



Making Podcasts of Lectures Using automated Open- Source Tools

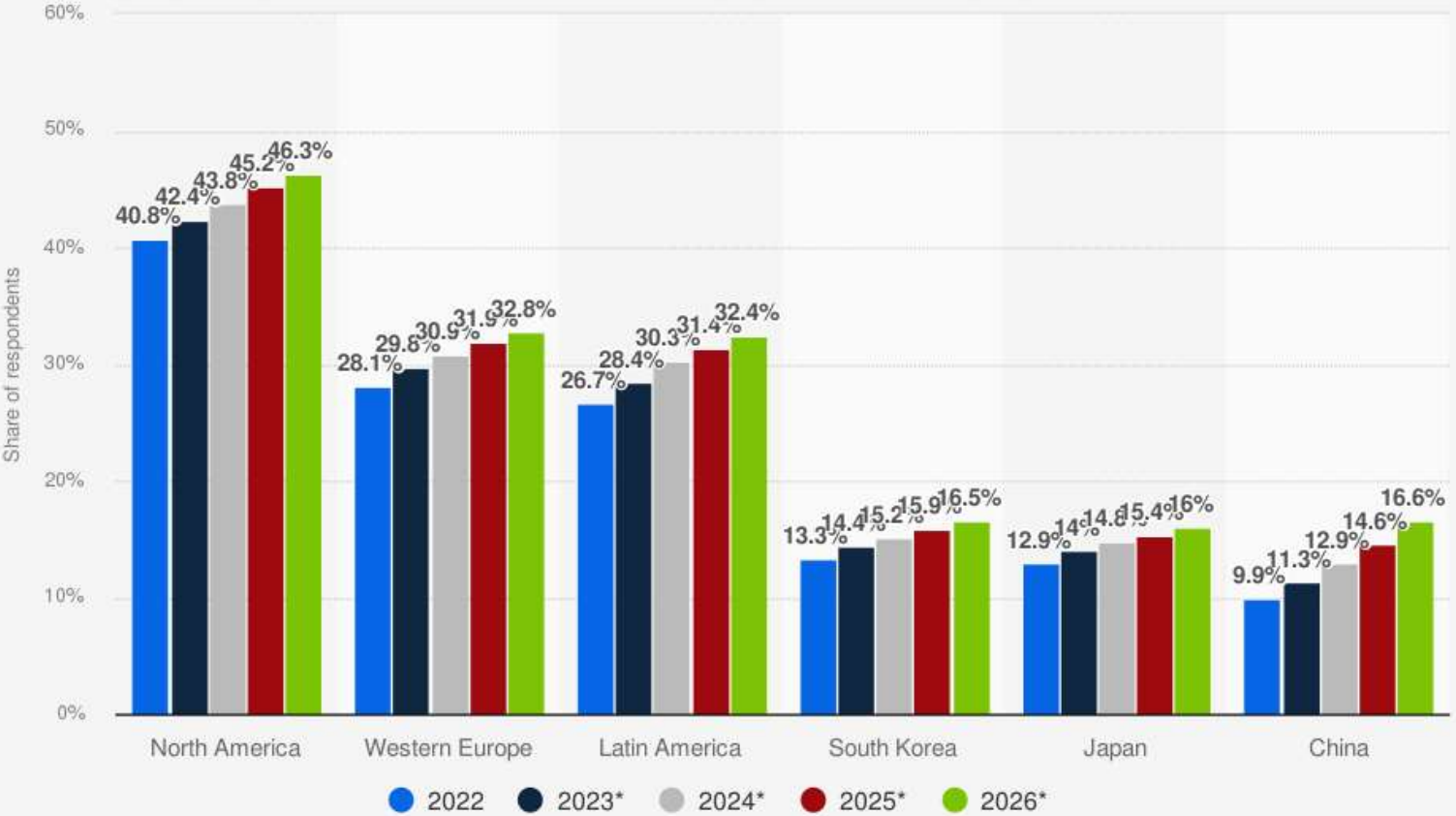
Carlos Turró, Universitat Politècnica de València
turro@cc.upv.es

