FILOZOFIA
Roč. 78, 2023, č. 8
DOI: https://doi.org/10.31577/filozofia.2023.78.8.7

MARK COECKELBERGH:

Etika umělé inteligence

Praha: Filosofia 2023, 267 s.

Začala nová doba. Nové doby ovšem začínají neustále, každým okamžikem. Tato doba je údajně opravdu nová: doba arteficiální inteligence (AI), v níž a jíž se všechno promění. Neměli bychom zapomínat, že označit něco za „nové“ či vůbec používat kategorii novosti, není jen tak. „Nové“ je ontologické určení a mluvíme-li o novém, rozšiřujeme tím sféru jsoucna o něco, co v ní dřív, před tímto naším určením, nebylo. AI v tomto smyslu „novost“ představuje: je oblastí technické novosti, v které probíhá stálá expanzivní diferenciace technických systémů do funkčně rozlišených systémových typů. Jako „to nové“ je produktem lidské činnosti a spadá do sféry výkonů člověka sledujících lidské účely, prvořadě převzetí strojové výroby, omezení namáhavosti práce a maximalizace produktivity. Není konstruována jako hračka pro akademické spekulace o fiktivní morálce digitalizovaných těl a myslí robotů.

AI jako problém nového tudíž nevzniká z vlastního účelu, ale jako vedlejší důsledek nezamýšlených důsledků. Fakt, že jsou to především civilizační rizika (mezi než se z kloubů vymknutá umělá inteligence počítá), jež vznikají ve vedlejším, nezamýšleném důsledku mnoha nedomyšlených lidských rozhodnutí, popsal nejvýstižněji Ulrich Beck v *Rizikové společnosti*. Jeho tezi je však možné snadno rozšířit: naše dějiny jsou souhrnem nezamýšlených důsledků vedlejších neočekávaných důsledků. Vedlejší důsledky nás formují, vytvářejí náš *lebenswelt*, definují naše snažení, očekávání, obavy a perspektivy. V dějinách najdeme skutečně všechno, vyjma předem zdůvodněné, jasné a bezproblémově naplňované intence, kterou by zformulovali vždy aktuálně kompetentní, plně zodpovědní lidé. Dějiny jsou dějinami toho, co se nepodařilo tak, jak jsme zamýšleli, a nyní to vyžaduje hasit jeden požár vedle druhého. Problém umělé inteligence a superinteligence se vyhrotil až do současných, místy bujně apokalyptických souvislostí, protože se stal ve svých nezamýšlených důsledcích uvědomovaným rizikem naší rizikové společnosti.

Mark Coeckelbergh je belgický filosof, jenž působí na vídeňské univerzitě, odkud pěstuje řadu odborných kontaktů, mimo jiné také s bratislavským Filosofickým ústavem SAV, který navštívil u příležitosti konferenční přednášky. Vyznačuje ho široké spektrum vědních zájmů, mezi něž spadají různé aspekty a souvislosti antropocénu.

Ve vedlejším plánu tak sleduje rovněž některé otázky arteficiální inteligence, jak dokládá český překlad jeho knihy *Etika umělé inteligence*, vydaný pražským nakladatelstvím Filosofia v roce 2023. Formálně je kniha rozvržena do dvanácti kapitol úzce provázaných opakující se otázkou po morálních (ne-)schopnostech uměle inteligentních systémů. Její kostru tvoří úvaha o etických výzvách, které představuje dnešní umělá inteligence a umělá inteligence nejbližší budoucnosti, a jejich dopadu na současné společnosti; jde o ovlivnění našeho života všudypřítomnými, výkonnými a čím dál inteligentnějšími technologiemi, o etické, právní a politické otázky, o problémy lidského poznání, konstituce lidské společnosti a o charakter lidské morálky. Na jednu knihu toho není málo, zvlášť pokud předpokládáme, že Coeckelbergh také objasní, co to AI je, jaké jsou její navazující body ke společnosti a individuu, jaká zobecnění nutná pro celokulturní (civilizační) ustanovení poskytuje, jaká rizika s ní spojená musíme předvídat a odkud brát reálné normy symbiózy společností a AI, aby se nejednalo o pouze voluntaristické projevy subjektivních morálních přesvědčení. V této oblasti kniha poněkud zaostává za případnými nároky vůči ní a získává místo analytického výrazně přehledový ráz.

