

Laskavé nové stroje budoucnosti⁽¹⁾

Jak by UI mohla změnit svět k lepšímu

Autor: [Dario Amodei](#)

Původní text: [Dario Amodei — Machines of Loving Grace](#)

Pokud vás v textu zaujme nějaká myšlenka, diskutujte o ní v komentářích. Strojový překlad¹ byl vytvořen primárně pro vzdělávací účely Sciopolis a komunity UčímsAI.cz. Pokud narazíte na nepřesnost nebo chybu, prosím, uveďte to také v komentáři. Díky.

Hodně přemýšlím a mluvím o rizicích spojených s vyspělou umělou inteligencí. Společnost Anthropic, které jsem CEO, se intenzivně věnuje výzkumu, jak tato rizika zmírnit. Lidé si proto někdy myslí, že jsem pesimista nebo skeptik, který vidí v UI především negativa a nebezpečí. Tak to ale vůbec není. Na rizika se soustředím proto, že představují jedinou překážku na cestě k tomu, co považuji v zásadě za pozitivní budoucnost. Myslím si, že většina lidí podceňuje jak mimořádný přínos UI, tak závažnost jejích rizik.

V této eseji se pokusím nastítnit, jak by tato pozitiva mohla vypadat – jak by mohl vypadat svět s vyspělou UI, pokud by se vše vyvíjelo ideálně. Nikdo samozřejmě nemůže znát budoucnost s jistotou a dopady vyspělé UI budou pravděpodobně ještě nepředvídatelnější než minulé technologické změny, takže vše bude nevyhnutelně jen spekulací. Snažím se však nabídnout alespoň informované a praktické odhady zachycující hlavní směry budoucího vývoje, i když se většina detailů nakonec ukáže jako mylná. Uvádím mnoho podrobností hlavně proto, že se domnívám, že konkrétní vize podněcuje diskuzi lépe než obecné a abstraktní úvahy.

Nejprve bych ale chtěl krátce vysvětlit, proč já a Anthropic o pozitivěch vyspělé UI tolik nemluvíme a proč se celkově budeme pravděpodobně i nadále spíše soustřeďovat na rizika. Toto rozhodnutí pramení z mé snahy:

- **Maximalizovat vliv.** Základní vývoj technologií UI a většina jejich přínosů (i když ne všechny) se jeví jako nevyhnutelné - pokud ovšem celý proces nezhatí právě rizika. Za tímto vývojem stojí silné tržní síly. Rizika naproti tomu předem daná nejsou a my svým jednáním můžeme výrazně ovlivnit jejich pravděpodobnost.

¹

Hrubý překlad: DeepL

Redakční úpravy: Gemini 1.5 Pro 002 a Claude 3.5 Sonnet

Prompting: Andrej Novik

- **Vyhnut se nařčení z propagandy.** Když technologické firmy mluví o úžasných výhodách UI, mohou působit jako propagandisté snažící se zakrýt negativa. Také věřím, že přílišné vychvalování vlastní práce člověku z principu škodí.
- **Vyhnut se grandióznosti.** Často mě odrazuje, jak mnozí veřejní experti na rizika UI (a samozřejmě i vedoucí představitelé firem zabývajících se UI) mluví o světě po singularitě, jako by byli proroky, jejichž posláním je dovést lidstvo do zaslíbené země. Myslím si, že je nebezpečné vnímat firmy jako jednostranné tvůrce světa a stejně tak nebezpečné je vnímat praktické technologické cíle v zásadě nábožensky.
- **Vyhnut se „sci-fi“ konotacím.** Ačkoli si myslím, že většina lidí pozitiva vyspělé UI podceňuje, malá komunita lidí, která o radikální budoucnosti UI diskutuje, tak činí často v přehnaně „sci-fi“ tónu (např. s tématy nahrávání mysli, dobývání vesmíru či kyberpunkové estetiky). Domnívám se, že to vede k tomu, že lidé berou tato tvrzení méně vážně a vnímají je jako nereálná. Abych to upřesnil, nejde o to, zda jsou popsané technologie možné či pravděpodobné (v eseji se tím podrobně zabývám) – jde spíše o „atmosféru“, která s sebou nese spoustu kulturních konotací a nevyslovených předpokladů o tom, jaká budoucnost je žádoucí, jak se budou vyvíjet různé společenské otázky atd. Výsledkem pak často bývá text, který připomíná fantazii pro úzkou subkulturu a většinu lidí spíše odradí.

I přes výše uvedené obavy si myslím, že je důležité diskutovat o tom, jak by mohl vypadat ideální svět s vyspělou UI, a zároveň se snažit vyvarovat se zmíněných úskalí. Dokonce si myslím, že je zcela zásadní mít skutečně inspirativní vizi budoucnosti, a ne jen plán pro řešení krizových situací. Mnoho aspektů vyspělé UI je problematických či nebezpečných, ale na konci cesty musí být něco, za co bojujeme, nějaký pozitivní výsledek, kde si všichni polepší, něco, co lidi sjednotí a pomůže jim povznést se nad malicherné spory a čelit budoucím výzvám. Strach je sice motivátor, ale sám o sobě nestačí: potřebujeme i naději.

Seznam pozitivních aplikací vyspělé UI je ohromující (zahrnuje robotiku, výrobu, energetiku a mnoho dalšího), já se ale zaměřím na několik oblastí, které mají dle mého soudu největší potenciál pro přímé zlepšení kvality lidského života. Pět kategorií, které považuji za nejzajímavější:

1. Biologie a fyzické zdraví
2. Neurovědy a duševní zdraví
3. Ekonomický rozvoj a boj s chudobou
4. Mír a správa věcí veřejných
5. Práce a smysl života

Mé předpovědi se můžou zdát podle většiny měřítek radikální (pomineme-li sci-fi vize „singularity“ (2), ale myslím je vážně a upřímně. Vše, co zde uvádím, se snadno

může ukázat jako mylné (jak už jsem zmínil), ale snažil jsem se alespoň podložit své názory poloanalytickým zhodnocením, jak moc by se mohl urychlit pokrok v různých oborech a co by to mohlo znamenat v praxi. Mám sice odborné zkušenosti v biologii a neurovědách a jsem informovaný laik v oblasti rozvojové ekonomiky, ale jsem si jistý, že se v mnoha věcech mýlím. Práce na této eseji mi mimo jiné ukázala, jak cenné by bylo shromáždit skupinu expertů z různých oborů (biologie, ekonomie, mezinárodní vztahy atd.), aby sepsali mnohem lepší a informovanější verzi toho, co jsem tu nastínil. Tento text by měl proto především posloužit jako odrazový můstek pro práci takové skupiny.

Základní předpoklady a rámec

Pro lepší pochopení a přesnější diskusi v této eseji je důležité nejprve definovat, co míníme vyspělou UI (a od kterého bodu se začne odpočítávat oněch 5–10 let). Zároveň si stanovíme rámec pro zhodnocení dopadů UI, jakmile jí bude dosaženo.

To, jak bude vyspělá UI (termín AGI se mi nelíbí) (3) vypadat a kdy (a zda vůbec) se objeví, je samo o sobě komplexní téma. Už jsem o něm veřejně hovořil a mohl bych mu věnovat celou samostatnou esej (a pravděpodobně to i udělám). Je zřejmé, že mnoho lidí je skeptických k brzkému vzniku vyspělé UI a někteří pochybují, že se objeví vůbec. Osobně si myslím, že by k tomu mohlo dojít už v roce 2026, ale jsou tu i scénáře, které počítají s mnohem delší dobou. V této eseji však podobné otázky ponechám stranou. Budu předpokládat, že vyspělá UI vznikne v relativně blízké budoucnosti, a soustředím se na vývoj v následujících 5-10 letech. Rád bych také vycházel z určité definice toho, jak bude takový systém vypadat, jaké budou jeho schopnosti a jak bude interagovat, i když se v tomto ohledu názory různí.

Pod vyspělou UI si představuji model UI – pravděpodobně formou podobný dnešním velkým jazykovým modelům (LLM), ačkoli by mohl být založen na odlišné architektuře, mohl by zahrnovat několik vzájemně propojených modelů a mohl by být trénován jinak – s těmito vlastnostmi:

- **Pokud jde o samotnou inteligenci**, (4) je chytřejší než nositel Nobelovy ceny ve většině relevantních oborů – biologie, programování, matematika, inženýrství, psaní atd. To znamená, že dokáže řešit dosud nevyřešené matematické problémy, psát vynikající romány, tvořit složité programy od nuly atd.
- Kromě toho, že je to „chytrá věc, se kterou si povídáte“, disponuje všemi „rozhraními“ dostupnými člověku pracujícímu virtuálně, včetně textu, audia, videa, ovládání myši a klávesnicí a přístupu k internetu. Může provádět jakékoli akce, komunikovat či provádět vzdálené operace umožněné tímto rozhraním, včetně aktivit na internetu, navigování lidí, objednávání materiálů, řízení experimentů, sledování a tvorby videí atd. Všechny tyto úkoly zvládá, opět s dovednostmi převyšující ty nejzdatnější lidi na světě.
- Nejen pasivně odpovídá na otázky; místo toho mu lze zadat úkoly, jejichž splnění trvá hodiny, dny nebo i týdny, a UI se jich pak autonomně ujme,

podobně jako chytrý zaměstnanec, a v případě potřeby si vyžádá další informace.

- Nemá fyzické tělo (existuje pouze na obrazovce počítače), ale dokáže ovládat skutečné fyzické nástroje, roboty či laboratorní vybavení prostřednictvím počítače. Teoreticky by si mohla dokonce navrhnout roboty nebo vybavení pro vlastní potřebu.
- Zdroje použité pro trénování modelu lze využít i ke spuštění milionů jeho instancí (což odpovídá předpokládané velikosti clusterů kolem roku 2027) a model dokáže zpracovávat informace a generovat akce zhruba 10–100krát rychleji než člověk. (5) Může však být limitován rychlostí reakce fyzického světa nebo softwaru, se kterým interaguje.
- Každá z těchto milionů kopií může fungovat nezávisle na sobě na různých úkolech, nebo v případě potřeby spolupracovat, stejně jako lidé, s různými podskupinami specializovanými na konkrétní úkoly.

Dalo by se to shrnout jako „stát géniů v datovém centru“.

Taková entita by bezpochyby dokázala řešit velmi složité problémy a to velmi rychle. Otázkou ale zůstává, jak rychle. Dva „extrémní“ názory se mi zdají mylné. První říká, že svět by se proměnil v řádu vteřin či dnů („singularita“), jakmile by se nadřazená inteligence začala sama rozvíjet a řešit všechny vědecké, inženýrské a operační úkoly téměř okamžitě. Problém je v tom, že narážíme na reálná fyzická a praktická omezení, například v oblasti konstrukce hardwaru nebo provádění biologických experimentů. I hypotetický stát géniů by na tyto limity narazil. Inteligence je sice velmi účinný nástroj, ale nemůžeme od ní čekat zázraky.

Druhý, opačný názor, tvrdí, že technologický pokrok je nasycený nebo omezený reálnými daty či společenskými faktory, a že inteligence převyšující lidskou k němu přidá jen velmi málo. (6) To se mi zdá stejně nepravděpodobné – dovedu si představit stovky vědeckých nebo i společenských problémů, u kterých by velká skupina skutečně chytrých lidí výrazně urychlila pokrok, obzvláště pokud by se neomezovali jen na analýzu, ale mohli by reálně zasahovat do světa (což by náš hypotetický stát géniů mohl, a to i skrze koordinaci a podporu lidských týmů).

Pravda bude pravděpodobně někde mezi těmito dvěma extrémy, bude se lišit v závislosti na úkolu a oboru a bude se skrývat v detailech. Domnívám se, že k produktivnímu zamyšlení se nad těmito detaily potřebujeme nové myšlenkové rámce.

