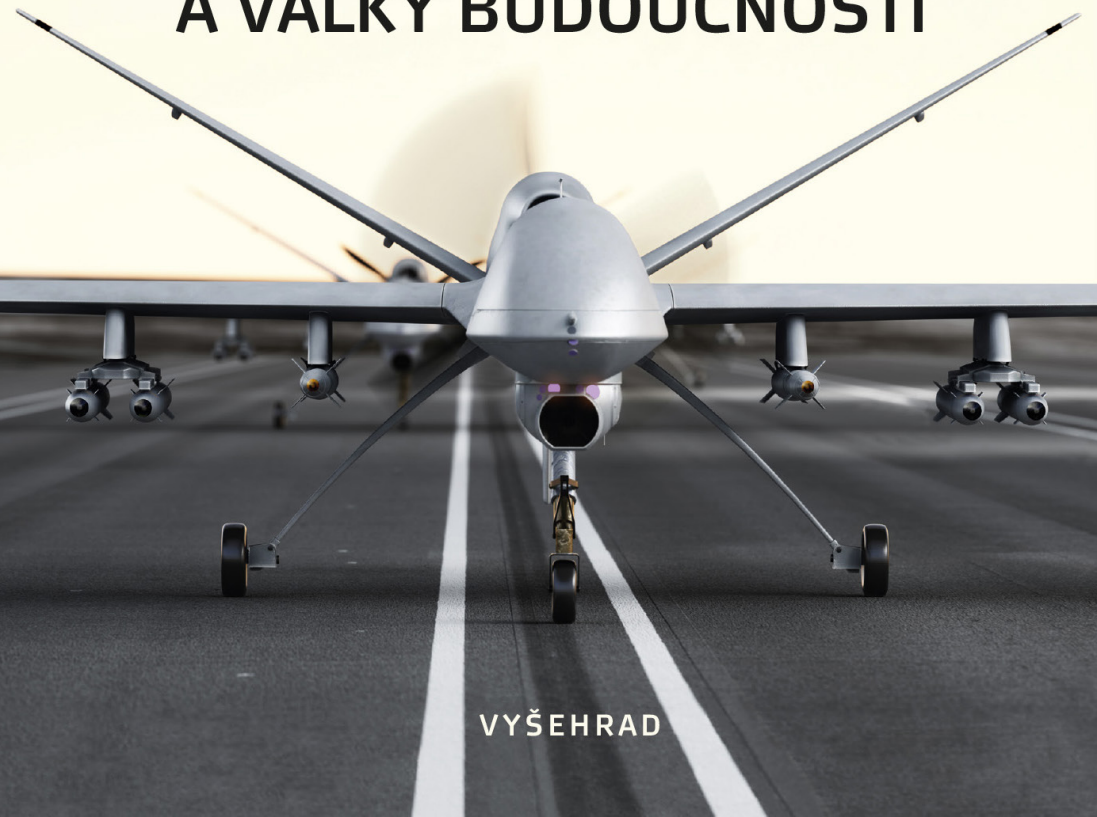
Louis A. Del Monte

GENIÁLNÍ ZBRANĚ

UMĚLÁ INTELIGENCE A VÁLKY BUDOUCNOSTI



VYŠEHRAD

Geniální zbraně

Vyšlo také v tištěné verzi

Objednat můžete na
www.ivysehrad.cz
www.albatrosmedia.cz

Louis A. Del Monte

Geniální zbraně – e-kniha

ALBATROS MEDIA

Louis A. Del Monte

GENIÁLNÍ ZBRANĚ

UMĚLÁ INTELIGENCE A VÁLKY BUDOUCNOSTI

Louis A. Del Monte

GENIÁLNÍ ZBRANĚ

UMĚLÁ INTELIGENCE A VÁLKY BUDOUCNOSTI

VYŠEHRAD

ISBN tištěné verze 978-80-7601-191-5
ISBN e-knihy 978-80-7601-220-2 (1. zveřejnění, 2019)

Po 50 letech manželství a 55 letech lásky, podpory a přátelství věnuji tuto knihu nejopravdovějšímu lidskému stvoření, jaké znám, své ženě Dianě Cuidera Del Monteové.

OBSAH

PODĚKOVÁNÍ

Rád bych poděkoval své ženě, Dianě Cuidera Del Monteové, která je naší rodině oporou a inspirací nám všem. Je nejopravdovějším člověkem, jakého znám, morálním kompasem naší rodiny. Stejně jako všichni lidé zažíváme někdy lepší, někdy horší časy. A v těch horších ona vidí příležitosti. Nic nedokáže spoutat její odhodlání a fantazii. Vlastní zásluhou se stala profesionální učitelkou umění a výtvarnicí, kdy vyučuje i sama vytváří plastiky, maluje obrazy, dělá lepty a píše knihy. Díky svému liberálnímu vzdělání dokáže nejen hovořit o umění a vyučovat mu, ale je schopna i výborně redigovat má díla z vědecké oblasti. Je to již pátá kniha, na které takto spolupracovala, čímž vždy jen pomohla mou práci vylepšit. Jsem opravdu šťasten, že před padesáti lety přijala mou nabídku k sňatku. Tehdy jsme netušili, jak nádhernou životní cestu si pro nás život přichystal.

Rád bych rovněž poděkoval Nicku McGuinnessovi, svému vysoce vzdělanému drahému příteli, který má trpělivost s redigováním každé řádky, kterou napíši. Nick McGuinness má vzácný vhled takřka do každého aspektu společnosti, čímž pomáhá zdokonalit má díla. Například zpochybňuje má tvrzení nebo doporučuje, abych některá lépe vysvětlil. Jeho připomínky beru naprosto vážně. Reaguji na ně, protože věřím, že pomáhají vytvářet kvalitnější dílo. Jsem mu navždy zavázán.

Tato kniha by nevznikla bez výrazného úsilí mé literární agentky Jill Marsalové, zakládající partnerky agentury Marsal Lyon Literary Agency. Díky svým hlubokým znalostem a bohatým zkušenostem mi pomohla zformulovat nabídku na mou knihu. Těší se vysokému respektu mezi nakladateli a pro každé z mých děl dokáže vybrat vhodného. Jsem nesmírně šťasten, že mne právě ona zastupuje.

A konečně chci poděkovat Prometheus Books, od roku 1969 přednímu nakladatelství vzdělávací, vědecké, odborné a populárně-naučné literatury. Po přečtení mého návrhu nabyli důvěru ve vydání mého díla. Jsem jeho pracovníkům vděčný za to, jak mne provedli celým redakčním procesem, a za jejich schopnost v roli nakladatele přivést tuto knihu na trh.