CT

Podcast Panorama



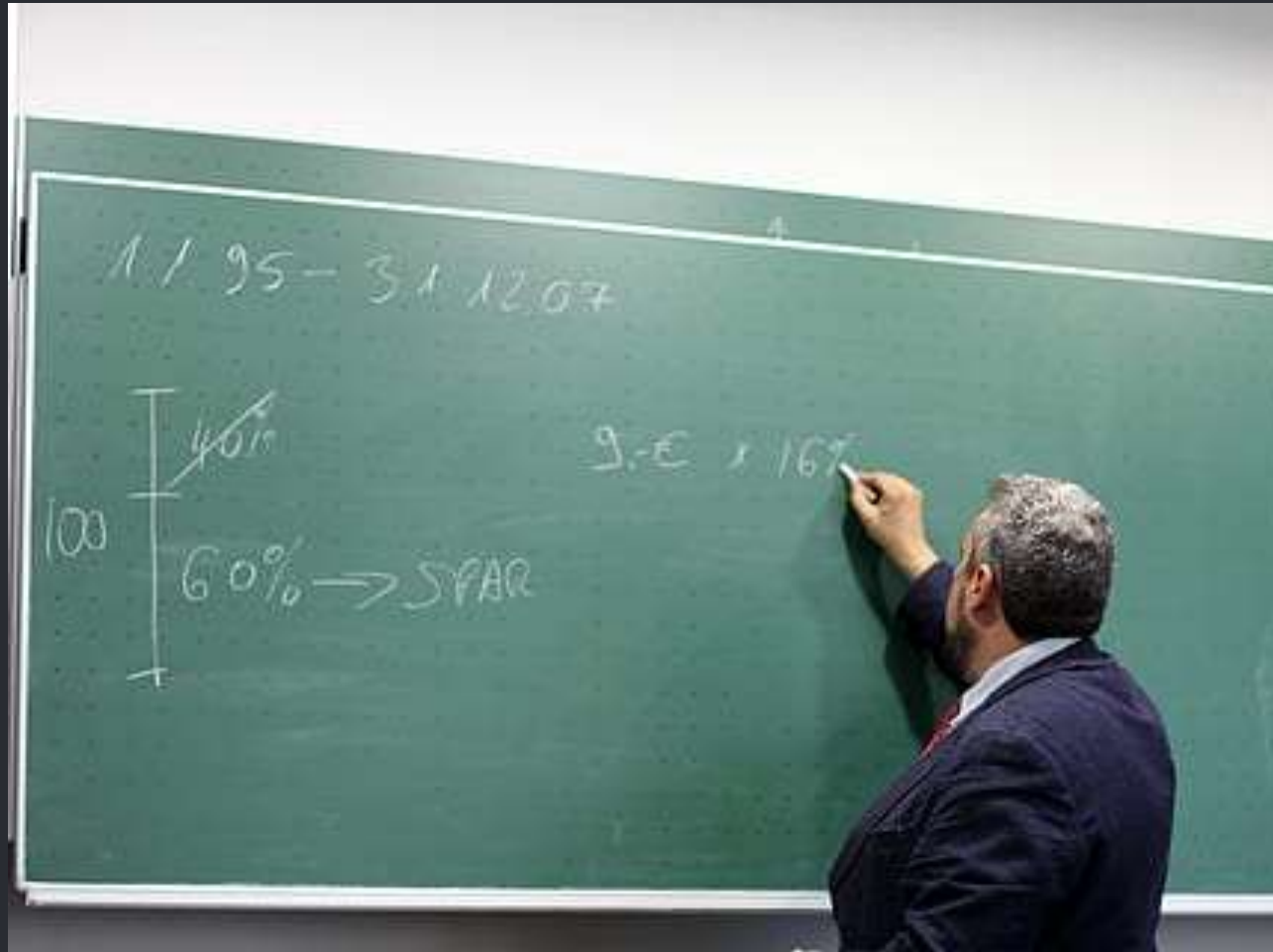
Share of podcast listeners among internet users in selected countries and regions worldwide 2022-2026



Source:
Insider Intelligence
© Statista 2025

Additional Information:
Worldwide; August 2022; among internet users of any age who listen to a podcast via digital stream or direct download on any device at least once per month

The idea



Why Turn Lectures into Podcasts?



Wider Reach

Reuse content beyond classroom walls



Flexible Learning

Support asynchronous mobile consumption



Resource Efficiency

Create value without additional effort



For fun !

