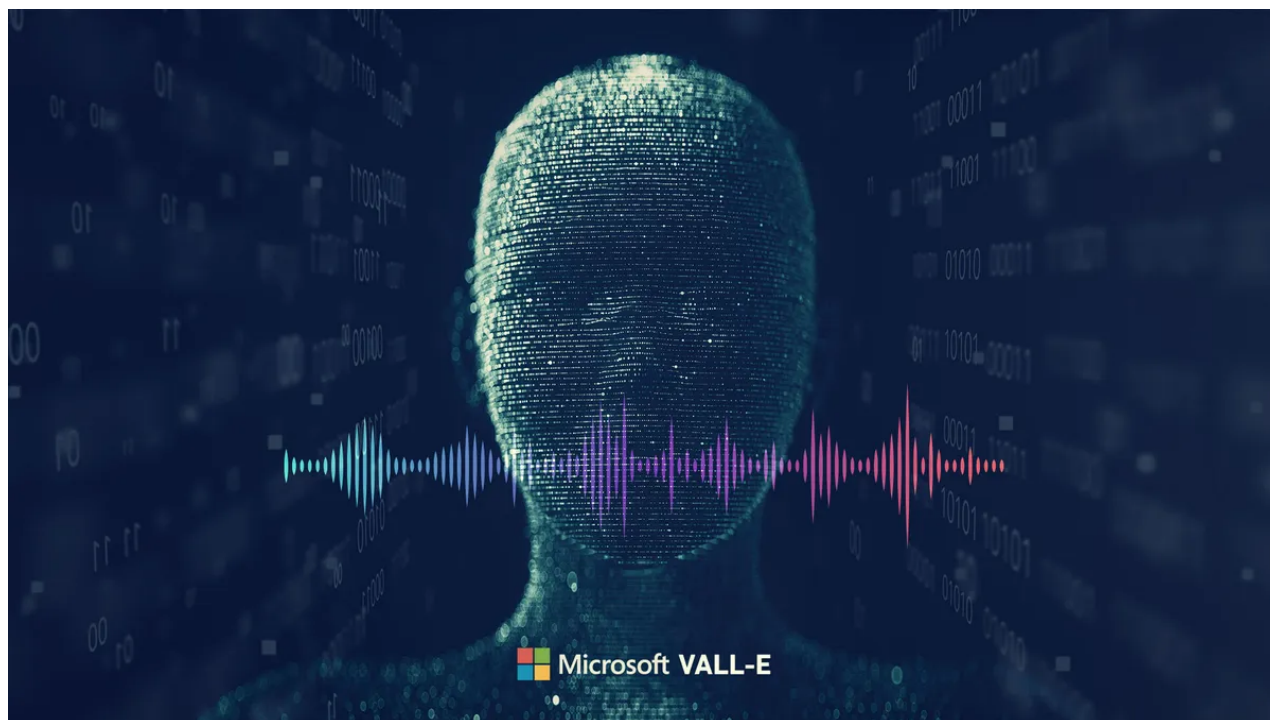


Microsoft představuje VALL-E, umělou inteligenci pro převod textu na řeč, kterou lze trénovat za pouhé 3 sekundy

[IE interestingengineering.com/innovation/microsoft-vall-e-text-to-speech](https://interestingengineering.com/innovation/microsoft-vall-e-text-to-speech)

10. ledna 2023



Výzkumní pracovníci společnosti Microsoft představili svůj nejnovější generátor převodu textu na řeč (TTS) VALL-E, který lze vycvičit tak, aby napodoboval hlas kohokoli během pouhých tří sekund. Na rozdíl od předchozích hlasových generátorů, které zněly roboticky, zní VALL-E přirozeně lidsky a to nemusí být moc dobrá věc.

Generátory převodu textu na řeč, které daly hlas jedné z největších myslí na planetě, Stephenu Hawkingovi, ušly dlouhou cestu. Od čtení zpráv na smartphonu až po čtení stránek z knihy – tyto služby jsou nyní všude a používají je všichni.

Viz také

Velké technologické společnosti, jako je Google, Meta a Microsoft, také pracují v tomto prostoru na zpřístupnění svých produktů. Tyto produkty však nejsou zaměřeny na napodobování hlasu uživatele a potřebují nespočet hodin školení, aby to dokázaly a špatně dopadly.

Ohromující schopnosti VALL-E

Generátory TTS se běžně spoléhají na manipulaci s průběhy při syntetizaci řeči. Na druhé straně VALL-E generuje diskrétní zvukové kodeky z textových a zvukových výzev a používá je k tomu, aby je porovnal s tím, co ví o tom, jak by hlas zněl, kdyby mluvil jinými frázemi.

Výzkumný tým tvrdí, že zvuková výzva by v tomto případě mohla trvat pouhé tři sekundy, což by pro VALL-E stačilo k tomu, aby vykonal svou práci. Díky tomu je VALL-E generátorem TTS zero-shot, kde software pozoruje vzorky, které během tréninku nepozoroval.

Zajímavé je, že školení VALL-E bylo provedeno pomocí LibriLight, zvukové knihovny, kterou dala dohromady Meta a obsahovala téměř 60 000 hodin anglické řeči z audioknih LibriVox, které jsou dostupné ve veřejné doméně.

VALL-E úspěšně porovná třísekundový zvukový vzorek s hlasem jednoho ze 7 000 lidí, které vycvičil, a poté doručí text hlasem podobným hlasu v trénovacích datech, aby bylo dosaženo přesné mimické odezvy.

Microsoft tvrdí, že VALL-E dokáže nejen simulovat hlasy v akustickém prostředí, jako je například telefonní hovor, ale také předat řeč v souladu s emocemi použitými ve výzvě reproduktoru, takže je mnohem personalizovanější a přirozenější.

Nejoblíbenější

K čemu by to mohlo vést

I když se jedná o velký technologický skok, není to příliš překvapivé. Může to být proto, že se blíží k úspěchu ChatGPT, kde algoritmus může chrlit eseje pro vysokoškoláky, a mohl by také napsat tento článek, kdyby nebyl tak obsazený.

Další produkt OpenAI, DALL:E, dokáže rozdávat obrázky v reakci na textové výzvy a nyní by technologie Microsoftu mohla oživit hlas dávno zesnulého herce v budoucím filmu. Zdá se, že podstatou těchto technologií je schopnost ušetřit peníze společností, které by mohly svou práci zvládnout zaplacením zlomku toho, co platí člověku.

Tato technologie by však mohla být také použita k podvržení jiného člověka uskutečněním nouzového volání nebo přístupem k citlivým informacím, které jsou uzamčeny hesly s podporou hlasu. Microsoft může v současné době držet klíče, aby se takové manipulaci vyhnul, ale jak jsme viděli u technologie AI dříve, netrvá příliš dlouho, než bude zkopírována a použita pro hanebný účel.

1. Domov
2. Inovace

 ZOBRAZIT KOMENTÁŘ (0) 