

Formy superinteligence

Co přesně tedy míníme „superinteligencí“? Jakkoliv nechceme zabředat do terminologických močálů, musíme svůj pojmový rámec několika slovy projasnit. V této kapitole rozlišíme tři formy superinteligence a budeme hájit tvrzení, že z praktického hlediska jsou navzájem ekvivalentní. Rovněž ukážeme, že potenciál strojové inteligence je větší než potenciál biologických mozků. Stroje mají oproti mozkům řadu zásadních výhod, díky nimž získají drtivou převahu. Biologičtí lidé, i pokud budou vylepšeni, budou poraženi na celé čáře.

V úzce vymezených oblastech se již nyní výkony mnoha strojů a zvířat pohybují na nadlidské úrovni. Netopýři interpretují sonarové signály lépe než člověk, kalkulačky nás předčí v aritmetice a počítačové programy nás porázejí v šachu. Množství specifických úloh, které software dokáže vykonávat lépe než my, se bude nadále zvětšovat. Přestože však specializované systémy na zpracovávání informací budou mít řadu uplatnění, některé hluboké otázky vyvstávají pouze v souvislosti s vidinou strojových intelektů, jejichž obecná inteligence bude postačovat, aby lidi zastoupily ve všech oblastech intelektuální činnosti.

Jak řečeno výše, termínem „superinteligence“ označujeme intelekty, které napříč mnoha velmi širokými kognitivními oblastmi výrazně překonávají nejlepší lidské duchy dneška. Podle této stále docela vágní definice by za superinteligenci mohly být považovány různé

druhy systémů se značně odlišnými výkonnostními charakteristikami. Abychom s analýzou pokročili, bude užitečné tento prostý pojem superinteligence rozložit a rozlišit různé balíky intelektuálních superschopností. Lze to udělat mnoha způsoby. Zde bude řeč o třech formách superinteligence: rychlostní, kolektivní a kvalitativní.

Rychlostní superinteligence

Jako rychlostní superinteligenci označujeme intelekt, který je stejný jako lidská mysl, ale rychlejší. Tuto formu superinteligence lze konceptuálně analyzovat nejnáz.¹⁷⁰ Můžeme ji definovat následovně.

Rychlostní superinteligence: *Systém, jenž dokáže dělat vše, co umí lidský intelekt, ale mnohem rychleji.*

„Mnohem“ zde znamená něco jako „o několik řádů“. Ale než bychom se snažili z této definice odstranit všechny stopy vágnosti, necháme na čtenáři, aby ji sám rozumně interpretoval.¹⁷¹

Nejjednodušším příkladem rychlostní superinteligence by byla úplná emulace mozku, která by běžela na rychlém hardwaru.¹⁷² Emulace, jež by byla desettisíckrát rychlejší než biologický mozek, by dokázala přečíst knihu během několika sekund a napsat disertaci za odpoledne. Pokud by byla oproti mozku zrychlená milionkrát, dokázala by celé tisíciletí intelektuální práce vykonat během jednoho pracovního dne.¹⁷³

Z pohledu takové rychlé mysli by se události ve vnějším světě odvíjely jako ve zpomaleném filmu. Předpokládejme, že by vaše mysl byla 10 000krát rychlejší. Pokud by pak váš přítel z masa a kostí třeba upustil na zem hrníček, mohli byste pozorovat, jak šálek po několika hodin pomalu klesá ke koberci, podoben kometě, jež tiše plachtí vesmírem na schůzku se vzdálenou planetou. Jak by se očkávání nárazu opožděně začalo šířit mozgovými závitky vašeho přítele a následně jeho periferní nervovou soustavou, viděli byste, jak jeho tělo postupně zaujímá postoj zmrazeného úleku – a za tu dobu

byste stačili nejen objednat náhradní hrníček, nýbrž si také zdřímnout a přečíst pár vědeckých článků.

Kvůli této zdánlivé dilataci času v hmotném světě by rychlostní superinteligence dávala přednost práci s digitálními objekty. Mohla by žít ve virtuální realitě a obchodovat v informační ekonomice. Případně by také mohla s fyzickým prostředím interagovat prostřednictvím nanoskopických manipulačních zařízení, neboť ta by mohla pracovat rychleji než makroskopické údy. (Charakteristická frekvence nějakého systému obvykle bývá nepřímo úměrná jeho rozměrům.¹⁷⁴) Rychlé mysli by se možná stýkaly raději mezi sebou než s pomalými a těžkopádnými lidmi.

