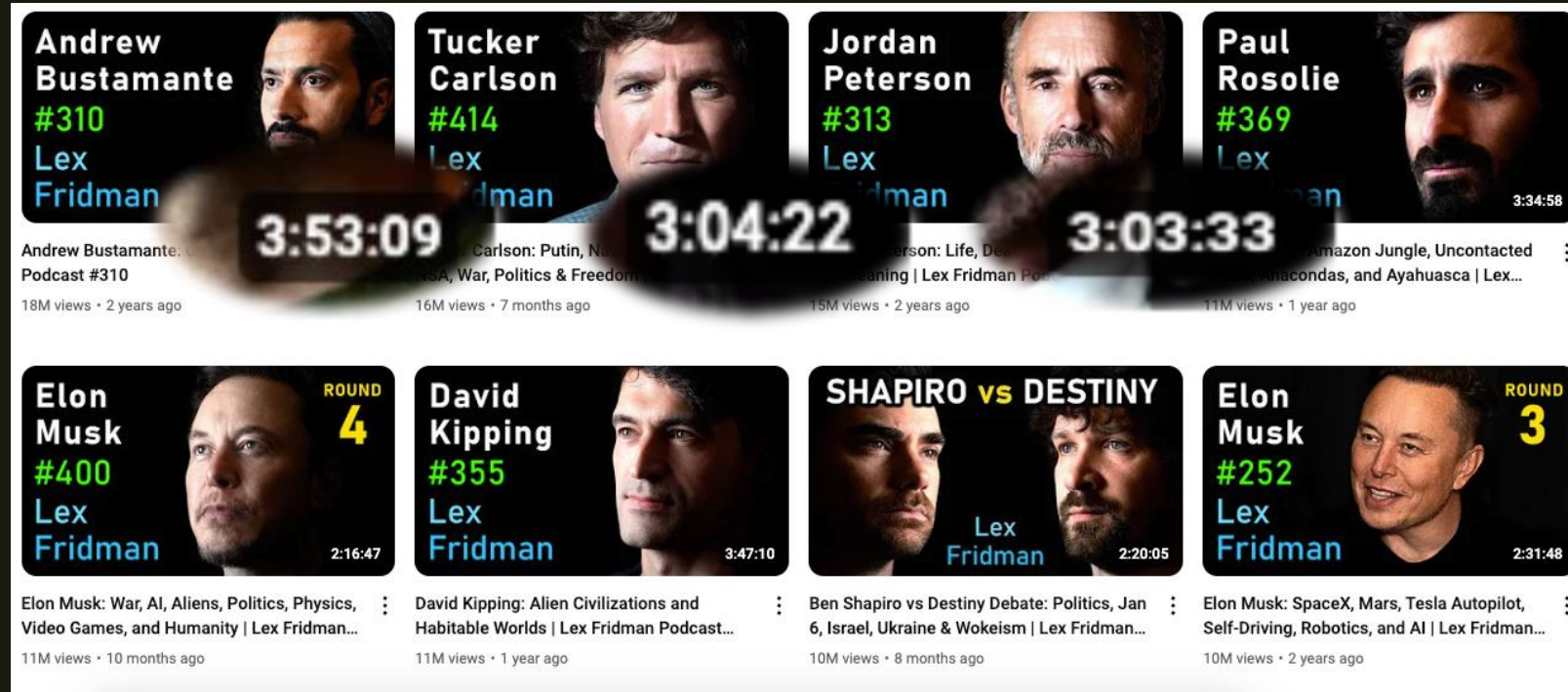




Long Text Summarisation

Lessons learned from Building [Lex Notes](#)

YEAH...BUT WHY?



That's about 20 days or 3 weeks non-stop listening

Average 3.4 hours per episode

Total 450 episodes

If I skip the opening and closing and listen at 2x speed, the total time would be 500 hours.

Lex Notes

The Dilemma

The Idea

Quick Stats

The Recipe

442 Donald Trump

441 Cenk Uygur

440 Pieter Levels

439 Craig Jones 2

438 Elon Musk and-Neuralink Team

437 Jordan Jonas

436 with Ivanka Trump

435 Andrew Huberman

434 Aravind Srinivas

433 Sara Walker

432 Kevin Spacey

431 Roman Yampolskiy

430 Charan Ranganath

429 Paul Rosolie

428 Sean Carroll 3

427 Neil Adams

426 Edward Gibson

424 Bassem Youssef

423 Tulsi Gabbard

422 Mark Cuban

Lex Notes

Discover concise and engaging summaries of the Lex Fridman Podcast. We simplify complex discussions on AI, science, philosophy, and technology into 1-minute reads, perfect for busy learners looking for insights without the time commitment. Stay informed, fast.

The Dilemma

Podcast Length

Episodes range from 1-8 hours, making them hard to fit into a busy schedule.

Overwhelming Quantity

With over 400 episodes available, it's difficult to know where to start.

Time Constraints

Even at 2x playback speed, episodes still take hours to finish.

The Idea

Get a Quick Episode Overview

Instantly scan key insights from each Lex Fridman podcast episode before committing to a full listen.