Everybody have a different way fo have fun



Steps to Transform Lectures



Record Lecture & Get a clear audio file

Capture original content



Generate Transcript

Convert speech to text



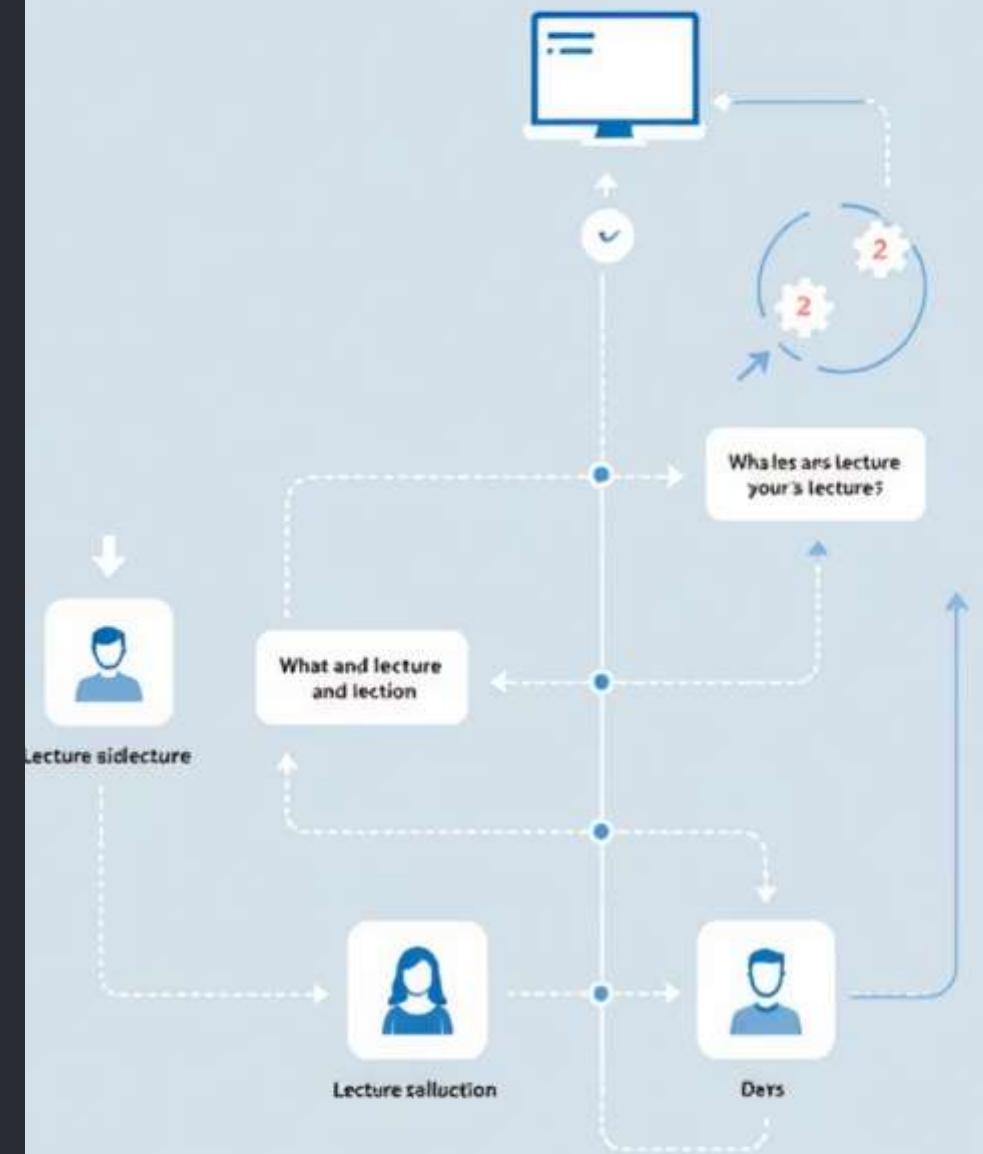
AI Processing

Create summaries and podcast scripts



Audio Production

Generate podcast from script



Audio recording for AI

Best practices

- AI has a different “hearing style” than human beings
- The higher the quality, the better, but ...
- Clipping will kill your AI -> Beware of sound levels
- Echo is also pretty bad
- Noise is not as bad as the other two



