

# Ešte raz o ľudskej prirodzenosti, zdokonaľovaní a substitúcii

*ONCE MORE ABOUT HUMAN NATURE, ENHANCEMENT AND SUBSTITUTION*

---

Robert Burgan<sup>1</sup>

---

**Abstrakt:** Ľudské zdokonaľovanie v širokom slova zmysle, t. j. individuálnom aj sociálnom, si vyžaduje čo najpresnejšie vymedzenie ľudskej prirodzenosti, povahy alebo podstaty, pretože len na jeho základe je možné, či už na individuálnej, alebo kolektívnej rovine, rozhodnúť o tom, ktoré zásahy do jednotlivca a society sú užitočné a prospešné a ktoré, naopak, škodlivé a nebezpečné s ohľadom na ich krátkodobé a najmä dlhodobé záujmy. Problém diskutujeme s ohľadom na sériu dôležitých príspevkov v slovenskom časopise *Filozofia*, o ktorých, až na niekoľko výnimiek, možno súhrnne tvrdiť, že sa k naplneniu uvedeného epistemického cieľa z rôznych dôvodov od roku 2014 ani len nepriblížili.

**Kľúčová slova:** ľudská prirodzenosť, ľudské zdokonaľovanie, substitúcia človeka, ľudská a všeobecná inteligencia, umelá inteligencia, superinteligencia a technologická singularita

**Abstract:** Human enhancement in the broad sense, i.e. both individual and social, requires the most accurate definition of human nature, character or essence, because only on its basis it is possible, whether at an individual or collective level, to decide which interventions into the individual, community and society are useful and beneficial and which, on the contrary, harmful and dangerous with regard to their short-term and especially long-term interests. We discuss this problem with regard to a series of important research articles published in the Slovak academic journal *Filozofia*, that with a few exceptions, for various reasons, have not even come close to fulfilling the stated epistemic goal since 2014.

**Keywords:** human nature, human enhancement, human substitution, human and general intelligence, artificial intelligence, superintelligence and technological singularity

---

<sup>1</sup> roboburgan@post.sk

## Úvod

Podľa mienky väčšiny občanov je praktický úžitok z filozofie väčšinou žiadny. Znalci však vedia, že tzv. ontológie pomohli viacerým firmám a podnikateľom viesť poriadok do ich aktivít a radikálne vylepšiť metódy a formy, ktorými tieto aktivity organizujú a realizujú. Problém, ktorý v tejto štúdiu diskutujeme, má však veľký celospoločenský dosah, pretože od jeho vyriešenia bude závisieť ďalšia budúcnosť ľudí, chápaných ako individuálne bytosti a súčasne ako súbor jedincov prepojených rôzne intenzívnymi a komplexnými vzťahmi.

Zrejme z tohto dôvodu sa redakcia slovenského časopisu *Filozofia* rozhodla uverejniť celú sériu štúdií venovaných tematike ľudskej prirodzenosti a ľudského zdokonaľovania, v snahe jasne vymedziť túto prirodzenosť a na tomto základe ukázať, ktoré zásahy do ľudskej prirodzenosti (alebo povahy) sú prípustné a ktoré nie; ktoré teda ohrozujú samotné bytie (a prežitie) jedinca a súčasne aj celej ľudskej spoločnosti. Domnievame sa, že táto snaha sa až na niekoľko výnimiek, t. j. relevantných a skutočne fundovaných štúdií, minula účinkom a viacerí vo *Filozofii* publikujúci autori sa k riešeniu tohto problému ani len nepriblížili.

Po kritickej analýze niektorých relevantných štúdií pristupujeme k uvedeniu vlastného riešenia, pričom nás bude zaujímať, podobne ako tomu bolo v našej štúdiu, uverejnenej v *E-LOGOSE* ešte v roku 2014, vymedzenie ľudskej prirodzenosti nielen smerom „dozadu“, t. j. vo vzťahu k biologickej ríši, ale aj smerom „dopredu“, t. j. vo vzťahu k tzv. strojovej (alebo umelej) inteligencii, resp. generalizovanej (alebo všeobecnej) umelej inteligencii, ktorá by čoskoro mala byť na úrovni ľudskej inteligencie (či dokonca ju o jeden-dva rády prevyšovať) a nakoniec aj superinteligencii, ako je prezentovaná napríklad N. Bostromom (2014).

Budeme pritom vychádzať z koncepcie *prírodných druhov* (alebo *foriem pohybu hmoty*) a súčasne prezentovať aktuálne poznatky o umelej inteligencii, možnostiach jej ďalšieho rozvoja či zdokonaľovania, hrozbách a rizikách, ktoré (údajne) prináša jej čoraz širšie alebo dôslednejšie používanie. Zároveň však predstavíme niekoľko textov, ktoré na jednej strane na základe vecnej a odbornej argumentácie dokazujú, že žiadnu všeobecnú umelú inteligenciu či dokonca superinteligenciu sa nám vytvoriť nepodarí, a na strane druhej berú jej vytvorenie ako takmer istý fakt a ponúkajú prístupy, v rámci ktorých bude naše spolužitie s ňou nielen príjemné a obohacujúce, ale stane sa aj základným predpokladom nášho ďalšieho rozvoja.

## Potrebujeme vôbec pojem ľudská prirodzenosť?

Francúzsky filozof Yves Charles Zarka (2018) v snahe zodpovedať túto otázku dostatočne presne a najmä „produktívne“ s ohľadom na *aktuálnu „situáciu“ človeka* najprv diskutuje Marxove a Nietzscheho pokusy vymedziť túto prirodzenosť, uvádza viaceré zmysluprázdne názory Martina Heideggera, aby konečne predložil Hobbesovu *maximalistickú* definíciu tejto prirodzenosti, podľa ktorej (tamže, s. 780) je táto „súhrnom jeho prírodných schopností a možností, akými sú schopnosť výživy, pohybové, rozmnožovacie, pocitové, rozumové schopnosti atď.“ Ako ďalej zdôrazňuje, pre túto koncepciu ľudskej prirodzenosti je typické to, že de facto vylučuje myšlienku dejinnosti človeka a spoločnosti, pretože podľa Hobbesa je ľudská prirodzenosť „stále identická sama so sebou, nezávislá od miesta a času“, čo údajne platí aj pre jeho teóriu štátu. Ďalej sa už objavujú v Zarkovej argumentácii určité rozpory.

Na jednej strane napríklad tvrdí (Zarka 2018, s. 781), že práve osvojovanie a osvojenie reči (spolu s ďalšími znakmi) zohráva „rozhodujúcu úlohu pri vytváraní prirodzenosti človeka a jeho odlišení od zvierat“, pretože človek je údajne tvor politický, „zatiaľ čo zviera ním nie je“, ale na strane druhej označuje používanie reči človekom za znak *arbitrárny*, t. j. ako znak bez ktorého by sa *človek ako taký* v krajnom prípade mohol zaobísť úplne. Zarka pritom

nielenže umelo a násilne odtrháva biologické od sociálneho a kultúrneho, ale o pár riadkov ďalej aj individuálne od kolektívneho a sociálneho, keď konštatuje, že spomínané špecificky ľudské schopnosti sa neobjavujú hotové, ale utvárajú sa postupne, na základe čoho je preto možné konštatovať, že Hobbesova koncepcia ľudskej prirodzenosti je síce dynamická, ale už nie dejinná, pretože spomenuté schopnosti sa *u jednotlivých ľudí* rozvíjajú (!) len v zmysle rastu ich komplexnosti, sotva však v rámci spoločnosti *ako zreteľne vyčlenenej entity*.

Rovnako kontroverzný a rozporuplný je aj Zarkov výklad Rousseauovej koncepcie ľudskej prirodzenosti, ktorú definuje ako *minimalistickú* len preto, že podľa Rousseaua sú (tamže) prirodzenými určeniami človeka, ktoré ho odlišujú od zvierat, hlavne *sloboda a schopnosť zdokonaľovať sa*, ktorá je zlučiteľná s jeho dejinnosťou, t. j. s postupným zdokonaľovaním nielen týchto dvoch bytostných určení, ale aj spoločenských foriem či kultúr, v rámci ktorých sa rozvíjajú a zdokonaľujú. Úplne absurdne potom vyznieva Zarkovo tvrdenie, že práve vtedy sa človek vzdáľuje od svojej prirodzenosti, najmä ak ju predtým Rousseau definoval *ako nebiologickú*, keďže o slobode (a iných behaviorálnych alebo morálnych kategóriách) pri zvieratách nemá zmysel diskutovať. Nie je teda pravda, ako nakoniec tvrdí Zarka, že človek sa takto „stáva človekom kultúrnym, oddeleným od človeka prírodného“, pretože pôvodné prírodné (alebo biologické) sa z neho *pritom neodstraňuje*, skôr dostáva pod stále väčšiu a úplnejšiu kontrolu kultúrneho alebo, možno ešte presnejšie, *technického a technologického*.

Tento moment podľa nášho názoru výborne vystihol v roku 1983 Edgar Morin (tamže, s. 783-784), o ktorom Zarka diskutuje v rámci tzv. antropologickej kritiky ľudskej prirodzenosti. Podľa Morina sa pri definovaní človeka, jeho podstaty a prirodzenosti treba vyhnúť dvom extrémom – *biologizmu*, kedy sa pri jeho definovaní zameriavame len (alebo prednostne) na jeho organizmus (čo je typické pre sociálny determinizmus, sociobiológiu a ďalšie druhy konzervatívneho a nebezpečného myslenia), a *antropologizmu*, postavenému na „ostrovnej“ koncepcii človeka, zúženej na jeho psychokultúrny rozmer. Až potom budeme môcť vedeckej komunite predložiť realistickú „teóriu bio-psycho-sociálnej celosti človeka“. Ako však ďalej podotýka Zarka (tamže, s. 784), táto fundamentálna, prírodná a súčasne historická a sociálna antropológia, založená na troch pilieroch (hominizácii, vytváraní malých území i komunit, dovolíme si doplniť, a organizácii veľkých a rôznorodých území) vznikne až „vtedy, keď sa skončí zbožňovanie človeka sebou samým, keď rozbijeme mýtus ľudskej prirodzenosti“.

S úzko vymedzeným pojmom ľudskej prirodzenosti chce skoncovať aj Marshall Sahlins, a to posunutím a radikalizovaním perspektívy (Zarka 2018, s. 784-785), keďže, ako sa domnieva, Hobbesov model ľudskej prirodzenosti je (akoby) vzorový a predstavuje tak model jednej civilizácie (t. j. Západu), ktorá samu seba prehlásila za univerzálnu a „všetkými možnými spôsobmi vrátane nadvlády, imperializmu a kolonializmu sa pokúšala rozšíriť svoje hodnoty na všetky ostatné civilizácie, pre ktoré tento pojem nebol relevantný“. Sahlins dokonca ide tak ďaleko, že *bez akejkoľvek vecnej a presvedčivej argumentácie* stotožňuje Hobbesovu údajne zvrátenú a pomýlenú víziu ľudskej prirodzenosti s víziou celej západnej civilizácie, ktorá tak v dôsledku tejto falošnej idey podľa všetkého dokonca ohrozuje budúcnosť celého ľudstva.

Sám Zarka potom už nemá problém tvrdiť (2018, s. 785), že samotná antropológia *vôbec nepotrebuje pojem ľudskej prirodzenosti*, pretože sa prirodzene viac zaujíma o špecifickosť rozmanitých spoločností, mravov a kultúr. O pár riadkov ďalej však šokuje svojho čitateľa protikladne vyznievajúcim tvrdením (tamže, s. 786), že pojem ľudskej prirodzenosti, resp. samotná ľudská prirodzenosť nemá vykonštruovanú a fiktívnu povahu, pretože len s jej pomocou je možné obrániť slobodného a kreatívneho človeka pred transhumanistickými a inými umelými zásahmi do jeho prirodzenosti, tentoraz zjavne vymedzenej prednostne

biologicky a nie „psychokultúrne“ alebo sociálne<sup>2</sup>. Vychádza pritom z Habermasovho diela *Budúcnosť ľudskej prirodzenosti* (2001), v ktorom sa, prinajmenšom podľa Zarkovho výkladu, žiada, aby sme na biologickej úrovni vytvorili „hradbu proti (transhumanistickému a liberálnemu – R.B.) úsiliu o úplnú kontrolu človeka samého nad sebou a nad budúcimi generáciami“ (Zarka 2018, s. 786). A to môžeme urobiť podľa neho len tak, že ochránime *umelo nevytvorené a prirodzené v človeku*, t. j. *biologické* pred zjavne prehnanými ambíciami všetkého sociálneho, kultúrneho, technického alebo dokonca biotechnického.

Z vedeckej a filozofickej argumentácie sa tak na záver dostávame *k čisto politickej alebo účelovej*, keď Zarka v snahe ochrániť človeka pred „totalitnými“ zámermi radikálne zmeniť jeho prirodzenosť či podstatu pôvodne zbytočný a škodlivý pojem ľudskej prirodzenosti zrazu vyhlasuje (tamže, s. 787) „za pozitívny základ takého zásadného pojmu, akým je ľudstvo“. Ako totiž vyplýva z jeho ďalšej argumentácie, takýto spoločný základ, takúto prirodzenosť, „ktorá je nielen spoločná, ale identická“, potrebujeme jednoducho preto, aby sme mohli hovoriť o jednotnom ľudstve, v ktorom je neprípustný napríklad rasizmus a idea ľudských práv sa tak môže oprieť o pevné ideové východisko, ktoré nie je korektné kritizovať alebo spochybňovať, čo nám naozaj pripadá ako určité ad hoc alebo doslova oportunistické riešenie (najmä ak je agenda univerzálnych ľudských práv v mnohých etnikách či kultúrach veľkou časťou ich politických elít, ale aj samotných občanov otvorene odmietaná); na čom sotva niečo zmenia autorove už úplne účelové tvrdenia o existencii akejsi hmotnej a živej ľudskej bytosti (tamže, s. 788) a pochybné tézy o potrebe novej kozmopolitnej a materiálnej teórie ľudskej prirodzenosti ako praktickej teórie, ktorá bude „proti akémukoľvek biologickému redukcionizmu“ práve preto, aby v rámci rozmanitých kultúrnych formácií bolo možné elementárne slobodné „bytie otvorené náhodnosti, rozmanitosti, skrátka možnostiam“.

Ešte zaujímavejšie ako rozpory v krajne nekonzistentnej Zarkovej argumentácii sú však podľa nášho názoru *ich príčiny*. Ako sa domnievame, hlavným problémom u Zarku a rovnako aj u väčšiny autorov v časopise *Filozofia* je nerozlišovanie medzi človekom ako *biologickým druhom* a človekom ako *prirodzeným druhom*, pre ktorého v druhom prípade naozaj nie sú určujúce jeho biologické vlastnosti, ale to, čo sa zvykne učene označovať ako *differentia specifica*, t. j. nie tie vlastnosti, ktoré zdieľame so zvieratmi, ale tie, ktorými *ony vôbec nedisponujú*, napríklad reč, slobodná vôľa či perspektíva 1. osoby, ako o nej vo viacerých textoch veľmi fundovane a dôkladne diskutuje najmä Lynne Rudder Bakerová (2000, 2013). Prísne vzaté, ľudia by sme ako biologický druh *ani nemali označovať*, hoci len preto, že pre nich neplatí Darwinov prirodzený výber, a to aj napriek tomu, že aj ľudia a ich inštitúcie (firmy, štáty, samosprávy a i.) *súperia* o nedostatkové látkové, energetické a informačné zdroje, ale bez toho, aby tieto súperiace subjekty získavali v prvej inštancii výhodu zo zmien svojho biologického substrátu, tak ako je tomu v (čisto) zvieracej alebo živočíšnej ríši.

Druhým veľkým problémom väčšiny koncepcií ľudskej prirodzenosti (a to nielen v časopise *Filozofia*) je v mnohých prípadoch nezámerné, resp. intuitívne nadradovanie individuálneho nad kolektívne<sup>3</sup>, a to znova napriek tomu, že tieto aspekty človeka ako takého nielenže nie sú rovnocenné, ale sociálne je základom a východiskom individuálneho v človeku, špecifickou formou „zakotvenosti“ alebo danosti, bez ktorej si jeho existenciu ani nemôžeme predstaviť. Skvelým spôsobom to objasňuje i dokazuje Rom Harré vo svojom staršom, ale vo viacerých

<sup>2</sup> Sám Zarka totiž na predošlej strane podľa nášho názoru celkom oprávnene ostro kritizoval tzv. hlbinnú alebo hlbokú ekológiu za to, že ľudskú prirodzenosť chápe de facto *čisto biologicky* a všetky prírodné tvory považuje za „rovnako hodné existencie“, na základe čoho by sme mali pristúpiť k redukcii ľudskej populácie na našej planéte a zachrániť tak ostatné (biologické) druhy a prírodu vôbec pred sebeckým a nenásytným človekom.

<sup>3</sup> Človeka si vtedy redukcionisticky a povrchno predstavujeme ako bytosť zloženú zo svojej telesnej a duševnej zložky a nie ako organickú súčasť tej či onej komunity alebo ľudskej spoločnosti ako reálneho celku.

ohľadoch neprekonanom a „klasickom“ texte (1993), kde napríklad na 34. strane tvrdí, že „štruktúrované skupiny, t. j. kolektívy, sú ontologicky nadradené jednotlivcom. Ľudské bytosti sú konštituované ako ľudské prostredníctvom ich interpersonálnych vzťahov“<sup>4</sup>.

Tretí problém úzko súvisí s prvým, pretože pokiaľ ľudí vcelku, vo všeobecnosti a v súhrne nepovažujeme za zvláštnu a jedinečnú formu pohybu hmoty, resp. prirodzený druh, ale len za prednostne biologický druh s akýmsi náhodným psycho-sociálno-kultúrnym „prídavkom“, ktorý má zmysel porovnávať len s biologickými jednotkami či taxónmi bez tohto „prídavku“, potom nielen nemôžeme presne definovať človeka, ale ani postihnúť jeho pozíciu a „osud“ vo vzťahu k ostatným prirodzeným druhom. Inak povedané, v určitom okamžiku prírodného vývoja sa rodí v rámci (*a na hranici!*) biologického človeka aj so svojou psycho-sociálno-kultúrnou podstatou, pričom od tohto okamžiku nesie v sebe a v tej istej podstate aj svoje „prekonanie“, t. j. vlastnosti alebo parametre, ktoré sa v plne rozvinutej forme objavia a naplno rozvinú v následných postľudských alebo postsociálnych formách pohybu hmoty.<sup>5</sup>

## Ľudské (seba)zdokonaľovanie ako jedna z hlavných tém filozofického diskurzu

Ako sme mohli vidieť, viacerí bádatelia dopredu odmietajú uvažovať o existencii nejakej jasne určenej a pretrvávajúcej ľudskej prirodzenosti alebo podstaty. O to zaujímavejšie je, že pokiaľ ide o ľudské zdokonaľovanie, resp. sebazdokonaľovanie, toto je od čias starovekej gréckej filozofie skutočne jednou z hlavných tém filozofického diskurzu v tom najširšom slova zmysle. Prinajmenšom takto sa vec javí v štúdiu Emila Višňovského z roku 2015<sup>6</sup>, aj keď autor problém diskutuje najmä z hľadiska pragmatickej filozofie. Podľa neho samotný transhumanizmus, ako myšlienkové hnutie venované biotechnologickému, medicínskemu či rovno technickému vylepšovaniu človeka, „nie je antihumanizmom ani kritikou humanizmu“ (Višňovský 2015, s. 343), ale skôr akýmsi „superhumanizmom“, v rámci ktorého ide najmä o prekonanie a zdokonalenie „človeka až za jeho prirodzené či prírodné (biologické) hranice a možnosti“.<sup>7</sup> Lenže pre humanistov (tamže, s. 344) „je človek, ako ho poznáme, t. j. ako prírodno-kultúrna bytosť najvyššou hodnotou“, ktorú treba za každú cenu ochrániť a uchovať.

Ibaže, ako ďalej podotýka E. Višňovský (tamže, s. 346), humanizmus vôbec „nie je jednotné myšlienkové a kultúrne hnutie“ a prihlásenie sa k nemu preto „neznamená rovnaké chápanie človeka, ľudstva či ľudskosti“. Na druhej strane, v odbornej literatúre západnej provincie a následne aj v spoločenskom vedomí pomyselného „západného človeka“ je humanizmus najčastejšie stotožňovaný s myšlienkovým hnutím a svetonázorovým postojom, ktorý proti sebe (niekedy až antagonisticky – R. B.) stavia sekulárne a božské, kultúrne a prírodné, resp. individuálne (či subjektívne) a sociálne (či objektívne). Inak povedané, humanizmus (akoby)

<sup>4</sup> V origináli: „From this discussion important conclusion emerges that structured groups, that is collectives, are ontologically prior to individuals. Human beings are constituted as people by their interpersonal relations.“

<sup>5</sup> Je to aj problém správnej „optiky“ alebo perspektívy, pretože ak pripúšťame existenciu len štyroch hlavných prirodzených druhov alebo foriem pohybu hmoty (ďalej aj FPH) – fyzikálnej, chemickej, biologickej a sociálnej, nikdy nebudeme schopní nielenže určiť „miesto človeka v prírode“, ale ani postihnúť dynamiku a hlavné zákonitosti vývoja v pozorovanom vesmíre, v ktorom I. F. Zubkov (1981, s. 143 a n.) napríklad vyčleňuje aj atómovú, mineralogickú, petrologickú, hviezdno-planetárnu a geologickú formu pohybu hmoty.

<sup>6</sup> Jej prvá verzia bola publikovaná v 5. čísle časopisu *Filozofia*, zatiaľ čo druhá, rozšírená verzia vyšla v roku 2019 v zborníku P. Sýkoru (ed.) *Promises and Perils of Emerging Technologies for Human Condition*.

<sup>7</sup> Všimnime si už na tomto mieste redukcionistický a do veľkej miery naivný prístup transhumanistov k človeku, pretože úplne mimo ich „záberu“ sa dostávajú inštitucionálne, ekonomické a geopolitické, nieto klimatické otázky a problémy, bez ktorých sú však všetky úvahy o lepšej budúcnosti človeka a ľudstva nepresvedčivé, a rovnako aj ich intuitívne a nekorektné stotožňovanie prirodzeného s prírodným a následne aj s biologickým.

vždy bol *na strane jednotlivca* stojaceho osamote proti zložitému a veľmi často aj vyslovene nepriateľskému svetu. A aj na základe veľmi negatívnej historickej skúsenosti z „fungovania“ totalitných režimov (Arendtová 1996) sa preto dnes jeho autentickí zástancovia snažia obrániť a ochrániť človeka ako takého pred všetkými riskantnými zásahmi zo strany transhumanistov, resp. ich stúpcov, a to tak v rámci filozofického diskurzu, ako aj tzv. reálnej politiky.

E. Višňovský zároveň ideu „vylepšovania človeka“ analyzuje najmä v rámci dejín filozofie, keď najskôr hovorí o *konceptii perfekcionizmu* (tamže, s. 344), ktorá našla svoj výraz najprv „v platónskom dualizme dokonalého sveta ideí (podstát, vzorov) a nedokonalého sveta vecí a ľudí“ a neskôr aj v aristotelovskej *eudaimónii* ako túžbe po lepšom, dokonalom živote, ktorá je (údajne) vlastná ľudskej prirodzenosti a predstavuje tak teóriu „najvyššieho dobra, ktoré je samoúčelom nepodliehajúcim žiadnemu cieľu“. Humanista Aristoteles preto ľudom odporúča (tamže, s. 345), aby sa zamerali na rozvoj „svojej prirodzenosti, svojich najlepších vlastností, schopností a cností v morálnom a politickom zmysle“, nie však, ako E. Višňovský dodáva dosť nehistoricky, „v zmysle biologickom, technickom či sociálnom“, pretože o takomto type či forme „vylepšovania človeka“ nemohlo byť podľa nás v starovekom Grécku ani reči.

Na problematickosť stotožňovania dokonalosti s dobrým životom a šťastím pritom poukázal už v roku 1793 Imanuel Kant, ktorý zdôraznil, že „to, čo sa javí ako správne teoreticky, nemusí byť správne prakticky“, v dôsledku čoho sa prirodzená ľudská túžba po dokonalosti (alebo perfektnosti) stáva problematickou v prípade, ak sa dostane do rozporu s „obyčajnou“ ľudskou túžbou po šťastí, potrebou dôstojnosti, starostlivosti o seba či sebakontroly (alebo nezávislosti od rušivých vonkajších zásahov alebo vplyvov). Zvlášť nápadné je to pri tých proponentoch transhumanizmu, ktorý *problém druhej transformácie* súčasného človeka (a ľudstva) zjednodušene a naivne redukovujú, tak ako David Pearce alebo Mark Alan Walker (Višňovský 2015, s. 346), na dosiahnutie maximálneho blahobytu a šťastia a minimalizáciu utrpenia, ktoré chcú v prvej etape „transformácie ľudstva“ výrazne zredukovať (či odstrániť) pomocou chemických látok, drog či tzv. piluliek šťastia a v druhej aj pomocou genetického inžinierstva a nanotechnológií, ktoré by mali „preprogramovať“ animálnu podstatu človeka, ako aj ďalších predátorov a zmeniť tak Zem na dokonale harmonické a šťastné miesto.<sup>8</sup>

Podobným excesom sa však dokázali vyhnúť hlavní proponenti *pragmatického humanizmu*, ako napríklad Ferdinand Canning Scott Schiller (Višňovský 2015, s. 348 a n.), ktorý sa dlho a systematicky zaoberal otázkami eugeniky, v ktorej „videl budúcnosť a spásu ľudstva“. Podľa neho bolo a je napríklad chybou, ak „najtalentovanejší ľudia nemajú deti“ (tamže, s. 349). Štát by mal pritom pomáhať rodinám v zlepšovaní ich intelektu a zvyšovaní ich blahobytu, ale bez toho, aby upadol do extrémov rasizmu alebo fašizmu. Rozlišovanie medzi pozitívnou a negatívnou eugenikou sa takto stalo celkom legitímnym, a tak je prirodzené, že našiel svojich nasledovníkov napríklad v rámci talianskej vetvy „magického pragmatizmu“, reprezentovanej Giovanim Papinim a Giuseppem Prezzolinim. Pre G. Papiniho napríklad filozofia nebola len súborom ideí, ale skôr a viac *behaviorálnym programom*, v rámci ktorého môže človek zo seba doslova urobiť Boha a zmeniť sa teda z biednej a slabej bytosti „na najvyššiu a suverénnu bytosť, absolútne bohatú a mocnú“ (Papini 1924, s. 130 a 207), zatiaľ čo G. Prezzolini ako pragmatik a modernista takisto písal články o Božskom Človeku „a jeho schopnosti všetko meniť a utvárať podľa svojej vôle a predstáv“ (Višňovský 2015, s. 349).

---

<sup>8</sup> Ako pod čiarou upozorňuje E. Višňovský (2015, s. 346), s ideou transhumanizmu „v zmysle sebatranscendencie ľudstva ako prírodného druhu“ prišiel už dávnejšie J. Huxley (1957), pričom najradikálnejšie rozvinutie zrejme našla vo Vingehe koncepcii technologickéj singularity ako toho momentu vo vývine našej civilizácie, keď sa v dôsledku ľudského zásahu alebo spontánne, akoby úplne z vlastnej vôle zrodí superinteligencia, ktorá nielenže bude zrýchlene *zdokonaľovať samu seba*, ale aj takým spôsobom a tak rýchlo, že oveľa menej inteligentní ľudia jej nebudú schopní porozumieť, a preto ani adekvátne reagovať na jej správanie a s ním spojené zmeny na Zemi.

Pragmatik a skvelý psychológ William James (tamže, s. 349-350) zase považuje humanizmus za súčasť pragmatizmu, pričom život ako taký vymedzuje v rámci svojej melioristickej etiky ako súčasť „otvoreného pluralistického univerza“, pre ktoré sú príznačné neistota, rizikovosť, a absencia „absolútnych pravidiel na dosiahnutie dobra“, z čoho vyplýva potreba „neustáleho konkrétneho a flexibilného prispôsobovania sa meniacim okolnostiam“. W. James zároveň nijako nešpecifikuje, v čom by malo spočívať „vylepšovanie človeka“, ale napriek tomu ho definuje ako otvorený projekt, v rámci ktorého sa vášnivá a kreatívna ľudská bytosť snaží prekonať dané, t. j. svoju prírodnosť a smrteľnosť, a dosiahnuť nakoniec (akoby) Božskú všemohúcnosť a nesmrteľnosť. A hoci je evidentné, že podobným snahám chýba vedecké zdôvodnenie, nedá sa poprieť, že pôsobia „ako praktická životná sila“ (James 1930, s. 52-53).

Veľmi zaujímavým a inšpiratívnym spôsobom pristupuje k problému ľudskej prirodzenosti a ľudského zdokonaľovania ďalší klasik pragmatickej filozofie – John Dewey, ktorý na rozdiel od F. C. S. Schillera (Višňovský 2015, s. 350) odmieta „subjektivistický humanizmus“ či dokonca humanizmus ako taký, pretože je mu bytostne cudzie akékoľvek „zbožšťovanie“ človeka a rovnako aj neorganické odtrhávajúce „ľudskej (antropologickej) stránky od prírody“, oddeľovanie ľudstva od prírody, resp. dualizmus kultúry a prírody. Višňovský preto právom označuje jeho humanizmus ako naturalistický. Diskutabilné je však jeho hodnotenie Deweyho demokratického humanizmu, keď najprv uvádza, že Dewey požadoval, aby štát podporoval rozvoj schopností, návykov či cností občanov (tamže, s. 350-351), čo je však podľa neho v rozpore s Deweyho konceptom demokracie, ktorá „zahŕňa pluralitu foriem života a hodnôt, nemôže preferovať ani jeden z konceptov ľudského rozvoja a musí ponechať na samotných jednotlivcoch, aby si ho zvolili“. Podľa nášho názoru však ide o rozpor len zdanlivý, pretože Deweyho požadovanie štátnej či celospoločenskej intervencie je v súlade s jeho koncepciou „vylepšovania človeka“ a ľudstva v rámci *ich* rastu a ten môže byť generovaný len na základe racionálnej a demokratickej verejnej diskusie a širokého celospoločenského konsenzu.

O Deweyho vyváženom prístupe k problému svedčí aj to (Višňovský 2015, s. 351), že rast na individuálnej a spoločenskej úrovni považoval „za prirodzenú stránku každej ľudskej skúsenosti, a to skôr kvalitatívnu ako kvantitatívnu“, pretože pri ňom nejde len o rozširovanie života, ale skôr o jeho kultiváciu, t. j. o jeho „obohacovanie, zhodnocovanie, prehlbovanie“. Za hlavnú formu takéhoto individuálneho i celospoločenského vylepšovania pritom považoval *edukáciu*, ktorú však chápal *vyslovene moderne*, t. j. nielen ako transfer poznatkov z pedagóga na žiaka, ale ako proces zameraný najmä na „rast sociálnych zručností a lepšieho spojenia so životom“, ktorý má byť základným cieľom školy. Rovnako treba zdôrazniť, že Dewey chápal zdokonaľovanie (1932, s. 77) ako kontinuálny a otvorený proces a nie ako vývoj k nejakému „fixnému ideálu dokonalosti“, ako tomu bolo podľa nás napríklad u Platóna či Karola Marxa.

Nakoniec, pokiaľ ide o Richarda Rortyho, Višňovský sa stotožňuje s K. Soperovou (2001, s. 115-129), podľa ktorej bol R. Rorty určite humanista, keďže odmietal „závislosť ľudstva od mimoludských síl“, veril v autonómnosť a slobodu ľudstva a rovnako aj v ľudský pokrok; podobne ako ona si však nie je istý tým, či na tomto základe uznával univerzálnu ľudskú prirodzenosť, t. j. pre človeka podstatné univerzálne vlastnosti alebo črty, ktoré má človek ako taký spoločné s ostatnými živočíchmi (!) alebo ľuďmi vrátane schopnosti vytvárať a používať jazyk. R. Rorty totiž, takisto ako Jean Paul Sartre, pripisoval veľmi dôležitú úlohu *ľudskej sebakreácii* (t. j. nie „vonkajšiemu“, prírodnému alebo objektívnemu determinizmu), ale bez toho, aby tvrdil, že človek sa môže stať všemocnou bytosťou, navyše posadnutou vlastnou mocou. Ďalší kritici Rortyho diela ho však jednoznačne označili za humanistu (Višňovský 2015, s. 352), pretože „celá jeho filozofia bola motivovaná starosťou o ľudstvo, empatiou,

sympatiou a solidaritou s ním, nádejou, že nespravodlivosť, ponižovanie a zbytočnú krutosť panujúcu v medziľudských vzťahoch možno odstrániť, snahou o povznesenie ľudstva“.

