



KLIMATICKÁ ZMENA  
GLOBÁLNE OTEPLOVANIE  
SKLENÍKOVÝ EFEKT  
ZLÝ STAV POĽNOHOSPODÁRSKEJ PÓDY  
ÚBYTOK BIOLOGICKÝCH DRUHOV  
ROZVRÁTENIE EKOSYSTÉMOV  
MIZNUTIE LEŠOV  
ZRÚTENIE EKOSYSTÉMOV  
EKOLOGICKÝ ŽIAĽ

Rudolf Sikora

Z cyklu Eko(ko)mix. 2016 - 2017

# (Z)mení umelá inteligencia našu budúcnosť?

**Martin Takáč**

(písané pre *Kritiku & Kontext*)

Väčšina ľudí donedávna poznala umelú inteligenciu (UI) najmä z katastrofických sci-fi filmov. Málokto si prepojí klišé predstavu robotov ničiacich ľudstvo s bežnou každodennou skúsenosťou inteligentného vyhľadávania na webe, automatického prekladu medzi jazykmi, GPS navigáciou, plánovaním dopravného spojenia, komunikáciou v prirodzenom jazyku so smartfónom či použitím inteligentného vysávača. V posledných rokoch nastal prudký - exponenciálny - nárast nielen počtu publikovaných vedeckých článkov o UI a počtu študentov zapísaných na UI odbory na univerzitách, ale aj počtu UI startupov a objemu peňazí investovaných do nich. Stále viac sa však ozývajú aj varovné hlasy, tentoraz nie od tvorcov sci-fi, ale od rešpektovaných vedcov ako Stephen Hawking alebo od technologických vizionárov ako Elon Musk. Podľa Stephena Hawkinga sú otázky ďalšieho vývinu UI technológií a ich vplyvu na ľudskú spoločnosť „najdôležitejšou konverzáciou dnešných čias“. Prečo? Prečo by nás vývin UI technológií mal zaujímať viac ako napr. vývin novej série automobilov? Preto, lebo ľudstvo má v rukách (áno, zatiaľ to ešte máme v rukách) nástroj, ktorý má potenciál významným spôsobom zmeniť život celého ľudstva. V nasledujúcom texte predstavím niektoré riziká a príležitosti spojené s vývojom UI a pokúsim sa načrtnúť aj možné riešenia.

Budem vychádzať z vlastných skúsenosti vedca skúmajúceho ľudskú myseľ a pracujúceho na vývoji umelej inteligencie pre súkromnú spoločnosť, ale najmä z knihy *Life 3.0* kozmológa Maxa Tegmarka z MIT [1]. Názov knihy odkazuje na tri štádiá života na Zemi: 1) biologické - životné formy, ktorých stavba tela („hardvér“) i vzorce správania („softvér“) sú evolučne vyvinuté a počas života jedinca nemenné, 2) kultúrne - formy s nemenným telom schopné učiť sa a modifikovať správanie/softvér skúsenosťou (napr. naučiť sa cudzí jazyk alebo jazdiť na bicykli), 3) technologické - formy schopné počas života modifikovať aj svoje telo/hardvér. Ľudstvo je momentálne niekde na stupni *Life 2.1*: napriek pokroku v medicíne a čiastočným „hardvérovým“ modifikáciami (protézy končatín, umelé kĺby, kochleárne implantáty, kardiostimulátor) sme stále limitovaní chorobami a smrteľnosťou. Hoci vymanenie z biologických obmedzení by nás podľa Yuvala Harariho postavilo na úroveň bohov [2], Tegmark očakáva vyriešenie tohto problému skôr od inej ako ľudskej inteligencie: od tzv. superinteligencie - UI, ktorá by

nás na rozdiel od súčasných úzko špecializovaných inteligentných technológií predstihla vo všeobecnej inteligencii – schopnosti riešiť problémy, samostatne si stanovovať ciele, učiť sa a rozumieť sama sebe.

