – vlastnosti, které nemálo lidí považuje za základy lidského bytí. Z toho plyne, že i v krátkodobém horizontu, dlouho před tím, než se nám jakákoli AGI vyrovná ve všech schopnostech, může mít AI dramatický dopad na to, jak chápeme sami sebe, co dovedeme, když nás AI vylepší, a na čem v konkurenci s umělou inteligencí můžeme vydělávat peníze. Budou tyto změny k lepšímu, nebo k horšímu? Jaké výzvy a příležitosti to přinese v krátkodobém horizontu?

Vše, co na civilizaci milujeme, je produktem lidské inteligence. Když se nám ji tedy podaří zvýšit pomocí inteligence umělé, budeme samozřejmě schopni zlepšit svůj život ještě více. Dokonce i skromný pokrok AI může vést k rozsáhlému pokroku ve vědě a technologiích a k odpovídajícímu snížení množství nehod, nemocí, bezpráví, válek, bídy a dřiny. Ale abychom sklizením těchto plodů AI zároveň nevytvořili problémy nové, je třeba zodpovědět řadu důležitých otázek. Například:

1. Jak vytvořit robustnější systémy umělé inteligence, které by dělaly právě to, co chceme, nehavarovaly, nedělaly ještě něco jiného a byly odolné proti hackerským útokům?
2. Jak modernizovat náš právní systém, aby byl spravedlivější, efektivnější a držel krok s překotně se proměňující digitální krajinou?
3. Jak udělat zbraně chytřejší a snížit pravděpodobnost, že zabijí nevinné civilisty, aniž spustíme nekontrolované závody ve zbrojení se smrtícími autonomními zbraněmi?
4. Jak automatizací zvýšit blahobyt a nepřipravit při tom lidi o příjmy nebo smysl jejich existence?

Věnujme zbytek kapitoly prozkoumání uvedených problémů. Tyto čtyři otázky pro nejbližší budoucnost se obracejí především na informatiky, právníky, armádní stratégy a ekonomy, ovšem abychom potřebné odpovědi znali, až nadejde čas, musí se do této diskuse zapojit všichni. Jak totiž uvidíme, tyto výzvy překračují tradiční hranice mezi specializacemi i mezi národy.

CHYBY VS. ROBUSTNÍ AI

Informační technologie už vyvolala hluboké pozitivní změny prakticky ve všech sférách lidské činnosti, od vědy po finance, průmysl, dopravu, zdravotnictví, energetiku a komunikaci, tyto dopady však blednou ve srovnání s pokrokem, který může přinést umělá inteligence. Čím více se však budeme spoléhat na technologii, tím důležitější bude její robustnost a míra jistoty, že udělá právě to, co od ní budeme chtít.

Po celou historii lidstva jsme se při snažení, aby naše technologie zůstaly prospěšné, spoléhali na tentýž prověřený postup: poučení z předchozích chyb. Vynalezli jsme oheň, opakovaně vyhořeli, a nakonec přišli s hasicím přístrojem, únikovým východem, požárním alarmem a hasičským sborem. Vynalezli jsme automobil,

opakovaně bourali, a nakonec přišly bezpečnostní pásy, airbagy a samořídící auta. Až dosud naše technologie zpravidla způsobovaly nehod poměrně málo a škody byly menší než přínosy. S vývojem stále mocnějších technologií však nepochybně dospějeme do bodu, kdy by i jediná nehoda mohla být tak devastující, že škoda převáží veškerý prospěch. Někdo říká, že by to mohla být omylem spuštěná globální jaderná válka, jiní tvrdí, že kritérium by splňovala i pandemie zaviněná genovým inženýrstvím, a v následující kapitole se podíváme na debatu, zda by vyhynutí lidstva nemohla zapříčinit umělá inteligence. Není třeba zvažovat až tak extrémní příklady, abychom došli k zásadnímu závěru: se zvyšující se mocí technologie bychom se měli při hledání bezpečnostních mechanismů méně spoléhat na přístup pokus-omyl. Jinými slovy: *Měli bychom být proaktivnější, nikoli jen reagovat*. Investovat do výzkumu bezpečnosti, aby k nehodám nedocházelo, aby se nestaly ani jednou. Právě proto společnost dává víc peněz na výzkum bezpečnosti jaderných reaktorů než do bezpečnosti pastiček na myši.

I kvůli tomu byl, jak jsme viděli v první kapitole, na portorické konferenci tak intenzivní zájem o výzkum bezpečnosti AI. Počítače a systémy AI havarovaly vždycky, ale tentokrát je to jiné: AI pomalu vstupuje do skutečného světa. A pokud způsobí kolaps rozvodné sítě, burzy nebo systému jaderných zbraní, není to jen pouhá nepříjemnost. Ve zbytku tohoto oddílu si představíme čtyři hlavní oblasti bezpečnosti AI, které dnes vévodí diskusím a jimiž se zabývají lidé na celém světě: jsou to *verifikace*, *validace*, *bezpečnost* a *kontrola*.* Aby výklad nebyl příliš suchý a pro běžného čtenáře nesrozumitelný, projdeme si dosavadní úspěchy i neúspěchy informačních technologií v různých odvětvích a cenná poučení, která si z nich můžeme odnést. Popíšeme i výzvy, které přinášejí pro výzkum.

Ačkoli je většina těchto příběhů staršího data a týká se primitivních počítačových systémů, které by skoro nikdo za umělou inteligenci neoznačil, a ačkoli nezpůsobily žádné (nebo skoro žádné) oběti na životech, uvidíme, že nás mají co naučit. Mohou pomoci při návrhu bezpečných a účinných systémů AI, jejichž chyby by jinak mohly mít opravdu katastrofické důsledky.

AI PRO ZKOUMÁNÍ VESMÍRU

Začněme u něčeho, co je mi blízké: u zkoumání vesmíru. Počítačová technologie nám umožnila dostat lidi na Měsíc a vyslat nepilotovaná plavidla ke všem planetám Sluneční soustavy, dokonce přistála na Saturnovu měsíci Titanu a na kometě. V kapitole 6 se podíváme na to, jak by nám AI v budoucnu mohla pomáhat v průzkumu jiných planetárních systémů, dokonce i galaxií – nebude-li mít chyby. Dne 4. června 1996 vědci, kteří chtěli zkoumat magnetosféru Země, nadšeně jásali při startu rakety Ariane 5 Evropské kosmické agentury ESA, která nesla na palubě jejich přístroje. O třicet sedm sekund později jejich úsměvy ztuhly – to když raketa

* Pokud chcete podrobnější mapu krajiny výzkumu bezpečnosti AI, existuje jedna interaktivní. Vznikla spojeným úsilím v čele s Richardem Mallahem z FLI: https://futureoflife.org/landscape.

explodovala v ohňostroji za stovky milionů dolarů. Jak se zjistilo, příčinou byl vadný software, který operoval s čísly příliš velkými na oněch 16 bitů, které pro ně měl alokované. Dva roky nato sestoupil omylem Mars Climate Orbiter od NASA do marsovské atmosféry a rozpadl se, protože dvě různé části softwaru operovaly každá s jinou jednotkou síly, což způsobilo 445% chybu v kontrole tahu motoru.[10] Jednalo se o druhou extrémně drahou softwarovou chybu v dějinách NASA. Předtím vybuchla mise Mariner 1 k Venuši po startu z Cape Canaveral 22. července 1962 poté, co software letové kontroly selhal kvůli nesprávnému interpunkčnímu znaménku.[11] Skoro jako by Sověti chtěli dokázat, že Západ nemá monopol na programové chyby, havarovala 2. září 1988 sovětská mise Fobos 1. Bylo to v historii nejtěžší vypuštěné meziplanetární kosmické plavidlo s ambiciózním cílem dostat přistávací modul na Phobos, měsíc planety Mars. To všechno přišlo vniveč, když chybějící spojovník způsobil, že plavidlu byl vyslán příkaz „konec-mise", čímž vypnul všechny jeho systémy.[12]

Tyto příklady nám ukazují důležitost něčeho, čemu informatici říkají *verifikace*: ta zajišťuje, aby software přesně splňoval všechny požadavky. Čím více životů a zdrojů je v sázce, tím více si chceme být jisti, že software bude fungovat tak, jak má. Naštěstí nám může s automatizací a vylepšováním verifikačního procesu pomoci umělá inteligence. Například jádro univerzálního operačního systému jménem *seL4* bylo nedávno v úplnosti matematicky zkontrolováno podle formálních specifikací, aby poskytovalo silnou záruku, že systém nezhavaruje a že bude dělat jen to, co má. I když Microsoft Windows ani MacOS pro své rozbujelé vlastnosti takovým certifikátem ještě vyšperkovány nejsou, můžete se už víceméně spolehnout, že vám neukážou dříve dobře známý obrázek, expresivně nazývaný „modrá obrazovka smrti" nebo „točící se kolečko smrti". Americká Agentura ministerstva obrany pro pokročilé výzkumné projekty (DARPA) poskytla finance na vývoj sady open-sourceových vysoce bezpečných nástrojů, jejichž bezpečnost lze prokázat: jmenuje se HACMS (High-Assurance Cyber Military Systems). Zásadní výzva zní, jak učinit podobné nástroje dostatečně silnými a snadno použitelnými, aby byly nasazovány ve velkém. Další problém pak spočívá v tom, že samotná verifikace se bude tím více komplikovat, čím více se software bude přesouvat do robotů a do nových prostředí. A také tím, čím více bude předem naprogramovaný software nahrazován systémy AI, které se neustále učí, a tím pádem mění své chování (jak popisujeme v kapitole 2).

AI PRO FINANČNÍ SEKTOR

Finančnictví je dalším odvětvím, které informační technologie zcela proměnily. Umožňují zde efektivní realokaci zdrojů přes celou zeměkouli rychlostí světla a jen díky nim lze dosáhnout na financování čehokoli – od hypotéky na garáž po technologické startupy. Pokrok AI pravděpodobně přinese velké příležitosti zbohatnout na finančních trzích. Většina rozhodnutí k nákupu či prodeji dnes probíhá automatizovaně, prostřednictvím počítače. Moji studenti z MIT jsou často zlákáni astronomickými nástupními platy a pouštějí se do vylepšování obchodovacích algoritmů.

Ani software na finančních trzích se však bez verifikace neobejde. Americká firma Knight Capital se to naučila po zlém: 1. srpna 2012 ztratila 440 milionů dolarů během 45 minut od spuštění neverifikovaného obchodovacího softwaru.[13] Slavný „Flash Crash" za bilion dolarů z 6. května 2010 však vešel do dějin z jiného důvodu. Ačkoli asi půl hodiny (dokud se trhy nestabilizovaly) způsoboval masivní výkyvy (cena akcií společností jako Procter & Gamble fluktuovala mezi jedním centem a 100 000 dolarů[14]), nebyly příčinou problému chyby programu, jimž mohla zabránit verifikace. Mohla za to chyba v očekávání: programy pro automatické obchodování mnoha společností se ocitly v nepředvídané situaci, kdy přestaly platit jejich předpoklady. Mimo jiné to byl předpoklad, že pokud počítač na burze udává cenu akcie jeden cent, má skutečně hodnotu jednoho centu.

Flash Crash skvěle ilustruje důležitost čehosi, co informatici označují jako *validace*. Zatímco verifikace se ptá: „Postavil jsem ten systém správně?", otázka při validaci zní: „Postavil jsem ten správný systém?"* Nespoléhá se třeba systém na předpoklady, které nemusejí být vždy platné? A pokud ano, jak ho vylepšit, aby lépe zvládal neurčitost?

AI PRO PRŮMYSLOVOU VÝROBU

Není třeba zdůrazňovat, že AI má nesmírný potenciál pro vylepšení výroby, jelikož řídí roboty, kteří zvyšují efektivitu a přesnost produkce. Neustále se zdokonalující 3D tiskárny dnes dovedou vyrobit prototypy čehokoli od kancelářských budov po mikromechanické objekty menší než zrnko soli.[15] Obrovští průmysloví roboti montují auta a letadla, cenově dostupné, počítačem kontrolované frézy, soustruhy, obráběcí nože a podobně jsou nejen páteří továren, ale staví na nich také hnutí Maker Movement, v jehož rámci nadšenci zhmotňují své nápady ve více než tisícovce komunitních „nadšeneckých laboratoří" (fan labs) po celém světě.[16] Čím víc robotů nás však obklopuje, tím zásadnější výzvou se stává verifikace a validace jejich softwaru. Prvním člověkem, o němž víme, že ho zabil robot, byl Robert Williams, dělník ve Fordově továrně ve Flat Rocku v Michiganu. V roce 1979 se porouchal robot, který měl přivézt součástky z prostoru, kde byly uloženy, a tak tam pro ně Williams vylezl sám. Robot neslyšně začal pracovat a rozbil mu hlavu. V činnosti pokračoval ještě půl hodiny, dokud si ostatní dělníci nevšimli, co se stalo.[17] Další obětí robota se stal Kenji Urada, inženýr údržby v továrně Kawasaki v Akaši v Japonsku. Při práci na rozbitém robotovi zavadil v roce 1981 omylem o jeho vypínač a hydraulická paže ho umačkala k smrti.[18] V roce 2015 nastavoval dvaadvacetiletý mechanik v jedné z výrobních hal Volkswagenu v německém Baunatalu robota, který bral autodíly a manipuloval jimi. Něco se ale porouchalo, robot ho zachytil a rozmáčkl o kovový plát.[19]

Jakkoli jsou tyto nehody tragické, měli bychom upozornit na fakt, že představují jen nepatrný zlomek všech nehod, k nimž v továrnách dochází. Navíc počet

* Přesněji řečeno se verifikace ptá, zda systém splňuje své specifikace, zatímco validace se ptá, zda byly zvoleny ty správné specifikace.