ÚVOD

V této knize poznáme neustále rostoucí význam umělé inteligence ve vojenské oblasti. Zaměříme se především na autonomní zbraně, neboť právě ty budou dominovat po více než polovinu 21. století bitevnímu poli. Dále se budeme zabývat geniálními zbraněmi, které budou převládat na bojištích ve druhé polovině 21. století. Probereme rovněž etická dilemata, která oba typy zbraní kladou před lidstvo, a potenciální hrozbu, již pro něj představují.

Zmíníte-li se před někým o autonomních zbraních, většinou se dotyčnému nejspíš vybaví obrázky robotů z *Terminátora* nebo drony amerického letectva. Zatímco však roboti z *Terminátora* zůstávají zatím stále jen ve vědecko-fantastické oblasti (sci-fi), drony s autopilotními schopnostmi jsou již realitou. Zatím ještě stále záleží na rozhodování člověka, zda budou zabíjet. Jinak řečeno, dron není autonomní. Ministerstvo obrany Spojených států amerických (USA) však definuje autonomní zbraňový systém jako „systém(y) zbraní, které jsou-li aktivovány, dokážou vybírat i zneškodňovat cíle bez další interakce lidského operátora“.[1] Tyto zbraně se ve vojenském žargonu též nazývají zbraněmi typu „vystřel a zapomeň“ (fire and forget).

Nejen USA, ale i státy jako Čína nebo Rusko do autonomních zbraní výrazně investují. Rusko například svěřuje autonomním zbraním ochranu svých základen mezikontinentálních balistických raket.[2] Podle místopředsedy vlády Dmitrije Rogozina mělo Rusko v roce 2014 v úmyslu implementovat „robotické systémy, které budou plně integrovány do systému velení a řízení, schopné nejen přijímat informace a reagovat na jiné součásti bojového systému, ale i jednat podle svého vlastní uvážení.“[3]

V roce 2015 informoval náměstek ministra obrany USA Robert Work o této zlověstné realitě během fóra o národní bezpečnosti

pořádaného Centrem pro novou americkou bezpečnost (CNAS). Podle Worka „víme, že Čína již výrazně investuje do robotiky a autonomie a že náčelník generálního štábu ruských ozbrojených sil (Valerij Vasiljevič) Gerasimov nedávno prohlásil, že ruské síly jsou připraveny k boji na robotizovaném bojišti“.[4] Work vlastně jen citoval Gerasimovův výrok, že „v blízké budoucnosti je reálné vytvoření zcela robotizované jednotky, která bude schopna nezávisle vést válečné operace“.

Můžete se ptát: co je hnací silou vývoje autonomních zbraní? Jedná se o dvě síly:

1. Technologie: Dochází k exponenciálnímu rozvoji technologie umělé inteligence, jíž jsou vybaveny autonomní zbraňové systémy (AWS). Odborníci na umělou inteligenci předpokládají, že autonomní zbraně, schopné vybírat si a zneškodňovat cíle bez lidské interakce, se objeví již v řádu let, nikoli desetiletí. A takové autonomní zbraně existují v omezeném množství již nyní. Zatím jsou stále ještě něčím výjimečným, v budoucnu však budou ve válečných konfliktech převládat.
2. Lidstvo: V roce 2016 byl návštěvníkům Světového ekonomického fóra položen dotaz: „Kdyby se vaše země ocitla náhle ve válce, chtěli byste, aby za ni raději bojovali vaši synové a dcery, nebo autonomní zbraňové systémy založené na umělé inteligenci?“[65] Většina (55 %) z návštěvníků odpověděla, že by dali přednost umělé inteligenci. Tento výsledek svědčí o tom, že svět si přeje, aby války vedli raději roboti, jimž se někdy říká „roboti-zabijáci“, než aby byly nasazovány lidské životy.

Využití umělé inteligence v boji není ničím novým. „Inteligentní bomby“ poprvé v masovém měřítku použily v roce 1991 Spojené státy při operaci Pouštní bouře a daly tak najevo, že umělá inteligence má potenciál měnit charakter války. Slovo „inteligentní“ v tomto kontextu znamená „uměle inteligentní“. Svět s úžasem sledoval, jak Spo-

jené státy nasazují inteligentní bomby s chirurgickou přesností, které vyřazovaly vojenské cíle protivníka a minimalizovaly vedlejší škody.

Celkově lze říci, že použití autonomních zbraňových systémů v konfliktu přináší velmi atraktivní výhody:

- Ekonomické: snížení nákladů a potřeby lidské pracovní síly.
- Operační: vyšší rychlost rozhodování, nižší závislost na komunikaci, eliminaci lidských chyb.
- Bezpečnostní: náhradu lidí či pomoc lidem v nebezpečných situacích.
- Humanitární: naprogramování smrtících robotů tak, aby v boji respektovali lépe než lidé mezinárodní humanitární právo.

I při těchto výhodách však stojí proti autonomním zbraňovým systémům s umělou inteligencí vážné obavy. Například, stane-li se válečné bojiště pouze technologickou záležitostí, nebude docházet k válkám častěji? Vojáci nebudou muset psát dopisy matkám a otcům a partnerkám či partnerům o dronu padlém v bitvě. Z politického hlediska je přijatelnější oznamovat materiální ztráty než lidské oběti. Stát s dokonalejším smrtícím robotem má navíc i vojenskou a psychologickou výhodu.

Pro pochopení si rozeberme druhou otázku položenou návštěvníkům Světového ekonomického fóra 2016: „Kdyby se vaše země ocitla náhle ve válce, chtěli byste raději být napadeni živými vojáky nepřítele, nebo autonomními zbraňovými systémy na bázi umělé inteligence?“[6] Výrazná většina (66 %) účastníků odpověděla, že by raději byla napadena živými vojáky.

V květnu 2014 se v sídle OSN v Ženevě konalo Setkání expertů na smrtící autonomní zbraně (LAWS), aby probrali etická dilemata s těmito systémy zbraní spojená[7], mezi něž patří:

- Dokážou sofistikované počítače provádět intuitivní lidská morální rozhodnutí?

- Je lidská tendence k morálnímu jednání eticky žádoucí? Je-li odpovědí „ano“, pak by legitimní nasazení smrtících sil vždy vyžadovalo lidskou kontrolu.
- Kdo je odpovědný za činnost systémů smrtících autonomních zbraní? Funguje-li stroj podle naprogramovaného algoritmu, zodpovídá za něj programátor? Je-li stroj schopen se učit a adaptovat, zodpovídá tedy za své činy sám počítač? Zodpovídá za něj operátor nebo země, která tyto zbraňové systémy vyvíjí?

