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

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.**

1. **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.**

1. **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.**

1. **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.**

1. **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.**

**آیا قصد دارید بر روی یک کاربرد خاص (مانند دستیار شخصی یا کاربری‌های دیگر) کار کنید؟**