

A close-up photograph of two young girls with light brown hair, wearing white polo shirts with horse logos. They are positioned side-by-side, facing each other and whispering. The girl on the left has her hand near her mouth. The background is dark and out of focus.

whispering to OpenAI

Introduction



Daniël Paulus

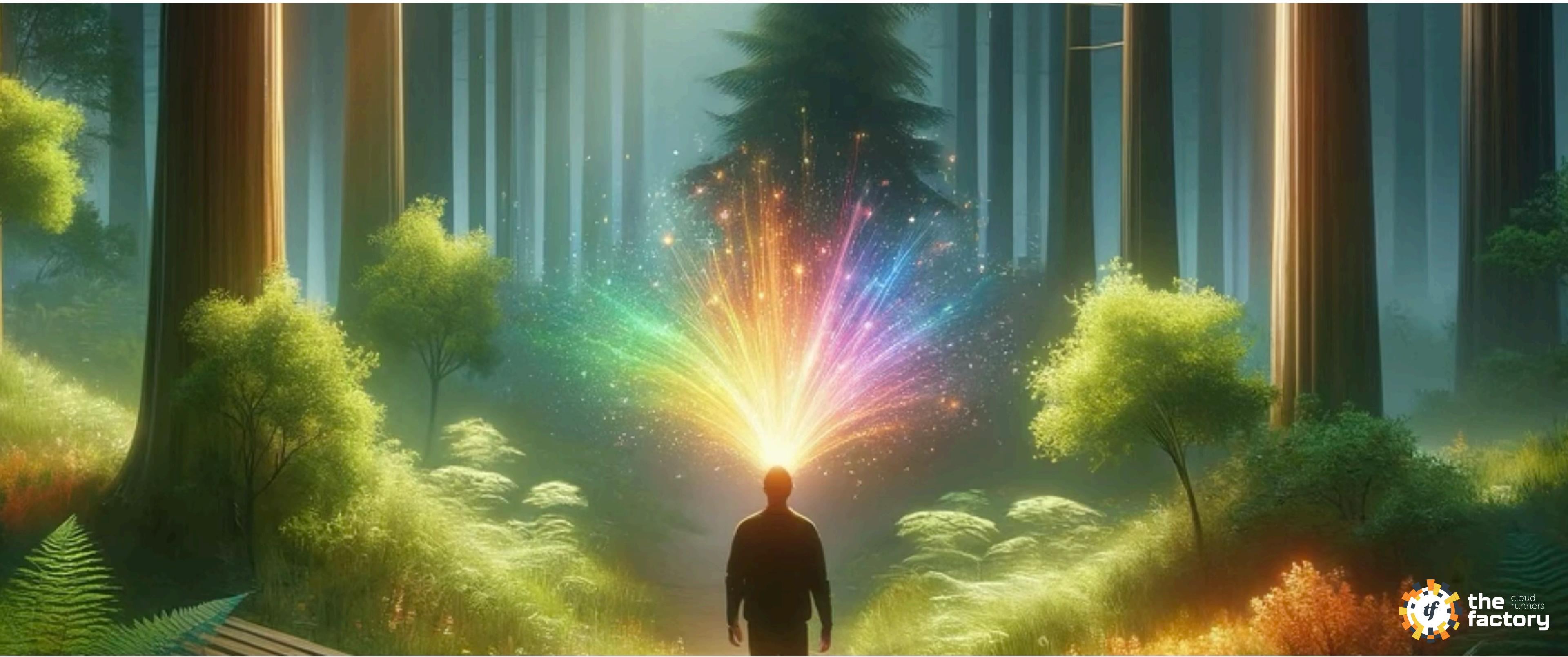
Head of Azure Solutions
The Factory



A Personal Quest



A Moment of Inspiration



The Challenge of Voice Memos



A photograph of a man with dark hair and a beard, wearing black-rimmed glasses and a red, white, and blue plaid shirt. He is leaning forward over a wooden desk, resting his head on his hand in a gesture of stress or exhaustion. On the desk in front of him is a laptop, an open notebook with a pen, and a large silver mug. The background is blurred, showing what appears to be a workshop or garage environment with various tools and equipment visible.

