

Žádný druh tu není navěky

S Tomášem Hříbkem o mysli a umělé inteligenci

Filosofa Tomáše Hříbka, který se věnuje filosofii vědy a bioetice, jsme se ptali, zda a jak lze současný vědecký a technologický pokrok využít k pochopení povahy mysli nebo jaké jsou hranice biologické a strojové inteligence. Mluvili jsme také o stvoření superinteligence a uploadu lidského vědomí do umělého systému.

MARTA MARTINOVÁ

Je nějaký zásadní rozdíl v tom, jak se filosofie k umělé inteligenci vztahovala v uplynulých osmdesáti letech a dnes? Co se týče počátků, máte asi na mysli abstraktní matematické úvahy Alana Turinga či Johnnyho von Neumanna, které se objevily už ve třicátých a čtyřicátých letech, ale explicitně se o umělé inteligenci začalo mluvit až v polovině let padesátých, čili zhruba před šedesáti roky. Tehdy se lidé z velmi různých oborů od logiky po kybernetiku začali zabývat otázkou, zda by počítače nezvládly stejné věci jako lidská mysl. Některé schopnosti – například uvažování, učení, řešení problémů nebo plánování – přitom označí za součást „inteligence“ každý, kdežto jiné – například vidění nebo motorickou kontrolu – málokdo. Ale všechno jsou to psychologické dovednosti, jejichž účelem je dosahování jistých cílů. Každopádně se ukázalo, že inteligence je mnoho-
rozměrný fenomén.

Byla filosofie dostatečně předvídatvá, nebo ji na současném vývoji technologií něco překvapuje?

Obory, které se věnují umělé inteligenci, mají technické i teoretické cíle. V těch prvních jde o tvorbu užitečných věcí, které nás dnes obklopují na každém kroku, od domácích spotřebičů přes bankovníctví až po internet. K teoretickým cílům patří využití prostředků umělé inteligence k pochopení povahy mysli. Tím se dostáváme k otázce vztahu mezi umělou inteligencí a filosofií: filosofové jako Hilary Putnam nebo Jerry Fodor už v šedesátých letech stáli u základů moderní matematické informatiky – computer science – a kognitivní vědy a formulovali nadlouho vlivný názor, že vztah mezi myslí a mozkem je podobný vztahu mezi softwarem a počítačem. Další filosofové, třeba u nás známý John Searle nebo Hubert Dreyfus, tento názor naopak kritizovali. Nevím, zda bych to nazval předvídatvostí, ale filosofie zkrátka byla u toho, když se diskuse o umělé inteligenci rozběhla. Ještě bych dodal, že filosofie mysli není jediná filosofická disciplína, která má co říci k umělé inteligenci. Jsou tu i etické otázky. Etika se dnes věnuje mimo jiné i tomu, jak zařídit, aby umělá inteligence byla morálně dobrá, jaké vůči ní máme případně závazky či jaká má práva – hovoří se o etice strojů, etice robotů a podobně.

Dá se říct, že se aktuální diskuse o budoucnosti umělé inteligence odehrávají mezi dvěma póly – umělou inteligencí coby prostředkem automatizace výroby a coby novou vědomou entitou?

Jde o podobné rozlišení, jako je to mezi technickými a teoretickými otázkami umělé inteligence. Co se týče prvního zmíněného pólu, svěříme-li výrobu umělé inteligenci, výsledkem zkrátka nebude pouze automatizovaná výroba. Automatizovaný systém je schopen provádět určitý úkol či soubor úkolů, kdežto inteligentní systém samostatně přijímá a realizuje rozhodnutí. Dalekosáhlým důsledkům takové změny ekonomiky se musí věnovat nejen ekonomové a sociologové, ale i filosofové. Teoretická otázka vztahu umělé inteligence a vědomí je odtazitější, protože k tomu, aby umělá inteligence řídila výrobu, vědomí asi nepotřebuje. Pro takové účely by vědomí mohlo být dokonce kontraproduktivní a zbytečně nákladné. Na druhé straně, pokud budou součástí ekonomiky i pečovatelské roboty, které se budou starat třeba o stárnoucí populaci, pak by u nich nějaké vědomí nebo alespoň emoční inteligence byly žádoucí.

Jak koncept umělé inteligence mění obsah a povahu toho, čím se zabývá filosofie mysli?