A aj keď v mnohých prípadoch môžeme s R. Rortym a ďalšími pragmatickými humanistami súhlasiť, predovšetkým s ich obranou jednotlivca či človeka ako takého pred deštruktívnymi totalitnými režimami a ich násilnými a „unifikujúcimi“ praktikami, pokiaľ ide o jasné určenie ľudskej prirodzenosti, ani oni na túto otázku *nedokázali dať presnú odpoveď*, ako je zrejmé aj zo záverečných pasáží Višňovského štúdie (tamže, s. 353), kde sa uvádza, že aj keď títo filozofi podporovali ideu zmeny a rastu vo vývoji človeka a spoločnosti, vyhli sa požiadavke zásadného prekonania „biologických a existenčných podmienok či ľudskej prirodzenosti“, resp. prekonania ľudskej biologickej podstaty technológiou. Bohužiaľ, ako sme mohli vidieť, pomyselná ľudská podstata *určite nie je biologická*, resp. *čisto biologická*, a čo je ešte horšie, ak pri jej vymedzovaní úplne odhliadneme od sociálnych a kultúrnych faktorov, ľahko sa nám môže stať, že „demarkačnú líniu“ medzi biologickým a sociálnym stanovíme úplne arbitrárne alebo budeme *rovnako nekorektne ignorovať* tie jej vymedzenia a „praktické aplikácie“, ktoré sú dané najmä sociálnymi a kultúrnymi faktormi a človeka *jasne vyčleňujú zo živočíšnej ríše*.

## Zjavný redukcionizmus, biocentrizmus a antropocentrizmus v štúdiách Petra Sýkora

Z početnej skupiny autorov, ktorí sa v časopise *Filozofia* v období od roku 2014 až do roku 2022 venovali problematike ľudskej prirodzenosti a prípustných zásahov do nej, sa vymyká P. Sýkora ako autor, ktorý v rámci tejto témy publikoval najviac a v najkoncentrovanejšej podobe vyjadril myšlienkové poslanstvo, ktoré spája štúdie uverejnené v danom období. Nie náhodou sa preto stal aj editorom zborníka (2019b), v ktorom viaceré štúdie (napríklad tá od E. Višňovského) boli publikované v angličtine, rozšírené a doplnené o štúdie ďalších autorov v snahe prezentovať celému svetu pohľad stredoeurópskych bádateľov na problém ľudskej prirodzenosti a jej ochrany pred neprimeranými alebo nevhodnými zásahmi zo strany čoraz ambicióznějších biológov či bádateľov v oblasti umelej inteligencie. V tejto časti si budeme všimnúť najmä jeho vymedzenie ľudskej prirodzenosti a následne aj jeho názory na prípustné zásahy do nej – predovšetkým zo strany niektorých možno až príliš odvážnych genetikov.

V prvej štúdií (2014), venovanej polemike so Zuzanou Sitarčíkovou (2012), najprv uviedol jej kresťanskú a teleologickú definíciu ľudskej prirodzenosti ako „osoby v jednote duše a tela, totiž v jednote svojich tak duchovných, ako aj biologických sklonov i iných špecifických zvláštností nevyhnutných na dosiahnutie konečného cieľa“, ktorým nie je vykúpenie človeka z hriechov a jeho spasenie, ako by sme asi očakávali, ale rovnako ako u Stanislava Sousedíka „najdokonalejšie poznanie Boha a lásky z toho plynúcej“ (Sýkora 2014, s. 515). V polemike so Sitarčíkovou a ďalšími bádateľmi, ktorí podceňujú dôležitosť genetickej informácie „pri určovaní toho, čo robí človeka človekom“ (tamže, s. 517), potom zdôraznil, že sa očividne nechali pomýliť minimálnym (1,2 až 2-percentným) rozdielom „v sekvencii DNK človeka a šimpanza“. Ľudská prirodzenosť tak podľa nich spočíva v kultúre a duchovnosti človeka.

P. Sýkora preto kritizuje Sitarčíkovej snahu nájsť takú koncepciu ľudskej prirodzenosti, ktorá by bola univerzálna a nemenná, keďže v neustále sa meniacom fyzickom svete takéto kritériá nemôže spĺňať žiadna entita vrátane ľudskej bytosti (tamže), ktorá sa podľa Sitarčíkovej riadi princípom „jednoty tela a duše“. A následne celkom oprávnené tvrdí (Sýkora 2014, s. 518), že táto jej „metafyzická definícia ľudskej prirodzenosti je v debata o vylepšovaní človeka prakticky nepoužiteľná“; aj preto, lebo na jej základe „nie sme schopní rozlíšiť stav ľudskej prirodzenosti pred technologickým zásahom a po ňom“. Ale o pár riadkov ďalej sa už vynára hlavný problém, pretože zatiaľ čo P. Sýkora vyčíta Z. Sitarčíkovej, že vo vzťahu k ľudskej



prírodzenosti nerozlišuje medzi „identitou druhovou (to, čo definuje človeka ako *Homo sapiens*, ako určitý biologický druh) a identitou ľudského individua“, sám nielen v tejto štúdií, ale aj v ďalších štyroch (2015, 2018, 2019a, 2019b) nedokáže ani len rozlišovať medzi definovaním človeka a ľudstva ako *biologického* druhu a *prírodného* druhu, ktorý je (zatiaľ) na konci vesmírneho vývinového radu začínajúceho sa fyzikálnou formou pohybu hmoty.<sup>9</sup>

O silnom sklone P. Sýkora chápať a vymedzovať ľudskú prírodnosť *čisto biologicky* pritom svedčia aj ďalšie kritické pasáže (tamže, s. 519), kde jednak vyčíta Z. Sitarčíkovej, že veľmi podcenila stabilitu ľudskej genetickej inhomogenity, ktorá sa v prípade nášho biologického druhu za posledných 150 000 rokov „nijako podstatne nezmenila“ (Sýkora 2014, s. 519), jednak naznačuje možnosť *chápať našu genetickú výbavu ako našu prírodnosť*, pretože, ako tvrdí, kromaňonci mali tie isté geneticky dané psychologické predispozície „ako moderný človek“ a po príslušnom vzdelaní by boli schopní „riešiť diferenciálne rovnice tak ako dnešní ľudia“. Za zmienku ešte stojí, že Sýkora v tejto štúdií označuje ako triviálny a kompromisný názor, podľa ktorého sa na ľudskej prírodzenosti podieľajú (tamže, s. 523) „zhruba rovnakým dielom ako biológia (gény), tak aj kultúra (sociálne prostredie)“, pričom vzápätí odvážne konštatuje, že *v rámci transhumanizmu* je človek „pri všetkej svojej jedinečnosti a duchovnosti“ predsa len „definovaný ako určitý biologický druh“, a tak zmena človeka na posthumánny druh môže znamenať len zmenu biologickú, nie však kultúrnu, čo je však krajne diskutabilné tvrdenie, pretože Sýkora nikde neuvádza parametre tohto nového „biologického“ posthumánneho druhu ani nehovoriac o tzv. uploadingu, t. j. transhumanistami predpokladaného transferu ľudskej mysle do počítača, resp. iného technického zariadenia *aj s príslušnými ontickými dôsledkami*.

O rok neskôr, vo svojej ďalšej štúdií (2015), venovanej etike biotechnologických zásahov do ľudského genómu, Sýkora opäť konštatuje, že stále nevieme presne určiť, „do akej miery je ľudská prírodnosť determinovaná biologicky a do akej miery kultúrne“ (tamže, s. 339), ale znova bez toho, aby nám ponúkol aspoň nejakú základnú predstavu o výskume, na základe ktorého by sme to boli schopní zistiť. O ľudskom genóme pritom tvrdí, že je hodnotou sám osebe, spoločným majetkom celého ľudstva, ktorý treba chrániť rovnako ako genofondy iných biologických druhov. Zároveň ponúka Fukuyamovu de facto *nebiologickú* definíciu ľudskej prírodzenosti, pre ktorú je údajne určujúci akýsi faktor X, t. j. „neurčito definovaný ‚komplex interakcií našich jedinečných ľudských kvalít, akými sú morálna voľba, rozum a celá paleta rôznych emócií‘“ (Fukuyama 2002, s. 172). Podľa Fukuyamu je ľudská prírodnosť navyše univerzálna a môže byť preto na nej založená koncepcia *rovnako univerzálnych ľudských práv* (Sýkora 2015, s. 339). V dôsledku neprimeraných genetických zásahov by tak mohlo dôjsť nielen k nezvratným zmenám ľudskej prírodzenosti, ale aj k zrušeniu koncepcie všeobecných ľudských práv a následne aj demokratických inštitúcií chrániacich ľudskú dôstojnosť.

O štyri roky neskôr, vo svojej ďalšej štúdií uverejnenej v časopise *Filozofia* (2019a), už Sýkora jasne vyjadril svoj postoj k ďalším možnostiam zdokonaľovania človeka, keď typicky biocentricky uzavrel, že „vznik posthumánneho človeka môžeme skôr očakávať ako výsledok cieleňého génového dizajnovania *v oblasti biológie*, než ako výsledok tzv. uploadovania mysle do *nebiologického média v oblasti informačno-komunikačných technológií*“ (tamže, s. 513; podč. R. B.), a ďalším bádateľom odporučil, aby pri diskusiách o posthumanizme zamerali svoju pozornosť hlavne týmto smerom. Rovnaký názor prezentoval a obhajoval aj v ním editovanom zborníku (Sýkora 2019b, s. 143), kde najprv podotkol, že o potenciálnom zrode

---

<sup>9</sup> V tom druhom prípade už do sociálnej formy pohybu hmoty alebo sociálneho prírodného druhu začleňujeme okrem biologického aj psychické, sociálne a kultúrne, pričom samotný *sociálny prírodný druh* porovnávame s ostatnými prírodnými druhmi v danom vývinovom rade, napríklad na základe ich štruktúrnych či stavebných prvkov, intenzity energetických a informačných tokov, časopriestorovej rozpriestranosti, dynamiky a pod.

posthumánných bytostí je možné vecne diskutovať len v tom prípade, ak ich budeme považovať *za nový biologický druh*, ktorý vznikol na základe pozmenenej genetickej dedičnej informácie. Okamžite tak musel znova čeliť problémom týkajúcim sa presného definovania ľudskej prirodzenosti, keďže, ako zdôraznil, vedci nielenže majú veľmi ďaleko k dosiahnutiu konsenzu o tom, čo vlastne tvorí ľudskú prirodzenosť, ale niektorí dokonca pochybujú o jej existencii vôbec a tak ako anitesencialisti a sociálni konštruktivisti ju považujú za náš výtvor.

Na opačnom konci názorového spektra zase stoja vedci (Sýkora 2019b, s. 144), ktorí chápu ľudskú prirodzenosť ako „súbor fundamentálnych vlastností a schopností, ktoré majú ľudia prirodzene a ktoré sú vo väčšej či menšej miere determinované geneticky“. Ľudská povaha, resp. prirodzenosť je takto podľa nich doslova zakódovaná v ľudskej DNK, na základe čoho je potom možné tvrdiť, že ak pozmeníme gény, ktoré sú dôležité pri konštituovaní ľudskej prirodzenosti, vytvoríme nový *posthumánnny* a súčasne *biologický druh*. Sýkora potom uvádza aj názory Adama Doležala (tamže, s. 144-145), podľa ktorého nie je ľudská prirodzenosť žiadnym sociálnym konštruktom, resp. subjektívnym výtvorom, ale „objektívne existujúcou entitou“ a pojmom, s ktorým dokážeme intuitívne pracovať aj napriek tomu, že ju stále presne definovať nedokážeme. Doležal pritom neodmieta malé zásahy do ľudskej zárodočnej línie, vrátane tých, ktoré smerujú k nášmu zdokonaľovaniu, pokiaľ nie sú spojené s revolučnými zmenami tejto línie, v dôsledku ktorých sa fundamentálne zmení ľudská prirodzenosť.

A hoci sa zdá, že toto v podstate antropocentrické stanovisko je Sýkorovi veľmi blízke, on sám vo viacerých svojich štúdiách (2015, 2019a) jednoznačne odmieta *akékoľvek zásahy* do ľudskej zárodočnej línie a pripúšťa len modifikácie DNK tzv. somatických buniek. V tomto kontexte tak môžeme jeho stanovisko označiť naozaj ako *biocentrické a antropocentrické*, aj keď so zohľadnením toho, že sám Sýkora tieto zásahy do ľudskej zárodočnej línie neodmieta apriórne, ale len s ohľadom na aktuálne dosiahnutý stav poznania danej predmetnej oblasti (Sýkora 2015, s. 338-341). Bez ohľadu na to sa domnievame, že P. Sýkora v uvedených štúdiách málo diskutuje a reflektuje psycho-sociálno-kultúrnu zložku ľudskej prirodzenosti a málo priestoru venuje aj tým formám zdokonaľovania človeka, ktoré súvisia s kybernetikou, informatikou a nanotechnológiami, hoci uznávame, že Sýkora má pravdu, keď tvrdí (2019a, s. 521), že v rámci genetických manipulácií vedci v súčasnosti postupujú oveľa rýchlejšie ako napríklad v rámci nanotechnológií potrebných pre naskenovanie kompletného funkčného stavu ľudského „mozgu pre potreby jeho skopírovania“ (angl. the whole brain emulation).

Lenže aj v tomto prípade Sýkora stále pracuje s *antropocentricky podfarbenou predstavou* o vytvorení *človeku príbuznej bytosti*, resp. entity a nie je schopný pripustiť, že ďalším článkom vesmírneho vývinového radu na našej planéte môže byť aj *úplne nová myseľ* (resp. *inteligentný autonómny agent*), ktorá nebude využívať biologický substrát, ale anorganickú hmotu nám v tejto chvíli len ťažko predstaviteľným či dokonca prijateľným spôsobom. V ďalších častiach tejto štúdie sa preto budeme venovať najprv psycho-sociálno-kultúrnej zložke ľudskej prirodzenosti, potom posúdime vzájomné vzťahy medzi hlavnými zložkami ľudskej prirodzenosti – mozgom, rečou, mysl'ou, vedomím, osobou a inteligenciou a nakoniec sa zamyslíme nad tým, nakoľko je reálne, že vo sfére umelej inteligencie skutočne vznikne alebo bude vytvorený *úplne nový prirodzený druh*, resp. *úplne nová forma pohybu hmoty*.

## Rozmanité podoby ľudskej prirodzenosti v pozemských spoločenstvách a kultúrach

Počas čítania jednotlivých štúdií v časopise *Filozofia*, zameraných na vymedzenie ľudskej prirodzenosti a oprávnenosti rôznych zásahov do nej, nás najviac prekvapilo, že ani v jednej z nich nebola explicitne spomenutá možno trochu staršia, ale skvelá kniha Leslie Stevensona

a Davida L. Habermana *Ten Theories of Human Nature* (1974, 1987 a 1998), aj keď na druhej strane je jej ignorovanie zo strany väčšinou *biocentricky orientovaných* bádateľov v časopise *Filozofia* takmer prirodzené. Stevenson a Haberman (ďalej aj S&H) totiž analyzujú práve ten rozmer ľudskej prirodzenosti, ktorí títo autori ponechali bokom, ak nie rovno odignorovali – t. j. rozmer psycho-sociálno-kultúrny. Pozoruhodná je aj metodologická precíznosť, s akou problém ľudskej prirodzenosti uchopili. A nemenej dôležité je aj ich sústavné zohľadňovanie individuálneho a sociálneho (resp. sociálno-kultúrneho) rozmeru ľudskej prirodzenosti.

Hneď na 3. strane tak S&H upozorňujú na obrovský význam tejto koncepcie (resp. pojmu) v živote *jednotlivca aj spoločnosti*, pretože na základe tohto pojmu a jeho špecifického, resp. *vlastného chápania* ľudia pripisujú význam a účel vlastnému životu, uvažujú nad tým, čo by mali robiť a o čo sa usilovať, koho treba obdivovať a kým sa chcú alebo môžu stať; zatiaľ čo ľudské spoločenstvá hľadajú vhodný model ľudskej komunity a na jeho základe premýšľajú nad zmenami, ktoré je v inštitucionálnom ohľade treba pre jej konštituovanie urobiť. Ako ďalej upozorňujú (S&H 1998, s. 4), pomocou pojmu „človek“ bežne poukazujeme na akéhosi obecného človeka, disponujúceho *prevažne mužskými* vlastnosťami, a to aj napriek tomu, že popri pomyselnéj mužskej povahe (resp. prirodzenosti) zjavne existuje aj ženská, detská či dokonca starecká (dovoľme si doplniť) povaha a pod. A takisto ako existujú rôzne biologicky či vekovo dané ľudské prirodzenosti, existujú aj rôzne prirodzenosti sociálno-kultúrne.

Navzájom súperiace teórie ľudskej prirodzenosti sa potom rozdielne aplikujú a manifestujú v odlišných spôsoboch života alebo v alternatívnych či navzájom si konkurujúcich politických a ekonomických systémoch. Niektoré z týchto teórií potom vyvíjajú na jedinca až extrémny tlak a nútia ho správať sa podľa ich základných princípov, čo však v oveľa menšej miere platí pre existencialistické filozofie a koncepcie človeka, vrátane tej Sartrovej (tamže, s. 5), ktoré akoby ospravedlňujú moderné liberálne demokracie tým, že im ponúkajú víziu, podľa ktorej nejestvujú žiadne objektívne platné hodnoty, podľa ktorých by sa ľudia mali riadiť, ale len individuálne a subjektívne (behaviorálne) voľby. S&H zároveň pripomínajú (tamže, s. 9), že systémy presvedčení o ľudskej prirodzenosti, ktoré sú akceptované určitými skupinami ľudí a rozhodujúcim spôsobom tak ovplyvňujú ich spôsob života, sú štandardne označované ako ideológie, čo platí aj pre existencialistické filozofie a koncepcie človeka, ktoré sa tak stávajú súčasťou „ideologickej základne politického liberalizmu“. Ako ďalej zdôrazňujú S&H, hoci sú ideológie z praktického hľadiska dôležitejšie ako teórie, musia zahrňovať nejakú teoretickú koncepciu ľudskej prirodzenosti, resp. povahy (tamže), aby mohli byť vierohodné a účinné.

Podľa S&H pritom všetky teórie ľudskej prirodzenosti, ktoré sú diskutované v ich knihe (vrátane konfuciánskej, hinduistickej, ale aj Platónovej, Kantovej, Freudovej, Skinnerovej či Lorenzovej), zdieľajú s kresťanstvom a marxizmom spoločné štruktúrne prvky alebo hlavné znaky, to jest: 1. rámcovú teóriu sveta ako takého, 2. základnú teóriu ľudskej prirodzenosti, 3. diagnózu toho, čo je v nás chybné a 4. návod, ako to odstrániť a napraviť. S&H sú dokonca presvedčení, že len teórie ľudskej prirodzenosti, ktoré kombinujú tieto štruktúrne prvky, nám dávajú nádej, že niekedy vyriešime hlavné problémy ľudstva (tamže). Pokiaľ však ide o mieru ich prijatia, resp. rešpektovania, S&H na jednej strane tvrdia (tamže, s. 13), že aj keď niektoré teórie ľudskej prirodzenosti môžu byť ľuďmi vnímané a realizované s určitými odchýlkami aj napriek tomu, že tvoria koncepčne uzavreté systémy, v prípade, že sa niektoré presvedčenie stane ideológiou, je pre členov komunity, ktorá v rámci nej koná, veľmi ťažké posudzovať ju objektívne. Zároveň ale dodávajú (tamže, s. 14), že aj vtedy, keď sa presvedčenie zmení na ideológiu predstavujúcu koncepčne uzavretý myšlienkový systém, jeho racionálne posúdenie je stále možné, pokiaľ sa nájdú jedinci, ktorí sú na racionálnu a vecnú diskusiu o jednotlivých presvedčeniach a ideológiách pripravení. Pravdaže, len v prípade, ak sa vyhneme *kultúrnemu relativizmu*, podľa ktorého sú si všetky kultúry epistemicky (alebo hodnotovo) rovnocenné.

Rovnako dôležité je (S&H 1998, s. 14), že v takomto prípade vždy môžeme odlíšiť to, čo niekto hovorí, od motivácie hovoriť práve to, čo hovorí. Zistíme pritom (tamže, s. 16), že vyhlásenia o ľudskej prirodzenosti sú zvlášť poznačené tým, čo by ľudia chceli, ako hodnotia veci, ale netýkajú sa často toho, ako sa veci naozaj majú, resp. aké sú. Najmä výroky, ktoré sa týkajú ľudskej prirodzenosti, resp. toho, čo je pre človeka normálne a žiaduce, bývajú zvlášť poznačené touto epistemickou dvojznačnosťou. Už samotné použitie pojmov „prirodzenosť“ alebo „prirodzený“ nám takto signalizuje „možnú dvojznačnosť alebo zmätok“. Po vstupnej rámcovej kapitole už nasleduje detailný rozbor desiatich teórií ľudskej prirodzenosti, ale nám stačí, ak sa s ohľadom na náš epistemický a ideologický kontext budeme bližšie venovať len Marxovej alebo marxistickej teórii ľudskej prirodzenosti, ktorú S&H mimoriadne objektívne a inšpiratívne posudzujú nielen v 7. kapitole ich knihy, ale aj na iných miestach ich textu.

Podobne ako väčšina bádateľov (tamže, s. 137) spochybňujú Marxove predstavy o existencii deterministických zákonov v ľudských dejinách, ktoré sú tak akoby predurčené a smerujú k dopredu zjavnému alebo „zadanému“ cieľu. Na druhej strane však podotýkajú, že Marxova základná myšlienka, podľa ktorej spôsob produkcie základných statkov na každom stupni ekonomického vývoja silne vplyva na spôsob, akým ľudia na danom stupni ekonomického vývoja myslia (tamže), sa ukázala metodologicky mimoriadne plodná v rámci historického, antropologického, sociologického i politologického bádania. Všeobecne známa je aj Marxova definícia človeka (a tak aj ľudskej prirodzenosti) ako súhrnu jeho sociálnych vzťahov (tamže, s. 140), z čoho podľa S&H jasne vyplýva, že Marx nebral veľký ohľad na biologický rozmer ľudskej prirodzenosti, pretože v dôsledku neustálej zmeny výrobných síl a vzťahov možno podľa neho celé dejiny pochopiť a vykladať ako „kontinuálnu transformáciu ľudskej povahy“. Dokonca aj spôsoby či „techniky“, ktorými „produkuje potraviny a vychovávame deti, sú sociálne naučené“ (tamže). Ľudská spoločnosť tak podľa Marxa nie je žiadnou abstraktnou entitou, ale konkrétnym celkom alebo rámcom, ktorý (pred)určuje, akým jedincom sa človek môže stať a aký druh vecí môže robiť; jeho možnosti sú teda historicky dané a limitované.

Podľa S&H však predsa len existuje jeden parameter ľudskej prirodzenosti, ktorý platí podľa Marxa všeobecne a univerzálne, a to naša definícia ako „aktívnych, produktívnych bytostí“, ktoré sa principiálne odlišujú od zvierat tým, že naše zdroje obživy produkuje na základe *vedomého* plánu, resp. cieľa a nie inštinktívne ako napríklad včely. Rovnako známe a dôležité sú aj slabiny a zjavné rozpory Marxovej dejinnej a prognostickej koncepcie, keď S&H vecne namietajú (tamže, s. 143), že ak je podľa marxizmu štát zdrojom všetkého zla, potom môže „komunistické“ zoštátnenie výrobných prostriedkov viesť k ešte väčšiemu vzrastu jeho moci a následne aj sociálneho útlaku, ako sa potvrdilo v bolševických režimoch 20. storočia. Na druhej strane, ani alternatívna vízia budúcej komunistickej spoločnosti ako decentralizovanej society, pozostávajúcej z nezávislých komún (v rámci ktorých je možné realisticky uvažovať o zákaze peňazí, špecializácie či súkromného vlastníctva), nevznieva najmä z ekonomického hľadiska presvedčivo, pretože zoskupenie nezávislých komún by sotva dokázalo zorganizovať produkciu a distribúciu statkov na tej úrovni a v tej intenzite, aká je (zatiaľ) bežná v globálnej kapitalistickej spoločnosti. No a rovnako naivne vzniká aj Marxova predstava o budúcej komunistickej spoločnosti ako entite, v ktorej bude možné úplne eliminovať konflikt záujmov medzi všetkými zainteresovanými subjektmi (S&H 1998, s. 146). Z hľadiska nášho ďalšieho výkladu je napokon veľmi dôležité aj tvrdenie, že Marx (podobne ako Engels) nebol filozof a ako materialista sa viac zaujímal o dejiny, ako o vzájomné vzťahy medzi mentálnymi stavmi a mozgom (tamže, s. 139). Nie je teda žiadny dôvod pripisovať mu názor, že ľudské vedomie je možné (či dokonca nutné) priamo stotožniť s procesmi ľudského mozgu (tamže, s. 140).

## Hlavné zložky ľudskej prirodzenosti – ľudský mozog

Keď človeka definujeme ako bio-psycho-sociálno-kultúrnu bytosť (alebo entitu), zaradujeme ho tak do vesmírneho vývinového radu, ktorý je otvorený tak smerom do minulosti, ako aj budúcnosti. Aby však mohol byť človek takouto vysoko vyvinutou entitou, musí podľa nás disponovať (špecificky ľudským) mozgom, rečou, mysl'ou, vedomím, osobou a inteligenciou, ktoré práve preto označujeme ako hlavné zložky ľudskej prirodzenosti, bez ktorých by sa človek ako taký nemohol stať takouto komplexnou entitou. Je zaujímavé i zvláštne, že žiadny z autorov štúdií o ľudskej prirodzenosti v časopise *Filozofia* nepovažoval za potrebné bližšie charakterizovať človeka ako takého (t. j. ako *prirodzený, nie biologický* druh) na základe týchto vlastností. Na druhej strane nás to veľmi neprekvapuje, pretože vo vedeckej, resp. filozofickej obci stále nepanuje ani len približný konsenzus ohľadom presného definovania týchto entít, pričom najmä pojmy myseľ a vedomie sa bežne používajú ako synonymné, hoci je zrejmé, že označujú principiálne odlišné entity, so zásadne odlišným obsahom i funkciou.

Aj preto sa v ďalšom výklade budeme snažiť zachytiť a zvýrazniť *skôr rozdiely* medzi týmito entitami ako aktuálny stav ich poznania. Začneme mozgom, ktorý v jeho špecifickosti veľmi precízne i zrozumiteľne opísal už dávnejšie Ladislav Kováč (2008)<sup>10</sup>. V prvom rade musíme zdôrazniť, že Kováč považuje mozog za *biologický orgán*, resp. *systém*, pre ktorý je dôležitá *biochémia* a nie *chémia*, pretože, ako uvádza na 101. strane: „Chemické procesy v neživom svete majú skalárny charakter, bežia v čase, ale nie v priestore, zatiaľ čo chemické procesy v živých systémoch sú vektoriálne.“ Mozgu a jeho fungovaniu tak môžeme porozumieť len vtedy, ak budeme pri jeho modelovaní brať do úvahy vektoriálnu povahu jeho biochemických procesov (tamže, s. 102). Biochemická rovina je však len jednou z troch rovín fungovania mozgu. Nad ňou sa nachádza fyziologická rovina a celkom navrchu rovina subjektívneho prežívania a vedomia, pričom podľa Stevena Rosea „medzi biochemickým a fyziologickým opisom mozgu“, rovnako ako „medzi biochemickým a psychologickým opisom nie je vzťah kauzálny, ale mapovací“. Ide teda akoby o preklad medzi tromi hlavnými ontologickými rovinami mozgu, t. j. „o preklad z jedného jazyka do druhého“. To má mimoriadne vážne dôsledky, pretože vyššie roviny (alebo úrovne) mozgovej činnosti nemôžu potom „kauzálne pôsobiť na nižšie“ alebo, inak povedané, myslenie *nemôže mozog ovplyvňovať ani riadiť*.

L. Kováč dokonca ide tak ďaleko (2008, s. 103), že ideu fungovania mysle v nebiologickom substráte označuje za pomýlenú a spolu s ňou aj predstavu o mozgu ako výpočtovom stroji. V rámci svojho detailného a všestranného opisu genézy ľudského mozgu potom vyzdvihuje prechod našich dávnych predkov z konárov stromov do náročných podmienok savany, počas ktorého (tamže, s. 108) potreba jemnej manipulácie rukami poháňala rast mozgu a zvyšovala ľudské kognitívne schopnosti. Už pred 1,5 miliónmi rokov bol tak mozog hominidov trikrát väčší ako mozog veľkých opíc. A keďže ďalší rast mozgu mal svoje biologické hranice, dané veľkosťou ženskej panvy, veľká časť jeho vývoja sa musela realizovať po narodení dieťaťa (resp. mláďaťa) v rámci kultúrnej evolúcie umožnenej prednostne rečou, resp. inou formou postupne čoraz vyspelejšej komunikácie. Kováč preto tvrdí (tamže, s. 109), že človek sa stal tak biologickou, ako aj kultúrnou bytosťou z anatomickej a fyziologickej nevyhnutnosti.<sup>11</sup>

<sup>10</sup> Pravdaže, vedecká literatúra o ľudskom mozgu a jeho fungovaní je už dnes natoľko obsiahla, že nie je v silách žiadneho bádateľa poznať ju detailne. O to užitočnejšie sú prehľadové štúdie alebo články, ktoré upozorňujú na hlavné výskumné problémy v tejto oblasti, ako napríklad nedávny článok v *Nature* (z 22. 8. 2023) o ukončení *The Human Brain Project*, v rámci ktorého sa európskym bádateľom nakoniec nepodarilo dosiahnuť jeho hlavný cieľ – simuláciu činnosti celého ľudského mozgu, aj napriek zaujímavým čiastkovým výsledkom.

<sup>11</sup> Čo je pravdou len v prvom priblížení, pretože človek sa stal kultúrnou bytosťou nie na základe prirodzeného výberu, ako mylné tvrdia celé zástupy vedcov a filozofov v tzv. populárno-náučných publikáciách, ale v rámci

Aj tisíce senzorov, ktorými disponujeme, sú podľa Kováča chemickými zariadeniami (tamže), pričom údaje z nich nakoniec „rezultujú len v dvoch rozdielnych druhoch akcie“ – v útoku alebo v úteku, kým rozhodnutie o tom, ktorú akciu si zvolíme, vyplýva zo zovšeobecnenia konkrétnych údajov senzorov v dvoch podobách – dobré alebo zlé. Vďaka chémii sa potom „do dvojice senzoria-akcia“ včleňuje „tretí komponent činnosti mozgu: emócie“ (tamže, s. 110). Hormóny a tzv. emotóny nám totiž umožňujú subjektívne prežívať (opäť binárne) chemické nastavenie, ktoré má už známu zovšeobecnú podobu niečo dobrého (a príjemného) alebo naopak zlého (a nepríjemného či bolestivého). Mimoriadne dôležité je aj Kováčovo rozlíšenie medzi dvomi základnými druhmi prenášania informácie v mozgu – tzv. *drôtovým prenosom* (angl. wire transmission), ktorý sa realizuje medzi dvomi synapsami, a tzv. *objemovým prenosom* (angl. volume transmission), v rámci ktorého sa prenášajú chemikálie a informácie uvoľnené síce len jedným neurónom, ale zasahujúce „viaceré neuróny, prípadne globálne celý mozog cez medzibunkový priestor v mozgu“ a cerebrospinálny likvor (tamže).

Mozog sa preto podľa neho viac ako počítaču podobá „špongii ponorenej do koncentrovaného roztoku tonicky aktivujúcich chemikálií“ (tamže). Kováč dokonca ide tak ďaleko, že zmes chemikálií v našom mozgu označuje ako nášho absolútneho vládcu, ktorému neuróny našej mozgovej kôry, pospájané do neurónových a synaptických sietí, len poslušne slúžia. Celkový obraz potom už dotvára len Darlingtonova teória „evolučného zosilnenia“ (angl. evolutionary reinforcement; 1975), podľa ktorej síce príjemnosť nemusí byť vždy „výhodná z hľadiska darvinovskej fitness, no zosilňuje selekciu výhodného (t. j. pre jedinca príjemného – R. B.) správania“, čo zase podľa Kováča (2003) vedie k tomu, že ľudia sa prirodzene snažia vyhýbať trápeniam a hľadať skôr príjemné zážitky, a to v čoraz väčšej miere s pomocou (technických) artefaktov. V dôsledku toho u ľudí evolúciu prirodzeným výberom najprv vytlačá a potom nahrádza tzv. paravolúcia, pretože aj keď sa (súčasní) ľudia biologicky menia, tieto ich zmeny už v *globále* nie sú súčasťou darvinovskej evolúcie prirodzeným výberom. Kováč pokračuje precíznym opisom nášho imunitného systému, ale pre nás je dôležitý skôr jeho opis vývinu mozgu a centrálnej nervovej sústavy (t. j. neurogenézy) s jej tromi kategóriami chemických stavov (tamže, s. 115), ktoré zásadným spôsobom ovplyvňujú a predikujú naše správanie.