Coeckelbergh zmiňuje v drobných poznámkách co nejvíce jednotlivých činitelů v problematice umělé inteligence a většinou jen naznačí, ale nerozvine zdůvodňující myšlenku. Velmi mnoho si uvědomuje, ale jen ojediněle předvádí víc než krátkou deskripci; text se tak mění v souhrn či podrobnou antologii otázek a názorů spojených s umělou inteligencí a s jí formovanou lidskou pospolitostí. Takový postup je jistě následováníhodný, pokud si uvědomíme, že autor mluví o dosud plně nezformulovaném problému s tekutými a nejasnými hranicemi, které je snadné věcně překračovat a dostávat se do předmětně odlišných souvislostí. Je to práce na půdorysu problému AI, jež bude nepochybně dál pokračovat a najde mnoho následovníků. Současně ale vytěsňuje komplexní analytiku zásad umělé inteligence a ponechává výkladu vnitřně neucelený, vlastně jen konfrontační charakter uvádění příkladů a protipříkladů, koncepcí a opačných koncepcí, návrhů a protinávrhů, názorů, které by chtěly být teoriemi, ale nemají ani zdůvodněnou hypotézu. Coeckelberghův popis jevových forem, v nichž vystupuje teoretická reflexe umělé inteligence, klade důraz na opozita. V základech je vždy rozpor: od prvotního, poněkud úsměvného protikladu pozdního osvícenství a romantiky, v němž se měla zrodit idea moderních technologií, přes rozpor humanismu a transhumanismu, podobný svár transhumanismu a posthumanismu až po další bipolární vztahová určení, jako je obecná inteligence vs. superinteligence, morální aktérství vs. morální trpnost, technologie vs. singulární technologie, člověk vs. stroj nebo mozek vs. počítač. Coeckelbergh sám zmiňuje, že diskusi o AI nesou narativy o všestranné konkurenci, přes které by se chtěl přenést k imanentnějším způsobům vyjádření smyslu umělé inteligence, ne vždy ale tohoto cíle dosahuje.

Má-li se k pochopení umělé inteligence využít filosofie a věda, jak Coeckelbergh požaduje, je nezbytné postupovat s využitím filosofických a vědeckých metod a poznávacích prostředků. Mezi ně nepatří tolik autorova otázka, zda je umělá inteligence možná (jako potence nepochybně, což ale neimplikuje závěr, že stroje budou mít stejné morální schopnosti jako lidé); pozornost by měl naopak vzbudit problém, *jak* je AI možná. Ten vyžaduje metodicky vyjasněné prostředí, v kterém bude položen jako prvořadě ontologický a noetický problém bez etických bezbřehých konotací. Příslušní specialisté musí nejprve promyslet otázku, jak vůbec klást problém umělé inteligence, mají-li se dobrat nějakých statusových tezí. Jak pokřivené je uvažování v tomto směru se dobře ukazuje na domnělé filosofii myšlenkových experimentů (Coeckelbergh mluví o „tramvajovém dilematu"): často se objevující dilemata typu autonomního automobilu, jenž se „rozhoduje" mezi přejetím dítěte a usmrcením řidiče, jsou pouhá karikatura filosofie. Umělá inteligence není otázkou dopravních situací, ale zásahu určitých lidských výtvorů do života jednotlivců a celých společností. Diskuse o ní ustrne, nebude-li objasňovat prvořadě, *co* do našeho života zasahuje, *jak* tyto zásahy probíhají a *za jakým účelem* nás formují. Co si s tím chceme počít, je odvozený problém, na který budeme reagovat tím adekvátněji, čím přesnější budou naše odpovědi.

Coeckelbergh otevírá ve svém přehledu řadu perspektiv, v nich může umělá inteligence vystupovat. Podle očekávání se jedna z nejsilnějších spojuje s informačními a komunikačními technologiemi, jejich využíváním, ale především zneužíváním ke sledování lidí, shromažďování velkých dat a ke skryté účinné manipulaci s jednotlivci a (zejména spotřebitelskými) masami. Na tuto souvislost jsme samozřejmě citliví, protože nás vede k úvaze o dohledu nad společností, z níž se může vyvinout některá z forem totality. Opírá se o ni také dnešní autoritářská politika, která se s využitím jejích možností dokáže stále lépe obejít bez demokracie. Zastíněna obavou z celospolečenské špionáže zůstává ale zcela bez povšimnutí jiná, bazální souvislost: politicko-ekonomická. Autor několikrát opakuje otázku, zda je AI „jen" stroj, nikde ale nevysvětluje, co stroj je; stranou zůstávají i další určení, která lze považovat za potřebná, nemá-li se diskuse o AI převést na mudrování o jejím morálním profilu.

Tak by odpověď vyžadoval problém produktivity a AI, vhodný by byl výklad pracovního výkonu a vůbec toho, co je v kontextu AI práce, co vícepráce, nutná práce a jak umělá inteligence vytváří nadhodnotu, jak ekonomicky expanduje atd. Od věci není ani připomínka, že odněkud se musí vzít materiály, na nichž umělá inteligence závisí; potřebné vzácné nerosty jsou již dnes předmětem válečných konfliktů a pokud se stanou obchodní protekcionistickou zbraní (příp. jejich zdroje budou vysychat), radikálně to změní celou ekonomickou souvislost našeho digitálního světa. Politická ekonomie umělé inteligence by o ní a její základně (AI jako technický systém

s dalšími navazujícími možnostmi) řekla víc než sveřepé úvahy o její morální svéhlavosti. Zřejmě se stále čeká na podněty ze strany ekonomů.

Recenzovanou knihu uzavírá kapitola o umělé inteligenci, klimatické změně a antropocénu. Je uvedena vcelku emotivně: zatímco obyvatelé jedné části světa bojují o přístup k pitné vodě, obyvatelé jiné části se obávají o své soukromí na internetu. To ovšem není problém umělé inteligence, ale naší mundánní civilizace ovládané globálním kapitálem. Mezi jeho důsledky patří také to, že nedýcháme stejný vzduch: zatímco někteří se dusí ve smogu velkoměst či v dýmu požárů, jiní odlétají do destinací se zdravou přírodou nebo používají výkonné čističky vzduchu. Únik ze zamořeného prostředí přináší další spotřebu energií a vyšší stupeň exhalací.

Coeckelbergh se neptá po etice této situace, bere ji jako fakt a zamýšlí se nad tím, zda nás umělá inteligence dokáže odtrhnout od problémů životního prostředí. Nejprve vylučuje, že by řešením mohlo být předání řízení našich životů na planetě Zemi umělé inteligenci, stejně jako spásu nepřinese exodus na jiné vesmírné těleso. Problematické je vůbec to, zda nějaké řešení stávající situace existuje. Tak trochu povinný optimismus ho nakonec vede ke zjištění, že cestou může být návrat k Zemi, k udržitelné umělé inteligenci.

Sousloví „udržitelná umělá inteligence“ není úplně samozřejmé. Ve všemožných deklaracích a agendách životního prostředí v posledních dekádách se udržitelnost objevuje v početných spojeních, stále častěji ale jako singulární udržitelnost; jako kdyby už nebylo třeba k ní nic dodávat, jako kdyby byla srozumitelná sama sebou. Spíš než o jasné náplni to svědčí o vyprázdněnosti tohoto termínu v důsledku jeho nadužívání. Coeckelbergh vymezuje udržitelnou umělou inteligenci způsobem, který není příliš šťastný: je to umělá inteligence, která umožňuje udržitelný způsob života lidí, přispívá k němu a neničí ekosystémy na Zemi. Těžko se ubránit pocitu, že se zde pohybujeme v tautologiích, jež nasvědčují tomu, že výraz udržitelná AI není dobrý nápad. Za takový lze naopak považovat autorovo závěrečné zjištění: při rozhodování o našich prioritách se nemůžeme obracet na umělou inteligenci, ale musíme rozvíjet praktickou moudrost, jež sice může využívat také abstraktní kognitivní procesy a analýzu dat, ale bude zásadně vycházet z vtělených zkušeností se vztahy a situacemi ve světě a s jinými lidmi, s kulturním a přírodním prostředím. AI se může rozvíjet jakkoli, lidé však musí pracovat na své praktické moudrosti; řečeno Coeckelberghovými slovy, „umělá inteligence dokáže rozpoznávat vzorce, ale moudrost strojům svěřit nemůžeme“ (s. 232).

Zdá se, že se zrodila nová aliance mezi etikou a AI; bude-li plodná tak, aby vytvořila adekvátní hodnotový systém odpovídající vrstvám a sférám vlivu umělé inteligence na kulturně-civilizační vztahy, se ještě ukáže. Obraz, který vykresluje ve své práci M. Coeckelbergh, lze brát jako relativně věcné a střízlivé vykročení

k problematice, která začíná již značně masivně vstupovat do mainstreamových sdělovacích prostředků. Vzhledem k citlivosti otázek, které AI bude do budoucna přinášet, je nezbytné, aby se veřejný prostor nezahltil senzačními informacemi, ale aby poskytoval dostatek odborně zdůvodněných a argumentačně vyzrálých záchytných bodů. Coeckelberghova kniha vyhovuje tomuto požadavku alespoň v tom smyslu, že nastiňuje směr dalších možných úvah a filosofických reflexí vztahu společnosti a AI.

Břetislav Horyna

Břetislav Horyna
Filozofický ústav SAV, v. v. i.
Klemensova 19
811 09 Bratislava
Slovenská republika
e-mail: filohory@savba.sk
ORCID-ID: https://orcid.org/0000-0002-6610-246X