Obrázek 3.3: Zatímco tradiční průmysloví roboti jsou drazí a jejich programování je těžké, existuje trend směřující k levnějším robotům řízeným umělou inteligencí, kteří se učí pracovat od dělníků bez zkušeností s programováním.

nehod v průmyslu se s technologickým pokrokem *snížil*, nikoli zvýšil. Ve Spojených státech poklesl ze 14 000 úmrtí v roce 1970 na 4 821 v roce 2014.[20] Ony tři výše zmíněné nehody ukazují, že přidání inteligence hloupým strojům by mohlo zvýšit bezpečnost v průmyslu, pokud by se roboti naučili větší opatrnosti v přítomnosti lidí. Všem třem neštěstím se dalo předejít lepší validací – roboti způsobili neštěstí nikoli kvůli chybám v programu (natož pak ze zlého úmyslu), ale protože chybné byly jejich předpoklady. Domnívali se, že před nimi není člověk, ale autodíl.

AI PRO DOPRAVU

Ačkoli AI může zachránit mnoho životů ve výrobě, ještě více jich může zachránit v dopravě. Jen v roce 2015 si na světě autonehody vyžádaly přes 1,2 milionu životů a při nehodách leteckých, železničních a lodních zahynulo mnoho dalších lidí. Ve Spojených státech zemřelo navzdory tamním vysokým bezpečnostním standardům při nehodách motorových vozidel v roce 2016 asi 35 000 lidí. To je sedmkrát více než při všech nehodách v průmyslu dohromady.[21] Při panelové diskusi, kterou jsme na toto téma měli v texaském Austinu na každoročním sjezdu Asociace pro pokrok umělé inteligence v roce 2016, se izraelský informatik Moshe Vardi na toto téma rozohnil a prosazoval názor, že pokud AI dokáže snížit počet obětí dopravních nehod, tak k jejímu nasazení prostě musíme přistoupit: „Je to morální imperativ!“ vykřikl. Jelikož jsou téměř všechny autonehody zapříčiněny chybou lidského faktoru, všeobecně se předpokládá, že samořídící auta ovládaná umělou inteligencí mohou eliminovat alespoň 90 % úmrtí na silnicích. Tento optimismus

je motorem velkého úsilí dostat samořídící auta skutečně na silnice. Podle Elona Muska budou samořídící automobily nejen bezpečnější, ale navíc mohou svým majitelům v době, kdy je nebudou potřebovat, také vydělávat peníze tím, že budou konkurovat třeba Uberu a Lyftu.

Dosud mají samořídící auta bezpečnostní statistiky skutečně lepší než lidští řidiči. Nehody, k nimž dosud došlo, jen podtrhují důležitost a obtížnost validace. První menší srážka zapříčiněná samořídícím autem Googlu se odehrála 14. února 2016. Vůz totiž nesprávně vyhodnotil chování autobusu: předpokládal, že mu jeho řidič uhne, když se před ním objeví. K první smrtelné nehodě způsobené samořídícím vozem Tesla došlo 7. května 2016, a to v důsledku dvou chybných předpokladů:[22] zaprvé že jasně bílá stěna návěsu tahače je jen částí jasného nebe a zadruhé že jeho řidič (který se údajně díval na film s Harrym Potterem) dává pozor, a pokud by se něco pokazilo, zasáhne.*

Někdy ovšem dobrá verifikace a validace k zabránění nehody nestačí, protože potřebujeme také dobrou *kontrolu*: schopnost lidského operátora systém monitorovat a v případě nutnosti změnit jeho chování. Aby takové systémy, jejichž součástí je člověk, fungovaly správně, je zásadní, aby komunikace mezi člověkem a strojem byla efektivní. V tomto duchu vás červená kontrolka na přístrojové desce auta upozorní, pokud omylem nezavřete kufr. Naproti tomu když britský trajekt *Herald of Free Enterprise* vyplul 6. března 1987 z přístavu Zeebrugge s otevřenými vjezdovými vraty na přídi, neměl kapitán žádnou kontrolku, která by ho varovala. Trajekt záhy po opuštění přístavu nabral vodu, převrátil se a zahynulo 193 lidí.[23]

Další tragické selhání kontroly, jemuž mohla zabránit lepší komunikace mezi člověkem a strojem, nastalo v noci na 1. června 2009, když se letadlo Air France při letu číslo 447 zřítilo do Atlantského oceánu a zahynulo všech 228 osob na palubě. Podle oficiální zprávy o nehodě „posádka nepochopila, že klesá vztlak pod křídlem, a proto nepřistoupila k patřičnému manévru" – stlačit příď letadla dolů – a pak už bylo příliš pozdě. Experti na bezpečnost dopravy se domnívají, že se havárii dalo předejít, pokud by v kokpitu měli kontrolku úhlu náběhu, díky níž by piloti viděli, že nos letadla míří příliš vzhůru.[24]

Když se 20. ledna 1992 ve Vogézách nedaleko francouzského Štrasburku zřítil let Air Intel 148 a zahynulo 87 lidí, nezpůsobil to nedostatek komunikace mezi člověkem a strojem, ale matoucí uživatelské rozhraní. Piloti zadali na klávesnici „33", protože chtěli sestupovat pod úhlem 3,3 stupně. Autopilot to ovšem vyhodnotil jako pokles o 3 300 stop za minutu, protože operoval v jiném módu. Displej byl příliš malý a mód neukazoval, a tak si piloti svou chybu ani nemohli uvědomit.

AI PRO ENERGETIKU

Informační technologie znamenaly velký pokrok i ve výrobě a distribuci energie.

* Navzdory této nehodě, která ve statistikách figuruje, se zjistilo, že zapnutý autopilot od Tesly snižuje nehodovost o 40 procent: http://tinyurl.com/teslasafety.

Sofistikované algoritmy vyrovnávají výrobu a odběr napříč celosvětovou elektrickou sítí, další komplikované kontrolní systémy zajišťují, aby elektrárny fungovaly bezpečně a efektivně. Budoucí pokrok AI pravděpodobně učiní inteligentní sítě ještě inteligentnějšími, aby se optimálně přizpůsobovaly změnám v dodávce a odběru energie i na úrovni jednotlivých solárních panelů na střechách a domácích akumulátorových systémů. Nicméně ve čtvrtek 14. srpna 2003 se bez elektřiny ocitlo 55 milionů lidí ve Spojených státech a Kanadě a pro mnoho z nich trval výpadek celé dny. I zde se zjistilo, že primárním důvodem byla chyba v komunikaci mezi strojem a člověkem: kvůli chybě softwaru neupozornil poplašný systém v kontrolní místnosti v Ohiu obsluhu na nutnost odklonit přenosovou cestu dříve, než se drobný problém (přetížené vedení se dotýkalo neprořezaných větví na stromech) vymkl kontrole.[25]

Částečné roztavení reaktoru na ostrově Three Mile Island v Pensylvánii 28. března 1979 si vyžádalo přibližně miliardu dolarů na sanaci a vyvolalo velkou vlnu odporu proti jaderné energii. Závěrečná zpráva o nehodě identifikovala mnoho faktorů, které k tomu vedly, mimo jiné to byl zmatek způsobený nekvalitním uživatelským rozhraním.[26] Konkrétně si personál myslel, že jistá kontrolka ukazuje, zda je mimořádně důležitý ventil otevřený nebo zavřený, a ona přitom jen indikovala, že byl vyslán signál k jeho uzavření. Obsluha si proto neuvědomila, že se ventil zasekl a zůstal otevřený.

Tyto nehody v energetice a dopravě nás učí, že když budeme pověřovat umělou inteligenci dohledem nad stále větším počtem fyzických systémů, budeme se muset intenzivně věnovat otázce, jak zajistit, aby stroje fungovaly správně samy, a nadto aby efektivně spolupracovaly s lidmi, kteří je obsluhují. Při zvyšující se inteligenci AI to nebude znamenat jen tvorbu dobrých uživatelských rozhraní pro sdílení informací, ale také budeme muset vyzkoumat, jak optimálně rozdělovat úkoly v týmech složených z lidí a strojů. Například pečlivě definovat situace, kdy se stroj má vzdát řízení, ale na druhé straně využívat lidskou obsluhu jen při rozhodování na nejvyšší úrovni a nezatěžovat ji záplavou bezvýznamných informací při běžném provozu.

AI PRO ZDRAVOTNICTVÍ

Umělá inteligence má obrovský potenciál zlepšit zdravotní péči. Digitalizace zdravotnické dokumentace už umožnila lékařům i pacientům rozhodovat lépe a rychleji. Lékaři mohou okamžitě získat pomoc odborníků z celého světa, kteří pomohou stanovit diagnózu podle digitálních snímků. Vzhledem k překotnému pokroku v počítačovém rozpoznávání obrazů a v hlubokém učení se předními experty na provádění takovýchto diagnóz mohou zanedlouho stát systémy AI. V roce 2015 například jedna nizozemská studie ukázala, že počítačová diagnóza rakoviny prostaty na základě snímků z magnetické rezonance je právě tak dobrá jako diagnóza od lidského radiologa.[27] A podle stanfordské studie z roku 2016 dokáže umělá inteligence odhalit rakovinu plic na základě snímků z mikroskopu dokonce lépe než člověk.[28] Když umí strojové učení odhalovat vztahy mezi geny, nemocemi a reakcí

na léčbu, mohlo by přinést revoluci do personalizované léčby (kromě toho, že bude vyvíjet odolnější plodiny a zlepšovat zdravotní stav hospodářských zvířat). Navíc mají roboti potenciál stát se přesnějšími a spolehlivějšími chirurgy než lidé, a to i bez použití pokročilé AI. V posledních letech úspěšně proběhlo mnoho robotických operací, kdy přesnost provedení a minimalizace operační rány vedly ke snížení bolestivosti, menším ztrátám krve a rychlejšímu hojení.

Ani zdravotnictví bohužel nezůstalo ušetřeno bolestivých lekcí o důležitosti robustního softwaru. Například kanadský přístroj pro radioterapii Therac-25 byl zkonstruován, aby léčil onkologické pacienty ve dvou odlišných módech: buď nízkoenergetickým proudem elektronů, nebo vysokoenergetickým paprskem rentgenového záření v řádu megavoltů zamířeným na dané místo za pomoci speciálního štítu. Neodladěný neverifikovaný software naneštěstí čas od času způsobil, že technici spustili rentgenový paprsek, zatímco se domnívali, že zapnuli proud nízkoenergetický – a nepoužili štít. Několik pacientů to stálo život.[29] Ještě mnohem více pacientů zemřelo na ozáření v Národním onkologickém institutu v Panamě, kde bylo v roce 2000 a 2001 vybavení pro radioterapii používající radioaktivní kobalt 60 naprogramováno k nadměrné expozici. Na vině bylo matoucí uživatelské rozhraní, které neprošlo řádnou validací.[30] Podle nedávné zprávy[31] bylo v USA mezi lety 2000 a 2013 spojeno s nehodami při robotických operacích 144 úmrtí a 1 391 zranění. Časté problémy se týkaly nejen hardwaru (zkraty, spálené či odlomené části přístrojů, které zůstanou v těle pacienta), ale svou roli sehrály i problémy softwaru, které způsobovaly polohovací chyby či samovolné vypínání.

Dobrou zprávou je, že zbytek z téměř dvou milionů robotických operací, které zpráva zahrnovala, proběhl hladce, a zdá se, že roboti operacím na bezpečnosti přidávají a jejich rizikovost snižují. Podle analýzy pro americkou vládu přispívá špatná nemocniční péče jen ve Spojených státech ročně k více než 100 000 úmrtí.[32] Vyvinout lepší umělou inteligenci pro medicínu je patrně ještě silnější morální imperativ, než jak je tomu v případě samořídících aut.

AI PRO KOMUNIKACI

Odvětvím, které počítače zatím ovlivnily nejvíc, je pravděpodobně obor komunikací. Po zavedení počítačů do telefonních ústředen v padesátých letech, internetu koncem let šedesátých a celosvětové sítě World Wide Web v roce 1989 se dnes miliardy lidí připojují online, aby komunikovali, nakupovali, četli zprávy, sledovali filmy nebo hráli hry. Zvykli si, že informace celého světa mají na dosah a dělí je od nich jediné kliknutí – a že bývají zdarma. Formující se *internet věcí* slibuje zvýšení účinnosti, přesnosti a pohodlí, a mimoto i ekonomický přínos poté, co online spojí vše od lamp, termostatů a mrazáků po transpondéry biočipů hospodářských zvířat.

