

عامل	توضیحات	ابزارها / تکنولوژی ها
کاربر (User)	فردی که با سیستم تعامل می کند و فرمان های صوتی به سیستم می دهد	میکروفون، رابط کاربری GUI (Interface Voice)
دستیار صوتی (VoiceAssistant)	سیستم هوشمندی که فرمان های کاربر را دریافت و پردازش می کند و به آن پاسخ می دهد	ابزارهای پردازش زبان طبیعی (NLP)، سیستم های مبتنی بر یادگیری ماشین
تشخیص گفتار (Speech Recognition)	عامل تبدیل صدای کاربر به متن جهت پردازش توسط سیستم	Microsoft Speech API، Google Speech API، Azure Speech Service، یا سایر API های تشخیص گفتار
پردازش زبان طبیعی (NLP)	وظیفه درک معنای فرمان های کاربر از طریق پردازش زبان طبیعی	مدل های ترانسفورمر مانند: SpaCy، NLTK، GPT-3، BERT
درک هدف (Intent Recognition)	بخش از سیستم که هدف فرمان کاربر را شناسایی می کند	مدل های یادگیری ماشین مانند Rasa، BERT
مدیریت مکالمه (Dialogue Management)	کنترل جریان مکالمه بین کاربر و سیستم، مدیریت مکالمات چندمرحله ای	BotPress، Rasa، طراحی های سفارش
تولید پاسخ (Response Generation)	تولید پاسخ (Response Generation)	مدل های ترانسفورمر مثل GPT-3 یا T5، الگوهای تولید متن
تبدیل متن به گفتار (Text-to-Speech)	تبدیل پاسخ متنی تولید شده به صدای قابل پخش برای کاربر	Google Text-to-Speech API، Microsoft Azure Text-to-Speech
پایگاه دانش (Knowledge Base)	دیتابیس یا منبع اطلاعاتی که دستیار برای پاسخ گویی به سوالات عمومی یا خاص از آن استفاده می کند	پایگاه های داده، API های جستجو مانند WolframAlpha، Wikipedia API
محیط یادگیری (Training Environment)	محیطی که در آن سیستم آموزش می بیند و مدل ها بهبود پیدا می کنند	Jupyter، Google Colab، TensorFlow، Notebook، PyTorch
پایگاه داده (Database)	برای ذخیره اطلاعات، تاریخچه مکالمات یا داده های آموزشی مدل ها	MongoDB، MySQL، PostgreSQL
زیرساخت سرور (Server Infrastructure)	بخش پردازش درخواست های کاربر و اجرای مدل های NLP و تشخیص گفتار	Google، AWS، Django، Flask، Microsoft Azure، Cloud
واسط کاربری (User Interface)	نمایشگر یا رابط صوتی برای نمایش پاسخ ها یا دریافت فرمان ها	واسط کاربری گرافیکی (GUI) یا Voice User Interface (VUI)
ابزارهای نظارت (Monitoring Tools)	ابزارهایی برای نظارت و بررسی عملکرد سیستم و عیب یابی مشکلات	Prometheus، Kibana، Grafana

1- تشخیص گفتار (Speech Recognition)

- ابزار Google Speech-to-Text API: یا ابزارهای مشابه) مانند Microsoft Azure Speech Service یا Wit.ai
- توضیح: این بخش وظیفه تبدیل صدای کاربر به متن را برعهده دارد. برای این کار می‌توانید از API‌های آماده مانند Google Speech API استفاده کنید که به صورت لحظه‌ای صدای کاربر را به متن تبدیل می‌کند.

2- پردازش زبان طبیعی (NLP)

- ابزارها:

- مدل‌های ترانسفورمر: مانند BERT، GPT (Generative Pre-trained Transformer)، و T5 برای درک بهتر فرمان‌های کاربر و تولید پاسخ.
- کتابخانه‌ها: NLTK (Natural Language Toolkit)، SpaCy برای تجزیه و تحلیل متن و استخراج اطلاعات معنایی.
- توضیح: پس از دریافت متن از سیستم تشخیص گفتار، این متن توسط مدل‌های NLP پردازش می‌شود تا معنی و هدف فرمان کاربر درک شود. مدل‌های ترانسفورمر، به‌ویژه BERT، در انجام وظایف مختلف NLP مانند درک متن، طبقه‌بندی جملات، و پاسخ به سوالات بسیار کارآمد هستند.

3- سیستم مدیریت مکالمه (Dialogue Management)

- ابزارها: Rasa، BotPress یا طراحی یک سیستم سفارشی.
- توضیح: این بخش مسئول مدیریت جریان مکالمه است و تعیین می‌کند که سیستم چه زمانی باید پاسخ دهد یا درخواست اطلاعات بیشتری از کاربر کند. می‌توان از چارچوب‌های آماده مثل Rasa برای این منظور استفاده کرد که قابلیت‌هایی برای مدیریت مکالمات پیچیده و مکالمات چندمرحله‌ای دارند.

4- پاسخ‌گویی و تولید متن (Text Generation)

- ابزار: مدل‌های ترانسفورمر مانند GPT یا T5 برای تولید پاسخ‌های طبیعی و بهینه.
- توضیح: پس از پردازش فرمان کاربر، این بخش وظیفه تولید پاسخ مناسب را دارد. مدل‌های GPT در تولید متن طبیعی بسیار موفق عمل می‌کنند.

5- تبدیل متن به گفتار (Text-to-Speech)

- ابزار Google Text-to-Speech API: یا ابزارهایی مثل Microsoft Azure Text-to-Speech.
- توضیح: بعد از تولید پاسخ متنی، برای بازگرداندن پاسخ به کاربر به صورت صوتی، از این بخش استفاده می‌شود.

مراحل کلی پیاده‌سازی:

1. دریافت صدا از کاربر: استفاده از میکروفون برای دریافت صدای کاربر.
2. تبدیل صدا به متن: استفاده از Google Speech API برای تبدیل صدای کاربر به متن.
3. پردازش متن و درک فرمان: استفاده از مدل‌های NLP مانند BERT برای تجزیه و تحلیل متن و فهمیدن معنای فرمان.
4. تولید پاسخ: تولید پاسخ به کاربر با استفاده از مدل‌های تولید متن مثل GPT.
5. تبدیل متن به صدا: استفاده از API تبدیل متن به صدا برای بازگرداندن پاسخ به صورت صوتی به کاربر.

نکات مهم:

- دقت در تشخیص گفتار: استفاده از مدل‌های به‌روز و داده‌های آموزش مناسب.
- سفارشی‌سازی مدل‌های NLP با توجه به دامنه کاربرد دستیار صوتی، مدل‌ها باید برای وظایف خاصی مثل پاسخ به سوالات یا انجام فرمان‌های خاص آموزش داده شوند.
- بهینه‌سازی مکالمات: سیستم باید بتواند مکالمات پیچیده و چندمرحله‌ای را به خوبی مدیریت کند.