

| /

ZRCADLO, ZRCADLO,
POVĚZ, KDO JE NEJCHYTŘEJŠÍ

Humbuk a obavy z umělé inteligence: Zrcadlo, zrcadlo, pověz, kdo je z nás všech nejchytřejší

Leemu Sedolovi se při vyhlášení výsledků zalesknou oči slzami. AlphaGo, umělá inteligence vyvinutá firmou DeepMind společnosti Google, ho právě porazila ve hře Go 4 : 1. Je březen 2016. Před dvaceti lety prohrál šachový velmistr Garry Kasparov se strojem Deep Blue a dnes počítačový program zvítězil nad osmnáctinásobným mistrem světa Leem Sedolem ve složité hře, u které se předpokládalo, že ji s využitím intuice a strategického myšlení může hrát jen člověk. Počítač nevyhrál proto, že se řídil pravidly, jež mu zadali programátoři, ale díky strojovému učení založenému na milionech minulých partií i díky tomu, že hrál sám proti sobě. Programátoři sice v podobném případě připravují shluky dat (datasety) a vytvářejí algoritmy, nemohou však vědět, jaké tahy program vymy-

slí. Umělá inteligence se učí sama. Po řadě jejích neobvyklých a překvapivých tahů musel Lee odstoupit (Borowiec 2016).

Umělá inteligence tak dosáhla úctyhodného úspěchu. Vzbuzuje však také obavy. Krása tahů vyvolala nejen obdiv, ale i smutek, a dokonce strach. Máme sice naději, že ještě chytřejší umělá inteligence nám umožní revolučně změnit zdravotní péči nebo najít řešení nejrůznějších společenských problémů, ale objevily se i obavy, že stroje převzou moc. Mohly by nás přechytračit a ovládnout? Je umělá inteligence stále pouhý nástroj, nebo se pomalu ale jistě stává naším pánem? Tyto obavy připomínají slova počítače HAL s umělou inteligencí ve vědecko-fantastickém filmu Stanleyho Kubricka 2001: *Vesmírná odysea*, když na lidský příkaz „Otevři dveře hangáru“ odpoví: „Obávám se, že to nemohu, Dave.“ A pokud nemáme obavy, můžeme cítit smutek či zklamání. Přesvědčení o naší výjimečnosti, náš pocit nadřazenosti a naše představy o moci svrhly z trůnu Darwin a Freud a dnes se zdá, že další ránu lidskému sebepjetí zasadí umělá inteligence. Pokud tohle dokáže stroj, co zbude nám? Co jsme my? Jen stroje? *Podřadné* stroje s příliš mnoha chybami? Co z nás bude? Stanou se z nás otroci strojů? Nebo ještě hůř, pouhé zdroje energie jako ve filmu *Matrix*?

Skutečný a všudypřítomný dopad umělé inteligence

Průlomové pokroky umělé inteligence se však neomezují jen na hry nebo na říše sci-fi. Umělá inteligence se objevuje již nyní a je všudypřítomná, často neviditelně zabudovaná do

našich denně používaných nástrojů i jako součást složitých technologických systémů (Boddington 2017). Vzhledem k exponenciálnímu růstu výkonu počítačů, dostupnosti (velkých) dat díky sociálním sítím a masivnímu využívání miliard chytrých telefonů a rychlých mobilních sítí dosáhla umělá inteligence, a zejména strojové učení, významného pokroku. Algoritmy tak mohly převzít řadu našich činností včetně plánování, řeči, rozpoznávání obličejů a rozhodování. Umělá inteligence má uplatnění v mnoha oblastech včetně dopravy, marketingu, zdravotnictví, financí a pojišťovnictví, bezpečnosti a armády, vědy, vzdělávání, kancelářské práce a osobní asistence (například Google Duplex),¹ zábavy, umění (mezi jiným při vyhledávání a skládání hudby), zemědělství a samozřejmě výroby.

Umělou inteligenci vytvářejí a využívají počítačové a internetové společnosti. Třeba společnost Google umělou inteligenci vždy využívala pro svůj vyhledávač. Facebook ji využívá pro cílenou reklamu a označování fotografií, společnosti Microsoft a Apple pro své digitální asistenty. Uplatnění umělé inteligence je však širší než jen v úzce vymezeném odvětví informačních technologií. Existuje například řada konkrétních plánů a experimentů s autonomními vozidly. I tato technologie je založena na umělé inteligenci. Využívají ji drony i autonomní zbraně, jež mohou zabít bez zásahu člověka. A umělou inteligenci již při svém rozhodování využily i soudy. Uvedme,

¹ Viz <https://www.youtube.com/watch?v=D5VN56jQMWM>.