Tegmark sa, ako správny kozmológ, nezastaví v blízkej budúcnosti a okrem súčasných problémov i potenciálnych scenárov vývoja spoločnosti po vzniku superinteligencie („najbližších 10000 rokov“) sa venuje kolonizácii vesmíru, odvráteniu energetickej smrti v dôsledku vyhasnutia Slnka („miliarda a viac rokov“) i našej údajnej morálnej povinnosti zachovať inteligentné vedomie vo vesmíre (aj keby nebolo naše). Keďže však hodnovernosť predpovedí klesá s časovou vzdialenosťou od predpovedaného obdobia, sústredím sa najmä na otázky, ktoré sú naliehavé už teraz.

### Legislatíva

Nové technológie prinášajú nové právne problémy. Kto je zodpovedný, ak samojazdiace auto zrazí chodca na prechode: výrobca auta, vlastník, používateľ, alebo dokonca auto samo? Autá by mohli vstupovať do právnych vzťahov, mať vodičské oprávnenie a povinné poistenie. Ak je však samo auto účastníkom poistného vzťahu i potenciálnym platiteľom, malo by mu byť umožnené vlastniť majetok. Od toho je však už iba krôčik k tomu, aby si chytré algoritmy zarábali obchodovaním na burze a nakupovali online služby. Ak si algoritmus bude môcť najat ľudí, aby preň pracovali, dokáže čokoľvek, čo ľudia. Ak by sa mu v investovaní darilo lepšie ako ľuďom, mohlo by sa stať, že väčšinu ekonomiky by po čase vlastnili a riadili stroje, tak ako to už teraz robia iné neživé entity – nadnárodné korporácie.

*Dnešné algoritmy umelej inteligencie zriedkakedy obsahujú explicitné pravidlá. Väčšinou ide o vysoko distribuované systémy, kde výsledné rozhodnutie je dané momentálnou súhrou mnohých paralelných pôsobiacich – a často aj konfliktných – faktorov.*

Samojazdiace autá by mohli medzi sebou komunikovať, uistiť sa, kto dáva komu prednosť, a tým eliminovať kolízne situácie na križovatkách, zrýchliť dopravu po státi v zápche synchronizovaným štartovaním či optimalizovať priepustnosť dopravy v meste koordinovanou navigáciou. To by mohlo fungovať, keby na cestách boli iba navzájom kompatibilné samojazdiace autá. Myslím si, že to je nereálne – vždy budú ľudia, ktorí šoférujú radi a nebudú sa chcieť vzdať kontroly v prospech autopilota. Potenciálne kolízne situácie s rôznymi účastníkmi premávky – vrátane chodcov – prinášajú nielen právne, ale aj etické dilemy. Ak je auto v kolíznej situácii, kde môže zabrániť zrazeniu skupiny piatich detí, ale pritom narazí do stromu

a ohrozí život šoféra, čo si má vybrať? Čí život má väčšiu cenu? Takéto dilemy sú v rôznych variantoch obetí (tehotná žena, mladí versus starí ľudia, kriminálnici versus bezúhonní) témou teoretických myšlienkových experimentov v morálnej filozofii. So samozjaziacimi autami však dostávajú zúfalo praktickú podobu. Výskumy ukázali, že väčšina ľudí si v našom prvom príklade myslí, že je správne zachrániť deti aj za cenu ohrozenia šoféra, ale sami by si auto, ktoré by bolo takto naprogramované, nekúpili a ani nesúhlasia, aby vláda zákonom nanútila implementáciu takejto morálky do samozjaziacich áut. Edmond Awad a jeho kolegovia z MIT Media Lab vykonali obrovský internetový prieskum postojov ľudí k rôznym variantom takejto dilemy [3]. Program Moral Machine získal a analyzoval 40 miliónov rozhodnutí od 2,3 milióna ľudí z 233 krajín. Výsledky ukázali, že aj keď sa ľudia úplne nezhodnú, isté všeobecné trendy sú, napr. preferencia záchrany života mladých pred starými, alebo ľudí pred zvieratami. Zaujímavé sú kultúrne odlišnosti: krajiny sa dali zhruba zoskupiť do troch klastrov: západný (Európa a Severná Amerika), južný (Latinská Amerika a niektoré bývalé francúzske kolónie) a východný (Japonsko, Taiwan a niektoré moslimské krajiny). Ľudia z východného klastra napríklad vykazovali najmenšiu preferenciu záchrany mladých pred starými.