Speech to Text Conversion

Whisper Benefits

- Extremely accurate results
- Free to use (not Open Source)
- Multiple output formats
- Supports many languages

```
# whisper my_lecture.mp4 --model medium --  
language English --output_format srt
```

Let's go

The screenshot shows a video player interface. The main video frame displays a lecture hall with a green chalkboard and a man standing in front of it. A digital clock on the wall shows 11:47. Overlaid on the right side of the video is a presentation slide titled "Ejercicio 11.1. Paginación sin MV" with the "fso" logo. The slide contains text about processor logical addresses and pagination. A "Descargas disponibles:" (Available Downloads) menu is open, listing various download options for video, presentation, audio, and captions. A "Settings" menu is also visible, showing options like "Datos y descripción de vídeo", "Calidad del vídeo", "Descargas", "Atajos de teclado", and "Versión del reproductor de Opencast". The video player controls at the bottom show a progress bar at 01:39 / 01:16:01, 1x speed, and a volume icon.

Ejercicio 11.1. Paginación sin MV fso

Un procesador maneja direcciones lógicas de 16 bits y soporta paginación. Se usan tamaños de página de 256, 1024 y 4096 bytes. El Operativo ha de gestionar la paginación según sus necesidades. El procesador contiene **2800** bytes de instrucciones a partir de 0x0000, **1198** bytes de datos a partir de 0x3000 y prevé una pila a partir de 0x9000. Se debe gestionar la tabla de páginas y el número de páginas necesarias para cada modelo de procesador

Descargas disponibles:

- presenter/delivery
 - [video/mp4] 1280x720@25
 - [video/mp4] 854x480@25
 - [video/mp4] 1280x720@25
- presentation/delivery
 - [video/mp4] 960x720@25
 - [video/mp4] 640x480@25
 - [video/mp4] 1024x768@25
- audio_ambient/upv
 - [audio/m4a] 48000 Hz
- audio_noisereduction/upv
 - [audio/m4a] 48000 Hz
- captions/delivery
 - [text/vtt] [español](#) [transcript]

Settings

- Datos y descripción de vídeo
- Calidad del vídeo
- Descargas
- Atajos de teclado
- Versión del reproductor de Opencast

01:39 / 01:16:01 1x

AI Summary Creation

"I've just uploaded the transcript of a live lecture, and I want you to provide a summary of the lecture, including the key points."