S rostoucí rychlostí inteligence se stává čím dál významnějším omezením rychlost světla, neboť rychlé mysli při spotřebě času na cestování nebo na komunikaci na dlouhé vzdálenosti čelí vyšším nákladům obětované příležitosti.¹⁷⁵ Světlo je asi milionkrát rychlejší než tryskové letadlo, takže digitálnímu agentovi s milionkrát urychlenou myslí by cesta kolem světa zabrala přibližně stejné množství subjektivního času jako dnešnímu lidskému cestovateli. Někoho na dálku telefonicky kontaktovat by trvalo stejně dlouho jako se k němu dostavit „osobně“, byť by zavolání vyšlo levněji, protože by nevyžadovalo tolik přenosové kapacity. Pokud by spolu značně urychlené mysli chtěly často komunikovat, mohlo by pro ně být výhodné se přestěhovat blízko k sobě. Extrémně rychlé mysli, jež by potřebovaly být často v kontaktu (například členové téže pracovní skupiny), by mohly obývat počítače umístěné v téže budově, aby se vyhnuly frustrujícím zpožděním.

Kolektivní superinteligence

Další formou superinteligence je systém, jenž vyšších výkonů dosahuje tím, že spojuje v jeden celek velké množství menších myslí.

Kolektivní superinteligence: *Systém složený z velkého množství menších intelektů takovým způsobem, že jeho celková výkonnost napříč mnoha*

velmi širokými oblastmi dalece překonává výkonnost jakéhokoliv dnešního kognitivního systému.

Z pojmového hlediska není kolektivní superinteligence tak jasná jako rychlostní superinteligence.¹⁷⁶ Jsme s ní však lépe obeznámeni empiricky. Zatímco s intelekty lidské úrovně, které by pracovaly výrazně jinou rychlostí, nemáme žádnou zkušenost, naše zkušenosti s kolektivní inteligencí – tedy se systémy sestávajícími z různého množství složek lidské úrovně, které navzájem s různou mírou efektivitly spolupracují – jsou hojné. Pokud zaujmeme poněkud abstraktnější hledisko, můžeme na firmy, pracovní týmy, sítě drbů, zájmové skupiny, akademické komunity, státy, dokonce i na lidstvo jako celek pohlížet jako na volně definované „systémy“, které dokážou řešit různé druhy intelektuálních úloh. Ze zkušenosti máme jisté povědomí, jak snadno si organizace rozličné velikosti a rozličného složení poradí s různými úkoly.

Kolektivní inteligence vyniká v řešení problémů, které se dají snadno rozložit na části tím způsobem, že řešení dílčích problémů lze hledat souběžně a ověřovat je nezávisle na sobě. Takové úlohy jako stavba raketoplánu nebo řízení řetězce rychlého občerstvení skýtají nesčetně příležitostí pro dělbu práce: různí inženýři pracují na rozdílných součástkách raketoplánu, různí zaměstnanci obsluhují různé restaurace. Způsob, jakým jsou v akademické sféře badatelé, studenti, časopisy, granty a ocenění rigidně rozděleni do samostatných a soběstačných oborů (což příliš nenapomáhá tomu typu zkoumání, který představuje tato kniha), můžeme – ovšem jen ve shovívavém a smířlivém rozpoložení – chápat jako nutný kompromis, jenž řeší praktické otázky spojené s tím, že velké množství rozličně motivovaných jednotlivců a týmů dostane možnost přispívat k růstu lidského poznání, a přitom pracovat relativně samostatně, každý na svém písečku.