Unveiling the Challenge

A Moment of Inspiration

- *Directness*: Voice memos capture immediate, unfiltered ideas and emotions more fluidly than written text.
- *Pause and Pace*: Speech pace and pauses reveal a speaker's thoughts and emotions, offering insights into their mindset.
- *Ease of Use*: Voice memos offer a convenient, hands-free way to record thoughts and emotions on-the-go, avoiding the disruption of writing.
- *Inclusive*: Voice memos are a vital tool for those who struggle with typing, such as individuals with dyslexia or visual impairments.
- *Brainstorming*: Voice memos foster creative thinking by allowing ideas to flow without worrying about grammar or structure.

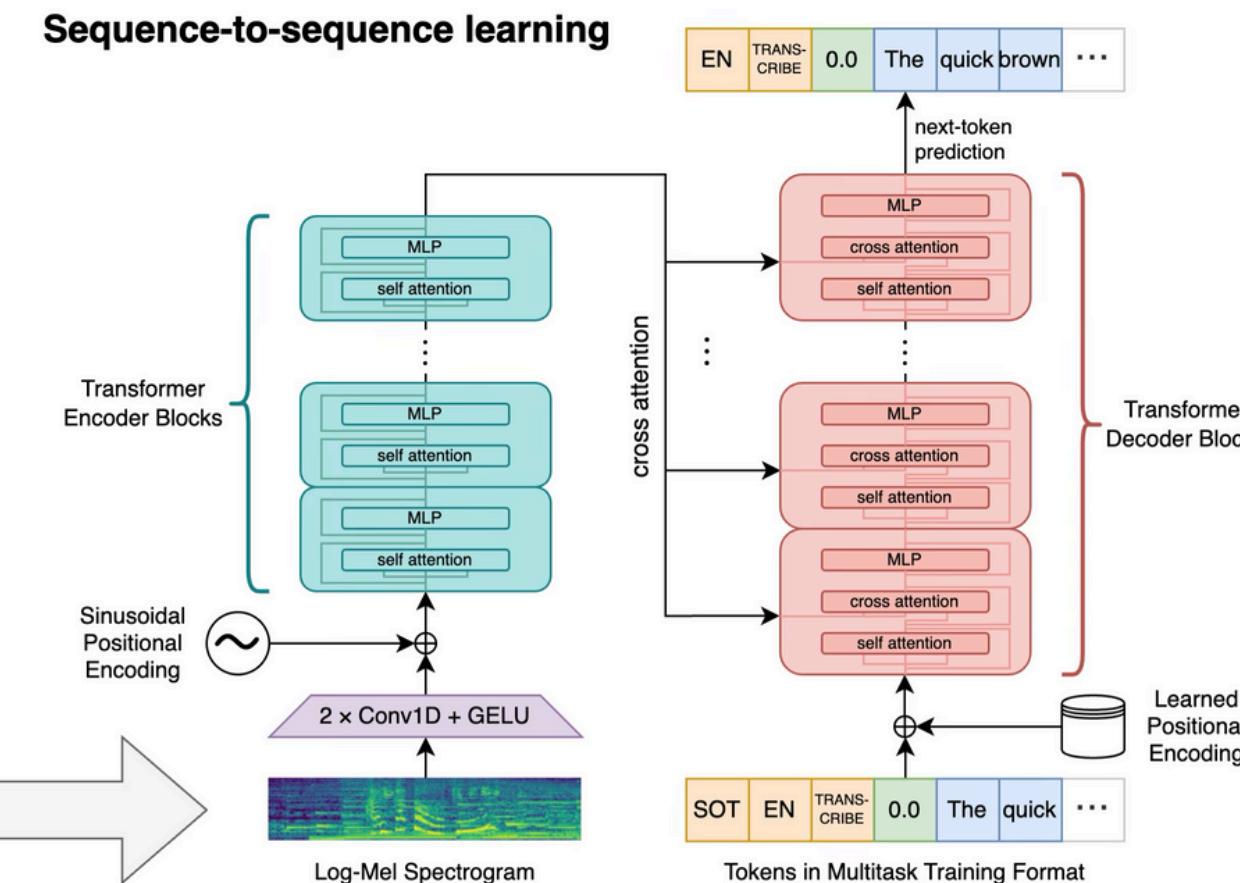
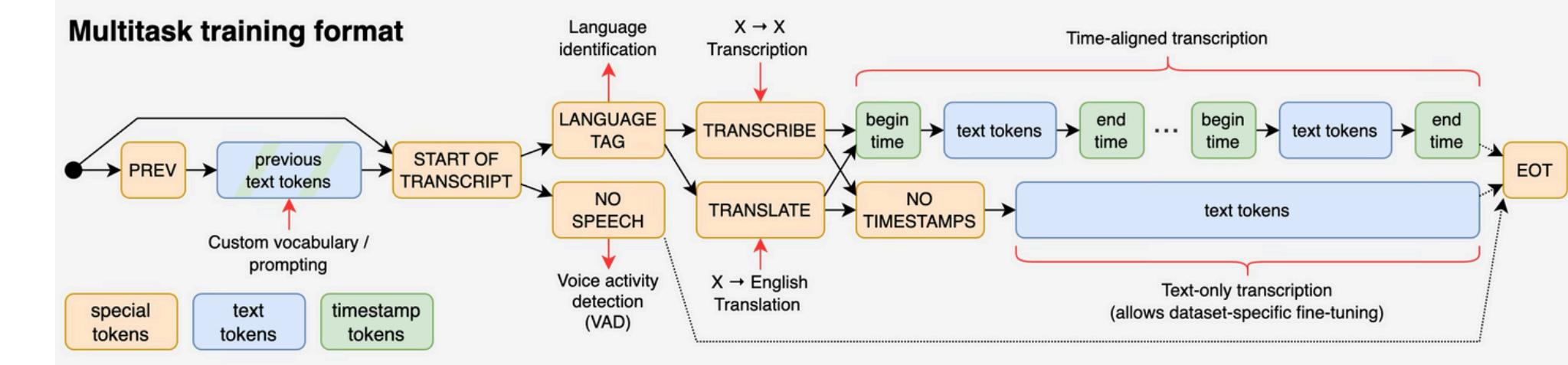
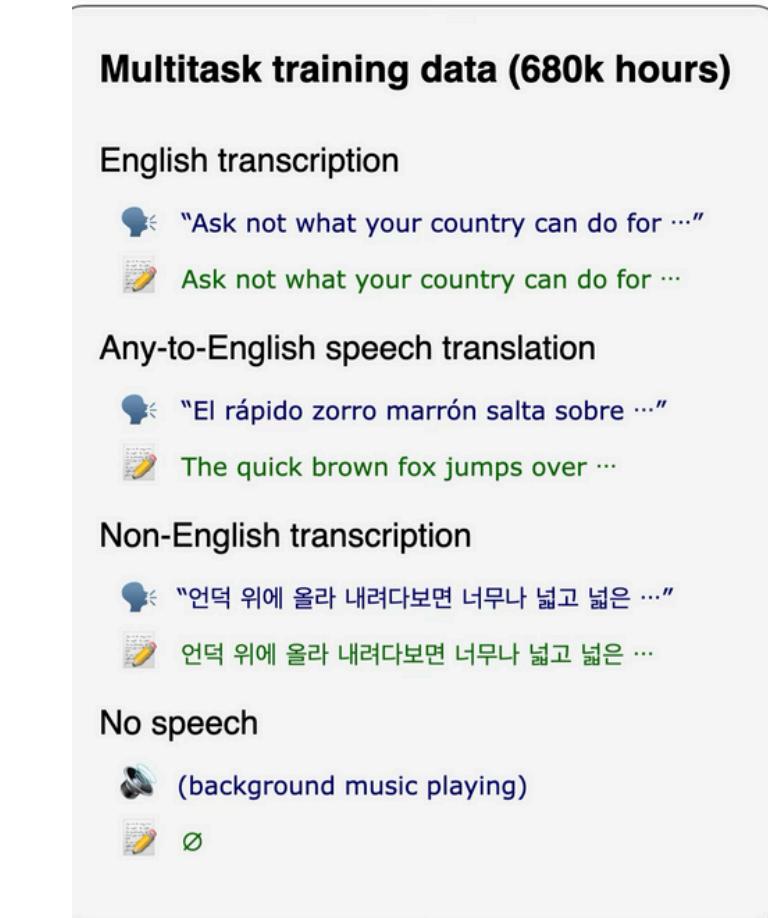
Encountering a Barrier