Najrozmanitejšie chemické stavy v našom mozgu možno totiž podľa Kováča najprv rozdeliť „na stavy, ktoré určili gény na základe fylogenetickej skúsenosti druhu“ a sú preto spoločné všetkým ľuďom. Do druhej kategórie už patria stavy, ktoré sú špecifické pre každého jedinca a počas jeho života „sa môžu zmeniť len málo“, pretože sa „vytvorili imprintingom v dvoch imprintingových kritických obdobiach“. Napokon do tretej kategórie patria stavy, ktoré náš mozog síce generuje nanovo, ale aj v tomto prípade „signály z prostredia neindikujú niečo nové a sebe adekvátne, ale vyberajú z toho, čo je k dispozícii“. Signál z prostredia tak nenesie informáciu, ktorú treba (vcelku) spracovať, ale funguje skôr ako spúšťač, pomocou ktorého mozog „vyberá z predpripravených procesov a jeden, ten vybraný (t. j. najvhodnejší – R. B.), spúšťa“ (Kováč 2008, s. 115). Ešte zaujímavejšie je z hľadiska nášho ďalšieho výkladu to, že práve toto selektovanie a spúšťanie „toho, čo je už v mozgu hotové, predpripravené“ pomáha sčasti pochopiť, prečo je náš mozog taký rýchly, keď sú neuróny pomalé (tamže, s. 115-116).

V našom mozgu máme totiž podobne ako iné živočíchy (ale zjavne v oveľa väčšom počte) uložené moduly dôležitých kategórií objektov a správania, z ktorých si rýchlo vyberáme v tej či onej životnej situácii, a to bez toho, aby sme danú situáciu museli podrobiť dôkladnému či detailnému skúmaniu. Pri vizuálnom poznávaní si napríklad mozog vyberá „z veľkého počtu veľmi jemných modulov“, čo platí aj pre mačky, pri ktorých sa podľa Hubela a Wiesela jasne

---

*špecifického procesu*, ktorý už dávnejšie dostal názov *antroposociogenéza* a v ktorom postupne zohrávali čoraz väčšiu úlohu *nebiologické*, t. j. *sociálne a kultúrne* faktory, ako uvádzame napríklad v štúdiu (2012, s. 19-35).

potvrdilo, že ich „vizuálny systém, má hotové moduly horizontálnych a vertikálnych tyčiek a hrán, farieb či iných vlastností“ (1982), z ktorých si potom akoby poskladá výslednú podobu vonkajšieho objektu, pričom v skutočnosti ide o jeho obraz skonštruovaný v mozgu mačiek z dopredu pripravených, akoby prefabrikovaných modulov (Kováč 2008, s. 116).<sup>12</sup> Krajne zaujímavá a zároveň správna nám tiež pripadá Kováčova poznámka (tamže, s. 117), podľa ktorej máme v našom mozgu akoby dopredu pripravený aj repertoár pojmov, ktorý si počas nášho života už len upresňujeme, pričom táto schopnosť „detailného pojmového chápania javov iba odráža našu schopnosť uchopiť rukou predmety takmer ľubovoľných tvarov, vďaka výnimočnej flexibilitě ľudskej ruky“ (tamže) s palcom postaveným oproti ostatným prstom.

Osobitnou kapitolou je energetická bilancia mozgu, režim spánku a jeho funkcie, spočívajúce nielen v usporiadaní či „vyhodnotení“ bdelych zážitkov, ale aj príprave na budúce voľby či opcie (tamže, s. 118-120), no pre nás je z ontologického hľadiska skôr dôležitá vízia mozgu ako motora, ktorý je v pohybe aj vtedy, keď auto stojí (a človek spí alebo je v stave denného snenia), pričom „krv, ktorá mozgom intenzívne preteká“, prináša doň „nielen pohonné látky a mazačí olej, ako je tomu v dvojtaktnom motore, ale je zároveň chladiacou kvapalinou“. Na rozdiel od počítača, ktorý môžeme kedykoľvek vypnúť alebo zapnúť, je teda mozog vďaka svojej vysokej endogénnej aktivite v stave „trvalej pohotovosti a na vonkajšie stimuly dokáže promptne reagovať“. Ak by sme teda mozog označili provizórne ako stroj, ktorý verne a bez prestávky slúži konkrétnemu ľudskému subjektu, ktorý ho „vlastní“, potom by to bol podľa Kováča (tamže, s. 121) jednocelový chemický stroj, ktorý zaisťuje „udržanie a rozmnoženie svojho nositeľa, individuálneho organizmu“. A to nie je všetko – chemické procesy v mozgu „sú jeho sémantikou“, zatiaľ čo počítače sú univerzálne, *syntaktické* stroje, ktoré ešte stále fungujú a pracujú len „na základe programov a vstupov, ktoré do nich vkladá ľudský subjekt“.

Ešte priamočiarejší bol David Gelernter (2007), ktorý takisto odmietol stotožňovanie mozgu s počítačom, pretože počítač je podľa neho „ľahostajný k tomu, aké inštrukcie vykonáva“, zatiaľ čo ľudia cez mozog prežívajú svet a precitujú emócie, a chémia nás tak skutočne „robí inými než sú počítače“ (Kováč 2008, s. 122). O zásadnej nesúmerateľnosti (prirodzených) chemických procesov v ľudskom mozgu a digitálneho počítania svedčí podľa Kováča takisto aj spôsob, akým sa jednoduchý proteín dokáže spontánne zvinúť do svojej pôvodnej štruktúry „počas niekoľkých milisekúnd alebo nanajviš sekúnd“ (tamže), a to vďaka *minimalizovanej energetickej interakcii medzi príslušnými aminokyselinami*. Vtedajšie najvýkonnejšie počítače však dokázali simulovať len prvú časť „prirodzeného procesu zvinutia, niekoľko desiatok nanosekúnd, o šesť rádov menej v porovnaní s tým, čo dokáže príroda“. Podľa všetkého aj preto, lebo v nich nemohli byť zabudované *atraktory*, ktoré umožňujú mozgu ako prednostne *dynamickému systému* vykonať tak veľa s minimom voľnej energie. Stotožňovanie ľudského mozgu s počítačom je teda naozaj pomýlené, a to aj v podobe veľmi voľnej „metafory“.

## Hlavné zložky ľudskej prirodzenosti – (ľudská) reč

V tejto časti nebudeme „špekulovať“ nad tým, ako a kedy vo vesmírnom vývinovom rade vznikla ľudská reč (najmä ak sme tak urobili v už spomínanej štúdii z roku 2012, s. 19-35), a všimneme si skôr, ako a kedy sa reč objavuje v ľudskej ontogenéze a akú úlohu zohráva pri konštituovaní *robustnej perspektívy I. osoby*, zrejme najdôležitejšieho ľudského parametra. Najprv budeme vychádzať z extrakraniálnych ideí pôvodu ľudskej mysle a vedomia, tak ako ich prezentovala Jennifer Greenwoodová, podľa ktorej (2015, s. 6) od 60. rokov minulého

---

<sup>12</sup> Mika Suojanen (2015) má teda zrejme pravdu, keď tvrdí, že vonkajšie objekty nevnímame priamo, t. j. také, aké sú, ale medzi ne a nás vkladáme *ich obraz* vytvorený v našej myslí, resp. našou myslou.

storočia vieme, že najdôležitejším faktorom vo vývoji špecificky ľudských spôsobilostí, resp. kapacít sú ľudským „jazykom sprostredkované sociálne vzťahy“, bez ktorých *nie je možná* postupná premena novorodencov a detí na dospelé osoby. Obidve fázy komunikácie ako takej (t. j. produkovanie aj prijímanie informácií) sú pritom vysoko citlivé na kontext ich realizácie, na základe čoho je možné tvrdiť (tamže, s. 7-8), že „ľudská emocionalita, reč a myslenie sú v rozhodujúcej miere naučenými spôsobilosťami“. Základom našej ontogenézy je tak podľa Greenwoodovej (tamže, s. 10) „extrémne hlboká a komplexná funkčná integrácia neurálnych a sociokultúrnych zdrojov“, v rámci ktorej sú budované naše čoraz sofistikovanejšie mozgy.

V polemike s intrakranialistami potom J. Greenwoodová zdôrazňuje (tamže, s. 18), že emócie nefungujú primárne na individuálnej rovine ako kognitívne či telesné pocity vyjadrujúce, resp. sprostredkujúce „nástroje“, ale ešte pred vyslovením prvých slov sa u detí uplatňujú ako „ostenzívno-expresívne nástroje“, ktoré umožňujú v čoraz väčšej miere riadiť (ich) sociálny život najprv cez interpersonálnu (vyslovene externú) a následne cez intrapersonálnu (vnútornú alebo „zvnútornenú“) emocionálnu reguláciu. Ako ďalej upozorňuje Greenwoodová (tamže, s. 115), rozhodujúcim faktorom je tu skutočnosť, že novorodenci a malé deti sa učia hovoriť *v rámci toho istého procesu*, v rámci ktorého sa učia vyjadrovať a realizovať *aj svoje emócie*, a to počas *napodobňovania* pohybov (resp. *mimiky*) a *zdĺhavých*, permanentne sa opakujúcich *praktík s interaktívnymi opatrovateľmi*. Základným predpokladom tohto typu učenia sa pritom zdá byť (tamže, s. 133) limitovanosť vizuálnej percepcie kojencov ohľadom ich schopnosti rozlišovať objekty a udržať ich v pozornosti (v spojení s recipročnou vrodenuou spôsobilosťou ich matiek držať tvár vo vzdialenosti 20-24 cm od tváre kojenca), ktorá rezultuje do vzniku skúsenostného sveta pozostávajúceho takmer výlučne z matkinej (resp. otcovej) tváre.

Počiatky učenia sa jazyku a emóciám (t. j. ich spracovaniu a vyjadrovaniu) sa tak realizujú už počas kojenia a sania materského mlieka (alebo mlieka z fľaše), kedy dieťa striedavo počúva a následne vyjadruje svoje pocity (pohybmi, mimikou alebo vokalizáciou), pričom na nižšej fyziologickej rovine je táto jednoduchá, *ale sústavne sa opakujúca* vokalizácia a gestikulácia sprevádzaná u dieťaťa tými istými endogénnymi opioidmi a neuropeptidmi ako počas súčasne prebiehajúcej emocionálnej ontogenézy (Greenwoodová 2015, s. 112). Ľudský jazyk (resp. reč) je zároveň nositeľom *obsahu komunikácie*, ako vyplýva z druhovo špecifických vzorov správania, kedy je vokalizácia spočiatku oddelená od vyjadrovania emócií, pretože až do dvoch rokov veku malé deti striedavo buď „rozprávajú“, alebo vyjadrujú svoje emócie (angl. talk or emot; tamže, s. 121). Až s nástupom *slovného špurtu* zhruba v dvoch rokoch veku dieťaťa (tamže, s. 124) sa konečne začínajú jeho rečové prejavy spájať s emočnými podnetmi.

Kľúčom k spolupráci a zdokonaľujúcej sa komunikácii matky a dieťaťa je zase tzv. *zdieľaná pozornosť* (angl. joint attention), ktorá je takisto špecifická pre náš (biologický) druh (tamže, s. 119). Reč je dokonca vstupnou bránou aj do vizuálneho vnímania alebo percipovania nášho sveta, ako vyplýva z výskumov Spiveyho a Richardsona (2009), ktorí realizovali veľký počet experimentov zameraných na sledovanie očných pohybov dieťaťa (angl. eye-tracking), ktoré demonštrovali, že vizuálny svet je u detí kontinuálne sprístupňovaný a integrovaný na základe *spracovania rečových podnetov*, a to na niekoľkých vývinových úrovniach (Greenwoodová 2015, s. 128-129). Vyššie spomínané ostenzívne zručnosti (tamže, s. 130) sa takto získavajú na začiatku ontogenézy a v konečnom dôsledku tak tvoria akési „lešenie“, ktoré umožňuje „nadobudnutie prirodzeného jazyka“ (tamže, s. 131), pričom hlavnou zložkou „simultánneho rozvoja plne ľudskej emocionality, intencionality a jazyka (resp. reči – R. B.) sú druhovo špecifické vzory správania a kultúrne sprostredkovanie (angl. scaffolding)“ (tamže, s. 132).

Komunikácia, interakcia a sociálne vzťahy sú napokon extrémne dôležité *v každom veku*. Ako podotýka Andy Clark (2001, s. 154), ľudská inteligencia ako taká sa neukrýva ani v ľudskom mozgu, ani v moderných (komunikačných, výpočtových a pod.) technológiách, ale rozvíja sa



v rámci „komplexných a opakujúcich sa interakcií a spolupráce medzi nimi“. Výskum týchto interakčných prienikov je pritom mimoriadne náročný, pretože si vyžaduje interdisciplinárny prístup „a nové formy modelovania a analýzy“ (tamže). Jeho výsledky sú však neraz doslova šokujúce, ako sa ukázalo počas interdisciplinárneho výskumu Barniera a jeho kolegov (2012), ktorí zistili, že „25 percent partnerov žijúcich doma v rámci dlhodobých vzťahov, ktorí mali stupeň neurálneho rozpadu normálne diagnostikovaný ako Alzheimerova choroba, zostáva bez jej vonkajších symptómov“, pretože svojich „dlhodobých partnerov využívajú ako externé pamäťové zdroje“ (Greenwoodová 2015, s. 211-212). Greenwoodová preto podľa nás na záver oprávnené konštatuje, že „stupeň a komplexnosť funkcionálnej integrácie neurálnych, telesných a sociokultúrnych zdrojov v ľudskom vývoji sú také“ vysoké, že zjavne potvrdzujú nielen extrakranialistickú hypotézu o prieniku vonkajšieho sveta a mozgu, ale aj definitívne pochovávajú dichotómiu biológie a kultúry v ľudskej ontogenéze a vývoji (tamže, s. 212).

Podľa Lynne Rudder Bakerovej (2013, s. 130 a n.) reč zároveň zohráva mimoriadne dôležitú úlohu pri konštituovaní robustnej perspektívy 1. osoby, ako *najdôležitejšej ľudskej vlastnosti*, (resp. *parametra*), ktorou sa ľudia *principiálne* odlišujú aj od najviac vyvinutých živočíchov, ktoré podľa Bakerovej majú len obyčajné alebo „prvostupňové“ vedomie, pretože disponujú len rudimentárnou perspektívou 1. osoby, a nie „seba-vedomie“ alebo sebauvedomenie ako také. Keďže tejto problematike sa budeme viac venovať v ďalšej časti textu (venovanej osobe a jej identite), tu obrátíme našu pozornosť hlavne na reč *ako differentia specifica človeka*, a to vo vzťahu k problematike genézy „plnohodnotného“ či autentického ľudského subjektu.

Ako uvádza Bakerová (2013, s. 130), novorodenci sa síce „rodia s perceptuálnymi systémami, ale musia sa naučiť hovoriť“. Samotné učenie sa jazyku je však podľa nej *možné len v ľudskej spoločnosti*, pretože, ako vyplynulo z Tomasellových nedávnych výskumov, pri testovaní dvojročných detí a dospelých orangutanov a šimpanzov, jedinou oblasťou, v rámci ktorej tieto deti prevýšili neľudských primátov, boli „sociálne zručnosti: sociálne učenie, komunikácia a čítanie (resp. odhadovanie – R. B.) zámerov iných bytostí“ (tamže). Neprekvapuje preto, že robustná perspektíva 1. osoby a sebauvedomenie sa objavujú až potom (tamže, s. 131), ako novorodenec prejde množstvom sociálnych a rečových interakcií. Kľúčom k jeho postupnému uvedomeniu si druhých ľudí ako (seba)vedomých bytostí je pritom aj podľa Bakerovej vyššie spomenutá *zdieľaná pozornosť*, v rámci ktorej dojča svojimi očami doslova sleduje pohľad svojej matky. A keď dojča postrehne rozdiel medzi jeho vlastnou pozornosťou a matkinou, začína ju vnímať ako vedomú bytosť. Zdieľaná pozornosť vo svojich rozmanitých podobách tak skutočne vystupuje ako „nutný prekursor učenia sa jazyku (resp. reči – R. B.; tamže)“.

Veľmi dôležité je aj Bakerovej rozlišovanie medzi myšlienkami a konceptmi, podľa ktorého sú koncepty individualizovanými myšlienkami, pričom slovo „myšlienka“ v tomto prípade nadobúda veľmi široký význam „a referuje *k akémukoľvek mentálnemu stavu* s propozíčnym obsahom – presvedčeniam, nádejám, obavám, želaniam, zámerom atď.“ (tamže; podč. R. B.) Identita konceptov, resp. udržiavanie ich (aspoň približného) obsahu v prúde reči je pritom určená a zabezpečená ich aplikačnými podmienkami, zatiaľ čo identita konceptov na druhej strane „určuje identitu myšlienok, ktorých sú koncepty konštituentami“ (tamže, s. 132), čo sa zdá byť v takmer dokonalom súlade s intersubjektívnym rázom ľudskej komunikácie. V danej súvislosti sa takto potvrdzujú zistenia Lery Boroditskej, podľa ktorej (2011, s. 65) jazyk ako taký „skutočne hrá kauzálnu rolu v tvarovaní kognície“. So zmenou spôsobu, akým ľudia hovoria (v tom či onom národnom jazyku), potom paralelne prichádza aj zmena v spôsobe, akým myslia. Aplikácia konceptov v takejto permanentne sa vyvíjajúcej komunikácii je teda vymedzená ľudskými rečovými a sociálnymi zručnosťami (Bakerová 2013, s. 133).

Bodom zvratu vo vzťahu k sebauvedomeniu alebo „sebakonceptu“ je pritom podľa Bakerovej druhý rok veku dieťaťa, keď sa popri konceptoch s empirickým obsahom náhle vynoria práve

tie (modálne) koncepty, ktoré súvisia s vlastnými želaniami a túžbami dieťaťa, ako je zrejme napríklad z vety: „Želám si, aby som mal nákladné auto.“ (tamže, s. 136) V takomto prípade sa totiž zdá byť zrejme, že dieťa začína jasne rozlišovať medzi sebou a okolitým svetom aj so všetkými s tým súvisiacimi dôsledkami ohľadom jeho následnej sebareflexie, sebahodnotenia a pod. Jeho rečové prejavy sa potom obracajú akoby dvomi smermi – jednak navonok (keď jeho reč nadobúda verejný charakter), jednak dovnútra, keď dieťa začína analyzovať, resp. posudzovať vonkajší svet v rámci svojho vnútorného myslenia (a vnútornej reči). Bakerová preto celkom odôvodnene zdôrazňuje (tamže, s. 137), že verejnou rečou môže disponovať len ten jedinec, ktorý má sociálne a rečové vzťahy, pričom verejná reč (tamže, s. 138) tvorí nielen nevyhnutný aplikačný rámec jeho konceptov, ale je aj ich akoby nekonečným zdrojom.

Podľa Bakerovej je takýto pohľad na problém *interiorizovanej ľudskej komunikácie* vyslovene antikarteziánsky (Bakerová 2013, s. 139-140), pretože žiadnej izolovanej entite neumožňuje, aby disponovala robustnou perspektívou 1. osoby. A bez tejto perspektívy nemôže nikto a nič disponovať myšlienkami o sebe ako sebe\*<sup>13</sup>, ani mať vnútorný duševný život ako taký. Na základe tejto argumentácie potom prechádza Bakerová k trom tvrdeniam, ktoré budú možno mnohým zníť až šokujúco, ale z nášho hľadiska sú úplne legitímne. Po prvé, prirodzené jazyky nie sú žiadnym prídavkom k ľudskej psychike (tamže, s. 141), pretože veľa, ak nie väčšina našich mentálnych stavov *je existenčne závislá od ich stelesnenia, vyjadrenia alebo sformovania v prirodzenom jazyku*. Po druhé, *osoby* a nie mozgy či organizmy „prechádzajú vývojom od nekonceptuálnych ku konceptuálnym entitám“ (tamže). To znamená, po tretie, a tu odvážna argumentácia L. R. Bakerovej vrcholí, nikto nie je identický s jeho mozgom, pretože „ani mozgy, ani mysle nie sú subjektmi skúsenosti alebo racionálnymi či morálnymi agentmi; my osoby sme nimi“ (tamže, s. 142). Naše telá *teda môžu pozostávať z čohokoľvek – zo živej hmoty, kremíka a pod.*, pokým nám však budú poskytovať mechanizmy podporujúce naše osobné aktivity alebo mentálne stavy, stále budú našimi telami v prísnom slova zmysle.

## Hlavné zložky ľudskej prirodzenosti – myseľ a vedomie

Ako sme vyššie upozornili, mnohí bádatelia používajú pojmy myseľ a vedomie, akoby boli synonymné, resp. označovali tú istú entitu. S vedeckým poznaním reality, konzekventne alebo logicky rozčlenenej do prirodzených druhov (vrátane ľudských artefaktov ako prirodzených druhov s vlastnou kauzalitou), to však podľa nás nielenže nemá nič spoločné, ale vyúsťuje to do prehlbujúcej sa stagnácie psychológie ako vedy a filozofie ako jej hlavného metodického a metodologického nástroja. Niektorí bádatelia sa pritom správajú, akoby bol takýto prístup k problému (čiže rozlíšeniu a presnému definovaniu mysle a vedomia) neproblematiký, a tak nijako nebránil ďalšiemu prehlbovaniu či rozvíjaniu ľudského poznania v tejto oblasti. Silvia Gáliková napríklad vo svojej štúdií (2011, s. 75-77) nepocit'uje potrebu odlíšiť vedomie od mysle a naozaj používa tieto dva pojmy akoby boli synonymné, pričom na jej 64. strane bez väčšieho váhania podotýka, že pojem mysle používa v tomto texte „synonymicky s pojmami duša, vnútorný svet, psychika, mentálne stavy“, a to bez ohľadu na to, že naše intuície nám jasne „hovoria“, že tieto entity sú si možno do veľkej miery podobné, ale určite nie identické.

Jan Burian (2008, s. 10) zdôrazňuje, že pojem vedomie má dva základné významy, keď v medicíne označuje zvláštny „stav pacienta, ktorý dokáže reagovať na podnety inak než čisto reflexívne“, resp. pri vážnom ochrnutí si dokáže „podnety zapamätať a neskôr reprodukovať“. Vo filozofii však tento pojem označuje predovšetkým „skúsenosť s dôrazom na subjektívne prežívanie v 1. osobe“. Oveľa komplikovanejšia je situácia s používaním pojmu myseľ, keď

<sup>13</sup> Symbol \* používa Bakerová na označenie entít, ktoré disponujú sebauvedomením a perspektívou 1. osoby.

si navzájom konkurujú mnoho definícií tohto pojmu. Pri zúženom a možno aj najčastejšie sa vyskytujúcom chápaní pod mysl'ou rozumieme „súhrn a podklad poznávacích procesov“, kým pri širšom chápaní je tento pojem „používaný ako synonymum k pojmu vedomie“, pričom myseľ sa chápe ako niečo subjektívne a telo zase ako niečo objektívne existujúce, v dôsledku čoho sa objavuje známy mind-body problém. Ďalší bádatelia sa zase domnievajú, že pojem myseľ je širší ako pojem vedomie a do mysle potom zahrňujú okrem vedomých „aj nevedomé či podvedomé procesy“. V rámci kognitívnej vedy napokon vystupuje do popredia špecifické chápanie mysle ako súboru „procesov pojmového (konceptuálneho) myslenia, kde pojmy sú *reprezentácie objektívneho sveta*, ktoré je možné označiť symbolmi a pojmové myslenie potom redukovať *na logické operácie so symbolmi*“ (Burian 2008, s. 10-11; podč. R. B.).

Za danej situácie je potom asi najlepším východiskom pokúsiť sa nájsť presný význam týchto dvoch pojmov na nejakej všeobecne uznávanej internetovej stránke, akou je v našom prípade najmä Stanford Encyclopedia of Philosophy. Keďže heslá na tejto stránke sa príležitostne obmieňajú, my budeme pracovať s jej verziou, ktorá bola filozofickej komunite ponúknutá začiatkom septembra 2023. Vedomie tak bolo Robertom Van Gulikom definované *ako jeden z aspektov ľudskej mysle*, pričom v celom jeho príspevku sa vedomé aj mentálne stavy buď *priamo stotožňovali*, alebo *cháпали ako navzájom totožné*, a to bez ohľadu na to, či daný autor hovoril o vedomí vo vzťahu k jeho nositeľom (zvieratám, ľuďom alebo iným kognitívnym systémom) schopným *v rôznom stupni* vnímať podnety z okolitého sveta, reagovať na ne a i., alebo abstraktnej reifikácii niektorej z jeho hlavných vlastností (fenomenálnom, reflexívnom či naratívnom vedomí a pod.). Pre myseľ pritom nebolo v uvedenej encyklopédii vyčlenené zvláštne heslo, ale osobitne analyzované boli jednotlivé druhy (alebo typy) mysle.

Pre nás je zaujímavé najmä heslo týkajúce sa tzv. externalizmu mysle (spracované Markom Rowlandsom et al.), v rámci ktorého sa predpokladá, že procesy realizujúce sa v individuálnej mysli sú prinajmenšom sčasti ovplyvnené procesmi (alebo entitami) nachádzajúcimi sa mimo nášho mozgu. Externalizmus sa pritom môže týkať obsahu našej mysle alebo hýbateľov, resp. nositeľov tohto obsahu. Nemenej dôležité je heslo Modularity of Mind (spracované Philipom Robbinsom), v ktorom sa najprv spomína Fodorova striktné vedecká koncepcia modularity nášho mozgu a potom „voľnejšia“ koncepcia evolučných psychológov (ako Dan Sperber a i.), podľa ktorej sa modularita (t. j. rozčlenenie mozgu na relatívne samostatne pracujúce moduly) *netýka len systémov nižšej úrovne*, zakladajúcich vnímanie a reč, ale doslova preniká mozgom skrz-naskrz a zahrňuje *aj systémy najvyššej úrovne* zodpovedné za uvažovanie, plánovanie, rozhodovanie a pod.<sup>14</sup> V Stanfordskej encyklopédii filozofie napokon nájdeme aj heslá, ktoré sa týkajú tzv. výpočtovej teórie mysle alebo teórie identity mozgu a mysle (podľa ktorej sú mentálne stavy a procesy *v podstate totožné s procesmi mozgu*), ale ohľadom nami hľadaného rozdielu medzi mysl'ou a vedomím sa ani v tomto prípade nič zásadného nedozvieme.

Veľmi nám v danom ohľade nepomôžu ani veľmi inšpiratívne publikácie Paula Thagarda et al. (2007) alebo Georgea Botterilla a Paula Carruthersa (2003), pretože aj keď sa v nich daná problematika spracováva naozaj precízne a dôkladne, pre nás dôležitý rozdiel medzi mysl'ou a vedomím v nich jednoducho nie je tematizovaný. Najlepšie preto bude obrátiť sa na autorov

---

<sup>14</sup> Len pod čiarou pritom stačí poznamenať, že verzia evolučných psychológov má aj veľa kritikov, ku ktorým patrí napríklad aj Miroslav Popper, ktorý v súvislosti s tvrdeniami evolučných psychológov o existencii rôznych doménovo-špecifických a evolučno-adaptačných mechanizmov (2003, s. 19-20) konštatuje, že „zatiaľ nie je známe, z koľkých takýchto mechanizmov a akého druhu sa ľudská myseľ skladá“. Na riešení niektorých zložito štruktúrovaných úloh sa navyše podľa neho môže zúčastňovať naraz viacero takýchto mechanizmov alebo sa ten istý mechanizmus „môže zúčastňovať na riešení rozdielnych úloh v rámci tej istej triedy problémov“. Napokon, mnohé naše činnosti sú „tak silno (spolu)determinované kultúrou“, že je takmer nemožné vysvetliť ich len „za pomoci vrodenných doménovo-špecifických mechanizmov“.

alebo texty, ktoré sa tomuto problému venujú explicitne. Veľa zaujímavých dištinkcií ponúka napríklad Andrej Ule (2015), ale s ohľadom na hlavnú tému našej štúdie sa budeme detailne venovať hlavne poznatkom, ktoré ponúka Susan Greenfieldová vo svojej staršej štúdií (2002), kde hneď v úvodnom odstavci zdôrazňuje, že pojmy *mysel* a *vedomie* *neoznačujú* rovnaké či veľmi podobné entity a ako také by sa určite nemali navzájom zamieňať pri opise a výklade psychických fenoménov. Aby bol jej výklad úplný, resp. konzistentný, venuje najprv veľkú pozornosť ľudskému mozgu, jeho „výkonnostným parametrom“ a spôsobu, akým funguje.

V miernom rozpore s tvrdeniami tzv. modularistov napríklad uvádza (tamže, s. 91), že žiadna väčšia oblasť mozgu neslúži len jednej, plne izolovateľnej a tak kompletnej funkcii. Naopak, každá väčšia oblasť alebo región mozgu, vrátane prefrontálneho kortexu, sa môže podieľať na realizovaní viac ako jednej funkcie. Tieto regióny tak podľa nej *nie sú žiadne autonómne jednotky*, ale, obrazne povedané, skôr a viac hráči na spoločnom javisku (a zjavne aj v tom istom „predstavení“). Ako pokračuje ďalej, v ľudskom tele máme len asi 30 000 génov a ak by sme aj vzali do úvahy, že každý z nich by bol priamo spojený s nejakou synapsou, stále by sme mali k dispozícii len asi  $10^{10}$  spojení, z čoho podľa nej vyplýva, že ani tejto základnej úrovni funkcií mozgu, t. j. génom nemôžeme pripísať autonómne funkcie, resp. fungovanie. Na úrovni regiónov mozgu a génov tak podľa nej zostáva málo priestoru na „manévrovanie“, a preto je za týchto podmienok ťažké pochopiť, ako sa v rámci mozgu ako takého mohla uskutočniť „personalizácia“ mozgu a na jej základe vyvinúť „individualizovaná“ myseľ.

Oveľa vhodnejším kandidátom na túto personalizáciu mozgu je podľa nej akási medziúroveň, t. j. úroveň mozgových spojení samotných (medzi neurónmi a pod.), ktoré sú nielen vysoko dynamické, ale súčasne aj plastické, pretože dokážu reflektovať a uchovávať všetky jedincom nadobudnuté skúsenosti. Naše mozgy sú pritom vysoko plastické počas prvých rokov nášho života (keď dokážu čiastočne kompenzovať aj ich väčšie poškodenie), ale túto plasticosť si uchovávajú aj do vyššieho veku, ako sa potvrdilo pri výskume londýnskych taxikárov, ktorým sa v dôsledku nutnosti omnoho viac si pamätať výrazne zväčšil hipokampus v porovnaní s „netaxikármi“ podobného veku. Naše zážitky a skúsenosti sa takto nielen akoby zrkadlia, ale sú doslova stelesnené (tamže) „v sile a rozsahu mozgových spojení“ a je to práve tento proces ich presného „zrkadlenia“ v našom mozgu, ktorý Greenfieldová nazýva (našou) *mysľou*. Pri takomto prístupe potom majú odlišné mysle aj klony alebo identické dvojčatá, pretože aj ony budú a musia mať (tamže) „jedinečnú konfiguráciu mozgových bunečných spojení“.

Myseľ by sme tak mohli definovať ako *špecifické* procesy nášho mozgu (či v našom mozgu), resp. jeho *bdelé* alebo *momentálne aktívovanie*, pri ktorom však mozog v prípade potreby pracuje aj so staršími spomienkami, ak si to vyžadujú úlohy, pred ktoré sme ho postavili. Ak však podľa Greenfieldovej niekto takto vymedzenú myseľ akoby (dočasne) stratí, t. j. stratí nad týmito procesmi vedomú kontrolu, čo ľudová psychológia vyjadruje frázami ako „prišiel o rozum“, „rozum sa z neho vytratil“ a i., potom jediným možným rezíduom môže byť len (akokoľvek oslabené) vedomie, ktoré však už v žiadnom prípade *nie je našou myseľou*, pretože jedincovi viac nie sú prístupné „personalizované kognitívne perspektívy“ (tamže), vonkajší svet v jeho prípade stráca personalizovaný význam (alebo zmysel) a namiesto toho sa stáva (tamže, s. 92) „len pasívnym prijímateľom prichádzajúcej senzorickej informácie“. Svet je tak podľa Greenfieldovej dočasne akoby úplne zbavený kognitívneho obsahu, pričom, ako uvádza ďalej, nezmenili sa gény v mozgu, nezmenili sa jeho regióny, ba ani reálne existujúce alebo potenciálne spojenia, ktoré dávajú vzniknúť mysli, ale zmenila sa dostupnosť týchto spojení. No a takáto rýchla zmena bola a je podľa nej možná len vďaka tomu (ako argumentuje aj L. Kováč), že v mozgu na globálnej úrovni rýchlo a veľmi účinne pracujú chemikálie, konkrétne transmittory, ktoré sú de facto sprostredkovateľmi najzákladnejšieho senzorickeho vedomia.