K nárůstu kolektivní inteligence systému může dojít buď tak, že se zvýší množství jeho složek, nebo tak, že selepší způsob jejich organizace.¹⁷⁷ Pokud bychom chtěli jakoukoliv současnou kolektivní inteligenci přeměnit v kolektivní *superinteligenci*, bylo by k tomu

zapotřebí dosti výrazných zlepšení. Bylo by nutné, aby výsledný systém dokázal dalece překonat jakoukoliv dnešní kolektivní inteligenci či jiný kognitivní systém napříč mnoha velmi širokými oblastmi. Vznik nového formátu konferencí, který by umožnil efektivnější výměnu informací mezi vědci, ani objev nového algoritmu pro společné filtrování informací, který by lépe předvídal uživatelské hodnocení knih a filmů, by samy o sobě zjevně ještě neznamenalý nástup něčeho, co by se blížilo kolektivní superinteligenci. Totéž platí pro 50% nárůst světové populace či pro takové zlepšení pedagogických metod, díky kterému by studenti mohli ve škole denně trávit čtyři hodiny namísto šesti. Ke splnění kritéria kolektivní superinteligence by byl zapotřebí mnohem výraznější nárůst kolektivní kognitivní kapacity lidstva.

Povšimněme si, že hranice pro kolektivní superinteligenci je definována ve vztahu k výkonnostní úrovni v současné době, tj. na začátku 21. století. V průběhu lidské prehistorie a potom opět v průběhu lidské historie již k mnohonásobnému zvýšení kolektivní inteligence lidstva došlo. Od pleistocénu například světová populace narostla přinejmenším tisíckrát.¹⁷⁸ Už proto můžeme říci, že současná úroveň kolektivní inteligence lidstva se *relativně vůči výchozí pleistocenní úrovni* blíží superinteligenci. Dalo by se rovněž tvrdit, že též některé pokroky v oblasti komunikačních technologií (zvláště vznik mluveného jazyka, ale možná také vznik měst, psaní a tisku) přinesly ohromné zlepšení kolektivní inteligence – takové zlepšení, že pokud by se někdy objevily další inovace se srovnatelným vlivem na naši kolektivní schopnost řešit intelektuální problémy, vedlo by to k příchodu kolektivní superinteligence.¹⁷⁹

Někteří čtenáři budou v pokušení nás na tomto místě přerušit a upozornit, že moderní společnost nepůsobí kdovíjak inteligentně. Možná politici v čtenářově zemi zrovna učinili nějaké nežádoucí rozhodnutí a jeho zjevná nerozumnost tíží čtenářovu mysl coby doklad nedostatečných mentálních schopností moderního lidstva. A není snad pravda, že dnešní lidé mají za modlu konzum, plýtvají přírodními zdroji, znečišťují životní prostředí, decimují druhovou rozmanitost a zároveň nedokážou napravit do očí bijící globální

nespravedlnosti a zanedbávají ty nejdůležitější humanistické či duchovní hodnoty? Ponecháme-li však stranou otázku, jak jsou na tom nedostatky současné doby ve srovnání se selháními předchozích epoch (jež také nebyla právě zanedbatelná), naše definice superinteligence nijak neimplikuje, že společnost s vyšší kolektivní inteligencí na tom nutně bude lépe. Tato definice dokonce neimplikuje ani to, že taková společnost bude *rozumnější*. Rozumnost můžeme chápat jako schopnost rozhodovat se v důležitých věcech přibližně správně. Můžeme si potom představit organizaci složenou z velmi velkého množství velmi efektivně koordinovaných znalostních pracovníků, kteří společně dovedou řešit intelektuální problémy napříč mnoha značně širokými oblastmi. Předpokládejme, že tato organizace dokáže řídit většinu druhů podniků, vynalézt většinu typů technologií a optimalizovat většinu druhů procesů. I tehdy by se mohla v několika klíčových otázkách dalekosáhlého významu zcela splést – například by mohla zanedbat náležitou obranu proti existenčním rizikům – a ve výsledku tak zažít krátký prudký růst, jenž neslavně skončí naprostým kolapsem. Taková organizace by mohla mít velmi vysokou kolektivní inteligenci; pokud dostatečně vysokou, šlo by o kolektivní superinteligenci. Měli bychom odolat pokušení spojovat všechny žádoucí atributy v jediný beztvary pojem mentálního fungování, jako kdyby jeden obdivuhodný znak musel vždy být doprovázen všemi ostatními. Namísto toho bychom měli uznat, že mohou existovat systémy na zpracovávání informací – inteligentní systémy –, které budou výkonnými nástroji, avšak nebudou zaručeně rozumné ani morálně dobré. K tomuto tématu se vrátíme v kapitole 7.

