

Roboti. Už po nás jdou

The Robots are Going After Us!

Josef Krob

<https://doi.org/10.5817/CZ.MUNI.P210-9688-2020-3>

Abstrakt:

Esej, ve které si autor všímá stále častějších obav souvisejících s možným nástupem umělé inteligence a jejími očekávatelnými důsledky. Poukazuje na to, že to v zásadě není zcela nová situace, že v historii již několikrát došlo k podobným narušením komfortní zóny člověka, a vycházeje z těchto historických příkladů se snaží domýšlet možné varianty dalšího vývoje člověka a společnosti.

Abstract:

An essay in which the author notes an increasing concern about the possible onset of artificial intelligence and its expected consequences. He points out that this is not a completely new situation. In fact, several similar violations of the human comfort zone have occurred in human history. Based on such historical examples, the author outlines possible variants of further development of man and society.

Keywords:

robot, man, artificial intelligence, perspective, threat

Patří to vlastně mezi věčná témata. Vynález se obrátí proti svému tvůrci. Golem přestane poslouchat, Frankenstein je hrozbou od samého vzniku a roboti? Ano, už po nás jdou. Situace začíná být pocítována za natolik vážnou, že se obavy z budoucího převalily ze sci-fi literatury do vědeckých článků a knih. Za všechny můžeme jmenovat jednu z nich – *Superintelligence. Až budou stroje chytřejší než lidé* od Nicka Bostroma.⁷⁴ Určitě je možné souhlasit s tím, že k inteligentní revoluci by mohlo dojít už v tomto století a že s ní mohou být spojena i nemalá existenční rizika, která lidstvo jistě podstoupí, protože potenciální zisky se zdají být velmi lákavé.

Nick Bostrom přirovnává současnou situaci k dětské hře s nevybuchlou bombou. Lidé se chovají jako děti a superintelligence je samozřejmě tou bombou, o které nikdo neví, jak může být přivedena k výbuchu, ale slabé tikání časovače je docela zřetelné. Dítě by se zachovalo rozumně, kdyby se bomby ani nedotklo, resp. ji opatrně odložilo, opustilo místo, kde se nachází, a vyhledalo někoho

74 BOSTROM, N.: *Superintelligence. Až budou stroje chytřejší než lidé*, Praha 2017.

dospělého. Bostrom však ví, že takto situace není nastavena. Nemáme co do činění s jedním dítětem, problém je v tom, že přístup k bombě má mnoho dětí a každé z nich má blízko alespoň k jednomu z mnoha spouštěcích mechanismů. Není tedy odpovědné spoléhat se na to, že všechny děti budou rozumné, určitě se najde nějaký malý pitomec, jak je Bostrom přesvědčen, který tu roznětku spustí. A není kam utéct a také není žádný dospělý v dosahu. Tuto téměř bezvýhodnou situaci představuje Bostrom jako výzvu, které je možné čelit tím, že si uchováme naši lidskost, tj. rozvážnost, zdravý rozum, veselost a slušnost.⁷⁵ Bostrom sám si dobře uvědomuje, že tento apel na lidskost je jenom část řešení, které musí mít mnohem komplexnější podobu, pokud chce mít naději na úspěch.

Nezdá se, že by v této věci byl Bostrom příliš konzistentní. Na jedné straně si realisticky uvědomuje, že spoléhat na rozumnost dětí by bylo pošetilé a současně rozumný dospělý není k dispozici a na straně druhé se snaží apelovat v rámci lidskosti na rozvážnost a zdravý rozum. Ale zdravý rozum koho, když dětem ho nepřiznáme a dospělé nemáme? Zkusme tedy dětskou metaforu dotáhnout do konce a pokusit se najít možná východiska a formulovat pravděpodobné scénáře.

Inspirací pro řešení nám může být minulost. Jak jsem předznamenal prvními větami, není to vlastně absolutně nová situace. Podobným existenčním výzvám byl člověk již vystaven, i když byly třeba jen hypotetické, ve fantazii románových autorů. Můžeme však toto ohrožení pojmout poněkud šířeji. Nikoli pouze jako existenční, apokalyptické ohrožení ze strany filmových terminátorů, predátorů a vetřelců, ale jako jakékoli významné narušení komfortní zóny lidstva, která byla často spojována se samou podstatou člověka.

Čím méně toho člověk o sobě věděl, tím více se opájel vlastní výjimečností. A tak na počátku zaznamenaných dějin byl středem vesmíru, byl stvořen bohy, kteří nad ním neustále bděli, vše se doslova i přeneseně točilo kolem něj. Následující dějiny člověka jsou dějinami ztrát jeho výsadního postavení. Samozřejmě se nejednalo o nějaké rychlé zvraty, změny byly postupné, dokonce mnohdy tak pomalé, že si je uvědomovali jen ti nejbystřejší myslitelé.