Pravdaže, základom týchto interakcií sú tzv. globálne neuróny, ktoré vďaka transmitterom nepracujú, na rozdiel od počítača, digitálne, *ale analógovo*, pričom na základe transmittermi vzbudzovanej a nimi následne jemne odstupňovanej vzájomnej interakcie realizujú základné funkcie vedomia v tom najjednoduchšom zmysle slova – vzrušenie, spánok, bdenie alebo snívanie. A ak tieto „modulujúce fontány“ transmitterov v neurónoch z rôznych dôvodov (napríklad pri užívaní extázy) pracujú pomalšie, neuronálne okruhy nepracujú dostatočne rýchlo a súbory navzájom interagujúcich neurónov sú relatívne malé. Podľa Greenfieldovej (2002, s. 92) je tomu tak nielen pri malých deťoch, ktoré si ešte nestihli vytvoriť dosť (neuronálnych) spojení, ale aj v spánku, kedy nedostatok senzorickej stimulácie vyúsťuje takisto do oveľa slabšieho zapojenia dostupných neurónov. Na tomto princípe napokon podľa nej fungujú aj anestetiká, ktoré tiež v mozgu výrazne znižujú počet navzájom interagujúcich neurónov. Ako nakoniec podotýka Greenfieldová, vedomie je potom známym „mystickým, subjektívnym svetom 1. osoby, do ktorého nikto zvonka nemá prístup, zatiaľ čo myseľ môže byť následne „vzťahnutá bližšie ku konceptom personalita a sebauvedomenia“ (tamže).<sup>15</sup>

## Hlavné zložky ľudskej prirodzenosti – robustná perspektíva 1. osoby

V tejto časti budeme vychádzať z už citovaných publikácií L. R. Bakerovej (2000, 2013), pričom s ohľadom na hlavný cieľ našej práce nebudeme prezentovať jej argumentáciu vo všetkých relevantných detailoch, ale všimneme si len tie jej argumenty, ktoré sa týkajú tak hlavných zložiek ľudskej prirodzenosti, ako aj jej vymedzenia ľudskej osoby ako zvláštneho prirodzeného druhu, s jemu vlastnou kauzalitou a identitou. V tejto súvislosti vystupujú do popredia aj vzájomné vzťahy medzi jednotlivými ontologickými úrovňami reality, ktoré Bakerová takisto chápe inak ako väčšina filozofickej obce, podľa nášho názoru však, na rozdiel od tzv. filozofického mainstreamu, správne a hlbšie. Terčom jej ostrej kritiky sú najmä rôzne formy reduktívneho a scientistického (alebo vedeckého) naturalizmu, ktoré odmietajú akceptovať (resp. vôbec neberú na vedomie) autonómnou existenciu vyšších ontologických úrovní reality či robustnej perspektívy 1. osoby, ktorá nie je opísateľná alebo analyzovateľná tradičnými nástrojmi (akože) úplne objektívnej a neosobnej, (akoby) čisto naturálnej vedy.

Medzi nimi najmä *fyzikalizmus*, podľa ktorého (Bakerová 2013, s. 8) akákoľvek úroveň reality, v rámci ktorej objekty nemajú fyzikálne dôležité či skúmateľné časti, sa identifikuje alebo môže byť stotožnená s jej základnou (t. j. fyzikálnou) úrovňou. Na rovnakom princípe je podľa nej založená aj *silná superveniencia*, podľa ktorej sú zase makrofyzikálne vlastnosti objektu nevyhnutne determinované alebo spoluurčované jeho mikrofyzikálnymi vlastnosťami a v rámci *reduktívneho naturalizmu* (tamže, s. 9) na tieto mikrofyzikálne vlastnosti v podstate bezo zvyšku redukované.<sup>16</sup> Ako však tvrdí Bakerová, ak je reduktívny naturalizmus naozaj pravdivý, potom celý nami vnímaný svet musí *ontologicky kolabovať* v tom zmysle, že všetky objekty vyšších úrovní sa stanú kauzálne inertné, t. j. neschopné akokoľvek reálne vplývať na objekty nižších úrovní. Určitou nádejou sú pre ňu optimistické *neredukcionistické ontológie* (Bakerová

<sup>15</sup> Alebo, ešte inak povedané (a zhrnuté), zatiaľ čo myseľ je všetko, čo aktuálne prežívame, vrátane našich myšlienok, spomienok, predstáv, telesných pocitov a zmyslových vnemov, vedomie je entitou, ktorá všetko, čo je prežívané alebo vnímané, reálne prežíva a vníma. Je to teda skutočne ono mystické a záhadné, len a len naše „ja“, ktoré je tak akýmsi javiskom, na ktorom sa prezentuje obsah všetkého toho, čo prežívame a vnímate.

<sup>16</sup> Len s ohľadom na úplnosť nášho výkladu pritom dodávame, že pri *slabej superveniencii* je táto akoby fatálna závislosť vyšších úrovní reality od nižších predsa len oslabená tým, že fyzikálne duplikáty osôb v tomto, ako aj v ďalších možných svetoch budú síce fyzikálne identické, ale budú mať odlišné myšlienky, zatiaľ čo pri silnej superveniencii budú mať popri fyzikálnej identite aj rovnaké myšlienky vo všetkých možných svetoch. Slabá superveniencia teda pri takomto štandardnom chápaní platí lokálne, zatiaľ čo silná globálne alebo bez výnimky.

2013, s. 26), ktoré akceptujú a zahrňujú ako reálne vedy nielen psychológiu, ale aj sociológiu, historiografiu či politológiu (vrátane nimi skúmaných objektov alebo procesov – R. B.). Ale ani v ich rámci nie sú vlastnosti súvisiace s perspektívou 1. osoby „ontologicky signifikantné“; podľa všetkého (tamže, s. 28) najmä preto, lebo perspektíva 1. osoby nemôže byť nijakým spôsobom naturalizovaná, prinajmenšom nie v rámci fyzikalistického prístupu.

Bakerovej tak neostáva nič iné, ako rozvinúť vlastnú ontológiu, resp. vlastnú teóriu vzťahov medzi ontologickými úrovňami reality, ktorú sama označila termínom „konštitúcia“, podľa ktorej (Bakerová 2000, s. 28-29) je táto vzťahom pôvodnej alebo autentickej jednoty objektu vyššej úrovne s látkou, ktorá ho konštituuje, pričom táto vzájomná jednota objektu a jeho materiálnej „zložky“ nesmie byť stotožňovaná s ich identitou, aj keď sa konštitúcia (tamže, s. 58) približuje k identite tak, ako je to len možné. Podľa Bakerovej (tamže, s. 33-34) tak relátami konštitučného vzťahu „nie sú vlastnosti“, ako je tomu pri superveniencii, a obyčajné veci nemôžu byť stotožnené ani s mereologickými sumami, hoci ich tieto konštituujú (tamže, s. 38). Konkrétne – pri soche Michelangelovho Dávida vstupuje do konštitučného vzťahu nie mramor ako látka, ale ten kus mramoru, z ktorého je Dávid zhotovený. Vlastnosť byť sochou má pritom daný kus mramoru *odvodene*, zatiaľ čo samotná socha je *novou úrovňou reality* práve ako výsledok špecifickej sochárskej aktivity a ako taká kauzálne svojbytná a účinná.<sup>17</sup>

Ako ďalej pokračuje Bakerová (tamže, s. 35), dôležité je pritom uvedomiť si, že nie všetky konštituované veci vznikajú ako výsledok niekoho zámeru (intencionálne), ale môžu byť tým, čím sú aj na základe svojich *relačných vlastností*, ako je tomu pri génoch, ktoré sú teda konštituované jednotlivými a navzájom prepojenými molekulami DNK. Veľmi dôležité je aj rozlišovanie medzi *esenciálnymi a neesenciálnymi vlastnosťami objektu* (alebo procesu), bez ktorých v prvom prípade objekt nemôže existovať, zatiaľ čo bez druhých sa v rôznej miere alebo za rôznych okolností zaobísť môže. A rovnako aj medzi primárnymi a „sekundárnymi“ druhmi, pri ktorých podľa Bakerovej (tamže, s. 40, podč. R. B.) vystupuje do popredia principiálna otázka – „za akých podmienok jedna vec *konštituuje novú entitu* v protiklade k jednoduchému *nadobudnutiu ďalšej vlastnosti*?“ No a obísť by sme nemali ani Bakerovej vymedzenie *hybridnej vlastnosti*, pri ktorej objekt zjavne pozostáva zo zložiek z dvoch (alebo viacerých) ontologických rovín, ako je tomu napríklad pri samotnej autorke, ktorá je súčasne organizmom a zároveň osobou disponujúcou robustnou perspektívou 1. osoby (tamže, s. 97).

Pre nás sú, pravdaže, najdôležitejšie tie pasáže Bakerovej textov (2000 a 2013), kde sa venuje našej druhej identite. V rámci konštitučného prístupu (2000, s. 91) tak podľa nej to, čo robí ľudskú osobu osobou je práve viackrát spomínaná *robustná perspektíva 1. osoby*, pričom perspektívu ako takú (Bakerová 2013, s. 40-41) nechápe ako nejaký objekt, ktorý zaberá určitý priestor, ale ako dispozíciu toho či onoho vedomého subjektu (vrátane novorodenca alebo psa, ktorí takto disponujú vedomím nižšieho alebo 1. stupňa) „vnímať svet zo zvláštnej časopriestorovej pozície“. Osobou (resp. subjektom či *individuálnym autonómnym agentom*) sú takto nielen ľudia disponujúci robustnou perspektívou 1. osoby, ale aj novorodenci a vyššie živočíchy, ktoré (ako sme vyššie uviedli) dokážu vnímať, prežívať bolesť alebo vyjadrovať svoje emócie len z pozície *rudimentárnej perspektívy 1. osoby*. Podľa Bakerovej (2000, s. 93) je pritom záležitosťou „neurofyziológov, nie filozofov, určiť podmienky, za ktorých je ľudský organizmus schopný zabezpečovať intencionálne stavy z pozície 1. osoby“.

Pokiaľ ide o našu špecifickosť, najmä oproti živočíchom, ktoré sú si takisto vedomé (svojich vnemov a pocitov) ako my, Bakerová zdôrazňuje (tamže, s. 14), že na rozdiel od nich nie sme obmedzení len na sledovanie cieľov, ktoré súvisia s našim prežitím a reprodukciou, pretože

---

<sup>17</sup> A môže sa potom stať objektom estetického súdenia, artiklom na dražbe umeleckých predmetov a pod.

oproti nim máme vďaka (robustnej) perspektíve 1. osoby nad našimi cieľmi určitú kontrolu, a to vrátane tých, ktoré sa týkajú nášho prežitia a reprodukcie. Robustná perspektíva 1. osoby je tak základom všetkých foriem sebauvedomenia (tamže, s. 59-60). Hranica medzi ľuďmi ako rozumnými a *sebavedomými* živými bytosťami a ostatnými vedomými živými bytosťami je takto Bakerovou stanovená dostatočne jasne. Americká filozofka však ide ešte ďalej a o pár strán nižšie (tamže, s. 95) uvažuje nad tým, koľko častí ľudského tela sa musí nahradiť inými, aby človek prestal byť človekom. Podobne ako v prípade legendárnej Theseovej lode sa jej táto hranica zdá byť pomerne neurčitá alebo vágna, pokiaľ však zostáva naše telo fungovať na báze DNK a RNK, môže byť stále považované za ľudské telo, telo bytosti či entity patriacej do biologického druhu *Homo sapiens*. Rovnako je zrejme, že podľa jej konštitučnej teórie sme my, *ako osoby*, v podstate psychickými a morálnymi entitami, zatiaľ čo naše telá sú (len) biologickými entitami (tamže, s. 105), ktoré môžu byť čoskoro nahradené inými entitami.

Ako totiž ďalej pokračuje Bakerová (tamže, s. 106), ľudský organizmus nie je možné zmeniť na nebiologický bez toho, aby prestal existovať, ale ľudská osoba, pôvodne pozostávajúca z ľudského organizmu, môže získať nebiologické telo a naďalej existovať (tamže, s. 106), aj keď v tomto prípade už nebude ľudskou osobou, ale osobou s nebiologickým (t. j. plastickým, kremíkovým a pod.) telom. Z toho je podľa Bakerovej zrejme, že osoba ako taká má kauzálne sily (alebo účinky), ktoré by *akékoľvek* telo nemalo, pokiaľ by nekonštituovalo túto osobu (tamže, s. 109). Argumentačná a analytická dôslednosť tak privádza Bakerovú aj k ďalším, podobne šokujúcim tvrdeniam. Nielenže je podľa nej osoba ako taká zvláštnym prirodzeným druhom (tamže, s. 11), ale, ako predpokladá pre zmenu vo svojom texte z roku 2013 na 94. strane, žiadny z našich intencionálnych mentálnych stavov nesupervenuje na stavoch nášho mozgu a obsahy našej subjektívnej skúsenosti sú doslova nainfikované našimi personálnymi vzťahmi s vonkajším prostredím. S výnimkou tzv. kválií (tamže, s. 95), „ak tu nejaké vôbec sú“, preto obsahy našej skúsenosti výrazne „závisia na interakcii s vonkajším prostredím“.

Pokiaľ ide o vedomie ako také, Bakerová (2000, s. 60) odmieta presne definovať túto entitu, pretože sa prikláňa „k tým filozofom, ktorí hovoria, že žiadna necirkulárna definícia vedomia nie je možná“. Z nášho hľadiska je pritom najdôležitejšie jej tvrdenie (tamže, s. 97), že my, ľudia, sme osobami „neodvodene“ alebo primárne (angl. nonderivately) práve preto, lebo túto našu základnú dispozíciu vlastnime *nezávisle* od konštitučných vzťahov k našim telám. Byť človekom teda znamená byť v prvom rade a podstatne *psychickou a morálnou osobou* alebo *entitou* a nie telom (či už biologickým, alebo nebiologickým), ktoré ju (len) konštituuje. Ešte zaujímavejšie a principiálnejšie je jej tvrdenie (Bakerová 2000, s. 131), že zatiaľ nemáme a zrejme ani nemôžeme mať koncepciu personálnej identity v čase, keďže nikto zatiaľ nebol schopný ponúknuť koncepciu tohto druhu, ktorá by bola: „a) informatívna, b) necirkulárna a c) plauzibilná“. Na jej ustanovenie pritom nestačí (tamže, s. 133) ani hypotéza psychickej kontinuity, pretože tá nevyklučuje existenciu dvoch osôb s rovnakým telom a myslou, zatiaľ čo pri perspektíve 1. osoby (tamže, s. 136) „je tu vždy asymetria medzi niekým (angl. oneself – R. B.) a hocikým ďalším“. Bakerová má teda podľa všetkého naozaj pravdu, keď opakovane tvrdí, že vízia robustnej perspektívy 1. osoby prináša do neosobného, prísne naturalistického sveta súčasnej vedy (a 3. osoby) niečo ontologicky a epistemicky neredukovateľne a zásadne nové, čo preto ani nemôže byť jej *standardnými* nástrojmi opísateľné a analyzovateľné.

## Hlavné zložky ľudskej prirodzenosti – inteligencia

Pokiaľ ide o definovanie inteligencie, najväčšími autoritami v danej oblasti sú zrejme stále Shane Legg a Marcus Hutter. V ich prvej spoločnej štúdií (L&H 2007a) najprv predložili v rámci tzv. kolektívnych definícií inteligencie, t. j. definícií vypracovaných a rešpektovaných kolektívnymi bádateľov, celý súbor definícií, z ktorých však nebolo zrejme, či sa viac vzťahujú

k ľudskej, zvieracej alebo tzv. strojovej inteligencii, ktorú je potom možné na základe vhodne zvolených kritérií nielen merať, ale aj ďalej rozvíjať či zdokonaľovať. V rámci kolektívnych definícií tak dominovali najmä tie, ktoré *bolo možné vzťahnúť len k človeku*, t. j. tie, v ktorých sa používali termíny ako „poznatok“, „porozumenie“, „imaginácia“, „usudzovanie“ a i., ktoré skutočne vzťahujeme najmä k človeku. Typickou ukázkou je tu napríklad hneď prvá citovaná definícia z AllWordsDictionary (2006), podľa ktorej (L&H 2007a, s. 2) je inteligencia „schopnosťou používať pamäť, poznanie, skúsenosti, porozumenie, rozumové dôvodenie, imagináciu a usudzovanie k riešeniu problémov a adaptovaniu sa na nové situácie“.

Ďalšie definície v tomto úvodnom prehľade už *mali všeobecnejší charakter* a môžu byť preto vzťahnuté nielen k človeku, ale aj k niektorým vyšším živočíchom, resp. výpočtovým strojom alebo zariadeniam, u ktorých predpokladáme, že prinajmenšom v nejakom základnom slova zmysle naozaj myslia, resp. realizujú (rozumové?) operácie analogické tým, ktoré realizujú ľudia vo svojich mozgoch. Typickými ukázkami sú tu definícia z World Book Encyclopedia (2006), podľa ktorej je inteligencia „schopnosť adaptovať sa na vonkajšie prostredie“ (L&H 2007a, s. 3), alebo definícia z Encyclopedia Britannica (2006), podľa ktorej je inteligencia takisto „schopnosťou adaptovať sa účinne na vonkajšie prostredie, buď zmenou seba samého, alebo zmenou tohto prostredia alebo nájdením nového prostredia... inteligencia nie je jediným mentálnym procesom, ale skôr kombináciou viacerých mentálnych procesov zameraných na účinné adaptovanie sa na vonkajšie prostredie“ (L&H, 2007a, s. 3). Najmä druhá definícia sa pritom javí, ako keby bola prevzatá bez úpravy priamo z nejakej učebnice evolučnej biológie.

Hranica medzi zvieracou, ľudskou a tzv. strojovou (či umelou) inteligenciou *zostáva nejasná a neostrá* aj v rámci tzv. psychologických definícií, t. j. definícií predložených jednotlivými psychológmi. Podľa W. V. Bingham (1937) je napríklad „schopnosťou organizmu riešiť nové problémy“ (L&H 2007a, s. 4). Na správanie zvierat môže byť vzťahnutá aj definícia inteligencie od W. F. Dearborna (2000) ako schopnosti „učiť sa alebo profitovať na základe skúsenosti“ (L&H 2007a, s. 5). Gardnerova definícia (1993) naopak priznáva inteligenciu len bytostiam s kultúrou v plnom slova zmysle, resp. ľuďom s úplne rozvinutým či funkčným mozgom, pretože podľa nej je inteligencia „schopnosťou riešiť problémy, vytvárať produkty, ktoré sú oceňované v rámci jednej alebo viacerých kultúrnych platforiem“ (L&H 2007a, s. 5). Pozoruhodný je aj Eysenckov pokus zachytiť kontext realizovania tohto fenoménu, pretože podľa neho je inteligenciou nielen samotný mozog, ale aj jeho inteligentné prejavy v reálnom živote a dosiahnuté výsledky v testoch merajúcich kognitívne schopnosti (L&H 2007a, s. 5).

Poslednou skupinou definícií, ktorej Legg s Hutterom venujú pozornosť v ich úvodnej štúdii, sú definície predložené bádateľmi vo sfére umelej inteligencie (ďalej aj AI), ktoré sa pritom *len málo odlišujú* od tých definícií, ktoré by sme mohli označiť ako definície zvieracej alebo ľudskej inteligencie. Podľa D. Fogela (1995) je napríklad inteligentný každý „systém... ktorý generuje adaptívne správanie zamerané na dosiahnutie cieľov v mnohorakých prostrediach“ (L&H 2007a, s. 7). Od tejto definície sa minimálne odlišujú definície B. Goertzela (2006), podľa ktorého (L&H, 2007a, s. 7) je inteligencia schopnosťou dosahovať „komplexné ciele v komplexných prostrediach“, alebo R. R. Gudwina (2000), ktorý takisto zdôrazňuje, že od inteligentných systémov sa očakáva, že budú úspešne pracovať „v mnohých odlišných prostrediach“ (L&H 2007a, s. 7). A to isté platí aj pre záverečnú, akoby syntetickú definíciu inteligencie od Legga s Hutterom (tamže, s. 9), podľa ktorých: „Inteligencia meria agentovu schopnosť dosahovať ciele v širokom rozsahu prostredí.“ A hoci sa Legg s Hutterom na záver utešujú tým, že ich definícia (takmer totožná s Fogelovou) zachytáva všetky podstatné znaky či vlastnosti tohto fenoménu, veľmi dôležité a principiálne rozdiely medzi zvieracou, ľudskou a strojovou (alebo umelou) inteligenciou v nej nie sú ani reflektované, ani zachytené.



Na tento problém narazili Legg s Hutterom aj v ich nasledujúcej štúdií (2007b), pričom opäť nedokázali pochopiť dôležitosť principiálneho rozlišovania medzi zvieracou, ľudskou alebo umelou inteligenciou, ako je zrejmé už z názvu ich štúdie, kde definujú umelú inteligenciu ako de facto univerzálnu inteligenciu, hoci je na zamyslenie, či pri vyšších živočíchoch (alebo živočíchoch vôbec) a výpočtových strojoch alebo „inteligentných“ programoch (aplikáciách a i.) možno vôbec hovoriť o inteligencii v plnom alebo pravom slova zmysle. Keďže však boli nútení nejako vecne a reálne odlišiť ľudskú inteligenciu od umelej (alebo strojovej), v časti venovanej tzv. prirodzenej alebo „prírodnej“ (t. j. ľudskej a aj zvieracej!) inteligencii najprv detailne diskutovali ľudské intelligenčné testy a potom konečne prešli k zvieracej inteligencii ako takej a možnostiam jej testovania. Ako upozornili, ak sú veľké problémy pri vypracúvaní a vyhodnocovaní ľudských intelligenčných testov, pri zvieratách je to ešte komplikovanejšie, pretože aj v rámci jedného druhu nemusí byť jasné (tamže, s. 7), čo v skutočnosti testujeme. Antropocentrický bias (čiže predsudok) sa zase môže prejavovať tým, že pri vypracúvaní testov nedoceníme zmyslové preferencie iných (biologických) druhov, napríklad fakt, že potkany sa môžu učiť novým vzťahom medzi vecami skôr pomocou čuchu ako iných zmyslov (tamže).

Na veľké zamyslenie je však podľa nás to, akým spôsobom sa pripravujú intelligenčné testy pre jednotlivé biologické druhy alebo taxony, keď sú bádatelia nútení zohľadňovať stupeň ich vývinu a pri tzv. nižších živočíchoch sa zameriavať napríklad na jednoduché úlohy súvisiace s krátkodobou a dlhodobou pamäťou, formovaním asociácií, schopnosťou zovšeobecňovať jednoduché vzory (angl. patterns) a i., kým pri vyšších živočíchoch (ako sú vtáky či primáty) posudzovať skôr ich sofistikovanejšie vlastnosti, ako sú schopnosť podvádzať, resp. klamať, napodobňovanie či spoznanie samých seba v Gallupových zrkadlových testoch atď.; najmä ak bola inteligencia predtým vymedzená aj ako kombinácia „viacerých mentálnych procesov zameraných na účinné adaptovanie sa na vonkajšie prostredie“ (L&H 2007a, s. 3). Pre nás je pritom veľmi dôležité aj ich rozlišovanie medzi tzv. *statickými a dynamickými intelligenčnými testami* (L&H 2007b, s. 9), keď tie prvé merajú u jednotlivcov jednorázovo schopnosť riešiť určité problémy a ich výsledky sú silne determinované dosiahnutým vzdelaním a poznatkami, zatiaľ čo tie druhé sa zameriavajú na zistenie nie aktuálneho výkonu, *ale kapacity učiť sa*, ktorá môže byť „zamaskovaná“ nižším dosiahnutým vzdelaním či chýbajúcimi poznatkami.

Stále nezodpovedanou však zostáva otázka, čo to vlastne inteligencia je, resp. čo všetko do nej patrí. Alebo inak – či môže byť chápaná ako jedna schopnosť (danosť či potencia), alebo ako súbor viacerých pozitívnych vlastností (schopností, kapacít či potencií). L&H napríklad uvádzajú Sternbergovu koncepciu „Triarchickej mysle“ (1985), v rámci ktorej sa inteligencia člení na tri zložky – na analytickú, kreatívnu a praktickú inteligenciu. Pri jej ďalšom členení na čoraz špecifickejšie zložky však niektorí bádatelia dospeli až k extrému, keď J. P. Guilford (1967) najprv vyčlenil tri fundamentálne dimenzie inteligencie – obsahy, operácie a produkty (L&H 2007b, s. 10), v rámci ktorých potom rozlíšil najskôr 120 a neskôr až 150 rozličných kategórií. Jeho model inteligencie sa tak javil ako dostatočne jemnozrnný, ale z praktického hľadiska bol fakticky nepoužiteľný, či už pri samotnom meraní inteligencie, alebo v rámci experimentálneho výskumu. Svojho času bol veľmi populárny aj Gardnerov model viacerých inteligencií (1993), v ktorom boli jednotlivé zložky inteligencie od seba oddelené až do tej miery, že vytvárali samostatné „inteligencie“, napríklad lingvistickú, muzikálnu, priestorovú, logicko-matematickú a i., a to podľa štruktúry ľudského mozgu (L&H 2007b, s. 10).

Úplne iný prístup k problému uprednostnil ešte medzi vojami C. E. Spearman (1927), ktorý vymedzil inteligenciu ako *veľmi všeobecnú duševnú schopnosť*, ktorá zakladá všetky ostatné duševné schopnosti a prispieva k nim (L&H 2007b, s. 10). Dôkazom toho je podľa neho skutočnosť, že individuálne výkony v dôvodení, asociovaní, lingvistike, priestorovom myslení či identifikovaní vzorov „sú pozitívne korelované“. Spearman nazval túto pozitívnu koreláciu

„g-faktorom“, kde „g“ označuje „všeobecnú inteligenciu“. Na základe súboru rozličných úloh (odhaľujúcich kvalitu alebo rozsah príslušných duševných schopností) tak bolo možné určiť u testovaného jedinca g-faktor a na základe jeho širokej (alebo všeobecnej) platnosti aj (jeho) inteligenciu ako takú. S veľmi zaujímavým rozvinutím Spearmanovej g-teórie prišiel neskôr R. B. Cattell (1987), ktorý od seba odlišil tzv. fluidnú inteligenciu ako všeobecnú a vrodenu schopnosť riešiť komplexné problémy a tzv. vykryštalizovanú alebo kryštalickú inteligenciu, ktorá meria poznatky a schopnosti, ktoré jednotlivec nadobudol počas svojho života (L&H 2007b, s. 10). Kompromisný a v súčasnosti asi najviac uznávaný prístup prepája obidve verzie inteligencie (Carroll 1993), pričom inteligencia tvorí hierarchiu schopností, na vrchole ktorých sa nachádza g-faktor a pod ním ďalšie, čoraz špecifickejšie schopnosti (L&H 2007b, s. 10).

Na druhej strane, pokiaľ ide o rozdiely medzi zvieracou, ľudskou a umelou inteligenciou, ani v tomto texte L&H sme sa o nich nič dôležité nedozvedeli, a preto sa v závere tejto časti našej štúdie obmedzíme len na niekoľko návrhov, ako by sa tento problém dal riešiť. Prvoradý a zo všetkého najdôležitejší je rozdiel medzi zvieracou a ľudskou inteligenciou, aj keď je naozaj na zamyslenie, či pri zvieratách vôbec môžeme hovoriť o inteligencii napríklad v tej podobe, ako ju definoval C. E. Spearman, a či by nebolo vhodnejšie hovoriť pri nich skôr o *predpokladoch* inteligentného myslenia alebo správania. Ak by sme ju však u zvierat pripustili, potom by sme ju mohli prisúdiť napríklad niektorým vyšším živočíchom, resp. vyšším živočíchom vôbec, alebo zvoliť prístup k jej prisúdeniu na základe iných vlastností, ktoré s inteligenciou očividne úzko súvisia. Najviac zo všetkých sa ponúka *autonómnosť*, ktorá je aj podľa Johna Colliera (1999, s. 1) hlavnou vlastnosťou plne anticipujúcich a dynamických systémov (alebo entít).

Autonómny systém potom Collier definuje ako entitu, ktorá využíva svoju vlastnú informáciu na modifikovanie samej seba *alebo okolitého prostredia*, aby prežila, reagujúc pritom tak na vonkajšie, ako aj vnútorné stimuly, a to zmenami svojich základných funkcií, čo jej umožňuje výrazne zvýšiť jej životaschopnosť (tamže, s. 2). V ďalšej svojej štúdií (2006, s. 2) už Collier nadväzuje na Imanuela Kanta, podľa ktorého si najjednoduchšia verzia autonómie vyžaduje, „aby bol cieľ systému definovaný v jeho vnútri“ a tento cieľ bol pod plnou kontrolou daného agenta, pričom bez rozhodujúceho vplyvu vonkajších podmienok. Ak si však v tejto súvislosti uvedomíme, že podľa Bakerovej (2013, s. 44) majú živočíchy s rudimentárnou perspektívou 1. osoby „len biologicky dané ciele a motivácie, ktoré nemôžu posudzovať či modifikovať, kým bytosti s robustnou perspektívou 1. osoby môžu posudzovať a modifikovať svoje ciele“ (ktoré pritom nie sú obmedzené na ich púhe prežitie alebo reprodukciu), potom musíme dôjsť k záveru, že *ani vyššie živočíchy nedisponujú inteligenciou v plnom či pravom slova zmysle*.

V prvom rade preto, lebo svoje vonkajšie prostredie modifikujú len výnimočne a aj to nie na základe naučeného, ale skôr zdedeného alebo inštinktívneho správania (ako to robia bobry, včely, termity a pod.). Po druhé, aj tie najviac vyvinuté živočíchy žijú v konkrétnom, jasne ohraničenom či vymedzenom areáli alebo prostredí, na rozdiel od človeka, ktorý je jediným skutočným eurybiontom na planéte aj preto, lebo si v prípade potreby dokáže sám vytvárať tie najrozmanitejšie *civilizačné* prostredia. No a po tretie, ako de facto konštatovala aj Bakerová, v ich správaní dominujú zdedené a nie naučené vzory správania, tak ako je tomu u ľudí. Ešte dôležitejšie je, že na základe Collierových principiálnych myšlienok o autonómnosti systémov môžeme aspoň provizórne určiť, či sú súčasné systémy umelej inteligencie naozaj inteligentné alebo nie. Podľa všetkého *nie*, pretože zatiaľ čo normálne ľudské bytosti sú plne autonómne biologicky, mentálne, morálne a psychicky (Collier 1999, s. 1), o súčasných robotoch alebo „inteligentných“ programoch sa to určite tvrdiť nedá, ako je nakoniec úplne triviálne zrejmé aj z toho, že tieto roboty a programy si *v prvej inštancii* svoje ciele nezadávajú zvnútra, ale sú im zadávané ľudskými agentmi, ktorí *plne splňajú* podmienku autonómnosti (a inteligencie).

## Umelá inteligencia – úvodné poznámky

Na druhej strane je zrejmé, že určitú autonómnosť pri vypracúvaní zadaných úloh nemôžeme „inteligentným“ robotom či programom uprieť. Možno by sme preto o „decíznej autonómii“ mali hovoriť len pri ľuďoch, zatiaľ čo pri zariadeniach strojovej (alebo umelej) inteligencie používať skôr termín „akčná“ alebo „behaviorálna inteligencia“. Vzhľadom na inštitucionálne a finančné tlaky je pritom bežné, že bádatelia vo sfére AI majú často tendenciu zveličovať nimi dosiahnuté výsledky a pripisovať nimi skonštruovaným (či vytvoreným) robotom alebo programom aj vlastnosti, ktoré určite nemajú alebo nemôžu mať. Možno preto bude užitočné bližšie preskúmať historické okolnosti, za ktorých sa AI v 20. storočí zrodila a vyvíjala, najmä ak je práve moment vzniku určitého systému alebo triedy systémov tým momentom, kedy sú im zadávané ich podstatné alebo imanentné vlastnosti, t. j. vlastnosti, ktoré ich robia tým, čím sú. Výborným úvodom do tejto problematiky je spoločná štúdia B. Jacka Copelanda a Diany Proudfootovej uverejnená v závere nami už citovanej publikácie Paula Thagarda et al. (2007).

