

Umělá inteligence a novinářina: Obavy o nahrazení roboty nejsou namístě

Sběr dat, kontrola faktů v reálném čase, automatizovaná videa a grafika nebo přesně cílené šíření informací. Umělá inteligence v žurnalistice už dávno není slovním spojením z žánru science fiction. Ve světových redakcích se již několik let úspěšně používá. Nástroje žurnalistiky umělé inteligence, které fungují na principech strojového učení a zpracování přirozeného jazyka, jsou schopny provádět nejen běžné zpravodajské úkoly, ale také vytvářet zprávy samotné.

● Sandra Štefaníková

Zmiňme některé zajímavé příklady. Program Truth Teller vyvinula společnost Washington Post s cílem ověřovat pravdivost výroků politiků v reálném čase. Washington Post od roku 2016 používá také nástroj Heliograf. Prvně byl využit pro tvorbu automatizovaných zpráv z olympiády v Rio de Janeiru. Prestižní americký deník používá Heliograf i nadále, konkrétně k pokrývání tzv. hyperlokálních zpráv (zpráv určených pro úzký okruh čtenářů – pozn. red.) z menších sportů, které by se jinak do médií dostávaly jen obtížně.

Platforma Wordsmith od společnosti Automated Insights zase stojí na principu zpracování přirozeného jazyka a dokáže převádět data do smysluplného způsobu vyprávění. V novinářském světě uvedený systém již několik let používá zpravodajská agentura Associated Press (AP) ve zpravodajství z oblasti byznysu. AP je také první novinářskou organizací, která má pracovní pozici „editor automatizace“, ten dohlíží na produkci automaticky vytvářených zpráv. Reuters k hledání zajímavých příspěvků z nepřehledné produkce na sociálních sítích používá nástroj News Tracer („sledovač zpráv“). Systém zastává roli prvotního gatekeepera (jakéhosi vrátného), který pracuje s principem zpravodajských hodnot a příspěvkům přiřazuje skóre, které ve výsledku může rozhodnout, zda bude zpravodajská agentura o události informovat.

Služba textového videa Wibbitz umožňuje uživatelům, mezi které patří USA Today, New York Times a Le Figaro, vytvářet videa během několika minut. Systém umělé inteligence Graphiq je schopen produkovat desítky interaktivních grafů denně bez zásahu člověka, s minimálními náklady a za velmi krátkou dobu, což z něj činí neocenitelný nástroj pro vytváření informačních grafů.

V současné době se automatizovaný obsah týká především rutinního zpravodajství, ke kterému jsou dostupná dobře čitelná a přesná strukturovaná data. Automatizace se proto dobře aplikuje například na výsledky sportovních klání, hospodaření na burze, na zprávy o počasí nebo na výsledky voleb. Poměrně složitější je potom používání v oblasti publicistiky, tedy textových útvarů, které vyžadují kritické myšlení a tvůrčí činnost. V této oblasti zatím k zavádění automatické žurnalistiky nedochází.

Prim anglofonního prostředí

Z výše zmíněných příkladů je také patrné, že dobře fungující systémy žurnalistiky umělé in-

teligence vznikají v první řadě v anglicky mluvícím prostředí. Anglický jazyk v tom hraje svou roli, ale trochu jinak, než by se mohlo na první pohled zdát. Principy angličtiny, její gramatika a syntax, mohou svádět k úvaze, že pro stroje je snazší se naučit spíše anglický než český jazyk. Na druhou stranu, pokud by toto platilo, měly by stroje být schopné lehce zvládnout i jiné „uživatelsky přátelštější“ jazyky, například indonéštinu nebo itaštinu. Pro strojové hluboké učení jazyková bariéra spočívá spíše než ve složitosti či náročnosti konkrétního jazyka v nedostatku vstupních dat. Ideální umělá inteligence je totiž taková, která je skutečně inteligentní a pracuje bez nebo s minimálním zásahem člověka. Nefunguje podle předem daných a vytvořených šablon jako běžné algoritmy, ale je schopná se na základě vstupních dat sama jazyk naučit a následně ho aktivně používat. Jedná se o počítačové zpracování přirozeného jazyka (tzv. Natural Language Processing – NLP).

To, do jaké míry se stroj bude schopný jazyk naučit, spočívá v tom, jaké množství textů mu nabídnete k učení. V anglickém jazyce se za tímto účelem používají rozsáhlé soubory textů (korpusy), které jsou veřejně dostupné. Z tohoto pohledu anglicky mluvící svět skutečně těží ze širokého využívání anglického jazyka, který se stal novým „lingua franca“.