Sigmund Freud v této souvislosti píše o tom, že „naivní sebeláska lidstva musela v průběhu věků snést dvě velké urážky. Tu první způsobuje Koperník a tvrzení, že naše Země není středem Kosmu, nýbrž nepatrnou částicí světového systému, jehož velikost je nepředstavitelná (tedy nejen Koperník, ale zejména jeho pokračovatelé, Kepler, Bruno — pozn. J. K.), druhou potom Darwin, jehož objevy připravily člověka o údajné privilegované postavení mezi tvorstvem, odhalily jeho původ z živočišné říše a nevyhladitelnost jeho animální přirozenosti. Freud sám představuje třetí ránu lidské ješitnosti vytvořením učení, které chce

75 Viz tamtéž, s. 389.

lidskému Já dokázat, že není pánem ani ve svém vlastním domě a je odkázáno na sporé zprávy o tom, co se nevědomě děje v jeho duševním životě⁷⁶.

Člověk se tak musel smířit s tím, že není středem vesmíru ani pozornosti kohokoli v něm, že není cílem ani účelem vesmírného či pozemského dění a že vlastně ani neví, co se děje v jeho vlastním nitru a co možná určuje jeho život mnohem více, než si zatím dokázal představit.

Mohli by inteligentní roboti způsobit lidem další podobnou urážku? Co by ještě mohli vzít bytosti, která si uvědomila, že je lhotejná celému vesmíru, o co by mohli připravit člověka, který dokonce uvažuje o tom, že svět, ve kterém žije, není fyzicky opravdový, ale je jen simulací naprogramovanou o řád vyšší civilizací? O co by ještě bylo možné takového tvora připravit? Jistě bychom mohli spekulovat o tom, jak by nás mohli inteligentnější roboti využívat, ale to už jsme viděli v mnoha filmových příbězích. O co by nás umělá inteligence mohla připravit, bude jistě záviset na tom, jakou bude mít podobu. Individuální humanoidní roboti? Centrální mozek? Rozptýlená síť?

Podívejme se tedy, jaké atributy si spojujeme s umělou inteligencí nebo inteligentními roboty. Na začátku cesty k umělé inteligenci je velmi pravděpodobně prosté napodobování. Snaha vytvořit umělou bytost, která by plnila lidské příkazy. Tato fantazie provází člověka od dob, kdy zvládnul základní řemesla – hrnčířství či zpracování kovů – a představoval si, že by si mohl vytvořit umělého sluhu, který by byl svému tvůrci bezmezně oddán.

Představy o modelování člověka jsou tak spojeny s řemeslnickým paradigmatickým, inspirují se aktuálně nejrozvinutějšími technologiemi, a tak tvor uplácený z hlíny přežíval ve fantaziích velmi dlouho – od biblických dob až do podoby golemů rudolfínské éry. Novovek a rozvoj kovozpracujících řemesel, zvládajících jemné hodinové mechanismy přivádí na scénu homunkuly v podobě mechanických strojů nevinně hrajících na cembalo. Rozvoj organické chemie dal podnět ke zrodu obávaného Frankensteina, který se stal varovným příkladem pro všechny, kteří vidí i v dnešních snahách o stvoření umělé bytosti spíše nebezpečí. Pošramocenou pověst nedokázala napravit ani mnohem mladší a pohlednější slečna Golem či Pán, který byl vdova. S rozvojem matematiky, formálních systémů a jejich praktické realizace v podobě počítačů se objevuje paradigma „myslicího programu“, tedy už umělé inteligence, jak ji dnes přibližně rozumíme.

Při podrobnějším zkoumání můžeme rozlišit dva až tři možné přístupy. V prvním se budeme držet původního a užšího významu umělé inteligence, která představuje formální systémy postihující modely chování v definovaném prostředí a jejich případnou realizaci a konkrétní aplikace. Širší, filosofický

76 Podle Freudovy přednášky k úvodu do psychoanalýzy cituje a komentuje J. Stromšík v doslovu k FREUD, S.: O člověku a kultuře. Odeon 1989, s. 398–399. Dostupné i on-line: <https://www.scribd.com/doc/37732697/Sigmund-Freud-O-Clověku-a-Kultuře>, citováno [20. 6. 2019].

význam se dotýká inteligence a jejího vztahu k mysli nebo vědomí. Science fiction literatura nebo nověji někteří soudobí filosofové by mohli ztělesňovat třetí přístup, který by byl jistým hybridem, zkřížením jedné větve aplikací, zejména robotiky s filosofickým problémem mysli. Na pořad se tak dostává otázka místa robotů v lidské společnosti a odvození i definice člověka. V těchto souvislostech se jasněji zobrazuje jedna z podstatných otázek spjatých s umělou inteligencí. Je problém umělé inteligence totožný s otázkou, zda umělý systém má vědomí, nebo jde pouze o pragmatické funkce robotů? Je tedy inteligence pouze funkcí, nebo souvisí s vědomím? Mohly by stroje myslet?