Ekonomové často hovoří o „výrobních faktorech“: jde o práci, půdu a kapitál. Pojem „mezní výnosy práce/půdy/kapitálu“ vyjadřuje myšlenku, že v dané situaci může být určitý faktor limitující – například letectvo potřebuje letadla i piloty, a když najme víc pilotů, ale nemá dost letadel, moc si nepomůže. Věřím, že ve věku UI bychom měli hovořit o mezních výnosech inteligence (7) a snažit se najít další faktory, které inteligenci doplňují a které se stávají limitujícími, jakmile je inteligence velmi vysoká. Nejsme zvyklí tímto způsobem přemýšlet – ptát se „o kolik nám v tomto úkolu

pomůže vyšší inteligence a za jak dlouho?“ – ale zdá se mi to být správný způsob, jak nahlížet na svět s velmi vyspělou UI.

Mezi faktory, které inteligenci omezují nebo doplňují, dle mého názoru patří:

- **Rychlost reálného světa.** Intelligentní systémy musí být v interakci s okolním světem, aby dosahovaly svých cílů a učily se.(8) Svět se ale mění jen určitou rychlostí. Buněčné procesy a životní cykly organismů mají své přirozené tempo, které nelze urychlit, a proto experimenty vyžadují určitý minimální čas. Totéž platí pro hardware, materiálové vědy, cokoli, co zahrnuje interakci s lidmi, a dokonce i naši současnou softwarovou infrastrukturu. Ve vědě je navíc často nutné provádět experimenty postupně, jeden po druhém, a každý další se učí z předchozích. To vše znamená, že i když se inteligence bude neustále zvyšovat, rychlost dokončení velkého projektu – například vývoje léku na rakovinu – může mít určité neredukovatelné minimum.
- **Potřeba dat.** Někdy chybí nezpracovaná data a v takovém případě ani vyšší inteligence nepomůže. Dnešní částicoví fyzici jsou velmi vynalézaví a vyvinuli řadu teorií, ale chybí jim data, aby se mezi nimi mohli rozhodnout, protože data z urychlovačů částic jsou omezená. Není jasné, jestli by si vedli o tolik lépe, i kdyby byli superintelligentní – snad kromě toho, že by urychlili konstrukci většího urychlovače.
- **Vnitřní složitost.** Některé jevy jsou ze své podstaty nepředvídatelné a chaotické (9) a ani sebelepší UI je nedokáže předpovídat nebo dešifrovat o mnoho lépe než dnešní lidé a počítače. Například ani neuvěřitelně výkonná UI nedokáže v obecném případě předvídat v chaotickém systému (například v problému tří těles) o moc dále než současné algoritmy.
- **Omezení daná lidmi.** Řada věcí by vyžadovala porušení zákonů, ublížení lidem nebo narušení společenského řádu. UI se správným nastavením cílů (*alignment*) by tyto věci dělat nechtěla (a pokud by UI správné cíle neměla, vracíme se k diskusi o rizicích). Mnoho lidských společenských struktur je sice neefektivních, nebo dokonce aktivně škodlivých, ale těžko se mění, pokud chceme respektovat omezení, jako jsou zákonné požadavky na klinické studie, lidskou vůli měnit zažité zvyklosti nebo chování vlád. Mezi příklady pokroku, které z technického hlediska fungují dobře, ale jejichž dopad byl výrazně omezen regulacemi nebo zbytečnými obavami, patří jaderná energie, nadzvukové lety a dokonce i výtahy.
- **Fyzikální zákony.** Jde o přísnější verzi prvního bodu. Fyzikální zákony stanovují hranice, které nedokážeme překročit. Nelze cestovat rychleji než světlo. Nemůžeme zvrátit přirozené fyzikální děje. Hustota tranzistorů na čipu má své hranice - při jejich překročení klesá spolehlivost čipu. Výpočet vyžaduje určité minimální množství energie na smazání bitu, což omezuje hustotu výpočtů.

Je třeba zohlednit také časovou perspektivu. Co nás dnes omezuje, můžeme v delším časovém horizontu pomocí inteligence překonat. Například inteligence by mohla být využita k vývoji nových experimentálních postupů, které by nám umožnily zjistit in vitro to, co dříve vyžadovalo pokusy na živých zvířatech. Inteligence může pomoci sestavit nástroje pro získávání nových dat (například větší urychlovače částic). V rámci etických mezí může také najít způsoby, jak překonat limity dané lidským faktorem. Může například zefektivnit systém klinických testů, pomoci vytvořit správné oblasti s menší byrokracií nebo vylepšit vědecké postupy tak, aby se snížila potřeba klinických testů na lidech nebo se testy zlevnily.

Měli bychom si tedy představit situaci, kdy je inteligence zpočátku silně limitována ostatními výrobními faktory, ale postupem času si sama nachází způsoby, jak tato omezení překonat, i když nikdy úplně nezmizí (a některé, jako fyzikální zákony, jsou absolutní). (10) Klíčová otázka zní, jak rychle se tohle všechno stane a v jakém pořadí.

S ohledem na tento rámec se pokusím odpovědět na tuto otázku v pěti oblastech zmíněných v úvodu.

1. Biologie a zdraví

Biologie je pravděpodobně oblastí, kde má vědecký pokrok největší potenciál přímo a jednoznačně zlepšit kvalitu lidského života. V minulém století se sice podařilo vymýtit některé prastaré nemoci jako neštovice, ale mnoho dalších stále sužuje lidstvo. Jejich překonání by znamenalo obrovský pokrok pro celé lidstvo. Kromě samotného léčení nemocí může biologická věda v principu zlepšit základní úroveň lidského zdraví – prodloužením aktivního života, zvýšením kontroly nad našimi biologickými procesy a řešením každodenních problémů, které v současnosti vnímáme jako nezměnitelné součásti lidské existence.

Použijeme-li koncept „omezujících faktorů“, který jsme popsali výše, hlavními výzvami při přímé aplikaci inteligence v biologii jsou data, rychlost fyzického světa a vnitřní komplexita (všechny tři spolu úzce souvisejí). Lidská omezení hrají roli také v pozdější fázi, v případě klinických studií. Pojdme si projít jednotlivé faktory.

Experimenty na buňkách, zvířatech, a dokonce i chemické procesy, jsou limitovány rychlostí fyzického světa: Mnoho biologických postupů vyžaduje kultivaci bakterií či jiných buněk nebo vyčkávání na dokončení chemických reakcí, což může trvat dny či týdny. V těchto případech není obvykle zjevné, jak by bylo možné průběh zrychlit. Pokusy na zvířatech zaberou měsíce (nebo i déle) a pokusy na lidech často trvají roky (či dokonce desetiletí v případě studií s dlouhodobým sledováním). S tím souvisí i nedostatek dat – ne ani tak kvantitativní, jako spíše kvalitativní. Vždy je nedostatek jasných a jednoznačných dat, která by izolovala studovaný biologický jev od ostatních 10000 rušivých faktorů, které kauzálně ovlivňují daný proces, nebo které přímo měří daný efekt (oproti nepřímému a nepřesnému odvozování jeho následků). I rozsáhlá kvantitativní molekulární data, jako jsou proteomická data, která jsem sbíral při práci na technikách hmotnostní spektrometrie, jsou nepřesná a leccos

opomíjejí (v jakých typech buněk se dané proteiny nacházely? V které části buňky? V jaké fázi buněčného cyklu?).

Za tyto problémy s daty může částečně i vnitřní komplexita systémů: Pokud jste někdy viděli diagram znázorňující biochemii lidského metabolismu, víte, jak těžké je izolovat vliv jakékoli jeho části, natožpak do něj přesně a předvídatelně zasahovat. A konečně, kromě času nutného k provedení experimentů na lidech zahrnují klinické studie spoustu byrokracie a regulačních požadavků, které (podle mnoha lidí, včetně mě) zbytečně prodlužují dobu výzkumu a vývoje.

Vzhledem k těmto faktorům byli mnozí biologové dlouhodobě skeptičtí k přínosu UI a „velkých dat“ v biologii. Matematici, informatici a fyzici, kteří v posledních 30 letech uplatňovali své znalosti v biologii, sice byli poměrně úspěšní, ale nedosáhli skutečně transformačního dopadu, v nějž se zpočátku doufalo. Část skepticismu se zmírnila díky zásadním a revolučním průlomům, jako je AlphaFold (který si právem vysloužil svým tvůrcům Nobelovu cenu za chemii) a AlphaProteo (11), ale stále přetrvává názor, že UI je (a bude) užitečná jen v omezeném okruhu situací. Často se říká: „UI sice dokáže lépe analyzovat vaše data, ale nedokáže jich nashromáždit víc, ani zlepšit jejich kvalitu. Z odpadu zlato neuděláte.“

Domnívám se však, že tento pesimistický pohled je chybný. Pokud je naše hypotéza o pokroku v UI správná, pak bychom o UI neměli uvažovat jako o metodě analýzy dat, ale jako o virtuálním biologovi, který vykonává všechny úkoly biologů, včetně navrhování a provádění experimentů (ovládáním laboratorních robotů, nebo prostým zadáváním úkolů lidským vědcům – stejně jako vedoucí výzkumné skupiny zadává úkoly svým studentům), vynalézáním nových biologických metod či měřicích technik atd. Právě urychlením celého výzkumného procesu může UI skutečně akcelarovat biologii. Chci to zdůraznit, protože se jedná o nejčastější mýlku, na kterou narážím, když mluvím o schopnosti UI transformovat biologii: Nemám na mysli UI jako pouhý nástroj pro analýzu dat. V souladu s definicí vyspělé UI na začátku této eseje mluvím o využití UI k provádění, řízení a zdokonalování téměř všeho, co biologové dělají.

Abych upřesnil, kde konkrétně očekávám urychlení výzkumu, překvapivě velká část pokroku v biologii pochází z velmi malého počtu objevů, často spojených s univerzálními měřicími nástroji nebo technikami (12), které umožňují přesné, a přitom generalizované nebo programovatelné zásahy do biologických systémů. Ročně se objeví zhruba jeden takový zásadní objev a dohromady pravděpodobně stojí za více než 50 % pokroku v biologii. Tyto objevy jsou tak mocné právě proto, že překonávají vnitřní složitost a omezení daná nedostatkem dat a přímo zvyšují naše chápání a kontrolu nad biologickými procesy. Několik objevů za desetiletí tak umožnilo nejen většinu našeho základního vědeckého poznání biologie, ale i vznik mnoha nejúčinnějších léčebných postupů.

Namátkou jmenujme:

- **CRISPR:** Technika umožňující editaci genů v živých organismech (nahrazení jakékoli genové sekvence jinou libovolnou sekvencí). Od vývoje původní

techniky dochází k neustálému zdokonalování v cílení na specifické buněčné typy, zvyšování přesnosti a snižování rizika nežádoucích editací – to vše je klíčové pro bezpečné využití u lidí.

- **Různé typy mikroskopie** pro pozorování dějů na buněčné úrovni: pokročilé světelné mikroskopy (s různými fluorescenčními technikami, speciální optikou atd.), elektronové mikroskopy, mikroskopy atomárních sil atd.
- **Sekvenování a syntéza genomu**, jejichž cena v posledních desetiletích dramaticky klesla.
- **Optogenetické techniky**, které umožňují aktivaci neuronů světlem.
- **mRNA vakcíny**, které v zásadě umožňují navrhnout vakcínu proti čemukoli a rychle ji adaptovat (mRNA vakcíny se proslavily během pandemie covidu-19).
- **Buněčné terapie** jako je CAR-T, které umožňují odebrat imunitní buňky z těla a „přeprogramovat“ je k útoku v podstatě na cokoli.
- **Konceptuální objevy** jako teorie choroboplodných zárodků nebo objev spojení mezi imunitním systémem a rakovinou. (13)

Všechny tyto technologie zmiňuji proto, abych zdůraznil jedno klíčové tvrzení: Věřím, že s větším počtem nadaných a kreativních vědců by se tempo objevů mohlo zdesetinásobit, možná i více. Jinak řečeno, investice do rozvoje inteligence mají u těchto objevů vysokou návratnost a vše ostatní v biologii a medicíně z nich do značné míry vyplývá.