Kolektivní superinteligence může být buď volněji, nebo těsněji integrovaná. Pro ilustraci prvního případu si představme planetu *Velezemi*, jejíž komunikační a koordinační technologie se nacházejí na stejné úrovni jako na naší Zemi, avšak jejíž populace je milionkrát početnější. Úměrně tomu by byla větší také celková intelektuální pracovní síla Velezemě. Předpokládejme, že vědecký génius Newtonova nebo Einsteinova kalibru se rodí alespoň jeden na deset miliard lidí: v tom případě by na Velezemi 700 000 takových géníů žilo současně, po boku úměrně ohromného množství trochu

menších talentů. Nové myšlenky a technologie by se rodily ve zběsilém tempu a globální velezemská civilizace by tvořila volně integrovanou kolektivní superinteligenci.¹⁸⁰

Budeme-li míru integrace systému postupně zvyšovat, možná nakonec dostaneme sjednocený *intelekt* – jedinou obří „mysl“ v protikladu k pouhému seskupení menších lidských myslí, které spolu volně interagují.¹⁸¹ Obyvatelé Velezemě by mohli podniknout kroky v tomto směru, kdyby zlepšovali své komunikační a koordinační technologie a vyvíjeli lepší způsoby, jak by na nějakém náročném intelektuálním problému mohlo pracovat mnoho jednotlivců zároveň. Takto by se kolektivní inteligence, až by dosáhla dostatečného stupně integrace, mohla stát „kvalitativní superinteligencí“.

Kvalitativní superinteligence

Můžeme rozlišit též třetí formu superinteligence, kterou nazveme kvalitativní.

Kvalitativní superinteligence: *Systém alespoň stejně rychlý jako lidská mysl a kvalitativně mnohem chytřejší.*

Pojem kvality inteligence, stejně jako pojem kolektivní inteligence, je poněkud nejasný. Tato obtíž je v daném případě dále umocněna tím, že nemáme žádné zkušenosti s kvalitou inteligence přesahující horní mez dnešního rozdělení lidské inteligence. Nějakou představu o tomto pojmu však získáme, když zvážíme několik souvisejících případů.

Zprvė můžeme zvětšit rozsah svých referenčních bodů a zaměřit se na živočichy, jejichž inteligence je nižší kvality. (To není míněno jako „speciestická“ poznámka, jež by vycházela z přesvědčení o nadřazenosti lidského druhu. Kvalita inteligence zebřičky je dokonale adaptována na požadavky jejího životního prostředí, zde je však namísto více antropocentrická perspektiva: zajímají nás výkony v těch složitých kognitivních úlohách, které jsou relevantní

z lidského pohledu.) Ostatní živočichové postrádají komplexní strukturovaný jazyk, nástroje nedokážou používat a vyrábět buď vůbec, nebo jenom primitivním způsobem, jejich schopnost dlouhodobého plánování a abstraktního uvažování je značně omezená. Pro žádná z těchto omezení přitom není dostatečným vysvětlením nižší rychlostní nebo kolektivní inteligence zvířat. Pokud jde o hrubý výpočetní výkon, lidské mozky pravděpodobně zaostávají za mozky některých velkých zvířat, třeba slonů nebo velryb. A přestože by naše komplexní technologická civilizace bez naší obrovské převahy v oblasti kolektivní inteligence nebyla možná, na této převaze nezávisí všechny charakteristiky lidské kognitivní schopnosti. Mnohé z nich jsou velmi rozvinuté i u malých a izolovaných skupin lovců-sběračů.¹⁸² Naproti tomu u vysoce organizovaných mimolidských živočichů, jako jsou šimpanzi či delfíni vytrénovaní lidskými instruktory nebo mravenci žijící ve svých vlastních rozsáhlých a dobře uspořádaných společenstvích, mnohé tyto schopnosti zdaleka tak dobře rozvinuté nejsou. Zásahu za pozoruhodné intelektuální úspěchy druhu *Homo sapiens* lze zjevně z velké části přičíst zvláštním charakteristikám architektury našeho mozku, jež závisejí na naší jedinečné genetické výbavě, kterou ostatní zvířata nesdílejí. Tento postřeh nám pomáhá ilustrovat pojem kvalitativní superinteligence: jde o inteligenci, jejíž kvalita přesahuje kvalitu lidské inteligence alespoň o tolik, o kolik ta přesahuje kvalitu slonů, delfínů nebo šimpanzů inteligence.