Umělá inteligence se objevuje již dnes a je všudypřítomná, často neviditelně zabudovaná do našich denně používaných nástrojů.

že ve Spojených státech se používá systém COMPAS, který předvídá, kdo pravděpodobně znovu spáchá trestný čin. Umělá inteligence vstupuje i do oblastí, které obecně považujeme za osobnější a intimnější. Stroje dnes například rozpoznají naše obličeje: dokáží nás nejen identifikovat, ale umějí také číst naše emoce a získávat nejrůznější informace.

Potřeba diskutovat o etických a společenských problémech

Umělá inteligence může mít celou řadu výhod. Lze ji využít ke zlepšení veřejných i komerčních služeb. Rozpoznávání obrazu je dobrou zprávou pro lékařství: může pomoci s diagnostikou nemocí, jako je rakovina nebo Alzheimerova choroba. Podobné každodenní využití umělé inteligence však také ukazuje, že nové technologie vyvolávají i obavy v souvislosti s etikou. Uvedu několik příkladů otázek z oblasti etiky umělé inteligence.

Měla by být v autonomních vozidlech zabudována etická omezení, a jestliže ano, jaká a jak by měla být stanovena? Pokud se například autonomní vůz dostane do situace, kdy si musí vybrat mezi nárazem do dítěte a nárazem do zdi, aby život dítěte zachránil, ale potenciálně zabil spolujezdce, co by si měl vybrat? A měly by se vůbec povolovat autonomní smrtící zbraně? Kolik rozhodnutí a jak velkou jejich část chceme umělé inteligenci svěřit? Kdo ponese odpovědnost, když se něco pokazí? V jednom případě soudci věřili víc algoritmu COMPAS než dohodám, kterých dosáhla obhajoba a obža-

loba.² Nebudeme se na umělou inteligenci spoléhat příliš? Algoritmus COMPAS je velmi kontroverzní i proto, že podle výzkumu byly falešně pozitivními případy (osobami, u nichž předpověděl spáchání dalšího trestného činu, které ho však nespáchaly) v nepoměrně větší míře černošští obyvatelé (Fry 2018). Umělá inteligence tak může posilovat předsudky a nespravedlivou diskriminaci. Podobné problémy mohou nastat u algoritmů, jež doporučují rozhodování o žádostech o hypotéku a zaměstnání. Nebo vezměme v úvahu takzvaný prediktivní policejní dohled: algoritmy se používají k předpovídání toho, kde pravděpodobně dojde k trestným činům (například v které části města) a kdo by je mohl spáchat, výsledkem však může být, že policejní dohled se nepřiměřeně zaměří na určité sociálně-ekonomické skupiny nebo skupiny určitého rasového původu. Prediktivní policejní dohled se již používá ve Spojených státech a podle nedávné zprávy organizace AlgorithmWatch (2019) také v Evropě.³ Ke sledování se pak často využívá technologie rozpoznávání obličejů založená na umělé inteligenci, která může narušovat soukromí lidí. Může také víceméně předvídat sexuální preference. K tomu nepotřebuje

² Viz případ Paula Zillyho, jak ho líčí Fry (2018, 71–72). Další podrobnosti Angwin, J., Larson, J., Mattu, S., Kirchner, L., Machine Bias, *ProPublica*, 23. 5. 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

³ Například v roce 2016 začala místní policejní zóna v Belgii používat prediktivní policejní software k předvídání vloupání a krádeží vozidel. (AlgorithmWatch 2019, 44)

žádné informace z telefonu ani biometrické údaje. Stroj svou práci vykonává na dálku. Díky kamerám na ulicích a dalších veřejných prostranstvích nás může identifikovat a celé „přečíst“ včetně naší nálady. Pomocí analýzy našich údajů lze předvídat naše duševní a fyzické zdraví – aniž bychom o tom věděli. Zaměstnavatelé mohou tuto technologii využívat ke sledování naší výkonnosti. Algoritmy, které jsou aktivní na sociálních sítích, mohou také šířit nenávistné projevy či nepravdivé informace; třeba političtí boti se mohou tvářit jako skuteční lidé a zveřejňovat politický obsah. Známým případem je chatbot společnosti Microsoft jménem Tay z roku 2016, který byl navržen pro rozpustilé konverzace na Twitteru, ovšem jakmile se poučil, začal tweetovat rasistické věty. Některé algoritmy umělé inteligence mohou dokonce vytvářet falešné videozáznamy projevů – například video vytvořené tak, aby klamně připomínalo projev Baracka Obamy.⁴