Proč si to myslím? Kvůli odpovědím na některé otázky, které bychom si měli klást, když se snažíme určit „výnosy z inteligence“. Za prvé, tyto objevy obvykle učiní malý počet výzkumníků, často titíž lidé opakovaně, což svědčí spíše o dovednosti než o náhodném hledání (náhodné hledání by naopak naznačovalo, že limitujícím faktorem jsou dlouhé experimenty). Za druhé, tyto objevy „mohly být učiněny“ o roky dříve: například CRISPR byl přirozenou součástí imunitního systému bakterií, známou už od 80. let, ale trvalo dalších 25 let, než si lidé uvědomili, že ji lze využít pro editaci genů obecně. Často se také stává, že objevy jsou zpožděny o mnoho let kvůli nedostatečné podpoře slibných směrů ze strany vědecké komunity (viz profil vynálezce mRNA vakcín; podobných příběhů je mnoho). Za třetí, úspěšné projekty jsou často skromné nebo se jedná o vedlejší produkty výzkumu, které původně nikdo nepovažoval za slibné, spíše než o projekty s velkým rozpočtem. To ukazuje, že objevy nepohání jen množství zdrojů, ale především vynalézavost.

A konečně, ačkoli některé z těchto objevů mají „sériovou závislost“ (k objevu B je třeba nejprve objevit A) – což může opět způsobit zpoždění – mnoho, možná i většina z nich, je vzájemně nezávislých, takže na nich lze pracovat paralelně. Z těchto faktů a také z mých obecných zkušeností biologa usuzuji, že na objevení čekají stovky takových objevů, jen kdyby byli vědci chytřejší a lépe propojovali obrovské množství biologických poznatků, které lidstvo nashromáždilo (vzpomeňme si na příklad CRISPR). Úspěch AlphaFold/AlphaProteo v řešení důležitých problémů

mnohem efektivněji než lidé, a to i přes desetiletí pečlivě navrhovaného fyzikálního modelování, je důkazem principu (i když s omezeným nástrojem v omezené doméně), který by měl ukazovat cestu vpřed.

Domnívám se tedy, že vyspělá UI by mohla alespoň desetinásobně zvýšit tempo těchto objevů, což by nám dalo 50–100 let biologického pokroku za pouhých 5–10 let. (14) Proč ne stonásobně? Možná je to možné, ale zde hraje roli sériová závislost a čas potřebný k provedení experimentů: Získat 100 let pokroku za jeden rok vyžaduje, aby se mnoho věcí podařilo hned napoprvé, včetně experimentů na zvířatech a konstrukce mikroskopů či drahých laboratoří. Připouštím i myšlenku (jakkoli může znít absurdně), že bychom mohli získat 1000 let pokroku za 5–10 let, ale jsem velmi skeptický k tomu, že bychom mohli získat 100 let pokroku za 1 rok. Jinak řečeno, určitému zpoždění se nelze vyhnout: experimenty a vývoj hardwaru mají určitou „latenci“ a je nutné je opakovat určitý „neredukovatelný“ početkrát, než se podaří zjistit věci, které nelze odvodit logicky. Masivní paralelizace je ale možná i v tomto případě. (15)

A co klinické studie? Přestože jsou s nimi spojeny byrokratické překážky a zpoždění, velká část (i když zdaleka ne všechna!) jejich pomalosti pramení z nutnosti důsledně testovat léky, které mají jen malý, nebo nejednoznačný efekt. To bohužel platí pro většinu dnešních terapií: Průměrný lék na rakovinu prodlouží přežití jen o několik měsíců a má přitom značné vedlejší účinky, které je nutno pečlivě monitorovat (podobně je tomu i u léků na Alzheimerovu chorobu). To vede k rozsáhlým studiím (kvůli dosažení statistické síly) a složitým rozhodnutím, která regulační orgány obvykle neumějí dobře vyřešit – opět kvůli byrokracii a složitosti kombinování protichůdných zájmů.

U skutečně účinných řešení je celý proces mnohem rychlejší: U prokazatelně účinné léčby lze využít zrychlenou cestu ke schválení. mRNA vakcíny proti covidu-19 byly schváleny za 9 měsíců – mnohem rychleji než je obvyklé. I tak byly klinické zkoušky zbytečně zdlouhavé – mRNA vakcíny mohly získat schválení už za přibližně 2 měsíce. Nicméně toto zpoždění (zhruba 1 rok pro vývoj léku) v kombinaci s masivní paralelizací a s potřebou určité, ale ne příliš velké iterace („několika pokusů“) je velmi slučitelné s radikální transformací za 5–10 let. Ještě optimističtější je možnost, že UI urychlí biologický výzkum natolik, že se sníží potřeba iterací v klinických testech, a to díky vývoji lepších zvířecích a buněčných modelů (nebo dokonce simulací), které budou přesněji predikovat, co se stane u lidí. To bude obzvláště důležité při vývoji léků proti stárnutí, které probíhá v horizontu desítek let, a proto zde potřebujeme rychlejší iterační smyčku.

A konečně, v souvislosti s klinickými testy a společenskými bariérami je třeba zdůraznit, že biomedicínské inovace mají v některých ohledech neobvykle dobrou historii co se týče úspěšného zavádění, na rozdíl od jiných technologií. (16) Jak jsem uvedl v úvodu, mnoho technologií je brzděno společenskými faktory, přestože technicky fungují dobře. To by mohlo naznačovat pesimistický pohled na to, čeho lze s UI dosáhnout. Biomedicína je ale v tomto ohledu specifická – ačkoli je vývoj léků

zbytečně komplikovaný, jakmile jsou léky vyvinuty, obvykle se úspěšně zavádějí a používají.

Abych to shrnul, moje základní předpověď zní, že biologie a medicína s využitím UI nám umožní zkondenzovat pokrok, kterého by lidští biologové dosáhli za příštích 50–100 let, do 5–10 let. Budu to nazývat „zkondenzované 21. století“. Po vývoji vyspělé UI dosáhneme během několika let takového pokroku v biologii a medicíně, jakého bychom jinak dosáhli za celé 21. století.

Přestože je předvídání schopností vyspělé UI v horizontu několika let ze své podstaty obtížné a spekulativní, určitá konkrétnost se skrývá v otázce „čeho by lidstvo mohlo dosáhnout bez pomoci UI v příštích 100 letech?“. K odhadu potenciální úrovně pokroku dosažitelného s pomocí UI stačí zvážit, čeho jsme dosáhli ve 20. století, jaký byl vývoj v prvních dvou desetiletích 21. století, nebo se zeptat, co by nám přineslo „10krát CRISPR a 50krát CAR-T“.

Níže uvádím seznam toho, co bychom mohli očekávat. Není založen na žádné rigorózní metodologii a téměř jistě se ukáže v detailech chybný, ale snaží se nastínit obecnou míru radikality, se kterou bychom měli počítat:

- **Spolehlivá prevence a léčba téměř všech (17) přirozeně se vyskytujících infekčních onemocnění.** Vzhledem k enormnímu pokroku v boji proti infekčním onemocněním ve 20. století není nereálné si představit, že bychom v „zkondenzovaném 21. století“ mohli tuto práci víceméně dokončit. mRNA vakcíny a podobné technologie nám ukazují cestu k „vakcínám proti čemukoli“. Zda se podaří infekční choroby zcela vymýtit (na rozdíl od pouhého vymýcení v některých oblastech), závisí na otázkách chudoby a nerovnosti, o kterých pojednávám v části 3.
- **Eliminace většiny typů rakoviny.** Úmrtnost na rakovinu v posledních desetiletích klesá zhruba o 2 % ročně; současným tempem vědeckého pokroku tedy směřujeme k eliminaci většiny typů rakoviny v 21. století. Některé podtypy již byly z velké části vyléčeny (například některé typy leukémie terapií CAR-T) a já se osobně těším zejména na vysoce selektivní léky, které cílí na rakovinu v raném stádiu a zabraňují jejímu dalšímu růstu. UI také umožní vývoj léčebných režimů přesně přizpůsobených individuálnímu genomu pacienta s rakovinou – to je sice možné už dnes, ale je to časově a finančně extrémně náročné. UI by nám měla umožnit škálovat tuto personalizovanou léčbu. Zdá se, že snížení úmrtnosti i incidence o 95 % a více je reálné. Rakovina je ale extrémně rozmanitá a adaptivní a pravděpodobně jde o nemoc, kterou bude nejtěžší zcela vymýtit. Nebylo by překvapivé, kdyby některé vzácné a obtížně léčitelné formy rakoviny přetrvávaly.
- **Velmi účinná prevence a léčba genetických onemocnění.** Výrazně vylepšený screening embryí pravděpodobně umožní prevenci většiny genetických onemocnění a bezpečnější a spolehlivější metody odvozené od CRISPR by

mohly vyléčit většinu genetických onemocnění u již narozených pacientů. Problémem ale můžou být onemocnění postihující velkou část buněk v těle.

- **Prevence Alzheimerovy choroby.** Stále se nám nedaří plně pochopit příčiny Alzheimerovy choroby (souvisí s beta-amyloidním proteinem, ale detaily jsou zřejmě velmi složité). Zdá se, že jde přesně o ten typ problému, který lze vyřešit s pomocí lepších měřicích nástrojů, které izolují biologické účinky. Jsem proto optimistický, že UI tento problém dokáže vyřešit. Je velmi pravděpodobné, že Alzheimerově chorobě půjde předcházet poměrně jednoduchými zásahy, jakmile pochopíme její skutečné příčiny. Poškození již způsobené Alzheimerovou chorobou ale bude pravděpodobně velmi obtížné zvrátit.
- **Vylepšená léčba většiny ostatních onemocnění.** Toto je zastřešující kategorie pro další nemoci, jako je cukrovka, obezita, srdeční choroby, autoimunitní onemocnění atd. Většina z nich se zdá být „snadněji“ řešitelná než rakovina či Alzheimerova choroba a v mnoha případech se jejich výskyt již rapidně snižuje. Například úmrť na srdeční choroby klesla o více než 50 % a jednoduché intervence jako agonisté GLP-1 už dosáhly velkého pokroku v boji proti obezitě a cukrovce.
- **Biologická svoboda.** Posledních 70 let přineslo pokrok v antikoncepci, metodách asistované reprodukce, regulaci hmotnosti a mnoha dalších oblastech. Domnívám se ale, že UI urychlí biologii natolik, že se možnosti v těchto oblastech výrazně rozšíří: hmotnost, fyzický vzhled, reprodukce a další biologické procesy budou plně pod kontrolou lidí. Budu to nazývat biologickou svobodou: Každý by měl mít možnost svobodně se rozhodnout, jakým směrem se bude jeho tělo a život vyvíjet. Samozřejmě se objeví důležité otázky týkající se celosvětově rovného přístupu k těmto technologiím; více v části 3.
- **Zdvojnásobení délky lidského života.** (18) Může se to zdát radikální, ale průměrná délka života se ve 20. století téměř zdvojnásobila (z ~40 na ~75 let), takže je „v souladu s trendem“ předpokládat, že by se v „zkondenzovaném 21. století“ mohla opět zdvojnásobit, na 150 let. Metody pro zpomalení samotného procesu stárnutí se samozřejmě budou lišit od těch, které v minulém století sloužily k prevenci (převážně dětských) předčasných úmrtí, ale samotný rozsah změny není bezprecedentní. (19) Už dnes existují léky, které prodlužují maximální délku života u potkanů o 25–50 % s minimálními negativními účinky. A některá zvířata (např. některé druhy želv) se dožívají 200 let, takže je zřejmé, že lidé se nenacházejí na nějaké teoretické horní hranici délky života. Odhaduji, že nejdůležitější by mohly být spolehlivé biomarkery stárnutí, které nepodléhají Goodhartovu zákonu. Ty by umožnily rychlou iteraci experimentů a klinických studií. Jakmile se lidský život prodlouží na 150 let, možná dosáhneme „únikové rychlosti“, získáme tím dostatek času a většina

dnes žijících lidí se bude moci dožít, jak dlouho bude chtít, i když samozřejmě neexistuje záruka, že je to biologicky možné

Stojí za to se na tento seznam podívat a zamyslet se nad tím, jak odlišný by byl svět, kdyby se všechny tyto body naplnily za 7–12 let (což by odpovídalo agresivnímu časovému harmonogramu vývoje UI). Jistě by se jednalo o neuvěřitelný humanitární triumf, vymýcení většiny neduhů, které lidstvo sužovaly po tisíciletí. Mnoho mých přátel a kolegů vychovává děti a doufám, že až tyto děti vyrostou, bude pro ně jakákoli zmínka o nemoci znít stejně archaicky jako pro nás kurděje, neštovice nebo mor. Tato generace bude také těžit ze zvýšené biologické svobody a možností sebevyjádření a se štěstím se bude moci dožívat takového věku, jaký si bude přát.