Zadruhé můžeme pojem kvalitativní superinteligence ilustrovat tak, že poukážeme na dílčí kognitivní deficity, jež mohou postihnout lidské jednotlivce, zvláště takové deficity, které nejsou zapříčiněny celkovou demencí ani jinými onemocněními spojenými s hromadnou destrukcí neurovýpočetních zdrojů. Vezměme si například jedince s poruchami autistického spektra, kteří mají nápadné nedostatky v oblasti sociální kognice, zatímco v ostatních kognitivních oblastech fungují dobře, nebo jedince trpící vrozenou amúzií, kteří nedokážou rozpoznat nebo si zabroukat ani jednoduché melodie, ale ve většině ohledů jsou jejich kognitivní výkony normální. Mohli bychom uvést mnoho dalších příkladů z neuropsychiatrické literatury,

jež překypuje případovými studiiemi pacientů s úzce vymezenými deficity způsobenými genetickými poruchami nebo poraněními mozku. Takové příklady ukazují, že normální dospělí lidé mají celou řadu pozoruhodných kognitivních vloh, které nejsou jednoduše funkcí dostatečného množství obecného výpočetního výkonu, ba ani dostatečné obecné inteligence: jsou pro ně zapotřebí také specializované nervové okruhy. Tento postřeh hovoří pro myšlenku *možných, ale nerealizovaných kognitivních vloh*, tedy vloh, které žádný skutečný člověk nemá, přestože jiné inteligentní systémy – ty s vyšším výpočetním výkonem, než má lidský mozek –, které by těmito vlohami disponovaly, by díky tomu dokázaly mnohem lépe vykonávat široké spektrum úkolů strategického významu.

Díky úvaze o zvířatech a o lidech s dílčími kognitivními deficity si tak můžeme utvořit jistou představu o různých kvalitách inteligence a jejich praktických dopadech. Pokud by *Homo sapiens* postrádal (například) ty kognitivní moduly, které umožňují komplexní jazykové reprezentace, možná by byl jen jedním z opičích druhů žijících v souladu s přírodou. A kdybychom naopak *získali* novou skupinu modulů, která by nám poskytla podobné výhody, jaké přináší schopnost tvořit komplexní jazykové reprezentace, stali bychom se superinteligentními.

Přímý a nepřímý dosah

Jakákoliv z těchto forem superinteligence by mohla časem vyvinout technologie potřebné k vytvoření těch ostatních. Všechny tudíž mají tentýž *nepřímý dosah*. V tomto smyslu náleží do stejné třídy ekvivalence také nepřímý dosah současné lidské inteligence (za předpokladu, že nakonec budeme schopni vytvořit nějakou formu superinteligence). V jiném smyslu jsou si však navzájem mnohem bližší ony tři formy superinteligence: každá z nich by k ostatním formám zvládla dospět mnohem rychleji, než to z našeho současného výchozího bodu dokážeme my.

Přímý dosah všech tří forem superinteligence lze porovnat obtížněji.

Možná mezi nimi není žádné jednoznačné pořadí. Schopnosti jednotlivých forem superinteligence závisejí na tom, v jakém stupni každá z nich uskutečňuje svou vlastní přednost – *jak* rychlá je rychlostní superinteligence, *jak* velkou kvalitativní převahu má kvalitativní superinteligence atd. Nanejvýš můžeme říci, že za jinak stejných podmínek rychlostní superinteligence vyniká v úlohách vyžadujících rychlé vykonání dlouhé řady kroků, které musejí být prováděny postupně, zatímco kolektivní superinteligence vyniká v úlohách připouštějících analytický rozklad na dílčí úkoly, jež lze vykonávat souběžně, a v úkolech vyžadujících spojení mnoha různých úhlů pohledu a souborů dovedností. V jistém vágním smyslu by kvalitativní superinteligence byla nejvýkonnější ze všech, neboť by dokázala chápat a řešit problémy, které by prakticky vzato byly mimo *přímý* dosah rychlostní a kolektivní superinteligence.¹⁸³