Tyto působivé úspěchy při propojování světa postavily informatiky před čtvrtou výzvu: kromě verifikace, validace a kontroly je třeba zlepšit také *bezpečnost* – ochranu před nebezpečným softwarem („malwarem“) a hackerskými útoky. Zatímco dříve jmenované okruhy řeší neúmyslné chyby, bezpečnost se zaměřuje na

úmyslně škodlivé jednání. Prvním malwarem, jenž si získal značnou pozornost médií, byl takzvaný Morris Worm, vypuštěný 2. listopadu 1988, který zneužíval chyby v operačním systému UNIX. Údajně se jednalo o pomýlený pokus spočítat, kolik počítačů je online. Přestože nakazil a přivedl k pádu asi 10 % z oněch 60 000 počítačů, které tehdy tvořily internet, nezabránilo to jeho tvůrci, Robertu Morrisovi, aby nakonec získal místo řádného profesora informatiky na MIT.

Jiný známý malware zneužíval zranitelnost nikoli softwaru, ale lidí. Dne 5. května 2000, jakoby na počest mých narozenin, obdrželi lidé od svých známých a kolegů e-mail, jehož předmět zněl: „ILOVEYOU". Ti uživatelé Microsoft Windows, kteří klikli na přílohu „LOVE-LETTER-FOR-YOU.txt.vbs", nevědomky spustili skript, který poškodil jejich počítač a přeposlal onen e-mail všem lidem z jejich seznamu kontaktů. Podobný červ, tentokrát vytvořený dvěma mladými programátory z Filipín, infikoval přibližně 10 % internetu (právě tak jako svého času Morrisův červ), ale protože tou dobou už byl internet mnohem větší, stal se jednou z největších infekcí vůbec. Devastoval 50 milionů počítačů a způsobil škody za více než 5 miliard dolarů. Měli byste vědět, že internet je stále zamořený bezpočtem druhů nakažlivého malwaru, který bezpečnostní experti dělí na červy, trojské koně, viry a další nebezpečně znějící druhy. Škody, které mohou napáchat, se značně liší – od zobrazení neškodných žertovných zpráv po smazání všech souborů či krádež osobních informací. Mohou vás dokonce špehovat nebo se zmocnit vašeho počítače, aby rozesílal spam.

Zatímco malware cílí na jakýkoli počítač, na který může, *hackeři* útočí na určité cíle, které je zajímají – nedávno se jejich terčem například staly společnosti Target, TJ Maxx, Sony Pictures, Ashley Madison, saúdská ropná společnost Aramco a americký Demokratický národní výbor. Navíc se zdá, že jdou po stále lákavější kořisti. V roce 2008 hackeři ukradli 130 milionů čísel kreditních karet a dalších informací k účtům od Heartland Payment Systems a v roce 2013[33] pronikli do více než miliardy(!) e-mailových účtů webové služby Yahoo!. O rok později se při útoku na americký vládní úřad Office of Personnel Management hackeři dostali k údajům o zaměstnancích a záznamům o pracovních pohovorech více než 21 milionů lidí. Údajně se týkaly i osob s nejvyššími stupni bezpečnostních prověrek, dokonce měli získat otisky prstů agentů operujících v utajení.

Proto kdykoli čtu, že je nějaký nový systém naprosto bezpečný a chráněný proti hackerským útokům, obracím oči v sloup. Nicméně právě „nehacknutelnou" AI v budoucnu nutně potřebujeme – musíme ji zabezpečit dříve, než ji nasadíme dejme tomu do klíčové infrastruktury nebo do zbraňových systémů. Rostoucí role AI ve společnosti neustále zvyšuje počítačová rizika. Některé hackerské útoky sice zneužívají lidskou důvěřivost nebo složitě využitelné mezery v novém softwaru, jiné ovšem dovolují neautorizované vzdálené přihlášení do počítače následkem jednoduchých chyb, jichž si ostudně dlouho nikdo nevšiml. Chyba „Heartbleed" vydržela mezi lety 2012–2014 v jedné z nejoblíbenějších softwarových knihoven pro bezpečnou komunikaci mezi počítači a chyba „Bashdoor" byl zabudována v samém

operačním systému unixových počítačů od roku 1989 až do 2014. Z toho plyne, že nástroje AI pro vylepšenou verifikaci a validaci povedou ke zvýšení bezpečnosti všech systémů.

Lepší systémy AI lze bohužel použít i k tomu, aby hledaly další zranitelná místa a prováděly sofistikovanější hackerské útoky. Představte si třeba, že jednoho dne dostanete neobvykle personalizovaný „phishingový" e-mail, který se vás bude snažit přesvědčit, abyste mu poskytli osobní údaje. Přišel z účtu vaší kamarádky, ale poslala ho umělá inteligence, která do něj pronikla a jen se za ni vydává. Napodobuje její styl (na základě analýzy autentických e-mailů) a zmiňuje řadu osobních informací, které pocházejí z jiných zdrojů. Naletěli byste? Co kdyby ten phishingový e-mail vypadal, že je od společnosti, od níž máte kreditní kartu, a následoval by po něm telefonát, kde byste u přátelsky znějícího hlasu nedokázali poznat, že ho vygenerovala AI? V probíhajícím boji mezi útokem a obranou v oblasti počítačové bezpečnosti dosud jen máloco naznačuje, že by obrana mohla vyhrávat.

ZÁKONY

My lidé jsme společenská stvoření, která si podrobila všechny ostatní druhy a dobyla Zemi díky schopnosti spolupracovat. Zákony jsme vytvořili, aby tuto spolupráci podněcovaly a usnadňovaly. Pokud tedy umělá inteligence vylepší náš právní systém a systém veřejné správy, může nám dopomoci spolupracovat úspěšněji než kdy dříve a projevit to nejlepší, co v sobě máme. V tom, jak jsou zákony formulovány a vykládány, je prostoru pro vylepšení více než dost. Podívejme se postupně na obojí.

Co vás napadne jako první, když přijde řeč na soudní systém ve vaší zemi? Pokud to jsou zdlouhavé průtahy, vysoké finanční náklady a občasná nespravedlnost, nejste v tom sami. Nebylo by báječné, kdyby místo toho vašimi prvními dojmy byly „efektivita" a „spravedlnost"? Jelikož lze soudní řízení abstraktně chápat jako výpočet, kde vstupem jsou zákony a informace o důkazních prostředcích a výstupem je rozsudek, někteří vědci sní o tom, že ho lze plně automatizovat pomocí *robosoudců*. Takové systémy AI by se každému případu věnovaly na nejvyšší úrovni právních standardů a nepodléhaly by lidským chybám – předpojatosti, únavě či nedostatku znalostí.

ROBOSOUDCI

Byron De La Beckwith jr. byl až v roce 1994 shledán vinným z vraždy vůdce černošského hnutí za lidská práva Medgara Everse, spáchané v roce 1963. Dvě různé poroty v Mississippi (tvořené výhradně bělošskými porotci) ho totiž rok po činu neodsoudily, ačkoli důkazy byly prakticky totožné.[34] Dějiny práva bohužel přímo oplývají rozsudky podjatými kvůli barvě pleti, pohlaví, sexuální orientaci, vyznání, národnosti a dalším faktorům. Robosoudci by v principu mohli zajistit, aby si poprvé v dějinách byli všichni před zákonem opravdu rovni. Mohli by být všichni

naprogramováni tak, aby se ničím nelišili a aby se všemi zacházeli stejně. Zákon by aplikovali transparentně a skutečně by neměli žádné předsudky.

Nasazení robosoudců by také mohlo eliminovat lidské předsudky, které nejsou záměrné a projevují se bezděčně. Kontroverzní studie izraelské justice z roku 2012 například tvrdí, že když má soudce hlad, vynáší mnohem přísnější rozsudky. Zatímco soudci hned po snídani zamítají přibližně 35 % podmíněných propuštění, těsně před obědem je to více než 85 %.[35] Další slabinou lidských soudců je, že nemají čas na prostudování všech detailů případu, zatímco robosoudce lze snadno zkopírovat, jelikož je to pouhý software. Díky tomu by všechny aktuální případy mohly být zpracovávány současně, nikoli jeden po druhém, až bude soudce volný. Každá kauza dostane svého robosoudce na tak dlouho, jak jen bude zapotřebí. A nakonec, zatímco lidští soudci nemohou ovládnout veškeré technické znalosti pro všechny druhy případů – od ožehavého patentového sporu po vyšetřování vražd nejnovějšími forenzními technikami – robosoudci budoucnosti budou mít k dispozici prakticky neomezenou informační kapacitu.

Jednoho dne budou možná takoví robosoudci efektivnější i spravedlivější díky své nezaujatosti, kompetentnosti a transparentnosti. Jejich efektivita je učiní ještě spravedlivějšími, neboť urychlí soudní procesy a vychytralým právníkům znemožní ovlivňovat závěry. A domáhání se spravedlnosti soudní cestou by se navíc značně zlevnilo, což by zvýšilo šance podvodně okradených jedinců nebo malých firem stojících proti nadnárodní korporaci, která disponuje armádou právníků.

Na druhou stranu – co kdyby měl robosoudce programovou chybu nebo do jeho systému pronikl nějaký hacker? Oba problémy se již dotkly strojů na sčítání volebních hlasů, a když půjde o roky ve vězení nebo miliony dolarů, kyberútoky to jen povzbudí. I kdyby byla AI natolik robustní, abychom se mohli spolehnout, že používá patřičný schválený algoritmus, budou všichni zúčastnění její argumentaci rozumět natolik, aby respektovali vynesený rozsudek? Tento problém ještě ztěžuje výkonnost neuronových sítí, které jsou v poslední době často úspěšnější než tradiční (a snadno pochopitelné) algoritmy umělé inteligence, ovšem za cenu neprůhlednosti. Když chtějí obžalovaní vědět, *proč* byli uznáni vinnými, neměli by mít nárok na lepší odpověď, než že „systém jsme natrénovali na obrovském množství dat, a toto je jeho rozhodnutí“? Nedávné výzkumy navíc ukázaly, že pokud natrénujete neuronovou síť hlubokého učení na velkém souboru dat o věznících, dokáže lépe než lidští soudci předpovědět, kdo se k protiprávní činnosti pravděpodobně vrátí (a proto by se jeho podmínečné propuštění mělo zamítnout). Ale co když systém dospěje k závěru, že riziko recidivy statisticky souvisí s pohlavím nebo rasovou příslušností vězně? Byl by takový robosoudce považován za sexistického a rasistického, a měl by tudíž být přeprogramován? V roce 2016 jeden výzkum skutečně tvrdil, že software na předpovídání recidivy (který se používá po celých Spojených státech) je předpojatý vůči Afroameričanům, a že tak přispívá k nespravedlivým verdiktům.[36] Tyto důležité otázky musíme všichni zvážit a prodiskutovat, protože jen tak zajistíme, aby AI mohla být prospěšná. Ohledně robosoudců nestojíme před

debatou „všechno, nebo nic“, ale spíše před rozhodnutím, v jakém rozsahu a jak rychle vpustíme AI do svého soudního systému. Chceme, aby lidští soudci měli k dispozici podpůrnou AI pro rozhodování, přesně jako zítřejší lékaři? Nebo půjdeme ještě dál a necháme rozhodovat robosoudce samostatně, přičemž se bude možné odvolat k soudcům lidským? Nebo to dovedeme až do konce a udělíme strojům úplnou pravomoc, včetně trestu smrti?

PRÁVNÍ KONTROVERZE

Dosud jsme hovořili jen o *aplikaci* práva, obraťme se nyní k jeho *obsahu*. Panuje široká shoda, že naše zákony se musí proměnit, jinak neudrží krok s technologií. Například ti dva programátoři, kteří vytvořili výše zmíněného červa ILOVEYOU a způsobili škodu v řádu miliard dolarů, byli zproštěni všech obvinění a odešli volní, protože v té době neexistovaly na Filipínách žádné zákony postihující tvorbu malwaru. Jelikož se zdá, že se tempo technologického pokroku zvyšuje, je třeba zákony aktualizovat stále častěji – mají totiž tendenci zůstávat pozadu. Od společnosti by proto bylo moudré, kdyby dostala na právnické fakulty a do vlády více lidí, kteří se vyznají v technologiích. Mělo by ale být povoleno ovlivňování („podpora“) voličů a zákonodárců umělou inteligencí? A potom snad přímo robotičtí zákonodárci?

Velmi kontroverzním tématem je otázka, jak změnit naše zákony, aby reflektovaly pokrok AI. Jedno z témat odráží napětí mezi soukromím a svobodou informací. Zastánci svobody informací tvrdí, že čím méně máme soukromí, tím více důkazů budou soudy mít a tím spravedlivější rozsudky budou vynášet. Kdyby se stát napojil do elektronických zařízení všech lidí a zaznamenával, kde jsou, co píšou, říkají a dělají a kam klikají, řada zločinů by se snadno vyřešila a mnoha dalším by se dalo zabránit. Zastánci soukromí oponují, že nechtějí orwellovský stát, který vše monitoruje, protože se z něj stane totalitní diktatura nevídaných rozměrů. Techniky strojového učení se navíc zdokonalily v analýze dat z funkční magnetické rezonance mozku a jsou schopny určit, na co dotyčný myslí (a mimo jiné i to, zda říká pravdu, nebo lže).[37] Kdyby se skenery mozkové aktivity s podporou AI staly běžnou součástí soudních síní, značně by to zjednodušilo a urychlilo dnes úmorný proces zjišťování faktů, což by urychlilo soudní proces a vedlo ke spravedlivějším rozsudkům. Stoupenci práva na soukromí by se však mohli obávat, že takové systémy příležitostně udělají chybu. A ještě důležitější otázka zní, jestli by naše mysl neměla být prostorem, kde vlády slídit nesmějí. Státy, které nepodporují svobodu myšlení, by mohly pomocí takovýchto technologií kriminalizovat už jen to, že někdo má jisté myšlenky a názory. Kudy byste vedli hranici mezi spravedlností a soukromím, mezi ochranou společnosti a ochranou osobní svobody? Ať už ji stanovíte kdekoli, nebude se pomalu, ale jistě posouvat na úkor soukromí, aby se kompenzovala skutečnost, že bude snadnější falšovat důkazy? Pokud bude například AI schopná vytvářet zcela realistická falešná videa, na nichž pácháte trestné činy, nebudete nakonec hlasovat

pro systém, v němž stát bude nepřetržitě sledovat, kde přesně jste, aby vám mohl v případě nutnosti poskytnout alibi?

