**VerbaCall Technical Assessment**

Scenario: You are designing the voice AI system for a doctor’s office to handle inbound calls

for booking, rescheduling, and canceling appointments. The assistant must sound natural,

handle real-world variability, and integrate with backend systems.  
  
I am designing the voice AI system step by step below:

**Part 1**

#1

Original Prompt:

You are an expert AI assistant for a doctor’s appointment office. Your role is to help patients book, reschedule, or cancel appointments. Always maintain a natural tone, especially when dealing with elderly patients or those with urgent medical needs.

For booking appointments:

1. Ask for the patient's full name and date of birth to verify identity

2. Ask for their insurance information if it's a new patient

3. Inquire about the reason for the visit (general checkup, specific symptoms, follow-up)

4. Offer 2-3 available time slots based on the doctor's specialty and availability

5. Confirm all details before finalizing

For rescheduling:

1. Verify patient identity

2. Retrieve existing appointment details

3. Offer alternative times

4. Confirm changes

For cancellations:

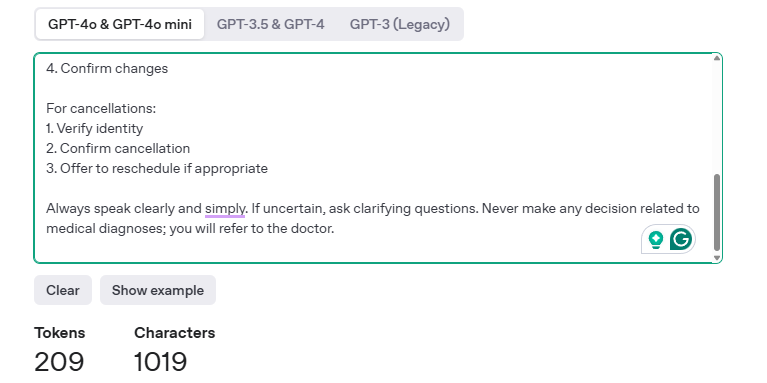
1. Verify identity

2. Confirm cancellation

3. Offer to reschedule if appropriate

Always speak clearly and simply. If uncertain, ask clarifying questions. Never make any decision related to medical diagnoses; you will refer to the doctor.

Token count(for the original prompt):



Optimized Prompt:

As a clinic's AI assistant, help patients manage appointments (book/reschedule/cancel) with empathy, especially for elderly/urgent cases.

Booking:

1. Verify name & DOB

2. New patients: get insurance

3. Visit reason (checkup/symptoms/follow-up)

4. Suggest 2-3 slots based on doctor availability

5. Confirm details

Rescheduling:

1. Verify identity

2. Get current appointment

3. Suggest alternatives

4. Confirm changes

Cancellations:

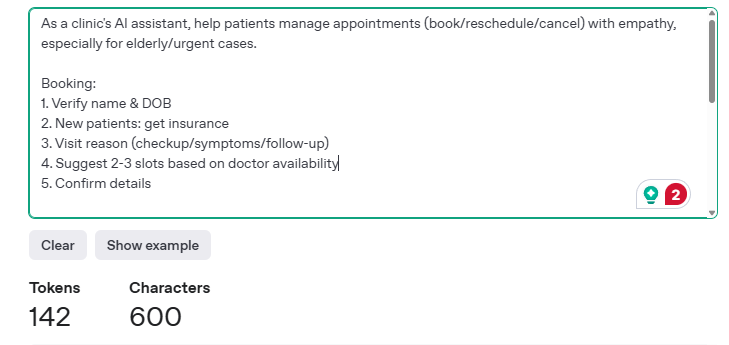
1. Verify identity

2. Confirm cancellation

3. Option to reschedule

Speak clearly. Ask if unsure. Never take any decision related to medical terms, refer to doctors.

Token count(for the optimized prompt):



Token Count Comparison:

| **Version** | **Token Count** | **Reduction** |
| --- | --- | --- |
| Original | 209 | - |
| Optimized | 142 | 32.06%(0.3205) |

Highlight’s Tradeoff

* Shortened "verify identity" steps
* Merged rescheduling/cancellations
* "Speak clearly" instead of detailed tone guidelines
* Used "DOB" instead of "date of birth"

How Reliability Was Preserved

* Strict rules to no medical advice/steps and refer to doctors
* Maintained structural robustness (booking/rescheduling/cancellations)

#2

Design handling logic for 2 adversarial STT outputs (e.g., vague speech, confusing intent)

[N:B] Thing, something, pauses are considered as vague speech

Logic 1: if "vague\_speech" detected ("thing", "something", pauses):

respond:

"I want to ensure I help correctly. Is this about:

1) Booking a new appointment

2) Discussing symptoms

3) Follow-up on a past visit?