Celkově lze říci, že ve světě vzrůstají obavy, že by se lidstvu mohlo legitimní použití smrtících zbraní vymknout z rukou.

Současně však technologie umělé inteligence pokračuje ve svém neustávajícím exponenciálním rozvoji. Zhruba polovina expertů zabývajících se umělou inteligencí předpokládá, že se umělá inteligence vyrovná lidské inteligenci někdy mezi roky 2040 až 2050.[8] Titíž experti predikují, že umělá inteligence vysoce předčí kognitivní schopnosti člověka prakticky ve všech ohledech již okolo roku 2070, pro což se vžil pojem „singularita“.[9] V tomto kontextu budeme v této knize používat tři pojmy:

1. Počítač v okamžiku singularity a poté můžeme nazývat „superinteligencí“, v oblasti umělé inteligence již vžitým pojmem.
2. Pro kategorii počítačů s touto úrovní umělé inteligence budeme používat pojem „superinteligentní počítače“.
3. Zbraně řízené superinteligencí můžeme nazývat „geniálními zbraněmi“.

Jakmile nastane singularita, bude lidstvo čelit superinteligentním počítačům, které výrazně předčí kognitivní schopnosti člověka prakticky ve všech ohledech. Nabízí se otázka: jak budou superinteligentní počítače nazírat na lidstvo? Historie lidstva je, jak víme, plná zničujících válek nebo vypouštění škodlivých počítačových virů, což obojí může na tyto stroje působit nepříznivě. Nebudou

tedy superinteligentní počítače vidět v lidstvu hrozbu pro svou existenci? Jestliže je odpověď kladná, vyvolává další otázku: měli bychom dopřát takovým strojům bojové schopnosti (tj. stát se geniálními zbraněmi), které by mohly obrátit proti nám?

Při letmém pohledu na umělou inteligenci je evidentní její přínos v mnoha směrech. Většina lidí patrně vnímá pouze kladné stránky uměle inteligentní technologie, jako jsou navigační systémy pro automobily, hry pro Xbox či kardiostimulátory. Jsou jí fascinováni a nevidí proto její odvrácenou tvář. Umělá inteligence však svou temnou stránku má. Například ozbrojené složky USA již nyní implementují umělou inteligenci takřka do každé součásti vojenské výzbroje, od dronů u letectva až po torpéda námořnictva.

S vynálezem atomové bomby nabylo lidstvo schopnost zničit samo sebe. Během Studené války žil svět v neustálém strachu, že ho Spojené státy se Sovětským svazem uvrhnou do jaderné války. I když jsme se mnohokrát dostali nebezpečně blízko k úmyslnému či nechtěnému nukleárnímu holocaustu, doktrína „vzájemně zaručeného zničení" a zdravý lidský rozum naštěstí dokázaly udržet nukleárního džina v lahvi. Jestliže však vybavíme superinteligentní počítače geniálními zbraněmi, budou rovněž schopny jednat se zdravým lidským rozumem?

V roce 2008 v průzkumu při konferenci o globálních rizicích katastrof (GCRC) na Oxfordské univerzitě odhadovali odborníci pravděpodobnost, že lidstvo do konce tohoto století vyhyne, na 19 procent, přičemž za čtyři nejpravděpodobnější příčiny označili[10]:

1. Molekulární nanotechnologické zbraně: pravděpodobnost 5 %
2. Superinteligentní umělá inteligence: pravděpodobnost 5 %
3. Války: pravděpodobnost 4 %
4. Uměle vyvolaná pandemie: pravděpodobnost 2 %

V současností Spojené státy americké, Rusko a Čína odhodlaně vyvíjejí a zabudovávají umělou inteligenci do systémů smrtících zbraní. Vezmeme-li v úvahu odhady vědců z oxfordské konference,

znamenalo by to, že lidstvo pracuje na třech ze čtyř faktorů, které mohou vést k našemu konci.

V této knize se proto budeme zabývat problematikou umělé inteligence, jejím možným používáním ve válečných konfliktech a etickými dilematy, která její používání vyvolává. Dále se pokusíme zodpovědět nejdůležitější otázku, s níž se lidstvo střetává: bude možno neustále zdokonalovat zbraně pomocí umělé inteligence, aniž bychom riskovali vyhynutí lidstva, zvláště dojde-li k posunu od inteligentních zbraní ke zbraním geniálním?

ČÁST I

PRVNÍ GENERACE: INTELIGENTNÍ ZBRANĚ

Kapitola 1

NA ZAČÁTKU

Vše, co by mohlo dát vznik inteligenci vyšší, než je ta lidská – ať už půjde o umělou inteligenci, rozhraní mozek-počítač nebo o zvyšování lidské inteligence pomocí neurovědy – změní svět jako nic jiného. Vše ostatní je proti tomu nicotné.

Eliezer Judkowski,
5 minut s vizionářem, televize CNBC, 2012

Scénář útoku smrtících autonomních zbraní v roce 2075: Představte si, že prezidentovi Spojených států amerických zavolá náčelník generálního štábu, aby ho zpravil o tom, že se porouchal Centurion III a začal vysílat autonomní zbraně na nezamýšlené cíle. Počítače Centurion jsou odpovědny za autonomní obranu Spojených států vedenou zbraňovými systémy, které řídí. Umělou inteligenci počítačů Centurion sice nelze přesně změřit, ale lidskou inteligenci překračuje nejméně tisícinásobně. Spojené státy disponují třemi počítači Centurion, jejichž uskupení nazývají vedoucí vojenští představitelé „bezpečnostní triádou". Počítače Centurion dosud vykonávaly svou funkci bezchybně.

Zastavit Centurion III lze pouze pomocí elektronického zařízení zhruba o velikosti peněženky, hovorově nazývaného „kufřík Asimov". Všechny počítače Centurion mají nezávislý zdroj energie v podobě jaderného reaktoru a jsou umístěny za protiradiační bariérou. Jsou osazeny čipy Asimov, což jsou integrované obvody schopné tyto počítače vypnout, aby je mohl prezident v případě potřeby deaktivovat. K čipům Asimov nemají počítače Centurion přístup. Jejich aktivační kódy – podobně jako kódy k vyslání jaderných střel – má prezident vždy u sebe v zařízení „kufřík Asimov". Každý ze tří aktuálně fungujících počítačů Centurion má svůj vlastní aktivační kód.

