**بسم الله الرحمن الرحیم**

موضوع: گزارش پروژه پیاده سازی تعامل صوتی با ChatGPT

استاد راهنما : سرکار خانم بهشید بهکمال

توسعه دهنده : رضا نصیری

فهرست

مقدمه

تکنولوژی های پروژه

- برنامه نویسی وب

- برنامه نویسی سرور

- سوکت

- ChatGPT

- تبدیل متن به صدا

- تبدیل صدا به متن

**مقدمه**

هوش مصنوعی این روزها به سراسر زندگی بشر نفوذ کرده است ؛ خودروهای خودران ، چت بات‌ها ، دستیارهای صوتی و ... نمونه

ای از این موارد می باشند چت‌ بات‌های هوش مصنوعی یکی از این موارد هستند که در آن انسان با یک سیستم مبتنی بر هوش

مصنوعی ارتباط برقرار می کند و سیستم به قدری هوشمندانه به سوالات کاربر پاسخ می دهد که کاربر به سختی می تواند

تشخیص دهد که طرف مقابل خود یک ربات می باشد.

تکنولوژی ChatGPT جدیدترین محصول در حوزه هوش مصنوعی می باشد که در دنیا فراگیر شده است و مردم را حیرت زده

کرده است . این ربات می تواند به تمام سوالات کاربران در تمام حوزه‌های مختلف مثل برنامه نویسی ، ریاضی ، تاریخی ، اطلاعات

عمومی و ... پاسخی جامع و کامل ارائه دهد.

این ربات که توسط شرکت OpenAI ساخته شده است در اخرین نسخه‌ی خود مبتنی بر مدل زبانی GPT 3.5 می باشد. این

مدل زبانی که بر پایه‌ی معماری شبکه های عصبی ساخته شده است قادر به پردازش زبان طبیعی و تولید متن می باشد.

اگر با ChatGPT کار کرده باشید خواهید فهمید که این بات مبتنی بر تعامل متنی می باشد و شاید شما علاقه ای به پرسیدن

سوالات خود به صورت متنی نداشته باشید و بخواهید که به صورت صوتی سوالات خود را بپرسید. سیستم پیاده سازی ای در این

پروژه قصد گزارش آن را داریم این امکان را به شما می دهد که سوالات خود را به صورت صوتی از ChatGPT بپرسید.

**ساختار پروژه**

این برنامه یک سیستم نرم افزاری تحت وب می باشد که بر پایه‌ی معماری client-server می باشد. معماری client-server

یک معماری رایج برای طراحی وب سایت و برنامه‌های کاربردی مبتنی بر وب می باشد که سمت client وظیفه‌ی ارتباط با کاربر و

سمت server وظیفه پردازش اطلاعات را بر عهده دارد.

فرایند کلی برنامه به این شکل است که کاربر زبان موردنظر خود را انتخاب می کند و صدای خود را ارسال می کند و سپس یک

پاسخ در قالب صوت دریافت می کند.

زبان های پشتیبانی شده عبارتند از:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| زبان | فرمت | زبان | فرمت | زبان | فرمت | زبان | فرمت |
| Arabic | AR | African | AF | English | EN | Persian | FA |
| Welsh | CY | Czech | CS | Catalan | CA | Bosnian | BS |
| Spanish | ES | Estonian | ET | Germany | DR | Danish | DA |
| Hindi | HI | French | FR | Finnish | FI | Greek | EL |
| Indonesian | ID | Italian | IT | Armenian | HY | Hungarian | HU |
| Polish | PL | Korean | KO | Kannada | KN | Japanese | JA |
| Latvian | LV | Latin | LA | Portuguese | PT | Malay | ML |
| Dutch | NL | Nepali | NE | Marathi | MR | Macedonian | MK |
| Turkish | TR | Romanian | RO | Russian | RU | Norwegian | NO |
| Urdo | UR | Swedish | SV | Slovak | SK | Serbian | SR |
|  |  |  |  |  |  | Chinese | ZH |

در ادامه به بررسی این دو بخش در پروژه موردنظر و تکنولوژی های به کار رفته در آن می پردازیم:

**قسمت server:**

قسمت سرور پروژه به کمک زبان پایتون پیاده سازی شده است . علت استفاده از پایتون ، پشتیبانی بالای کتابخانه‌های تبدیل متن

به صوت و صوت به متن از پایتون می باشد. تکنولوژی‌های به کار رفته در سمت سرور عبارتند از:

1.**تبدیل صوت به متن**:

تبدیل صوت به متن فرایندی است که در آن به کمک الگوریتم‌های پیچیده و شبکه‌های عصبی آموزش دیده شده صدای کاربر

تشخیص داده می شود و تبدیل به متن می شود و چون هدف این برنامه تعامل صوتی کاربر با ChatGPT می‌باشد بنابراین باید از

این تکنولوژی استفاده کنیم.

برای آن که بتوانیم صدا‌ی کاربر را به متن تبدیل کنیم از کتابخانه‌ی OpenAI و از ماژول whisper استفاده می کنیم.

تکنولوژی whisper یکی از جدیدترین مدل های تبدیل صوت به متن می باشد که توسط شرکت OpenAI توسعه داده شده

است. این مدل از اکثر زبان‌های انسانی و به خصوص زبان فارسی پشتیبانی می کند که علت اصلی استفاده از کتابخانه همین بوده

است. این کتابخانه از 7 فرمت صوتی پشتیبانی می‌کند که این فرمت ها عبارتند از:

mp3, mp4, mpeg, mpga, m4a, wav, webm

2.**تبدیل متن به صوت:**

در این تکنولوژی هدف ، این است که به کمک الگوریتم‌های پیچیده‌ی هوش مصنوعی ، یک صدای مصنوعی برای خواندن یک

متن را فراهم کنیم. از این تکنولوژی در این پروژه برای این استفاده می شود که جواب ChatGPT را به صورت صوتی به کاربر

بدهیم.

کتابخانه های و API های مختلفی به صورت رایگان و پولی برای انجام این کار وجود دارد که در نهایت 2 کتابخانه برای این کار

انتخاب شده است:

1.gTTS: سرویس تبدیل متن به صوت گوگل که برای زبان پایتون ارائه شده است اما متاسفانه از زبان فارسی پشتیبانی نمی‌کند.

2.سرویس ابری اریانا: برای پر کردن خلأ زبان فارسی از سرویس پولی اریانا استفاده می کنیم.

برای استفاده از این سرویس باید به سایت <https://asr-gooyesh.com/fa> مراجعه کنید و اقدام به ساخت اکانت و خرید

سرویس مورد‌نظر خود بکنید

**قسمت کلاینت:**

1.کتابخانه‌ی React: یکی از محبوب ترین کتابخانه های موجود برای طراحی رابط کاربری وب کتابخانه‌ی React می باشد.

این کتابخانه بر پایه Component ها می‌باشد برنامه نویسان می توانند با ساخت کامپوننت‌های مستقل مانند کامپوننت دکمه ،

ضبط صدا ، دریافت ورودی و ... کامپوننت‌های ترکیبی بسازند و با ترکیب این کامپوننت های ترکیبی می شود صفحات مختلف

یک وبسایت را ساخت.

2.فریمورک Remix: این فریمورک که بر پایه‌ی کتابخانه‌ی React ساخته شده است استفاده از React با فراهم کردن قابلیت

هایی مانند مسیریابی ، بهینه سازی و ... تسهیل کرده است.

3.کتابخانه‌ی Tailwindcss: