

Magdaléna Ostárková

Gymnázium Olgy Havlové, Ostrava–Poruba

Téma č. 1: Umělá inteligence a lidská práva

Jak zajistit, aby umělá inteligence rozhodovala spravedlivě?

V této eseji se zaměřím na problematiku umělé inteligence v souvislosti s lidskými právy. Přestože se má umělá inteligence rozhodovat bez předsudků, dnes je čím dál tím více jasné, že tomu tak není. Ale jak proti tomu bojovat?

V dnešní moderní době žijeme ve světě technologií, které nám mají pomáhat s každodenními činnostmi, zlepšovat kvalitu našeho života, zrychlit a zjednodušit práci. Vývoj umělé inteligence vzbudil v mnoha lidech z dnešního pohledu již naivní očekávání, že pomůže lidstvu zbavit se negativ a špatností, které v sobě lidské plémě má zakořeněné od nepaměti. Postupem času se ale začalo ukazovat, že stejné chyby, které dělá člověk, dělá také umělá inteligence, která je člověkem naprogramovaná. V počítačovém odvětví tomu vědci říkají: „garbage in, garbage out,”¹ v překladu: smetí dovnitř, smetí ven. To má znamenat, že to, co člověk vloží do algoritmu, z něj také vyjde – tudíž zakódují-li se do algoritmu například prvky rasismu, bude umělá inteligence ovládaná algoritmem také rasistická.

Rasismus je samozřejmě nepřijatelný. Základní listina lidských práv říká, že příslušnost ke kterékoli národnostní nebo etnické menšině nesmí být nikomu na újmu. Podle mě to ale umělá inteligence porušuje. Lze to dokázat na příkladu Brishy Bordenové z Floridy v USA. Jedná se o Afroameričanku, která v roce 2014 ukradla dětské kolo. Byla policií zatčena a obviněna z drobné krádeže a vloupání. Rok předtím byl policií zadržen Vernon Prater za to, že ukradl zboží z nedalekého obchodu v přibližně stejné hodnotě jako Brisha. Jednalo se o muže bílé pleti. Zarážející je na tom ale fakt, že když byli přijati do věznice, počítačový algoritmus vyhodnotil Brishu jako člověka s vysokým rizikem recidivy, nicméně Vernona jako člověka s rizikem nízkým a to jen z důvodu jejich barvy pleti. O dva roky později se ukázalo, že Brisha již žádný podobný čin nespáchala, za to Vernon byl opět ve vězení za podobné činy jako předtím. Takže to počítač vyhodnotil přesně naopak. Na základě

¹ BURANYI, Stephen. Rise of the racist robots – how AI is learning all our worst impulses. The Guardian [online]. 2017, 8. 8. 2017 [cit. 2021-02-14]. Dostupné z: <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>

Magdaléna Ostárková

Gymnázium Olgy Havlové, Ostrava–Poruba

Téma č. 1: Umělá inteligence a lidská práva

Jak zajistit, aby umělá inteligence rozhodovala spravedlivě?

tohoto označení se pak rozhoduje o změně či snížení trestu. A lidé označení za vysoce rizikové dostávají často vyšší tresty či mají menší šanci na propuštění dříve.²

Nejedná se samozřejmě o ojedinělý případ. Takových bychom mohli zejména ve Spojených státech amerických najít spoustu. Umělou inteligenci tam nepoužívá jen vězeňská služba, ale také policie, která je umělou inteligencí posílaná častěji do lokací se zvýšenou kriminalitou, bohužel se tady zase jedná o chudé černošské čtvrti. Umělou inteligenci využívají ale také soudci při rozhodování o trestech – proto se stává, že Afroameričané dostávají tvrdší tresty než lidé bílé pleti.³ Afroameričané bývají také znevýhodňováni při poskytování lékařské péče.⁴

Myslím si, že takové jednání umělé inteligence je zapříčiněno vývojáři algoritmů. Většina vývojářů je totiž bílé pleti a také jsou to muži. Také 11 z 15 firem, které vlastní digitální platformy, je z USA.⁵ Jedno řešení se zde nabízí – zajistit, aby týmy vývojářů nebyly homogenně zastoupeny pouze muži bílé pleti, ale také ženami a lidmi různého etnického zastoupení, ale také různého věku. Domnívám se, že by to mohlo pomoci ke konfrontaci názorů, docílilo by se různých pohledů na věc a pomohlo by to odbourat určité předsudky a stereotypy, které potom přejímá také umělá inteligence.