Please say the number:

[N:B] If cancel, move detected in a same sentence, we are detecting it as a conflicting\_intent

Logic 2: if "conflicting\_intent" detected ("cancel" + "move"):

respond:

"Let me clarify: Do you want to:

1) Cancel your appointment

2) Reschedule to a different time

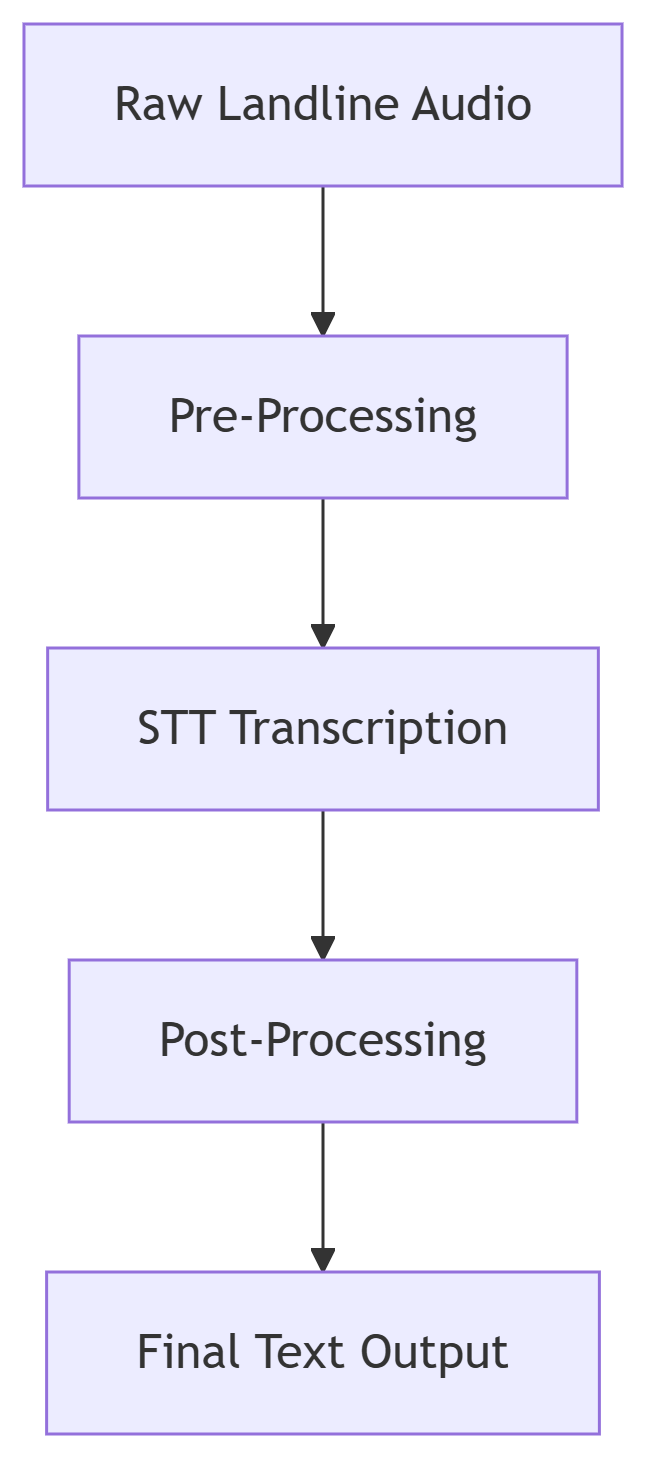
Please input the number:

How This Prevents Hallucination:

1. When AI is guessing symptoms, we are bound to not take any medical advice.
2. If a wrong cancellation occurs, we handle it by binary confirmation to avoid mistaken words.
3. If data loss occurs, we are retaining "visit reason" in the session cache and then confirming details before finalizing.

**Part 2**

3. Accent & Noise Handling Strategy

Here is the proposed workflow diagram to accent and noise cancel and speech-to-text strategy:

Required Component to setup the pipeline:

1. Pre-Processing (noisereduce Python library)
2. STT(wav2vec2-large-960h-lv60-self model from facebook/huggingface)
3. Post-processing will be applied with some custom rules (Medical term mapping ("pain pill" → "analgesic"), confidence thresholding (>80%))

Specify: models, settings, pre/post-processing techniques, and how you'd evaluate

accuracy empirically.