Znamená to, že pre rôzne trhy by sa mali vyrábať samozjaziace autá vybavené rôznymi morálnymi pravidlami? Alebo že by sa mali všeobecne záväzné morálne pravidlá prijímať hlasovaním? Alebo umožniť, aby si každý nastavil morálne správanie auta podľa svojich preferencií? Výsledky naznačujú, že ide o ťažký problém a ani my v ňom nemáme jasno. Napriek tomu dennodenne sadáme do auta a nakoniec sa inštinktívne vždy nejako rozhodneme. A takto to možno nakoniec bude aj s rozhodovaním áut, keďže dnešné algoritmy umelej inteligencie zriedkakedy obsahujú explicitné pravidlá. Väčšinou ide o vysoko distribuované systémy, kde výsledné rozhodnutie je dané momentálnou súhrou mnohých paralelne pôsobiacich - a často aj konfliktných - faktorov.

Vráťme sa teraz ešte do oblasti legislatívy a ľudských práv. Ďalšou dilemou, ktorú nové technológie prinášajú, je sledovanie versus bezpečnosť: Aké údaje je dovolené o ľuďoch zbierať a ukladať a kto a ako je oprávnený ich používať? Má právo na súkromie prednosť pred právom na slobodný prístup k informáciám či právom na bezpečie? Všadeprítomné bezpečnostné kamery a možnosť čítať z telefónov údaje o pohybe osôb a záznamy hovorov by predišli viacerým trestným činom, na druhej strane sú zneužiteľné na totalitnú manipuláciu. S rozvojom neurozobrazovacích metód bude možné zostrojiť presnejšie detektory lži či dokonca čítať niektoré myšlienkové obsahy. To už je na hranici orwellovskej dystópie. Už teraz je možné pomocou funkčnej magnetickej rezonancie (fMRI) zhruba zrekonštruovať kategóriu (ľudská tvár, postava, krajinka) alebo približný vizuálny tvar toho, na čo sa človek pozerá na videu [4] alebo určiť kľúčové slová témy, o ktorej číta [5]. Dnes sa ale pomocou technológií dajú ľahko

vytvoriť falošné dôkazy v podobe zvukového alebo obrazového záznamu či odtlačkov prstov: záznam zo sledovacieho systému alebo obsahu vlastnej mysle by mohol slúžiť ako objektívne alibi.

V niektorých krajinách sa začínajú používať algoritmy UI na tzv. *predictive policing*. Ide o odhad rizika spáchania trestných činov za účelom preventívneho zásahu, tak ako sme to videli v sci-fi filme *Minority Report*. Vo viacerých amerických štátoch predpovedá riziko recidívy zatknutého algoritmus UI. Na základe výsledku potom sudca rozhodne, či zatknutého prepustia na kauciu alebo nie. Ak by v budúcnosti takéto algoritmy úplne nahradili sudcov, eliminovali by sa subjektívne faktory a rozhodnutia by boli nestranné, zvýšila by sa dostupnosť a skrátila doba súdnych rozhodnutí. Ale bolo by to naozaj tak? Štatistická analýza rozhodnutí vyššie spomínaného algoritmu ukázala, že po kontrole faktorov, ako sú trestné záznamy z minulosti, vek a pohlavie, predpovedal vysoké riziko spáchania násilného trestného činu u zatknutých čiernej pleti o 77% častejšie ako u belochov. Navyše, keď sa predpovede neskôr porovnali so skutočnosťou, algoritmus sa mýlil u belochov a černochoch rôznymi spôsobmi: chybu 1. druhu (falošný poplach, teda predpoveď vysokého rizika a potom žiadna recidíva) urobil u 44,9% černochoch a u 23,5% belochov, chybu 2. druhu (falošná negativita, teda predpoveď nízkeho rizika a potom recidíva) u 47,7% belochov a 28% černochoch [6].

Už spomínaný Harari píše, že sledovacie systémy, ktoré používajú algoritmy spracovania veľkoobjemových dát (*Big Data*) na analýzu vzorcov ľudského správania a osobnosti, môžu ohroziť nielen naše súkromie a individualitu, ale aj naše demokratické inštitúcie [7]. Ak budú algoritmy rozumieť našej motivácii a rozhodovaniu lepšie, ako im rozumieme my sami, demokratické voľby a voľný trh sa stanú zastaranými. Algoritmy, ktoré budú vedieť predpovedať naše emócie, manipulovať naše túžby a robiť rozhodnutia za nás, sú veľmi nebezpečnou zbraňou, či už v rukách marketinových spoločností, ktoré nám chcú predať svoj tovar, alebo v rukách politických manipulátorov, ktorí chcú uchopiť moc.

