

POSUDEK OPONENTA PÍSEMNÉ KVALIFIKAČNÍ PRÁCE

Název práce: KRITICKÁ REFLEXE GENERATIVNÍ UMĚLÉ INTELIGENCE
Autor/ka práce: Matěj Martinec
Studijní program: FAMU CAS
Typ studijního programu: [bakalářský]

Vymezení cíle a jeho naplnění:

Cíl je srozumitelný, ale v úvodu by se měl zakládat na skutečných člancích (které pak v textu jsou) a ne na „dopoledním čtení“. Cíle byly naplněny částečně.

Aktuálnost tématu (a relevance zvolené metodologie v případě diplomové práce):

Aktuálnost je vysoká.

Odborný přínos, původnost práce a její případné využití v praxi:

Autor nepřesně ocitoval knihu Objekt, jejímž autorem je Václav Jánoščík.

Formální úprava a náležitosti práce včetně jejího rozsahu:

Rozsah je těsně za doporučenou délkou textu.

Jazyková, stylistická a terminologická úroveň, Práce s informačními zdroji, Logická stavba a členění práce:

Stylistická úroveň je nedostačující. Autor používá větné konstrukce typu „Dnes dopoledne jsem se dočetl“, „V dalším článku jsem si přečetl“ či „Nakonec ale zjistím“ bez uvedení citace článku a s bezdůvodným časovým údajem kdy onen článek četl. Toto je obzvláště zářející, jelikož je to hned v úvodu práce a poté autor píše, že se tato tvrzení pokusí vyvrátit, aniž bychom věděli odkud jsou.

Autor vynechává části tezí. Citace Brynjolfssona vyznívá opačně, než ji Brynjolfsson zamýšlí. Martinec uvádí na základě četby Brynjolfssona, že „stroje nejsou vytvářeny proto, aby rozšiřovaly schopnosti lidí, ale aby schopnosti lidí automatizovaly a nahrazovaly, což vyřazuje lidi z procesu vytváření hodnoty“. Tím by se dalo usuzovat, že si Brynjolfsson myslí, že veškerá AI má tento problém. Ten ale ve svých člancích dodává, že „Stroje jsou vyráběny k tomu, aby augmentovali naše schopnosti, a tím produkovali daleko větší hodnotu, než snaha o vytváření AI, která by člověka nahradila zcela.“ Myšlenou Turingovou pastí je tedy pouze ta AI, která se snaží o úplné nahrazení člověka, ne veškerá AI. Navíc se nejedná pouze o nahrazení schopností lidí jak uvádí Martinec, ale celého lidského myšlení, tedy alespoň ve smyslu překonání Turingových zákonů.

Dále se dopouští nepodložených závěrů – např. u služby Amazon Mechanical Turk tvrdí, že se jedná o „podvod“. Martinec tvrdí, že „Uživatelům se může zdát, že aplikace využívá k plnění úkolů

pokročilou umělou inteligenci“, když přímo v popisu služby je napsáno, že služba nabízí schopnosti reálných lidí. Služba se běžně používá k výzkumu právě v případech, kdy potřebujeme zaručit, že za počítačem sedí člověk a tedy minimálně tito uživatelé si rozhodně nemyslí, že by se jednalo o nějakou umělou inteligenci či podvod. Ačkoliv je relevance námezdní práce v souvislosti např. s případy použití lidského faktoru ke kategorizaci dat do datasetů relevantní, nemá tuto souvislost se službou Mechanical Turk. Zmínka o této službě tak může souviset s tím, že právě tito námezdní pracovníci jsou často využíváni pro výzkum umělé inteligence a často si pomáhají tím, že použijí umělou inteligenci k generování problémů, které mají zodpovědět jako lidé. To ale autor již nezmiňuje.