Je těžké docenit, jak překvapivé tyto změny budou pro všechny kromě malé skupiny lidí, kteří očekávali příchod mocné UI. Například tisíce ekonomů a politických expertů v USA v současné době debatují o tom, jak zajistit finanční stabilitu systémů sociálního zabezpečení a zdravotní péče (Medicare) a obecněji jak udržet nízké náklady na zdravotní péči (kterou z velké části spotřebovávají lidé starší 70 let, a obzvláště pak pacienti s terminálními onemocněními, jako je rakovina). Situace se u těchto programů pravděpodobně radikálně zlepší, pokud se tyto předpovědi naplní (20), jelikož se dramaticky změní poměr lidí v produktivním věku k důchodcům. Tyto výzvy sice nepochybně nahradí jiné, například jak zajistit široký přístup k novým technologiím, ale stojí za to se zamyslet nad tím, jak moc se svět změní, i kdyby se UI podařilo urychlit pouze oblast biologie.

2. Neurovědy a mysl

V předchozí části jsem se zaměřil na fyzická onemocnění a biologii obecně a nevěnoval jsem se neurovědám ani duševnímu zdraví. Neurovědy jsou ale podoborem biologie a duševní zdraví je stejně důležité jako to fyzické. V některých ohledech má duševní zdraví na lidský blahobyt dokonce ještě přímější vliv. Stovky milionů lidí trpí nízkou kvalitou života kvůli problémům jako závislost, deprese, schizofrenie, nízkofunkční autismus, posttraumatická stresová porucha, psychopatie (21) nebo mentální postižení. Miliardy dalších se potýkají s každodenními problémy, které lze interpretovat jako mírnější verze těchto závažných klinických poruch. Stejně jako v případě obecné biologie je ale možné jít nad rámec pouhého řešení problémů a zaměřit se na zlepšení základních parametrů lidské zkušenosti.

Základní principy, které jsem nastínil pro biologii, platí i pro neurovědy. Tento obor je poháněn malým počtem objevů, které se často týkají nástrojů pro měření nebo cílené intervence. V seznamu výše uvedené optogenetika byla objevem v neurovědách a novější pokroky, jako CLARITY a expanzní mikroskopie, jdou ve stejném duchu. Mnoho obecných metod buněčné biologie je navíc přímo aplikovatelných i v neurovědách. Domnívám se, že tempo těchto pokroků bude s pomocí UI podobně akcelеровáno, a proto se i na neurovědy vztahuje koncept „100 let pokroku za 5–10 let“, a to ze stejných důvodů jako v biologii. Stejně jako v biologii byl pokrok v neurovědách ve 20. století enormní – do 50. let jsme například ani nechápali, jak a

proč neurony vysílají signály. Zdá se proto rozumné očekávat, že neurovědy urychlené UI zaznamenají v příštích letech prudký rozvoj.

K tomuto základnímu obrazu je třeba dodat jednu věc: některé poznatky, které jsme v posledních letech získali o UI (nebo se je učíme), pravděpodobně pomohou i v neurovědách, a to i v případě, že se jimi budou nadále zabývat pouze lidé. Zřejmým příkladem je interpretovatelnost: Ačkoli biologické neurony fungují na první pohled zcela odlišně od umělých neuronů (komunikují prostřednictvím impulzů a frekvencí impulzů, takže zde hraje roli časový faktor, který u umělých neuronů chybí, a řada detailů týkajících se buněčné fyziologie a neurotransmiterů jejich fungování podstatně modifikuje), základní otázka, „jak distribuované, trénované sítě jednoduchých jednotek, které provádějí kombinované lineární/nelineární operace, spolupracují na provádění důležitých výpočtů,“ je stejná. Silně se domnívám, že detaily komunikace jednotlivých neuronů budou ve většině zajímavých otázek o výpočtech a obvodech abstrahovány. (22) Například výpočetní mechanismus objevený výzkumníky interpretovatelnosti v systémech UI byl nedávno znovuobjeven v mozcích myší.

Protože je mnohem snazší provádět experimenty na umělých neuronových sítích než na skutečných (u těch druhých je často nutné zasahovat do mozku zvířat), interpretovatelnost se může stát užitečným nástrojem pro pochopení fungování mozku. Navíc samotné vyspělé UI budou pravděpodobně schopné tento nástroj vyvíjet a používat efektivněji než lidé.

Kromě interpretovatelnosti by poznatky z UI o tom, jak trénovat inteligentní systémy, měly (i když si nejsem jistý, zda se tak už stalo) způsobit revoluci v neurovědách. Když jsem pracoval v neurovědách, mnoho lidí se zaměřovalo na otázky, které bych dnes považoval za chybné, protože koncept hypotézy škálování / hořké lekce ještě neexistoval. Myšlenka, že jednoduchá účelová funkce a velké množství dat mohou vést k neuvěřitelně komplexnímu chování, činí zajímavějším zkoumáním účelových funkcí a architektonických zkreslení a méně zajímavé je studium detailů emergentních výpočtů. V posledních letech jsem tento obor detailně nesledoval, ale mám dojem, že výpočetní neurovědci si toto ponaučení dosud plně neosvojili. Můj postoj k hypotéze škálování byl vždy „aha – tohle je vysvětlení, na obecné úrovni, jak funguje inteligence a jak se tak snadno vyvinula“, ale nemyslím si, že to je názor průměrného neurovědce, zčásti proto, že hypotéza škálování jako „klíč k inteligenci“ není plně uznávaná ani v rámci výzkumu UI.

Myslím si, že neurovědci by se měli snažit propojit tento základní poznatek se specifiky lidského mozku (biofyzikální omezení, evoluční historie, topologie, detaily motorických a senzorických vstupů/výstupů), aby se pokusili rozluštit některé klíčové záhady neurověd. Někteří to pravděpodobně dělají, ale mám podezření, že to nestačí a že UI neurovědci budou schopni tento úhel pohledu lépe využít k urychlení pokroku.

Očekávám, že UI urychlí pokrok v neurovědách čtyřmi různými cestami, které všechny mohou doufejme spolupracovat na vyléčení duševních nemocí a zlepšení kognitivních funkcí:

- **Tradiční molekulární biologie, chemie a genetika.** Toto je v podstatě stejný princip jako u obecné biologie v části 1 a UI ji může pravděpodobně urychlit stejnými mechanismy. Existuje mnoho léků, které modulují neurotransmitery, a tím mění mozkovou aktivitu, ovlivňují bdělost či vnímání, mění náladu atd. UI nám může pomoci objevit mnoho dalších. UI pravděpodobně také urychlí výzkum genetických příčin duševních onemocnění.
- **Detailní měření a ovlivňování neuronů.** Jde o schopnost měřit aktivitu mnoha jednotlivých neuronů nebo neuronových obvodů a zasahovat do ní s cílem změnit jejich chování. Optogenetika a neuronové sondy jsou technologie, které umožňují jak měření, tak i ovlivňování neuronů v živých organismech. Navržena byla i řada dalších, velmi pokročilých metod (jako jsou molekulární záznamové pásky pro čtení aktivačních vzorců velkého množství jednotlivých neuronů), které se zdají být principiálně možné.
- **Pokročilé výpočetní neurovědy.** Jak už jsem zmínil, specifické poznatky i celkový přístup moderní UI lze pravděpodobně s užítkem aplikovat na otázky systémových neurověd, včetně odhalování skutečných příčin a dynamiky komplexních onemocnění, jako jsou psychózy nebo poruchy nálady.
- **Behaviorální intervence.** Vzhledem k zaměření na biologickou stránku neurověd jsem se o behaviorálních intervencích příliš nezmínil, ale psychiatrie a psychologie samozřejmě vyvinuly širokou škálu behaviorálních intervencí během 20. století. Je logické, že UI by mohla jejich rozvoj a aplikaci urychlit, a to jak vývojem nových metod, tak i pomocí pacientům s dodržováním stávajících postupů. Obecněji řečeno, velmi slibně se jeví koncept „UI kouče“, který vám pomáhá dosáhnout vašeho plného potenciálu, studuje vaše interakce s okolím a pomáhá vám stát se efektivnějšími.

Můj odhad je, že tyto čtyři cesty pokroku, pokud budou fungovat synergicky, by mohly vést k vyléčení nebo prevenci většiny duševních nemocí v příštích 100 letech, a to i bez zapojení UI – a proto by s pomocí UI mohly být dosažitelné za 5–10 let.

Konkrétně si myslím, že se stane zhruba toto:

- **Většina duševních nemocí bude pravděpodobně vyléčitelná.** Nejsem expert na psychiatrická onemocnění (během svého působení v neurovědách jsem se zabýval konstrukcí sond pro studium malých skupin neuronů), ale odhaduji, že onemocnění jako posttraumatická stresová porucha, deprese, schizofrenie, závislosti atd. bude možné rozluštit a efektivně léčit pomocí výše zmíněných čtyř směrů. Řešení pravděpodobně bude kombinací „biochemické poruchy“ (i když možná velmi komplexní) a „poruchy v neuronové síti na obecné úrovni“. Jinými slovy, jedná se o otázku systémových neurověd – což ale nepopírá důležitost behaviorálních intervencí, o kterých jsem se zmínil výše. Nástroje

pro měření a intervence, obzvláště u živých lidí, pravděpodobně povedou k rychlému opakování pokusů a pokroku.

- **Strukturální onemocnění mozku budou pravděpodobně obtížněji léčitelná, ale ne nemožná.** Existují důkazy o tom, že psychopatie je spojena se zjevnými neuroanatomickými rozdíly – některé oblasti mozku jsou u psychopatů menší nebo méně vyvinuté. O psychopatech se také soudí, že jim od útlého věku chybí empatie; ať už je na jejich mozku cokoli odlišného, pravděpodobně to tak bylo vždy. Totéž může platit pro některé formy mentálního postižení a možná i pro další onemocnění. Přestavba mozku zní sice složitě, ale zároveň se zdá být úkolem s vysokou návratností investic do inteligence. Možná existuje způsob, jak stimulovat dospělý mozek k návratu do dřívějšího, plastičtějšího stavu, ve kterém by ho bylo možné přetvořit. Nejsem si jistý, nakolik je to reálné, ale můj instinkt mi velí být v tomto ohledu optimistický a věřit ve schopnosti UI.
- **Efektivní genetická prevence duševních nemocí se zdá být možná.** Většina duševních nemocí je alespoň zčásti dědičná a studie celogenomové asociace začínají identifikovat relevantní faktory, kterých bývá často mnoho. Pravděpodobně bude možné většině těchto onemocnění předcházet screeningem embryí, podobně jako u fyzických onemocnění. Rozdíl je v tom, že psychiatrická onemocnění mají spíše polygenní charakter (přispívá k nim mnoho genů), vzhledem ke složitosti celého systému roste riziko, že nevědomky potlačíme i pozitivní vlastnosti, které s nemocí korelují. Studie GWAS z posledních let ale kupodivu naznačují, že tyto korelace mohly být nadhodnoceny. UI nám v každém případě může pomoci tyto záležitosti objasnit. Screening embryí pro komplexní znaky samozřejmě vyvolává řadu etických a společenských otázek a bude kontroverzní, i když se domnívám, že většina lidí by podpořila screening pro závažná a oslabující duševní onemocnění.
- **Každodenní problémy, které nepovažujeme za onemocnění, budou také vyřešeny.** Většina z nás má každodenní psychologické problémy, které se obvykle nepovažují za klinickou diagnózu. Někdo se snadno rozčílí, jiný má problém se soustředěním nebo je často ospalý, další trpí strachem a úzkostí, nebo špatně reagují na změny. V dnešní době už existují léky, které pomáhají s bdělostí a soustředěním (kofein, modafinil, ritalin), ale stejně jako v jiných oblastech i zde UI pravděpodobně otevře cestu k mnohem lepším řešením. Pravděpodobně se podaří najít mnoho dalších takových léků, které ještě nebyly objeveny, a UI by také mohla vést k objevu zcela nových způsobů léčby, jako je cílená světelná stimulace (viz optogenetika výše), nebo magnetická pole. Vzhledem k tomu, kolik léků ovlivňujících kognitivní funkce a emoční stav bylo vyvinuto ve 20. století, jsem velmi optimistický ohledně „zkondenzovaného 21. století“, kde si každý bude moci optimalizovat fungování svého mozku a prožívat naplňující každodenní zkušenosti.