Primární cílem při konstrukci strojů je funkčnost, co stroj „cítí“, když funguje, není v tomto případě důležité. Samozřejmě nejde o jednoduché primitivní funkce, ale o poměrně složitou síť rozhodovacích procesů, která by mohla mít za následek, jako vedlejší produkt, vědomí. Inteligence jako funkce tak není dnes technicky nespílitelný problém a v mnoha ohledech by se již dalo říci, že jej již máme realizovaný. Od seriálních rozhodovacích softwarových systémů přes šachové programy a jiné herní strategie až po například pana Sponku z jedné sady kancelářských programů. Vlastně jde jen o míru složitosti, která nám postačí k uznání inteligence.

Pokud jde o vztah inteligence a vědomí, zdá se být zřejmé, že nelze jednoduše hovořit o jejich přímé úměře. Méně inteligentní člověk nemá přece méně vědomí. Naopak není jasno, jaká by mohla být hranice mezi vědomím a nevědomím nebo jakým způsobem se vědomí poprvé probudí – je to skokový proces, nebo pomalé přirůstání? Není tak možné jednoznačně určit ani hranici mezi strojovým systémem bez vědomí a systémem, který již vědomí má či se v něm probouzí postupným připojováním dalších prvků.

Můžeme se tedy vrátit k otázce, o co by nás umělá inteligence nebo inteligentní roboti mohli připravit. Protože ale víme, že UI může nabývat různých podob od formálních systémů až po humanoidní roboty, a plnit tak nejrůznější funkce, ponechejme stranou poměrně reálné hrozby týkající se zejména změn na pracovním trhu. Lidé jistě přijdou o některé pracovní pozice či typy práce. Odhaduje se, že v roce 2026 bude polovina veškerého objemu práce na světě vykonána roboty. Pro naši úvahu není důležité, jak moc přesné tyto odhady jsou, v jakém pořadí změny půjdou (soustružníci, pokladní, písaři, pracovníci na přepážkách úřadů) ani jak budou tyto pozice nahrazovány novými pracovními místy podle nových požadavků a potřeb. To jsou otázky pro ekonomy a sociology. Otázka zní, zda můžeme přijít o něco, co je nenahraditelné, a čeho si tedy vážíme (nebo bychom si měli vážít) mnohem více.

Přenecháme-li tedy starosti s pracovním trhem a strukturováním společnosti specialistům, můžeme se vrátit k urážkám, které lidstvo již utrhlo, a pokusit se o extrapolaci. Na čem nám ještě záleží? O co se nejvíce staráme? Co si nejvíce vzájemně přejeme? Přece zdraví. Tělesné i duševní. Cvičíme a vzděláváme se a vedle osobnostního rozvoje máme na mysli i prosperitu společenství. Pečujeme

tak o tělo, duši a společenství, na kterém si ceníme zejména svobody a zajištění soukromí. Otázka tedy zní: může nás umělá inteligence připravit o naše tělo, rozum, svobodu a soukromí? A odpověď se zdá být poměrně jasná. Ano, a již se tak děje. Respektive v neustálém kontaktu s chytrými stroji se těchto cenností postupně zbavujeme sami.

Tělo. Na vylepšování těla není evidentně nic nebezpečného. Zejména, když je to vlastně jen kompenzace nějakého nedostatku. Protetika či posílení lidských smyslů je vcelku běžná věc. Bude stejně tak běžnou věcí vylepšování zdravého normálního těla? Třetí ruka pro bubeníky, genetické vylepšení lidských schopností, prostě postupná kyborgizace těla člověka?

Duše. Rozum, mysl. Ani zde nevnímáme vylepšování jako hrozbu, a ještě nějakou dobu vnímat nebudeme. Naopak, usilovně pracujeme na celoživotním zdokonalování. Jenom se bude měnit pohled na možné a přijatelné podpůrné prostředky. Chemické, elektronické, genetické a mezitím bude postupovat kyborgizace rozumu člověka.