Asi největší nedostatky pak spatřuji v Martincově tvrzení, že „...v 19.století britští eugenici vyvinuli metody korelace a lineární regrese, datovou analýzu ...“ Britští eugenici rozhodně nevyvinuli datovou analýzu. Datová analýza je obecný pojem a provádí se již staletí, ne-li tisíciletí. Dále pak spatřuji jako problematické Martincovo spojení lineární regrese a korelace a eugeniky. Autor uvádí že “korelace položila základy pro představu univerzálních zákonů, pro eugeniku tak důležitých“. Problém vidím v tom, co již autor dále neuvádí. Lineární regrese je legitimní statický nástroj a ten jako takový nezapřičiňuje to že se při jeho používání dopustíme eugeniky. Autoři tohoto nástroje nejdříve tuto metodu vyvinuli při množení hrášku. To že z toho Mengele udělá závěry na celé lidstvo již pak není problém metody.

Dále pak autor tvrdí, že „lineární diskriminační analýza, vyvinutá na počátku dvacátého století za účelem rozlišování mezi třídami a rasami lidí.“ Metoda byla poprvé použita při analýze květů kosatce a dodnes slouží jako spolehlivý nástroj s širokým využitím ve vědě a výzkumu. Martinec se pak pouští do závěrů že „Strojové učení je podmíněné statistickými postupy eugeniků z 19. století, ..., neboť celá její podstata je vybudována na myšlenkách diskriminace a segregace.“ Zde je nutné poznamenat, že uvedené metody skutečně byly zneužity pro myšlenky eugeniků, to ale bylo již kritizováno samotnou akademickou obcí, která již tehdy kritizovala to, že jsou závěry použité ve prospěch eugeniky nevědecké nebo hrubě zjednodušené. Ostatně tato zaujatost je problém, který je pojmenován a aktivně se řeší. V žádném případě se ale nedá tvrdit že „celá podstata“ strojového učení stojí na těchto slepých uličkách vědy. Ačkoliv jednoznačně mají současné formy umělé inteligence v sobě zaujatost, nedá se tvrdit, že za to zcela můžou nástroje, které k tomu byly použity. Domnívám se tedy, že zde tkví problém v nepochopení statistických nástrojů. Jako každý nástroj mají své využití a pouze skutečný odborník si může dovolit tvrdit, kde se tyto nástroje hodí a kde ne. Pokud tedy budeme dělat zjednodušené závěry ze složitých problémů, a ještě za použití nástrojů, kterým nerozumíme, skutečně můžeme dopadnout špatně. Mnoho vědců tak používá nástroje, kterým nerozumí, či nejsou v jejich používání odborníky a počítačové vědy nejsou výjimkou. Příkladem může být prováděná statistika pomocí zmíněné lineární regrese nad náhodnými daty, která bude mít vždy nějaký trend. Relevantní otázkou by pak bylo, komu může sloužit tato špatná praxe. Svět je zkrátka více náhodný než regresivní analýza

shrnutí hodnotitele:

Jsem si vědom že na katedře CAS je kladen větší důraz na teoretické práce než na jiných uměleckých školách. Jsem rád že autor používá citace a je schopen pojmenovat aktuální problém. Nejsem ale spokojen s jeho závěry, které jsou mnohdy nepodložené, zavádějící či nepravdivé. Doporučil bych přidat do citovaných autorů a literatury více lidí s technickým vzděláním a nedělat unáhlené závěry. Vzhledem k náročnosti tématu a zmíněným zvýšeným požadavkům na tyto teoretické práce si po diskuzi dovedu představit zlepšení závěrečné známky.

Otázky a náměty k diskuzi při obhajobě:

Co je to datová analýza a jaká je její historie?

Jaké jsou v současnosti nejpravděpodobnější důvody různé předpojatosti současných AI modelů?

Komu slouží špatná vědecká praxe?

Doporučení práce k obhajobě: ANO

Navrhovaná klasifikace: D- uspokojivě (přijatelný výkon, ale se značnými nedostatky)

Vojtěch Radakulan, 24.6.2024