- **Základní lidská zkušenost se může výrazně zlepšit.** Mnoho lidí zažilo mimořádné stavy zjevení, kreativní inspirace, soucitu, naplnění, transcendence, lásky, krásy nebo meditativního klidu. Charakter a četnost těchto zážitků se u jednotlivých lidí i u téhož člověka v různých dobách velmi liší a někdy mohou být vyvolány i drogami (často ale s vedlejšími účinky). To naznačuje, že „prostor možných zážitků“ je velmi široký a že by se z těchto mimořádných momentů mohla skládat větší část lidských životů. Pravděpodobně je také možné zlepšit různé kognitivní funkce napříč celým spektrem. Dalo by se to nazvat neurovědeckou verzí „biologické svobody“ nebo „prodloužené délky života“.

Jedno téma, které se často objevuje ve sci-fi vizích UI, ale o kterém jsem se zde záměrně nezmínil, je „nahrávání mysli“ – myšlenka zachycení vzorců a dynamiky lidského mozku a jejich přenesení do softwaru. Tomuto tématu by se dala věnovat celá eseje, ale postačí říct, že ačkoli si myslím, že nahrání mysli je principiálně možné, v praxi čelí i s vyspělou UI značným technologickým a společenským výzvám, které ho pravděpodobně odsouvají za horizont 5–10 let, o kterém zde diskutuji.

Stručně řečeno, neurovědy urychlené UI pravděpodobně výrazně zlepší léčbu, nebo dokonce vyléčí většinu duševních nemocí a také výrazně rozšíří „kognitivní a mentální svobodu“ a kognitivní a emocionální schopnosti lidí. Bude to stejně radikální změna jako zlepšení fyzického zdraví popsáná v předchozí části. Svět se navenek možná viditelně nezmění, ale prožívání světa lidmi se výrazně zlepší. Bude to humánnější místo s většími příležitostmi k seberealizaci. Domnívám se také, že zlepšení duševního zdraví zmírní řadu dalších společenských problémů, včetně těch, které se zdají být politické, nebo ekonomické.

3. Ekonomický rozvoj a chudoba

Předchozí dvě části se zabývaly vývojem nových technologií pro léčbu nemocí a zlepšení kvality lidského života. Nabízí se ale zřejmá otázka z humanitárního hlediska: „Bude mít k těmto technologiím přístup každý?“

Jedna věc je vyvinout lék na nemoc, druhá věc je vymýtit nemoc z celého světa. Mnoho existujících zdravotnických intervencí se dosud neaplikuje celosvětově, a totéž platí pro technologický pokrok obecně (i mimo oblast zdraví). Životní úroveň v mnoha částech světa je stále zoufale nízká: HDP na obyvatele v subsaharské Africe je zhruba 2 000 dolarů, ve srovnání s 75 000 dolary ve Spojených státech. Pokud UI dále zvýší ekonomický růst a kvalitu života ve vyspělém světě, ale rozvojovým zemím nepomůže, měli bychom to považovat za obrovské morální selhání a černou skvrnu na jinak pozitivních humanitárních vítězstvích popsáných v předchozích dvou částech. V ideálním případě by vyspělá UI měla pomoci rozvojovému světu dohnat vyspělý svět a zároveň ho posunout na novou úroveň.

Nejsem si tak jistý, že UI dokáže vyřešit problémy nerovnosti a ekonomického růstu, jako si jsem jistý, že dokáže vyvinout přelomové technologie. Technologie mají zjevně vysoké výnosy z inteligence (včetně schopnosti obejít složitosti a nedostatky

dat), zatímco ekonomika je ovlivněna mnoha lidskými faktory a také velkou mírou vnitřní komplexity. Pochybuji, že by UI dokázala vyřešit známý „problém ekonomické kalkulace v socialismu“ (23) a nemyslím si, že by vlády měly (nebo chtěly) svěřit svou hospodářskou politiku takové entitě, i kdyby to dokázala. Existují také problémy, jako je přesvědčování lidí k léčbě, která je sice účinná, ale které nedůvěřují.

Výzvy, kterým rozvojový svět čelí, jsou ještě složitější kvůli všudypřítomné korupci jak v soukromém, tak ve veřejném sektoru. Korupce vytváří začarovaný kruh: zhoršuje chudobu a chudoba plodí další korupci. Plány ekonomického rozvoje poháněné UI se musí s korupcí, slabými institucemi a dalšími lidskými problémy vypořádat.

Přesto vidím i důvody k optimismu. Některé nemoci byly vymýceny a mnoho zemí se z chudoby dostalo k bohatství. Rozhodnutí, která k tomu vedla, vykazují vysoké výnosy z inteligence (navzdory lidským omezením a složitosti). UI by tedy tyto úkoly pravděpodobně zvládla lépe, než se daří v současnosti. Mohou také existovat cílené intervence, které obcházejí lidská omezení a na které by se UI mohla zaměřit. Je třeba jednat. Do řešení se musí zapojit jak firmy vyvíjející UI, tak politické špičky vyspělých zemí, aby rozvojový svět nezůstal pozadu. Morální imperativ je v tomto případě příliš silný. V této části tedy budu i nadále optimistický, ale mějte na paměti, že úspěch není zaručen a závisí na našem společném úsilí.

Níže uvádím několik předpovědí ohledně možného vývoje v rozvojovém světě v horizontu 5–10 let po vzniku mocné UI:

- **Distribuce zdravotnických intervencí.** V této oblasti jsem asi nejvíce optimistický. Nemoci se skutečně daří vymýtit pomocí centralizovaných kampaní: neštovice byly zcela vymýceny v 70. letech 20. století a dětská obrna a drakunkuliáza jsou téměř vymýceny, s méně než 100 případy ročně. Matematicky sofistikované epidemiologické modelování hraje v kampaních na vymýcení nemocí aktivní roli a zdá se velmi pravděpodobné, že systémy UI chytřejší než lidé by v tomto ohledu mohly dosáhnout lepších výsledků. Také logistika distribuce léčiv se pravděpodobně dá výrazně optimalizovat. Z mých zkušeností s dárcovstvím organizací GiveWell vím, že některé zdravotnické charity jsou mnohem efektivnější než jiné. Doufáme, že projekty urychlené UI budou ještě efektivnější. Některé biologické pokroky navíc logistiku distribuce usnadňují: například malárii bylo obtížné vymýtit, protože vyžaduje léčbu při každém novém nakažení. Vakcína, která stačí podat jen jednou, logistiku výrazně zjednodušuje (a takové vakcíny proti malárii se skutečně vyvíjejí). Můžeme využít i jednodušší metody distribuce: Některé nemoci by bylo možné vymýtit zacílením na jejich zvířecí přenašeče, například vypuštěním komárů infikovaných bakteriemi, které blokují přenos nemoci (a kteří pak nakazí ostatní komáry), nebo jednoduše pomocí genových pohonů k vyhubení komárů. To vyžaduje jeden, nebo několik centralizovaných zásahů, spíše než koordinovanou kampaň, která musí ošetřit miliony lidí individuálně. Celkově vzato si myslím, že 5–10 let je realistický časový rámec pro to, aby se alespoň

polovina zdravotních benefitů plynoucích z UI rozšířila i do nejchudších zemí. Dobrým cílem by mohlo být, aby se zdravotní stav rozvojového světa 5–10 let po vzniku vyspělé UI alespoň výrazně zlepšil oproti současnému stavu vyspělého světa, i kdyby za ním i nadále zaostával. Dosažení tohoto cíle si samozřejmě vyžádá obrovské úsilí v oblasti globálního zdraví, filantropie, politické advokacie a dalších oblastech, na kterých by se měli podílet jak vývojáři UI, tak i tvůrci politik.

- **Ekonomický růst.** Dokáže rozvojový svět rychle dohnat vyspělý svět nejen ve zdravotnictví, ale i v ekonomice jako celku? Známe podobné případy z minulosti: V posledních dekadách 20. století dosáhlo několik východoasijských ekonomik trvalého růstu reálného HDP okolo 10 % ročně, což jim umožnilo dohnat vyspělý svět. Lidští ekonomičtí plánovači se na tomto úspěchu podíleli nikoliv přímým řízením ekonomik, ale ovlivněním několika klíčových faktorů (například politikou exportně orientovaného růstu a odmítáním spoléhat se na bohatství z přírodních zdrojů). Je možné, že „ministři financí a guvernéři centrálních bank pohánění UI“ by mohli tohoto 10% růstu dosáhnout také, nebo ho dokonce překonat. Důležitou otázkou je, jak přesvědčit vlády rozvojových zemí, aby tyto technologie přijaly, a zároveň respektovat princip sebeurčení – některé vlády budou nadšené, jiné spíše skeptické. Z optimistického hlediska pravděpodobně mnoho zdravotnických intervencí zmíněných v předchozím bodě povede k organickému růstu ekonomiky: vymýcení AIDS, malárie a parazitických červů by mělo transformační dopad na produktivitu, nemluvě o ekonomických benefitech, které by plynuly z některých neurovědeckých intervencí (například zlepšení nálady a soustředění) jak ve vyspělém, tak v rozvojovém světě. Technologie urychlené UI v jiných než zdravotnických oblastech (jako je energetika, drony pro přepravu, vylepšené stavební materiály, lepší logistika a distribuce atd.) by mohly do světa proniknout přirozeně; například i mobilní telefony se v subsaharské Africe rychle rozšířily díky tržním mechanismům, bez nutnosti filantropického úsilí. Na druhou stranu, přestože UI a automatizace přináší řadu potenciálních výhod, představují také výzvy pro ekonomický rozvoj, zejména pro země, které ještě neprošly industrializací. Pro ekonomy a politiky je důležité najít způsoby, jak zajistit, aby se tyto země mohly i nadále rozvíjet a zlepšovat v éře rostoucí automatizace. V ideálním scénáři – a možná i cíli, o který bychom se měli snažit – by rozvojový svět dosahoval 20% ročního růstu HDP, přičemž 10 % by pocházelo z ekonomických rozhodnutí poháněných UI a 10 % z přirozeného šíření technologií urychlených UI (nejen ve zdravotnictví). Pokud by se to podařilo, subsaharská Afrika by za 5–10 let dosáhla současného HDP Číny na obyvatele a většina ostatních rozvojových zemí by překonala současné HDP USA. Ještě jednou připomínám, že se jedná o ideální scénář a ne o automatický vývoj. Je to něco, o co se musíme všichni společně snažit.