### Autonómne zbrane

Ďalšou témou sú skutočné zbrane schopné zabíjať a fyzicky ničiť. V tejto oblasti už prebieha čulý výskum, armáda je vítaným sponzorom UI projektov v mnohých krajinách. Nahradenie živých vojakov autonómnymi/samoriadenými zbraňami by pomohlo znížiť straty na životoch na vlastnej strane a v prípade veľmi presných systémov možno aj na strane civilistov u nepriateľa.<sup>1</sup> Zároveň však nižšie potenciálne straty znižujú prah pre rozpútanie vojnového konfliktu. Výskum ktorejkoľvek veľmoci v tejto oblasti by

1 Podľa údajov publikovaných v roku 2014 v *American Journal of Public Health* [8] v súčasnosti na jedného zabitého vojaka vo vojnových konfliktoch pripadá desať civilných obetí, často v dôsledku zničenia infraštruktúry a následných humanitárnych katastrof.

odštartoval nové preteky v zbrojení (na rozdiel od jadrových zbraní, inteligentné technológie nevyžadujú nákladné alebo ťažko dostupné suroviny, preto sú dostatočne lacné na masovú výrobu). Tým by sa však skôr či neskôr dostali na čierny trh a do rúk rôznym diktátorom a teroristom. Aby sa predišlo takémuto scenáru, vznikli rôzne výzvy a iniciatívy za zákaz autonómnych zbraní medzinárodnou chartou, podobne ako je to v prípade chemických a biologických zbraní. Výzvu iniciovanú v roku 2015 Tegmarkom a Stuartom Russelom podpísalo viac ako 3000 vedcov z oblasti UI vrátane špičiek z firiem Google, Facebook, Microsoft a Tesla a 17 000 ďalších signatárov, medzi nimi napríklad Stephen Hawking.

### Trh práce a vzrast nerovnosti

Technológie, ktoré od priemyselnej revolúcie nahrádzali najmä mechanickú prácu, začínajú prenikať do oblasti služieb – digitálni asistenti budú poskytovať konzultačný a zákaznícky servis, právne poradenstvo, vzdelávanie, medicínsku diagnostiku, osobné trénerstvo, roboty prenikajú do ošetrovateľstva a starostlivosti o seniorov.<sup>2</sup> Softvér nie je nikdy unaveňný, možno ho mnohokrát replikovať či v okamihu nahradit novšou verziou, preto sa technologické riešenia postupne stanú lacnejšími ako služby poskytované ľuďmi. To môže demokratizovať služby a zvýšiť ich dostupnosť tam, kde by boli inak nedostupné alebo príliš drahé, ale môže viesť aj k ich dehumanizácii či k nerovnosti. Podobne ako si ekonomicky slabšie vrstvy nemôžu dovoliť zdravé biopotraviny, nebudú si môcť dovoliť ľuďmi poskytovanú starostlivosť.

Erik Brynjolfsson a Andrew McAfee [9] tvrdia, že technológie zvyšujú nerovnosť týmito tromi spôsobmi:

1) kvalifikácia: Trhová cena práce klesá dostupnosťou lacnejšej pracovnej sily. Je možné, že lacné technológie ju v niektorých odvetviach stlačia pod hranicu životných nákladov. Optimisti tvrdia, že so zánikom niektorých povolání vzniknú nové, tvorivejšie a zaujímavejšie, a nie také úmorné. To však povedie k ďalšiemu zvyšovaniu nerovnosti: dopyt po vyššie kvalifikovanej pracovnej sile bude rásť a po nižšej klesať; mladí, flexibilní a vzdelanejší, schopní rýchlejšej rekvalifikácie, budú vo výhode. Údaje o trendoch vo výške príjmu za posledných 50 rokov z USA ukazujú, že rýchlosť rastu príjmu je priamo úmerná dosiahnutému vzdelaniu: plat vysokoškolsky vzdelaného Američana vzrástol v priemere o 25 %, plat Američana s nedokončeným stredoškolským vzdelaním klesol o 30 %.