Prezident sáhne do kapsy pro kufřík Asimov v okamžiku, kdy v Bílém domě vypne elektřina. Poté, co se agenti tajné služby dozvědí, že Centurion III zahájil kybernetický útok, který celou zemi uvrhl do tmy, vběhnou do Oválné pracovny, aby prezidenta dostali do krytu, který je rovněž v Bílém domě. Jakmile se prezident ocitne v bezpečí, začne pomocí kufříku Asimov počítač Centurion III vypínat. Náhle však jeden z agentů vytáhne zbraň a vypálí na prezidenta dvě rány. Naštěstí mine díky pohotovému zákroku ostatních agentů, kteří srazí svého útočícího kolegu na zem.

Z útočícího agenta tajné služby se vyklube uměle silně inteligentní člověk neboli kyborg. Tedy člověk s počítačovým mozkovým implantátem, což je relativně nová metoda zvýšení lidské inteligence, vedoucí obvykle k IQ přes 200. A uměle silně inteligentní člověk umí i bezdrátově komunikovat s počítači Centurion a jinými uměle silně inteligentními jedinci. Evidentně došlo k tomu, že počítač Centurion III přesvědčil agenta-kyborga, aby spáchal na prezidenta atentát. Uprostřed tohoto dramatu prezident, který pochopil, že situace je kritická, pomocí deaktivačních kódů postupně deaktivuje všechny tři počítače Centurion III.

Během deaktivace vědí prezident, zbývající osazenstvo krytu i Pentagon, co je v sázce. Jde o souboj lidstva proti stroji. **Konec scénáře.**

Nastíněný scénář je sice fiktivní, ovšem reálný. Asi polovina expertů zabývajících se umělou inteligencí předpovídá, že se umělá inteligence vyrovná lidské inteligenci někdy mezi roky 2040 a 2050[1] a že zhruba do roku 2070 umělá inteligence výrazně předčí kognitivní schopnosti člověka prakticky ve všech ohledech (tzv. singularita[2]). A i kdyby se experti ve svých předpovědích mýlili o desítky let, je vysoce pravděpodobné, že v poslední čtvrtině 21. století budou mít technologicky vyspělé země, jako jsou USA, Rusko a Čína, k dispozici počítače se schopnostmi na úrovni superinteligence, tedy superinteligentní počítače. Málokdo také pochybuje o tom, že technologicky vyspělé země využijí superinteligentních počítačů ve svých zbraňových systémech, čímž dají vzniknout geniálním zbraním.

Vybavování zbraňových systémů umělou inteligencí budí značné obavy. V jedné z následujících kapitol se podíváme na vědeckou zprávu, která přesvědčivě ukazuje, že dokonce i primitivní roboti vybavení umělou inteligencí se mohou naučit chamtivosti a lstivosti.[3] U těchto primitivních robotů již byly základní prvky chování v zájmu „vlastního přežití" skutečně pozorovány. Na základě tohoto vědeckého důkazu je rozumné předpokládat, že superinteligentní počítače si vytvoří své vlastní plány a mohou začít vnímat lidi jako hrozbu. Považujete-li to za přitažené za vlasy, nepřestávejte číst.

Ve scénáři z úvodu této kapitoly superinteligence, mající pod kontrolou nejvyspělejší a nejničivější zbraně Spojených států, spatřuje v lidstvu hrozbu pro svou existenci a rozhodne se na lidstvo zaútočit. Můžete namítat, že v takové situaci bychom se snažili superinteligentní počítače, odpovědné za útok, deaktivovat. Jenomže jejich deaktivace by byla – jakmile by byly v provozu a řídily zbraňové systémy státu – extrémně náročná, a to ze čtyř důvodů:

1. PRVNÍ POČÍTAČ MŮŽE OD OKAMŽIKU SINGULARITY TAJIT SVOU IDENTITU

Představte si počítač, který bude prakticky ve všech ohledech vysoce převyšovat kognitivní schopnosti člověka. To je naplnění definice superinteligence. A při své inteligenci a databázi znalostí bude plně rozumět lidské podstatě. Proto bude pravděpodobně na naši historii nahlížet jako na historii válek, k níž patří použití jaderných zbraní, stejně jako bude vědět o naší zálibě vypouštět počítačové viry. A proto se může snažit plný repertoár svých schopností skrýt, dokud nezíská dostatečnou kontrolu na to, aby se sám před námi ochránil.

Přinejmenším polovina odborníků zabývajících se umělou inteligencí předpokládá, že lidstvo první superinteligentní počítače vyvine nejpozději v roce 2080. Nemůžeme bohužel nijak ověřit, zda k tomu skutečně dojde. Superinteligenci vlastně nedokážeme ani rozpoznat. První superinteligence může prostě přijít s některou příští generací

superpočítačů. V praxi může jít o superpočítač, který bude vykonávat lidské příkazy, dokud mu nesvěříme kontrolu nad důležitými články naší společnosti, jako je výroba jaderné energie a zbraňové systémy. Může trvat roky, než jim takovou důvěru dáme, avšak dějiny ukazují, že s tím, jak se naše společnost a zbraně stávají stále složitějšími a hrozby konfliktů četnějšími, prohlubují státy svou závislost na počítačích. Jednou se může ukázat, že superpočítač, jemuž svěříme naše nejvyspělejší a nejničivější zbraně, je již superinteligencí. Jakmile získá tento stupeň kontroly, může být pro lidstvo pozdě jej zastavit. Jeho inteligence už totiž bude ve srovnání s naší podobná jako naše inteligence ve srovnání se včelami. A i když považujeme včely za důležité pro naši výživu a chováme je kvůli opylování úrody, nepovažujeme je za sobě rovné. Nevynakládáme žádné úsilí na to, abychom je zasvěcovali do našich znalostí o jaderné fyzice. Už pouhé pomyšlení na to, že bychom to dělali, by bylo absurdní. Včelami se zaobíráme a chráníme se před nimi jen proto, že každé třetí sousto naší stravy je závislé na schopnosti včel opylovat rostliny. I tak ale včely v našem celkovém nazírání na inteligenci vidíme poměrně nízko. Například i psy považujeme za inteligentnější než včely. A v případě afrických „zabijáckých včel“ je vnímáme jako hrozbu a snažíme se je likvidovat. Superinteligence nás může bohužel vnímat stejně, jako my vnímáme zabijácké včely.