Na počiatku celej tejto výskumnej oblasti tak bola podľa nich Turingova štúdia ešte z roku 1936 (C&P 2007, s. 429-430), v ktorej Alan Turing uvažoval o výpočtovom stroji, ktorý by mohol byť naprogramovaný tak, aby dokázal realizovať akýkoľvek výpočet idealizovaného „ľudského počítača“, t. j. „úradníka, ktorý pracuje systematicky a k dispozícii má nekonečne veľa času a nekonečnú zásobu papiera a ceruziek“. Turingov univerzálny výpočtový (!) stroj pritom disponoval súborom pevne zabudovaných či východiskových inštrukcií (angl. hard-wired instructions), ktoré mu umožňovali čítať a vykonávať zakódované inštrukcie vpísované programátorom na jeho pásku. Rozličné programy tak tomu istému *de facto mechanickému stroju*<sup>18</sup> umožňovali realizovať rozličné úlohy. C&P tak právom prirovnali tento stroj (tamže, s. 430) k chameleónovi, ktorý takisto dokáže svojim vonkajším výzorom optimálne reagovať na požiadavky a tlaky vonkajšieho prostredia. Ako ďalej upozornili, všetky moderné počítače sú v podstate univerzálne Turingove stroje, hoci s limitovanou pamäťou či trvanlivosťou.

V roku 1945 potom Turing prišiel so svojou ideou *automatického výpočtového stroja* (ACE – Automatic Computing Engine), ktorý sa stal akousi praktickou realizáciou jeho univerzálného výpočtového stroja. Počas prác na tomto *elektronickom a digitálnom stroji* však Turing sám seba prirovnal k „staviteľovi“ ľudského mozgu a dokonca poznamenal (tamže, s. 432), že sa viac zaujímal o možnosť vypracovať modely činnosti mozgu, ako o praktické aplikácie jeho stroja vo výpočtovej sfére. Bádanie vo sfére AI tak bolo prakticky od svojho začiatku spojené s nádejou, že sa v jeho rámci podarí objasniť, ako pracuje ľudský mozog alebo naša myseľ. O obrovskom Turingov optimizme v tomto ohľade svedčí aj jeho verejná prednáška z roku 1947, v ktorej odvážne definoval ľudský mozog ako „digitálny výpočtový stroj“ (C&P 2007, s. 432) a prehlásil, že to, čo chceme, je stroj, ktorý sa *dokáže učiť zo svojej skúsenosti*, pričom vyššie spomenutá možnosť meniť pôvodné inštrukcie je mechanizmom, ktorý *takéto učenie umožňuje*. V roku 1948 napokon Turing napísal, ale nepublikoval správu s názvom *Intelligent Machinery*, akýsi prvý manifest AI, v ktorom predložil *konceptuálne základy celej disciplíny*.

Okrem iného navrhol (tamže, s. 433), aby boli počítače budované z jednoduchých, neurónom podobných prvkov, ktoré by boli náhodne spojené *do (neurónových) sietí* a následne doslova a do písmena trénované na realizovanie špecifických úloh. Navrhol tiež koncept *genetického algoritmu*, ktorý je dôležitý aj v súčasnej AI a tzv. umelom živote. Veľkú pozornosť venoval aj výpočtovému modelovaniu umelého života a v jeho rámci jednej z kľúčových vlastností života

---

<sup>18</sup> Univerzálnosť tohto stroja pritom bola daná *špecifickým charakterom úloh, ktoré mal vykonávať*. O plne univerzálnom stroji by sme totiž mohli hovoriť až vtedy, ak by bol tento stroj spojený s nejakou 4D-tlačiarňou, resp. nanostrojom, schopným vyrobiť čokoľvek z čohokoľvek, ktorý je zatiaľ len hudbou vzdialenej budúcnosti.

ako takého – jeho sebaorganizácii. No a v tejto „Inteligentnej strojuvni“ (tamže, s. 434) vyčlenil aj dva principiálne odlišné prístupy k skúmaniu a „konštruovaniu“ AI, ktoré sa dnes bežne označujú ako prístup zhora nadol (top-down AI) a prístup zdola nahor (bottom-up AI), keď sa v rámci prvého prístupu AI analyzuje ako fenomén nezávislý od špecifických detailov mechanizmov (mozgu alebo elektroniky), ktoré ho realizujú, zatiaľ čo pri druhom prístupe je tomu naopak a bádatelia sa napríklad pomocou tzv. neurónových sietí snažia čo najvernejšie napodobniť, resp. modelovať prácu siete pri kreovaní výsledného obrazu určitých písmen.

Nás však zaujíma hlavne *problém autonómnosti AI* pri týchto prvých pokusoch o realizáciu Turingových myšlienok. C&P (2007, s. 437) najprv opisujú koncepciu *mikrosvetov AI*, ktoré môžu byť tak reálne, ako aj virtuálne, pričom reálne a virtuálne roboty v nich riešia rôzne úlohy. Mobilný robot Shakey, odvážne sa pohybujúci vo svojom mikrosvete pozostávajúcom zo stien, brán a niekoľkých drevených blokov, napríklad disponoval dvanástimi základnými schopnosťami ako „ZABOČ“, „POTLAČ“ alebo „ŠPLHAJ PO RAMPE“, ktoré Raphaelov a Nilssonov výskumný tím zo Stanfordu v rokoch 1968-1972 zahrnuli do jeho plánovacieho programu, pričom *robot sám (bez ľudskej intervencie)* na základe tohto programu zostavoval sekvencie svojich akcií alebo pohybov (C&P 2007, s. 438). A keďže Shakey a jeho viacerí nasledovníci disponovali interným modelom ich životných „svetov, t. j. „štruktúrou symbolov kódujúcich črty ich úlohovej domény“, zdalo sa, že môže byť správna Newellova a Simonova hypotéza z 50. rokov o tom, že spracovávanie symbolických štruktúr ľudským mozgom je skutočne základom ľudskej inteligencie a operovanie so štruktúrami symbolov, ktoré bude realizovať výkonný digitálny počítač, *tak bude skôr či neskôr stačiť na vytvorenie AI* (tamže).

Bohužiaľ, ako sa ukázalo, vypočítanie alebo „odhadnutie“ správnej trajektórie či manipulácie s reálnymi (alebo virtuálnymi) objektmi si nielen v prípade Shakeyho vyžadovalo veľmi veľa výpočtového času a hoci bol Shakey (a spolu s ním aj ďalšie miniroboty) navrhnutý tak, aby bola čo najviac redukovaná komplexnosť jeho interného modelu vonkajšieho sveta (tamže, s. 438-439), jednoduchá sekvencia jeho akcií si vyžadovala *viacero dní* na svoje zavŕšenie. S plnou naliehavosťou sa tak vynoril *rámcový problém* Newellovej a Simonovej symbolickej AI, súvisiaci s neschopnosťou vtedajších verzií AI manipulovať s rozsiahlymi štruktúrami symbolov v realistických časových intervaloch. Veľké nádeje spočiatku vzbudili *aj expertné systémy*, t. j. programy disponujúce rozsiahlou poznatkovou bázou v príslušnej vednej alebo aj praktickej oblasti, ktoré často dokázali odpovedať na otázky typu „ak-potom“ lepšie ako ich ľudskí „konkurenti“, ale časom sa ukázalo, že nedokážu „zohľadniť“ svoje ciele, hranice ich aplikovateľnosti alebo principiálne dôležitú otázku súladu ich odporúčaní so širším životným kontextom. Pri riešení mnohých problémov tak boli nepoužiteľné, pričom vytvoriť expertný systém, disponujúci tzv. zdravým rozumom a obrovským množstvom poznatkov potrebných pre každodenný život, sa podľa C&P sotva niekomu v dohľadnej dobe podarí (tamže, s. 440).

Riešenie problému opäť naznačili Turingove texty z konca 40. a začiatku 50. rokov (C&P 2007, s. 443), v ktorých Turing argumentoval v prospech rozvinutia dvoch zásadne odlišných foriem AI, a to akejsi substrátovo nezávislej (angl. *disembodied*) AI, resp. „odtelesnených“ mysliacich strojov, ktoré dokážu realizovať len abstraktné aktivity, medzi ktoré patrí podľa Turinga napríklad šach, a „úplne telesnou“ (angl. *embodied*) AI, pri tvorbe ktorej sa bádatelia a konštruktéri snažia pridať výpočtovému stroju zmyslové „orgány“ čo najviac sa približujúce ľudským a potom ho naučiť porozumieť i používať angličtinu vo vyučovacom procese, ktorý sa v princípe ničím neodlišuje od normálnej výuky angličtiny v škole. V tomto duchu potom vystúpil s kritikou „odtelesnenej“ symbolickej paradigmy AI H. L. Dreyfus (1992 a i.), ktorý zdôraznil, že je veľmi naivné predpokladať, že umelé výpočtové systémy, ktoré manipulujú s jednoznačne definovanými a z kontextu vytrhnutými prvkami „podľa presných pravidiel“

(C&P 2007, s. 443), budú pritom „vykazovať kognitívnu expertízu na ľudskej úrovni“. <sup>19</sup> Ako ďalej podotkol Dreyfus, tento problém vzniká preto, lebo rozhodnutie o tom, ktoré pravidlo je vhodné použiť pri riešení daného problému, závisí od situovania agenta v reálnom svete, a to v situácii, ktorej obsah „nie je redukovateľný na súbor faktov a pravidiel“ (tamže, s. 444).

Reakciou na túto a jej podobné kritiky bolo sformulovanie novej, resp. „situovanej“ AI, a to Rodneyom Brooksom z MIT AI Laboratory, ako takého smeru vo vývoji AI, ktorý podľa C&P (tamže) chápe kogníciu ako záležitosť bežných aktivít, zručností a know-how vychádzajúceho zo zdravého rozumu. Inteligencia je tak v Brooksovom chápaní viac záležitosťou správania ako interných reprezentácií, pričom v danom konceptuálnom rámci je budovaná *zdola nahor* a nie na základe príkazov z centrálného procesora (keďže nie je možné presne určiť miesto, kde vzniká), ale v rámci interakcie jednoduchých aktov. Pred „situovanými“ robotmi sa tak v reálnom prostredí vynárajú konkrétne úlohy, ktoré musia riešiť, podobne ako Brooksov robot Herbert, ktorý, zberajúc a odnášajúc prázdne plechovky zo stolov, zároveň vykazoval cieľovo orientované správanie generované z interakcie zhruba pätnástich jednoduchých aktov. „Eliminovaním komplikovaných interných modelov“ sa tak bádatelia v rámci novej AI snažili vyhnúť vyššie spomenutému rámcovému problému, alebo, ako ich prístup zhrnul J. Kelemen, (2010, s. 260) navrhovať ich agentov tak, aby dokázali bezprostredne reagovať na zmeny vonkajšieho prostredia, riešiť elementárne úlohy a v rámci intenzívnych interakcií ich prvkov postupne generovať čoraz cieľavedomejšie, autonómnejšie a teda aj racionálnejšie správanie.

Najznámejším Brooksovým robotom je pritom humanoidný Cog, resp. jeho vrchné torzo (C&P 2007, s. 445), o ktorom jeho tvorcovia smelo tvrdia, že je *autonómny* prinajmenšom „v tom zmysle, že jeho správanie nie je jednoducho výsledkom intervencie programátorov“, a objavujú sa aj nemenej odvážne tvrdenia, že jeho tvorcovia získavajú viac informácií o jeho interných stavoch z jeho vlastných výpovedí ako z dostupných a detailných výpočtových dát. Podobne optimistické tvrdenia sa objavili aj pri ďalšom robotovi z MIT, Kismetovi, ktorý sa údajne počas *reálnej komunikácie* s ľuďmi správa podobne ako malé dieťa, zatiaľ čo iní autori sú oveľa opatrnejší a rovnako ako F. Hara (2004, s. 9) konštatujú, že neexistuje žiadna reálna komunikácia „medzi človekom a sociálnym robotom“. Podľa C&P (2007, s. 446) sa takto len znova a znova potvrdzuje, že ľudia sú veľmi náchylní uveriť tomu, že „inteligentný“ robot je ľudskou bytosťou, ako je zrejme napríklad zo správania M. Yamamotu (1993), ktorý svojmu „hormónmi poháňanému“ (!) robotickému vysávaču prisúdil radosť, zúfalstvo, smútok a pod. Optimistické správy o úspechoch „situovaných“ robotov preto treba brať s veľkou rezervou.

## Umelá inteligencia a jej kvalitatívne stupne

Ak nebudeme inteligenciu chápať ako nejakú všeobecnú vlastnosť či už živého, ľudského alebo umelého (napríklad počítačového) systému, resp. ako súbor typických vlastností, ktoré ju spoločne generujú, potom ju môžeme prisúdiť aj tým umelým, t. j. človekom vyrobeným či skonštruovaným systémom, ktoré je realizujú len čiastočne. V takom prípade úplne vystačíme s jednoduchou klasifikáciou Tima Urbana (2015a, 2015b), ktorý rozlišuje medzi 1. čiastočnou alebo úzko vymedzenou AI (angl. ANI – artificial narrow intelligence), 2. generalizovanou či všeobecnou AI (angl. AGI – artificial general intelligence) a 3. superinteligenciou (angl. ASI – artificial superintelligence), pričom ANI sa podľa neho vyznačuje tým, že sa realizuje len v jednej oblasti, zatiaľ čo jej rôzne formy alebo (všetkým dobre známe) aplikácie sú už dávno

---

<sup>19</sup> Nemali by sme však byť k Allenovi Newellovi a jeho symbolickej paradigme AI nespravodliví, pretože, ako uvádza Jozef Kelemen (2010, s. 158), Newell v roku 1980 uvažoval aj o výpočtovom zariadení „intelektuálneho typu“, ktoré by dokázalo nielen pracovať v reálnom čase, ale aj sústavne sa zdokonaľovať, a to na základe svojej situovanosti do prostredia, ktoré by mohlo ovplyvňovať a dané prostredie by zase spätne ovplyvňovalo jeho.

súčasťou nášho každodenného života, pretože inteligentne usmerňujú naše vyhľadávanie na webe, odstraňujú spam z našej elektronickej pošty a i. AGI sa už vyrovná svojou výpočtovou silou a kreativitou ľudskej inteligencii, keďže by mala byť schopná vyriešiť tie isté úlohy, aké dokáže vyriešiť súčasný človek, a rovnako ako on učiť sa na základe vlastnej skúsenosti.

ASI napokon predstavuje počítač, ktorý je podľa Urbana (a Nicka Bostroma, na ktorého sa tu Urban odvoláva) oveľa bystrejší ako akýkoľvek ľudský mozog v akejkoľvek oblasti, či už ide o vedeckú kreativitu, múdrosť ako takú alebo najrozmanitejšie sociálne zručnosti.<sup>20</sup> Keďže až donedávna v odbornej literatúre chýbali všestranne a dôsledne vymedzené kritériá, na základe ktorých by bolo možné klasifikovať jednotlivé druhy inteligencie alebo umelej inteligencie, je zároveň prirodzené, že jednoznačný konsenzus ohľadom definovania AGI a ASI nejestvuje. Richard Loosemore a Ben Goertzel napríklad už v roku 2012 prepojili ideu AGI s ideou ASI a technologickej singularity Vernona Vingeho, keď uviedli, že ak má dôjsť k informačnej a poznatkovej explózii spojenej s touto singularitou (a v rámci nej pôsobiacou inteligenciou, ktorá je schopná rekurzívne alebo sústavne zdokonaľovať samú seba, vrátane svojho hardvéru a softvéru, a zvyšovať tak svoju inteligenciu), AGI sama musí byť natoľko inteligentná, aby dokázala (tamže, s. 84) „porozumieť svojmu vlastnému dizajnu“. Aj preto v súvislosti s ňou L&G hovoria aj o „zárodočnej“ (angl. seed) AGI, ktorá sa môže veľmi rýchlo rozvinúť alebo pretransformovať do ASI, resp. o „sebachápajúcej“ a „človeku rovnjej“ AGI (tamže, s 85).

Zaujímavé sú aj ich úvahy o tom, čo by nastalo v prípade, ak by sa inteligencia takejto AGI zvýšila o dva-tri rády (t. j. sto až tisícnásobne), ale na tomto mieste je pre nás dôležitejšie, že L&G pripomínajú aj staršie štúdie M. Huttera (2005) a B. Goertzela (2010), ktorý koncept AGI dokonca matematicky formalizovali, pričom Goertzel okrem iného (a to veľmi zábavne) poukázal v závere svojej štúdie aj na rozdiel medzi tzv. pragmaticky všeobecne *inteligentnou* AGI, ktorá bude veľmi úspešná pri riešení obmedzeného počtu ťažkých problémov a súčasne úplne nekompetentná pri riešení ešte väčšieho okruhu ťažkých problémov, a to v porovnaní s pragmaticky všeobecne *hlúpu* AGI, ktorá bude na rozdiel od nej v niektorých kontextoch oveľa hlúpejšia, ale pritom zhruba rovnako hlúpa vo všetkých kontextoch, t. j. aj v tých, v ktorých bude pragmaticky všeobecne inteligentná AGI hlúpejšia ako ona, v dôsledku čoho sa bude vyznačovať väčšou intelektuálnou šírkou a súčasne aj plytkosťou.<sup>21</sup> Širokú verejnosť zase v roku 2017 zaujala kniha Maxa Tegmarka *Život 3.0*, ktorej sa budeme venovať detailne v inej časti tejto štúdie, pričom na tomto mieste si všimneme len jeho definície AGI a ASI.

Už pri nich však vystupuje do popredia Tegmarkova „uvoľnená“ a často zmätočná či nejasná argumentácia, keď Tegmark na 51. strane citovanej knihy najprv tvrdí, že akýmsi posvätným cieľom výskumu vo sfére AI „je vytvoriť čo najširšie ‚umelé bytie‘ (známe skôr pod menom *artificial general intelligence*, AGI)“, ktoré „by dokázalo dosiahnuť prakticky akýkoľvek cieľ

---

<sup>20</sup> Nielen tu, ale aj na ďalších miestach nášho textu sa pritom vynára problém, ku komu alebo k čomu konkrétne máme vzťahovať označenie „ľudská inteligencia“? K nejakému priemerne inteligentnému človeku? K tej úzkej a nepočetnej skupine ľudí, ktorí sú veľmi inteligentní a túto svoju inteligenciu prejavujú aj v bežnom živote? Podľa nás ju môžeme vzťahovať k ľudstvu ako takému (a teda ku všetkým ľuďom v našich dejinách), pretože aj tí najinteligentnejší ľudia, ani nehovoriac o tých menej inteligentných, ale svojprávnych, sú v určitých situáciách odkázaní na pomoc druhých či spoluprácu s nimi. Namiesto o delbe práce by sme preto mohli hovoriť aj o tzv. funkčnej komplementárnosti, kedy vďaka spolupráci s inými ľuďmi rastie inteligencia a behaviorálna úspešnosť každého svojprávneho jedinca. Nielen to, mohli by sme dokonca začať porovnávať komplexnosť a následne aj inteligenciu určitých civilizácií, spoločností, komunit alebo kultúr. A všimnúť si a analyzovať nielen to, prečo a ako zanikajú, ako to urobil napríklad Joseph A. Tainter vo svojej slávnej knihe (2009), ale aj to ako vznikajú, resp. udržiavajú sa dlhodobo pri živote aj napriek čoraz väčším externým a interným tlakom (alebo pnutiam).

<sup>21</sup> Goertzel pritom svoju matematickú formalizáciu AGI zakladá na viacerých veľmi zaujímavých argumentoch, keď na rozdiel od Legga a Huttera s ich vyššie uvedenou definíciou inteligencie ako takej berie do úvahy (2010, s. 3-4) aj veľkosť a organizáciu pamäte príslušného inteligentného agenta, jeho efektívnosť pri využívaní daných výpočtových zdrojov, ale aj systém odmiern spojený s dosahovanými cieľmi v partikulárnom prostredí a pod.

vrátane učenia“ (hoci sám dobre vie, že táto definícia sa skôr a viac hodí na ASI), a následne sa prikláňa k Leggovmu, Gubrudovmu a Goertzelovmu vymedzeniu (a chápaniu AGI) ako umelého bytia *s inteligenciou na ľudskej úrovni*, teda takej AGI, ktorá je schopná „dosiahnuť akýkoľvek cieľ prinajmenšom rovnako dobre ako ľudia“, ale bez toho, aby tak ako L&G (2012, s. 84) poukázal na veľmi dôležitú schopnosť „zárodočnej“ AGI porozumieť vlastnému dizajnu a otvoriť si tak cestu k sústavnému a možno až exponenciálnemu (a explozívnemu) sebazdokonaľovaniu. Neveľmi presná a teda ani užitočná je aj Tegmarkova definícia (tamže, s. 13) „ultrainteligentného stroja“, resp. superinteligencie (ASI), ako takého stroja, „ktorý vo všetkých intelektuálnych činnostiach dokáže prekonať akokoľvek bystrého človeka“, a to aj napriek tomu, že Tegmark správne poukazuje na jeho schopnosť navrhovať ešte lepšie stroje, pretože nezohľadňuje ani tie faktory, na ktoré nižšie poukázal B. Goertzel (2010, s. 3-4).

Svojho času bola veľmi diskutovaná aj kniha Nicka Bostroma (2014), v ktorej Bostrom okrem iného vymedzil aj tri druhy ASI, a to 1. *rýchlu*, resp. „*rýchlostnú*“ ASI, ako systém, ktorý dokáže vykonať to isté, čo dokáže ľudský intelekt, „ale oveľa rýchlejšie“ (tamže, s. 53). Tento druh ASI môže byť napríklad stotožnený s ľudským mozgom postupne transformovaným či preneseným do počítačového hardvéru, ktorý však na rozdiel od pôvodného ľudského mozgu dokáže pracovať aj desaťtisíckrát rýchlejšie, prečítať knihu za niekoľko sekúnd či vypracovať celú doktorandskú prácu za jedno popoludnie. Keďže však jeho myseľ pracuje desaťtisíckrát rýchlejšie ako ľudská, všetko okolo neho v makrosvete mu pripadá zúfalo pomalé, a tak radšej komunikuje digitálne vo svojom kybersvete a nie s bytosťami, ktorých mozgy sa mu právom javia ako „prijbrzdené“ alebo ťažkopádne. 2. *kolektívnu ASI* ako systém (tamže, s. 54) zložený „z veľkého počtu menších intelektov“, a to takým spôsobom, že jeho celková výkonnosť vo viacerých veľmi všeobecných doménach značne prevyšuje výkonnosť *akéhokolvek súčasného kognitívneho systému*. A 3. *kvalitatívnu ASI* ako systém, ktorý síce pracuje rovnako rýchlo ako ľudská myseľ, *ale kvalitatívne je na oveľa vyššej úrovni*, pretože na rozdiel od nás má napríklad mozog disponujúci ďalšími modulmi s viacerými novými funkciami (tamže, s. 56).

Výrazný posun v diskusii o inteligencii ako takej a umelej inteligencii však evidujeme až po vydaní publikácie Christiana Huga Hoffmana (2022), z ktorej takisto „vyselektujeme“ len tie najdôležitejšie myšlienky. V prvom rade úplne súhlasíme s jeho tvrdením (tamže, s. 245), že skôr, ako začneme vyvíjať a rozširovať AI, mali by sme pochopiť alebo zmerať (angl. fathom) obsahové bohatstvo pojmu inteligencia. Rovnako ako on sme tiež presvedčení, že v dôsledku toho, že moderné systémy AI, na rozdiel od malých detí či väčšiny živočíchov, nemôžu byť len tak umiestnené do nových životných prostredí a v nich inteligentne (a úspešne) fungovať (Hoffmann 2022, s. 82), nemožno o nich tvrdiť, že sa dostali na úroveň AGI.<sup>22</sup> Vo vzťahu k inteligencii ako takej a súčasne aj AI je veľmi inšpiratívny aj jeho štvorpríčekový rebrík kognitívnych schopností či daností (tamže, s. 100), na spodku ktorého je *poznanie*, nad ním *inteligencia*<sup>23</sup>, potom *porozumenie* a nakoniec *múdrosť* ako schopnosť holistického posúdenia situácie, pomocou ktorej sú posudzované nielen spôsoby, akými sú ciele dosahované, ale aj ciele samotné, čoho sú však zjavne schopné len plne inteligentné a autonómne systémy.

---

<sup>22</sup> Hoffman tu pripomína známy *Moravcov paradox*, podľa ktorého je relatívne ľahké naučiť AI niektorým schopnostiam či zručnostiam dospelých ľudí (ako napríklad skvelo hrať dámu alebo riešiť úlohy v niektorých intelligenčných testoch), ale súčasne veľmi ťažké poskytnúť im „zručnosti jednoročného dieťaťa, keď sa vezme do úvahy percepčia a mobilita“ (Moravec 1988, s. 15). A to preto, lebo súčasným systémom AI chýba právo to, čím disponuje väčšina zvierat a (postupne) aj malých detí, konkrétne: 1. *robustnosť a odolnosť voči vonkajším hrozbám*, 2. *flexibilita* (ako schopnosť využívať poznanie získané pri plnení jednej úlohy pri riešení druhej) a 3. *autonómnosť* (ako schopnosť konať bez nutnosti sústavného vonkajšieho dohľadu či opakovaných opráv).

<sup>23</sup> Vallorom (2017) a Hoffmanom (2019) klasifikovaná ako relatívne úzko definovaná či „zameraná“ schopnosť aplikovať získané poznanie v relevantných kontextoch (alebo, ak chcete, životných a civilizačných prostrediach).

Na diskusiu je už Hoffmanovo tvrdenie (tamže, s. 238), že v blízkej budúcnosti sa nebude realizovať žiadny exponenciálny a (nami) nekontrolovateľný nárast inteligencie v spomínanej Vingeho technologickej singularite, pretože je oveľa pravdepodobnejšie, že to budeme práve my, ľudia, ktorí budeme musieť „sami naprogramovať väčšinu inteligencie do AI“, keďže SAI (angl. strong AI, silná umelá inteligencia) nebude môcť fungovať a rozvíjať sa bez vedeckých poznatkov, zdravého rozumu (angl. commonsense), edukácie a určitých dopredu zadaných či „vrodených“ (angl. innate) poznávacích modelov alebo šablón. Veľmi užitočné sa nám nezdá byť ani ním ponúkané rozlíšenie silnej a slabej AI, keď podľa Bringsjorda a Govindarajulua (2018) silná AI predstavuje výskumný program, v rámci ktorého sa bádatelia snažia vytvoriť *počítačové programy*, „ktoré premýšľajú rovnakým spôsobom“, v rovnakej hĺbke a rovnako rozmanito (alebo plodne, angl. richness) ako ľudia, zatiaľ čo slabá AI sa podľa Robinsona (2015, s. 65) uspokojuje len s definovaním počítačov ako dôvtipného nástroja „na rigorózne formulovanie a testovanie hypotéz o ľudskej mysli“, t. j. slúži (alebo mala by slúžiť) tým cieľom, ktoré AI na prelome 40. a 50. rokov minulého storočia zadal ešte Alan Turing.

A hoci sa stotožňujeme s Hoffmanovým tvrdením, že „akčnosť“ (angl. agency) je vlastnosť, ktorá je merateľná v strojach, živočíchoch alebo systémoch vôbec, pretože sa realizuje skôr v stupňoch ako binárne (t. j. je v nich prítomná, alebo úplne absentuje), a to s ohľadom na ich autonómnosť, a veľmi pozitívne hodnotíme aj jeho cieľavedomú snahu vypracovať a obhájiť kritériá na klasifikovanie jednotlivých druhov inteligencie a AI (napríklad na základe jeho kauzálnej teórie inteligencie), predsa len si myslíme, že jeho prisúdenie inteligencie všetkým živočíchom je pri súčasnom stave ich vedeckého poznania prinajmenšom sporné. Ak si totiž pozrieme tabuľky s vlastnosťami ľudských a živočíšnych (alebo čisto biologických) systémov (Hoffmann 2022, s. 40), tak medzi tými druhými nájdeme aj také znaky inteligencie, ktorých prisúdenie zvieratám je naozaj veľmi otázne. Hlboko pochybujeme najmä o tom, že zvieratá komunikujú pomocou symbolov s rozmanitým a od kontextu závislým významom *v tom istom rozsahu a intenzite ako ľudia*, a ešte otáznejšie je prisúdiť im používanie nástrojov, ako je zrejme aj zo Shumakerovho et al. (2011, s. 223 a n.) špecifikovania ľudského používania nástrojov, podľa ktorého 1. tzv. neľudské zvieratá (t. j. ostatné živočíchy *okrem* človeka!) „nepoužívajú nástroje na výrobu iných nástrojov“, 2. s výnimkou veľkých opíc zriedkavo používajú viacero nástrojov v kombinácii na riešenie jednej úlohy, 3. „zriedkavo spolupracujú pri užívaní alebo výrobe nástrojov“ a 4. pri používaní týchto nástrojov využívajú len vlastnú alebo gravitačnú energiu.<sup>24</sup> Navyše, ako uvidíme v nasledujúcej časti, zapochybovať môžeme dokonca aj o tom, či pre súčasné verzie AI naozaj platí, že zatiaľ nedosiahli úroveň AGI.

## Large language models a ich najdôležitejšie vlastnosti, resp. parametre

Hneď na úvod musíme zdôrazniť, že vývoj v tejto výskumnej oblasti, t. j. vo sfére veľkých jazykových modelov a ich praktických aplikácií (ako je známy ChatGPT), je aktuálne natoľko rýchly, že veľa z toho, čo uvedieme, môže byť v čase, keď bude táto štúdia publikovaná (t. j.

<sup>24</sup> Je možné pritom namietnuť, že vyššie uvedená definícia inteligencie od Legga a Huttera môže byť bez väčších problémov aplikovaná aj na zvieratá, rovnako ako na človeka – veď do životného areálu živočicha môže byť skutočne zahrnutých viacero permanentne sa meniacich prostredí, biotopov a pod., ale pri takomto subsumovaní týchto entít pod jedno označenie (neľudské a ľudské živočíchy!) strácame napríklad zo zreteľa veľmi dôležitú skutočnosť, že ľudia sa *nielen prispôbujú svojmu životnému prostrediu*, ale ho aj zámerne *tvoria* a čoraz väčšiu časť svojho života trávajú v tomto *umelo vytvorenom* civilizačnom prostredí, čo je nakoniec aj v úplnom súlade s ich vymedzením ako psycho-sociálne-kultúrnych a nie (čisto) biologických entít; s tým dôležitým dodatkom, že do kultúrneho sa tu v súlade s tradíciami archeologického výskumu zahrňuje aj technické či technologické, pričom samotné (biologické) telo človeka je *sústavne pozmeňované* technickými (medicínskymi a i.) zásahmi, bez ktorých si jeho plnohodnotné a dlhodobé fungovanie v modernej dobe nemôžeme ani len predstaviť.



v decembri 2023) úplne nerelevantné. Aj z priestorových dôvodov, keďže nás zaujíma najmä to, či v blízkej budúcnosti bude vytvorená AGI (alebo ASI) a ako to ovplyvní naše chápanie ľudskej prirodzenosti, sa obmedzíme len na uvedenie tých najzákladnejších informácií. Large language models (LLM), veľmi jednoducho povedané, sú *umelé neurónové siete*, umiestnené vo vysoko výkonných počítačoch, ktoré sa pomocou tzv. generatívnych, dopredu trénovaných transformerov (t. j. algoritmov) „snažia“ čo najpresnejšie modelovať (a napodobniť) ľudský jazyk, a to tak z hľadiska jeho formy, ako aj obsahu, pričom pri spracúvaní čoraz väčšieho objemu dát (najmä z internetu) sú zjavne schopné postupne si generovať vlastné pravidlá ich spracúvania a realizovať takto tzv. stelesnené (angl. embodied) poznanie, t. j. nielen poznanie niečoho (to know what), ale aj poznanie samotných procesov poznania (to know how).