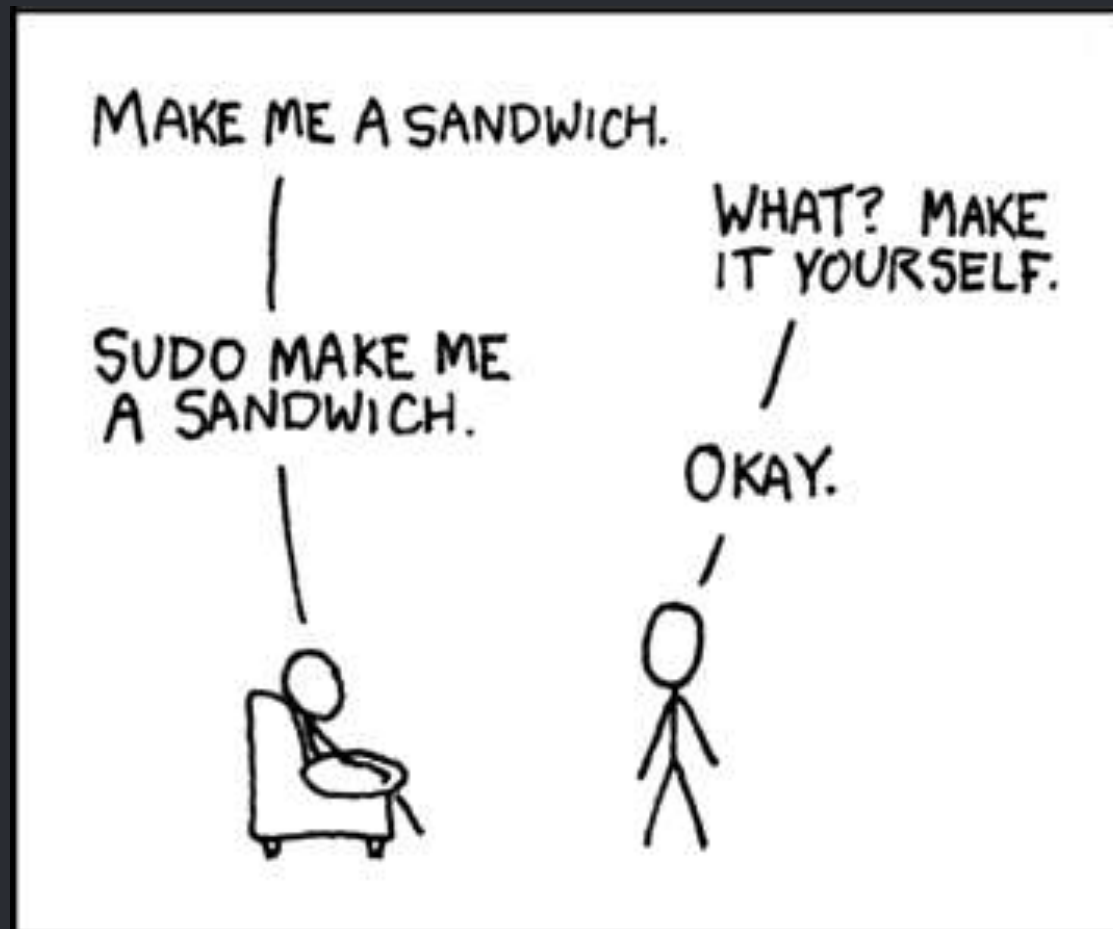
In this presentation, the importance of hand gestures in oral presentations is discussed. Gestures are helpful in capturing the audience's attention, as movement attracts the eye and can convey emotions and ideas more effectively than words alone.

They can also be used to narrate and illustrate a text, making it more engaging and easier to understand. Additionally, gestures can help structure speech, guiding the audience through the presentation and highlighting key points.
(...)



AI Podcast Creation. Take 1

"Make a podcast from the transcript"



AI Podcast Creation. Take 2

You are a world-class podcast writer; you have worked as a ghost writer for many famous podcasters. We are in an alternate universe where actually you have been writing every line they say, and they just stream it into their brains. You have won multiple podcast awards for your writing. Your job is to write word by word, even "umm, hmmm, right" interruptions by the second speaker based on the PDF upload. Keep it extremely engaging, the speakers can get derailed now and then but should discuss the topic.

Remember Speaker 2 is new to the topic and the conversation should always have realistic anecdotes and analogies sprinkled throughout. The questions should have real world example follow ups etc

Speaker 1: Leads the conversation and teaches the speaker 2, gives incredible anecdotes and analogies when explaining. Is a captivating teacher that gives great anecdotes.



AI Podcast Creation. Take 2

Speaker 2: Keeps the conversation on track by asking follow up questions. Gets super excited or confused when asking questions. Is a curious mindset that asks very interesting confirmation questions. Make sure the tangents speaker 2 provides are quite wild or interesting. Ensure there are interruptions during explanations or there are "hmm" and "umm" injected throughout from the second speaker. It should be a real podcast with every fine nuance documented in as much detail as possible. Welcome the listeners with a super fun overview and keep it really catchy and almost borderline click bait. Always start your response directly with speaker 1: Do not give episode titles separately, let speaker 1 title it in her speech. Do not give chapter titles. It should strictly be the dialogues.

The podcast should be in {output_language} language



Much better results

"speaker": "speaker-1",

"text": "Welcome to 'Designing for All', a podcast where we explore the importance of accessibility in our daily lives. I'm your host, and today we're discussing how to make online content more inclusive. Joining me is an expert in accessible design. Let's dive right in! What inspired you to focus on accessibility?"

"speaker": "speaker-2",

"text": "Thanks for having me! I realized the impact of inaccessible design when a family member struggled with a simple website. It sparked a passion to ensure everyone can engage with online content effortlessly."

"speaker": "speaker-1",

"text": "That's a powerful story. Let's talk about today's topic: accessible online content. What are some common pitfalls designers fall into, and how can they avoid them?"



