For a **Doctor Appointment Chatbot**, the model we choose should handle tasks like appointment booking, symptom checking, and providing general health advice. Here’s a ranking of models based on accuracy and performance for such tasks:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Architecture** | **Accuracy Range (%)** | **Key Strengths** |
| **GPT-4 (OpenAI)** | Transformer | **90–95%** | Best for natural conversations, context retention |
| **Gemini Pro (Google** | Transformer | **88–94%** | Great for multilingual and healthcare Q&A |
| **LLaMA 2 (Meta)** | Transformer | **85–92%** | Open-source, adaptable for medical domains |
| **BioBERT (Fine-tuned)** | Transformer | **83–90%** | Specialized in biomedical text understanding |
| **Claude (Anthropic)** | Transformer | **80–88%** | Good for task-based conversations |
| **Rasa (DIET model)** | Hybrid NLU | **75–85%** | Open-source, customizable with healthcare intents |
| **Dialogflow CX** | Proprietary | **70–80%** | Easy-to-use, Google Cloud integration |

* **Accuracy Insights**
* **GPT-4**: Achieves up to **95%** accuracy in structured medical dialogues with sufficient domain-specific training.
* **BioBERT**: Fine-tuning on medical datasets like **MIMIC-III** can boost performance to **90%**.
* **Rasa DIET**: Performance depends heavily on well-labeled intents and entities.
* **Model Selection Guidelines**
* For cloud-based and multilingual capabilities → **GPT-4** or **Gemini Pro**
* For privacy and customization → **LLaMA 2** or **BioBERT**
* For cost-effective, open-source solutions → **Rasa DIET**