- *Quick Capture:* Voice memos capture fleeting thoughts, building up a substantial audio archive over time.
- *Lack of Metadata:* Voice memos often miss tags or categories, which makes systematic organization challenging.
- *Underutilized Information:* Valuable insights in voice memos often go unused due to their difficult-to-search nature.
- *Missed Opportunities:* The difficulty in finding past insights in voice memos can result in overlooked opportunities.
- *Physical Library Analogy:* Searching through unlabeled voice memos is like looking for a specific untitled book in a disorganized library.

Discovering Whisper

Introducing Whisper

- Automatic Speech Recognition (ASR) system
- Designed to convert spoken language into written text
- Multilingual Capabilities
- Open-Source Availability
- Works in Varied Conditions
- Advancing Accessibility
- Cost-Effectiveness
- Support for Complex Tasks



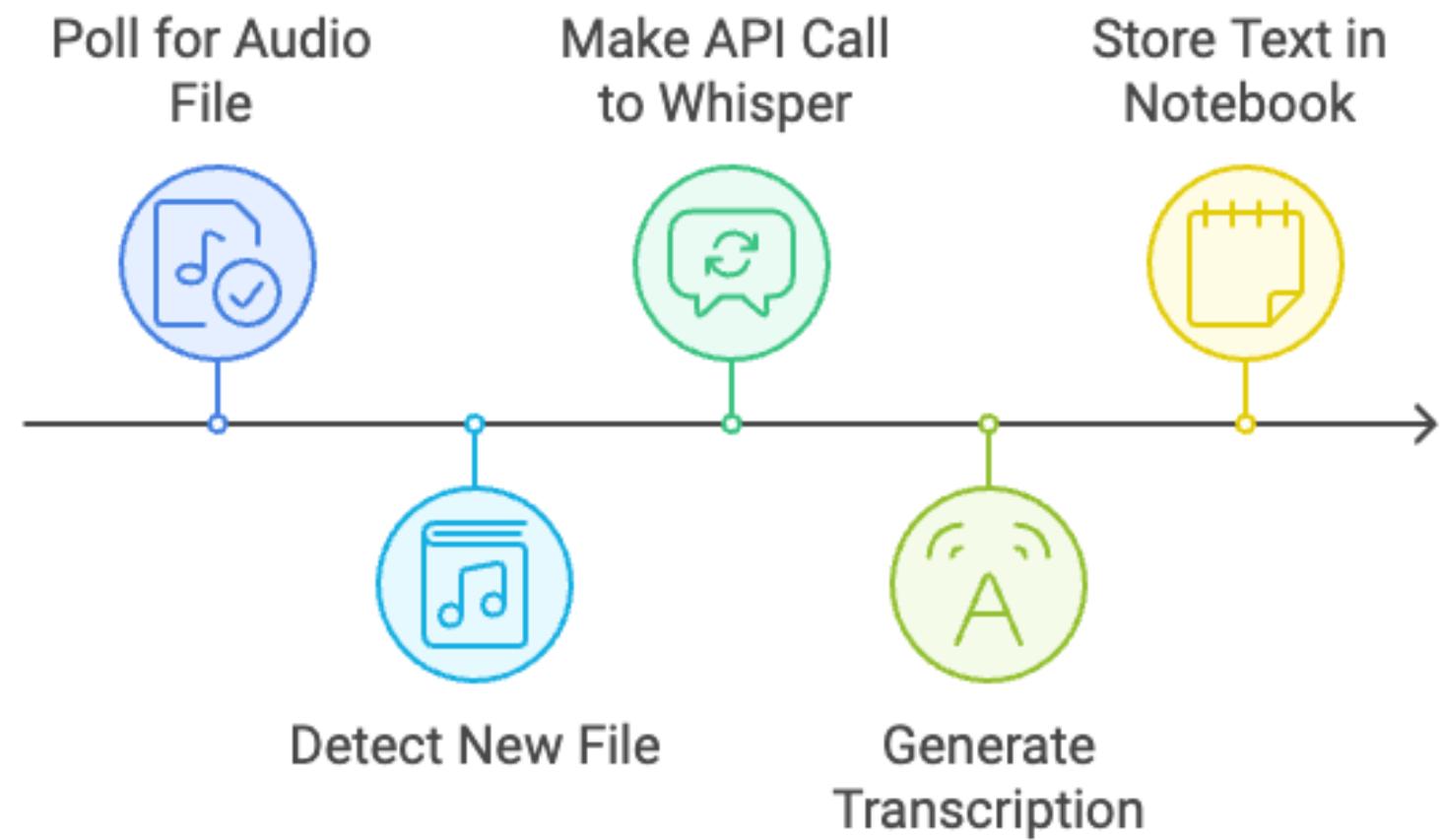
The Integration Journey

The power of API

```
curl -v 'https://testi7440-whisper.openai.azure.com/openai/deployments/whisper/audio/transcriptions?api-version=2023-09-01-preview' \
-H "api-key: b07d876349723463388e" \
-H "Content-Type: multipart/form-data" \
-F file="@test.m4a"
```