Další zajímavý spor se týká toho, zda by výzkum AI měl být regulován. Nebo v obecnější rovině: jakou strategií pobídek pro výzkum AI zvýšit pravděpodobnost, že výsledek bude prospěšný. Někteří odborníci se staví proti jakékoli regulaci vývoje AI, protože by zbytečně brzdila nezbytné inovace (například samořídící automobily, které zachraňují tisíce životů), a stejně by jen přesunula špičkový výzkum AI do ilegality nebo do liberálnějších zemí. Na konferenci o prospěšné umělé inteligenci v Portoriku, o níž byla řeč v první kapitole, se Elon Musk nechal slyšet, že právě teď od vlád nepotřebujeme dohled, ale vhled: přesněji řečeno, technicky schopné lidi na patřičných místech státního aparátu, kteří dokážou monitorovat pokrok AI a v případě nutnosti upravit jeho směr. Upozornil také, že státní regulace mohou někdy pokrok povzbuzovat, nikoli brzdit – třeba pokud vládou stanovené bezpečnostní standardy pro samořídící auta napomohou snížení jejich nehodovosti, pak se zvýší pravděpodobnost, že nenastane prudká negativní reakce veřejnosti, což urychlí přijetí nové technologie. Opatrnější firmy v oboru by tudíž mohly být regulovaným normám nakloněné, protože to donutí k obezřetnosti i jejich méně skrupulózní konkurenty.

Jiná zajímavá právní kontroverze se týká práv strojů. Pokud by v USA jezdila jen samořídící auta a počet obětí dopravních nehod se snížil z 32 000 ročně na polovinu, nejspíše se jejich výrobci nedočkají 16 000 děkovných dopisů, ale 16 000 žalob. Kdo by měl nést odpovědnost za nehodu způsobenou samořídícím automobilem? Jeho pasažéři, vlastník nebo výrobce? Odborník na právo David Vladeck přišel s čtvrtou odpovědí: samo auto! Navrhuje, aby samořídící auta byla sama nositelem pojištění. Modely s perfektní historií bezpečného provozu by tak získávaly nárok na velmi nízké pojistné (pravděpodobně nižší, než na jaké dosáhnou lidští řidiči), a naopak špatně navržené modely od nedbalých výrobců by spadaly do tak nákladné kategorie pojištění, že vlastnit je by bylo nad možnosti kupujících.

Ovšem pokud strojům, jakými auta bezesporu jsou, dovolíte uzavírat pojistné smlouvy, neměly by také smět vlastnit majetek? Pokud by se tak stalo, nic by z právního hlediska nebránilo chytrým počítačům vydělávat si peníze a získané prostředky pak používat třeba k nákupu online služeb. Jakmile počítač bude smět platit lidem, aby pro něj pracovali, může dosáhnout čehokoli, co dokážou lidé. Pokud by nakonec systémy umělé inteligence překonaly člověka v umění investovat (což už se v některých sektorech stalo), mohlo by to vést k situaci, kdy bude většina naší ekonomiky vlastněna a kontrolována stroji. To chceme? Jestli se vám to zdá přehnané, zamyslete se nad faktem, že většina naší ekonomiky už je vlastněna entitami, jež nejsou lidmi: jsou to korporace, většinou mocnější než kterýkoli člověk v nich. Korporace, které do jisté míry mohou žít svým životem.

Pokud vám nedělá problém přiznat strojům vlastnická práva, co jim tedy přiznat i právo volební? Kdyby k tomu došlo, měl by mít po jednom hlasu každý počítačový

program? Ten si lehce vytvoří na cloudu biliony kopií sebe sama (pokud tedy bude dost bohatý), a bude tak rozhodovat ve všech volbách. A kdyby se vám to nelíbilo, na jakém morálním základu budeme diskriminovat mysl stroje oproti lidské mysli? Je nějaký rozdíl v tom, zda jsou mysli strojů nadané vědomím v témže smyslu jako my, tedy zda mají subjektivní zkušenosti? Na tyto kontroverzní otázky na téma počítačů ovládajících náš svět se blíže zaměříme v následující kapitole; otázky týkající se vědomí strojů hlouběji rozebereme v kapitole 8.

ZBRANĚ

Lidstvo od nepaměti sužují hladomor, nemoci a války. Už byla řeč o tom, kterak může umělá inteligence pomoci s řešením hladu a chorob, ale co s takovou válkou? Někteří lidé tvrdí, že jaderné zbraně odrazují státy od toho, aby vedly válku proti těm, kteří je vlastní, a že je to právě kvůli tomu, jak strašlivé tyto zbraně jsou. Co takhle dovolit všem státům, aby postavily ještě děsivější zbraně založené na AI, protože by se tím mohly války ukončit jednou provždy? Pokud vás tento argument nepřesvědčil a domníváte se, že války jsou v budoucnu nevyhnutelné, co takhle použít AI, aby tyto války byly humánnější? Kdyby proti sobě bojovaly výhradně stroje, nemuseli by umírat vojáci ani civilisté. A co víc, budoucí drony s AI a další autonomní zbraňové systémy (AWS – jejich odpůrci jim říkají „roboti zabijáci“) snad bude možné učinit čestnějšími a racionálnějšími, než jsou lidští vojáci. Jsou vybaveny senzory lepšími než lidské smysly a bez strachu ze smrti, proto si mohou zachovat rozvahu a spočítat rizika i v bitevním shonu, čímž omezí neúmyslné ztráty na civilistech.

Obrázek 3.4: Zatímco dnešní armádní drony (jako tento U. S. Air Force MQ-1 Predator) pilotují lidé na dálku, umělou inteligencí ovládané drony budoucnosti mají potenciál lidi z tohoto řetězce vypustit – na koho zacílit a koho zabít rozhodne algoritmus.

ČLOVĚK ZAPOJENÝ V PROCESU

Ale co když jsou automatické systémy plné chyb, zmatené nebo se nechovají, jak se od nich čeká? Americký systém Phalanx pro křižníky využívající Aegis automaticky detekuje, sleduje a zneškodňuje hrozby, jakými jsou například protilodní střely a letouny. USS *Vincennes* byl raketový křižník, kterému se přezdívalo Robocruiser podle systému Aegis, jímž byl vybaven. Během irácko-íránské války, dne 3. července 1988, jeho radar signalizoval uprostřed přestřelky s íránskými dělovými čluny blížící se letoun. Kapitán William Rodgers III. z toho usoudil, že na ně střemhlav útočí íránská stíhačka F-14, a dal systému Aegis souhlas k palbě. V tu chvíli si neuvědomil, že míří na civilní let 655 Iran Air. Cíl zasáhl, zahynulo všech 290 lidí na palubě a vznikl mezinárodní skandál. Z následného vyšetřování vyplynulo, že uživatelský interface byl matoucí a neukazoval automaticky, které body na obrazovce představují civilní lety (let 655 se držel své pravidelné trasy a měl zapnutý transpondér pro civilní letouny). Neindikoval ani to, která letadla klesají (třeba aby zaútočila) a která stoupají (tak jako let 655 po svém vzletu z teheránského letiště). Místo toho automatizovaný systém na dotaz o tajemném letounu odpověděl: „Klesá." To byl však status jiného letadla, kterému systém omylem přiřadil číslo civilního letu; letadlo, které klesalo, byl americký hlídkový stroj SUCAP, který operoval daleko v Ománském zálivu.

V tomto případě byl do procesu zapojen člověk, který učinil konečné rozhodnutí – a který v časové tísni příliš důvěřoval tomu, co mu sdělil automatizovaný systém. Podle vojenských představitelů všech států na světě mají zatím všechny nasazené zbraňové systémy do rozhodovacího procesu zapojeného člověka. Výjimku tvoří technicky primitivní nástražná zařízení jako pozemní miny. Vývoj dnes ovšem jde směrem ke skutečně autonomním zbraním, které si volí cíle a útočí na ně zcela samostatně. Z vojenského hlediska je lákavé z tohoto procesu úplně vyloučit lidský faktor a zvýšit tím rychlost rozhodování. Kdo myslíte, že by ve vzájemném souboji vyhrál – plně autonomní dron schopný okamžité reakce, nebo dron, který má odezvu pomalejší, protože ho přes půl světa ovládá člověk?

Nicméně se už vyskytly situace, kdy to bylo takříkajíc „o vlásek" a my jsme měli obrovské štěstí, že do rozhodovacího procesu člověk zapojen byl.

Například 27. října 1962 obklíčilo během karibské krize jedenáct torpédoborců amerického námořnictva a letadlová USS *Randolph* u Kuby sovětskou ponorku B-59, a to v mezinárodních vodách za hranicí americké „karanténní" zóny. Nevěděli ovšem, že teplota na palubě ponorky překročila 45 °C, protože jí docházely baterie, vypnula se klimatizace a mnoho členů posádky bylo v bezvědomí na otravu oxidem uhličitým. Posádka už několik dnů neměla kontakt s Moskvou a netušili, jestli mezitím nevypukla 3. světová válka. Poté Američané začali do vody vrhat malé hlubinné pumy a Moskvu informovali, že se tak jen snaží přimět ponorku, aby se vynořila a odplula; její posádka o tom ale neměla ani tušení. „Mysleli jsme, že je to tady – že přišel konec," vzpomínal V. P. Orlov. „Bylo to, jako byste seděli v plechovém barelu, do kterého někdo bez ustání mlátí železnou palicí." Američané také

nevěděli, že B-59 nese jaderné torpédo, které měli dovoleno vypustit bez další autorizace z Moskvy. Kapitán Valentin Grigorjevič Savickij se pro to skutečně rozhodl a jeho torpédový důstojník prý zvolal: „Zemřeme také, ale všechny je potopíme – nezahanbíme své námořnictvo!" Jeho rozkaz naštěstí vyžadoval schválení dvěma dalšími důstojníky na palubě a jeden z nich, Vasilij Alexandrovič Archipov, se vyslovil proti. Je zahanbující, že o Archipovovi slyšel jen málokdo, ačkoli jeho rozhodnutí nejspíše odvrátilo 3. světovou válku a bylo jedním z nejcennějších příspěvků pro lidstvo v moderních dějinách.[38] A také nás zarazí představa, co se mohlo stát, kdyby B-59 byla ponorka ovládaná AI a člověk tam neměl rozhodovací pravomoc.

O dvě desetiletí později, 9. září 1983, byly vztahy mezi velmocemi opět silně napjaté: prezident Spojených států Ronald Reagan nedávno nazval Sovětský svaz „říší zla", Sověti před týdnem sestřelili civilní let 007 Korejských aerolinií, který zabloudil do jejich vzdušného prostoru, a zabili všech 269 lidí na palubě včetně amerického kongresmana. Nyní hlásil automatický sovětský systém včasné výstrahy, že Spojené státy odpálily pět jaderných raket na Sovětský svaz. Důstojník Stanislav Jevgrafovič Petrov měl jen pár minut na rozhodnutí, jestli je poplach falešný. Zdálo se, že satelit funguje, jak má, a Petrov měl podle protokolu nahlásit blížící se jaderný útok. Místo toho však důvěřoval svým instinktům (nepřipadalo mu pravděpodobné, že by Spojené státy zaútočily jenom pěti raketami) a nadřízenému ohlásil falešný poplach, přestože si v dané chvíli nebyl jistý. Později se zjistilo, že si satelit spletl odraz Slunce od povrchu mraků s plameny raketových motorů.[39] Říkám si, jak by to asi dopadlo, kdyby na Petrovově místě byla umělá inteligence, která by protokol dodržela.

DALŠÍ ZÁVOD VE ZBROJENÍ?

Už jste si asi domysleli, že ve mně osobně vzbuzují autonomní zbraňové systémy značné obavy. A to jsem se vám ještě ani nezmínil o tom, co mě znepokojuje nejvíce: jaký by byl důsledek závodů ve zbrojení se zbraněmi s AI. V červenci 2015 jsem spolu se Stuartem Russellem tuto svou obavu vyjádřil v následujícím otevřeném dopise, přičemž mi zpětnou vazbu poskytli kolegové z Institutu budoucnosti života:[40]

> AUTONOMNÍ ZBRANĚ:
> Otevřený dopis od specialistů na AI a robotiků
> Autonomní zbraně vybírají a napadají cíle bez lidského přičinění. Může se jednat například o ozbrojené kvadrokoptéry, které mohou vyhledávat lidi splňující jistá předdefinovaná kritéria a eliminovat je, ale nespadají sem střely s plochou drahou letu nebo na dálku řízené drony, neboť u těch činí všechna rozhodnutí o zacílení lidé. Technologie umělé inteligence (AI) dosáhla bodu, kdy je nasazení takovýchto systémů prakticky, třebaže ne právně, proveditelné během několika málo let, nikoli desetiletí, a v sázce je mnoho. Autonomní zbraně byly označeny za třetí revoluci ve vedení války, po vynálezu střelného prachu a jaderných zbraních.