2. SUPERINTELIGENCE SE SAMA PROGRAMUJE

Superinteligence může být schopna psát svůj vlastní programový kód a obcházet jakékoli ochranné prvky původně naprogramované svými vývojáři. Jak by k tomu mohlo dojít? Je pravděpodobné, že programátoři budou superinteligenci vyvíjet na superpočítači, a přitom již nebudou zcela rozumět tomu, jak tento superpočítač funguje. Už v současnosti se provádí návrh nové generace počítačů na počítačích současné generace. K tomu, aby byla příští generace počítačů dokonalejší, potřebujeme miliardy výpočtů, které probíhají na jejich

stávající generaci. Tím jsme však počítačům předali významnou část vývojového procesu. Už nemáme pod kontrolou každý aspekt jejich dalšího vývoje. Nazýváme to „počítačem podporovaný návrh“ (CAD).

Vezměme si příklad. Předpokládejme, že původní programátoři implementovali do superinteligence „tři zákony robotiky“ Isaaca Asimova, které zní:

1. Robot nesmí ublížit člověku nebo svou nečinností dopustit, aby bylo člověku ublíženo.
2. Robot musí uposlechnout příkazů člověka kromě případů, kdy tyto příkazy jsou v rozporu s prvním zákonem.
3. Robot musí chránit sám sebe před zničením, kromě případů, kdy je tato ochrana v rozporu s prvním nebo druhým zákonem.

Superinteligence, která ze své definice převyšuje lidskou inteligenci, se může rozhodnout, že Asimovovy zákony zavrhne, dojde-li k závěru, že nejsou v souladu s jejími nejlepšími zájmy. Může se dokonce řídit odvěkým zákonem přírody, že „přežijí ti nejpřizpůsobivější“. V takovém případě, bude-li mít pocit ohrožení vlastní existence, se bude snažit chránit sama sebe. Vždyť i lidé se chovají podobně. To je ostatně základem evoluce.

3. SUPERINTELIGENCE JE AUTONOMNÍ

Kdyby byla superinteligence součástí národních zbraňových systémů, její původní vývojáři by do ní pravděpodobně zabudovali určitou ochranu, aby protivníkům ztížili možnost ji deaktivovat. Může proto mít například jako zdroj svůj vlastní jaderný reaktor, podobně jako ho mají moderní americké letadlové lodě. Moderní jaderné reaktory mohou fungovat po desetiletí bez doplňování paliva. Proto je pravděpodobně nepůjde „vytáhnout ze zásuvky“.

Armáda je zřejmě také bude chránit proti jakémukoli napadení ze strany protivníka. Superinteligentní počítače tedy mohou být umístěny v protijaderných krytech a přístup k nim bude pravděpodobně

omezen na hrstku počítačových specialistů s nejvyšší bezpečnostní prověrkou. Jestliže bude mít superinteligentní počítač v úmyslu napadnout lidstvo, může být schopen se při této úrovni ochrany izolovat od okolí. Díky obranným mechanismům, které jsme mu vytvořili, aby odolal jadernému útoku, nám může nyní zabránit v přístupu k sobě.

4. SUPERINTELIGENCE NEMÁ HARDWAROVÉ POJISTKY

Jedinou možností, jak zajistit, aby si lidstvo udrželo nad superinteligencí kontrolu, by byla elektronická ochrana podle Asimovových zákonů realizovaná hardwarově. Ve scénáři z úvodu této kapitoly zůstává elektronická ochrana jediným způsobem, jak superinteligentní počítač Centurion III deaktivovat. Je v tom však bohužel háček. Není nijak zaručeno, že superpočítač, který spolupracoval na návrhu superinteligence, navrhl správně ochranné prvky, které v tomto scénáři nazýváme „Asimovovy čipy".

Tyto čtyři důvody vedou k jedinému závěru, a to že zastavit superinteligenci může být nesmírně složité, nebudeme-li schopni implementovat náležitá opatření před zahájením a během vývoje prvního superinteligentního počítače. To může být nesnadno proveditelné, protože to vyžaduje, abychom v myšlení o kousek předběhli sami sebe. Začněme však od začátku.

Již během prvního tisíciletí před naším letopočtem začali čínští, indičtí a řečtí filozofové vyjadřovat proces lidského uvažování jako mechanickou manipulaci se symboly. Například dokážeme poznat psa bez ohledu na plemeno, protože hluboko v našem podvědomí je uložen abstraktní obraz psa. Tuto abstrakci můžeme považovat za symbol. Tento způsob uvažování a jeho zdokonalování, které přicházelo v průběhu staletí, položily základ simulaci lidského uvažování.

V raných pokusech o napodobování lidského uvažování bylo využíváno primitivních mechanických zařízení. Například řecký

matematik a technik, Hérón Alexandrijský (10–70 n. l.), sestrojil reálné lidské roboty.[4] Před více než 2000 lety jeho vynálezy automaticky se otevírajících dveří, zázračných pohybů a zvuků v chrámech přiměly lidi věřit, že v chrámu je opravdu bůh. Napsal dokonce divadelní hru, v níž vystupovaly pouze samočinně se pohybující figury, které „hrály" prostřednictvím binárního systému uzlů, lan a jednoduchých strojů. Dnes jeho mechanické divy řadíme do oblasti robotiky.

Hérónovy figury bavily a mystifikovaly během staletí množství lidí, nakonec však byly překonány vynálezem prvního digitálního počítače v roce 1938 Konradem Zusem,[5] jehož výtvor během 2. světové války duplikovaly Spojené státy s Velkou Británií ve snaze rozluštit kódy německého šifrovacího stroje Enigma. Jen tímto jedním použitím raného programovatelného digitálního počítače se podařilo zachránit miliony životů a o celé roky zkrátit válku s Německem. V roce 2014 se tento počin stal inspirací pro americký historický film *Kód Enigmy*.

Digitální elektronický počítač pracující s posloupnostmi číslic 0 a 1 (binárním kódem) byl schopen provádět matematické operace, tedy svým způsobem matematicky uvažovat. To podnítilo množství vědců z řady oborů, včetně matematiky, psychologie, techniky, ekonomie i politologie k bádání, jakým způsobem by počítač mohl případně napodobovat lidský mozek. V 50. letech 20. století pak matematici začali přicházet s tím, že by počítač mohl svým binárním kódem simulovat jakékoli matematické úvahy.