Zde se ale naskýtá další problém a to, jak zajistit, aby vývojářské týmy nebyly složeny pouze z mužů či pouze z jedné etnické skupiny lidí. Zaměřím se nejprve na zastoupení muži x ženy. Samozřejmě zákon zaručuje rovné pracovní podmínky a příležitosti. Praxe je ale jiná. A nehraje zde roli jen a pouze lidský faktor. Zde se už zapojuje i umělá inteligence. Protože

² ANGWIN, Julia, Jeff LARSON, Surya MATTU a Lauren KIRCHNER. Machine Bias. ProPublica [online]. 2016, 23. 5. 2016, (-), - [cit. 2021-02-14]. Dostupné z:

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

³ ANGWIN, Julia, Jeff LARSON, Surya MATTU a Lauren KIRCHNER. Machine Bias. ProPublica [online]. 2016, 23. 5. 2016, (-), - [cit. 2021-02-14]. Dostupné z:

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

⁴ LEDFORD, Heidi. Millions of black people affected by racial bias in health-care algorithms. Nature [online]. 2019, 26. 10. 2019, (-), - [cit. 2021-02-14]. Dostupné z:

<https://www.nature.com/articles/d41586-019-03228-6>

⁵ Bias, racism and lies: facing up to the unwanted consequences of AI. UN news [online]. 2020, 30. 12. 2020, (-), - [cit. 2021-02-14]. Dostupné z: <https://news.un.org/en/story/2020/12/1080192>

Magdaléna Ostárková

Gymnázium Olgy Havlové, Ostrava–Poruba

Téma č. 1: Umělá inteligence a lidská práva

Jak zajistit, aby umělá inteligence rozhodovala spravedlivě?

ještě předtím, než se nějaká žena či dívka rozhodne pro své budoucí zaměstnání, je vystavena reklamám na sociálních sítích, které jsou řízené algoritmy. Ty samozřejmě nyní vytváří převaha mužů, proto například se ženám tolik nezobrazují nabídky práce například v IT sektoru. Proto ženy mnohdy ani nezačnou nad takovou kariérou přemýšlet. A opět zde hrají roli stereotypy. A v podstatě stejný princip se dá použít také v otázce etnického zastoupení. Nehledě na to, že si poté zaměstnavatel volí své zaměstnance a jeho rozhodnutí také nemusí být naprosto nestranné a neovlivněné stereotypy a předsudky.

Další možností, jak regulovat předpojaté rozhodování umělé inteligence je ověřování a úprava dat, které využívají algoritmy. Ty totiž často pracují s daty, které jsou sice pravdivé, ale ovlivněné stereotypy implementované lidmi. Například výše zmíněný algoritmus vyhodnocující lokality se zvýšenou kriminalitou, kam pak posílá hlídky policie častěji než jinde, pracuje s historickými daty, které označují místa zatčení. Bohužel počty zatčených jsou v černošských lokalitách vyšší než v jiných sousedstvích, protože předsudky se řídící policisté častěji zatýkají osoby černé pleti. Algoritmus tedy pracuje s opravdovými daty, nicméně ty jsou ovlivněné lidským předpojatým úsudkem.