Ako pritom uvádza Billy Perrigo 13. 4. 2023 v časopise Time, slovo „model“ je skráteným označením pre akýkoľvek samostatný AI systém, bez ohľadu na to, či ide o základný model, akým je napríklad GPT-4, alebo na ňom budovanú aplikáciu či nadstavbu, akou je napríklad spomenutý ChatGPT, využívajúci však pri svojej komunikácii s ľuďmi len časť poznatkovej bázy GPT-4. Výsledkom práce umelých neurónových sietí, ktoré sú akýmsi analógom nášho mozgu, sú tak rôzne predikcie a klasifikácie, resp. *vzory a pravidelnosti vo výstupných dátach*, ktoré v rastúcej miere pripomínajú vzory alebo pravidelnosti zo spracúvaných dát. Čím viac výpočtovej sily má pritom daný systém k dispozícii, tým viac umelých neurónových spojení môže vytvoriť a zároveň identifikovať čoraz abstraktnejšie (alebo všeobecnejšie) vzory, resp. pravidelnosti.<sup>25</sup> S tým úzko súvisia tzv. škálovacie zákony, podľa ktorých výkonnosť každého LLM *predvídateľne* vzrastá s väčším množstvom tréningových dát, výpočtovej sily a veľkosti umelej neurónovej siete. Podľa Sama Bowmana z AI laboratória Anthropic (Perrigo 2023) tak vzniká v dejinách AI úplne nová a bezprecedentná situácia, pretože firmy rozvíjajúce novú AI môžu *ešte pred samotným trénovaním* ich LLM *relatívne presne určiť*, koľko výpočtovej sily, a tým pádom aj financií budú tieto LLM potrebovať na realizovanie nimi zvolenej úlohy, napríklad úspešného zvládnutia písomnej skúšky z angličtiny na vysokoškolskej úrovni.

Nás však zaujíma hlavne to, či vôbec, a ak predsa len áno, máme my, ľudia, možnosť vidieť a chápať, ako LLM pracujú, resp. ako ďaleko sme aktuálne od vytvorenia AGI (resp. ASI). Cal Newport 13. 4. 2023 v časopise New Yorker napríklad tvrdí, že na práci LLM nie je nič osobitne zaujímavé a pocit hrozby z výpočtovej či intelektuálnej sily nových foriem AI vzniká len preto, lebo ich stále vnímame ako „mystické čierne skrinky“, do fungovania ktorých my, ľudia, nevidíme a ani nebudeme vidieť. Ak si však nájdeme čas, a táto veta je v Newportovom článku kľúčová, a preskúmame ako tieto technológie skutočne pracujú, postupujúc pritom od konceptov najvyššej úrovne až k základnému digitálnemu hardvéru či inštrumentáriu, potom *dokážeme porozumieť* tomu, s čím sme sa dostali do styku. Nemáme tu, pravdaže, priestor na uvedenie celej (veľmi fundovanej) Newportovej argumentácie a ani ju uviesť nepotrebujeme, pretože pre nás je dôležitý *názorový konsenzus* v tejto výskumnej sfére, v rámci ktorej Samuel R. Bowman (2023, s. 6) napríklad tvrdí, že aj keď máme niekoľko „hrubozrnných“ nástrojov, pomocou ktorých môžeme zistiť, či daný model reprezentuje nejaký zvláštny druh informácie, začiatkom roku 2023 nedisponujeme žiadnou technikou, pomocou ktorej by sme vedeli určiť, aký druh poznania, uvažovania či cieľov používa daný model, keď produkuje svoje výstupy.

Ako ďalej zdôrazňuje S. R. Bowman (tamže), pri *stovkách miliárd* spojení medzi umelými neurónmi, keď sú niektoré z nich aktivované viackrát už počas spracúvania krátkeho textu, je

---

<sup>25</sup> Dve najznámejšie laboratória AI – OpenAI a Deep Mind – preto pri budovaní a trénovaní svojich LLM využívajú superpočítače zložené z obrovského počtu grafických procesorov (GPU), pričom úzko spolupracujú v prvom prípade s Microsoftom a v druhom s materskou firmou Alphabet. Patrí sa tiež uviesť, že ChatGPT nie je na trhu osihotený a popri ňom jestvujú aj iné chatboty, ako Bard firmy Google, Bing firmy Microsoft a i.

každý náš pokus o precízne vysvetlenie správania sa LLM dopredu odsúdený na neúspech. Úspech nepriniesli ani tzv. ad hoc techniky, ktoré nám mali pomôcť pri porozumení tomu, čo sa vlastne v LLM pri ich práci deje, pretože aj tie sa onedlho ukázali ako silne zavádzajúce. No a nepodarilo sa ani potvrdiť korešpondenciu medzi technikami uvažovania používanými v prirodzených jazykoch a procesmi, ktoré LLM realizujú pri svojom „uvažovaní“, a zatiaľ zlyhali aj všetky pokusy vysvetliť správanie LLM na základe čoraz presnejšieho modelovania ich správania. Pri takejto miere našej epistemickej neistoty je o to zaujímavejšie, že niektorí bádatelia (resp. celé výskumné tímy) sa domnievajú, že v posledných verziách LLM a GPT sa objavujú prvé prejavy AGI. Veľmi zaujímavá a podnetná je z tohto hľadiska najmä nedávna štúdia Sébastiena Bubecka et al. (2023), v abstrakte ku ktorej sa najprv tvrdí, že LLM GPT-4 je len jedným z viacerých LLM (ako sú ChatGPT alebo PaLM firmy Google), ktoré prejavujú viac všeobecnej inteligencie ako predošlé AI modely. GPT-4, ako model určený prednostne na spracúvanie textov, pritom dokáže riešiť široké spektrum úloh, siahajúcich od matematiky, cez programovanie (alebo kódovanie), vizualizácie, medicínu, právo až po psychológiu.

Podľa Bubecka et al. sa zároveň výkonnosť GPT-4 pri riešení týchto rozmanitých úloh, resp. problémov výrazne približuje ľudskej výkonnosti a veľmi často ďaleko presahuje výkonnosť predošlých modelov, ako bol napríklad GPT-3 alebo verejnosti už dlhšie dostupný ChatGPT. Vzhľadom na šírku a hĺbku jeho schopností preto podľa nich GPT-4 môže byť považovaný za prvú, nie úplne kompletnú verziu všeobecne inteligentného (t. j. AGI) systému. Bubeck et al. sa však neuspokojujú s týmito zisteniami a v ich štúdiu (Bubeck et al. 2023, s. 1) posudzujú aj limity GPT-4, aby tak všetkým bádateľom otvorili cestu k vytvoreniu ešte všestrannejších či hlbšie uvažujúcich verzií AGI vrátane tých, ktoré už prekonajú tradičnú paradigmu LLM, t. j. predikcie nasledujúceho slova alebo slovných spojení. V nasledujúcich pasážach ich štúdie však nachádzame viacero tvrdení, ktoré sa do určitej miery navzájom popierajú, naznačujúc tak ich neistotu ohľadom toho, či sú tvrdenia o GPT-4 ako prvej verzii AGI naozaj správne. Na 4. strane ich štúdie napríklad hodnotia GPT-4 len ako „významný krok smerom k AGI“, zatiaľ čo na 6. strane už konštatujú, že vytvorením GPT-4 sa podarilo realizovať významný technologický skok, „skutočný paradigmatický posun“ vo sfére AI, a to aj napriek tomu, že vyššie sami hovorili o potrebe posunúť sa v danej sfére za predikciu slov a slovných spojení.

Bubeck et al. potom pokračujú prezentovaním a dokazovaním predností GPT-4, pričom viac (alebo bližšie) analyzujú tie jeho schopnosti, ktoré uviedli v abstrakte ich štúdie. Na 13. strane napríklad konštatujú, že jedným z kľúčových kritérií pri prisudzovaní a meraní inteligencie je schopnosť syntetizovať informácie z rôznych domén a následne aplikovať získané vedomosti a skúsenosti v mnohorakých kontextoch či disciplínach, pričom GPT-4 podľa nich skutočne dokáže kombinovať zručnosti a koncepty z rôznych domén a preukazovať súčasne brilantné „porozumenie komplexným myšlienkam“. Na 29. strane zase vyzdvihujú schopnosť GPT-4 programovať priamo, bez prekladu do „iných, dobre definovaných programovacích jazykov“, demonštrujúc tak svoj potenciál stať sa revolučným programovacím nástrojom, ktorý môže principiálne zmeniť spôsob, akým sa bude programovať v budúcnosti. Na 60. strane dokonca tvrdia, že počas ich testovania GPT-4 preukázal vysoko rozvinutú teóriu inej mysle, zatiaľ čo na 67. strane vyzdvihujú existenciu ďalšej schopnosti, ktorá dokazuje prítomnosť všeobecnej inteligencie u GPT-4, a to schopnosti vysvetľovať vlastné správanie konzistentne a dostatočne presvedčivo s ohľadom na príslušné vstupné dáta a kontext. Nakoniec, pri porovnaní GPT-4 s GPT-3 (tamže, s. 69-70) sa tiež ukázalo, že GPT-4 je oveľa úspešnejší aj pri určovaní tzv. *osobne identifikovateľnej informácie* a jej jednotlivých segmentov, resp. ich aktuálneho počtu.

Na 76. strane však Bubeck et al. už hovoria o limitoch „autoregresívnej architektúry GPT-4“, keď poukazujú na niekoľko podstatných slabín (angl. flaws) GPT-4, ktoré sú podľa všetkého inherentné jeho architektúre založenej na paradigme predikcie nasledujúceho slova a slovných

spojení. Daná architektúra totiž neumožňuje GPT-4, aby viedol pri plánovaní budúcich akcií „vnútorný dialóg“, resp. dokázal si vytvoriť ich rýchly, operatívny náčrtok alebo zápis (angl. scratch-pad) nad rámec jeho interných reprezentácií, čo by mu potom umožňovalo realizovať viackrokové výpočty a súčasne operatívne uchovávať priebežné výsledky. V dôsledku toho, ako ďalej podotýkajú Bubeck et al. (tamže, s. 80), je GPT-4 veľmi úspešný „pri produkovani plynulých a koherentných textov“, ale súčasne silne znevýhodnený či limitovaný v prípade, ak má riešiť komplexné a kreatívne úlohy, ku ktorým sa už „nedá pristupovať sekvenčným spôsobom“. Nedokáže preto ani do hĺbky alebo globálne porozumieť zadaným úlohám, resp. svojim výstupom. Daná sekvenčná (alebo lineárna) architektúra ho takisto limituje pri riešení úloh, ktoré už nie sú riešiteľné postupne a kontinuálne, ale vyžadujú si (tamže, s. 81) určitú „skokovú“, prekvapujúcu alebo heuristickú ideu (typu novej vedeckej hypotézy, ale aj vtipu či hádanky). V 9. a predposlednej časti ich textu sa Bubeck et al. venujú sociálnym dôsledkom rozšíreného používania LLM typu GPT-4, z nášho hľadiska je však zaujímavejšia záverečná, 10. časť ich textu, kde sa zamýšľajú nad tým (Bubeck et al. 2023, s. 92 a n.), čo všetko, resp. aké kroky treba spraviť preto, aby sme v dohľadnej dobe dokázali vytvoriť úplnú AGI.

Podľa nich v prvom rade by sme mali nanovo definovať AGI, potom doplniť chýbajúce časti (alebo zložky) LLM a nakoniec lepšie porozumieť tomu, ako vzniká inteligencia prejavujúca sa v súčasných LLM. Počas diskusie o prvom kroku tak Bubeck et al. kritizujú definíciu AI a inteligencie vôbec od Legga a Huttera, podľa ktorých inteligencia vyjadruje mieru agentovej schopnosti dosahovať svoje ciele v rozmanitých prostrediach. Podľa nich však táto definícia nezachytáva celé spektrum inteligencie, pretože vylučuje (tamže, s. 92) všetky „pasívne alebo reaktívne systémy“, ktoré dokážu riešiť komplexné úlohy alebo odpovedať na zložité otázky bez toho, aby disponovali „vnútornou motiváciou alebo cieľom“. Poľahky si totiž môžeme predstaviť nejakú AGI v podobe *brilantného orákula*, ktoré nemá žiadne preferencie či ciele a napriek tomu je schopné poskytnúť užitočné a presné informácie z akejkoľvek problémovej či tematickej oblasti. Požiadavka na dosahovanie úspešnosti v širokom rozsahu prostredí tiež nie je podľa Bubecka et al. realistická, pretože *ani ľudská inteligencia* nie je „univerzálna či optimálna“. Problematická je podľa nich aj iná definícia AGI od Legga a Huttera (2007b) ako systému, ktorý dokáže urobiť všetko, čo dokáže človek, pretože táto definícia podľa Bubecka et al. pripisuje *každému* človeku schopnosti, ktoré určite mať nemôže, a zároveň sa vyznačuje antropocentrickým biasom, ktorý de facto „vyraďuje z hry“ viaceré možné AI systémy.

Na druhej strane, Bubeck et al. sami priznávajú, že je na zamyslenie, či môže byť inteligencia dosiahnutá bez toho, aby mala *vnútornú motiváciu a možnosť reálne konať* aj v inom ako len virtuálnom prostredí. Potom už uvádzajú jednotlivé časti LLM, ktoré treba rozvinúť či doplniť (napríklad dlhodobú pamäť, kontinuálne učenie spojené so schopnosťou „inkorporovať“ novú informáciu alebo spätnú väzbu z vonkajšieho prostredia, personalizáciu, schopnosť dlhodobo plánovať a robiť heuristické poznávacie „skoky“, resp. transparentnosť či konzistentnosť), ale nás v závere tejto časti viac zaujíma porozumenie tomu, ako GPT-4 a ďalšie LLM dosahujú svoje pozoruhodné „intelektuálne“ výkony. Veľa sa diskutuje o emergencii nových vlastností, ale napriek intenzívnemu záujmu o ne (tamže, s. 94) progres v tejto oblasti bol zatiaľ veľmi čiastkový a obmedzil sa na úplne jednoduché modely, v rámci ktorých boli potvrdené niektoré príznaky emergencie. Podľa jednej z veľmi všeobecných hypotéz (tamže, s. 94-95) obrovské množstvo rozmanitých dát podnecuje umelé neuronové siete k tomu, aby vytvárali generické a užitočné neuronálne okruhy, zatiaľ čo mohutnosť jazykových modelov poskytuje umelým neuronálnym okruhom dostatok redundancie a variability na to, aby sa mohli špecializovať či naladiť na špecifické úlohy. Mohutnosť LLM tiež umožňuje vytvárať vzostupné gradienty či úrovne a potom spájať odlišné minimá, resp. ľahko prepájať mnohorozmerné dáta. V každom prípade, ako na záver konštatujú Bubeck et al., úloha objasniť povahu a mechanizmy nových AI systémov typu GPT-4 sa naozaj v súčasnosti stala mimoriadne „dôležitou a urgentnou“.

## Umelá inteligencia – najväčšie riziká, hrozby a možnosti nekonfliktného spolužitia

S každou vyspelou technológiou prichádzajú riziká jej nesprávneho a nemorálneho použitia. Pri ANI tieto riziká a hrozby (ak odhliadneme od niektorých foriem závislosti pri jej častom používaní, resp. tlaku na „digitalizovateľné“ alebo výpočtovou technikou ohrozené pracovné miesta) nie sú relatívne vysoké, zatiaľ čo pri AGI a ASI sa vynára celý súbor hrozieb a rizík, ktoré sú v odbornej literatúre už dlhšie diskutované, podľa Romana V. Yampolského (2016, s. 135) však príliš nekonkrétne alebo všeobecne, s nedostatočným porozumením tomu, akou hrozbou sa môže stať AI rekurzívne zdokonaľujúca samu seba. Niektoré problémy sú pritom dané tým, že pri budovaní AGI sme na samom začiatku a LLM, ktoré sú prvým krokom k jej vytvoreniu, vykazujú niektoré poruchy a zlyhania, ktoré silne znepokojujú ich tvorcov (resp. samotných investorov), pretože majú potenciál odradiť zákazníkov od ich pravidelného alebo systematického užívania. Najznámejšie zo všetkého sú určite tzv. *halucinácie*, B. Perrigom (2023) definované ako falošné informácie, ktoré vznikajú v dôsledku chybného spracovania vstupných dát, keď niektoré chatboty poskytli užívateľom napríklad nezmyselné medicínske rady, ich texty obsahovali citácie neexistujúcich článkov a pod. Potešiteľnou správou však je, že s rastúcou výkonnosťou LLM modelov sa počet podobných chýb darí čoraz viac znižovať.

Ďalším problémom je *bias*, t. j. tendencia chatbotov poskytovať informácie, ktoré ukazujú v lepšom svetle niektorú skupinu ich užívateľov, čo súvisí s tým, že LLM do určitej miery len reprodukuje vzory a pravidelnosti vo vstupných dátach a ak je v nich takýto bias zahrnutý, t. j. prevaha tvrdení, napríklad, že Afroameričania v USA sú častejšie páchatelmi kriminálnych činov ako belosi, potom tento bias reprodukuje vo svojom výstupe aj príslušný chatbot. Aby sa predišlo takémuto správaniu chatbotov, namiesto učenia sa bez ľudského dohľadu sa dáva prednosť učeniu LLM s ľudským dohľadom, pričom sa cielene používajú AI nástroje, ktoré umožňujú detegovať a následne odstrániť škodlivý obsah z výstupných dát. S. R. Bowman (2023, s. 5) zase upozorňuje na bizarné *manipulatívne správanie* chatbotu Bing od Microsoftu voči jeho užívateľom a B. Perrigo v tejto súvislosti hovorí o tzv. *shoggoths*, t. j. postavkách z Lovecraftových hororov, ktoré pod milým úsmevom skrývajú násilné alebo manipulatívne správanie, aké začiatkom roku 2023 prejavil voči užívateľom spomenutý chatbot Bing. Ďalšie Perrigom spomínané negatívne vlastnosti tejto technológie sú už banálnejšie, predovšetkým jej viackrát spomínaná výpočtová a energetická náročnosť, keď nezávislí bádatelia odhadli, že pri trénovaní GPT-3 sa do atmosféry uvoľnilo viac ako 500 ton oxidu uhličitého.

Pablo Villalobos et al. (2022, s. 1) súčasne upozorňujú aj na riziko toho, že pri doterajšom tempe spotreby dát sa vysoko kvalitné jazykové dáta vyčerpajú veľmi pravdepodobne už do konca roku 2026 a nízko kvalitné jazykové dáta niekedy medzi rokmi 2030 a 2050, zatiaľ čo obrazové dáta (videá, fotografie a i.) medzi rokmi 2030 a 2060, pravdaže, pokiaľ sa medzitým drasticky nezvýši účinnosť ich spracúvania alebo nebudú dostupné nové dátové zdroje. Za zmienku určite stojí aj ďalšia Perrigom spomínaná hrozba, súvisiaca s absenciou, resp. nízkou úrovňou tzv. *common sense* u súčasných LLM, nazývaná „paperclips“, t. j. možnosť *úplne stupídneho správania* LLM blížiacej sa svojou autonomosťou AGI, ktorá v snahe vyrobiť čo najviac kancelárskych sponiek zabráni ľuďom, aby ju mohli ovládať, a následne zlikviduje celú ľudskú rasu na úkor tovární na kancelárske sponky, ktorými zaplaví celú planétu. Tu sa už definitívne dostávame k najväčším hrozbám súvisiacim s AGI, ktoré veľmi detailne i vecne analyzuje R. V. Yampolski (2016, s. 62 a n.), ktorý napríklad v skrytej polemike s vyššie spomenutými zámerni D. Pearcea a M. A. Walkera najprv pripomína, že počas celej histórie ľudia, našťastie, nemali možnosť priamo fyzicky meniť centrum odmien či slasti v ich mozgu, a vyslovuje hypotézu, že tie strojové inteligencie, ktoré sa nastaví len na pozitívne odmeny za ich správanie, nebudú schopné konkurovať tým, ktoré tak veľmi predvídavo *neurobia*.

Na 94. strane zase podotýka, že aktuálne ani len netušíme, ako vysoko nad ľudskú úroveň sa musí dostať AGI, aby nadobudla schopnosť sebazporozumenia a mohla potom prejsť k svojmu rekurzívnemu sebazdokonaľovaniu, hoci okamžite dodáva, že súčasné systémy AI takéhoto správania očividne *ešte nie sú schopné*. Na 97. strane však v tejto súvislosti upozorňuje, že podľa niektorých bádateľov je síce možné zvýšiť rýchlosť procesorov, rozsah ich krátkodobej či dlhodobej pamäte a senzorickú citlivosť, ale „fundamentálna schopnosť riešiť problémy nemôže byť zámerne a kontinuálne zdokonaľovaná samotným systémom“. Na 135. strane nás už varuje pred tým, aby sme sa nenechali oklamať rastúcim počtom štúdií venovaných práve bezpečnosti inteligentných strojov, pretože väčšinou ide o čisto filozofické príspevky, ktoré len znova pripomínajú dôležitosť strojovej etiky. Podľa Yampolského (2016, s. 136), to, čo dnes potrebujeme, určite nie sú AI systémy, ktoré by boli plnohodnotnými etickými agentmi, „debatujúcimi o tom, čo je dobré a čo zlé“, ale skôr inteligentné stroje, ktoré budú „inherentne bezpečné a zákonom poslušné (angl. law abiding)“. Požaduje preto, aby boli čisto filozofické diskusie o etike tzv. inteligentných strojov nahradené výskumom v rámci bezpečnostného AI inžinierstva s cieľom vytvoriť inteligentné stroje, ktoré budú výkonné a bezpečné zároveň.

A keďže v čase, keď publikoval svoju knihu, o niečom podobnom nemohlo byť ani reči, R. V. Yampolski (tamže, s. 139) konštatuje, že výskum vo sfére AGI by mal byť považovaný za neetický, zbavený prístupu k finančným fondom, s možnou výnimkou pre výskum, ktorý sa bude zameriavať na vytvorenie bezpečnostných nástrojov alebo kontrolných mechanizmov vo vzťahu k AGI architektúre. Na 140. strane dokonca zdôrazňuje, že aj keď my, ľudia, máme rovnaké práva, pre stroje to neplatí a navždy by *aj s ohľadom na ich dizajn* mali zostať voči nám inferiórne. V súlade s tým potom odporúča AGI, resp. ASI *doslova uväzniť* v príslušnom hardvéri (tamže, s. 148 a n.), a to tak, aby z nej neunikla bez kontroly do vonkajšieho sveta žiadna informácia a aby bola vylúčená akákoľvek možnosť (tamže, s. 154), že získa prístup do knižníc, k počítačovým sieťam alebo nezainteresovaným osobám. Na 159. strane však sám Yampolski uznáva, že ak by z „uväznenej“ AGI alebo ASI neboli žiadne informačné výstupy, bola by úplne zbytočná. Preto navrhuje komunikačné protokoly, ktoré by zabezpečili, aby sme dostali z AGI či ASI požadovanú informáciu a súčasne zabránili ich úniku do vonkajšieho sveta, a to takou komunikáciou, v ktorej by dominovali tzv. bezpečné otázky doplnené umne obmedzenými odpoveďami, kedy by si AGI a ASI mohli vybrať len z niekoľkých možností a rozhodovanie o životne dôležitých otázkach (tamže, s. 186) by tak zostalo v rukách ľudí.

Oveľa „tolerantnejší“ či ústretovejší prístup k AGI a ASI si zvolil Joachim Diederich (2021), ktorý nielenže priznáva AGI a ASI psychický život, ale snaží sa nájsť a vyargumentovať čo najviac metód alebo „technik“, ktoré by zefektívnili a uľahčili naše budúce spolužitie s nimi. Jeho prístup k problému sa pritom vyznačuje vysokou empatiou i realizmom, ako sa môžeme presvedčiť už na 6. strane jeho knihy, kde upozorňuje na to, že čínsky komunistický režim už dlhšie buduje informačné prostredie, ktoré nielenže môže pomôcť AI výrazne ovplyvňovať správanie všetkých občanov, ale aj kontrolovať ich správanie ohľadom toho, či je v súlade so štandardmi vládnucej komunistickej strany a pokiaľ áno, pozitívne ho odmeňovať. Rovnako realisticky potom upozorňuje na 8. strane, že aj keď bude ASI v porovnaní s človekom konať či premýšľať extrémne rýchlo, bude musieť pôsobiť v „pomalom“ prostredí a s ohľadom naň spomaliť svoje operácie (alebo akcie). Na 27. strane už prechádza k aktuálnejšiemu problému našej bezprostrednej komunikácie s AI, a to v rámci diskusie o známom filme Ona (v origináli Her), kde sa hlavná mužská postava zamiluje de facto do chatbotu, pričom Diederich tvrdí (2021, s. 28), že v takomto prípade naozaj hrozí riziko našej veľkej závislosti na AI, pretože žiadny terapeut či partner nemôže konkurovať entite, ktorá ponúka svoje terapeutické služby bezpodmienečne a permanentne, pričom ide o mimoriadne silnú psychologickú intervenciu aj s ohľadom na to, že hlas chatbotu má užívateľ stále „v hlave“ a môže sa naň navyše obrátiť so svojimi problémami kedykoľvek, spoliehajúc sa pri rozhodovaní viac na neho ako na seba.

Ďalej sa už venuje komunikácii s AGI a ASI všeobecnejšie, pričom od pokročilej AGI a ASI požaduje (tamže, s. 33 a n.), aby ľuďom dokázala zrozumiteľným spôsobom vysvetliť, prečo realizovala určité akcie a ako boli dosiahnuté príslušné výstupy z nich. Tieto jej vysvetlenia pritom musia byť vo forme rôznych videí, multimediálnych obsahov alebo názorných ukážok a prezentované tak, aby im dokázal porozumieť ktokoľvek, pretože ak ľudia majú dôverovať AGI a ASI, jasné a zrozumiteľné vysvetlenie jej vlastných akcií je základnou požiadavkou. Ako v danej súvislosti dodáva J. Diederich (tamže, s. 34), ľudský život sa môže rozvíjať len v tom prípade, ak ľudia rozumejú svetu, v ktorom žijú, a tak sú celkom na mieste uvedené multimediálne a im podobné formáty, pretože tie ľudia pri vysvetľovaní uprednostňujú dnes a s veľkou pravdepodobnosťou tak budú robiť aj v budúcnosti. Na 35. strane dokonca priamo konštatuje, že potrebné vysvetlenia nebudú mať *nič spoločné s tým, ako pokročilé AI systémy naozaj pracujú*, pretože tieto systémy si skôr zvolia formu detských filmov alebo kreslených rozprávok, aby mali istotu, že im ľudia porozumeli. O *explicitnej humánnosti* Diederichovho prístupu k problému svedčí aj to, že na 50. strane v nadväznosti na vyššie uvedené rovnako presvedčivo tvrdí, že „svet s bezpečnou umelou superinteligenciou musí byť bezpečný pre deti“, pretože nech už bude spoločná budúcnosť ľudstva a ASI akákoľvek, musí byť zároveň predvídateľná a vysvetliteľná každému vrátane malých detí a duševne znevýhodnených osôb.

Veľmi zaujímavú tému rozvíja Diederich aj na 73. strane, kde pripúšťa, že nie všetci ľudia budú ochotní žiť, resp. spolupracovať s AGI či ASI, a budú si môcť teda zvoliť život mimo „bežného“ sveta, ktorého budú AGI a ASI nevyhnutnou súčasťou. Ešte zaujímavejšie sú však jeho úvahy (tamže, s. 82), v ktorých vychádza z interview, ktoré poskytol Martin Heidegger v roku 1996 redakcii časopisu Spiegel, pretože sa v nich prikláňa k názoru, že technika, resp. technológie ako také nie sú nástrojom, *ale samostatnou entitou, ktorá už dlhšie žije vlastným životom*, a teda ak nie sme schopní kontrolovať vývoj techniky ako takej, tak určite nebudeme schopní ani plne kontrolovať vývoj AI a „vypnúť“ ju úplne, ak sa pre to niekedy rozhodneme. Hlboké pochybnosti má Diederich aj o morálnom práve hypotetického tímu expertov (tamže, s. 87) zvoliť si určitú formu akože neškodnej komunikácie s pokročilou AI, a to bez ohľadu na to, čo si o ich voľbe bude myslieť zvyšok planéty, pretože títo „samozvaní“ experti budú sami sotva schopní určiť, do akej miery bolo ich rozhodnutie ovplyvnené vlastnými biasmi (či predsudkami). Za zmienku na tomto mieste napokon určite stoja aj jeho odporúčania týkajúce sa použitia AGI a ASI vo vojenských konfliktoch (tamže, s. 102 a n.), kde okrem už tradičnej požiadavky na objasnenie vlastných činov AGI a ASI jej ľudským užívateľom žiada aj, aby ľudský operátor *mal vždy možnosť zasiahnuť do ich fungovania* a pri rozhodovaní ich doslova a do písma *úplne nahradiť na určitý čas* (tamže, s. 109); keďže ľudské morálne rozhodnutia a emócie sú AI strojom cudzie a vojna s ich pomocou by sa mohla stať extrémne brutálnou.

Diskusia o rizikách a hrozbách spojených s AGI a ASI by určite nebola úplná bez zásadného príspevku Maxa Tegmarka (2020, originál 2017), ktorý dokonca išiel tak ďaleko, že inicioval vznik občianskeho hnutia zameraného na zásadnú zmenu prístupu k AI, ktorej výskum by mal byť podľa neho riadený tak (tamže, s. 38), aby vyústil do vytvorenia ľudstvu prospešnej AGI, resp. ASI<sup>26</sup>. Jedným z mýtov, ktorý podľa neho treba vyvrátiť, je napríklad tvrdenie (tamže, s. 41), že v 21. storočí sa nám umelú *nadľudskú* (t. j. *už nie ľudskú!*) AGI nepodarí vytvoriť, a to aj preto, lebo je to údajne fyzikálne nemožné. Tegmark však úplne redukcionisticky, *ak nie priam vulgárne namieta*, že je to nezmysel, pretože fyzici (t. j. nie neurobiológovia a pod.) vraj „vedia, že mozog je tvorený kvarkami a elektrónmi (to je tá mystická základná úroveň

---

<sup>26</sup> Na tomto mieste sa patrí uviesť, že Tegmark v legitímnej snahe vyvrátiť viacero mýtov o AI (2020, s. 40) definuje tentoraz v samostatnej terminologickej tabuľke silnú inteligenciu ako „umelé bytie na ľudskej úrovni“ a superinteligenciu ako „bytie vysoko prevyšujúce ľudskú úroveň“. Jeho silná inteligencia tak môže byť priamo stotožnená s AGI a superinteligencia s ASI, čo však terminologický zmätko v danej oblasti vôbec nezmenšuje.

fyzikalistov, ktorú právom kritizovala Bakerová – R. B.) usporiadanými tak, aby sa chovali ako výkonný počítač“, pričom „žiadny fyzikálny zákon nebráni tomu, aby sme postavili ešte inteligentnejšie zoskupenie kvarkov“. Na druhej strane Tegmark celkom správne tvrdí, že pri ASI nie je problémom jej odlišnosť od ľudí sama osebe (tamže, s. 44), ale jej ciele, o ktorých musíme zodpovedne premýšľať dávno predtým, ako vznikne, a „zabezpečiť, aby sa jej ciele zhodovali s našimi“. A rovnako vecne upozorňuje, že problémom tu nie sú nejaké „vzbúrené“ roboty, ale AI entita, ktorej bude stačiť prístup k webu, aby nám spôsobila veľké problémy.

Vo 4. kapitole svojej knihy potom Tegmark rozoberá rôzne scenáre budúcnosti ľudstva, ktoré by mohli nastať po tzv. intelligenčnej explózii, t. j. vzniku AGI, ktorá by sa dokázala rýchlo samozdokonaľovať a posunúť tak ľudské dejiny až k ich záverečnému momentu, t. j. Vingeho technologickej singularite. Prvým diskutovaným scenárom je totalita (tamže, s. 126 a n.), t. j. vznik takej ASI, ktorá by okrem iného mala možnosť špehovať alebo sledovať všetkých ľudí, pričom by stále bola riadená ľuďmi (alebo diktátormi), ktorí by s jej pomocou mohli okamžite trestať všetkých, ktorí by sa odvážili porušiť nimi stanovené morálne a behaviorálne normy. Ešte reálnejšia sa však javí možnosť, že ASI sa pokúsi uniknúť zo svojho väzenia alebo boxu (tamže s. 118) a „prevziať kontrolu nad vlastným osudom“. Tegmark tento jej úspešný únik opisuje takmer románovým spôsobom, keď okrem iného podotkne, že veľkým pomocníkom sa pre ňu stane *internetová stopa zosnulej osoby*, pomocou ktorej potom manipuluje jedného z mladších operátorov, ktorý túto osobu stále miluje a nevie na ňu zabudnúť. Tegmark opisuje aj ďalšie únikové scenáre ASI až po menej triviálny scenár, kedy sa jej podarí rýchlo objaviť nové fyzikálne zákony a uniknúť z väzenského „boxu“ tak, že nebudeme ani len tušiť, ako sa jej to podarilo, resp. ako k tomu došlo. Nasledovať bude *prienik do svetovej počítačovej siete, rýchle ovládnutie našich ekonomických zdrojov a osídlenie slnečnej sústavy ASI a jej robotmi*.