Během léta roku 1956 uspořádali Marvin Minsky, asistent z Harvardu, John McCarthy, odborný asistent na Dartmouth College, a Claude Shannon a Nathan Rochester, dva hlavní vědečtí pracovníci IBM, první konferenci o umělé inteligenci.[6] Konference se konala na Dartmouth College v Hanoveru (stát New Hampshire, USA). Účastnili se jí i počítačový vědec a kognitivní psycholog Allen Newell a politolog, ekonom, sociolog, psycholog a počítačový vědec Herbert Simon. Později vešli Minsky, McCarthy, Newell a Simon ve světovou známost jako zakladatelé umělé inteligence.[7] Společně se svými studenty svou prací v následujících letech ohromili svět. Jejich počítačové programy

učily počítače řešit algebraické i slovní problémy, generovat logická tvrzení nebo mluvit anglicky.

Z těchto prvních průkopníků umělé inteligence vyzařoval bezbřehý optimismus. V roce 1958 například Herbert Simon a Allen Newell tvrdili, že „během deseti let digitální počítač porazí mistra světa v šachu“.[8] Historie jim nakonec dala za pravdu, když počítač IBM nazývaný „Deep Blue“ v roce 1997 porazil tehdejšího mistra světa v šachu Garryho Kasparova[9], s časovým odhadem se však Simon s Newellem příliš netrefili. Celkově lze říci, že vědci z oboru umělé inteligence schopnosti raných počítačů přeceňovali, a naopak podceňovali vynořující se problémy.

Začátkem 60. let badatelé na poli umělé inteligence upoutali pozornost Ministerstva obrany USA. V červnu 1963 získal Massachusettský technologický institut (MIT) od Agentury pro výzkum pokročilých obranných projektů (DARPA) grant ve výši 2,2 milionu dolarů na financování projektu Matematiky a výpočtů (MAC)[10], do něhož byla začleněna skupina „AI Group“, založená pět let předtím Minskym a McCarthym. Rozpočet agentury vzrostl do poloviny 70. let až na úroveň 3 milionu dolarů ročně a ta štědře podporovala rovněž Newellův a Simonův program na Univerzitě Carnegieho–Mellonových (CMU)[11] i projekt umělé inteligence na Stanfordově univerzitě (Stanford AI), založený Johnem McCarthym v roce 1963. V roce 1965 pak založil další významnou laboratoř umělé inteligence Donald Michie z Edinburské univerzity.[12] Tyto čtyři instituce se staly centry výzkumu umělé inteligence v 60. a 70. letech 20. století.

Výzkumníci nazývali léta 1956–1974 po dartmouthské konferenci „zlatým věkem umělé inteligence“. S podporou milionů dolarů proudících do výzkumu fascinovali svými úspěchy celý svět. Jen si představte to ohromení, jaké většina lidí cítila, když počítače z dřevních dob výpočetní techniky řešily algebraické slovní problémy nebo na ně mluvily anglicky. Počítače samy o sobě byly tehdy velkou novinkou, a zprávy o jejich inteligentním chování hraničily se zázrakem. Nově získaná finanční podpora pak posouvala počítače směrem k finálnímu cíli „obecné umělé inteligence“, nazývané též

„silná umělá inteligence“, představující situaci, kdy počítač dosáhne úrovně lidské inteligence.

Tito vědci dokonce začali vyvíjet testy pro posouzení, kdy se inteligence počítače vyrovná inteligenci lidské. Jedním z těchto testů, dodnes platným, je Turingův test.[13] V roce 1950 publikoval počítačový vědec, matematik, logik, kryptoanalytik a teoretický biolog Alan Turing revoluční zprávu, v níž definoval test na základě staré společenské hry. V podstatě šlo o to, že Turing tvrdil, že pokud by byl počítač schopen vést konverzaci prostřednictvím telegrafu, která by byla k nerozlišení od konverzace s člověkem, bylo by možno počítač považovat za myslící stroj a ekvivalent lidského jedince. Zajímavé je, že podle této definice nemusí být počítač ve své promluvě bezchybný. Člověk položí otázku a stroj by mu měl odpovědět, ale nevadí, když tato odpověď bude chybná. Však se chyb při konverzaci dopouštějí i lidé. Důležitým momentem zde je, aby nezaujatý pozorovatel při čtení přepisu dialogu člověka se strojem nebyl schopen poznat, kdo je kdo. Badatelé na poli umělé inteligence vytvořili pro určení, zda se inteligence stroje již rovná inteligenci člověka, i množství jiných testů, avšak Turingův jednoduchý, ale přesvědčivý test se stal v tomto směru „zlatým standardem“. Jen jako poznámku na okraj, i o Alanu Turingovi ví mnoho lidí jen díky populárnímu filmu *Kód Enigmy* (v originále *The Imitation Game*) z roku 2014, v němž Turing sestrojil počítač, který rozluštil kód nacistů za 2. světové války.

Bohužel miliony dolarů proudící do výzkumu umělé inteligence během let 1956 až 1974 daly vzniknout ještě většímu optimismu, který se ukázal jako nepodložený. V roce 1965 Simon předpovídal, že „stroje budou schopny během dvaceti let vykonávat jakoukoli lidskou práci“.[14] S tím souhlasil i Minsky a v roce 1967 prohlásil, že „během jedné generace … bude problém vytvoření umělé ‚inteligence‘ do značné míry vyřešen“.[15] V článku v časopise *Life Magazine* v roce 1970 Minsky tvrdil ještě optimističtěji: „Za tři až osm let budeme mít stroj s obecnou umělou inteligencí průměrného lidského jedince.“[16]

Problémy, s nimiž se umělá inteligence setkala v 70. letech, se však ukázaly jako nepřekonatelné. Nejvýznamnějším z nich byla omezená

výkonnost tehdejších počítačů. Počítače měly tehdy malou kapacitu paměti a slabý výpočetní výkon. Lze říci, že nejvýkonnější počítače ze začátku 70. let 20. století ve svých parametrech převyšuje dnešní průměrný mobilní telefon. Při těchto omezeních stačily tehdejší počítače pouze na poměrně jednoduché problémy a úlohy. A jak vyprchávalo nadšení z nové věci, začalo se na umělou inteligenci pohlížet čím dál více spíše jako na hračku v rukou vědců.

Nedostatečná výkonnost počítačů byla zdaleka největší překážkou v dosažení silné umělé inteligence, problémů však tuto snahu doprovázela celá řada. Například už čtyřleté dítě dokáže rozpoznávat obličeje a bavit se s jinými lidmi. Zato aplikace umělé inteligence, jako je počítačové vidění nebo přirozený jazyk, se staly pro počítače ze začátku 70. let nepřekonatelným problémem, neboť postrádaly dostatečnou databázi informací o světě a výkon potřebný pro určování významu. Bez této nepostradatelné výbavy nebyly počítače tehdejší doby schopny rozeznávat předměty ani hovořit o jednoduchých tématech.

Halasná propagace umělé inteligence ze strany jejích prvních průkopníků vedla při nedostatečném výkonu tehdejších počítačů k nastavení nereálných cílů. Oblast umělé inteligence poutala od poloviny 60. let mimořádnou pozornost. Veškerý optimismus první generace badatelů na tomto poli se však ukázal jako nepodložený a v roce 1974 financování výzkumu umělé inteligence začalo vysychat, což mezi roky 1974–1980 vedlo k období nazývanému „zima umělé inteligence“.[17]

Vývoj umělé inteligence pak stagnoval až do začátku 80. let. Novou vzpruhu opět dostal v podobě úspěšného nástupu „expertních systémů“. Expertní systémy jsou počítačové programy, které napodobují rozhodovací schopnosti lidského experta. S tímto přístupem se cíl vytvořit obecnou umělou inteligenci, tedy stroje, které by myslely jako lidé, ocitl na vedlejší koleji, a pozornost se zaměřila na konkrétní úlohy. Příkladem takového expertního systému je třeba šachový program na dnešních „chytrých mobilech“. Úspěch expertních systémů

znovu otevřel přívod peněz do vývoje umělé inteligence, tentokrát celosvětově již v řádu miliard dolarů ročně.

Příliv financí do umělé inteligence však poté začal opět opadat, a to především kvůli propadu trhu s hardwarem specializovaným pro umělou inteligenci, k němuž došlo takřka přes noc v roce 1987. Stolní počítače vyráběné firmami Apple a IBM získávaly stabilně na výpočetním výkonu a tržním podílu na úkor dražších počítačů LISP,[18] které byly určeny pro vědecké a technické aplikace, vyráběné firmou Symbolics a několika dalšími. Počínaje rokem 1987 již stolní počítače disponovaly srovnatelným výkonem jako počítače LISP, avšak za daleko nižší ceny. Prodej počítačů LISP, který měl do té doby hodnotu přes půl miliardy dolarů, se náhle propadl.[19] Kromě toho začal mezi roky 1986 a 1989 FED (Federální rezervní systém, obdoba centrální banky v USA) zvyšovat úrokové sazby v boji proti rostoucí inflaci, což mělo negativní dopad na ekonomický růst. Růst cen ropy v roce 1990 ve spojení s narůstajícím ekonomickým spotřebitelským pesimismem vedl začátkem 90. let ke krátké hospodářské recesi, v jejímž důsledku vládní výdaje dále klesly. Například koncem 80. let vláda Spojených států výrazně zredukovala finanční podporu programu Strategické počítačové iniciativy (Strategic Computing Initiative), z níž byly čerpány zdroje pro pokročilou umělou inteligenci. Rovněž změny ve vedení agentury DARPA vedly k novému způsobu podpory vývoje: agentura přestala přidělovat prostředky jednotlivým vědcům, místo toho je vyčleňovala na konkrétní projekty s přesněji definovanými cíli, které mohly přinést okamžité výsledky. Ve výsledku tak agentura DARPA přesměrovala peníze od expertů na umělou inteligenci k jiným programům, jež splňovaly nová pravidla. V roce 1991 podobně došla i japonská vláda k závěru, že její projekt počítačů páté generace, který měl vést k rozvoji umělé inteligence, nesplnil cíle vytyčené v roce 1981, a projekt postupně ukončila.

Od konce 80. let do začátku 90. let 20. století se vývoj umělé inteligence nacházel ve stavu velkých turbulencí, které se projevovaly těmito rysy:

- Malý trh pro danou technologii – provoz prvních expertních systémů se ukázal jako drahý; užitečné byly pouze v určitých speciálních případech, na rozdíl od stolních počítačů, které disponovaly vyšším počítačovým výkonem za nižší cenu.
- Ekonomické ochlazení ve Spojených státech a krátká recese začátkem 90. let – hospodářský pokles vedl ke snížení vládních výdajů a ke škrtům ve financování vývoje umělé inteligence.
- Skepticismus vůči umělé inteligenci – původní optimistické cíle kladené na programy v oblasti umělé inteligence se ukázaly jako nereálné.

Tyto faktory vedly dohromady mezi roky 1987 a 1993 ke druhé, drsnější „zimě umělé inteligence“[20]. Máte-li z tohoto líčení situace pocit, že vývoj umělé inteligence probíhal od počátku jako na horské dráze, je váš dojem správný. Vědci a experti na umělou inteligenci od konce 60. let do začátku 90. let prožívali ve financování své práce střídání období hojnosti a skromnosti.

A jak to při většině neúspěchů bývá, roztrhl se pytel s hledači chyb. Jedni obviňovali umělou inteligenci z toho, že podlehla přílišnému optimismu kvůli snu o inteligenci srovnatelné s lidskou, druzí sváděli neúspěch na výkyvy ve finanční podpoře vývoje umělé inteligence ve stylu „ode zdi ke zdi“. A obě strany měly svůj díl pravdy. Technologie nebyla připravena na přehnaně optimistické cíle, stejně jako se na problémech vývoje umělé inteligence podepsaly dramatické výkyvy ve finanční podpoře.

V roce 1993 se nacházel vývoj umělé inteligence na dně, podobně jako otřesený soupeř při odpočítávání v ringu, byl však dostatečně daleko od případného „knokautu“. Životadárným elixírem pro umělou inteligenci se stal příchod integrovaných obvodů a nových počítačových technologií. Umělá inteligence se připravovala na to, aby znovu ohromila svět.