- **Potravinová bezpečnost.** (24) Pokrok v zemědělských technologiích, jako jsou lepší hnojiva a pesticidy, větší automatizace a efektivnější využití půdy, v 20. století dramaticky zvýšil výnosy plodin a zachránil miliony lidí před hladomorem. Genetické inženýrství v současnosti dále vylepšuje mnoho plodin. Nalezení dalších cest, jak toho dosáhnout – a jak zefektivnit zemědělské dodavatelské řetězce – by nám mohlo přinést druhou zelenou revoluci poháněnou UI, která by pomohla snížit rozdíly mezi rozvojovým a vyspělým světem.
- **Zmírnění klimatických změn.** Klimatické změny budou mít mnohem větší dopad na rozvojový svět a budou brzdit jeho rozvoj. Můžeme očekávat, že UI povede ke zlepšení technologií, které zpomalují, nebo zabraňují klimatickým změnám – od odstraňování uhlíku z atmosféry a technologií pro čistou energii až po laboratorně pěstované maso, které snižuje naši závislost na uhlíkově náročném velkochovu. Technologie samozřejmě není jediným faktorem, který brzdí boj proti klimatickým změnám – jak už jsem zmínil v této eseji, lidské a společenské faktory hrají důležitou roli. Máme ale dobrý důvod se domnívat, že výzkum s využitím UI nám dá nástroje k tomu, abychom zmírňování dopadů klimatických změn učinili mnohem levnějším a méně rušivým, čímž se mnoho námitek stane bezpředmětnými a rozvojové země budou moci dosáhnout většího ekonomického pokroku.
- **Nerovnost uvnitř zemí.** Mluvil jsem hlavně o nerovnosti jako o globálním jevu (který považuji za jeho nejvýznamnější projev), ale nerovnost samozřejmě najdeme i uvnitř jednotlivých zemí. S pokročilými zdravotnickými intervencemi, a zejména s radikálním prodloužením délky života či s léky na kognitivní vylepšení, se jistě objeví oprávněné obavy, že tyto technologie budou „jen pro bohaté“. Jsem ale optimističtější ohledně nerovnosti v rámci jednotlivých zemí, obzvláště ve vyspělém světě, a to ze dvou důvodů. Zaprvé, trhy ve vyspělém světě fungují lépe a trhy obvykle dokážou snižovat cenu vysoce hodnotných technologií v průběhu času. (25) Zadruhé, politické instituce vyspělého světa lépe reagují na potřeby svých občanů a mají větší kapacitu realizovat programy pro všeobecný přístup – a očekávám, že občané budou požadovat přístup k technologiím, které tak zásadně zlepšují kvalitu života. Není samozřejmě předem dané, že tyto požadavky uspějí – a zde je další oblast, kde se musíme všichni společně snažit o vytvoření spravedlivé společnosti. Samostatným problémem je také majetková nerovnost (oproti nerovnosti v přístupu k technologiím, které zachraňují a zlepšují životy), který se zdá být složitější a o kterém mluvím v části 5.
- **Problém odporu vůči změnám.** Jednou z obav, jak ve vyspělém, tak v rozvojovém světě, je odmítání benefitů, které UI přináší (podobně jako u hnutí proti očkování, nebo obecněji u luddistických hnutí). Mohlo by vzniknout negativní sebezposilující se chování. Například lidé, kteří jsou nejméně schopni se správně rozhodovat, by odmítli právě ty technologie, které by jim s rozhodováním pomohly, což by vedlo k prohlubování propasti a dokonce

k vytvoření dystopické společenské periferie (někteří vědci tvrdí, že by to podkopalo demokracii, což rozebírám v následující části). To by vrhlo stín na pozitivní přínosy UI. Je to složitý problém, protože si nemyslím, že je etické lidi k něčemu nutit, ale můžeme se alespoň snažit o zlepšení jejich vědeckého porozumění – a možná nám v tom pomůže i samotná UI. Pozitivním signálem je, že historicky se hnutí proti technologiím projevovала spíše slovy než činy. Kritika moderních technologií je sice populární, ale většina lidí je časem stejně přijme, alespoň když je to věc osobní volby. Lidé většinou přijímají zdravotnické a spotřební technologie, zatímco technologie, které skutečně narážejí na odpor, jako je jaderná energie, jsou spíše záležitostmi kolektivních politických rozhodnutí.

- Celkově jsem optimistický ohledně rychlého rozšíření biologických pokroků UI do rozvojového světa. Doufám, i když si nejsem jistý, že UI také umožní nebývalé tempo ekonomického růstu a rozvojovým zemím alespoň dohnat vyspělý svět. Znepokojuje mě problém s odmítáním technologií v obou zmíněných částech světa, ale předpokládám, že časem odezní a UI tento proces urychlí. Nebude to dokonalý svět a ti, kteří jsou pozadu, vyspělé země plně nedohoní, alespoň ne v prvních letech. S velkým úsilím z naší strany se nám ale může podařit nastartovat vývoj správným směrem – a rychle. Pokud se nám to povede, splatíme alespoň část dluhu v podobě důstojnosti a rovnosti, který dlužíme každému člověku na Zemi.

4. Mír a správa věcí veřejných

Předpokládejme, že vše v prvních třech částech dopadne dobře: nemoci, chudoba a nerovnost se výrazně sníží a základní úroveň lidské zkušenosti se podstatně zvýší. Z toho ale nevyplývá, že tím vyřešíme všechny hlavní příčiny lidského utrpení. Lidé pro sebe navzájem stále představují hrozbu. Přestože se objevuje trend, kdy technologický pokrok a ekonomický rozvoj vedou k demokracii a míru, je to velmi volný trend s častými (a nedávnými) odchylkami. Na začátku 20. století si lidé mysleli, že války mají za sebou; pak přišly dvě světové války. Před třiceti lety psal Francis Fukuyama o „Konci dějin“ a konečném vítězství liberální demokracie; to se nestalo. Před dvaceti lety tvůrci politik v USA věřili, že volný obchod s Čínou povede k její liberalizaci s tím, jak poroste její bohatství; to se nestalo a nyní zřejmě směřujeme k nové studené válce se silícím autoritářským blokem. Vznikly i přesvědčivé teorie, které naznačují, že internetové technologie mohou ve skutečnosti nahrávat spíše autoritářství než demokracii, jak se původně předpokládalo (např. v období „arabského jara“). Zdá se důležité pochopit, jak vyspělá UI ovlivní tyto otázky míru, demokracie a svobody.

Bohužel nevidím žádný pádný důvod se domnívat, že UI bude přednostně, nebo strukturálně podporovat demokracii a mír, tak jako si myslím, že podpoří lidské zdraví a zmírní chudobu. Lidské konflikty jsou konfrontační a UI může v principu pomoci jak „kladům“, tak „záporákům“. Některé strukturální faktory jsou dokonce znepokojivé: UI nejspíš umožní daleko účinnější propagandu a dohled nad obyvatelstvem, což

jsou dva hlavní nástroje v rukou autokratů. Je proto na nás, jednotlivých aktérech, abychom vývoj nasměrovali správným směrem: Pokud chceme, aby UI podporovala demokracii a lidská práva, musíme za to bojovat. Cítím to ještě silněji než v případě mezinárodní nerovnosti: Vítězství liberální demokracie a politické stability není zaručeno, a možná ani pravděpodobné. Bude vyžadovat velké oběti a úsilí z naší strany, jako tomu bylo v minulosti.

U tohoto problému vnímám dvě části: mezinárodní konflikty a vnitřní uspořádání států. V mezinárodním měřítku je velmi důležité, aby demokracie měly při vzniku vyspělé UI navrch. Představa autoritářství posíleného UI je příliš děsivá, než abychom o ní uvažovali. Demokracie proto musí být schopny nastavit podmínky pro zavedení vyspělé UI do světa tak, aby předešly porážce od autoritářských režimů a zároveň předešly porušování lidských práv v autoritářských zemích.

Můj současný odhad nejlepšího postupu je strategie „dohody“, (26) kde se koalice demokratických států snaží získat jasnou výhodu (i kdyby jen dočasnou) v oblasti vyspělé UI zajištěním svého dodavatelského řetězce, rychlým rozvojem a blokováním či zpomalením přístupu protivníků ke klíčovým zdrojům, jako jsou čipy a polovodičové komponenty. Tato koalice by na jedné straně využila UI k dosažení výrazné vojenské převahy („bič“), a zároveň by nabízela sdílení benefitů vyspělé UI („cukr“) se stále širším okruhem zemí výměnou za podporu strategie koalice na podporu demokracie (obdoba iniciativy „Atomy pro mír“). Cílem koalice by bylo získat podporu co největšího počtu zemí, izolovat naše nejhorší nepřátele a nakonec je dostat do situace, kdy pro ně bude výhodnější přijmout stejnou dohodu jako zbytek světa: vzdát se soupeření s demokraciemi výměnou za přístup ke všem benefitům a vyhnout se tak boji s nadřazeným protivníkem.

Pokud se nám tohle všechno podaří, budeme žít ve světě, kde budou demokracie udávat směr na globální scéně a budou mít ekonomickou a vojenskou sílu k tomu, aby se ubránily podkopávání, dobývání či sabotážím ze strany autokracií. Svou převahu v oblasti UI by pak mohly proměnit v trvalou výhodu. V optimistickém scénáři by to mohlo vést k trvalému stavu podobnému roku 1991 – světu, ve kterém mají demokracie navrch a Fukuyamovy sny se naplňují. Opakuji, že toho bude velmi těžké dosáhnout. Bude to vyžadovat úzkou spolupráci mezi soukromými společnostmi vyvíjejícími UI a demokratickými vládami a také mimořádně moudrá rozhodnutí ohledně rovnováhy mezi „cukrem“ a „bičem“.

I když se tohle všechno podaří, zůstává otázkou boj mezi demokracií a autokracií v rámci jednotlivých zemí. Budoucí vývoj lze samozřejmě těžko odhadovat, ale jsem optimista a věřím, že v globálním prostředí, kde demokracie ovládají nejmocnější UI, může UI strukturálně nahrávat demokracii po celém světě. V takovém prostředí mohou demokratické vlády využít svou převahu v UI k vítězství v informační válce: Mohou čelit operacím, které se snaží ovlivňovat veřejné mínění a šířit propagandu ze strany autokracií, a mohou dokonce vytvořit globálně svobodné informační prostředí tím, že budou poskytovat informační kanály a služby UI způsobem, který autokracie z technických důvodů nedokážou blokovat ani monitorovat. Pravděpodobně nebude

nutné šířit vlastní propagandu, stačí jen bránit se útokům a zajistit volný tok informací. Ačkoli se efekt nedostaví okamžitě, takové rovné podmínky mají dobrou šanci postupně naklonit misky vah globální politiky ve prospěch demokracie, a to z několika důvodů.

Zaprvé, zvýšení kvality života popsané v částech 1–3 by mělo za jinak stejných podmínek vést k posílení demokracie. Historicky tomu tak alespoň do určité míry bylo. Očekávám, že zlepšení duševního zdraví, blahobytu a vzdělání posílí demokracii, protože všechny tyto tři faktory negativně korelují s podporou autoritářských vůdců. Obecně platí, že když jsou základní potřeby lidí uspokojeny, touží po větším sebevyjádření a demokracie je mimo jiné i jeho formou. Naopak autoritářství těží ze strachu a nespokojenosti.

Zadruhé, je velmi pravděpodobné, že svobodný přístup k informacím skutečně oslabuje autoritářské režimy, pokud je autoritáři nemohou cenzurovat. A necenzurovaná UI může jednotlivcům poskytnout mocné nástroje k podkopávání represivních režimů. Represivní vlády si udržují moc tím, že brání lidem v přístupu k informacím a znemožňují jim uvědomit si, že „císař je nahý“. Srdja Popović, který se podílel na svržení Miloševićova režimu v Srbsku, sepsal řadu publikací o technikách, jak psychologicky zbavit autoritáře moci, prolomit jejich kouzlo a shromáždit podporu proti diktátorovi. Nadlidsky efektivní UI verze Popoviće (jehož dovednosti, jak se zdá, mají vysokou návratnost z hlediska inteligence) v kapse každého, neblokovaná a necenzurovaná diktátory, by mohla dodat vítr do plachet disidentům a reformátorům po celém světě. Zopakuji – bude to dlouhý a vleklý boj a vítězství není jisté, ale pokud UI navrhne a vytvoříme správným způsobem, může to být alespoň boj, ve kterém budou mít zastánci svobody všude výhodu.

Podobně jako u neurověd a biologie se můžeme ptát, jak by se věci mohly vyvíjet lépe než za běžných okolností – nejen jak se vyhnout autokracii, ale jak zlepšit demokracie oproti jejich současnému stavu. I v demokraciích dochází k bezpráví. Společnosti založené na právním státu slibují svým občanům rovnost před zákonem a právo na základní lidská práva, ale tyto sliby se v praxi ne vždy naplňují. Už jen to, že jsou alespoň částečně plněny, je důvod k hrdosti, ale může nám UI pomoci dosáhnout lepších výsledků?