2) globálna konkurencia: V mnohých povolaniach berie víťaz konkurencie skoro všetko. Digitálne technológie umožňujú lacno replikovať nielen

2 EU v rámci grantovej výzvy Active Assisted Living programu Horizon 2020 alokovala viac ako 600 miliónov eur na projekty zvýšenia kvality života seniorov pomocou asistívnych technológií.

informačné produkty, ale do istej miery aj služby a obchodné procesy. To spôsobuje, že superhviezda v nejakej oblasti, či už je to spisovateľ, hudobná skupina, manažér alebo tvorca počítačových aplikácií, je dostupná globálne, čím konkuruje a vytláča z trhu horších, aj keď lokálne najlepších, hráčov.

3) kapitál versus práca: Keďže produkty ako knihy, filmy, účtovné nástroje sa digitalizujú, keď sa raz vyrobia, je ich možné replikovať prakticky s nulovými nákladmi a bez nutnosti najat ďalších pracovníkov. Preto v digitálnej ére pripadá väčšina vygenerovaného zisku majiteľom kapitálu a investorom, nie pracovníkom. Podľa správy organizácie Oxfam [10] 82% celosvetovo vygenerovaného bohatstva v roku 2017 pripadlo najbohatšiemu 1% obyvateľov planéty, zatiaľ čo chudobnejšia polovica ľudstva (3.7 miliardy) nedostala nič. Globálny koláč bohatstva bude pravdepodobne ďalej rásť, ale nebude sa deliť pre všetkých.

Ruka trhu zrejme ekonomickú nerovnosť sama nevyrieši - budú potrebné rozumné zásahy štátu. Jedným z možných riešení, o ktorých ekonomickí teoretici uvažujú, je univerzálny príjem. Ďalším je redistribúcia časti bohatstva vyprodukovaného technológiami na zmiernenie dopadov na najviac postihnuté sociálne skupiny. Nemusí to byť priamou finančnou podporou, ale napr. tým, že spoločnosť poskytne svojim členom zadarmo alebo v dotovanej podobe cesty, mosty, parky, verejnú dopravu, vzdelanie, zdravotnú starostlivosť, starostlivosť o deti a seniorov či prístup na internet. Na rozdiel od univerzálného príjmu, takéto riešenie okrem zníženia životných nákladov aj vytvára pracovné miesta. Ak by aj stroje dokázali vykonávať všetku prácu lacnejšie a lepšie ako ľudia, vlády sa môžu rozhodnúť dotovať povolania ako starostlivosť o deti a starých ľudí (napr. formou nižšej dane z príjmu), radšej ako ich vykonávanie prenechať strojom. Naopak, firmy masívne profitujúce z technológií možno daňovo zatažiť viac (tzv. digitálna daň, daň z robotov, atď.)

Práca a výkon povolania nemajú iba ekonomické aspekty, ale súvisia s pocitom sebahodnoty a zmyslu života. Podľa Tegmarka pozitívna psychológia identifikovala, že na naše duševné blaho vplyvajú faktory ako sieť priateľov a kolegov, zdravý a poctivý život, úcta, sebaúcta, pocit zvládania a *flow*, keď robíme niečo, v čom sme dobrí, pocit, že sme potrební a záleží na nás, pocit, že sme súčasťou a slúžime väčšiemu celku. Tieto faktory sú do istej miery sýtené prácou (aspoň v niektorých povolaniach), avšak dobrá správa je, že ich možno sýtiť aj mimopracovnými aktivitami. Spoločnosť, ktorá nechce generovať frustrovaných, nezamestnaných a zbytočných občanov (a voličov populistických a extrémistických strán), by mala investovať prostriedky vygenerované technológiami a výskumné kapacity ekonomov, ale najmä sociológov a psychologov, do zabezpečenia aktivít sýtiacich duševné blaho pre všetkých. Podporovať občianske, susedské a komunitné aktivity - oblasti vyžadujúce a podporujúce empatiu, sociálnu inteligenciu a medziľudské vzťahy.