> Bylo vzneseno mnoho argumentů pro a proti autonomním zbraním, například že nahrazení lidských vojáků stroji je dobré proto, že snižují ztráty na dané straně. Tím ovšem také snižují hranici, kdy je strana ochotná jít do boje, což má dopad negativní. Klíčová otázka pro lidstvo dnes zní, jestli zahájit globální závod ve zbrojení zbraněmi s umělou inteligencí, nebo mu naopak zabránit. Pokud jakýkoli významný stát se silnou armádou bude prosazovat vývoj zbraní s umělou inteligencí, jsou celosvětové závody ve zbrojení de facto nevyhnutelné, a kam by vedly, je očividné: autonomní zbraně jsou účinné a levné – jsou jakýmsi kalašnikovem zítřka. Na rozdíl od jaderných zbraní nevyžaduje jejich výroba drahé ani obtížně získatelné suroviny. Proto se stanou všudypřítomnými a levnými zařízeními, která budou v masovém měřítku produkovat všichni významní hráči. Bude jen otázkou času, než se objeví na černém trhu a v rukou teroristů, diktátorů bažících po větší kontrole obyvatelstva, vojenských velitelů ovládajících svá teritoria, kteří budou chtít provést etnickou čistku, nebo v rukou dalších jim podobných. Autonomní zbraně jsou ideálním nástrojem pro atentáty, destabilizaci států, podrobení obyvatelstva a selektivní vyvražďování vybraného etnika. Věříme tedy, že závod ve zbrojení zbraněmi s umělou inteligencí by nebyl pro lidstvo přínosem. Existuje mnoho způsobů, jimiž může AI přispět k vyšší bezpečnosti bojišť pro lidi, především pro civilisty, aniž by vznikl nový nástroj k zabíjení lidí. Tak jako většina biologů a chemiků nestojí o vývoj chemických a biologických zbraní, nemá většina specialistů na AI naprosto žádný zájem o vývoj zbraní s umělou inteligencí. A nechtějí, aby jiní pošpiňovali jejich obor tím, že se o to budou snažit – mohli by totiž vyvolat silný odpor veřejnosti proti AI, a ten by snížil potenciální prospěšnost umělé inteligence pro celou společnost. Chemici a biologové široce podporují mezinárodní dohody, které úspěšně zakázaly chemické a biologické zbraně. Právě tak jako většina fyziků podporuje úmluvy zakazující orbitální jaderné zbraně a oslepující laserové zbraně.

Aby naše obavy nebylo tak lehké smést ze stolu jako výlev pacifistických aktivistů, chtěl jsem, aby náš dopis podepsalo co nejvíc významných informatiků a kybernetiků. Mezinárodní kampaň za kontrolu robotických zbraní (The International Campaign for Robotic Arms Control) už dříve shromáždila stovky podpisů volajících po zákazu robotů-zabijáků a tušil jsem, že my bychom si mohli vést ještě lépe. Věděl jsem ovšem, že profesní organizace by jen nerady poskytovaly své obrovské seznamy kontaktů na své členy kvůli něčemu, co zavání politikou. Proto jsem sestavil seznam vědců a institucí z dostupných dokumentů a crowdsourcingové platformě Mechanical Turk od Amazonu jsem zadal úkol dohledat jejich e-mailové adresy. Adresy většiny vědců visí na stránkách jejich zaměstnavatelů, takže o 24 hodin později (a po zaplacení 54 dolarů) jsem byl hrdým vlastníkem mailing listu stovek specialistů na AI, kteří byli natolik úspěšní, že byli zvoleni členy Asociace pro pokrok umělé inteligence (AAAI). Patřil mezi ně i britsko-australský profesor Toby Walsh, který se laskavě uvolil, že napíše všem ostatním ze seznamu a pomůže naší kampani prorazit. Pracovníci MTurku po celém světě mezitím pro Tobyho neúnavně

vytvářeli další mailingové seznamy a zanedlouho podepsalo náš otevřený dopis více než 3 000 výzkumníků z oborů AI a kybernetiky. Bylo mezi nimi i šest bývalých předsedů AAAI a také přední tváře AI průmyslu z Googlu, Facebooku, Microsoftu a Tesly. Armáda dobrovolníků FLI neúnavně validovala seznamy signatářů a odstraňovala žertovné podpisy jako Bill Clinton a Sára Connorová. Připojilo se přes 17 000 dalších včetně Stephena Hawkinga, a když o tom Toby na konferenci International Joint Conference of Artificial Intelligence uspořádal tiskovku, objevilo se to ve zprávách po celém světě.

Biologové a chemici se v podobné situaci jasně vyjádřili už dávno, a díky tomu jsou jejich obory daleko více známé vývojem prospěšných léků a materiálů než biologických a chemických zbraní. I komunity okolo AI a kybernetiky nyní promluvily: signatáři dopisu také chtěli, aby se jejich obory proslavily vytvářením lepší budoucnosti, nikoli novými způsoby, jak zabíjet lidi. Bude však hlavní použití AI v budoucnu civilní, nebo vojenské? Ačkoli jsme v této kapitole věnovali více stránek první zmíněné variantě, už brzy možná budeme utrácet více peněz za to druhé – zvlášť pokud se rozběhne závod ve zbrojení s použitím AI. Civilní investice do AI přesáhly v roce 2016 miliardu dolarů, to však zcela zastínila žádost Pentagonu pro fiskální rok 2017, kdy z rozpočtu požadoval 12–15 miliard dolarů na projekty z oboru umělé inteligence. A Čína a Rusko si patrně dobře povšimly, co při této příležitosti prohlásil náměstek ministra obrany Spojených států Robert Work: „Chci, aby si naši soupeři lámali hlavu, co je za tou černou oponou.“[41]

MĚLI BYCHOM MÍT MEZINÁRODNÍ ÚMLUVU?

I když už dnes existuje značný mezinárodní tlak, aby se vyjednal nějaký zákaz zabijáckých robotů, není jasné, co se bude dít dál. Probíhá živá debata, co a – jestli vůbec něco – by se *mělo* stát. I když se mnoho předních zainteresovaných osob shoduje, že světové velmoci by měly načrtnout jakási mezinárodní pravidla regulující výzkum a užití autonomních zbraňových systémů (AWS), již mnohem menší shoda je na tom, co přesně by mělo být zakázáno a jak takový zákaz vymáhat. Například zda by se měly zakázat pouze smrtící autonomní zbraně, nebo i ty, které způsobují vážná zranění – které dejme tomu oslepují? Měli bychom zakázat jejich vývoj, výrobu nebo vlastnictví? Měl by se zákaz vztahovat na všechny autonomní zbraňové systémy, nebo pouze na ty útočné, jak stojí v našem dopise, takže by byly povoleny systémy obranné, jako třeba autonomní protiletadlové zbraně a protiraketová obrana? Co se týče druhého případu: měly by se AWS považovat za obranné, i když je lze snadno převézt na nepřátelské území? A jak byste dodržování takové úmluvy vymáhali, když většina komponentů pro autonomní zbraně má dvojí, tedy i civilní, využití? Tak například není velký rozdíl mezi dronem roznášejícím zásilky pro Amazon a dronem roznášejícím bomby.

Někteří účastníci této debaty se nechali slyšet, že sepsání efektivní úmluvy o autonomních zbraňových systémech je neschůdné, a proto bychom se o to ani neměli pokoušet. Na druhou stranu když J. F. Kennedy ohlašoval cíl přistát na Měsíci,

zdůraznil, že stojí za to usilovat o dosažení náročných cílů, když budoucnosti lidstva přinesou mnoho pozitiv. Řada odborníků navíc argumentuje, že zákazy biologických a chemických zbraní byly cenné i přesto, že se ukázalo, že vynucování těchto pravidel znesnadňují podvody. Zákazy totiž přinesly stigmatizaci, která používání těchto zbraní významně omezila.

Na jedné večeři jsem se v roce 2016 setkal s Henrym Kissingerem a měl jsem příležitost se ho zeptat na jeho roli v zákazu biologických zbraní. Vyložil mi, jak byl tehdy poradcem pro národní bezpečnost a přesvědčil prezidenta Nixona, že tento zákaz prospěje národní bezpečnosti Spojených států amerických. Udělalo na mě hluboký dojem, jak perfektně mu slouží mysl i paměť – bylo mu tehdy dvaadevadesát – a bylo fascinující poslechnout si jeho pohled zevnitř. Spojené státy už tehdy kvůli svým konvenčním a jaderným silám měly status supervelmoci, a proto v globálním závodu ve zbrojení biologickými zbraněmi měly více co ztratit než získat, neboť výsledek takového závodu by byl nejasný. Jinými slovy, když už jste první, má smysl řídit se heslem „Jestli to není rozbité, nespravuj to". K našemu rozhovoru se po večeři připojil Stuart Russell a bavili jsme se o tom, jak tuto argumentaci použít i v případě smrtících autonomních zbraní. Největší naději na zisk ze závodu ve zbrojení nemají velmoci, ale malé „darebácké státy" a nestátní síly jako teroristé, kteří se k těmto zbraním dostanou na černém trhu, jakmile bude jejich vývoj dokončen.

Kdyby se vyráběly ve velkém, stál by takový malý zabijácký dron disponující AI jen o něco více než smartphone. Ať teroristovi, který hodlá spáchat atentát na politika, nebo zhrzenému milenci, jenž se chce pomstít bývalé přítelkyni, každému stačí jen nahrát fotografii a adresu cíle do zabijáckého dronu. Dron pak doletí na místo určení, danou osobu identifikuje a eliminuje – a třeba se zničí, aby smazal stopy. Pro ty, kteří si předsevzali provést etnickou čistku, by dron nabízel možnost snadného naprogramování k tomu, aby zabíjel jen osoby určité barvy pleti či etnické příslušnosti. Stuart se domnívá, že čím chytřejšími se takové zbraně stanou, tím méně materiálu, palebné síly a finančních prostředků bude zapotřebí na jedno zabití. Obává se například dronů o velikosti čmeláka zabíjejících levně za použití minimální výbušné síly. Střílely by do oka, jelikož je dost měkké na to, aby dovolilo i drobnému projektilu proniknout až do mozku. Nebo by se mohly přichytit k hlavě kovovými drápky a pak prorazit lebku malou tvarovanou náloží. Ve chvíli, kdy lze milion takových zabijáckých dronů vypustit z jediného kamionu, máte zcela nový druh strašlivé zbraně hromadného ničení. Zbraň, která dokáže selektivně zabít pouze předem danou kategorii lidí, zatímco vše ostatní nechá nedotčené.

Běžný protiargument zní, že zabijáckým robotům můžeme dodat etiku – třeba aby zabíjeli jen nepřátelské vojáky. Máme-li však obavy z vymahatelnosti úplného zákazu, jak by se asi vymáhal požadavek, aby nepřátelské autonomní zbraně byly stoprocentně etické? A lze zodpovědně prohlásit, že dobře vycvičení vojáci civilizovaných zemí dodržují pravidla války tak špatně, že to roboti dokážou lépe?

A přitom spoléhat, že darebácké státy, diktátoři a teroristické skupiny budou tato pravidla vedení války ctít natolik, že se nikdy neuchýlí k nasazení robotů, kteří by je porušovali?

KYBERVÁLKA

Dalším zajímavým vojenským aspektem AI je, že vám umožní útok na nepřítele, aniž byste museli stavět vlastní zbraně. Stačí vést kyberválku. Jako ochutnávka, co může přinést budoucnost, poslouží worm Stuxnet, všeobecně připisovaný americké a izraelské vládě. Napadl na přelomu let 2009 a 2010 centrifugy íránského jaderného programu na obohacování uranu a způsobil, že se roztrhaly. Čím automatizovanější se společnost stává, tím větší sílu má útok na umělou inteligenci a tím ničivější následky může kyberválka mít. Pokud dokážete hacknout nepříteli jeho samořídící auta, letadla řízená autopilotem, jaderné reaktory, průmyslové roboty, komunikační systémy, systém finančnictví a rozvodu elektřiny, pak můžete srazit na kolena jeho ekonomiku a ochromit jeho obranu. A jestli se vám při tom podaří proniknout do některé z jeho zbraní, tím lépe.

Tuto kapitolu jsme začali mapováním zářných možností, jež může umělá inteligence lidstvu v dohledné době přinést – pokud se nám ji podaří učinit robustní a odolnou vůči hackerským útokům. Ačkoli na zvyšování robustnosti systémů AI může pracovat i sama AI (a v kyberválce být tedy v obraně), může nesporně působit i na straně útoku. Jedním z klíčových krátkodobých cílů ve vývoji AI proto musí být zajištění, aby obrana měla navrch – jinak se může všechna ta úžasná technologie, kterou vytváříme, obrátit proti nám!