Practical Issues with Long Lectures

Size Limitations

Subtitle files may exceed LLM context window

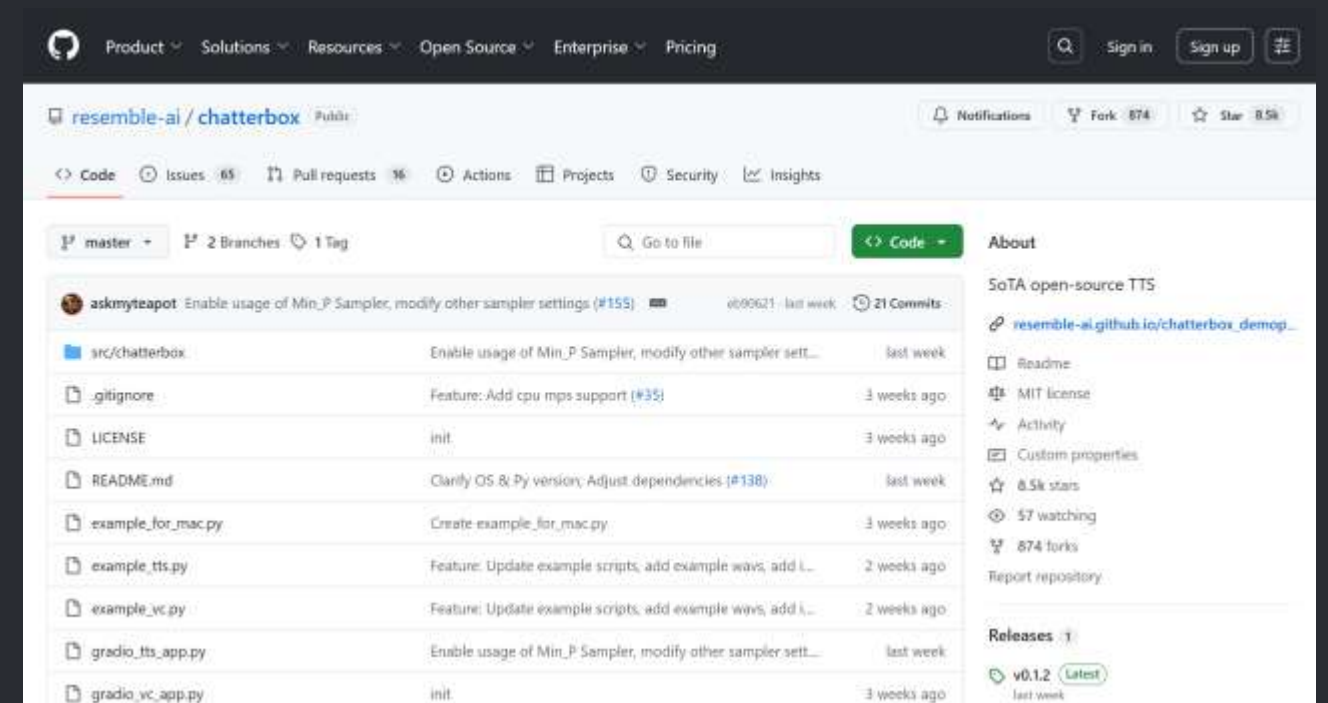
Solutions

- Create summary of entire lecture
- Process section by section
- Use chunking techniques
- Implement iterative processing

Text to Speech using Open Source

Best practices

- Each week there is a new tool in Huggingface (e.g. CoquiTTS).
- This week I like Chattebox.
https://resemble-ai.github.io/chatterbox_demopage/
- If your text to speech doesn't support diarization (multiple speakers), you need to make a sound file for each speaker intervention
- Many of the TTS include voice cloning, so you may use the real voice of your teacher(s) (!!)



The Final Demo

Demonstration available at: <https://paellaplayer.upv.es/player.html?id=ai-tool>

- 1 — **Original Lecture**
Full-length video recording
- 2 — **Generated Transcript**
Text version with timestamps
- 3 — **AI Summary**
Condensed key points
- 4 — **Podcast Version**
Audio-optimized content





Final Considerations

1 A proper evaluation is needed

Measure actual educational benefit

2 Quality Control

Each step requires human review

3 Purpose Questions

Define appropriate use cases

4 Pedagogical Integration

Align with educational objectives