

Voicebotok a valóság próbáján: tanulságok valós ügyféltapasztalatokon keresztül

2026. 09. 07. 15:40

MKIK Prompt Aréna

HUN · 20 perc



Andrejkovics Zoltán
Senior AI Expert · VCC Live

Andrejkovics Zoltán, a VCC Live vezető MI szakértője, keretet állít fel a hangalapú MI jelenlegi korlátainak és gyakorlati alkalmazásainak megértéséhez a vevőszolgálatban.

MIRŐL VOLT SZÓ

A prezentáció különbséget tesz a hang-hang modellek és a beszéd-szöveg (STT), nagy nyelvi modellek (LLM) és szöveg-hang (TTS) hagyományos csővezetékek között. Míg a Google és az OpenAI szolgáltatóktól származó hang-hang modellek natív hangfeldolgozást kínálnak, jelentősen kisebb kontextusablakkal rendelkeznek, amelyek a gyakorlatban gyakran körülbelül 3000 karakterre korlátozódnak. Például egy hang-hang modell nehezen tudja fenntartani az összefüggő hosszú beszélgetést egy LLM-alapú csővezetékhez képest.

A valós tesztek jelentős akadályokat fednek fel, például az emberi háttérzajt, amelyet az AI gyakran nem szűr ki olyan hatékonyan, mint az emberi agy. Továbbá specifikus nyelvi kihívások merülnek fel a magyar nyelv esetében, ahol az STT modellek 18-30%-os hibaarányt mutatnak telefonhívások során. Ezek a technikai korlátok azt jelentik, hogy bár a jelenlegi megoldások képesek egyszerű feladatok elvégzésére, például az ügyfélminősítés vagy az IVR útvonalválasztás terén, még nem képesek teljesen természetes, emberhez hasonló élményt nyújtani. Ezek az elemek együttesen azt jelzik, hogy az iparág jelenleg a hibrid modellek felé vezető átmeneti szakaszban van.

KIEMELT MEGÁLLAPÍTÁSOK

- A hang-hang modellek jelenleg súlyos kontextusablak-korlátokkal szembesülnek a szabványos LLM-ekhez képest.
- A telefonos környezetekben a magyar STT hibaarányai továbbra is magasak, elérve a 18% és 30% közötti értéket.
- Az emberi háttérzaj továbbra is kritikus hibahely a jelenlegi hangalapú botoknál, mivel hiányzik belőlük az emberi kognitív szűrés.

ÜZLETI HATÁS

Ezeknek a technológiáknak a gyakorlati alkalmazása jelenleg a nagy volumenű, alacsony komplexitású feladatoknál a leginkább életképes, például 50 000 potenciális ügyfél felhívása érdeklődés felmérésére vagy intelligens IVR útválasztás. A hang-hang modellek magas költsége miatt azonban jelentős üzleti döntés, hogy az AI vagy az emberi operátorok közül melyiket választják hosszú beszélgetésekhez. A vállalatok jelenleg olyan hibrid struktúrákat tudnak alkalmazni, ahol a TTS kezeli az első köszöntéseket, és a hang-hang modulok kezelik a hívás rövid kontextusú szegmenseit.

KÖVETKEZTETÉS

Andrejkovics Zoltán szerint az „ChatGPT-pillanat” még nem érkezett el az audio számára, mivel a jelenlegi rendszerek továbbra is igénylik az emberek beavatkozását a szünetek és a logika kezelésében. A jövő a hibrid modellekben rejlik, amelyek az LLM-ek hosszú kontextusú képességeit ötvözik a hang-hang rendszerek kifinomult audio-értésével. A beszélő arra ösztönzi a közönséget, hogy ismerje fel ezeket a jelenlegi korlátokat, miközben felkészül a kifinomultabb, gyártásra kész audio AI integrációjára.

Voicebots in the Real World: Lessons from Practice - lessons from real customer experiences

2026. 09. 07. 15:40

Prompt Arena

HUN · 20 perc



Zoltán Andrejkovics
Senior AI Expert · VCC Live

Zoltán Andrejkovics, Senior AI Expert at VCC Live, discusses the current state and practical limitations of voice-based AI in customer service environments.

WHAT WAS DISCUSSED

Zoltán Andrejkovics distinguishes between voice-to-voice models and traditional pipelines consisting of Speech-to-Text (STT), Large Language Models (LLM), and Text-to-Speech (TTS) components. While voice-to-voice models from providers like Google and OpenAI offer more natural interaction, they suffer from significantly smaller effective context windows, often struggling to manage more than 3,000 characters of prompt data. For example, a voice-to-voice model may fail to maintain complex instructions that a standard LLM would easily process.

The presentation highlights significant hurdles in real-world deployment, specifically regarding background noise and linguistic nuances in the Hungarian language. Zoltán Andrejkovics notes that AI often fails to filter out ambient sounds like street noise or television audio, which humans naturally ignore. Furthermore, STT models currently show high error rates of 18-30% for Hungarian phone calls, making it difficult for AI to accurately capture specific names or addresses. These factors combined indicate that while the technology is advancing, it is not yet at a level where it can perfectly replicate human conversational fluidity.

KEY TAKEAWAYS

- Voice-to-voice models currently have much smaller functional context windows compared to standard text-based LLMs.
- Background noise remains a critical failure point because many customers use older mobile devices lacking advanced active noise cancellation.
- Hungarian language processing in STT models currently faces an error rate of 18-30% in real-world telephone conditions.

BUSINESS IMPACT

The practical application of these technologies is currently limited to high-volume, low-complexity tasks such as initial interest screening for 50,000 leads or basic IVR routing. Zoltán Andrejkovics warns that the cost of using voice-to-voice models for long conversations can be prohibitively expensive, sometimes approaching the cost of human operators. Businesses must therefore balance the desire for a "magical" AI experience against the current technical limitations of prompt following and linguistic accuracy.

CONCLUSION

Zoltán Andrejkovics concludes that the industry is moving toward hybrid models that combine the long-context capabilities of LLMs with the nuanced audio processing of voice-to-voice systems. He suggests that while we are not yet at the "ChatGPT moment" for audio, these hybrid architectures will eventually provide the production-ready, natural experiences that customers expect.