Možným řešením této situace je pozměňování dat, které vývojáři do algoritmu vkládají. Data tak například nehledí na věk osob, jejich pohlaví a rasu či další faktory. Vědci poté zjišťují, do jaké míry jsou výsledky proměnnými ovlivněny. Nikdo ale neví, jestli takové pozměňování opravdu změní zaujatost výsledků nebo to akorát povede k jiným závěrům, ale stejně zaujatým. Nikdo si totiž dnes není jistý, do jaké míry a jak stereotypy rozhodování ovlivňují. Je to příliš komplexní problematika, která se nedá jednoznačně zkoumat. Toto pozměňování by poté mohlo vést k jiným problémům. Mohly by vznikat nereálné scénáře, data by totiž již nereflektovala realitu, ale utopickou fikci. Mnoho vědců se ale domnívá, že programy s takto pozměněnými daty by se staly nefunkčními.⁶

⁶ BURANYI, Stephen. Rise of the racist robots – how AI is learning all our worst impulses. The Guardian [online]. 2017, 8. 8. 2017 [cit. 2021-02-14]. Dostupné z: <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>

Magdaléna Ostárková

Gymnázium Olgy Havlové, Ostrava–Poruba

Téma č. 1: Umělá inteligence a lidská práva

Jak zajistit, aby umělá inteligence rozhodovala spravedlivě?

Dalším problémem je netransparentnost programů. Mnoho firem nechce v rámci ochrany značky sdílet informace o daném algoritmu, například na základě jakých dat vyhodnocuje. To se snažila Evropská unie regulovat Obecným nařízením o ochraně osobních údajů. Přesto ale firmy nejsou povinny sdílet všechny informace, které by byly nutné k vyhodnocení správnosti výsledků daného programu. Toto je problém hlavně v případě, kdy soukromá firma poskytuje služby nějaké vládní instituci. Ta její program využívá, ale přesně neví, na základě čeho program výsledky vyhodnocuje. Myslím si, že by stálo za zvážení zavedení určitých regulací, které by byly schopny ještě více zajistit transparentost algoritmů a informace o datech, která vyhodnocují.⁷

Bohužel doba jde stále kupředu a často se stává, že zákony vydávané státem nestačí rychlosti vývoje moderních technologií. Státy sice prosazují zákaz diskriminace, když se ale jedná o diskriminaci technologiemi, často nejsou schopny tuto diskriminaci adresovat a účinně proti ní bojovat. Možná o nich státníci ani nevědí. Nicméně to neznamená, že tato forma diskriminace neexistuje. Stát by měl proti ní účinněji bojovat a zasazovat se o změny v zákonech, které by diskriminaci skrze data objevovaly a postihovaly.

Myslím si, že nelze jednoznačně říci, co by se mělo udělat, aby umělá inteligence rozhodovala spravedlivě. Algoritmy jsou totiž ovlivňovány lidskými úsudky, které mnohdy spravedlivé a nezaujaté nejsou. Protože jsou algoritmy nastavovány lidmi, často se to projeví. Jistě by se na vývoji algoritmu mělo podílet co možná nejrozmanitější spektrum lidí, aby mohli sdílet své různé názory, které by se v algoritmu následně projevíly. Morální hodnoty jsou pro každého člověka jiné, proto nelze stanovit jednotná pravidla, protože bychom poté žili v totalitním světě. Na druhou stranu jsou kritéria, podle kterých bychom se všichni měli řídit. Ty vychází nejen ze zákonů jednotlivých států, ale také z mezinárodních smluv a dohod,

⁷ BURANYI, Stephen. Rise of the racist robots – how AI is learning all our worst impulses. The Guardian [online]. 2017, 8. 8. 2017 [cit. 2021-02-14]. Dostupné z: <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>

Magdaléna Ostárková

Gymnázium Olgy Havlové, Ostrava–Poruba

Téma č. 1: Umělá inteligence a lidská práva

Jak zajistit, aby umělá inteligence rozhodovala spravedlivě?

jako například Listina základních práv a svobod. Myslím si, že podle těchto měřítek by se měla řídit také umělá inteligence a lidé by měli zajistit, aby tak činila.