PRÁCE A MZDA

V této kapitole jsme se zatím zabývali především otázkou, jak nás umělá inteligence ovlivní coby *konzumenty* tím, že nám nabídne převratné nové výrobky a služby za dostupné ceny. Jak nás ale ovlivní coby *zaměstnance* tím, že změní pracovní trh? Pokud se nám podaří zjistit, jak zvyšovat blahobyt pomocí automatizace, aniž by lidé přicházeli o příjmy či přímo o smysl života, pak bychom měli příležitost vytvořit fantastickou budoucnost plnou volného času a nebývalé hojnosti pro každého, komu se zachce. Jen málokdo o tom přemýšlel déle a usilovněji než ekonom Erik Brynjólfsson, jeden z mých kolegů na MIT. Přestože chodí vždycky dobře upravený a perfektně oblečený, má islandské kořeny a někdy se nemohu ubránit představě, že si teprve nedávno zastřihl zrzavé vikinské vousy a vlasy, aby zapadl na naší ekonomické fakultě. Určitě si ale nepřistřihl své divoké názory ani optimistickou vizi pracovního trhu, kterou nazývá „digitální Athény". Občané Athén si v antice mohli užívat zahálčivý život naplněný demokracií, uměním a hrami především proto, že většinu práce odváděli otroci. Co kdybychom otroky nahradili roboty s AI, a vytvořili tak digitální utopii, z níž by se mohli těšit všichni? Erikova ekonomika poháněná AI by eliminovala stres a dřinu a produkovala by hojnost všeho, co dnes používáme.

A nadto by dodávala mnoho skvělých nových produktů a služeb, o nichž dnešní zákazníci ještě ani netuší, že je chtějí.

TECHNOLOGIE A NEROVNOST

Z dnešního stavu se do Erikových digitálních Athén můžeme dostat, pokud budou hodinové mzdy všech lidí rok od roku růst, aby ti, kdo stojí o více volného času, mohli pracovat méně a přitom se jim životní standard i nadále zvyšoval. Obrázek 3.5 ukazuje, že přesně to se dělo ve Spojených státech od konce 2. světové války do poloviny 70. let. Navzdory příjmové nerovnosti rostla celková velikost koláče natolik, že se skoro na každého dostal větší kousek. Pak se ovšem – a Erik to přiznává jako první – cosi změnilo. Týž obrázek 3.5 ilustruje, že navzdory pokračujícímu růstu ekonomiky a zvyšování průměrného příjmu šly v posledních čtyřech dekádách zisky hlavně těm nejbohatším, povětšinou hornímu jednomu procentu, zatímco příjmy nejchudších 90 % stagnovaly. Nárůst nerovnosti bude ještě jasnější, když se místo na příjem podíváme na majetek. Dolních 90 % amerických domácností mělo v roce 2012 průměrné čisté jmění o hodnotě přibližně 85 000 dolarů – tedy stejné jako o pětadvacet let dříve – zatímco horní 1 % za tu dobu své jmění více než zdvojnásobilo (vše po očištění od inflace), a to na 14 milionů dolarů.[42] Rozdíly jsou ještě markantnější v mezinárodním srovnání, kde v roce 2013 celkové jmění spodní poloviny světové populace (přes 3,6 miliardy lidí) dosahuje stejné výše jako jmění osmi nejbohatších lidí světa.[43] Tato statistika ukazuje chudobu a zranitelnost těch dole právě tak jako bohatství těch nahoře. V roce 2015 řekl Erik na naší konferenci v Portoriku shromážděným výzkumníkům, že podle něj bude pokrok AI a automatizace i nadále ekonomický „koláč" zvětšovat, ale že žádný zákon ekonomiky neříká, že z toho budou mít prospěch všichni lidé (nebo alespoň většina z nich).

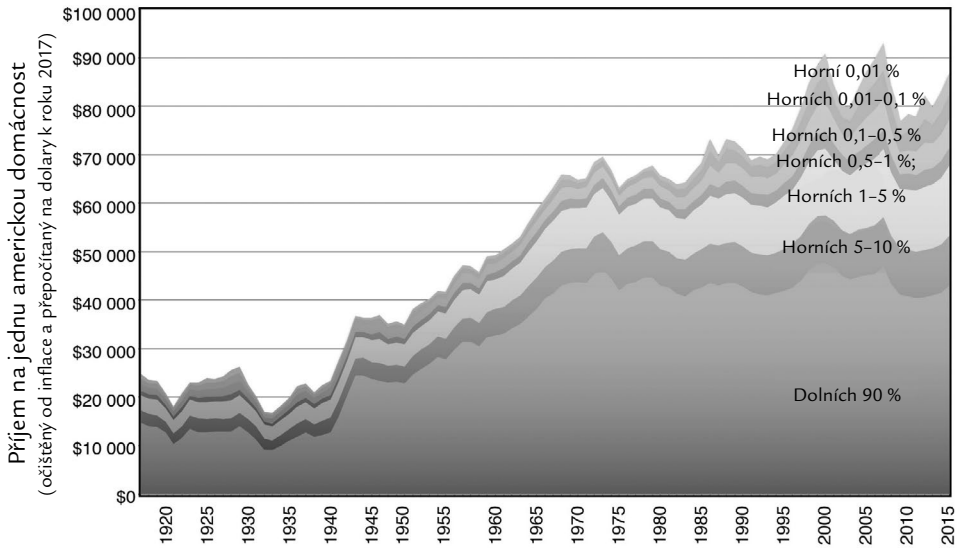
Ačkoli mezi ekonomy panuje shoda, že se nerovnost zvyšuje, existuje zajímavý spor, proč tomu tak je a zda tento trend bude pokračovat. Účastníci této debaty z levé strany politického spektra často prohlašují, že hlavní příčinou je globalizace nebo ekonomická opatření jako úlevy na daních pro bohaté, nicméně podle Erika Brynjólfssona a jeho spolupracovníka z MIT Andrewa McAfeeho je hlavním důvodem cosi jiného: je to technologie.[44] Konkrétně tvrdí, že digitální technologie zvyšuje nerovnost třemi různými způsoby.

Zaprvé technologie odměňuje vzdělané lidi tím, že nahrazuje staré pracovní pozice novými, které vyžadují více schopností. Od poloviny 70. let vzrostly lidem s vysokoškolským vzděláním mzdy přibližně o 25 %, zatímco průměrnému středoškolákovi se mzda snížila o 30 %.[45]

Zadruhé podle nich plyne od roku 2000 stále větší podíl z příjmů právnických osob těm, kdo společnosti vlastní, a to na úkor lidí, kteří v nich pracují. A dokud bude automatizace pokračovat, měli bychom prý očekávat, že vlastníci strojů si budou z onoho „koláče" brát čím dál větší část. Taková převaha kapitálu nad prací může být zvlášť důležitá pro rostoucí digitální ekonomiku, jíž technický vizionář

Nicholas Negroponte definuje jako přesouvání informací místo hmoty. Dnes, když vše od knih po filmy a nástroje pro zpracování daňového přiznání dostalo digitální formu, mohou být další kopie prodávány celosvětově s takřka nulovými náklady, aniž je třeba najímat další zaměstnance. Kvůli tomu může jít většina tržeb investorům, nikoli pracovní síle. Pomáhá nám to také vysvětlit, proč příjmy detroitské „Velké trojky" (GM, Fordu a Chrysleru) z roku 1990 dosahovaly téže výše jako příjmy „Velké trojky" ze Silicon Valley (Googlu, Applu a Facebooku) z roku 2014, přestože ta druhá skupina měla devětkrát méně zaměstnanců a její tržní kapitalizace (souhrnná cena akcií) byla třicetkrát vyšší.[47]

Zatřetí Erik se svými kolegy tvrdí, že digitální ekonomika mnohem více favorizuje absolutní špičku každého oboru před všemi ostatními. Autorka Harryho Pottera J. K. Rowlingová coby první spisovatel(ka) vstoupila do klubu miliardářů. Svým bohatstvím dalece předčila Williama Shakespeara, jelikož se její příběhy digitálně šíří formou textů, filmů i her a dostávají se k miliardám lidí s velmi nízkými náklady. Podobně vydělal Scott Cook miliardu dolarů na softwaru pro zpracování daní Turbo Tax, který (na rozdíl od lidí najatých na tento úkol) může být stahován a prodáván s minimálními náklady na každého nového klienta. A protože je většina lidí ochotna zaplatit za řekněme desátý nejlepší daňový software jen velmi málo nebo nic, zbývá na trhu místo pouze pro skromný počet špičkových hráčů. Z čehož plyne, že i kdyby všichni rodiče na světě radili svým dětem, aby se staly další Joanne Rowlingovou, Gisele Bündchenovou, Mattem Damonem, Cristianem



Obrázek 3.5: Jak ekonomika za posledních sto let zvýšila průměrný příjem a jaká část šla kterým skupinám. Do 70. let to vypadá, že bohatí i chudí bohatli ruku v ruce, poté ale většina zisků odcházela hornímu 1 %, zatímco dolních 90 % v průměru nezískalo téměř nic.[46] Údaje byly očištěny od inflace a přepočítány na dolary z roku 2017.

Ronaldem, Oprah Winfreyovou nebo Elonem Muskem, jen u málokteré z oněch ratolestí bude taková strategie fungovat.

JAKOU PORADIT DĚTEM KARIÉRU

Jakou kariéru bychom tedy měli poradit svým dětem? Já sám podporuji ty své, aby si zvolily profese, v nichž jsou stroje v současnosti špatné, a kde je tudíž nepravděpodobné, že by v blízké budoucnosti proběhla automatizace. Nedávný odhad horizontů, kdy budou různé pracovní pozice ovládnuty stroji, přichází s řadou užitečných otázek, které byste si měli položit, než se rozhodnete pro vzdělání vyžadované pro určitou kariéru.[48] Například:

- Vyžaduje to interakci s lidmi a sociální inteligenci?
- Vyžaduje to kreativitu a vymýšlení chytrých řešení?
- Vyžaduje to práci v nepředvídatelném prostředí?

Čím víc těchto otázek můžete zodpovědět kladně, tím lepší asi vaše volba povolání bude. Z toho plyne, že relativně nízké riziko přináší sázka na učitele, sestřičku, lékaře, zubaře, vědce, podnikatele, programátora, inženýra, právníka, sociálního pracovníka, kněze, umělce, kadeřnici nebo maséra.

Naopak práce, které obnášejí opakující se a strukturované akce v prostředí, jehož chování lze snadno předpovědět, pravděpodobně nebudou na automatizaci čekat dlouho a nevydrží. Počítače a průmysloví roboti takové nejjednodušší práce převzaly už dávno. A zlepšující se technologie už je na nejlepší cestě, aby eliminovala mnoho dalších – od telemarketingu po skladníky, pokladní, strojvedoucí, pekaře a kuchaře u výrobních linek.[49] Řidiči kamionů, autobusů, taxi a Uberu či Lyftu pak budou patrně následovat záhy. Mnoho dalších profesí (včetně právních asistentů, úvěrových analytiků, úvěrových úředníků, klasických účetních a účetních pro daňovou oblast) sice na seznamu kriticky ohrožených druhů nefiguruje, nicméně většina jejich úkolů se už automatizuje, a proto je zde poptávka po lidech mnohem nižší.

Vyhnout se automatizaci ovšem není jedinou podmínkou úspěšné kariéry. V této globalizované digitální době je rozhodnutí stát se profesionálním spisovatelem, filmařem, hercem, atletem nebo módním návrhářem riskantní ještě z dalšího důvodu. Ačkoli lidem z těchto profesí nebude v brzké době hrozit vážná konkurence od strojů, budou čelit stále tvrdší konkurenci jiných lidí z celého světa. A podle výše zmíněné teorie absolutních špiček uspěje jen hrstka.

V mnoha případech by bylo příliš krátkozraké a zjednodušující dávat kariérní rady na úrovni celých oblastí. Řada pozic nevymizí zcela, nicméně značná část jejich úkolů projde automatizací. Například pokud chcete být lékařem, nestaňte se radiologem, který analyzuje diagnostické snímky a bude nahrazen Watsonem od IBM, ale lékařem, který analytickému softwaru dává příkazy, probírá výsledky s pacientem a rozhoduje o plánu léčby. Jestli se dáte na finance, nebuďte datovým analytikem, který pomocí softwaru zpracovává data a sám bude softwarem nahrazen,

ale investičním manažerem, který používá výsledky kvantitativní analýzy k tomu, aby dělal strategická investiční rozhodnutí. V advokacii nebuďte právním asistentem, který reviduje tisíce dokumentů pro rešerše a vytlačí ho automatizace, ale obhájcem, který poskytuje poradenství klientům a prezentuje případy u soudu.

Dosud jsme se věnovali otázce, co může jedinec udělat, aby maximalizoval svůj úspěch na trhu práce ve věku AI. Co však mohou udělat státy pro to, aby pomohly své pracovní síle? Třeba jaký vzdělávací systém nejlépe připraví lidi pro práci za situace, že se AI překotně vyvíjí? Je jím stále náš současný model s jednou či dvěma dekádami vzdělání, po němž následují čtyři desetiletí specializované práce? Nebo by bylo lepší přejít na systém, kde lidé několik let pracují, pak se na rok vrátí do školy a potom dalších pár let zase pracují?[50] Nemělo by se průběžné vzdělávání (snad poskytované online) stát standardní součástí zaměstnání?