Například, mohla by UI přispět k vylepšení našeho právního a soudního systému zavedením nestrannějšího rozhodování a spravedlivějších procesů? Dnes se lidé v právním a soudním kontextu obávají hlavně toho, že systémy UI povedou k diskriminaci, a tyto obavy jsou důležité a je třeba se jimi zabývat. Zároveň ale vitalita demokracie závisí na využití nových technologií ke zdokonalení demokratických institucí, nejen na reakci na rizika. Skutečně vyspělá a úspěšná implementace UI má potenciál snížit míru zaujatosti a být spravedlivější pro všechny.

Po staletí se právní systémy potýkaly s dilematem, že zákon sice usiluje o nestrannost, ale je ze své podstaty subjektivní, a proto musí být interpretován lidmi s vlastními předsudky. Snaha o plně mechanický zákon selhala, protože realita je komplexní a nelze ji vždy popsat matematickými vzorci. Právní systémy se proto

spoléhají na notoricky nepřesná kritéria, jako je „krutý a neobvyklý trest“, nebo „zcela postrádající vykupitelskou společenskou hodnotu“, která pak interpretují lidé – a často tak činí způsobem, který je zaujatý, protežující, nebo svévolný. „Chytré smlouvy“ v kryptoměnách nezpůsobily revoluci v právu, protože běžný kód není dostatečně inteligentní na to, aby dokázal rozsoudit většinu sporných bodů. UI by na to ale mohla stačit. Je to první technologie, která je schopna dělat komplexní a nepřesné úsudky opakovaně a mechanicky.

Nenavrhuji nahradit soudce systémy UI, ale spojení nestrannosti se schopností chápat a vyhodnocovat složité situace z reálného světa by mohla mít významně pozitivní dopad na právo a spravedlnost. Minimálně by takové systémy mohly fungovat po boku lidí jako pomůcka při rozhodování. V každém takovém systému by byla klíčová transparentnost a vyspělá věda o UI by ji mohla zajistit: Proces trénování takových systémů by mohl být důkladně prozkoumán a pokročilé techniky interpretovatelnosti by nám mohly umožnit nahlédnout do finálního modelu a posoudit ho z hlediska skrytých předsudků, což u lidí jednoduše není možné. Nástroje UI by se také daly využít k monitorování porušování základních práv v soudním a policejním kontextu, a ústavy by se tak staly samovymáhacími.

Obdobně by UI mohla sloužit k agregaci názorů a budování konsenzu mezi občany, k řešení konfliktů, hledání společné řeči a kompromisů. Některé rané projekty v tomto směru už probíhají v rámci iniciativy *computational democracy*, včetně spolupráce s Anthropic. Informovanější a uvážlivější občané by samozřejmě posílili demokratické instituce.

UI by také mohla zjevně pomoci s poskytováním vládních služeb – jako jsou zdravotní pojištění nebo sociální služby – které jsou sice principiálně dostupné všem, ale v praxi často chybí, a v některých místech více než v jiných. Jde například o zdravotní péči, služby ministerstev a úřadů, daně, sociální zabezpečení, stavební úřady atd. Mít k dispozici promyšlenou a informovanou UI, jejímž úkolem by bylo poskytovat vám vše, na co máte od státu ze zákona nárok, a to srozumitelným způsobem, a zároveň vám pomáhat s dodržováním často matoucích pravidel, by bylo velkým přínosem. Zvýšení kapacity státu pomůže naplnit slib rovnosti před zákonem a posílí respekt k demokratickému vládnutí. Špatně fungující služby jsou v současnosti hlavním zdrojem cynismu vůči vládě. (27)

Všechny tyto nápady jsou poněkud vágní a jak jsem řekl na začátku této části, nejsem si zdaleka tak jistý jejich proveditelností, jako v případě pokroků v biologii, neurovědách a zmírňování chudoby. Mohou být nereálně utopické. Důležité ale je mít ambiciózní vizi, nebát se velkých snů a zkoušet nové věci. Vize UI jako garanta svobody, individuálních práv a rovnosti před zákonem je příliš lákavá na to, abychom za ni nebojovali. Státní zřízení 21. století, obohacené o UI, by se mohlo stát silnějším ochráncem individuální svobody a zároveň majákem naděje, který by pomohl šířit liberální demokracii jako formu vlády, kterou si přeje přijmout celý svět.

5. Práce a smysl života

I kdyby všechno v předchozích čtyřech částech dopadlo dobře – nejenže bychom zmírnili nemoci, chudobu a nerovnost, ale liberální demokracie by se stala dominantní formou vlády a stávající demokratické státy by se staly lepšími verzemi sebe samých – přesto by zůstala alespoň jedna důležitá otázka. Někdo by mohl namítnout: „Je skvělé, že žijeme v tak technologicky vyspělém a zároveň spravedlivém a slušném světě, ale když UI bude dělat všechno, jaký smysl pak bude mít lidský život? A jak budou lidé ekonomicky přežívat?“

Myslím si, že tato otázka je složitější než ty předchozí. Neznamená to, že bych byl nutně pesimističtější (i když vidím výzvy). Spíše je tato otázka mlhavější a hůře se na ni hledají odpovědi předem, protože se týká makroskopických otázek o uspořádání společnosti, které se obvykle řeší samy časem a decentralizovaným způsobem. Například historické lovecko-sběračské společnosti si asi jen těžko dokázaly představit, že život bez lovu a různých s ním spojených náboženských rituálů může mít smysl, a naši blahobytnou technologickou společnost by vnímaly jako bezúčelnou. Pravděpodobně by také nechápaly, jak může naše ekonomika zajistit blahobyt pro všechny, a jakou roli v ní můžou hrát lidé v mechanizované společnosti.

Přesto stojí za to se k ní alespoň krátce vyjádřit, ale mějte, prosím, na paměti, že stručnost této části není známkou toho, že bych tyto otázky bral na lehkou váhu – naopak, je to známka toho, že na ně neexistují jasné odpovědi.

Pokud jde o smysl života, myslím si, že je chybou domnívat se, že činnosti, kterým se věnujete, ztrácejí smysl jen proto, že by je UI mohla dělat lépe. Většina lidí není v ničem nejlepší na světě a zdá se, že jim to nijak zvlášť nevadí. Dnes samozřejmě můžou stále přispívat do společnosti skrze komparativní výhodu a smysl života můžou nacházet v ekonomické hodnotě, kterou vytvářejí, ale lidé si také užívají činnosti, které nemají žádnou ekonomickou hodnotu. Trávím spoustu času hraním videoher, plaváním, procházkami v přírodě a rozhovory s přáteli, z čehož nic nemá ekonomickou hodnotu. Můžu strávit den snahou zlepšit se ve videohře, nebo zdolat horu na kole rychleji, a přitom mi vůbec nevadí, že je někdo v těchto činnostech mnohem lepší než já. V každém případě si myslím, že smysl života pramení především z lidských vztahů a spojení s ostatními, a ne z ekonomické práce. Lidé touží po pocitu zadostiučinění a po soutěžení, a i ve světě s UI bude možné věnovat roky úsilí náročným úkolům a komplexním strategiím, podobně jako to dnes dělají lidé ve výzkumu, herectví, nebo podnikání. (28) To, že (a) někde existuje UI, která by daný úkol zvládla lépe, a (b) tento úkol už není ekonomicky ohodnocen v rámci globální ekonomiky, mi nepřijde nijak zvlášť důležité.

Ekonomický aspekt se mi zdá být palčivější než otázka smyslu života. „Ekonomickým aspektem“ v této části myslím potenciální problém, že většina, nebo dokonce všichni lidé nebudou schopni smysluplně přispívat k dostatečně rozvinuté ekonomice poháněné UI. To je mnohem širší problém než otázka nerovnosti, obzvlášť nerovnosti v přístupu k novým technologiím, o které jsem hovořil v části 3.

V krátkodobém horizontu souhlasím s argumenty, že díky komparativní výhodě zůstanou lidé i nadále relevantní a jejich produktivita se dokonce zvýší a v některých ohledech se i srovnají podmínky mezi lidmi. Dokud bude UI lepší jen v 90 % dané práce, zbývajících 10 % povede k tomu, že lidé budou moci lépe využít svůj potenciál a zvýší se i jejich příjmy. V praxi tak vznikne mnoho nových pracovních míst, která budou doplňovat a rozvíjet to, v čem je UI dobrá, takže oněch „10 %“ se rozšíří a bude i nadále zaměstnávat téměř každého. I kdyby UI dokázala dělat 100 % věcí lépe než lidé, ale u některých úkolů by byla neefektivní, nebo drahá, případně by se zdroje nutné pro fungování lidí a UI lišily, logika komparativní výhody by i nadále platila. Oblastí, kde si lidé pravděpodobně udrží relativní (nebo i absolutní) výhodu po delší dobu, je fyzický svět. Proto si myslím, že lidská ekonomika může dávat smysl i po tom, co dosáhneme úrovně „státu géníů v datovém centru“.

Domnívám se však, že z dlouhodobého hlediska se UI stane natolik efektivní a levnou, že tato logika přestane platit. V tom okamžiku naše současné ekonomické uspořádání ztratí smysl a bude nutné vést širší společenskou diskusi o tom, jak by měla být ekonomika organizována.

I když to může znít bláznivě, civilizace v minulosti úspěšně prošla řadou velkých ekonomických změn: od lovu a sběru k zemědělství, od zemědělství k feudalismu a od feudalismu k industrialismu. Tuším, že bude potřeba něco nového a neobvyklého, co si dnes nikdo nedokáže ani představit. Možná půjde o něco tak jednoduchého, jako je vysoký univerzální základní příjem pro všechny, i když se domnívám, že to bude jen část řešení. Možná vznikne kapitalistická ekonomika systémů UI, které budou rozdělovat zdroje (v obrovském množství, protože ekonomický koláč bude gigantický) lidem na základě nějaké sekundární ekonomiky, která bude odměňovat to, co systémy UI považují za smysluplné u lidí (na základě úsudku odvozeného z lidských hodnot). Možná bude ekonomika fungovat na základě bodů Whuffie. Nebo možná budou mít lidé i nadále ekonomickou hodnotu, a to způsobem, který si dnes nedovedeme představit. Všechna tato řešení mají řadu potenciálních problémů a bez experimentování a opakování pokusů nelze s jistotou říct, zda budou fungovat. A stejně jako v případě jiných výzev budeme muset o pozitivní výsledek bojovat. Vykořisťovatelské a dystopické scénáře jsou reálné a je třeba jim zabránit. O těchto otázkách by se dalo napsat mnohem více a doufám, že se k nim v budoucnu vrátím.

Shrnutí

V této eseji jsem se pokusil představit vizi světa, která je jednak možná, pokud se vývoj UI bude ubírat pozitivním směrem, a jednak mnohem lepší než ten současný. Nevím, jestli je tento svět realistický, ale i kdyby byl, jeho dosažení si vyžádá obrovské úsilí a úsilí mnoha statečných a oddaných lidí. Všichni (včetně firem zabývajících se UI!) se budou muset zapojit, jak do prevence rizik, tak i do plného využití jejich benefitů.

Je to ale svět, za který stojí za to bojovat. Pokud se během příštích 5-10 let podaří vymýtit většinu nemocí, zvýšit biologickou a kognitivní svobodu, vymanit miliardy lidí

z chudoby a zapojit je do nového světa technologií, liberální demokracie a lidských práv - věřím, že to bude pro všechny, kdo to budou sledovat, zcela ohromující zážitek. Nemám na mysli pouze osobní prospěch z nových technologií, i když ten bude jistě obrovský. Spíše mám na mysli prožitek z toho, jak se naplňují naše ideály. Myslím, že to může mnohým z nás vehnat slzy do očí.