Úmysly jsou často dobré. Tyto etické problémy jsou však obvykle nezamýšlenými důsledky technologií: většinu z nich, jako je třeba předpojatost či nenávistné projevy, vývojáři ani uživatelé technologie nezamýšleli. Navíc je třeba si vždycky položit jednu zásadní otázku: Komu bude zlepšení sloužit? Vládě, nebo občanům? Policii, nebo těm, kteří jsou v jejím hledáčku? Prodejcům, nebo zákazníkům? Soudcům, nebo ob-

⁴ BuzzFeedVideo, You Won't Believe What Obama Says in This Video! (Neuvěříte, co Obama říká v tomhle videu!) https://www.youtube.com/watch?v=cQ54GDm1eL0&fbclid=IwAR1oD0AlopEza00XH03WNcey_qNn-NqTsvHN_aZsNb0d2t9cmsDbm9oCfX8A.

viněným? Do hry vstupují otázky týkající se moci, kupříkladu když technologii vyvíjí jen několik megakorporací (Nemitz 2018). Kdo utváří budoucnost umělé inteligence?

Tato otázka poukazuje na společenský a politický význam umělé inteligence. Etika umělé inteligence se zabývá nejen technologickými změnami a jejich dopadem na životy jednotlivců, ale i transformací společnosti a ekonomiky. O společenském významu umělé inteligence svědčí již otázky předpojatosti a diskriminace. Umělá inteligence však mění také ekonomiku, a tím zřejmě i sociální strukturu naší společnosti. Podle Brynjolfssona a McAfeeho (2014) jsme vstoupili do druhého věku strojů, v němž stroje člověka nejen doplňují, jako tomu bylo v průmyslové revoluci, ale také jej nahrazují. Vzhledem k tomu, že umělá inteligence ovlivní povolání a práce všeho druhu, předpokládá se, že se naše společnost dramaticky změní, protože do reálného světa vstoupí technologie kdysi popisované ve sci-fi (McAfee, Brynjolfsson 2017). Jaká je budoucnost práce? Jaký život budeme mít my, až umělá inteligence obsadí pracovní místa? A kdo je to „my“? Kdo z této transformace získá a kdo bude trpět?

Tato kniha

Na základě velkolepých průlomů je kolem umělé inteligence velký humbuk. Umělá inteligence se přitom využívá v celé řadě znalostních oblastí a lidských postupů již dnes. V tom prvním případě její využití vedlo k divokým spekulacím o technologické budoucnosti a zajímavým filosofickým diskusím o tom,

Etika umělé inteligence se zabývá nejen technologickými změnami a jejich dopadem na životy jednotlivců, ale i transformací společnosti a ekonomiky.

co znamená být člověkem. V tom druhém případě vyvolalo pocit, že etici a tvůrci strategií musejí naléhavě zajistit, aby nám tato technologie byla ku prospěchu a jednotlivcům a společnosti nepřinášela nepřekonatelné problémy. Tyto obavy více souvisejí s praxí a jsou naléhavější.

Tato kniha, napsaná akademickým filosofem, který má zkušenosti i s poradenstvím pro tvůrce strategií, se zabývá oběma aspekty: etiku pojímá jako téma související se všemi těmito otázkami. Jejím cílem je poskytnout čtenáři dobrý přehled o etických problémech spojených s umělou inteligencí, kterou chápe v širokém slova smyslu od vlivných narativů o budoucnosti umělé inteligence a filosofických otázek o povaze a budoucnosti člověka až po etické obavy týkající se odpovědnosti a předpojatosti i toho, jak se s praktickými problémy, jež tato technologie v reálném světě způsobuje, vypořádat prostřednictvím strategií – pokud možno dříve, než bude pozdě. Co se stane, až bude „příliš pozdě“? Některé scénáře jsou dystopické a utopické zároveň. Začnu několika sny a nočními můrami o technologické budoucnosti, vlivnými narativy, jež se alespoň *na první pohled* pro hodnocení potenciálních přínosů a nebezpečí umělé inteligence zdají být relevantní.