

Umělá inteligence načetla knihu hlasem Karla Gotta. Vznikl na Fakultě aplikovaných věd ZČU

31.3.2024 - Pavel Korelus | Západočeská univerzita v Plzni

V roce 1939 se v Plzni narodil Karel Gott, v roce 2023 se na Západočeské univerzitě v Plzni (ZČU) znovu zrodil jeho hlas. Český rozhlas (ČRo) a Karel Gott Agency představí ve čtvrtek 13. července, v předvečer zpěvákových nadožitých 84. narozenin, projekt GOTT NAVŽDY, který by bez ZČU nevznikl. Gottova autobiografická kniha Má cesta za štěstím z roku 2021 získala audiopodobu, v níž hlas Karla Gotta rozezněla umělá inteligence. Knihu načetla spolu s hercem Igorem Barešem.

ČRo se na tým syntézy řeči, který na katedře kybernetiky a ve výzkumném centru NTIS Fakulty aplikovaných věd ZČU vede Jindřich Matoušek, obrátil koncem loňského roku. *„Byla to výzva. Jsme sice perfekcionista a asi nikdy nebudeme úplně spokojeni, ale myslím, že v podmínkách a časových možnostech, jaké byly, jsme dosáhli výborného výsledku,”* říká Jindřich Matoušek.

Katedra kybernetiky má s technologií syntézy řeči bohaté zkušenosti. *„Řečové technologie nejsou žádná novinka, jsou tu s námi už 30 nebo 40 roků a na katedře s nimi pracujeme od devadesátých let. Z poslední doby je známý například náš projekt automatické konzervace hlasu, který vznikl zejména jako pomoc pro nemocné, jimž hrozí, že přijdou o hlas. Spolu s našimi partnery, 1. Lékařskou fakultou Univerzity Karlovy v Praze a softwarovými firmami SpeechTech a CertiCon, jsme navrhli proces, v rámci kterého si lidé doma nahrají svůj hlas, dokud ještě mohou hovořit, a my z nahrávek vytvoříme jeho kopii, již pak promlouvají z různých přístrojů. Tyto zkušenosti jsme využili i při vytváření hlasu Karla Gotta,”* vysvětluje Jindřich Matoušek.

Fakultu aplikovaných věd oslovil ČRo prostřednictvím její spinoff firmy SpeechTech, s níž spolupracuje na zavádění hlasových, na umělé inteligenci založených technologií do každodenního provozu. Vědci museli během jednoho až dvou měsíců dokázat, že co nejvěrnější kopii Gottova hlasu vytvoří. Data jim poskytl rozhlasový archiv, neboť zpěvák v letech 2011–2015 na stanici Dvojka moderoval vlastní pořad Zpátky si dám tenhle film.

„Dostali jsme k dispozici velké množství nahrávek v délce kolem 200 hodin. Pro syntézu řeči je běžně potřeba několik desítek, ale to jde o člověka, který text načítá přímo za tímto účelem, a nahrávání je pod kontrolou. Tentokrát byla situace jiná, po prostřihání zbylo asi 20 hodin záznamu bez šumu, slov v cizím jazyce nebo písní v podkresu. Díky nim jsme vyrobili první ukázkou, již Český rozhlas představil paní Ivaně Gottové, a ta pak rozhodla, že se bude pokračovat. To bylo někdy v prosinci 2022, v lednu a únoru jsme už odevzdávali první a začátkem června poslední kapitolu,” pokračuje Jindřich Matoušek.

Že se projekt chystá, oznámili Ivana Gottová a generální ředitel ČRo René Zavoral 13. června. Odezva fanoušků byla okamžitá a názory veřejnosti budou podle Jindřicha Matouška rozhodně zajímavé: *„Když děláte syntézu hlasu, který všichni moc dobře znají, je to něco úplně jiného, výsledek bude každý kriticky hodnotit. Velmi důležitá je samozřejmě etická stránka věci, neboť bez souhlasu rodiny Karla Gotta by takový projekt vzniknout nemohl. Ale když už svolení máte, je to velká příležitost oslovit širokou veřejnost, předvést jí technologii syntézy řeči i možnosti umělé inteligence, a ukázat, že se jí nemusí bát.”*

Na rekonstrukci hlasu Karla Gotta pracoval šestičlenný tým. Kromě těch, kteří se zabývali zvukem, se zapojili i anotátoři společnosti SpeechTech, jelikož nezbytnou součástí byl i text. *„Kromě nahrávek jsme potřebovali jejich kompletní textový přepis. 200 hodin znamená asi 200 tisíc vět, a protože rozhlasový pořad neměl scénář a v časové tísní nebylo možné přepis obstarat ručně, pomohla firma SpeechTech a její rozpoznávač řeči, vyvinutý ve spolupráci s našimi kolegy z centra NTIS. Řečové nahrávky převedl na velmi kvalitní text, v němž anotátoři pouze opravovali některá slova. Na základě nahrávek a odpovídajících přepisů jsme pak mohli natrénovat model syntézy řeči založený na moderních hlubokých neuronových sítích tak, že hlasem Karla Gotta přečte zadaný psaný text,“* vysvětluje Jindřich Matoušek.

Hlas nakonec vznikl v šesti verzích, ze kterých mohl ČRo vybírat. *„Díky tomu, že zvuk ještě prošel náročnou postprodukcí, zní finální produkt asi nejlépe, jak znít mohl,“* pokračuje Jindřich Matoušek s tím, že na vědce čekají další výzvy.

„Bylo nám potěšením na tak zajímavé úloze pracovat a doufám, že jsme i tímto projektem ukázali, že v České republice umíme dělat výzkum a vývoj řečových technologií a potažmo umělé inteligence na světové úrovni. Byla by škoda, kdybychom v budoucnu o toto postavení měli přijít například kvůli hrozící menší podpoře výzkumu, která by se dotkla nás i všech kolegů a kolegyní z ostatních oborů. I když technologie syntézy řeči udělala v posledním desetiletí obrovský pokrok a v běžných úlohách, jako je čtení textu, funguje výborně, neznámá to, že je zcela vyřešena. Aby byl počítačový hlas plnohodnotným partnerem člověka v hlasovém dialogu, nestačí jen přečíst daný text. Syntéza musí také dokonale porozumět tématu konverzace, umět vyjádřit postoj mluvčího, náladu, emoce, mít paměť... Takové úlohy na svá řešení teprve čekají,“ dodává Jindřich Matoušek.

Četbu na pokračování Má cesta za štěstím odvysílá ČRo Dvojka během prázdnin, bude mít 49 dílů. Celkem je však dílů 73 a v nezkrácené podobě vyjdou v audioknize na podzim. ČRo připravil projekt GOTT NAVŽDY jako jeden z dáreků posluchačům ke 100. výročí zahájení pravidelného vysílání na našem území, více informací lze nalézt na gott.rozhlas.cz.

<https://info.zcu.cz/Umela-inteligence-nacetla-knihu-hlasem-Karla-Gotta--Vznikl-na-Fakulte-aplikovany-ch-ved-ZCU/clanek.jsp?id=5571>