Soukromí a svoboda. Soukromí se rovněž zbavujeme dobrovolně a mnohdy za to ještě i platíme. Pořizujeme si nejrůznější elektronické hračky, jejichž funkčnost je vykoupena ztrátou velké části soukromí. A není to jen trend posledních desetiletí, je to v podstatě princip všech technologií, počínaje technologií ohně. V okamžiku, kdy člověk ovládl oheň a zvolil k večeři pečenou kýtu místo syrové, zvolil i světlo, ve kterém ho všichni ostatní viděli. Pečená kýta sněžená před svědky byla lepší než kus syrového masa potmě, i když beze svědků. A nebyla cesta zpět. A nikdo se dlouho ani zpět neohlížel, vřdyť nové vynálezy byly tak úžasné, že ztráta té trochy soukromí byla zanedbatelná. Dohled a kontrola v relativní prosperitě dostaly přednost před svobodným umíráním.

Vypadá to tedy, že nemusíme čekat na roboty a superinteligenci, abychom se cítili ohroženi. Pokud vůbec přistoupíme na to, že uvedené jevy nás nějakým způsobem ohrožují, je zřejmé, že zatím jsme největší hrozbou sami sobě. Je to všechno naše volba. Vidím zde pouze jeden problém, který bych nenazval nebezpečím, i když jistá ztráta to může být, narušení komfortní zóny určitě. S tím, jak bude člověk vylepšovat sám sebe a současně bude zdokonalovat i umělé stroje, může se stát, že v určitém okamžiku přestaneme být schopni rozeznat, zda za konkrétní aktivitou stojí člověk, nebo robot. Zůstává otázkou, nakolik by nám mělo vadit, když nebudeme vědět, zda nám jídlo v restauraci uvařil a servíroval stroj, humanoidní robot nebo člověk, zda se za přepážkou v bance nedomluvíme s úřednicí nebo bankovním systémem, ale je velmi pravděpodobné, že nová situace může být zdrojem xenofobních reakcí.

Fantastická literatura zpracovává tento námět v nesčetných příbězích ve snaze najít nějaký podstatný rozdíl mezi člověkem a robotem a konstruuje zapeklité situace, ve kterých se roboti, kteří musí dodržovat robotické zákony, jak je zformuloval I. Asimov, ukáží jako neschopní řešit „bezvýchodné“ situace na rozdíl od člověka, který má intuici, cit apod.

Karel Čapek přišel s prvním návrhem na rozlišení robota a lidské bytosti. Člověk je podle něj tvor, který je obdařen láskou. City mají být tím rozlišovacím znamením. Pokud však budeme na emoce usuzovat podle verbálních projevů, řeči těla, jednání, slz a smíchu, zjistíme, že toto kritérium bude selhávat s tím, jak se roboti naučí tyto projevy emulovat. Stanislav Lem nabízí jako rozlišovací znak chybovost a váhání. Umělá inteligence je přece vytvořena proto, aby byla lepší a rychlejší v úsudku, snad dokonce bezchybná, a chybička a váhání tedy budou známkou lidskosti. Ovšem i váhání lze předstírat a chyby lze dělat úmyslně. Snad že by nám v rozlišení člověka a robota pomohl determinismus? Umělá inteligence a roboti jsou stvořeni proto, aby efektivně zvládali nějakou činnost. Jsou tak naprogramováni. Neměli by neefektivně plýtvat energií. Člověk dělá nepředloženosti, youtube je plný videí zaznamenávajících lidské aktivity, nad kterými zůstává rozum stát. Ale dostatečně složitý program by zřejmě zvládl i generování náhodných činností, a i robot by tak mohl produkovat nepředloženosti. Co zatím umělé systémy nezvládají dobře je ironie a humor. Pokud se začnete do automaticky generovaných vtipů, moc se ještě nezasmějete. Ale možná i toto jednou stroje zvládnou.

Ale i když žádné z uvedených kritérií není absolutně spolehlivé, přece jen je to určitá indicie a zde spatřuji pro člověka největší nebezpečí ve spojení s roboty a umělou inteligencí. Ve snaze odlišit se od svého výtvoru, vsadí člověk na emoce, chyby, nedeterminovanost, ironii a vtip a výsledkem může být plačtivý, chybující tvor produkující nepředloženosti, o kterých si myslí, že jsou vtipné. Snaha zachránit podstatu lidství skončí v jeho karikatuře.

Literatura

- BOSTROM, N.: Superintelligence. Až budou stroje chytřejší než lidé. Praha: Prostor, 2017.
- STROMŠÍK, J.: Doslov. In: FREUD, S.: O člověku a kultuře. Praha: Odeon, 1989. Dostupné i on-line: <https://www.scribd.com/doc/37732697/Sigmund-Freud-O-Člověku-a-Kultuře>