Při psaní této eseje jsem si všiml zajímavého rozporu. V jistém smyslu je prezentovaná vize extrémně radikální - neodpovídá očekáváním téměř nikoho ohledně toho, co se stane v příštím desetiletí, a pro mnohé bude znít jako absurdní fantazie. Někteří ji možná ani nebudou považovat za žádoucí; ztělesňuje hodnoty a politické postoje, se kterými se ne všichni ztotožní. Zároveň je v ní ale cosi neoddiskutovatelně zřejmého – jakási nevyhnutelnost – jako by mnoho různých pokusů o představu ideálního světa nevyhnutelně vedlo zhruba sem.

V knize Iaina M. Bankse *Hráč her* (29) protagonista – člen společnosti zvané Kultura, která je založena na principech velmi podobných těm, které jsem zde nastínil – cestuje do represivní, militaristické říše, kde je vůdcovství určováno soutěží ve složité válečné hře. Hra je ale natolik komplexní, že strategie hráče odráží jeho politické a filozofické přesvědčení. Protagonistovi se podaří porazit císaře ve hře a ukázat tak, že jeho hodnoty (hodnoty Kultury) představují vítěznou strategii i ve hře vytvořené společností založené na nelítostné soutěži a přežití nejschopnějších. Známý článek Scotta Alexandera prezentuje stejnou tezi – že soutěž je kontraproduktivní a vede ke společnosti založené na soucitu a spolupráci. „Oblouk morálního vesmíru“ je podobný koncept.

Myslím si, že hodnoty Kultury představují vítěznou strategii, protože jsou souhrnem milionu malých rozhodnutí, která mají jasnou morální sílu a která všechny spojují. Základní lidské intuice ohledně spravedlnosti, spolupráce, zvědavosti a autonomie se těžko zpochybňují a jsou kumulativní, na rozdíl od našich destruktivnějších impulzů. Je snadné argumentovat, že děti by neměly umírat na nemoci, pokud tomu můžeme zabránit, a stejně tak je snadné argumentovat, že na toto právo mají nárok všechny děti stejnou měrou. Odtud už je jen krok k myšlence, že bychom se měli spojit a využít naši inteligenci k tomu, abychom takového výsledku dosáhli. Jen málokdo by nesouhlasil s tím, že lidé by měli být potrestáni za útoky a zbytečné ubližování druhým, a odtud už není daleko k myšlence, že tresty by měly být konzistentní a systematické. Stejně tak intuitivně chápeme, že by lidé měli mít autonomii a nést zodpovědnost za svůj život a rozhodnutí. Tyto jednoduché intuice, pokud je dovedeme do důsledku, vedou k právnímu státu, demokracii a hodnotám osvícenství. I když nevyhnutelně, tak alespoň statisticky vzato lidstvo tímto směrem míří. U nás jednoduše nabízí příležitost, jak se tam dostat rychleji – aby logika byla zřetelnější a cíl jasnější.

Je to něco nesmírně krásného. Máme příležitost se na tom alespoň malou měrou podílet.

Poděkování patří Kevinu Esveltovi, Paragovi Mallickovi, Stuartovi Ritchiemu, Mattu Yglesiasovi, Eriku Brynjolfssonovi, Jimovi McClavovi, Allanu Dafoeovi a mnoha dalším lidem z Anthropic za zpětnou vazbu při psaní této eseje.

Vítězům Nobelovy ceny za chemii v roce 2024 za to, že nám ukázali cestu.

Poznámky pod čarou

1. <https://allpoetry.com/All-Watched-Over-By-Machines-Of-Loving-Grace>
2. Jsem si vědom, že někdo může tento přístup považovat za příliš opatrný. Těmto kritikům bych však doporučil zachovat střízlivý pohled na realitu. Co je podstatnější - umírněnost ve společenských změnách má své opodstatnění. Společnost dokáže vstřebat jen určité množství proměn v daném časovém úseku. A tempo, které navrhuji, se pravděpodobně již blíží hranici toho, co můžeme zvládnout, aniž by došlo k významným společenským otřesům.
3. Termín AGI (obecná umělá inteligence) mi přijde nepřesný a zavádějící, s nádechem sci-fi. Preferuji „vyspělá UI“ nebo „UI na expertní úrovni v oblasti vědy a inženýrství“, což lépe vystihuje, co tím myslím. (*V originále „powerful AI“*).
4. V této esejí používám pojem 'intelligence' jako zastřešující termín pro obecnou schopnost řešit problémy napříč různými doménami. Tato schopnost zahrnuje komplexní kognitivní procesy jako uvažování, učení, plánování a kreativitu. Jsem si však vědom, že tento zjednodušující přístup má svá úskalí - otázka povahy intelligence totiž představuje složité a intenzivně diskutované téma jak v kognitivní vědě, tak ve výzkumu umělé inteligence. Zatímco někteří vědci zastávají názor, že intelligence není jednotným konceptem, ale spíše souborem distinktivních kognitivních schopností, jiní argumentují ve prospěch existence obecného faktoru intelligence (g-faktoru), který tvoří základ různorodých kognitivních dovedností. Tato fascinující debata si však zaslouží samostatnou pozornost.
5. To je zhruba současná rychlost systémů UI – například dokážou přečíst stránku textu za několik sekund a napsat stránku textu za zhruba 20 sekund, což je 10–100krát rychleji než u lidí. Větší modely sice pracují pomaleji, ale výkonnější čipy tento efekt kompenzují; k dnešnímu dni se tyto dva vlivy zhruba vyrovnaly.
6. Ačkoli může tento argument působit až schematicky, zaslouží si pozornost - renomovaní myslitelé jako Tyler Cowen a Matt Yglesias ho předkládají jako závažný problém hodný diskuze, přestože se s ním plně neztotožňují. Na základě toho se domnívám, že tento pohled nelze jednoduše odmítnout jako nerealistický.

7. Nejrelevantnější ekonomický výzkum v této oblasti se zaměřuje na technologie obecného určení a s nimi související nemotné investice, které tyto technologie doplňují a podporují.
8. Toto učení se může zahrnovat dočasné učení v kontextu, nebo tradiční trénování; obojí bude limitováno fyzickým světem.
9. V chaotickém systému se malé chyby sčítají exponenciálně, takže i enormní zvýšení výpočetního výkonu vede jen k malému zlepšení v predikci budoucího vývoje a v praxi může chyba měření výsledek ještě zhoršit.
10. Dalším faktorem je samozřejmě i to, že vyspělá UI může být potenciálně použita k vytvoření ještě vyspělejší UI. Vycházím z předpokladu, že k tomu může (a pravděpodobně i dojde), ale že efekt bude menší, než by se mohlo zdát, a to právě kvůli „klesajícím mezním výnosům z inteligence“, o kterých zde mluvím. Jinými slovy, UI se bude i nadále rychle zdokonalovat, ale její vliv bude nakonec omezen faktory, které s inteligencí nesouvisí, a právě analýza těchto faktorů je klíčová pro pochopení rychlosti vědeckého pokroku mimo oblast UI.
11. Tyto úspěchy jsou pro mě inspirací a možná i nejlepším existujícím příkladem toho, jak lze UI využít k transformaci biologie.
12. „Pokrok ve vědě závisí na nových technikách, nových objevech a nových náparech, pravděpodobně v tomto pořadí.“ – Sydney Brenner
13. Děkuji Paragovi Mallickovi za upozornění na tento bod.
14. Nechtěl jsem zahlcovat text spekulacemi o konkrétních budoucích objevech, které by UI mohla umožnit, ale zde je brainstorming několika možností:
 - * Návrh lepších výpočetních nástrojů, jako jsou AlphaFold a AlphaProteo – tj. obecný systém UI urychlující vývoj specializovaných výpočetních nástrojů UI pro biologii.
 - * Efektivnější a selektivnější CRISPR.
 - * Pokročilejší buněčné terapie.
 - * Průlomové objevy v materiálové vědě a miniaturizaci vedoucí k lepším implantovaným zařízením.
 - * Lepší kontrola kmenových buněk, buněčné diferenciaci a dediferenciaci a výsledná schopnost regenerovat nebo přetvářet tkáň.
 - * Lepší kontrola imunitního systému: jeho selektivní aktivace k léčbě rakoviny a infekčních nemocí a jeho selektivní deaktivace k léčbě autoimunitních onemocnění.
15. UI samozřejmě může také pomoci s chytřejším výběrem experimentů: s vylepšením designu experimentů, lepším zužitkováním informací z prvního kola experimentů pro druhé kolo atd.
16. Děkuji Matthewovi Yglesiasovi za upozornění na tento bod.
17. Rychle se vyvíjející nemoci, jako jsou kmeny s mnohočetnou lékovou rezistencí, které v podstatě využívají nemocnice jako evoluční laboratoř k

- neustálému zlepšování své odolnosti vůči léčbě, by mohly být obzvláště problematické a mohly by nám zabránit v dosažení 100% úspěšnosti.
18. Je třeba poznamenat, že v horizontu 5–10 let nemusí být jasné, zda se nám podařilo zdvojnásobit délku lidského života. I když bychom toho dosáhli, nemusíme si toho v daném časovém rámci ještě všimnout.
19. V tomto bodě jsem ochoten, navzdory zjevným biologickým rozdílům mezi léčením nemocí a zpomalením samotného procesu stárnutí, se podívat na statistický trend z větší perspektivy a říci: „I když se detaily liší, myslím si, že lidská věda by pravděpodobně našla způsob, jak v tomto trendu pokračovat; koneckonců, plynulé trendy ve složitosti jsou nutně tvořeny kombinací velmi různorodých faktorů.“
20. Například mi bylo řečeno, že zvýšení ročního růstu produktivity o 1 %, nebo dokonce o 0,5 %, by mělo transformační dopad na prognózy těchto programů. Pokud by se nápady prezentované v této eseji uskutečnily, zisky produktivity by mohly být mnohem větší.
21. Média ráda vykresluje psychopaty s vysokým společenským postavením, ale průměrný psychopat je spíše člověk se špatnými ekonomickými vyhlídkami a slabou sebekontrolou, který stráví značnou část života ve vězení.
22. Myslím si, že je to do jisté míry analogické s tím, že mnoho (i když pravděpodobně ne všechny) výsledků, které získáváme z výzkumu interpretovatelnosti, by zůstalo relevantních, i kdyby se některé architektonické detaily našich současných umělých neuronových sítí, jako je mechanismus pozornosti, nějak změnily, nebo byly nahrazeny.
23. Mám podezření, že je to trochu jako klasický chaotický systém – plný neredukovatelné složitosti, kterou lze zvládat převážně decentralizovaným způsobem. Jak ale uvádím dále v této části, možné jsou i skromnější intervence. Ekonom Erik Brynjolfsson mi oponoval argumentem, že velké firmy (jako Walmart, nebo Uber) začínají mít dostatek centralizovaných znalostí k tomu, aby rozuměly spotřebitelům lépe než jakýkoli decentralizovaný proces, což by nás mohlo donutit přehodnotit Hayekovy názory na to, kdo má nejlepší lokální znalosti.
24. Poděkování Kevinu Esveltovi za návrh tohoto bodu.
25. Například mobilní telefony byly zpočátku technologií pro bohaté, ale rychle zlevnily díky meziročnímu vylepšování, které probíhalo tak rychle, že smazalo jakoukoli výhodu nákupu „luxusního“ telefonu a dnes má většina lidí telefony podobné kvality.
26. Toto je název připravovaného článku od RAND, který zhruba popisuje strategii, kterou zde navrhuji.
27. Když se průměrný člověk zamyslí nad veřejnými institucemi, pravděpodobně si vybaví své zkušenosti s úřady, finančním úřadem, zdravotní pojišťovnou a podobnými institucemi. Zlepšení těchto zkušeností by mohlo být účinným způsobem, jak bojovat s neopodstatněným cynismem.
28. Ve světě poháněném UI bude spektrum takových výzev a projektů mnohem širší, než je dnes.

29. Porušuji své vlastní pravidlo nezmiňovat sci-fi, ale je pro mě těžké se jí alespoň trochu vyhnout. Pravdou je, že sci-fi je jedním z mála zdrojů myšlenkových experimentů o budoucnosti. Myslím, že je škoda, že je tak úzce spjata s určitou subkulturou.