V některých oblastech je kvantita chabou náhražkou za kvalitu. Jediný osamělý génius ukrytý ve své korkem obložené ložnici dokázal napsat *Hledání ztraceného času*. Mohlo by vzniknout podobné mistrovské dílo, pokud bychom si za tím účelem najali budovu plnou námezdnicích pisálků?¹⁸⁴ I v rámci současného rozdělení inteligence můžeme vidět, že u některých úkolů se práce jednoho výjimečného génia vyplatí více než námaha spousty průměrných fušerů. Rozšíříme-li záběr i na *superinteligentní* mysl, budeme muset připustit, že pravděpodobně budou existovat intelektuální problémy, jež budou řešitelné jenom pro superinteligenci, nikoliv pro jakýkoliv, ať jakýkoliv velký kolektiv nevylepšených lidí.

Možná tedy existují úlohy, jež dokáže vyřešit jenom kvalitativní a snad i rychlostní superinteligence, ale nikoliv volně integrovaná kolektivní superinteligence (pokud nejprve nezvýší svou vlastní inteligenci).¹⁸⁵ Nemůžeme mít jasno, jaké všechny problémy to jsou, ale můžeme je obecně charakterizovat.¹⁸⁶ Obvykle by šlo o problémy zahrnující množství vzájemně provázaných složek, jejichž řešení nesestávají ze samostatně ověřitelných kroků, tedy o problémy, které nelze řešit kus po kuse a které by mohly vyžadovat kvalitativně nové způsoby porozumění nebo nové reprezentační rámce příliš hluboké nebo komplikované na to, aby je dnešní smrtelníci dokázali objevit

nebo efektivně používat. Do této kategorie by mohly spadat některé typy umělecké tvorby, strategického uvažování a možná i některé vědecké objevy. Lze také spekulovat, že lidský pokrok v mnoha filozofických „věčných problémech“ je tak pomalý a kostrbatý, protože lidská kůra mozková se na filozofickou práci nehodí. Podle tohoto pohledu se naši nejproslulejší filozofové podobají psům chodícím po zadních – jenom stěží dosahují výkonnostní úrovně potřebné, aby se dané činnosti mohli *vůbec* věnovat.¹⁸⁷

Výhody digitální inteligence

Jak je patrné ze vzájemného srovnání intelektuálních a technologických úspěchů lidí a ostatních lidoopů, drobné změny v objemu a konektivitě mozku mohou mít následky zásadního významu. Dopady těch změn ve výpočetním výkonu a výpočetní architektuře, které přinese strojová inteligence, budou pravděpodobně ještě významnější. Vytvořit si intuitivní představu o schopnostech superinteligence je pro nás obtížné, ne-li nemožné; nějaké ponětí o možnostech, jež se nabízejí, však můžeme získat, když se podíváme na některé výhody dostupné digitálním myslím. Nejsnáž viditelné jsou přednosti jejich hardwaru:

— *Rychlost výpočetních prvků.* Biologické neurony pracují na maximální frekvenci asi 200 Hz, tedy o celých sedm řádů pomaleji než moderní mikroprocesory (~ 2 GHz).¹⁸⁸ Následkem toho je lidský mozek odkázán na masivní paralelizaci a nedokáže rychle provést žádný výpočet, který vyžaduje postupné vykonání velkého množství operací.¹⁸⁹ (Pro úkony trvající méně než sekundu nemůže mozek použít o moc více než stovku sekvenčních operací – snad jenom několik tuctů.) Mnohé z algoritmů, které mají největší praktický význam v programování a informatice, však nelze snadno paralelizovat. Leckteré kognitivní úlohy by se daly vykonávat mnohem efektivněji, pokud by mozek kromě paralelizovatelných algoritmů na vybírání podle vzoru podporoval také rychlé sekvenční procesování.