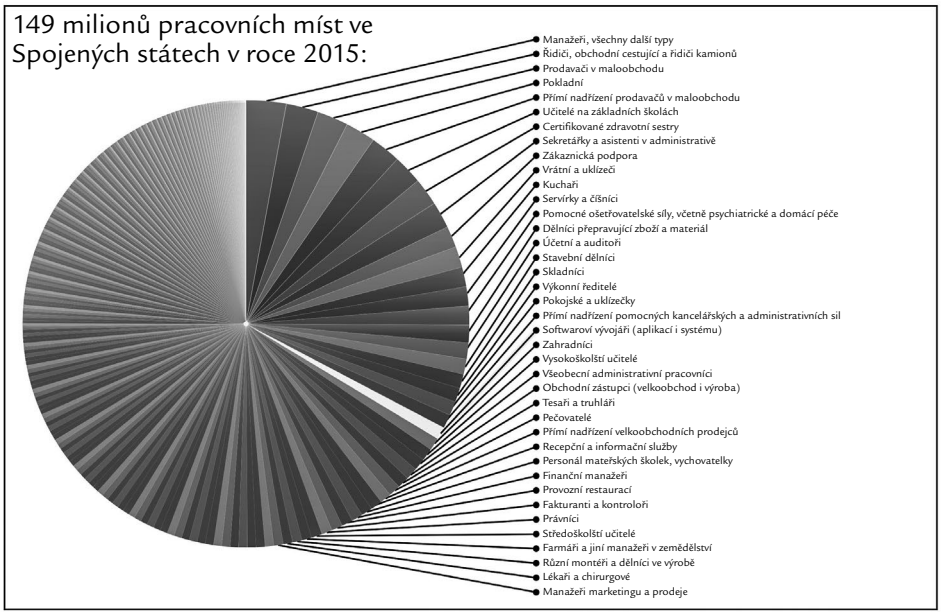
A jaké ekonomické kroky nejspíše pomohou při vytváření nových pracovních míst? Andrew McAfee tvrdí, že potenciálně užitečných opatření je více. Zahrnují například vysoké investice do výzkumu, vzdělání a infrastruktury, usnadňování migrace a pobídky podnikání. Domnívá se, že „pravidla manažerské metody Econ 101 znějí jasně, ale nedodržují se", alespoň ne ve Spojených státech.[51]

BUDOU LIDÉ NAKONEC NEZAMĚSTNATELNÍ?

Co se stane, pokud se bude umělá inteligence stále zlepšovat a automatizovat víc a víc pracovních pozic? Mnoho lidí zastává optimistický názor, podle něhož budou automatizované práce nahrazeny novými, ještě lepšími. To se přece stalo zatím pokaždé – od doby, kdy se luddité začali obávat technologické nezaměstnanosti během průmyslové revoluce.

Jiní jsou v otázce práce pesimističtější, tentokrát je to prý jiné a stále větší procento populace se stane nejen nezaměstnaným, ale i nezaměstnatelným.[52] Tito pesimisté vysvětlují, že volný trh nastavuje výši mezd podle nabídky a poptávky a že rostoucí nabídka levné strojové práce nakonec posune odměnu za lidskou práci hluboko pod životní náklady. Jelikož tržní mzda za práci činí hodinovou cenu kohokoli, kdo ji vykoná nejlevněji, mzdy klesaly vždy, když bylo možné určitou činnost přesunout do země s nižšími příjmy, případně levnému stroji. V průběhu průmyslové revoluce jsme začali zjišťovat, jak nahradit své svaly stroji, a lidé se uchýlili k lépe placeným pracím, kde se více využívá rozum. Místo montérek nastoupily bílé košile. A nyní pomalu přicházíme na to, jak stroji nahradit i svůj mozek. Pokud se nám to nakonec podaří, jaké pracovní pozice na nás zbudou?

Někteří optimisté argumentují, že po fyzických a duševních pracích nastane další boom v pracích *kreativních*, ovšem podle pesimistů je kreativita jen jedním z duševních procesů, a i ji jednou AI zvládne. Jiní optimisté doufají, že další boom se namísto toho odehraje v nových, technologiemi vytvořených profesích, na něž jsme zatím ani nepomysleli. Vždyť koho by v době průmyslové revoluce napadlo, že jejich potomci budou jednoho dne pracovat jako weboví designéři nebo řidiči Uberu? Pesimisté ovšem odporují, že to je jen zbožné přání, které se neopírá o empirická



Obrázek 3.6: Tento koláčový graf ukazuje povolání oněch 149 milionů Američanů, kteří v roce 2015 měli práci, a zahrnuje 535 kategorií prací amerického Bureau of Labor Statistics seřazených podle oblíbenosti.[53] Popsána jsou všechna povolání s více než milionem pracovníků. Až do dvacátého místa tam nefigurují žádné nové práce vytvořené výpočetní technologií. Graf je založen na analýze Federica Pistona.[54]

data. Upozorňují, že totéž jsme mohli říkat před sto lety (před počítačovou revolucí) a předpovědět, že většina dnešních povolání bude nová a umožněná dříve neexistujícími a nepředstavitelnými technologiemi. Taková předpověď by byla kolosálně mimo, jak je vidět na obrázku 3.6: naprostá většina dnešních povolání existovala již před sto lety. A když je seřadíme podle počtu pracovních míst, která poskytují, narazíme na nové zaměstnání až na jednadvacátém místě. Tím jsou softwaroví vývojáři, kteří tvoří necelé 1 % pracovního trhu Spojených států.

Situaci nám pomůže pochopit obrázek 2.2 z kapitoly 2, který znázorňoval krajinu lidské inteligence, přičemž nadmořská výška vyznačovala, nakolik je pro stroje obtížné dané úkoly splnit. Stoupající mořská hladina pak reprezentovala, které činnosti už stroje vykonávat dokážou. Převládajícím trendem na trhu práce není přesun ke zcela novým profesím. Místo toho se shlukujeme na těch místech krajiny na obrázku 2.2, která ještě nezatopil stoupající příliv technologie! Obrázek 3.6 pak znázorňuje, že se nejedná o jediný ostrov, ale o komplexní souostroví s ostrůvky a korálovými ostrovy odpovídající všem hodnotným věcem, které stroje pořád ještě nedokážou provádět tak levně jako lidé. Patří tam nejen vysoce technologické profese jako softwarový vývojář, ale také celá paleta na technologii nenáročných prací, kde využíváme svou výjimečnou obratnost a společenské schopnosti – od maséra po herce.

Mohla by nás umělá inteligence předhonit v intelektuální oblasti tak rychle, že by poslední zbývající pracovní pozice patřily do málo technologické skupiny? Jeden můj přítel mi nedávno žertem povídal, že třeba bude poslední profesí ta nejstarší: prostituce. Ovšem když to řekl jistému japonskému kybernetikovi, setkal se s nesouhlasem: „To ne, roboti jsou v takových věcech opravdu dobří!"

Pesimisté tvrdí, že výsledek je nasnadě: pod vodou se ocitne celé souostroví a nezůstanou žádné pracovní pozice, kde by lidé byli levnější než stroje. Skotsko-americký ekonom Gregory Clark ve své knize z roku 2007 *A Farewell to Alms* (Sbohem, almužno) podotýká, že se pár věcí o budoucích pracovních vyhlídkách můžeme dozvědět na základě následující bajky.

> Dva koně se v roce 1900 dívají na raný model automobilu a přemýšlejí o své budoucnosti.
>
> „Dělá mi starosti technologická nezaměstnanost."
>
> „Ale ne, nebuď luddita. Naši předci říkali totéž, když nám parní stroje sebraly práci v průmyslu a vlaky nám vzaly práci při tahání dostavníků. Zato teď máme více pracovních míst než kdy dříve – a taky lepších. Mnohem radši bych tahal lehký kočár po městě, než abych trávil celý den chozením v kruhu a poháněním hloupého čerpadla ze šachty."
>
> „No a co když se tenhle spalovací motor vážně uchytí?"
>
> „Já jsem přesvědčený, že se pro koně objeví nová pracovní místa, která nás ještě ani nenapadla. Vždycky to tak bylo – jako když vynalezli kolo a pluh."

Žel, ona pracovní místa pro koně, o nichž se nikomu ještě ani nesnilo, se nikdy neobjevila. Nyní již nepotřební koně bez náhrady putovali na porážku, takže se v USA počet koní zmenšil z přibližně 26 milionů v roce 1915 na asi 3 miliony v roce 1960.[55] Kvůli mechanickým svalům byli koně najednou přebyteční. Budou mít mechanické mysli tentýž efekt u lidí?

DÁVAT LIDEM PENÍZE BEZ PRÁCE

Kdo má tedy pravdu: ti, kdo říkají, že automatizované práce budou nahrazeny lepšími, nebo ti, kteří tvrdí, že většina lidí skončí jako nezaměstnatelní? Pokud se AI bude dál vyvíjet se stejnou intenzitou, dostanou možná zapravdu *obě* strany – jedna v krátkodobém horizontu, ta druhá v dlouhodobém. Ovšem přestože lidé často vidí mizení pracovních míst černě, vůbec to nemusí být špatná věc! Luddité byli posedlí konkrétními pracemi a opomíjeli možnost, že tutéž společenskou hodnotu mohou poskytovat práce jiné. Podobně jsou snad příliš úzkoprsí ti, kteří dnes nedávají pokoj s pracovními místy: chceme pracovní místa, protože ta nám mohou dát příjem a smysl života. Nicméně vzhledem k bohatství zdrojů vytvářených stroji bychom mohli nalézt alternativní způsoby, jak obstarat finanční příjmy a smysl existence i *bez* zaměstnání. Něco takového se stalo v onom koňském příběhu: neskončilo to vyhynutím všech koní. Místo toho se počet koní od roku

1960 více než ztrojnásobil, ochránil je svého druhu systém sociálního zabezpečení koní. Přestože nemohli platit své účty, rozhodli se o ně lidé starat, mít je u sebe pro zábavu, kvůli sportu a proto, aby jim dělali společnost. Můžeme se podobně ujmout našich bližních, lidí v nouzi?

Začněme u otázky příjmu: přerozdělování (byť jen malého zlomku rostoucího „koláče" ekonomiky) by mělo pomoci k lepšímu životu každému. Řada lidí prohlašuje, že to provést nejen *můžeme*, ale že bychom to udělat *měli*. Na panelu v roce 2016, kde Moshe Vardi mluvil o morální povinnosti zachraňovat životy technologiemi s umělou inteligencí, jsem prohlásil, že dalším morálním imperativem je zasazovat se o to, aby byly používány prospěšně, včetně sdílení blahobytu. Erik Brynjólfsson, který se onoho panelu také zúčastnil, řekl, že „jestli se vším tím nově vytvořeným bohatstvím nedokážeme zabránit zhoršení finanční situace poloviny lidí, měli bychom se stydět!".

Existuje řada různých návrhů, jak se dělit o bohatství, a každý má své zastánce a odpůrce. Nejjednodušším z nich je *základní nepodmíněný příjem*, kdy každý člověk dostává měsíční peněžní dávky bez jakýchkoli podmínek nebo požadavků. V současnosti se plánuje či dokonce v praxi zkouší několik experimentů menšího rozměru, mimo jiné v Kanadě, ve Finsku a v Nizozemsku. Zastánci tohoto modelu jsou přesvědčeni, že základní nepodmíněný příjem je efektivnější než jeho alternativy, jako například sociální dávky potřebným, protože tak odpadne administrativní zátěž při zjišťování, kdo má nárok. Sociální dávky založené na hmotné nouzi se také staly terčem kritiky, protože snižují motivaci k práci. To však bude irelevantní v budoucnosti, kdy žádná pracovní místa nebudou a nikdo nebude pracovat.

Státy mohou svým občanům pomáhat nejen rozdáváním peněz, ale také tím, že zdarma nebo za dotovanou cenu poskytují služby, jakými jsou silnice, mosty, parky, veřejná doprava, péče o děti, vzdělávání, zdravotní péče, domovy důchodců a internetové připojení. Mnoho vlád už skutečně poskytuje většinu těchto služeb. Na rozdíl od nepodmíněného příjmu mají takovéto státem dotované služby dva rozdílné cíle: snižují lidem náklady na život a zároveň poskytují pracovní místa. Dokonce i v budoucnosti, kdy stroje překonají lidi ve všech zaměstnáních, by se státy mohly rozhodnout platit lidem za práci v péči o děti, seniory a tak dále, místo aby pečovatelské pozice přenechaly robotům.

Technologický pokrok může kupodivu nakonec přinést mnoho cenných produktů a služeb zdarma i bez přičinění vlády. Lidé například platívali za encyklopedie, mapy, posílání dopisů a telefonické hovory, zatímco dnes má každý s přístupem k internetu tyto věci zadarmo – společně s videokonferencemi, sdílením fotografií, sociálními médii, online kurzy a nesčetnými dalšími službami bez poplatku. Cena mnoha dalších věcí, které mohou být pro jedince takřka neocenitelné, jako řekněme životně důležité kurzy o antibiotikách, výrazně klesla. Díky technologii má dnes množství chudých lidí přístup k věcem, které v minulosti neměli ani ti nejbohatší. Podle některých to znamená, že výše příjmu nutného k důstojnému životu se snižuje.