Jinými slovy, ptáte se, zda lidská mysl v souvislosti s umělou inteligencí nenaráží na nějaké nepřekročitelné hranice? Je zajímavé, že v diskusi o umělé inteligenci hrály jistou roli takzvané gödelovské argumenty, nazvané podle našeho brněnského rodáka Kurta Gödela, jednoho z největších logiků všech dob a zakladatele matematické informatiky. Gödel a jeho pokračovatel britský matematik Roger Penrose tvrdili, že lidská mysl má komputační schopnosti, které Turingovým strojům, jež jsou abstraktním základem moderních počítačů, z principu chybějí. Gödel a jeho následovníci jsou sice v menšině, ale pokud by měli pravdu, pak by spíš než lidská inteligence byla ohraničená ta umělá. Patrně více stoupenců má ovšem v současnosti názor, že dříve či později sestrojíme



Tomáš Hříbek. Foto Honza Šípek

umělou inteligenci, která pak stvoří superinteligenci, jež ve všech ohledech překoná tu lidskou. Filosof to může považovat za poslední fázi vyvrácení antropocentrismu – představy, že člověk je pupek světa.

Snaha vytvořit umělou inteligenci, která by měla i něco na způsob lidského vědomí či sebevědomí, je někdy považována za projev lidské pýchy a arogance: člověk se chce přiblížit Bohu tím, že stvoří inteligentní bytost, podobně jako Bůh údajně stvořil člověka. Má k tomu co říct filosofie?

Tyto populární úvahy ve filosofické literatuře nehrají žádnou roli, nebo alespoň ne v té, která je mi známa. Zmíněná kritika předpokládá oprávněnost teologické představy, že jisté činnosti jsou člověku zapovězeny, protože je má v režii stvořitel. Pokud ale není žádný dobrý důvod si něco podobného myslet, pak tato úvaha odpadá. Kromě toho se mi zdá, že zmíněný populární typ kritiky si představuje umělou inteligenci jako Terminátora nebo Frankensteinovo monstrum, ačkoli mnohem spíš jde prostě o software.

Ještě poznámka k tomu „vědomí a sebevědomí“: hovoří-li filosofové o vědomí, mají na mysli takzvané fenomenální vědomí, jež se obvykle chápe tak, že jde o vlastnosti zkušenosti, které lze stěžejně popsat a vysvětlit jazykem objektivní vědy. Některé filosofové, třeba David Chalmers, se dokonce domnívají, že je to nemožné. Naopak sebevědomí předpokládá schopnost reflexe, jejíž podmínkou je pojem vlastního já. Koně nebo novorozenci takový pojem nemají, a tudíž nemají ani sebevědomí, ale asi bychom jim neupřeli vědomí. Podobně musíme rozlišovat pojmy jako

inteligence, vědomí a sebevědomí i v případě umělých systémů. „Hostitelé“ v seriálu Westworld možná dosáhli sebevědomí. Ale mají fenomenální vědomí?

Může být umělá inteligence dalším stupněm evoluce?

Platí-li teorie evoluce, pak tu žádný druh není navěky, a to se týká i homo sapiens. A je možné, že dalším krokem evoluce, v tomto případě poprvé záměrným, bude umělá inteligence. Některé filosofové a technologičtí vizionáři dokonce předpokládají, že pokud existují nějaké mnohem starší mimozemské civilizace, už dávno překonaly biologickou fázi vývoje, a v podstatě se tedy jedná o civilizace inteligentních strojů. Otázkou je, zda původně biologická inteligence v těchto

cizích světech zanikla a byla nahrazena umělou inteligencí, nebo zda biologická inteligence s tou umělou časem splynula. Pokud totiž vývoj umělé inteligence dříve či později povede k vytvoření superinteligence v tom smyslu, jak o tom byla řeč, je otázka, jaký vztah bude taková inteligence mít k lidem. Bude vůči nim lhostejná, nepřátelská, anebo s nimi bude spolupracovat? Momentálně je pro nás větším existenčním rizikem klimatická změna, ale za pár desetiletí to může být superinteligence, a v tom případě by mohlo být řešením splynutí – upload lidského vědomí do sítě a podobně. Tomuto tématu se mimochodem věnuje filosof Nick Bostrom v knize Superinteligence, která předloni vyšla i česky.

Tomáš Hříbek (nar. 1965) vystudoval filosofii na FF UK, na Bowling Green State University a na University of Minnesota a v současnosti působí na Filosofickém ústavu Akademie věd, kde se zabývá zejména filosofií mysli a jazyka, filosofií vědy a etikou. Je autorem monografií Metafyzika antiindividualismu (2008) a Jaké to je, nebo o čem to je? Místo vědomí v materiálním světě (2017). Spoluzaložil Centrum Karla Čapka pro studium hodnot ve vědě a technice (cevast.org).