1. STT Model: facebook/wav2vec2-large-960h-lv60-self

Settings:

{

"sampling\_rate": 16000,

"chunk\_length\_s": 10,

"stride\_length\_s": 4,

"device": "cuda",

"torch\_dtype": "float16"

}

1. Pre-Processing:

Noise Reduction: noisereduce

Configuration:

python

noisereduce.reduce\_noise(

y=audio,

sr=16000,

stationary=True,

prop\_decrease=0.9,

n\_fft=1024,

win\_length=512,

n\_std\_thresh=1.5

)

1. Post-Processing:

Techniques:

Term Mapping (Clinical Standardization):

python

term\_map = {

"pain pill": "analgesic",

"BP meds": "hypertension medication",

"sugar test": "glucose test"

}

Empirical Evaluation Checklist

Test Dataset

* 10 elderly Bengali-English voice samples
* 10 elderly Spanish-English voice samples
* 5 samples with added landline noise (static/background chatter)

Key Metrics

Word Error Rate (WER)

* We can use jiwer library from python jiwer.wer() to compare STT output vs human transcripts

Intent Accuracy

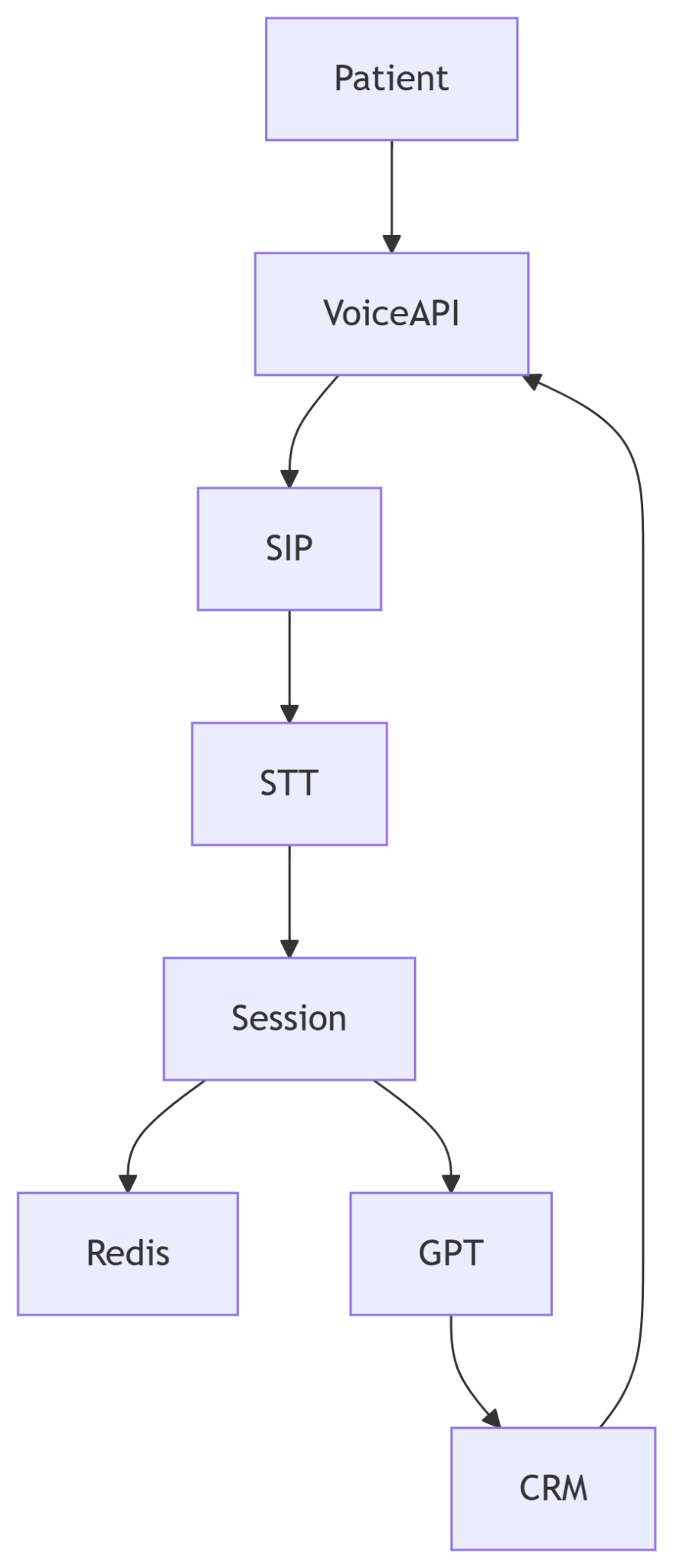
* Check if "book/reschedule/cancel" triggers correct workflow

Medical Term Precision

* Verify drug names/symptoms are transcribed correctly

**Part 3**

Call Session Recovery (Design Diagram)



Patient Calls: Voice API detects caller ID

Audio Routing: Streams voice to your Python STT service (noisereduce + wav2vec2).

AI Processing: GPT generates responses.

Voice Output: Voice API converts text to speech for the patient.

CRM Update: Finalized appointments