Popularita a rozšířenost angličtiny přitom s sebou nese ještě jeden důležitý aspekt, který je nezbytně nutný pro vývoj žurnalistiky umělé inteligence. Jedná se o finanční zdroje na komerční i výzkumné (akademické) úrovni. Zde se logika zcela podřizuje trhu. Širší čtenářská obec znamená potenciálně širší zásah, vyšší zisky i vyšší investice do inovací. Proto se mezi průkopníky žurnalistiky umělé inteligence řadí zvučná jména novinářského světa. Data agentury AP ukazují, že po zavedení platformy Wordsmith se zisky v oblasti byznysu zdesetinásobily. Podobná úměra přitom platí i ve vý-



Foto pixabay

zkumné oblasti, která je v anglicky mluvícím prostředí logicky nesrovnatelně větší co do počtu vědeckých pracovníků i financí než třeba v Maďarsku, Norsku nebo České republice.

Situace v České republice

Bez ohledu na podfinancování a nepříliš vysokou návratnost investovaných zdrojů se žurnalistika umělé inteligence pomalu rozvíjí také v České republice. Cílem těchto snah je přitom vyvinout a v praxi v žurnalistice využívat takovou umělou inteligenci, jakou jsme popsali výše. Pravdou je, že většina dosavadních snah v oblasti finančního, sportovního či volebního zpravodajství se vydala spíše cestou automatizované žurnalistiky, jejíž aplikace je vhodná především na oblasti, kde se využívá velké množství strukturovaných dat.

O co v tomto případě jde? Fungování automatizované žurnalistiky se na rozdíl od ideálu žurnalistiky umělé inteligence neobejde bez předdefinovaných šablon a pomoci člověka. Principy strojového učení a počítačové zpracování přirozeného jazyka nejsou v případě automatizované žurnalistiky vůbec využívané nebo jen omezené. Nejedná se tedy o počítačového novináře, který by dokázal jakkoliv konkurovat novinářům z masa a kostí. Naopak pomoc člověka a kontrola z jeho strany představuje pro automatizovanou žurnalistiku naprostou nezbytnost.

Na zmiňovaném principu funguje často uváděný příklad volebního automatu RUR, který v roce 2018 využívala Česká tisková kancelář (ČTK) u příležitosti voleb do Senátu Parlamentu ČR a zastupitelstev obcí. ČTK využívala jednoduchých postupů tím, že byla v průběhu sčítání hlasů zpracovávána data od Českého statistického úřadu (ČSÚ). Pro tento účel byly vytvořeny šablony pro tři typy zpráv, které se automaticky generovaly. V databázi byly připraveny číselníky stran, měst, volebních obvodů, klíčová slova ke zprávám a další údaje. V průběhu načítání dat od ČSÚ se zpracovávaly a detekovaly okamžiky vhodné pro vydávání zpráv, tj. po sečtení 50 % hlasů v daném obvodu a ve chvíli, kdy byl zřejmý vítěz voleb. Následně se vydávala také souhrnná zpráva, která shrnovala výsledky voleb. To, co algoritmus vygeneroval, se v prostředí redakčního systému automaticky propojilo s předem připravenými událostmi. Systém dokázal přebírat metadata (data poskytující informace o jiných datech – pozn. red.) od navázané události a přiřazovat je ke automaticky vytvářeným zprávám. Následně šel text do rukou editora, který jej případně upravil a vydal.

Práce tuzemských vědců

Také na české akademické půdě se pracuje na vývoji a aplikaci žurnalistiky umělé inteligence. Středisko žurnalistiky umělé inteligence vzniklo v roce 2019 při katedře žurnalistiky IKSŽ FSV UK. „V počátcích našeho prvního grantu z Technologické agentury ČR jsme experimentovali s formátem „shrnutí“, říká Václav Moravec, vedoucí střediska. „ČTK shrnutí používá k získávání klíčových informací ze sledu zpráv o stejné

O Středisku žurnalistiky umělé inteligence

Středisko žurnalistiky umělé inteligence vzniklo na počátku roku 2019 při katedře žurnalistiky IKSŽ, Fakulty sociálních věd UK. Jeho cílem je především rozvoj interdisciplinárního výzkumu na poli žurnalistiky, etiky a počítačových věd. Pracovníci střediska odborně zajišťují výuku volitelných předmětů, jež úzce souvisejí s nástupem automatizované žurnalistiky a žurnalistiky umělé inteligence; provozují portál Žurnalistika 5.0 popularizující nejnovější poznání v oblasti žurnalistiky umělé inteligence a automatizované žurnalistiky; realizují kurzy a workshopy v rámci celoživotního vzdělávání novinářů; vyhledávají doplňkové financování výzkumu a popularizace předmětné problematiky. V současnosti se tým podílí na řešení těchto výzkumů podpořených Technologickou agenturou ČR: Proměna etických aspektů s nástupem žurnalistiky umělé inteligence, Multimediální platforma Beey a Umělá inteligence, média a právo.

události. Jeho tvorba je časově náročná a vyžaduje velké úsilí ze strany novináře. Naším cílem bylo vytvořit systém, který by práci redaktorům ulehčil,“ dodává Moravec.