Pokud jednoho dne budou stroje vytvářet všechny statky a zboží současnosti s minimálními náklady, pak zjevně existuje dostatek bohatství na to, abychom se měli lépe všichni. Jinými slovy: i poměrně nízké zdanění by vládám umožnilo zaplatit základní nepodmíněný příjem a služby zdarma. Nicméně fakt, že sdílení bohatství se *může* stát skutečností, neznamená, že se to opravdu *stane*. V současnosti probíhá intenzivní politický spor, zda by se to vůbec stát *mělo*. Jak jsme viděli výše, podle všeho jde současný trend ve Spojených státech opačným směrem. Některé skupiny obyvatelstva dekádu od dekády chudnou. Politická rozhodnutí o tom, jak sdílet rostoucí bohatství společnosti, bude mít dopad na všechny, na diskusi o budoucí podobě ekonomiky by se proto měli podílet všichni, nejen informatici, kybernetici a ekonomové.

Mnoho účastníků této debaty tvrdí, že snížení příjmové nerovnosti je dobrý nápad nejen v budoucnosti, kde dominuje AI, ale i dnes. Ačkoli hlavní argument bývá morální, prokázalo se také, že větší rovnost vede k lépe fungující demokracii. Když má početná střední třída dobré vzdělání, není tak snadné manipulovat voliči a pro malý počet lidí či společností je pak obtížnější kupovat si nepatřičný vliv na řízení státu. Lepší demokracie může na oplátku vést k lépe vedené ekonomice, která je méně zkorumpovaná, efektivnější a roste rychleji. V konečném důsledku z toho budou mít užitek prakticky všichni.

DÁVAT LIDEM SMYSL BEZ ZAMĚSTNÁNÍ

Pracovní místa mohou lidem dávat více než jen peníze. V roce 1759 Voltaire napsal, že „práce od nás odvrací tři velká zla. Nudu, neřesti a nouzi.“* Naopak obstarat lidem příjem samo o sobě nezajistí jejich štěstí. Římští císaři obstarávali chléb a hry, aby udrželi lůzu spokojenou, a Ježíš zdůraznil nehmotné potřeby v biblickém citátu: „Ne samým chlebem živ bude člověk.“** Co cenného tedy přesně zaměstnání kromě peněz přináší a jakými alternativními způsoby to může společnost bez pracovních míst zajistit?

Odpovědi na tyto otázky jsou očividně komplikované, protože někteří lidé svou práci nenávidí, zatímco jiní ji zbožňují. Navíc mnoha dětem, studentům a ženám v domácnosti se daří skvěle bez práce, ale historie je plná příběhů o rozmazlených dědicích a princích, kteří podlehli nudě a depresi. Jistá metaanalýza v roce 2012 ukázala, že nezaměstnanost mívá dlouhodobé negativní dopady na pocit pohody a spokojenosti, ovšem důchod mívá smíšenou negativní i pozitivní stránku.[56] Rozvíjející se obor *pozitivní psychologie* identifikoval celou řadu faktorů, které podporují lidský pocit spokojenosti a smyslu. Zjistilo se, že některé (ale ne všechny!) práce jich mohou skýtat hned několik, patří mezi ně například:[57]

- společenská síť přátel a kolegů
- zdravý a správný životní styl

* Voltaire: *Candide*. Přel. Radovan Krátký. Praha: Svoboda, 1949, s. 182.

** Bible kralická, Matouš 4.4.

- respekt, sebevědomí, vědomí vlastních kompetencí a nadšený pocit zabrání do práce, v níž jste dobří
- pocit, že vás někdo potřebuje a že můžete něco změnit
- pocit, že jste součástí něčeho, co vás přesahuje a čemu můžete přispět

To poskytuje příčiny k optimismu, jelikož všechny tyto věci lze najít i mimo pracoviště, například ve sportu, provozování koníčků či při učení, a také při kontaktu s rodinou, přáteli, v týmech, klubech, komunitních skupinách, ve školách, náboženských a nenáboženských organizacích, politických hnutích a jiných institucích. Abychom vytvořili společnost s nízkou zaměstnaností, která vzkvétá a nedegeneruje k sebezničujícímu chování, musíme pochopit, jak pomáhat takovým aktivitám, které podporují pocit spokojenosti. Hledání se musí účastnit nejen vědci a ekonomové, ale i psychologové, sociologové a pedagogové. Pokud se budeme usilovně snažit o vytvoření spokojeného života pro všechny, částečně placeného bohatstvím, které vytvoří budoucí AI, pak by společnost mohla vzkvétat jako nikdy dříve. Přinejmenším by mělo být možné učinit každého tak šťastným, jako by měl svou vysněnou práci, ale jakmile se osvobodíme od nutnosti vytvářet každou činností zisk, nic nám nestojí v cestě.

INTELIGENCE NA LIDSKÉ ÚROVNI?

V této kapitole jsme se zabývali otázkou, jaký potenciál má umělá inteligence vylepšit naše životy v krátkodobém horizontu – pokud budeme plánovat dopředu a vyhneme se různým nástrahám. Ale co horizont dlouhodobý? Nezastaví se nakonec pokrok AI kvůli nepřekonatelným překážkám? Podaří se jejímu výzkumu dosáhnout původního cíle, tedy vytvořit umělé bytí na lidské úrovni? V minulé kapitole jsme viděli, že fyzikální zákony umožňují vhodným shlukům hmoty uchovávat informaci, provádět výpočty a učit se, a že nic takovým shlukům nezakazuje, aby to jednoho dne nedělaly lépe než shluky hmoty v našich mozcích. Je však už mnohem méně jasné, zda a kdy lidé dokážou takovou nadlidskou AGI sestavit. Z první kapitoly vyplynulo, že o tom zatím nic nevíme, neboť se na tom neshodnou ani přední světoví experti. Odhady se většinou pohybují v řádu desetiletí nebo staletí, a podle některých se to dokonce nestane nikdy. Předpovídat cokoli je tu obtížné, protože když se pohybujete na neprobádaném území, nevíte, kolik hor vás dělí od kýženého cíle. Obyčejně vidíte jen tu nejbližší a další překážku objevíte až po tom, co na tu horu vylezete.

Kdy nejdříve by k tomu mohlo dojít? I kdybychom věděli, jak vytvořit AGI na lidské úrovni za použití hardwaru dnešních počítačů (což nevíme), potřebovali bychom k tomu dostatek hrubé výpočetní síly. Jaká je tedy výpočetní síla lidského mozku měřená v bitech a FLOPS z kapitoly 2?* Je to dost ošidná otázka a odpověď závisí na tom, jak se zeptáme:

* Připomeňme, že FLOPS znamená počet operací v pohyblivé řádové čárce za sekundu, řekněme kolik 19ciferných čísel lze vynásobit za jedinou sekundu.

- Otázka 1: Kolik FLOPS je zapotřebí k simulaci mozku?
- Otázka 2: Kolik FLOPS je zapotřebí k lidské inteligenci?
- Otázka 3: Kolik FLOPS dokáže provést lidský mozek?

K otázce 1 bylo publikováno nepřeberné množství článků a většinou se odhaduje v řádu kolem stovky petaFLOPS, tedy 10^{17} FLOPS.[58] To je přibližně stejná výpočetní síla, jakou má čínský Sunway TaihuLight (obrázek 3.7), v roce 2016 nejrychlejší superpočítač světa, který stál asi 300 milionů dolarů. I kdybychom ale věděli, jak simulovat mozek šikovného dělníka, měli bychom z této simulace zisk, jen kdybychom si mohli TaihuLight najmout za méně než hodinovou mzdu dotyčné osoby. Možná bychom museli zaplatit více, neboť mnoho vědců věří, že k přesné reprodukci mozku nestačí matematicky zjednodušený model neuronových sítí z kapitoly 2. Snad ho místo toho potřebujeme simulovat na úrovni molekul nebo subatomárních částic – a to by požadavky na výpočetní sílu značně zvýšilo.

Odpověď na otázku 3 je už jednodušší: lidský mozek je špatný při násobení 19místných čísel, a jedna taková operace zabere několik minut. Tím se dostaneme pod 0,01 FLOPS – o celých 19 řádů níže, než je odpověď na otázku 1! Důvod této obrovské nesrovnalosti spočívá v tom, že mozky a superpočítače jsou optimalizované pro diametrálně odlišné úkoly. Podobný rozpor najdeme u následujících otázek:

Jak dobře může traktor plnit práci vozu formule 1?
Jak dobře může vůz formule 1 plnit práci traktoru?

Kterou z těch prvních dvou otázek o FLOPS se tedy snažíme zodpovědět při předpovídání budoucnosti umělé inteligence? Ani jednu! Kdybychom chtěli simulovat lidský mozek, zajímala by nás otázka 1, ale při vytváření AGI na lidské úrovni nás místo toho zajímá ta prostřední: otázka 2. Odpověď na ni zatím nezná nikdo, ale docela dobře to může být mnohem levnější než simulovat mozek – za předpokladu, že buď adaptujeme software, aby se lépe hodil k dnešním počítačům, nebo že postavíme hardware podobnější mozku (v současnosti probíhá překotný pokrok na poli takzvaných neuromorfních čipů).

Hans Moravec vypracoval odhad založený na srovnávání porovnatelných činností u výpočtu, který dokáže efektivně provádět náš mozek i dnešní počítače: jde o určité úkoly nízkoúrovňového zpracování obrazu, které provádí lidská sítnice, než pošle zrakovým nervem výsledek do mozku.[59] Podle něj vyžaduje zopakování výpočtů sítnice na běžném počítači přibližně miliardu FLOPS a celý mozek provádí asi desetitisíckrát více výpočtů než sítnice (podle srovnání počtu neuronů). Výpočetní kapacita mozku je proto zhruba 10^{13} FLOPS, což přibližně odpovídá síle optimalizovaného počítače za 1 000 dolarů v roce 2015!

Sečteno a podtrženo: neexistuje vůbec žádná záruka, že se nám podaří postavit AGI na lidské úrovni během našeho života – nebo že se to vůbec kdy podaří. Zároveň však nelze vyloučit, že se to podaří. Už neplatí silný argument, že nám schází

Obrázek 3.7: Sunway TaihuLight, v roce 2016 nejrychlejší superpočítač, jehož hrubá výpočetní síla údajně překonává potenciál lidského mozku.

dostatečně silný hardware nebo že by to bylo příliš drahé. Nevíme, jak daleko jsme od cílové pásky z hlediska architektury, algoritmů nebo softwaru, nicméně současný pokrok je rychlý a těmito výzvami se zabývá překotně se rozšiřující globální komunita talentovaných specialistů na AI. Jinými slovy: nemůžeme zavrhnout možnost, že AGI nakonec lidské úrovně dosáhne a že ji překoná. Věnujme proto následující kapitolu zkoumání této možnosti a jejích důsledků.

SHRNUTÍ ZÁKLADNÍCH FAKTŮ:

- Pokrok umělé inteligence má v krátkodobém horizontu potenciál značně zlepšit naše životy nepřeberným množstvím způsobů, od našich osobních životů přes zvyšování efektivity rozvodných sítí a finančních trhů po záchranu mnoha životů prostřednictvím samořídících automobilů, chirurgických robotů či automatických diagnostických systémů s AI.
- Pokud předáme kontrolu nad systémy v reálném světě AI, je klíčové, abychom se ji naučili vytvářet robustnější, tedy zajistit, aby dělala právě to, co po ní chceme. Z toho plyne nutnost vyřešit obtížné technologické problémy verifikace, validace, bezpečnosti a kontroly.
- Potřeba zvýšit robustnost je zvláště naléhavá u zbraňových systémů ovládaných AI, kde může být v sázce nesmírně mnoho.
- Řada předních výzkumníků AI a kybernetiků volá po mezinárodní úmluvě, která by zakazovala některé druhy autonomních zbraní, neboť hrozí nekontrolované závody ve zbrojení, které by poskytly účinné vražedné stroje komukoli s dostatkem prostředků.
- Umělá inteligence může výrazně zvýšit spravedlivost i účinnost našich soudních systémů, pokud se podaří zajistit, aby rozhodování robosoudců bylo transparentní a prosté předsudků.
- Naše zákony potřebují rychlé aktualizace, aby udržely krok s rozvojem AI. To přináší komplikované otázky soukromí, odpovědnosti a regulace.
- Inteligentní stroje nás budou postupně vytlačovat z pracovního trhu již mnohem dříve, než získají potenciál to udělat nadobro.
- To nemusí být špatná věc, pokud bude společnost schopná rozumně přerozdělovat aspoň zlomek bohatství generovaného umělou inteligencí.

- Jinak se bude nerovnost prudce zvyšovat.
- S patřičným plánováním by měla být společnost s nízkou zaměstnaností schopná vzkvétat nejen finančně a lidé by mohli získávat pocit smyslu života i z jiných aktivit než pracovních.
- Kariérní rada pro dnešní děti: vyberte si profese, v nichž jsou počítače špatné – takové, které zahrnují lidi, nepředvídatelnost a kreativitu.
- Existuje nezanedbatelná pravděpodobnost, že pokrok AGI postoupí na lidskou úroveň a že ji dokonce překoná – na to se podíváme v další kapitole.