## Technológie a ľudskosť

Zaujímavou otázkou je vplyv technológií na kultúru, trávenie voľného času, medziľudské vzťahy, komunikáciu, na našu ľudskosť ako takú. Internet nám umožňuje byť v okamžitom spojení s ľuďmi kdekoľvek na svete. Plánovať cestu do inej krajiny či nájsť najbližšiu dobrú reštauráciu je hračkou - informácie sú dostupnejšie a algoritmus ich často za nás aj spracuje a ponúkne nám, čo si máme vybrať. Mailovací program nám ponúka, že na maily odpovie za nás. Počítačové hry sa automaticky adaptujú na úroveň schopností hráča, a tým ho udržujú v pocite *flow*. So smartfónom po ruke sa už nikdy nemusíme nudiť. Protagonista transhumanizmu Ray Kurzweil ide ešte ďalej a predpovedá, že časom budeme vedieť eliminovať akékoľvek utrpenie rozšírenou realitou, chytrými drogami a nanorobotmi, ktoré v ľudskom tele „vyregulujú“ okruhy generujúce nepríjemné pocity. Kde je hranica medzi uľahčením si života a únikom z reality, vytváraním ilúzie? Namiesto riešenia problémov ich len prestať vnímať? Podľa švajčiarskeho futurológa Gerda Leonharda [11] stieranie hranice medzi virtuálnym a reálnym svetom a ľahká dostupnosť virtuálneho uspokojenia môžu viesť k otupeniu a strate ľudskosti. Cestou cez rómsku osadu si môžeme v rozšírenej realite premietiť na špinavé chatrče nádhernú krajinku s vodopádmi. Harari v knihe *Homo deus* hovorí, že v modernej dobe prestali byť meradlom morálky externé - Bohom dané - pravidlá, ľudia sa orientujú podľa vlastného cítenia. Čo bude naším morálnym kompasom, keď bude možné cítenie upraviť podľa potreby?

*Technológie nám môžu život uľahčiť, ale nemali by diktovať formu našej komunikácie a vzťahov. Ich prenikanie do všetkých oblastí života môžeme využiť na to, aby sme zistili, kto sme, akí sme a akí chceme byť. Aby sme zistili, čo je to podstatné ľudské, čo si chceme chrániť.*

V novom filme *Bohemian Rhapsody* je scéna, kde sa Freddie Mercury po pokusoch so sólovými projektmi vracia k členom skupiny Queen. Hovorí im: „Chlapci, s muzikantmi, ktorí ma doprevádzali, bol problém. Povedal som im, ako majú hrať, aká je moja predstava, a oni to presne tak zahrali. Všetci súhlasili, nikto nič proti.“ Rastieme a učíme sa aj cez konflikty a zlyhania. Rodič, ktorý dieťaťu odstráni z cesty všetky prekážky a vyrieši všetky problémy za neho, mu nerobí dobrú službu. Je to tenká hranica, na ktorú by sme mali byť citliví: technológie nám môžu život uľahčiť, ale nemali by diktovať formu našej komunikácie a našich vzťahov. Ich prenikanie do všetkých oblastí života môžeme využiť na to, aby sme zistili, kto sme, akí sme a akí chceme byť. Aby sme zistili, čo je to podstatné ľudské, čo si chceme chrániť.



## Moc a spoločnosť

Hoci som slúbil, že ostanem pri otázkach súčasnosti alebo blízkej budúcnosti, Tegmarkove scenáre usporiadania spoločnosti po vzniku superinteligencie upozorňujú na dôležitý aspekt prepojenia inteligencie s ekonomickou silou a politickou mocou. Záujmová skupina vlastniaca veľmi chytrú UI by ju mohla využívať na riešenie úloh pre platformy ako Amazon Mechanical Turk.<sup>3</sup> Neskôr by mohla zarobené prostriedky znásobiť investovaním na burze či kupovaním nehnuteľností. Chytré algoritmy, ktoré by vedeli predikovať ľudské správanie či správanie sa trhu, by rýchlo poskytli konkurenčnú výhodu. S akumuláciou ekonomickej sily prichádzajú javy, ktoré poznáme z nadnárodných spoločností a od oligarchov: lobing a pokusy o prispôsobovanie si zákonov a pravidiel, ovládnutie nezávislých printových i elektronických médií či ovplyvňovanie volieb personalizovanými kampaňami. To sme napokon mohli sledovať v kauze Cambridge Analytica, kde únik dát z Facebooku a ich následné využitie na vytváranie psychologických profilov voličov zasiahlo do posledných amerických volieb či do referenda o brexite.