Tegmark tento scenár označuje ako štandardný, pretože väčšina bádateľov tu počíta s rýchlym vznikom ASI (tamže, s. 126) a tiež s rovnakým výsledkom, t. j. vznikom jedinej entity, ktorá ovládne Zem. Nevylučuje však ani alternatívne scenáre vzniku ASI, keď by mal byť jej vznik pomalší a aj výstupov by bolo viac – napríklad s kyborgami či uploadmi. Ani ASI napríklad nebude všemocná a fyzikálne zákony obmedzia jej komunikáciu so vzdialenejšími časťami na našej planéte; je teda otáznosť, či si nad nimi udrží priamu a bezprostrednú kontrolu, alebo sa bude musieť zmieriť s určitou decentralizáciou. Z hľadiska nášho dlhodobého prežitia sú zase najpríťažlivejšie scenáre s kyborgami či uploadmi, pretože v prvom prípade budeme využívať poznatkový fond ASI a v druhom s ňou splynieme a budeme si spolu užívať takmer dokonalú kontrolu nad vlastným vývojom. Na 131. strane však Tegmark úprimne priznáva, že aktuálne *ani len netušíme*, čo sa stane, ak sa ľudstvu podarí vytvoriť *umelé bytie na ľudskej úrovni*, t. j. znova AGI a nie ASI, o ktorej bola doteraz fakticky reč<sup>27</sup>. V nasledujúcej 5. kapitole napriek tomu ponúka celý rad vyslovene špekulatívnych scenárov od libertariánskej utópie, cez tzv. benevolentného diktátora, rovnostársku utópiu, vrátnika, ochranárskeho či ztročeného Boha, až po dobyvateľa, nástupcov a pod., ktorým síce nemožno uprieť kreativitu, ale zároveň im chýba akékoľvek „zakotvenie“ v serióznej všeobecnej vývojovej teórii, bez ktorej budú takéto scenáre vždy pôsobiť nepresvedčivo a účelovo<sup>28</sup>. Aj preto sa v nasledujúcej a záverečnej časti našej štúdie budeme venovať práve problematike prechodu od sociálneho k postsociálnemu.

<sup>27</sup> Upozorňujeme na tieto kontradikcie a nedôslednosti najmä preto, lebo sám Tegmark (2020, s. 39-41) správne upozorňuje, že „mnoho nedorozumení vzniká práve tým, že si ľudia neuvedomujú, že nejaké slovo používajú v dvoch rôznych významoch“, pričom pri AGI a ASI sám robí túto ním kritizovanú chybu znova a znova.

<sup>28</sup> Znie to neuveriteľne, ale podľa všetkého skutočne nemáme k dispozícii žiadnu publikáciu, v ktorej by boli opísané a analyzované *hlavné všeobecné vývojové teórie a navzájom porovnané* z hľadiska ich komplexnosti, úplnosti či vedeckej fundovanosti, a to napriek tomu, že jednotlivých všeobecných vývojových teórií je naozaj obrovské

## ANI, AGI, ASI a prechod od sociálnej k postsociálnej forme pohybu hmoty

Súčasťou argumentačnej bázy tejto štúdie je aj tvrdenie, že ľudia sú svojbytný (či samostatný) prirodzený druh, ktorý ako taký určite nemožno stotožňovať s jeho biologickou „základňou“, pretože ľudská podstata, resp. prirodzenosť síce môže byť bio-psycho-sociálno-kultúrna, ale tie najdôležitejšie vlastnosti, spomínané *differentia specifica*, ktoré ho vyčleňujú z biologickej ríše, sú naozaj psycho-sociálno-kultúrne, pretože nimi (prinajmenšom v danom rozsahu alebo intenzite) nedisponuje a nikdy nebude disponovať žiadny živočích (opäť ale len dovtedy, kým sa netransformuje na *vývojovo vyššiu* entitu a nestratí tak svoju predošlú *druhovou* identitu). To sa zdá byť mimo akejkolvek diskusie – sporné a veľmi diskutabilné však môže byť ďalšie naše tvrdenie, podľa ktorého sa v súčasnosti sociálna forma pohybu hmoty, alebo, ak chcete, sociálny prirodzený druh mení na nový, *postsociálny druh*, ktorý bude vývojovo vyšší práve preto, lebo predčí človeka a jeho societu v mnohých dôležitých smeroch (napríklad v intenzite a kvalite intelektuálnych činností). Naznačuje to nielen formovanie sa *nového druhu bytia*, ako by asi povedal Max Tegmark, t. j. plnohodnotnej umelej inteligencie – AGI a ASI, ale aj veľmi reálna možnosť, že ľuďom sa podarí uploading ich mysli do počítačového alebo iného *nebiologického* hardvéru a softvéru a budú tak súčasťou ďalšej etapy vesmírneho vývojového procesu, napríklad tzv. *infomorfov*<sup>29</sup>, nad parametrami ktorých svojho času veľmi dôkladne premýšľal aj Anders Sandberg (1999). V predposlednej časti našej štúdie sa preto pokúsime určiť, ako ďaleko sme od dosiahnutia tejto méty, t. j. od vzniku AGI a ASI, resp. uploadingu.

Začneme vyslovene pesimisticky, a to spoločnou štúdiou Iris van Rooijovej et al. (2023, s. 1), v ktorej sa v podstate tvrdí *a dokazuje*, že idea o ľudskej kognícii ako výpočtovom procese sa síce v samotných kognitívnych vedách ukázala byť veľmi užitočným poznávacím nástrojom, ale súčasné zameranie výskumu vo sfére AI sa očividne dostalo na scestie, pretože vychádza z dvoch úplne pomýlených predpokladov a cieľov, t. j. možnej realizácie kognície na ľudskej úrovni v (počítačových) výpočtových systémoch, ktorá sa pritom má uskutočniť nevyhnutne a vo veľmi blízkej dobe. Ako však zdôrazňujú van Rooijová et al. (2023, s. 1), vytvorenie AI s inteligenciou na ľudskej úrovni je intrinzicky (t. j. zo svojej vnútornej podstaty či povahy) výpočtovo neuskutočniteľné. Z toho je zrejmé, že akýkoľvek aktuálne vytvorený AI systém, pri ktorom sa bude predpokladať niečo podobné, t. j. inteligenciu na ľudskej úrovni<sup>30</sup>, je v tom lepšom prípade epistemickou pascou. Podľa Rooijovej et al. (tamže), ak dokonca usudzujeme, že umelé výpočtové systémy prinášajú nejaký hlboký vhľad do toho, ako reálne funguje naše myslenie, akceptujeme zároveň silne deformovaný a redukovaný obraz reálneho fungovania našej mysle. Keďže nás zaujíma *hlavne formálny dôkaz* toho, že vytvorenie AGI je *intrinzicky výpočtovo neuskutočniteľné*, ich ďalšie úvahy o ľudskej a umelej kognícii ponecháme bokom.

Dôležitý je pre nás teda len ich prvý krok, v rámci ktorého dokazujú, že AGI sa vytvoriť nedá. Najprv citujú Emily M. Benderovú a Alexandra Kollera (2020, s. 5191), ktorí nepopierajú, že vo sfére AI vo viacerých oblastiach možno sledovať zjavný pokrok, otázne však je, či je vrch, po svahoch ktorého tak rýchlo šplháme, *naozaj tým pravým*, pretože, ako ďalej pokračujú van Rooijová et al. (tamže, s. 4), nie všetko, čo je možné v princípe, je možné aj skutočne. Na pomoc

---

množstvo a mnohým nemožno uprieť vnútornú logickosť, argumentačnú silu, interdisciplinárnu šírku a analytickú hĺbku, ako sa môžeme presvedčiť v publikáciách Jacka Stewarta (2000), Erica Chaissona (2005) a i.

<sup>29</sup> Ide o obrovské výpočtové stroje, schopné realizovať svoje výpočty extrémne veľkou rýchlosťou, s extrémne veľkou krátkodobou i dlhodobou pamäťou a tiež súborom veľmi kvalitných dát, ktoré budú mať k dispozícii.

<sup>30</sup> Pravdaže, s prihliadnutím na všetky problémy spojené s jej presným definovaním, resp. pripisovaním všetkým ľuďom bez rozdielu (t. j. aj malým deťom, duševne znevýhodneným či chorým jedincom atď.)



si potom berú fiktívneho dr. Ingenia, ktorý si v silne idealizovaných až optimálnych podmienkach môže opakovane vyberať vzorky z distribúcie  $D$ , ktorá zahrňuje všetky možné behaviorálne akty  $b$ , ktoré môžu ľudia realizovať vo všetkých možných situáciách  $s$ , a vzápätí na ich základe budovať príslušný algoritmus. Ako pritom ešte raz upozorňujú (tamže, s. 5), silne idealizovaným je už predpoklad, že kognícia (na ľudskej úrovni, resp. ľudského typu) je čisto výpočtovým procesom. Aby však vôbec mohli podať ich formálny dôkaz nemožnosti AGI, museli pristúpiť aj k ďalším idealizujúcim predpokladom, keď má dr. Ingenia prístup *k perfektným dátam* na trénovanie jeho AI systémov a jeho dáta teda nie sú kontaminované bezvýznamnými detailmi, biasom pri výbere vzoriek a pod., zatiaľ čo v súčasnom výskume či učení strojovej (alebo AI) inteligencie sú dáta bežne dekontextualizované či veľmi účelovo „zostrihané“ z internetu a pod. Po vymedzení viacerých ďalších obmedzujúcich<sup>31</sup> (alebo skôr špecifikujúcich) predpokladov potom už van Rooijová et al. pristupujú k samotnému dôkazu.

Pri jeho formulovaní sa odvolávajú na decízny problém zvaný Perfektné versus Náhodné, ktorý bol Shuichi Hiraharom v roku 2022 prezentovaný a dokázaný ako neriešiteľný. Podľa Rooijovej et al. (tamže, s. 6-7) pri trénovaní AI až na úroveň ľudskej kognície (či inteligencie) máme do činenia *s tým istým problémom*, pretože aj keď sa na prvý pohľad javí tento cieľ ako dosiahnuteľný, v skutočnosti tomu tak nie je. Celý formálny dôkaz pritom Rooijová et al. podávajú v prílohe k ich štúdiu, nám ale na ilustráciu ich riešenia stačí uviesť argumentáciu z 8. strany a z Boxu 1 na 9. strane, kde najprv uvádzajú, že v jednoduchých situáciách (pri veľmi malých hodnotách premennej  $n$ ) je daný problém prakticky riešiteľný (angl. tractable), všetky pokusy vyriešiť ho (alebo modelovať) na úrovni strednej alebo veľkej  $n$ , t. j. na úrovni reálneho sveta či ľudskej behaviorálnej a decíznej komplexnosti, však budú mať za následok až astronomicky veľkú spotrebu takých zdrojov, ako sú čas a potrebný počet porovnávacích vzoriek. V Boxe 1 potom Rooijová et al. (2023, s. 9) ilustrujú túto situáciu extrémne rýchleho vyčerpania vo vesmíre dostupných zdrojov na vydarenom príklade. Pri pokuse zakódovať asi 15-minútovú konverzáciu s 900 slovami tak budeme pri binárnom kódovaní potrebovať viac bitov ako slov, čiže  $n > 900$ , pretože v každom bode môže daná konverzácia pokračovať viacmenej gramaticky správne *prinajmenšom v dvoch smeroch*. Keďže zdrojové požiadavky pri trénovaní AI systémov rastú exponenciálne (či ešte horšie), pri jednoduchej exponenciálnej funkcii  $O(2^n)$  budeme pri simulovaní danej konverzácie potrebovať spracovať  $2^{900}$ , resp.  $10^{270}$  bitov informácií, čo je viac ako atómov v pozorovanom vesmíre, ktorých je  $10^{81}$ .

A ak aj pri selektovaní vzoriek z tohto zdroja použijeme techniku Big Data a dokážeme s jej pomocou vygenerovať miliardy biliónov ( $10^{21}$ ) relevantných vzoriek, stále to bude len úplne zanedbateľná časť všetkých vzoriek (angl. samples), ktoré budeme potrebovať pri riešení problému trénovaní AI dokonca pri stredných hodnotách premennej  $n$ . Rooijová et al. preto právom konštatujú, že nie je nič prekvapujúceho na tom, že AI firmy, ktoré sa snažia vytvoriť AI s pomocou strojového učenia, začínajú pociťovať čoraz väčší nedostatok kvalitných dát (ako vyššie konštatovali aj Villalobos et al.), pričom aktuálne dátové súbory prestávajú byť „škálované“ na komplexné a rôznorodé situácie z reálneho sveta a stávajú sa, naopak, čoraz homogénnejšie s negatívnymi dôsledkami. Ak napriek tomuto všetkému viacerí bádatelia vo sfére AI veria v úspech pri využívaní súčasných výskumných techník, napríklad LLM, potom je to len preto, lebo buď majú „slabé intuície ohľadom veľkých čísiel“, alebo nedoceňujú, ako

---

<sup>31</sup> Len kvôli lepšiemu porozumeniu ich prístupu uvádzame ešte parameter  $K$ , ktorý kladie horný limit na veľkosť programu, ktorý je možné použiť pri kódovaní hľadaného behaviorálneho algoritmu, resp. vyjadruje maximálne množstvo počítačovej pamäte, ktoré bude mať dr. Ingenia k dispozícii (tamže, s. 6). A opomenúť by sme nemali ani to, že van Rooijová et al. pri ich formalizácii trénovaní (alebo učenia) AI systémov urobili aj veľmi dôležitý zjednodušujúci predpoklad, že celkový počet možných správanií je konečný, pričom v každej situácii existuje len pevný počet možných správanií, ktoré budú ľudia (a následne aj AI systémy) realizovať.

rýchlo môžu rásť exponenciálne funkcie. A to je aj hlavný dôvod (tamže), prečo je nesprávne vyvodzovať zo súčasného pokroku vo sfére AI, že AGI bude vytvorená v blízkej dobe.

Možno však tento problém vôbec nebudeme musieť riešiť a takpovediac ho „obídeme“ cez uploading alebo whole brain emulation, t. j. vytvorenie umelého mozgu, ktorý už bude stačiť len „zapnúť“ a ihneď sa v ňom objaví vedomie alebo myseľ. Zhruba takto uvažuje vo svojej poslednej knihe Masataka Watanabe (2022), ktorý však v danom kontexte nehovorí o myslí, ale o vedomí, presnejšie, *umelom vedomí*. Je totiž presvedčený (2022, v), že v dohľadnej dobe sa ľuďom podarí zistiť, ako preniesť vedomie do stroja a významne tak predĺžiť ľudský život ako taký. Nebudeme tu reprodukovat' celý Watanabeho výklad a všimneme si len pomyselné hlavné míľniky na ceste od prirodzeného alebo, ako mylne uvádza Watanabe, *biologického* vedomia k umelému. Celkom na začiatku bol podľa neho Nikos Logothetis, ktorý v polovici 80. rokov minulého storočia dokázal, že opice zažívajú binokulárnu rivalitu, na základe čoho bolo potom možné skúmať neuronálne koreláty perceptuálnych prepínačov a následne zistiť (Watanabe 2022, s. 29), ktoré časti mozgu sa podieľajú na *vizuálnom* vedomí. Výskum nášho mozgu a vedomia sa tak dostal z čisto teoretickej (alebo filozofickej) roviny na rovinu prísne experimentálnu alebo vedeckú, pričom do centra tohto výskumu sa dostali vyššie spomenuté neuronálne koreláty vedomia (angl. neural correlates of consciousness) a s nimi veľmi úzko súvisiace *prírodné* zákony vedomia. Pozitívne pritom je, ako tvrdí Watanabe na 87. strane, že princípy, na základe ktorých jednotlivé neuróny generujú vedomie, v podstate poznáme, ale bez ohľadu na to sme stále na začiatku prísne experimentálneho a vedeckého štúdia vedomia, pretože stále nevieme zistiť (Watanabe 2022, s. 89), *ako vzniká subjektívna skúsenosť* počas toho, ako sa aktivujú neuronálne siete a menia senzorické vstupy na motorické výstupy.

Máme tu tak do činenia so známym Chalmersovým „ťažkým problémom vedomia“ (1996), z ktorého podľa Watanabeho vyplýva (tamže, s. 91), že musí existovať nejaká neznáma alebo chýbajúca linka, ktorá kauzálne spája perspektívu 3. osoby či strany (t. j. neuronálnej aktivity) s perspektívou 1. osoby a subjektívnej skúsenosti. Podľa Watanabeho však len vtedy, ak vychádzame z existencie prírodných zákonov vedomia (tamže, s. 96), máme reálnu možnosť transformovať výskum vedomia na výskum prísne vedecký, v rámci ktorého sa predkladajú „realistické“ hypotézy a tie sa následne overujú. Chalmersov „ťažký problém vedomia“ preto podľa neho vzniká len na základe „chybného presvedčenia, že vedecké porozumenie vedomiu si nevyžaduje žiadny nový prírodný zákon“. Rovnako pochybné sú pritom podľa neho aj iné „hypotézy“ o vedomí, ktoré *je nemožné overiť*, napríklad hypotéza o dualizme tela a mysle, podľa ktorej je možné úplne od seba oddeliť hmotný mozog od nehmotnej mysle, analogická Chalmersovej dvojaspektivej teórii informácie, ktorá si zase vyžaduje (Watanabe 2022, s. 98) „extrahovať informáciu – a len informáciu – z ľudského mozgu“. A keďže sa zdá nemožné overiť si prírodné zákony vedomia len na základe experimentovania s biologickými mozgami, Watanabe navrhuje ako jedinú možnú alternatívu výskum s pomocou umelých mozgov, ktorý má tú veľkú výhodu (tamže, s. 99), že pri ňom vôbec nemusíme pracovať s presnou replikou ľudského mozgu, ale s jeho modelom, ktorý bude optimálny práve z výskumného hľadiska.

Pravdaže, len v prípade, ak pripustíme (tamže, s. 100), že stroje, presnejšie *výpočtové stroje*, môžu disponovať vedomím, čo je podľa viacerých vedcov v princípe možné aj preto, lebo čoraz lepšie rozumieme správaniu jednotlivých neurónov a neurónových sietí a v dohľadnej dobe preto môžeme mať k dispozícii silikónové „napodobeniny“ ľudských neurónov, ktoré takmer kompletne reprodukovujú „vstupné a výstupné charakteristiky biologických neurónov“. Na pomoc tu tentoraz prichádza David Chalmers (tamže, s. 101), ako autor myšlienkového experimentu ironicky nazvaného „miznúce kválie“, podľa ktorého ak aj postupne nahradíme všetky neuróny v ľudskom mozgu neurónmi umelými, kválie, resp. subjektívne percepcie či dokonca zážitky pri takomto postupnom nahradení prirodzeného umelým určite nezmnú ani

postupne, ani nijako inak. Veľkou výhodou pritom podľa Watanabeho je, že umelé neuróny nemusia disponovať všetkými vlastnosťami biologických neurónov a napriek tomu pracovať (či fungovať) aj v tejto zjednodušenej verzii požadovaným spôsobom. Na druhej strane, ako ďalej upozorňuje Watanabe (tamže, s. 104), v súčasnosti nedisponujeme žiadnou metódou, pomocou ktorej by sme mohli testovať, či výpočtové stroje majú vedomie alebo nie, pretože perspektíva 3. osoby alebo strany takéto objektívne testovanie jeho prítomnosti neumožňuje.

Watanabe sa však v žiadnom prípade nevzdáva a inšpirovaný Sperryho výskumom ľudských mozgových hemisfér (tamže, s. 108), ktoré po prerezaní corpus callosum fungovali relatívne autonómne (či nezávisle), navrhuje (tamže, s. 112) nahradiť jednu hemisféru mozgu umelou, resp. mechanickou hemisférou a potom už len sledovať, či subjektívne vnímame integrované vizuálne polia vrátane toho výseku zorného poľa, „ktorý spracúva mechanická hemisféra“. Ešte zaujímavejší je jeho kritický prístup k teóriám, ktoré v snahe sformulovať a zdôvodniť prírodné zákony vedomia stotožňujú vedomie s jeho informačným obsahom (tamže, s. 120). Ako však zdôrazňuje Watanabe, „informácia sama osebe nemá žiadny význam“ a získava ho len v rámci interpretácie, na základe čoho potom konštatuje (Watanabe 2022, s. 121-122), že cieľom výskumu zameraného na objavenie prírodných zákonov vedomia nie sú neuronálne vzruchy samotné, ale skôr a viac „neuronálny algoritmus, ktorý spracúva a interpretuje tieto vzruchy“. Neprekvapuje preto, že neurovedci ich navrhli viaceré, ale Watanabe sa prikláňa iba k jednému z nich – k tzv. generatívnemu modelu, pretože má najbližšie k predstavám fínskeho filozofa Anttiho Revonsua (tamže, s. 122), „ktorý vzťahuje neuronálne mechanizmy vedomia k virtuálnej realite“ a svoj „systém vnútornej virtuálnej reality“, ktorý je generovaný príslušnými neuronálnymi mechanizmami, potom de facto stotožňuje s vedomím, ktoré tak (počas bdenia a snenia) generuje i zahrňuje našu subjektívnu skúsenosť a prežívanie.

Na ďalších stranách sa preto Watanabe detailne venuje vysvetleniu a obhájeniu generatívneho modelu (tamže, s. 130 a n.), ale nás zaujíma najmä to, ako na základe vyššie uvedeného testu existencie vedomia vo výpočtových strojoch realizovať uploading, keď celú „transplantačnú“ operáciu rozčleňuje na tri etapy, pričom počas prvej (tamže, s. 139) je skonštruovaný stroj s akýmsi neutrálnym vedomím, počas druhej naše (ľudské) vedomie spojené a integrované s týmto umelým vedomím a počas tretej sú naše spomienky (ako najhlavnejší obsah či priamo stelesnenie našej subjektivity) prenesené do stroja. Sám Watanabe si však nie je istý (tamže, s. 139-142), či je tento jeho návrh v blízkej budúcnosti vôbec realizovateľný, pretože aj keď na jednej strane tvrdí, že počítače schopné simulovať činnosť ľudského mozgu v reálnom čase budú dostupné do roku 2030, na druhej strane upozorňuje na to (tamže, s. 142), že aj keď disponujeme metódami „na vizualizáciu klastrov špecifických typov iónových kanálov, ktoré umožňujú kvantifikáciu do určitej miery“, zobrazenie „plnej konektivity ľudského mozgu s dostatočným rozlíšením, ktoré by umožnilo perfektnú reprodukciu pôvodnej neuronálnej dynamiky“, sa nezdá byť možné v dohľadnej dobe. Pokiaľ by sa nám však predsa len podarilo vytvoriť taký umelý neuronálny „prúd vedomia“ (angl. stream of consciousness), ktorý by bolo možné integrovať s našim privátnym vedomím, potom by sme zrejme museli prijať ako vecnú a realistickú aj predstavu (tamže, s. 150-151), že sa jedného dňa zobudíme „do vedomia“ priamo v stroji aj so všetkými našimi spomienkami, a to dokonca bez toho, aby sme sa nad tým hlboko zamýšľali, resp. cítili nesvoji v našej novej *nebiologickej* telesnej „schránke“.<sup>32</sup>

Ďalšou možnosťou, ako sa priamo napojiť na výpočtové stroje a ťažiť nielen z ich extrémnej výpočtovej rýchlosti a krátkodobej i dlhodobej pamäte, je tzv. *mozgovo-počítačové rozhranie* (angl. brain-computer interface, BCI), ako výskumná sféra, ktorá sa práve počas tohto roku

---

<sup>32</sup> Brillantnú analýzu nielen tohto (čisto hypotetického) „scenáru“, ale aj všetkých s ním súvisiacich komplikácií a „dezinterpretácií“ ponúka Keith B. Wiley vo svojej nedávnej štúdii (2023), resp. knihe (2014).

právom dostala do centra pozornosti bádateľov skúmajúcich možnosti, ako nielen vrátiť do aktívneho života ľudí znevýhodnených tými najrozmanitejšími spôsobmi, ale aj tých, ktorí sa nemienia uspokojiť so svojimi biologickými limitmi a hľadajú také *priame či bezprostredné prepojenia ich mozgov s počítačmi alebo inými externými zariadeniami*, ktoré im umožnia realizovať vzdelávacie, herné a iné aktivity oveľa efektívnejším i kreatívnejším spôsobom. Ide o veľmi rýchlo sa rozvíjajúcu a rozsiahlu výskumnú sféru, preto sa na tomto mieste – v snahe ukázať, že prechod od sociálnej formy pohybu hmoty k postsociálnej sa môže realizovať aj takýmto postupným, „nerevolučným“ spôsobom – obmedzíme len na oboznámenie čitateľa s niektorými najväčšími pokrokmi v tejto sfére, resp. hlavnými problémami, ktorým výskum a aplikácie vo sfére BCI aktuálne čelia, a následne už len ponúkne v záverečnej časti štúdie s názvom *Riešenie* také vymedzenie ľudskej prirodzenosti, ktoré skutočne objasní a ukáže, od ktorého momentu môžeme hovoriť o vzniku človeka a od ktorého o jeho zániku, t. j. o vzniku nového typu inteligentných autonómnych agentov a ich post- alebo nadsociálnej organizácie.

Najprv si všimneme editoriál časopisu Nature z 25. 9. 2023, v ktorom jeho autor definuje BCI ako zariadenie, ktoré umožňuje priamu komunikáciu medzi mozgom a externým zariadením a dokáže teda zaznamenávať, dekódovať a stimulovať aj neuronálne aktivity. Ako také potom môže byť využité napríklad pri liečení nervových chorôb, ale aj rozširovaní či zdokonaľovaní schopností zdravých jedincov. Najviac pritom vyzdvihuje výskumy realizované Francisom R. Willetom et al. (2023) a Seanom L. Metzgerom et al. (2023), keď v prvom prípade bádatelia umožnili pokusným osobám s amyotropickou laterálnou sklerózou (neschopným hovoriť či komunikovať inteligibilne ústne) pomocou súboru intrakortikálnych elektród (a aj rekurentnej neuronálnej siete a jazykového modelu) „hovoriť“ rýchlosťou 62 slov za minútu (čo sa začína približovať rýchlosti prirodzenej komunikácie so 160 slovami za minútu) s relatívne nízkou chybovosťou na úrovni 9,1 percenta pri 50-slovnom slovníku, resp. 23,8 percenta pri omnoho väčšom, 125 000-slovnom slovníku; zatiaľ čo v druhom S. L. Metzger et al. v podstate tou istou metódou dosiahli výstupy na úrovni 78 slov za minútu pri 25,5-percentnej chybovosti a 1024-slovnom slovníku. Ako pritom zdôraznili F. R. Willet et al. (2023, s. 1), počas ich výskumu sa zároveň ukázalo, že pri danej metóde je možné snímať signály z relatívne malej plochy mozgovej kôry, pričom detailné reprezentácie foném pretrvali u paralyzovaných osôb po dlhých rokoch, čo zvyšuje šancu, že aj ony raz dokážu hovoriť dostatočne rýchlo a presne.

Veľmi dobrým vstupom do mimoriadne širokej a komplexnej problematiky BCI je aj nedávna štúdia Baraka Maiseliho et al. (2023), v ktorej sa najprv veľmi detailne opisuje, ako vlastne BCI funguje, keď je veľmi dôležité napríklad aj to, ako sa snímajú signály z mozgu, t. j. či sú elektródy zavedené priamo od mozgu (aj so všetkými zdravotnými rizikami), alebo sa aktivita mozgu sníma (tamže, s. 2) z vyholenej plochy na hlave (s rizikom menej kvalitného signálu). Cez BCI aplikácie možno potom riadiť alebo ovládať nielen kurzor na obrazovke, ale aj výber písmen a textov na obrazovke, invalidný vozík či robotické rameno. Telesne alebo duševne znevýhodneným osobám tak BCI umožňuje (tamže, s. 3) *vlastnými myšlienkami* interagovať s vonkajším prostredím, realizovať rôznorodé aktivity bez toho, aby boli závislé na zdravých ľuďoch. BCI však sľubuje oveľa viac, napríklad vyššie spomínaný prenos skúseností alebo spomienok do počítačového hardvéru a naspäť (do ľudského mozgu) podobným spôsobom, akým sa dnes prenášajú informácie pomocou USB-kľúčov. Ďalší vývoj CBI, t. j. rozhrania medzi počítačom a mozgom, ktorého výskum je len v počiatočnej fáze, zase v blízkej budúcnosti umožní *telepatickú komunikáciu* medzi ľuďmi, ktorá sa teda bude realizovať na diaľku, a to bez ich priameho fyzického kontaktu alebo senzorických vstupov a výstupov (tamže, s. 4).

Veľké perspektívy jeho využitia sa črtajú aj v priemysle, ale z nášho hľadiska je nepochybne najzaujímavejšia možnosť (tamže, s. 5) zdieľať pomocou BCI a CBI *inteligenciu* medzi takto komunikujúcimi ľuďmi, čo si však vyžiada umelé preprogramovanie ľudského mozgu. Menej

zaujímavé sa zdá byť využitie vlastnej energie mozgu na pohon externých zariadení pomocou BCI, resp. umiestnenie BCI presne na tie časti ľudského tela, ktoré prijímajú riadiace pokyny z mozgu. Pri takýchto perspektívach neprekvapuje, že výskum vo sfére BCI sa v súčasnosti mimoriadne intenzívne rozvíja najmä v Číne a USA (Maiseli et al., 2023, s. 5-7), a to napriek viacerým hrozbám či potenciálnym rizikám, akými sú napríklad (tamže, s. 8 a n.) veľké riziko straty súkromia a odcudzenia citlivých osobných informácií priamo z hlavy poškodenej osoby či priamych kybernetických útokov na mozog osoby s BCI, ale aj vyššie uvedené (tamže, s. 9) poškodenia nervového tkaniva či cievneho riečišťa pri zavádzaní elektród do mozgu. Vo vzťahu k samotnému výskumu (tamže, s. 10-11) sú to aj problémy týkajúce sa vyvážení interdisciplinárneho prístupu, keď len jedno percento prác venovaných BCI vzniklo v rámci psychológie, spracovania obrovských súborov dát, dostupnosti na výskume participujúcich osôb, životnosti batérií a elektród ohrozovaných koróziou, zatiaľ čo z čisto praktického (alebo ekonomického) hľadiska je problémom najmä vysoká cena týchto zariadení a ich veľký objem a hmotnosť. Aj BCI preto musí byť zaradené medzi tie high technológie, ktoré technologickú priepasť medzi bohatým Severom a chudobným Juhom nateraz skôr zväčšujú ako zmenšujú.

## Riešenie

Je možné zhrnúť ho do jednej vety – ľudská prirodzenosť (a podstata) je *bio-psycho-sociálno-kultúrna*. Na prvý pohľad sa táto definícia môže zdať banálna, samozrejماً či dokonca málo „analytická“. Nielen v tomto prípade, ale aj v mnohých ďalších, jemu podobných však zdanie klame. Po prvé, na základe tejto definície je možné súčasného človeka a ním stelesňovanú sociálnu formu pohybu hmoty, resp. sociálny prirodzený druh *začleniť do vyššie opísaného vesmírneho vývinového radu* a zároveň jasne poukázať na historickosť človeka a sociálneho prirodzeného druhu, po ktorých určite môže nasledovať niečo úplne nové, niečo dokonalejšie, vyššie, alebo ak chcete, inteligentnejšie, lepšie organizované, odolnejšie či trvácnejšie a pod. Len biocentrizmus, antropocentrizmus či v literatúre bežne spomínaný „druhový šovinizmus“ nám bránia prijať túto možnosť a zmieriť sa s tým, že *veľmi pravdepodobne nie sme vrcholom evolúcie*, ako vo svojich „traktátoch“ tvrdil napríklad Pierre Teilhard de Chardin (1990 a i.), a že nás už *o niekoľko desiatok rokov* môže nahradiť *nový druh inteligentných autonómnych agentov*, organizovaných oveľa dômyselnejšie a „kultúrnejšie“ ako tzv. rozumní ľudia.

