Flow Diagram

```
                          ┌──────────────┐
                          │  Client UI   │
                          └──────┬───────┘
                                 │
                                 ▼
                          ┌──────────────┐
                          │ WebSocket/   │
                          │    ask       │
                          └──┬────┬────┬─┘
                 ┌───────────┘    │    └───────────┐
                 ▼                ▼                ▼
          ┌────────────┐   ┌────────────┐   ┌────────────┐
          │   Audio    │   │    Text    │   │   Image    │
          │  Handling  │   │  Handling  │   │ Generation │
          └─────┬──────┘   └─────┬──────┘   └─────┬──────┘
                │                │                │
                ▼                ▼                ▼
          ┌────────────┐   ┌────────────┐   ┌────────────┐
          │   FFmpeg   │   │   Gemini   │   │  Hugging   │
          │            │   │  AI(Text)  │   │  Face API  │
          └─────┬──────┘   └─────┬──────┘   └─────┬──────┘
                │                │                │
                ▼                ▼                │
          ┌────────────┐   ┌────────────┐         │
          │  Whisper   │   │  TTS(Text) │         │
          └─────┬──────┘   └─────┬──────┘         │
                │                │                │
                ▼                │                │
          ┌────────────┐         │                │
          │   Gemini   │         │                │
          │  AI(Audio) │         │                │
          └─────┬──────┘         │                │
                │                │                │
                ▼                │                │
          ┌────────────┐         │                │
          │ TTS(Audio) │         │                │
          └─────┬──────┘         │                │
                └────────┐       │       ┌────────┘
                         ▼       ▼       ▼
                       ┌──────────────┐
                       │    JSON      │
                       │  Response    │
                       └──────┬───────┘
                              │
                              ▼
                       ┌──────────────┐
                       │  Client UI   │
                       │   Output     │
                       └──────────────┘
```
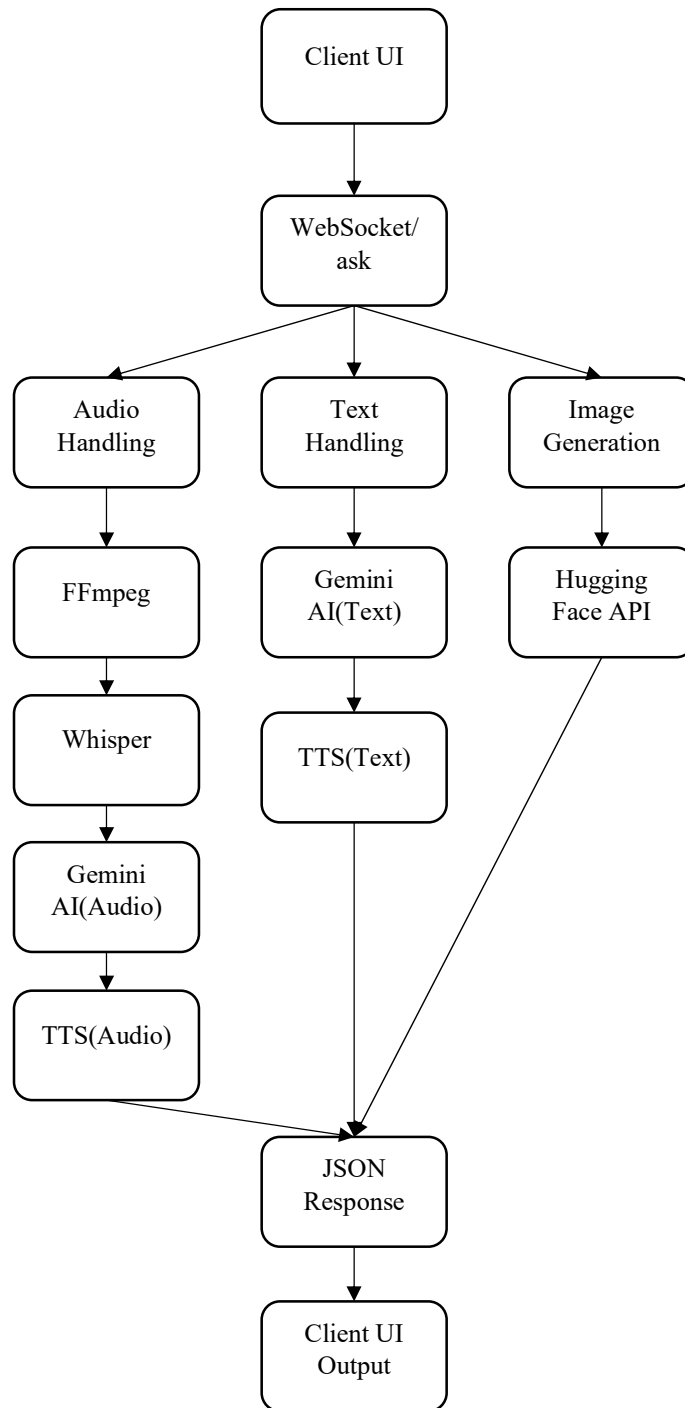
Test Scenarios

Here are structured test cases you should run:

1. WebSocket Audio Flow

- **Scenario**: Send valid audio stream via WebSocket.
- **Expected Result**:
    - Audio saved as. Webm → converted to .wav.
    - Whisper transcribes correctly.
    - Gemini returns text answer.
    - TTS generates MP3.
    - Hugging Face returns image.
    - JSON response includes transcription, answer, audio URL, image URL.
- **Edge Case**: Send >5MB audio.
    - **Expected**: Connection closed with warning.
- **Edge Case**: Send no audio.
    - **Expected**: Response includes {"error": "No audio received."}.

2. Text Question via /ask

- **Scenario**: POST form with question="What is AI?".
- **Expected Result**:
    - Gemini returns answer.
    - TTS generates MP3.
    - Hugging Face generates image.
    - JSON response includes all fields.
- **Edge Case**: Empty question.
    - **Expected**: Error 500 or validation error.
- **Edge Case**: Malformed JSON (if you extend to JSON input).
    - **Expected**: Error 400.

3. TTS Generation

- **Scenario**: Generate audio for a short answer.
- **Expected Result**: MP3 file saved at /static/live_answer.mp3.
- **Edge Case**: Long text (>5000 chars).
    - **Expected**: Engine still generates audio, but performance may degrade.

4. Image Generation

- **Scenario**: Prompt "diagram of neural network".
- **Expected Result**: Hugging Face returns PNG/JPEG.
- **Edge Case**: HF_API_TOKEN missing.

- o **Expected**: Fallback placeholder image returned.
- **Edge Case**: Hugging Face timeout.
    - o **Expected**: Placeholder returned.

5. Error Handling

- **Scenario**: FFmpeg not installed.
- **Expected Result**: WebSocket returns error JSON with "FFmpeg failed" message.
- **Scenario**: Whisper model fails.
    - o **Expected**: "No clear speech detected." returned.

6. Routes

- **Scenario**: GET /static/live_answer.mp3 before generation.
    - o **Expected**: 404 error.
- **Scenario**: GET /favicon.ico when file exists.
    - o **Expected**: Returns icon.
- **Scenario**: GET /favicon.ico when file missing.
    - o **Expected**: Returns 204 empty response.