

Slovo algoritmus bylo pro většinu z nás donedávna součástí hantýrky informatiků, kterému bylo třeba věnovat v běžném životě asi tolik pozornosti jako pojům z geologie nebo lingvistiky – totiž v podstatě žádnou. Dnes je ale vše jinak. Německá verze Google News (služba umožňující prohledávat internetová média) nabízí při vyhledávání slova algoritmus na zpravodajských stránkách skoro 50 tisíc výsledků. Nadpisy článků jsou namátkou tyto: „Algoritmus pozná vyhané e-mail“, „Transparentnost: Algoritmus Amazonu je pro zákazníky matoucí“ nebo „Tipy pro psaní motivačního dopisu: Takhle přelstíte algoritmus“. To poměrně dobře dokládá, jakou roli začaly algoritmy najednou hrát v našich každodenních životech.

# Plíživá nadvláda ALGORITMŮ

Algoritmy jsou zjednodušeně řečeno zadáním pro činnost jakýchkoli počítačových programů, diktují jim, jaké kroky a v jaké posloupnosti mají učinit. Jako takové tedy nejsou ničím novým, základem programování jsou odjakživa (první algoritmus k programování předchůdce dnešních počítačů vznikl již v 19. století). S tím, jak se naše životy přesunují do digitálního prostoru a zanechávají po sobě hromady digitálních informací (tzv. big data), se nicméně s analýzou těchto informací otevírá algoritmům nové pole působnosti. Právě algoritmy představují jakési zadání, podle kterého jsou vyhodnocována data. A protože jsou z těchto analýz čím dál častěji vyvozovány důsledky pro náš digitální, ale i reálný život, problematika algoritmů se nás všech začíná bezprostředně dotýkat.

## Živé personalisty už v USA téměř nepotřebují

Algoritmy například rozhodují o tom, co se nám zobrazuje na internetu – ať už jde o reklamu šitou na tělo uživateli, nebo ceny na Amazonu, které jsou prokazatelně za stejnou věc pro různé uživatele různě vysoké. Algoritmy většinou určují, zda a za jakých podmínek dostaneme půjčku nebo který pojišťovací tarif nám bude nabídnut. Ve Spojených státech zase používají algoritmy k předpovědi pravděpodobnosti recidivy u vězňů.

Velké pole využití mají algoritmy v personalistice. Odhaduje se, že ve Spojených státech i Velké Británii projde kolem 70 procent uchazečů nějakým typem automatizovaného testu, při kterém algoritmus automaticky podle jejich odpovědí oddělí ty pro danou pozici vhodné od těch nevhodných. Zřejmě nejdále se v této oblasti dostala firma Psyware, která vyvinula software, který dokáže podivuhodně přesný psychologický profil člověka vytvořit na základě patnáctiminutového rozhovoru s ním. Ovšem ne z toho, co říká, ale jak to říká – jakým tempem mluví, kolik slov používá a jak jsou v jeho projevu zastoupeny jednotlivé slovní

druhy, jak nahlas mluví či kolik používá záporů.

Kolem 70 procent životopisů ve Spojených státech již vůbec neprojde rukama živého personalisty, i u nich provedou selekci právě algoritmy, a to podle toho, zda obsahují zvolená klíčová slova, nebo také podle toho, zda by uchazeči o práci na pracoviště museli dojíždět z dálky, protože u takových je statisticky větší pravděpodobnost, že rychleji dají výpověď. Algoritmy kromě toho rozhodují o tom, komu se při prohlížení internetu ukazuje v políčku pro reklamu nebo při hledání ve vyhledávači inzerát na jakou pracovní pozici.

## Snaha eliminovat lidský faktor a její úskalí

Algoritmy se pro analýzu big data využívají samozřejmě hlavně kvůli objemu informací, s nimiž by se člověk buď jen velmi obtížně vypořádal, anebo by to v jeho silách nebylo vůbec. Například zmíněný program Psyware dokáže za 0,2 vteřiny posoudit účastníka rozhovoru podle neuvěřitelných 180 000 různých kritérií. Zároveň mají algoritmy také eliminovat tzv. lidský faktor – je kupříkladu dokázáno, že soudci rozhodují přísněji, když mají hlad, nebo že personalisté nevybírají uchazeče o zaměstnání jen na základě skutečných kompetencí, ale spíše podle toho, kdo je jim samým nejpodobnější. A protože algoritmu hlad úsudek nezatěmní, věříme v jejich objektivitu a racionalitu.

Někteří pozorovatelé začínají tento předpoklad zpochybňovat. Často se totiž zapomíná, že i algoritmy někdo sestavil a že ten někdo nemá neutrální či objektivní náhled na svět. Způsob, jakým takový člověk algoritmus sestaví, musí zákonitě obsahovat zkresení zapříčiněné jeho subjektivním vnímáním reality nebo může reprodukovat rozšířené stereotypy, kterým vývojář sám podléhá. Typickým příkladem je automatizovaná práce algoritmů s rádoby objektivními kritérii, jako je dosažené vzdělání či bydliště. Vypovídá formálně dosažený

## Sondy Revue vám ve spolupráci s nadací



přináší seriál článků zaměřených na témata, o kterých se v ČR nediskutuje a nepíše.

stupeň vzdělání skutečně něco o vhodnosti uchazeče nebo uchazečky na danou pozici? Nevylučujeme slepým rozřazením podle těchto kategorií lidí, kteří neměli stejné výchozí podmínky jako jiní, a neznevýhodňujeme je tak znovu? Firma, která vyvinula výše zmíněný algoritmus, jenž z uchazečů o zaměstnání vyřazoval ty, kteří měli do práce delší dojezd, zjistila, že se ve většině případů jedná o hůře sociálně postavené, kteří bydlí na periferii nebo si nemohou dovolit se kvůli práci přestěhovat. Vyřadit je kvůli tomu z výběrového řízení znamenalo zastavět jim cestu z chudoby, která způsobovala jejich vyloučení a uvrhla je tím do začarovaného kruhu. Firma po tomto zjištění kritérium z algoritmu odstranila.