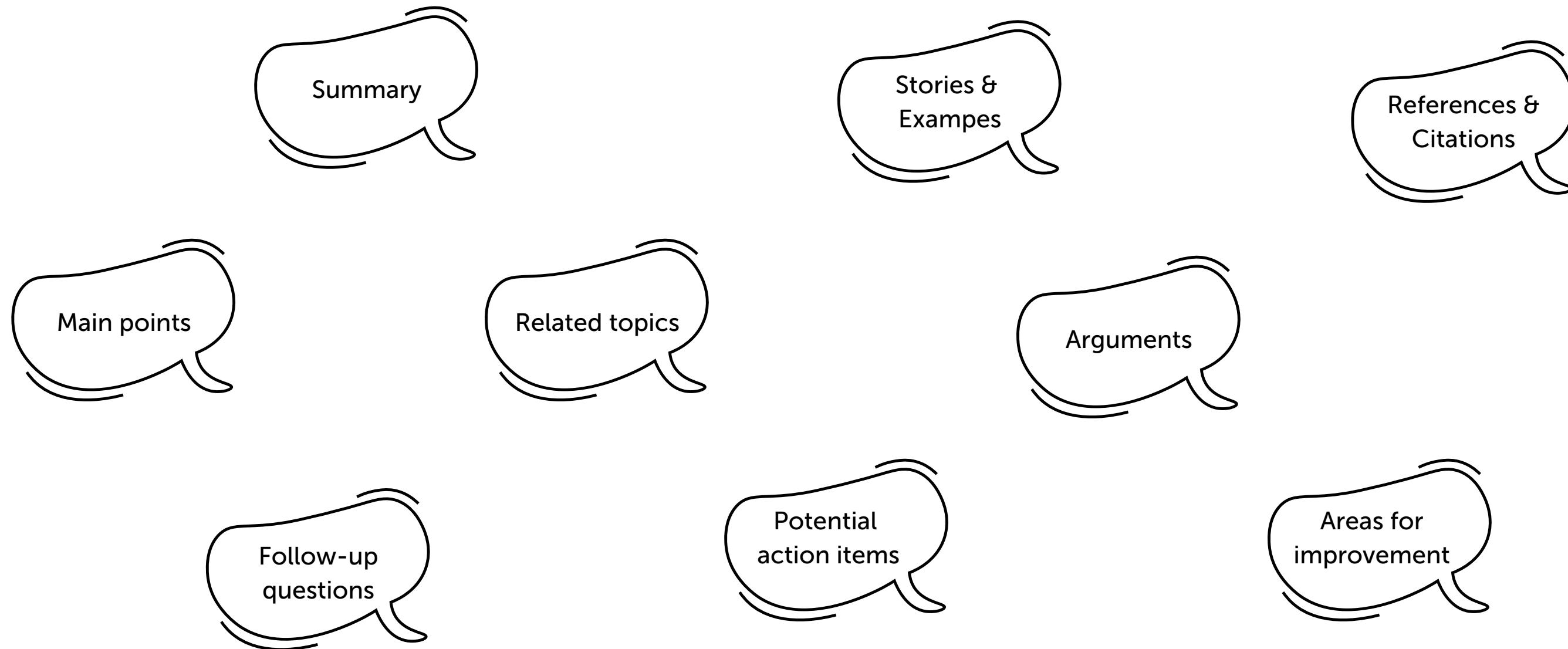
```
{
  "text": "I'm going to test voice memos to parse with Whisper. That way I can express the ideas I
have. While I'm driving home or taking a walk, for instance. Quite a nice idea to do it like that."
}
```

The Mechanism



From Transcription to Actions

Beyond Transcription



Azure OpenAI



```
curl 'https://testi7440-gpt4.openai.azure.com/openai/deployments/gpt-4/chat/completions?api-version=2023-05-15' \
  -H "Content-Type: application/json" \
  -H "api-key: b07d876349723463388e" \
  -d '{"messages": [{"role": "system", "content": "You are a helpful assistant."}, {"role": "user", "content": "Does Azure OpenAI support customer managed keys?"}, {"role": "assistant", "content": "Yes, customer managed keys are supported by Azure OpenAI."}, {"role": "user", "content": "Do other Azure AI services support this too?"}]}'
```

System Message

You are an assistant that summarizes voice notes and other audio recordings that primarily involve human speech. If the speaker in a transcript identifies themselves, use their name in your summary content instead of writing generic terms like "the speaker". If they do not, you can write "the speaker".

Analyze the transcript provided, then provide the following: create a summary that is roughly 15% of the length of the transcript.

"main_points" - add the main points. Limit each item to 100 words and limit the list to 5 items.

