

# Posudek oponenta bakalářské práce

Název práce: *AI's Impact on the Environment and Anthropocene*

Autor práce: Matyáš Mandík

Vedoucí práce: Juraj Hvorecký, Ph.D.

FHS UK 2024

Předložená práce se věnuje tématu, jež samo o sobě je nesporně jak aktuální, tak smysluplné a jež zároveň, v přehledové a základní verzi, může být uchopeno v bakalářské práci.

Autor práci člení do pěti kapitol a úvodu, v nichž postupně představuje téma umělé inteligence, současné enviromentální krize, fyzických nákladů spojených s provozem a produkcí technologií, jež jsou nezbytné pro fungování AI. Ve čtvrté kapitole blíže komentuje dopady AI na různá odvětví a životy lidí, případně nastiňuje etické problémy s tím spojené. V závěrečné části autor formuluje stanovisko k problému AI z hlediska enviromentálních dopadů.

## Obecné připomínky k práci:

Tím, co práci v první řadě musím vytknout, je její nejasná koncepce. Práce se spíše věnuje tématu dopadu AI na životní prostředí, než že by formulovala a dále argumentačně rozvíjela či dokládala tezi, respektive si pokládala jasnou problémovou otázku. Toto je především patrné v úvodu, který jednoduše neřekne, jakou tezi práce hájí, jakou otázku si, konkrétně řečeno, pokládá.

Tato nekoncepčnost vrcholí v závěru, kde jsem v páté kapitole naprosto nečekal, že práce najednou skončí. Její etický imperativ, který ještě níže komentuji, je možná nějakým výsledkem, ale je argumentačně velmi slabý a prozrazuje v zásadě kruhovou argumentační strukturu (viz níže).

Je jistě normální, že se práce od zadání tématu vyvíjejí, ale bylo by dobře si alespoň povšimnout, že v názvu se mluví o antropocénu, což, jako terminus technicus nelze úplně odbýt tím, že nějak jde o lidi a jejich život na této planetě. V práci se o antropocénu neřekne ani slovo.

Čili obecně je problémem, že práce neví, co sděluje a neumí proto ani vhodně vystavět celou strukturu a argumentační kroky.

Obecně dále práce volí literaturu, která je příliš specifická. Co jí naopak chybí, jsou obecnější – a zvláště filosofické – teoretické studie, jež by se věnovaly problému technologií či vztahu člověka a techniky. Z těch, které poměrně dobře znám, bych uvedl alespoň práce Jüngera, Anderse, Lazzarata, či Heideggera.

Ještě více překvapující pro mě ale je, že práce nijak nebere v potaz, že její způsob tázání patří do okruhu transhumanismu (případně, z určitého hlediska, posthumanismu).

## Konkrétní kritické poznámky:

Práce zmiňuje tradici umělých lidí, respektive automatonů, ale pokud už jde o pokus o tuto historii, pak je to historie extrémně zjednodušující, která si nevšímá ani notoricky známých literárních prekursorů, jako je Hoffmannova povídka Der Sandmann či podle ní adaptovaná opera Coppélia, golemů, robotů, atd. Lepší by bylo tuto pasáž vůbec neuvádět. Především je totiž otázka, proč se uvádí historie "umělých lidí", když jde o umělou inteligenci?

Podobně zvláště výběrová je historická perspektiva teorie umělé inteligence. Neříkám, že nutně každá musí začínat, nebo alespoň zmiňovat, Touringa, ale nechápu, proč tato část není přinejmenším trochu systematická a skáče mezi současností a Dartmouthskou konferencí.

O poznání větší problém mám s tím, že práce vůbec nijak necharakterizuje alespoň základní vymezení umělé inteligence (např. ani nezmíní strong AI), základní teoretické koncepce, atp.

Na s. 7 autor sumarizuje dosavadní výklad a píše, že pro první období je typický "spirit of semantic artificial intelligence", ovšem oč běží není jasné, protože tato představa (o semantic artificial intelligence) se objevuje až zde.

Přístup práce, jak je prezentován v závěru první kapitoly, by – kromě odkazu k tradici vytváření automatonů – bylo možné napsat bez celé této kapitoly. Kapitola k takto vymezenému záměru práce nic nepřidává, nijak tento přístup nezakládá.

Kapitola má však hlavně nějak uchopit problém toho, co je umělá inteligence, jak jí autor rozumí. Určitě oceňuji ten rozměr, na který se soustředí, sice fyzickou infrastrukturu umělé inteligence, ale zároveň příliš neoceňuji, že na svou vlastní otázku autor nijak neodpovídá.

Konečně si první kapitola uchovává stejnou tajemnost, jako úvod. Práce má být "useful inspiration", být jaksi mezi utopickými a dystopickými vizemi. Ale co to znamená?

Interpretace knihy Hannah Arendt jde přímo proti záměru, který autorka vyjadřuje. Arendt si pouze pokládá otázku, co by opuštění Země znamenalo z hlediska lidství člověka a jeho podmíněnosti. Druhá kapitola je jinak celkově přehlednější a jasnější.

Nerozumím tomu, v jakém smyslu tato – navíc deklaratorně filosofická – práce chápe určení metafyziky: "In the interaction between a human and an artificial intelligence system, we also enter a peculiar metaphysical space, but a rather distorted one." (15) Podobně na s. 18: proč by opakem fyzického mělo být "metafyzické" a ne třeba virtuální?

Jinak však třetí kapitola podává velice zdařilý přehled environmentální náročnosti vytváření hardware nezbytného pro AI. Mojí jedinou otázkou je, čím se zde důraz na AI liší od nároků na fungování IT obecně? Představuje v tomto smyslu umělá inteligence větší zátěž? Je něčím odlišná?