Problém ale nastává i s výběrem dat, která mají být podrobena algoritmické analýze. I do něj se může promítat specifické vidění světa vývojáře algoritmu, které je dále přeneseno do výsledků jejich analýzy. A konečně je třeba počítat s tím, že i samotná data obsahují zkreslení reality nebo sociální vzorce, na jejichž reprodukci nebo zesilování nemáme zájem, ale u nichž naopak usilujeme o odstranění. Marginalizované skupiny jsou například kvůli svému znevýhodnění obsaženy v datech méně nebo tak, jak to odpovídá jejich diskriminaci. Třeba při analýze obsazení rozhodovacích pozic vychází, že ženy jsou v nich zastoupeny méně, to je ale něco nežádoucího a algoritmy by z toho neměly slepě vyvozovat své předpově-

di nebo doporučení ohledně budoucnosti nebo tímto směrem nenápadně směřovat něčí chování.

## Kde už končí legrace

Konkrétní příklady podobných potíží s algoritmy se začínají kupit – k humornějším z nich patří pokus uspořádat soutěž o „objektivně“ (vybraného strojem) nejkrásnějšího člověka. Lidé byli vyzváni k zaslání svých fotografií, z nichž měl algoritmus

vybrat toho nejkrásnějšího. Fotky nicméně zaslalo daleko méně lidí s tmavou barvou pleti, a tak byli mnohem méně zastoupeni mezi vítězi. Algoritmus si z jejich chybějícího zastoupení ve vstupních datech odvodil, že jsou jejich vzhledové znaky spíše negativní odchylkou od obecně sdílené představy o kráse.

O něco méně humorné je, když se ženám na internetu zobrazuje málo inzerátů na obsazení vyšších pozic, protože si algoritmus z faktu, že jsou v takových

pozicích méně zastoupeny, odvodí, že jim nejspíš nepřísluší o ně usilovat, čímž stávající nespravedlivé poměry dál reprodukuje. K těm nejotřesnějším příkladům se řadí kritické zhodnocení výsledků algoritmické předpovědi pravděpodobnosti recidivy u amerických vězňů. Vyplývá z něj, že kriminálníci tmavé pleti mají dvakrát vyšší pravděpodobnost, že u nich bude riziko recidivy předpovězeno mylně, zatímco

vězňům bílé pleti je dvakrát častěji mylně předpovězena bezproblémovost.

Čím rozšířenější je algoritmická analýza dat a čím častější je, že má pro naše životy reálný dopad třeba ve formě toho, zda dostaneme práci nebo ne, tím více je potřeba o jejich úskalích přemýšlet. Takovou reflexi ale komplikuje, že ne vždy se dozvíme, že byla v konkrétním případě provedena. Odmítnutí uchazeči o zaměstnání se většinou nikdy nedozví, zda je neodmítl algoritmus na základě diskriminačních předpokladů. Přesná podoba algoritmů a dat, z nichž vychází, je u většiny programů a aplikací přísně střeženým obchodním tajemstvím a je těžké je podrobit kritickému přezkoumání. Alespoň v případě jejich využití ve veřejných službách – jako u zmíněného algoritmu na predikci recidivy – bychom měli na odkrytí algoritmického mechanismu a příslušných dat trvat.

## „Objektivní“ informace versus nealgoritmizovatelné sympatie

Další výzvou pak bude využití algoritmů, které se budou postupným zpracováním dat samy učit a osamostatňovat se. Tam už bude v podstatě nemožné vysledovat východiska pro jejich rozhodnutí a budeme jim vydání napospas. Měli bychom si tedy dobře rozmyslet, na co je využíváme a kde bychom možná přece jen měli zůstat u rozhodnutí lidí, i když třeba budou méně rychlá nebo přesná. Lidský faktor má koneckonců kromě negativních efektů i pozitivní stránku – totiž lidskost. Kupříkladu proces zaměstnávání nového kolegy by možná měl i přes všechny možnosti techniky zůstat lidskou interakcí, ve které se na rozhodnutí podílejí i nealgoritmizovatelné sympatie nebo lidský příběh, který přebije všechny objektivní informace ze životopisu.

Z celé věci se dá vyvodit ještě jeden závěr: především v české diskusi vyvolává rychlý technologický pokrok ještě častější volání po technicky vzdělaných lidech. Sociální dopady nesprávného využití algoritmů jsou ovšem poměrně dobrým příkladem, proč bude v digitální éře možná více potřeba lidí humanitně vzdělaných, kteří budou na podobná společenská nebezpečí upozorňovat. Teprve správným směrem namířený technologický vývoj můžeme označovat za pokrok. Od lidí čistě technicky vzdělaných můžeme takové usměrnění očekávat jen těžko. Pro budoucnost bude ještě zásadnější užší mezioborová spolupráce humanitních a technických oborů.

*První verze textu vyšla na portálu A2larm.cz*



**Kateřina Smejkalová je politoložka, pracuje jako vědecká pracovnice v organizaci Friedrich-Ebert-Stiftung**