

مقدمه (Overview)

نرم افزار مطلوب یک ChatGPT اختصاصی برای سازمان است که رابط کاربری (UI) شخصی سازی شده دارد و از طریق API اختصاصی به مدل GPT OpenAI متصل می شود.

نرم افزار به صورت وب اپلیکیشن درون سازمانی (Intranet Web Application) پیاده سازی خواهد شد، به گونه ای که تنها کاربران متصل به شبکه داخلی شرکت قادر به دسترسی و استفاده از آن باشند.

این نرم افزار الزاما با Open WebUI پیاده سازی خواهد شد و تسلط و تجربه در آن شرط همکاری است.

محدوده (Scope)

نرم افزار در شبکه داخلی شرکت deploy می شود و از طریق زیردامنه ای که به آن تخصیص داده خواهد شد قابل دسترسی خواهد بود.

نرم افزار فاقد صفحات ورود یا ثبت نام است؛ کاربران متصل به شبکه سازمان با ورود کردن به سیستم هایی که سازمان در اختیار آنان قرار داده است، احراز هویت شده و می توانند به نرم افزار از طریق مرورگر دسترسی داشته باشند.

امکان گفت و گو درباره مسائل غیرمرتبط با سازمان و اطلاعات سازمانی ارائه شده وجود ندارد.

کاربران فقط می توانند سوالات عمومی مرتبط با سازمان را از طریق برنامه مطرح کرده و پاسخ دریافت نمایند. برای این منظور، سیستم از RAG (Retrieval-Augmented Generation) استفاده خواهد کرد تا اطلاعات سازمانی معتبر و به روز را بازیابی و در پاسخها لحاظ نماید.

کاربران تنها قادر به ارسال پیام متنی خواهند بود و ارسال صدا، تصویر یا فایل های دیگر امکان پذیر نیست.

در صورت امکان باید مکانیزمی تعبیه شود تا پیام‌های غیرمرتبط شناسایی و نادیده گرفته شوند و پیامی از سیستم به کاربر برای اطلاع او از غیرمرتبط بودن Prompt او ارسال شود؛ به عبارت دیگر پیام‌های نامربوط اصلاً نباید به مدل ارسال شوند.

اطلاعات سازمان بسته به خواست طراحان به صورت‌های گوناگون از جمله PDF، Excel، و Data Base ارائه می‌شوند که برای پیاده‌سازی RAG، استخراج و به صورت Vector Database ذخیره خواهند شد. این ساختار امکان بازیابی سریع و دقیق اطلاعات سازمانی را برای تولید پاسخ‌های مدل فراهم می‌کند.

الزامات فنی (Technical Requirements)

همانطور که ذکر شد، نرم‌افزار باید به وسیله **Open WebUI** پیاده‌سازی شود.

اطلاعات کاربران سازمان از Active Directory و به وسیله پروتکل LDAP قابل دسترسی است و کاربر بعد از ورود به سیستم خود که مجهز به سیستم عامل ویندوز است، می‌تواند بدون نیاز به ورود مجدد از وب‌سرویس مذکور استفاده کند.

درست مانند ChatGPT Web، نیاز است تا چت‌های مختلف هر کاربر ذخیره، نگهداری و هنگام استفاده از نرم‌افزار بازیابی شوند؛ کاربر می‌تواند به هر کدام از گفت‌وگوهای گذشته خود رفته و در آن صفحه با مدل گفت‌وگو کند و در عین حال، مدل پیام‌های گذشته آن صفحه گفت‌وگو را به خاطر دارد. به عبارت دیگر برنامه سمت سرور در هر پیغام API GPT باید تعدادی از پیام‌های سابق آن گفت‌وگو را، بسته به اندازه Context Window، ارسال کند.

برای سهولت استفاده از وب‌سرویس نیاز است تا برای هر بخش از اطلاعات سازمان و اطلاعات RAG شده مربوط به اطلاعات آن بخش، یک پایگاه داده مجزا در نظر گرفته شود. برای مثال فرض کنید نرم‌افزار مجهز به حالت‌های زیر است:

- سوالات اداری
- سوالات مالی

• سوالات ارتباطی

کاربر بسته به سوال خود، می‌تواند یکی از حالات ذکر شده را انتخاب کند. اگر کاربر سوال اداری داشته باشد، تنها پایگاه داده بخش اطلاعات اداری بررسی می‌شود؛ این امر بسیار بهتر از بررسی یک پایگاه داده بزرگ و مجتمع و یا استخراج اطلاعات نامرتبب احتمالی و ارسال آن به مدل GPT است، که در نهایت منجر به کاهش کیفیت یا خطا در پاسخ تولید شده می‌شود.

رابط و تجربه کاربری (UI/UX)

UI/UX نرم‌افزار مشابه با ChatGPT Web است اما با تفاوت‌هایی که در زیر لیست شده‌اند:

- لحاظ دکمه‌هایی بر رفتن به حالت‌های مختلف ذکر شده در بخش الزامات فنی.
- عدم قرار دادن بخش‌هایی که مربوط به بارگذاری هرگونه فایلی است.
- عدم قرار دادن بخش‌های نامرتبب به نیازمندی‌های ذکر شده مانند دکمه Share.

نمونه‌های مشابه (Samples)

نمونه‌های زیر مثال‌های بسیار خوبی برای درک انتظارات از نرم‌افزار مطلوب هستند:

[موچت \(muchat\)](#)

[گپ‌جی‌پی‌تی \(GapGPT\)](#)