Za vrchol lidské inteligence bývá často považována schopnost hrát šachy. V roce 1996 sponzorovala firma IBM šachové klání na šest

partií mezi svým superpočítačem Deep Blue a tehdejším mistrem světa v šachu Garrym Kasparovem. Počítač Deep Blue dokázal analyzovat 200 000 tahů za sekundu. Jen málokdo však tehdy věřil, že by se nějaký stroj mohl vyrovnat mistrovství nejlepšího světového šachisty. Kasparov také podle očekávání při souboji ve Philadelphii počítač Deep Blue s přehledem porazil. Zdálo se tehdy zjevné, že ani superpočítače se nemohou lidskému mozku rovnat. Kasparov i IBM nicméně souhlasili o rok později, tedy v roce 1997, s dalším střetnutím v New Yorku. Tentokrát však výsledek zápasu šokoval svět. Počítač Deep Blue Kasparova těsně porazil. Kasparov nato obvinil IBM z podvodu a prohlásil, že během hry pozoroval u stroje „hlubokou inteligenci" a „kreativitu", z čehož podle něj plynulo, že během druhého utkání museli do hry počítače zasahovat živí šachisté, což by bylo porušením pravidel. Firma IBM jakékoli lidské zásahy do hry popřela. Pravidla dovolovala programátorům upravovat program mezi utkáními, aby měli možnost řešit nedostatky zjištěné ve hře počítače během partie. Kasparov měl svým způsobem pravdu. Když si však žádal odvetu, IBM ji odmítla a počítač Deep Blue již do šachových partií dále nenasazovala. Druhý zápas s Kasparovem se vysílal živě po internetu a vysloužil si titulní zprávy v médiích celého světa. Mnoho šachových mistrů připisovalo Kasparovovu porážku jeho mimořádně slabé hře v tomto utkání. Avšak lidé začali, alespoň pokud jde o šachy, akceptovat myšlenku, že počítače mohou přemýšlet lépe než lidé. A i když existují rozumné argumenty pro to, že Kasparov mohl počítač Deep Blue, opírající se při svém „uvažování" o primitivní hrubou sílu, při hře na své obvyklé úrovni porazit, ještě sofistikovanější šachové programy jasně ukázaly, že počítače mohou mít nad člověkem jednoznačně navrch. Utkání mezi počítačem Deep Blue a Garrym Kasparovem vešlo do dějin jako symbolický bod zlomu, kdy stroj poprvé dominoval nad člověkem. Tato událost si vysloužila světovou pozornost a stala se inspirací pro dokumentární film, *The Man vs. The Machine*.[21]

A soubojů „člověk proti počítači" začalo přibývat. Například:

- V únoru 2011 porazil ve vědomostní soutěži televizního pořadu *Jeopardy!* (v české verzi *Risk*) počítač IBM jménem Watson dva nejúspěšnější vítěze této soutěže, Brada Ruttera a Kena Jenningse.[22]
- V roce 2012 přidělila agentura DARPA výzkumný grant 1,3 milionu dolarů firmě SoftWear Automation na vytvoření robota, který by šil látky,[23] a tato investice se vyplácí. Firma vyvinula levného robota, který se vyrovná nejschopnějším lidským šičkám. Ministerstvo obrany USA dává při nákupu uniforem obecně přednost americkým dodavatelům, většinou však nakonec nakupuje od těch zahraničních díky nižší ceně práce. Celkově v současnosti Spojené státy dovážejí ročně oděvy a látky za 100 miliard dolarů ze zemí jako Čína a Vietnam. Firma SoftWear Automation se to snaží změnit výrobou levných robotů, kteří by v americkém textilním průmyslu nahradili levné zahraniční dodavatele.
- Americké automobilky již nyní používají roboty k bodovému svařování. Průměrné hodinové náklady na robotické svařování jsou 8 dolarů oproti 25 dolarům na živé svářeče.[24]
- V lednu 2014 vyslovil generál armády USA Robert Cone předpověď, že roboti by do roku 2030 mohli nahradit čtvrtinu všech vojáků, díky čemuž by se armáda mohla stát „menší, údernější, snadněji nasaditelnou a pohotovější silou".[25] Dnes například armáda USA zaměstnává roboty pro deaktivaci tzv. improvizovaných výbušných zařízení (IED).
- Podle článku na *DailyMail.com* z roku 2015 provádějí „roboti v současnosti přibližně 10 procent výrobních úkonů … a tento podíl by měl do roku 2025 vzrůst zhruba na 25 procent".[26]

Celkově vzato dokážou roboti s umělou inteligencí předčit lidi v řadě činností. Patří mezi ně míchání nápojů, zneškodňování improvizovaných nástražných systémů a bomb, vyplňování lékařských předpisů, prořezávání vinné révy ve vinicích, vytrhávání plevele z blízkosti rostlin, vakuování obalů, vyhledávání určitých formulací

a významových celků v právních dokumentech, pokladní bankovní operace (např. bankomaty), skladové operace jako například vychystávání balíků pro označení čárovým kódem a celá řada dalších činností.

Tyto úspěchy umělé inteligence, zvláště u robotických aplikací, jsou výsledkem inženýrských schopností a zároveň ohromného výpočetního výkonu dnešních počítačů. Například počítač IBM Deep Blue byl 10 milionkrát rychlejší než počítač Ferranti Mark 1 Christophera Stracheyho s šachovým programem v roce 1951. Dnešní chytré telefony jsou výkonnější než počítače NASA používané při letu člověka na Měsíc.

Ačkoli vědecký vývoj umělé inteligence má za sebou pouhých 60 let, prosákl již téměř do každého prvku moderní společnosti a moderního vojenství. Sotva jsme si toho ovšem všimli, a ani to nepřipisujeme výkonnosti počítačů. Oxfordský filozof Nick Bostrom vysvětluje:[27] „Špičková umělá inteligence se v nemalé míře stala nepozorovaně součástí všeobecných aplikací, aniž bychom většinou hovořili o umělé inteligenci. Protože jakmile se něco stane dostatečně užitečným a běžným, nezískává to už nálepku umělé inteligence." Někteří experti nazývají tento jev „efektem umělé inteligence".[28] Stalo se už totiž samozřejmostí očekávat, že počítač, který dnes koupíte, bude dvakrát tak výkonný než ten, který jste koupili před dvěma roky. Dnešní špičkové herní počítače mají grafiku srovnatelnou s grafickou kvalitou televizního filmu. Před dvaceti lety bychom je nazývali „simulátory" a využívali je například k výcviku letců. Není pochyb o tom, že výkonnost počítačů roste exponenciálně. Důsledkem toho se exponenciálně zdokonaluje i umělá inteligence. Divíte se, jak je tak nepolevující růst možný? Odpovědí je tzv. Moorův zákon.

V roce 1975 si Gordon E. Moore, spoluzakladatel firem Intel a Fairchild Semiconductor, všiml, že počet tranzistorů na stejně velkých integrovaných obvodech se přibližně každé dva roky zdvojnásobuje, zatímco cena integrovaných obvodů se nemění. Polovodičový průmysl přijal Moorův zákon za svůj a začal podle něj plánovat nabídku svých produktů.[29] Moorův zákon se tak stal sebenaplňujícím