ČTK v současnosti publikuje desítky shrnutí každý den, takže do jejich produkce jsou zapojeny značné lidské zdroje. Při experimentu, který prováděli odborníci z oblasti NLP Západočeské univerzity v Plzni (ZČU), dostal počítač za úkol vytvořit automatické shrnutí souvisejících zpráv s využitím materiálů vybraných z archivu ČTK. Vědci použili dvě metody vytváření automatických shrnutí. Zatím ale nebyly získány zcela uspokojivé výsledky, a proto postupy ještě nejsou zcela připraveny k použití v praxi.

Druhý experiment iniciovaný střediskem při pražské katedře žurnalistiky byl již úspěšnější. Tentokrát se odborníci zaměřili na automatické generování zpráv z výsledků obchodování na Burze cenných papírů Praha. V tomto případě byl použit ověřený systém předdefinovaných šablon. „Návrh šablony je založen na stávajících zprávách ČTK, kdy počítač analyzuje věty ze svých předchozích zpráv. Pomocí metod strojového učení a tzv. neuronových sítí také experimentujeme s automatickým generováním šablon na základě starších zpráv a souvisejících strukturovaných dat,“ uvedl Jakub Sido, člen expertního týmu ze ZČU.

Výsledkem je plně funkční systém, který generuje novinky z dat poskytnutých Burzou cenných papírů Praha. „Naše algoritmy jsou nyní schopny generovat sedm základních druhů novinek, včetně titulků, bez zásahu novinářů během jedné sekundy. Kromě ČTK a České televize se do našeho výzkumu zapojil od letošního června i ekonomický deník E15,“ uvedl Moravec.

Pozitiva automatizace

Výhody automatizace žurnalistiky jsou podobné jako v jiných oblastech. Elektronicky definované postupy nejenže mohou vytvářet tisíce zpravodajských příběhů pro konkrétní téma, ale také to dělají rychleji, levněji a potenciálně s menším počtem chyb než jakýkoliv lidský novinář. Nad rámec těchto zjevných výhod můžeme jmenovat další potenciály automatizované žurnalistiky. Stejná data mohou být použita pro tvorbu textů vyprávěných z různých úhlů pohledu, čímž se přizpůsobují individuálním preferencím čtenáře. Algoritmy tak mají například potenciál generovat zprávy na vyžádání a vytvářet zprávy v reakci na požadavky čtenářů.

Druhou stranou mince jsou potom některé limity, případně potenciální hrozby vyplývající z použití automatizované žurnalistiky. V základní rovině je to nedostatečná jazyková vybavenost a obratnost, která vede k horší kvalitě automatizovaných zpráv. Zde je ale nutné podotknout, že automatizace se v daném ohledu neustále zdokonaluje. Největší problém pro automatizovanou žurnalistiku představuje její hlavní pracovní nástroj, a sice pevně semknutí se strukturovanými daty, která mohou být chybna či zkreslená. Výsledný text tak může obsahovat závažné dezinformace, případně může být automatizace zneužita k šíření dezinformací. Algoritmy totiž zatím nedokáží (a nejspíš dlouho ani nebudou umět) odhalovat příčinné souvislosti, nekladou otázky, ani nevytvářejí nové jevy.

Nese robot odpovědnost?

Jedna z důležitých diskusí v oblasti žurnalistiky umělé inteligence se nevyhnutelně týká také odpovědnosti za automatizovaný obsah. Nese ji novinář, který ho použije pro svůj článek, editor, šéfredaktor nebo celé vydavatelství? Je třeba, aby redakce používající automatizovaný obsah byly transparentní a vytvořily základní strukturovaných dat,“ uvedl Jakub Sido, člen expertního týmu ze ZČU.

Hojně se diskutuje také problematika autorství. Jak mají být označeny články, které jsou částečným nebo kompletním produktem umělé inteligence? Měli bychom pro ně vytvořit univerzální zkratku (UI, AI, RUR atd.), stejně jako její konkrétní tiskové agentury, například ČTK nebo AP? Tak či onak je jasné, že umělá inteligence proniká do redakcí a pravděpodobně v nich také zůstane. Jaké poučení z toho plyne pro mediální organizace, novináře, publikum a společnost?

Novinář se zatím nemusí bát o živobytí, pokud tedy náplní jejich práce není strohé zpracování strukturálních dat. Spíše než jako konkurenci lze umělou inteligenci a automatizovaný obsah chápat jako prodlouženou ruku novináře. Tedy nástroj, který může novináře osvobodit od zdoluhavých a pro mnohé i nudných činností, jako je analýza velkého počtu dat, a uvolnit jim ruce pro jiné činnosti. Novinář si tak musí být vědomí toho, jakým způsobem automatizovaný obsah vzniká, jaké jsou jeho přínosy a limity, aby ho dokázal vhodně využít a korigovat. Z toho vyplývá, že při využívání automatizovaného obsahu je stěžejní pozice editora, který dohlíží na kvalitu, vyváženost a informační přesnost.