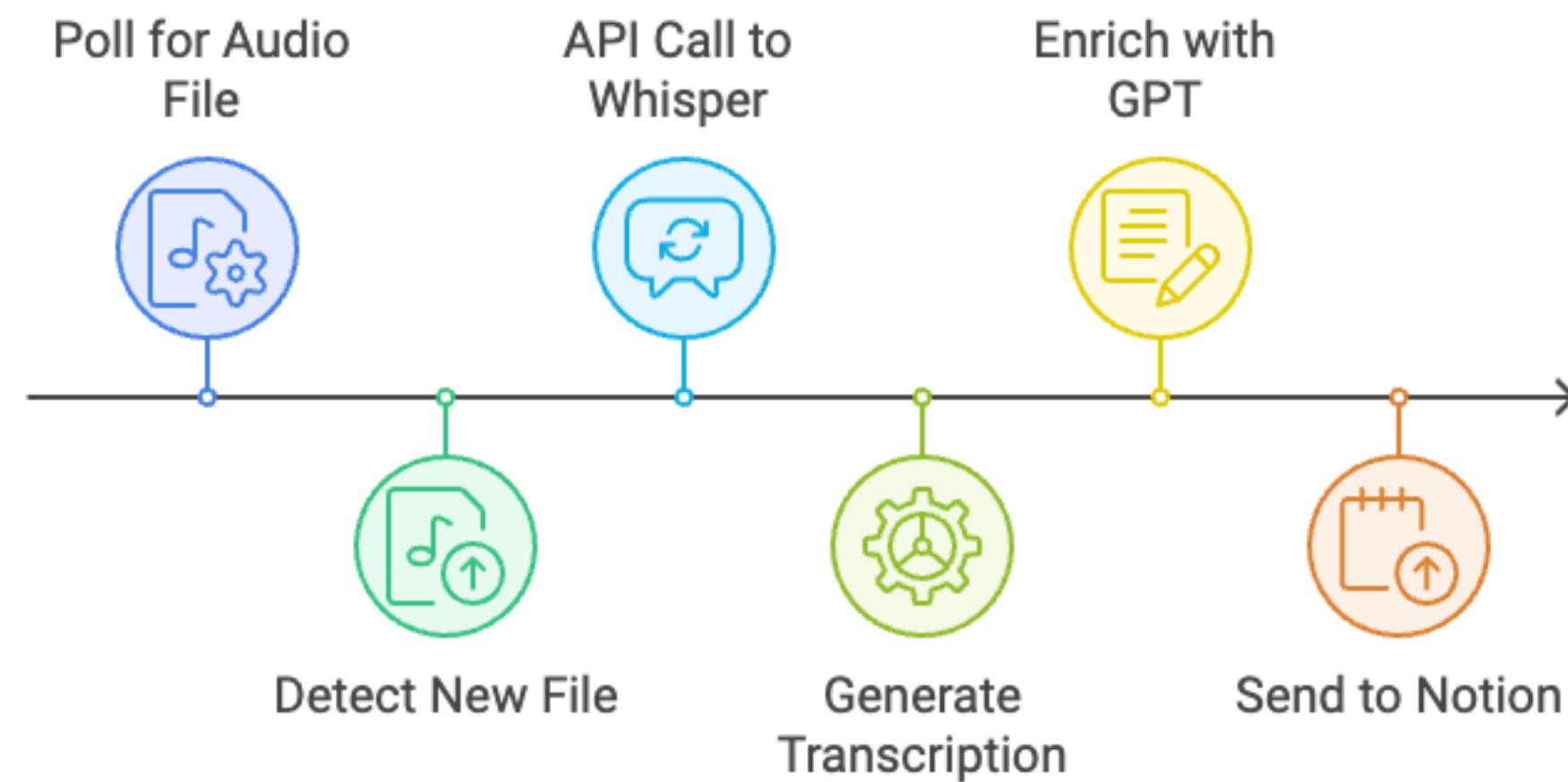
"action_items:" - add action items. Limit each item to 100 words and limit the list to 10 items. The current date will be provided at the top of the transcript; use it to add ISO 601 dates in parentheses to action items that mention relative days (e.g. "tomorrow").

"follow_up:" - add follow-up questions. Limit each item to 100 words and limit the list to 5 items.

"stories:" - add stories or examples found in the transcript. Limit each item to 200 words and limit the list to 3 items.

"references:" - add references made to external works or data found in the transcript. Limit each item to 100 words and limit the list to 3 items.

A System of Enrichment



A group of six diverse students are gathered around a table, focused on a robotics project. In the foreground, a student wearing a white hijab and a striped shirt is holding a colorful, multi-layered ring. Behind her, another student in a blue denim jacket is working on a small electronic device. To the right, a student in a light blue denim shirt is pointing at something off-camera. The background shows other students, one with glasses and another with dark hair tied back, also engaged in the activity. The scene is set in a bright, modern classroom environment.

Demo Time

Broader Implications





A Call to Innovation

The End