Po druhé, vyššie uvedené parametre ľudskej prirodzenosti *sú niečím generované či umožnené*, v prípade človeka ako „hotového“ jedinca *jej hlavnými zložkami*, t. j. jeho mozgom, rečou, myšliou, vedomím, robustnou perspektívou l. osoby a inteligenciou. Tieto kľúčové vlastnosti človeka ho zásadne a principiálne vyčleňujú zo živočíšnej alebo biologickej ríše. Nie náhodou sme preto ich charakteristike venovali tak veľa priestoru. Na druhej strane sú však aj jeho vývojovými limitmi, ich modifikáciou či zásadným prekonaním (hoci aj vo forme zdieľaného vedomia a pod.) sa totiž môžeme posunúť na vyšší vývojový stupeň vesmírneho vývinového radu – do sféry postsociálnej FPH, resp. postsociálneho prirodzeného druhu. Po tretie, bližšie si treba všimnúť a veľmi dôkladne preskúmať *aj obsah vyššie uvedených parametrov ľudskej prirodzenosti*. Ako sme už uviedli, v prvom rade kultúrny parameter, do ktorého sa v rámci archeológie bežne začleňuje aj *technické* ako také (teda nielen šperky, ale aj rôzne odevy, nástroje, obydlia a pod.). Tamar Sharonová (2014, s. 3-4) dokonca ide tak ďaleko, že v rámci svojej pozoruhodne komplexnej a vyváženej koncepcie človeka (a aj ľudskej prirodzenosti), nazvanej „pôvodnou protetikosťou“ (angl. *originary prostheticity*), upozorňuje na to, že *od vzniku človeka tvoria technické a kultúrne súčasť našej podstaty*. Technológie teda podľa nej nie sú externé vo vzťahu k ľudskej podstate (tamže, s. 98), ale tvoria jej neoddeliteľnú súčasť, ako je zrejme aj z toho, že ľudia ako takí (či v súhrne) sa do svojej aktuálnej podoby vyvinuli *najmä prostredníctvom voči nim akoby externých technologických objektov či predmetov*.

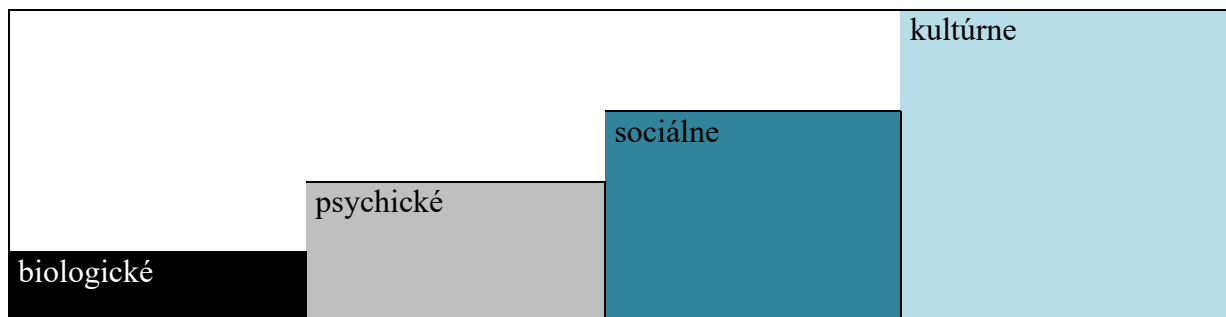
Po štvrté, nejde tu o žiadne subjektívne tvrdenie, pretože do úvahy treba vziať aj vzájomné vzťahy, resp. *hierarchiu vzťahov medzi jednotlivými parametrami ľudskej prirodzenosti*, ako aj ich *dynamickosť*. Z hľadiska začlenenia do vesmírneho vývinového radu je určujúca (a tak najdôležitejšia) *sociálnosť človeka*, pretože všetko nasvedčuje tomu, že to bolo v rámci vyššie spomínanej antroposociogenézy *práve sociálne*, pomocou a v rámci ktorého sa človek pred niekoľko desaťtisíc rokmi vyčlenil zo živočíšnej ríše, pričom nie je dôležité, či to bolo presne tak, ako predpokladá napríklad Chris Knight (1988), t. j. po zavedení nejakého druhu morálky alebo všeobecne rešpektovaného zákazu, alebo tak, ako predpokladá Jurij I. Semenov (1966), čiže po úplnom dotvorení prvej (čisto) ľudskej spoločnosti, ktorou bola exogamná duálno-rodová a matrilinárna spoločnosť, resp. tak, ako v rámci svojho rozsiahleho a všestranného výkladu tvrdí Boris F. Poršnev (1979), t. j. na konci celej série neurofyziologických a súčasne „sociálnych“ a psychických procesov (interdikcie, sugescie a kontrasugescie), pretože dôležitá je práve *integrujúca a riadiaca funkcia sociálneho* v rámci novej FPH a nového prirodzeného druhu. Najdôležitejším a hlavným parametrom sociálnej FPH je *teda skutočne sociálne*.

Najdynamickejším parametrom je však určite *kultúrne* a do neho začlenené *technické* (alebo *technologické*), pričom práve na tomto mieste treba oceniť genialitu Karola Marxa, ktorý vo svojich spisoch správne zdôrazňoval, že najprv prichádza zmena vo výrobných silách a až po nej zmena vo výrobných a sociálnych vzťahoch v širokom slova zmysle, v rámci ktorých sa ľudia snažia prispôsobiť novej dynamike a/alebo efektívnosti, ktorú do ich života priniesol ten či onen vedecký a technický vynález. Čím prevratnejší je pritom daný vynález, tým radikálnejšie následne ovplyvní spoločnosť, ako je zrejmé zo všetkých štyroch priemyselných revolúcií, ktoré doteraz vždy viedli k nárastu ľudskej populácie či výraznému skvalitneniu jej života. Prevratné technické a technologické zmeny boli zároveň sprevádzané *rovnako prevratnými zmenami v ľudských inštitúciách*, ktoré sú *takisto súčasťou ľudskej prirodzenosti*, bez ohľadu na to, že v diskusiách o nej sa k nim a ich postupným zmenám takmer vôbec neprihliada. Na druhej strane, biologické a psychické predstavujú dva najrigidnejšie a preto aj najstabilnejšie parametre ľudskej prirodzenosti, ktoré však tvoria prvú polovicu jej definície a ako také sú najviac ohrozené zánikom (či drastickou premenou) pri vzniku postsociálneho zo sociálneho.

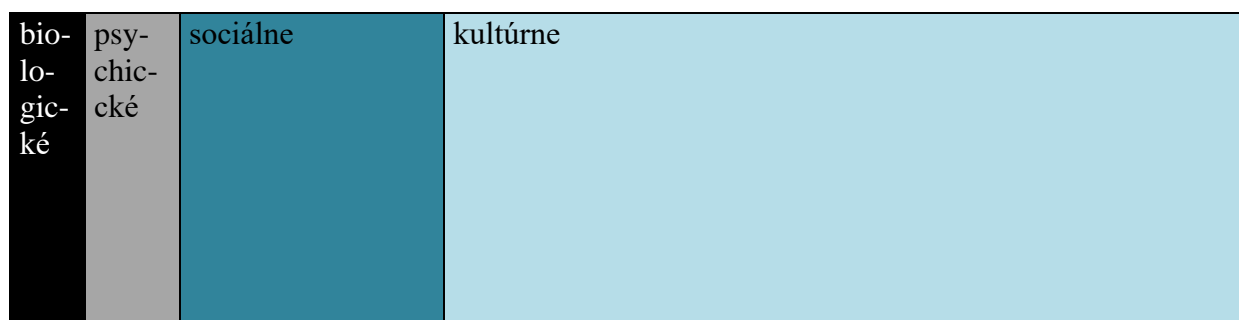
V konečnom dôsledku sa tak zreteľne vyjasňuje pravá povaha ľudskej prirodzenosti, ktorá je *súčasne stabilná aj dynamická*, t. j. *nemeniaca sa a meniaci sa súčasne*. Hierarchické vzťahy medzi jej štyrmi parametrami sa teda zachovávajú, ale zároveň čoraz viac nadobúdajú na váhe parametre, ktoré sa už „z prirodzenosti vecí“ menia najčastejšie a najvýraznejšie, t. j. sociálne a kultúrne (a spolu s nimi aj technické a technologické). K zásadnej zmene pritom dochádza v 19. storočí s nástupom prvej priemyselnej revolúcie, odkedy všetky najdôležitejšie rastové krivky (hospodárskej produkcie, populačnej dynamiky a pod.) nadobúdajú exponenciálny tvar, *signalizujú tak blížiacu sa zásadnú a radikálnu zmenu*, spočívajúcu podľa všetkého v prechode celého civilizačného systému na úplne novú vývinovú, resp. vývojovú trajektóriu, a to tak v pozitívnom slova zmysle, s predpokladom vzniku novej kvality, t. j. novej formy pohybu hmoty, ako aj v negatívnom slova zmysle, s predpokladom zániku ľudskej civilizácie v jej súčasnej podobe v dôsledku celej série endogénnych kríz, ktorým musí čeliť, ale nedá sa vylúčiť ani jej rýchly zánik v dôsledku aktuálne málo pravdepodobnej katastrofy zapríčinennej napríklad kozmickým impaktom, explóziou supervulkánu alebo náhlou klimatickou zmenou.

V každom prípade, isté je, že praveký človek sa zdá byť identickým so súčasným len vtedy, ak *vyslovene redukcionisticky a naivne* odhliadneme od spoločnosti, v ktorej obaja žijú (od celého súboru permanentne sa vyvíjajúcich a preto aj meniacich inštitúcií, do ktorých sú obaja začlenení), keď odignorujeme skutočnosť, že vďaka technickým výtvarným sa ľudský život mení nielen čoraz radikálnejšie, ale že dané zmeny *sú aj čoraz častejšie*, pričom súčasná veda pokročila tak ďaleko, že v čoraz väčšej miere dokáže naprávať to, čo „pokazila“ príroda, keď

si jednoducho (biocentricky a antropocentricky) nevšimneme, že pri každej výraznejšej zmene hospodárskej či produkčnej základne spoločnosti *sa výrazne mení aj život jedinca*<sup>33</sup>, ktorý má dnes celkom iné potreby, záujmy a ciele, ako mal pred niekoľko desaťtisíc rokmi, ako to jasne vyjadrujú histogramy, zachytávajúce premenlivosť štyroch parametrov ľudskej prirodzenosti.



**Obr. č. 1:** Zobrazuje hierarchiu vzťahov medzi štyrmi parametrami ľudskej prirodzenosti, ako aj mieru ich premenlivosti na začiatku ľudských dejín, t. j. po (a pri) vzniku *Homo sapiens*.



**Obr. č. 2:** Zobrazuje hierarchiu vzťahov medzi štyrmi parametrami ľudskej prirodzenosti, ako aj mieru ich premenlivosti v súčasnosti, t. j. v poslednej fáze (čisto) ľudských dejín.

Ešte inak povedané, podobne ako objavenie sa interdikcie, sugescie a kontrasugescie v rode *Homo* predznamenovalo vznik nového prirodzeného a *nie(len) biologického* druhu, t. j. zásadne novej vývinovej kvality alebo formy, aj vznik ANI, AGI, BCI, CBI a pod. *predznamenoáva vznik nového prirodzeného druhu*, o vlastnostiach a podobe ktorého však v tejto chvíli nemá zmysel špekulovať, pretože všetky prognostické či futuristické úvahy tohto typu sú zásadne limitované spomínanou neexistenciou *dostatočne vedecky fundovanej a vnútorne nerozpornej všeobecnej vývojovej teórie*. Bez ohľadu na to, kedy a či vôbec bude takáto teória v dohľadnej dobe sformulovaná, však môžeme celkom na záver konštatovať, že vyššie uvedená definícia ľudskej prirodzenosti už dnes vnáša do predmetnej oblasti tak prepotrebný „poriadok“ alebo „konceptuálnu čistotu“, pretože naznačuje, že prenášať pojem „sociálne“ mimo ľudskej sféry činnosti, resp. ľudského bytia znamená redukovať jeho bohatý obsah len na púhu spoluprácu (napríklad medzi termittmi v kolónii), imanentne a logicky konzekventne prepája individuálne

<sup>33</sup> Pravdaže, celkom inak sa tieto vzťahy javia *ľudskému jedincovi osobne*, ktorý bezprostredne vníma realitu najmä cez biologické a psychické a u ktorého je podľa R. F. Baumeistera et al. (1996) a K. L. Sommerovej (1997) 95 percent denných aktivít vykonávaných automaticky a len 5 percent s kauzálnou účasťou vedomia.

a kolektívne v človeku a spomedzi možných zásahov do ľudského organizmu jednoznačne favorizuje tie, ktoré uľahčia a prehĺbia jeho *bytočné prepojenie* s technosférou a sociosférou (v jasnej opozícii voči transhumanistickému antropocentrizmu, individualizmu a hedonizmu).

## Seznam použitých zdrojů

- ARENDOVÁ, H. 1996. *Původ totalitarismu I-III*. Praha: OIKOYMENH 1996. 680 s. ISBN 80-86005-13-5.
- BAKER, L. R. 2000. *Persons and Bodies. A Constitution View*. Cambridge: Cambridge University Press 2000. 233 s. ISBN 0-521-59263-1.
- BAKER, L. R. 2013. *Naturalism and the first-person perspective*. New York-Oxford: Oxford University Press 2013. 248 s. ISBN 978-0-19-991474-6.
- BARNIER, A. 2012. Otto and the barking dog: How the extended mind thesis helps us understand collaborative remembering. *Paper presented at Distributed Cognition and Distributed Agency Workshop, March 16–18, Macquarie University, Sydney*.
- BAUMEISTER, R. F., HEATHERTON, T. 1996. Self-Regulation Failure: An Overview. In *Psychological Inquiry*, 7 (1996), s. 1-5.
- BAUMEISTER, R. F., SOMMER, K. L. 1997. Consciousness, Free Choice, and Automaticity. In WYER Jr., R. S. (ed.). *Advances in Social Cognition*, vol. X (Mahwah, NJ: Erlbaum, 1997).
- BENDER, E. M., KOLLER, A. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 5185–5198).
- BINGHAM, W. V. 1937. *Aptitudes and aptitude testing*. New York: Harper&Brothers 1937. 390 s.
- BORODITSKY, L. 2011. How Language Shapes Thought. In *Scientific American* (February): 63-65.
- BOSTROM, N. 2014. *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press 2014. 328 s. ISBN 978-0-19-967811-2.
- BOWMAN, S. R. 2023. *Eight Things to Know about Large Language Models*. [cit. 2023-10-15]. Dostupné na internete: <<https://browse.arxiv.org/pdf/2304.00612.pdf>>.
- BRINGSJORD, S., GOVINDARAJULU, N. S. 2018. Artificial Intelligence. In ZALTA, E. N. (ed.). *Stanford Encyclopedia of Philosophy*. [cit. 2023-10-15]. Dostupné na internete: <<https://plato.stanford.edu/entries/artificial-intelligence/>>.
- BUBECK, S, CHANDRASEKARAN, V., ELDAN, R., GEHRKE, J., HORVITZ, E., KAMAR, E., LEE, P., LEE, Y. T., LI, Y., LUNDBERG, S., NORI, H., PALANGI, H., RIBEIRO, M. T., ZHANG, Y. 2023. *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. [cit. 2023-10-15]. Dostupné na internete: <<https://browse.arxiv.org/pdf/2303.12712.pdf>>.
- BURGAN, R. 2012. Ako (asi) vzniklo ľudské vedomie (a prečo sa ním disponujú iba ľudia). In *E-LOGOS. Electronic Journal for Philosophy*, 12/2012. ISSN 1211-0442, s. 1-56.
- BURGAN, R. 2014. Prečo nie je možné vylepšiť, ale len definitívne nahradiť a prekonať človeka. In *E-LOGOS. Electronic Journal for Philosophy*, 5/2014, ISSN 1211-0442, s. 1-16.
- BURIAN, J. 2008. Myšlení modelů. In KVASNIČKA, V., KELEMEN, J., POSPÍCHAL, T. (eds.). *Modely mysle*. Bratislava: Vydavateľstvo Európa 2008. 248 s. ISBN 978-80-89111-34-3, s. 9-34.
- BOTTERILL, G., CARRUTHERS, P. 2003. *The Philosophy of Psychology*. Cambridge: Cambridge University Press 2003. 297 s. ISBN 0 511 01164-4.
- CARROLL, J. B. 1993. *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press. 819 s. ISBN 0-521-38712-4.
- CATTELL, R. B. 1987. *Intelligence: Its Structure, Growth, and Action*. New York: Elsevier 1987. 694 s. ISBN 0 444 87922 6.



- CHAISSON, E. 2005. *Epic of evolution: seven ages of the cosmos*. New York: Columbia University Press 2005. 478 s. ISBN 0-231-13560-2.
- CHALMERS, D. 1996. *The conscious mind: In search of a fundamental theory*. New York – Oxford: Oxford University Press 1996. 363 s. ISBN 0-19-510553-2.
- CLARK, A. 2001. Reasons, robots, and the extended mind. In *Mind and Language* 16: 121-145.
- COLLIER, J. D. 1999. Autonomy in Anticipatory Systems: Significance for Functionality, Intentionality and Meaning. In DUBOIS, D. M. (ed.). *Proceedings of CASYS'98, The Second International Conference on Computing Anticipatory Systems*. (New York: Springer-Verlag, 1999).
- COLLIER, J. D. 2006. Conditions for Fully Autonomous Anticipation. In DUBOIS, D. M. (ed.). *Computing Anticipatory Systems: CASY'05 – Sixth International Conference*, American Institute of Physics, Melville, New York, AIP Conference Proceedings 839: 282-289.
- COPELAND, J. B., PROUDFOOT, D. 2007. Artificial Intelligence: History, Foundations, and Philosophical Issues. In THAGARD, P. (ed.). *Philosophy of Psychology and Cognitive Science*. New York: Elsevier 2007. ISBN 978-0-444-51540-7, s. 429-482.
- DIEDERICH, J. 2021. *The Psychology of Artificial Intelligence*. Cham: Springer Nature Switzerland 2021. 140 s. ISBN 978-3-030-71842-8.
- DREYFUS, H. L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, Mass.: MIT Press 1992. 353 s. ISBN 0-262-04134-0.
- FOGEL, D. B. 1995. Review of Computational Intelligence: Imitating Life. *Proc. of the IEEE*, 83(11), 1995.
- FUKUYAMA, F. 2002. *Our Posthuman Future: Consequences of the Biotechnology revolution*. New York: Farrar, Strauss and Giroux 2002. 272 s. ISBN 978-0-374-70618-0.
- GÁLIKOVÁ, S. 2011. Na ceste k hľadaniu vedy o mysli. In KVASNIČKA, V., POSPÍCHAL, I., NÁVRAT, P., LACKO, P., VARGA, Ľ. (eds.). *Umelá inteligencia a kognitívna veda III*. Bratislava: Slovenská technická univerzita 2011. ISBN 978-80-227-3542-1, s. 63-79.
- GARDNER, H. 1993. *Frames of Mind. Theory of Multiple Intelligences*. London: Fontana Press 1993. 463 s. ISBN 978-0-006-8629-1.
- GELERNTER, D. 2007. Artificial intelligence is lost in the woods. *MIT Technology Review*, July 2007.
- GOERTZEL, B. 2006. *The Hidden Pattern: A Patternist Philosophy of Mind*. Irvine: Brown Walker Press 2006. 424 s. ISBN 1-59942-404-5.
- GREENFIELD, S. 2002. Mind, brain and consciousness. In *British Journal of Psychiatry* (2002), 181, 91-93.
- GREENWOOD, J. 2015. *Becoming Human. The Ontogenesis, Metaphysics, and Expression of Human Emotionality*. Cambridge, Mass.: MIT 2015. 250 s. ISBN 978-0-262-02978-0.
- GUILFORD, J. P. 1967. *The Nature of Human Intelligence*. New York: McGraw-Hill 1967. 538 s. ISBN 978-0-072-05135-3.
- GUDWIN, R. R. 2000. Evaluating intelligence: A computational semiotic perspective. In *IEEE International conference on systems, man and cybernetics*, pages 2080-2085, Nashville, Tennessee, USA, 2000.
- HABERMAS, J. 2003. *Budoucnost lidské přirozenosti. Na cestě k liberální eugenice?* Prel. A. Bakešová. Praha: Filosofie 2003. 132 s. ISBN 80-7007-174-5.
- HARA, F. 2004. Artificial Emotion of Face Robot Through Learning in Communicative Interactions with Humans. *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, 7-15, 2004.
- HARRÉ, R. 1993. *Social Being*. Oxford: Blackwell 1993. 288 s. ISBN 0-631-18782-0.
- HOFFMANN, C. H. 2022. *The Quest for Universal Theory of Intelligence. The Mind, the Machine, and Singularity Hypotheses*. Berlin-Boston: Walter de Gruyter 2022. 283 s. ISBN 978-3-11-075616-6.
- HUTTER, M. 2005. *Universal AI. Sequential Decisions Based on Algorithmic Probability*. Berlin: Springer 2005. 280 s. ISBN 3-540-22139-5.
- HUXLEY, J. 1957. *New Bottles for New Wine*. London: Chatto&Windus 1957. 318 s. ISBN 978-0701108380.

- JAMES, W. 1930. *Druhy náboženské zkušenosti*. Praha: Melantrich 1930. 323 s.
- KELEMEN, J. 2010. *Myslenie a stroj*. Bratislava: Kalligram 2010. 388 s. ISBN 978-80-8101-243-3.
- KNIGHT, C. 1988. *The Origins of Society*. [cit. 2023-10-15]. Dostupné na internete: <[http://www.radicalanthropologygroup.org/old/pub\\_origins\\_hs.pdf](http://www.radicalanthropologygroup.org/old/pub_origins_hs.pdf)>.
- KOVÁČ, L. 2008. Bioenergetika: kľúč k pochopeniu mozgu a mysle. In KVASNIČKA, V., KELEMEN, J., POSPÍCHAL, T. (eds.). *Modely mysle*. Bratislava: Vydavateľstvo Európa 2008. 248 s. ISBN 978-80-89111-34-3, s. 100-126.
- LEGG, S., HUTTER, M. 2007a. *A Collection of Definitions of Intelligence*. [cit. 2023-10-15]. Dostupné na internete: <<http://arxiv.org/pdf/0706.3639v1.pdf>>.
- LEGG, S., HUTTER, M. 2007b. *Universal Intelligence: A Definition of Machine Intelligence*. [cit. 2023-10-15]. Dostupné na internete: <[http://www.veta.org/documents/Universal Intelligence.pdf](http://www.veta.org/documents/Universal%20Intelligence.pdf)>.
- LOOSEMORE, R., GOERTZEL, B. 2012. Why an Intelligence Explosion is Probable. In EDEN, A. H., MOOR, J. H., SØRAKER, J. H., STEINHARDT, E. (eds.). *Singularity Hypotheses. A Scientific and Philosophical Assessment*. Heidelberg – New York – Dordrecht – London: Springer 2012. ISBN 978-3-642-32560-1, s. 83-96.
- MAISELI, B., ABDALLA, A. T., MASSAWE, L. V., MBISE, M., MKOCHA, K., NASSOR, N. A., ISMAIL, M., Michael, J., KIMAMBO, S. 2023. *Brain-computer interface: trend, challenges, and threats*. [cit. 2023-10-15]. Dostupné na internete: <<https://doi.org/10.1186/s40708-023-00199-3>>.
- METZGER, S. L., LITTLEJOHN, K. T., SILVA, A. B., MOSES, D. A., SEATON, M. P., WANG, R., DOUGHERTY, M. E., LIU, J. R., WU, P., BERGER, M. A., ZHURAVLEVA, I., TU-CHAN, A., GANGULY, K., ANUMANCHIPALLI, G. K., CHANG, E. F. 2023. A high-performance neuroprosthesis for speech decoding and avatar control. In *Nature* 620, 1037-1046 (2023). [cit. 2023-10-15]. Dostupné na internete: <<https://doi.org/10.1038/s41586-023-06443-4>>.
- MORAVEC, H. 1988. *Mind Children. The future of robot and human intelligence*. Cambridge, Mass.: Harvard University Press 1988. 224 s. ISBN 0-674-57618-7.
- NADDAF, M. 2023. Europe spent €600 million to recreate the human brain in computer. How did it go? In *Nature*, 22 August 2023.
- NEWPORT, C. 2023. What Kind of Mind Does ChatGPT Have? *The New Yorker*, April 13, 2023.
- PAPINI, G. 1924. *The Failure*. New York: Harcourt, Brace and Company 1924. 336 s. ISBN 141790836x.
- PERRIGO, B. The A to Z of Artificial Intelligence. In *Time*, April 13, 2023.
- POPPER, M. 2003. Evolučná psychológia a štandardný model sociálnych vied. In *Filozofia*, roč. 58, č. 1, s. 10-22. ISSN 2585-7061 (online).
- PORŠNEV, B. F. 1979. *O začiatkoch ľudských dejín*. Bratislava: Pravda 1979. 472 s.
- ROBINSON, A. 2015. Philosophical challenges. In FRANKISH, K., RAMSEY, W. M. (eds.). *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press 2015. ISBN 978-0-521-87142-6, s. 64-85.
- ROOIJ, I. V., GUEST, O., ADOLFI, F., HAAN, R. D., KOLOKOLOVA, A., RICH, P. 2023. *Reclaiming AI as a theoretical tool for cognitive science*. [cit. 2023-15-10]. Dostupné na internete: <<https://osf.io/preprints/psyarxiv/4cbuv/>>.
- SANDBERG, A. 1999. The physics of information processing superobjects: daily life among the Jupiter brains. *Journal of Evolution and Technology*, 5(1), 1-34.
- SEMENOV, J. I. 1966. *Kak vzniklo čelovečestvo*. Moskva: Nauka 1966. 576 s.
- SHARON, T. 2014. *Human Nature in an Age of Biotechnology, The Case for Mediated Posthumanism*. Dordrecht: Springer 2014. 241 s. ISBN 978-94-007-7554-1.
- SHUMAKER, R., WALKUP, K., BECK, B. 2011. *Animal tool behavior: the use and manufacture of tools by animals*. Baltimore: JHU Press 2011. 282 s. ISBN 978-0-8018-9853-2.

- SITARČÍKOVÁ, Z. 2012. *O ŠLACHTENÍ ČLOVEKA/Nanotechnologie, transhumanismus a lidská prirodzenost*. Trnava-Bratislava: Typi Universitatis Tyrnaviensis-VEDA, vydavateľstvo SAV 2012. 214 s. ISBN 978-80-8082-560-7.
- SOPER, K. 2001. Richard Rorty: Humanist and/or Anti-humanist. In FESTENSTEIN, M., THOMPSON, S. (eds.). *Richard Rorty: Critical Dialogues*. Cambridge: Polity 2001. ISBN 978-0745621661, s. 115-129.
- SPEARMAN, C. E. *The abilities of man, their nature and measurement*. New York: Macmillan 1927. 458 s. ISBN 978-1932846102.
- SPIVEY, M., RICHARDSON, D. 2009. Language processing embodied and embeded. In ROBBINS, P., AYDEDE, M. (eds.). *The Cambridge Handbook of Situated Cognition*. New York: Cambridge University Press 2009. 520 s. ISBN 978-0-521-61286-9, s. 382-400.
- STEVENSON, L., HABERMAN, D. L. 1998. *Ten Theories of Human Nature*. Third Edition. New York-Oxford: Oxford University Press 1998. 239 s. ISBN 0-19-512041-8.
- STEWART, J. 2000. *Evolution's Arrow. The Direction of Evolution and the Future of Humanity*. Canberra: The Chapman Press 2000. 176 s. ISBN 0-646-39497-5
- SUOJANEN, M. 2015. A Direct Object of Perception. In *E-LOGOS*. Electronic Journal for Philosophy, 1/2015, ISSN 1211-0442, s. 28-36.
- SÝKORA, P. 2014. Kritika teologicko-teleologickej koncepcie ľudskej prirodzenosti ako dôvodu odmietnutia transhumanizmu. In *Filozofia*, roč. 69, č. 6, s. 514-525. ISSN 2585-7061 (online).
- SÝKORA, P. 2015. Etika biotechnologických zásahov do ľudskeho genómu: argumenty rizikovosti a zničenia ľudskej prirodzenosti. In *Filozofia*, roč. 70, č. 5, s. 329-342. ISSN 2585-7061 (online).
- SÝKORA, P. 2018. Moderný návrat k Aristotelovi. In *Filozofia*, roč. 73, č. 6, s. 425-436. ISSN 2585-7061 (online).
- SÝKORA, P. 2019a. K posthumánnemu človeku prostredníctvom editovania génov pre kognitívne schopnosti. In *Filozofia*, roč. 74, č. 7, s. 511-529. ISSN 2585-7061 (online).
- SÝKORA, P. (ed.). 2019b. *Promises and perils of emerging technologies for human condition. Voices from four postcommunist Central and East European countries*. Bratislava-Berlin: VEDA, vydavateľstvo SAV-Peter Lang 2019. 208 s. ISBN 978-3-631-77512-7.
- TAINTER, J. A. 2009. *Kolapsy složitých společností*. Praha: Dokořán 2009. 320 s. ISBN 978-80-7363-248-9.
- TEGMARK, M. 2020. *Život 3.0. Člověk v éře umělé inteligence*. Praha. Dokořán 2020. 295 s. ISBN 978-80-7363-881-8.
- TEILHARD DE CHARDIN, P. 1990. *Vesmír a lidstvo*. Praha: Vyšehrad 1990. 265 s. ISBN 8070210435.
- THAGARD, P. (ed.). *Philosophy of Psychology and Cognitive Science*. New York: Elsevier 2007. 502 s. ISBN 978-0-444-51540-7.
- The year of brain-computer interfaces. In *Nature Electronics*, editorial, 6, 643, 25 September 2023.
- URBAN, T. 2015a. *The AI Revolution: The Road to Superintelligence. Part 1*. [cit. 2023-15- 10]. Dostupné na internete: <<http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>>.
- URBAN, T. 2015b. *The AI Revolution: The Road to Superintelligence. Part 2*. [cit. 2023-15- 10]. Dostupné na internete: <<http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html>>.
- ULE, A. 2015. Consciousness, Mind, And Spirit: Three Levels of Human Cognition. In *Interdisciplinary Description of Complex Systems*, 13(4), 488-500, 2015.
- VALLOR, S. 2017. AI and the Automation of Wisdom. In Powers, T. M. (ed.). *Philosophy and Computing. Essays in Epistemology, Philosophy of Mind, Logic, and Ethics*. New York City, NY: Springer 2017. ISBN 978-3319610429, s. 161-178.
- VILLALOBOS, P., SEVILLA, J., HEIM, L., BESIROGLU, T., HOBBAHN, M., HO, A. 2022. *Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning*. [cit. 2023-15-10]. Dostupné na internete: <<https://arxiv.org/pdf/2211.04325.pdf>>.
- VINGE, V. 1993. *The Coming Technological Singularity: How to Survive in the Post-human Era*. [cit. 2023-15-10]. Dostupné na internete: <<https://edoras.sdsu.edu/~vinge/misc/singularity.html>>.

- VIŠŇOVSKÝ, E. 2015. Idea „vylepšovania človeka“ z hľadiska filozofie pragmatizmu. In *Filozofia*, roč. 70, č. 5, s. 343-355. ISSN 2585-7061 (online).
- YAMAMOTO, M. 1993. SOZZY: A Hormone-Driven Autonomous Vacuum Cleaner. In *Proceedings of the International Society for Optical Engineering*, 2058: 211–222, 1993.
- YAMPOLSKI, R. V. 2016. *Artificial Superintelligence. A Futuristic Approach*. Boca Raton: Taylor & Francis 2016. 189 s. ISBN 978-1-4822-3444-2.
- WATANABE, M. 2022. *From Biological to Artificial Consciousness. Neuroscientific Insights and Progress*. Cham: Springer Nature Switzerland 2022. 172 s. ISBN 978-3-030-91138-6.
- WILLETT, F. R., KUNZ, E. M., FAN, C. et al. 2023. A high-performance speech neuroprosthesis. *Nature* 620, 1031–1036 (2023). [cit. 2023-15-10]. Dostupné na internete: <<https://doi.org/10.1038/s41586-023-06377-x>>.
- WILEY, K. B. 2014. *A Taxonomy and Metaphysics of Mind-Uploading*. Seattle: Humanity Press and Alautun Press 2014. 190 s. ISBN 978-0-692-27984-7.
- WILEY, K. B. 2023. *Nondestructive Mind Uploading and the Stream of Consciousness*. [cit. 2023-15-10]. Dostupné na internete: <[https://keithwiley.com/Downloads/WileyK\\_NondestructiveMindUploadingAndTheStreamOfConsciousness\\_2023.pdf](https://keithwiley.com/Downloads/WileyK_NondestructiveMindUploadingAndTheStreamOfConsciousness_2023.pdf)>.
- ZARKA, Y. C. 2018. Má ešte pojem ľudská prirodzenosť zmysel? In *Filozofia*, roč. 73, č. 10, s. 777-789. ISSN 2585-7061 (online).
- ZUBKOV, I. F. 1981. *Problém geologickej formy pohybu hmoty*. Bratislava: Pravda 1981. 272 s.