Nemohla by však umelá inteligencia znásobiť dobré sily a pomôcť ľudstvu popasovať sa so súčasnými globálnymi výzvami, ako sú hroziace ekologické katastrofy či práve tu už viackrát spomínaná narastajúca sociálna nerovnosť? Globálne problémy potrebujú globálne riešenia, na ktoré naša kmeňová mentalita „my versus oni“ nestačí. Tegmark začína svoju knihu hypotetickým scenárom umelej superinteligencie, ktorá postupným ovládnutím médií a moci zabráni ľuďom zničiť svet, drancovať prírodu, utláčať menšiny - jednoducho stane sa osvieteným diktátorom, ktorý sa bude starať o naše blaho. Možno by sme takému diktátorovi odovzdali vládu radi a dobrovoľne. Ak by bol dostatočne múdry, držal sa v pozadí a ani by sme o ňom nevedeli... Ale keďže pokrok závisí od investícií, a tie sú väčšinou motivované ziskom, nie je zaručené, že výsledok technologického pokroku bude, utilitaristicky povedané, čo najväčšie blaho čo najväčšieho počtu ľudí.

### **Môžeme regulovať výskum, aby to všetko malo šťastný koniec?**

Je možné napriek všetkým opísaným rizikám ostať optimistom? Tegmark rozhodne optimistom je a vzdať sa výskumu v umelej inteligencii by podľa neho znamenalo premárniť šancu radikálne novým spôsobom pomocou technológií vyriešiť mnohé ťažké problémy, ktorým ľudstvo čelí. Je však múdre poznať riziká a podľa možnosti sa im vyhnúť. V osvietenejších krajinách preto vznikajú interdisciplinárne fóra, ktorých cieľom je príprava

---

3 Crowdsourcingová platforma s ponukou platených úloh, ktoré v súčasnosti počítače nevedia dobre riešiť.

na budúcnosť. V prvom rade analýzou situácie a potom aj prijatím preventívnych opatrení. Napríklad na Novom Zélande, kde pôsobím, je takéto fórum (NZ AI Forum) zložené z vedcov z oblasti UI, ale aj sociológov a ekonómov, právnikov, zástupcov vlády a zákonodarnej moci, ako aj investorov a kľúčových firiem podnikajúcich v oblasti UI technológií. Každá strana má svoje záujmy, napr. cieľom vlády je regulovať trh v prospech občanov a cieľom biznismenov je minimalizovať regulácie, resp. zachovať si dostatočný podnikateľský priestor. Predsa je však pre všetky strany výhodné komunikovať, spolupracovať a hľadať akési Nashovo ekvilibrium.

Umelá inteligencia je obrovskou komerčnou príležitosťou. V konkurencii ide aj o čas, preto môžu mať firmy vyvíjajúce UI tendenciu zanedbávať bezpečnosť a etické aspekty. Stretol som sa s dravým názorom „Kto sa samoobmedzuje, tomu ujde vlak“. Regulovať je v tomto prípade predovšetkým úlohou zákonodarcov. Štáty sú však dnes často slabšie ako globálne korporácie, preto jedinou šancou je riešiť problémy na nadnárodnej úrovni, čoho príkladom je napr. GDPR legislatíva Európskej únie. Ďalšou možnosťou sú medzinárodné charty a dohovory, podobne ako je to v prípade zákazu niektorých typov zbraní či v prevencii globálneho otepľovania.

Významnú úlohu zohráva tlak verejnej mienky a verejná kontrola zo strany nezávislých médií a mimovládnych organizácií. Ešte lepšie je, ak sa k problematike vyjadrujú tí, čo jej najviac rozumejú, teda rešpektovaní vedci a akademické či akademicko-priemyselné združenia. Tých našťastie pribúda, spomeniem napríklad Future of Humanity Institute (Oxford, riaditeľ Nick Bostrom), The Centre for the Study of Existential Risk (Cambridge UK, Huw Price), Future of Life Institute (Cambridge MA, Jaan Tallinn a Max Tegmark), Center for Human-Compatible Artificial Intelligence (Berkeley, Stuart Russell), Open AI (Elon Musk) či The Partnership on AI to Benefit People and Society (Amazon, Facebook, Google, Microsoft a IBM).