- **Rychlost vnitřní komunikace.** Axony přenášejí akční potenciály rychlostí asi 120 m/s nebo nižší, zatímco elektronická procesorová jádra mohou komunikovat opticky, rychlostí světla (300 000 000 m/s).¹⁹⁰ Pomalost nervových signálů omezuje maximální velikost mozku, který by ještě fungoval jako jedna procesorová jednotka. Nemá-li například obousměrné zpoždění mezi jakýmkoliv dvěma prvky systému přesáhnout 10 ms, nesmí mít biologický mozek větší objem než 0,11 m³. Elektronický systém by oproti tomu mohl mít $6,1 \times 10^{17}$ m³, což zhruba odpovídá objemu trpasličí planety: mohl by tedy být o osmnáct řádů větší než biologický mozek.¹⁹¹
- **Množství výpočetních prvků.** Lidský mozek má o něco méně než sto miliard neuronů.¹⁹² Je tedy asi třiapůlkrát větší než mozek šimpanze (ale pětkrát menší než mozek vorvaně obrovského).¹⁹³ U biologických organismů jsou hranice pro množství neuronů dány v první řadě velikostí lebky a metabolickými omezeními, ale u větších mozků mohou hrát roli i další faktory (jako chlazení, doba vývoje a zpoždění při vedení signálu – viz předchozí bod). Oproti tomu počítačový hardware lze zvětšovat prakticky neomezeně, dokud nenarazíme na (velmi vzdálené) fyzikální meze.¹⁹⁴ Superpočítače mohou být rozměrné jako skladiště (nebo ještě rozměrnější) a jejich kapacitu lze ještě zvětšit prostřednictvím vysokorychlostního spojení se vzdálenými počítači.¹⁹⁵
- **Paměťová kapacita.** Lidská pracovní paměť nedokáže v jakémkoliv daném okamžiku podržet více než asi čtyři nebo pět jednotek informací.¹⁹⁶ Jakkoliv by bylo zavádějící velikost lidské pracovní paměti přímo srovnávat s množstvím RAM dostupným digitálnímu počítači, je jasné, že vzhledem ke svým hardwarovým výhodám bude digitální inteligence disponovat větší pracovní pamětí. Díky tomu by taková mysl mohla být schopna intuitivně chápat složité vztahy, s nimiž si lidé dovedou poradit jenom tápavě s pomocí těžkopádných výpočtů.¹⁹⁷ Lidská dlouhodobá paměť má rovněž omezenou kapacitu, ačkoliv není jasné, jestli ji za dobu jednoho běžného lidského života zvládne vyčerpát – tak pomalé je tempo, jímž informace hromadíme. (Podle jednoho odhadu je v dospělém lidském mozku uložena asi jedna miliarda bitů dat, tedy o dva řády méně než v levném chytrém

telefonu.¹⁹⁸) Množství skladovaných informací i rychlost přístupu k nim by tedy byly u strojového mozku větší než u biologického.

- **Spolehlivost, životnost, senzory atd.** Strojová inteligence by mohla mít různé další hardwarové výhody. Biologické neurony například nejsou tak spolehlivé jako tranzistory.¹⁹⁹ Jelikož jsou u výpočtů doprovázených informačním šumem zapotřebí nadbytečné znakové kódy, které jediný bit informací kódují prostřednictvím několika různých prvků, mohla by se efektivita digitálního mozku zvýšit tím, že by používal vysoce přesné výpočetní prvky. Mozky se po několika hodinách práce unaví a po několika desetiletích subjektivního času začínají nenávratně chátrat; tato omezení se mikroprocesorů netýkají. Tok dat přicházejících do strojové inteligence by se dal zvětšit připojením milionů senzorů. Strojový mozek by (v závislosti na použitých technologiích) mohl mít rekonfigurovatelný hardware, který by šlo optimalizovat pro proměnlivé požadavky různých úloh. Naproti tomu velká část architektury mozku se od narození nemění buď vůbec, nebo jenom pomalu (ačkoliv v detailech se synaptická konektivita během kratších časových období, třeba dnů, proměňovat může).²⁰⁰

V současnosti srovnání mezi výpočetním výkonem biologických mozků a digitálních počítačů stále vyznívá ve prospěch prvně jmenovaných, ačkoliv výkonnost nejlepších superpočítačů již spadá do rozmezí, v němž se pohybují hodnověrné odhady výpočetního výkonu mozku.²⁰¹ Hardware se ale rychle zlepšuje a horní hranice jeho možné výkonnosti se nachází mnohem výše než u biologických výpočetních substrátů.