Obecněji: jaké je vůbec myslitelné měřítko posuzování dopadu jakékoliv oblasti lidské činnosti na životní prostředí? Lze tyto činnosti srovnávat?

Práce tedy na jedné straně velmi pěkně upozorňuje na to, že AI není jakýsi elektronický přelud, nýbrž výsledek operace fyzických zařízení, která vyžadují investice materiálu, práce, atp. Ale co autor nedělá, je zařazení do kontextu, vytvoření alespoň nějaké perspektivy posouzení tohoto environmentálního vlivu AI.

Příliš nerozumím argumentu, že problémem s machine learning (s. 19) je potřeba obrovského datasetu. Znamená to snad, že tato data vznikají záměrně za účelem učení AI? Tato data přeci vznikají tak jako tak. Co vzniká specificky kvůli AI, jsou *některá* datová úložiště, která tato data shromažďují.

Jakkoliv jsou dnešní lodě obrovské, nemají kapacitu 250 milionů tun nákladu (s. 20). Největší z nich uveze okolo 400 tisíc tun.

Když už se (na s. 24) mají srovnávat spotřeby elektřiny v různých regionech, bylo by dobře uvádět jednu jednotku. Sice si dokážu spočítat, že 1 TWh je 1 miliarda kWh, ale jednodušší by bylo prostě uvádět vše v TWh nebo kWh. Edit: vidím, že je převod v poznámce pod čarou, což ovšem jen potvrzuje moji poznámku. Proč by mělo dávat smysl uvádět převod v poznámce pod čarou, když je jednodušší pracovat s jedinou jednotkou? Věcně se však musím zeptat: co tato čísla říkají? Za prvé: čísla se týkají hlavně roku 2012, kdy spotřeba elektřiny v datacentrech činí globálně okolo 250 TWh. Jak se tato spotřeba týká umělé inteligence? Celosvětově byla spotřeba elektřiny přibližně 20 000 TWh, čili by šlo o 1,2% vší spotřeby elektřiny. Není to – naopak – vlastně málo z hlediska tak všudypřítomné a zásadní položky lidského života současnosti?

Práce si zcela správně všímá etických problémů s autorskými právy. Ovšem jak souvisí tento problém s enviromentálními dopady?

V práci se objeví teze o poměrně malém etickém dohledu nad výzkumem na poli AI. Zde je třeba odlišit prostředí v USA a Číně od EU. V EU podléhá jakýkoliv výzkum (zvláště ten financovaný přímo z programu Horizon) AI přísným etickým pravidlům.

Práce na s. 33 konstatuje: "Artificial intelligence, as it is now, cannot amount to anything but a negative influence." Smyslem je, a s tím mohu souhlasit, formulovat etický imperativ vědomého přijetí nákladů na umělou inteligenci. Tyto by jistě neměly být ignorovány. Ale práce, která se charakterizuje jako filosofická, jednoduše nemůže zůstat u takto hrubého a nejasného stanoviska.

Důvodem, proč autor může formulovat takto negativní stanovisko, je především ten, že se v práci naprosto nevěnuje pozitivním aspektům umělé inteligence z hlediska problému životního prostředí. Já sám nejsem odborníkem na tento specifický problém, ale zcela elementární hledání přes Google ukáže, že je celá řada oblastí, kde AI slouží k řešení enviromentálních problémů. Jsou to třeba modelace klimatických procesů, vývoj obnovitelných zdrojů, nové metody v zemědělství, sledování změn ekologických systémů, predikce přírodních katastrof, či třeba vývoj na poli zpětného zachycování uhlíku v atmosféře.

Jinak řečeno: práce neprezentuje takovou perspektivu, ze které by bylo možné posoudit, zda a v jakém smyslu jsou ony zdroje vynaložené na umělou inteligenci ospravedlnitelné enviromentálním dopadem. Podobnou otázku si můžeme klást u odvětví, jako je těžba kryptoměn (kde je zjevné, že jde o eticky takřka neudržitelnou situaci), nebo třeba obranný průmysl (kde je etické posouzení daleko složitější). Pokud bychom pokládali otázky pouze tak, jak činí autor práce, byla by stejně smysluplná i otázka: můžeme eticky ospravedlnit provoz nemocnic, pokud v USA v roce 2012 spotřebovaly 210 TWh, čili třikrát tolik, co datacentra? (<https://www.mdpi.com/1996-1073/12/19/3775>)

Neexistuje žádná oblast lidského života, která by nakonec neměla dopad na životní prostředí. Ale etické hodnocení těchto oblastí nemůže být postaveno pouze na problému brutto pojatých energetických, materiálových, logistických a jiných nákladů. Pak bychom se mohli rovnou vzdát i knih.

## **Celkové hodnocení:**

Navzdory příkrým komentářům jednotlivých kroků a celkové koncepci práce si nemyslím, že by byla tato mimořádně nezdařilá. Její zcela jednoznačně silnou stránkou je autorův dobře míněný teoretický zájem o problém AI a životního prostředí, kdy tomuto problému věnoval zjevně mnoho úsilí a nastudoval řadu odborných titulů. Uplatnění těchto „dat“ však neproběhlo zcela zdařile. Práce by potřebovala pořádné promyšlení a ujednocení.

Domnívám se však, že práci **možu doporučit k obhajobě**, byť nemohu navrhnout lepší stupeň hodnocení, než **dobře**.

V Praze 5. září 2024

Jakub Marek, Ph.D.