Vedci, vyvíjajúci technológie UI v spolupráci so súkromnými firmami, pokiaľ nie sú motivovaní iba ziskom, ale majú etickú reflexiu a zodpovednosť za využitie znalostí, ktorými disponujú, môžu byť hlasom svedomia zvnútra - ovplyvňovať dianie, určovať hranice, podporovať vnútrofirémnu diskusiu a vytvárať etické kódexy firiem. Verejnosť, investigatívni novinári a nezávislí auditori môžu sledovať dodržiavanie takýchto etických kódexov. Certifikačné organizácie môžu udeľovať certifikáty bezpečnej či etickej UI podobne, ako dnes máme certifikáty Fair Trade alebo certifikáty biokvality na ekologicky nezávadných potravinách. A spotrebiteľ sa rozhodne, aké výrobky/technológie bude kupovať.

Na záver to najdôležitejšie: aby sme mali vedcov so zdravým charakterom, hodnotovou orientáciou a etickou reflexiou, potrebujeme zmeniť vzdelávanie: dnes máme filozofov a právnikov, ktorí nerozumejú technológiám, a technokratov a IT špecialistov s nulovým humanitným zákla-

dom. Budúcnosť si vyžaduje interdisciplinárny prístup, preto je dôležité integratívne vzdelávanie (výučba etiky a iných humanitných predmetov v rámci IT a naopak) na všetkých úrovniach. Štúdium ľudskej psychológie by sa malo stať povinným predmetom na všetkých stredných školách, aby sa mladí ľudia učili porozumieť ľudskej psychike od detstva až po dospelosť. Čo túto psychiku formuje (a deformuje) a ako formovať sám seba, aby nás nemanipulovali ani ľudia ani UI. Príkladom (vysokoškolského) študijného odboru prepájajúceho interdisciplinárne prístupy k skúmaniu ľudskej mysle s technológiami, je v našich podmienkach úspešný medzinárodný magisterský program kognitívna veda na Fakulte matematiky, fyziky a informatiky Univerzity Komenského v Bratislave, v rámci ktorého pôsobím. Ak urobíme všetko preto, aby výchova k prosociálnym hodnotám, líderstvu, spoločenskej angažovanosti, sociálnej citlivosti a inkluzivite, kritickému mysleniu a historickej pamäti nezaostávali za technickými zručnosťami, budúcnosť nemusí byť taká temná, ako sa zdá.

## Odkazy na literatúru

- [1] M. Tegmark: *Life 3.0*. Allen Lane, 2017.
- [2] Y. N. Harari: *Homo Deus: A Brief History of Tomorrow*. Random House, 2016.
- [3] E. Awad a kol.: The Moral Machine experiment. *Nature* 563, 59–64, 2018.  
<https://www.youtube.com/watch?v=KMA23JJ1M1o>
- [4] Pereira et al.: Generating text from functional brain images.  
*Frontiers in Human Neuroscience* 5, 2011.
- [5] J. Angwin a kol.: Machine Bias. *Pro publica*, 2016. [tinyurl.com/robojudge](http://tinyurl.com/robojudge)
- [6] Y. N. Harari: Life 3.0 by Max Tegmark review - we are ignoring the AI apocalypse.  
*The Guardian*, 22. sept. 2017.
- [7] W.H. Wiist: The Role of Public Health in the Prevention of War: Rationale and Competencies.  
*American Journal of Public Health*, 2014.
- [8] E. Brynjolfsson, A. McAfee: Why Workers Are Losing the War Against Machines.  
*The Atlantic*. 2011. [tinyurl.com/wagedrop](http://tinyurl.com/wagedrop)
- [9] An Economy for the 99%. Oxfam briefing paper. 2017, [tinyurl.com/oxfam2017](http://tinyurl.com/oxfam2017)
- [10] G. Leonhard: *Technológia vs. humanita.*, SIEA, 2018.
- [11]