Digitální mozky budou rovněž těžit z významných softwarových výhod:

- **Možnost úprav.** V softwaru lze s hodnotami různých parametrů experimentovat snáze než v nervovém „wetwaru“. U úplné emulace mozku bychom například mohli snadno vyzkoušet, k čemu dojde, když do nějaké oblasti mozkové kůry přidáme více neuronů nebo když zvýšíme či snížíme jejich vzrušivost. Provádět takové experimenty na živých biologických mozcích by bylo mnohem obtížnější.

- **Kopírování.** Software nám umožňuje dostupnou hardwarovou základnu rychle zaplnit libovolným množstvím velmi přesných kopií programu. Biologické mozky se oproti tomu dají reprodukovat jenom velice pomalu – a každá nová kopie se na začátku nachází v bezradném stavu, bez jakékoliv vzpomínky na to, co se během svého života naučili její předci.
- **Koordinace cílů.** Výkonnost lidských kolektivů značně omezuje skutečnost, že je téměř nemožné dosáhnout toho, aby všichni členové velké skupiny sdíleli přesně tytéž cíle – alespoň dokud lidi nebude možné ve velkém měřítku činit povolnými pomocí drog nebo genetické selekce. U „klanu kopií“ (skupiny skoro nebo úplně identických programů sdílejících společný cíl) by tyto problémy s koordinací zmizely.
- **Sdílení paměti.** Biologické mozky musejí za účelem zisku nových vzpomínek a dovedností podstoupit dlouhodobou výuku a výcvik, zatímco digitálním myslím stačí si vyměnit datové soubory. Populace sestávající z miliardy kopií jednoho UI programu by mohla periodicky synchronizovat své databáze, takže by každá kopie věděla vše, co se jakákoliv jiná naučila během předchozí hodiny. (Přímý přenos paměti vyžaduje standardizované formáty reprezentace. Snadná výměna vysokoúrovňového kognitivního obsahu by tak nebyla proveditelná mezi jakýmkoliv dvěma strojovými inteligencemi – zvláště ne mezi dvěma úplnými emulacemi mozku první generace.)
- **Nové moduly, modality a algoritmy.** Zrakové vnímání nám připadá jako něco zcela snadného, zcela rozdílného od řešení úloh z učebnice geometrie – a to vzdor skutečnosti, že aby z dvourozměrných vzorců na sítnici zrekonstruoval trojrozměrnou reprezentaci světa zalidněného rozpoznatelnými předměty, musí mozek provést ohromné množství výpočetních operací. Důvodem, proč se nám tento proces zdá snadný, je to, že na zpracovávání vizuálních informací máme vyhrazenou nízkoúrovňovou nervovou mašinerii. Toto nízkoúrovňové zpracování informací probíhá nevědomě a automaticky, aniž by vyčerpávalo naši mentální energii nebo vědomou pozornost. Zvláštní neurovýpočetní moduly jsou zřejmě zasvěceny také vnímání hudby, práci s jazykem, sociální kognici a dalším formám zpracovávání

informací, které jsou pro nás lidi „přirozené“. Umělé mysli, jež by měly specializovanou podporu pro jiné kognitivní oblasti, které v dnešním světě nabyly na významu – jako pro inženýrství, počítačové programování a obchodní strategie –, by měly velké výhody oproti myslím, jako jsou naše, které jsou při přemýšlení o takových věcech odkázány na nemotornou víceúčelovou kognici. Mohly by také být vyvinuty nové algoritmy, které by využívaly specifických možností digitálního hardwaru, například jeho podpory pro rychlé sériové procesování.

Sečteme-li je, jsou softwarové a hardwarové výhody, kterých by strojová inteligence *nakonec* mohla dosáhnout, ohromné.²⁰² Avšak jak rychle by se tyto potenciální výhody mohly stát skutečností? K této otázce se obrátíme nyní.