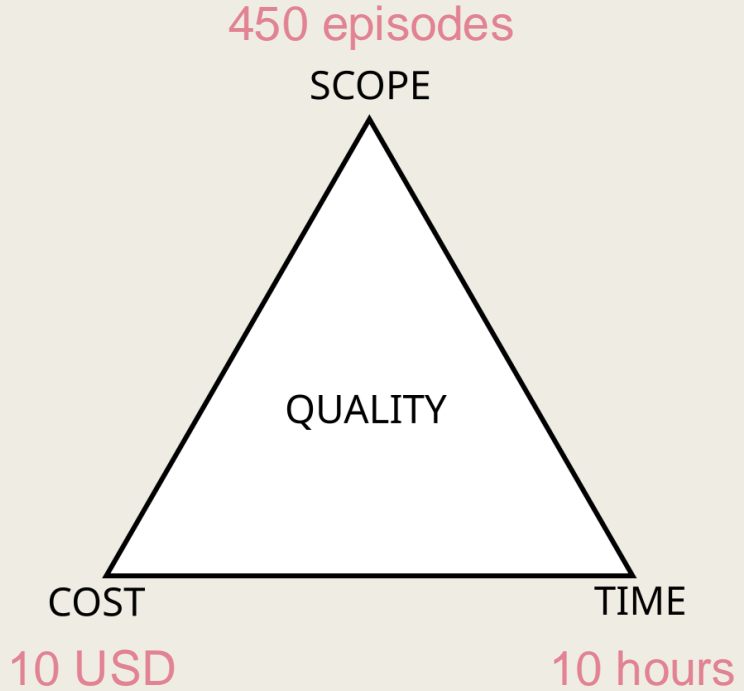
Easily Shortlist Episodes

I can effectively perform this task manually in NotebookLM

- Built-in Briefing Doc feature, or
- Custom prompts, or
- Generated Audio Overview

The screenshot displays the NotebookLM interface. On the left, a sidebar shows 'Select all sources' with a checkmark and a source 'Ed Barnhart: Maya, Az...' also checked. The main area shows a 'New note' dialog box. The dialog has a title 'New note' and a subtitle 'Saved response'. Below this, it states 'SAVED RESPONSES ARE VIEW ONLY'. The main content of the dialog is a briefing document titled 'Decoding Ancient Civilizations: A Conversation with Ed Barnhart'. The document text includes: 'This briefing document summarizes key themes and insights from the Lex Fridman Podcast #446 featuring archaeologist Ed Barnhart.' followed by a section 'Main Themes:' with four bullet points: 'Mysteries of the Past', 'Ancient Knowledge and Practices', 'The Power of Storytelling and Myth', and 'The Human Experience Across Time'. At the bottom of the dialog, there is a 'View Chat' button and a text input field with '1 source' and 'Start typing...'. A footer note at the bottom of the interface reads: 'NotebookLM may still sometimes give inaccurate responses, so you may want to confirm any fact'.

But with 450 episodes, can I solve this effectively, within budget, and at “scale”?



Source	Instant Data Scraper: html to csv Listly – Web Scraping: html to csv + fetch Wget & html2text: podcast transcript to txt yt-dlp: youtube subtitle to txt
Transform	Jupyter Pandas Python
Summarise	OpenAI LangChain LM Studio Ollama
Measure	Anthropic Claude AI Google NotebookLM
Deploy	mkdocs GitHub Page Cloudflare

4o vs 4o mini



```
Analysis the podcast episode from the Lex Fridman Podcast series. Summarise the key points discussed in the episode with a focus on: ~  
## The main arguments presented ~  
## Any notable quotes from the guest ~  
## Relevant topics or themes covered~  
Ensure the summary is clear and concise, with bullet points under the specified headings~
```

```
Provide a detailed analysis of the Lex Fridman Podcast episode. ~  
Your summary should be comprehensive yet well-structured, covering the breadth and depth of the discussion. ~  
Organize your response under the following headings:~  
~  
# [Episode Title] [Guest Name]~  
~  
## The main arguments ~  
~  
----  
instruction for this section~  
- List at least 5 key arguments or points made during the episode~  
- provide a brief explanation and its significance in the context of the discussion~  
- Highlight any counterarguments or alternative perspectives mentioned~  
----  
~  
## Any notable quotes ~  
~  
----  
instruction for this section~  
- Include at least 5 direct quotes that encapsulate important ideas or memorable moments~  
- provide a sentence of context explaining its relevance or impact~  
----  
~  
## Relevant topics or themes~  
~  
----  
instruction for this section~  
- Identify and explain at least 5 major themes or topics discussed in the episode~  
- provide examples of how it was explored in the conversation~  
- Note any connections between different themes or how they relate to broader societal issues~  
----  
~  
Additional Guidelines:~  
~  
Use clear and concise bullet points under each heading~  
Provide brief explanations for any specialized terms, events, or concepts mentioned~  
Highlight any unique aspects of the episode, such as the interviewing style or guest's background~  
Include a mix of factual information, personal anecdotes shared by the guest, and any hypothetical scenarios discussed~  
If applicable, note any evolution in the guest's perspectives throughout the conversation~
```

GPT-4o used up my OpenAI credits
5 USD for 2 hours of coding in Cursor with GPT-4o

AI tools are expensive!! Thanks REA for my GitHub Copilot subscription

RAG vs No RAG



RAG is super helpful especially when the raw text isn't good enough, or the guest or topic isn't very well-known

```
being kbr halbert and they imported workers in from places like nepal and sri lanka and the philippines to come in and cook food and work at this dining facility um i got curious about how much how much it cost us as taxpayers and so i started asking around some of the people and i think at that time it was like well every time a soldier or a service member walks through the door if i were to go in for breakfast and grab a banana and walk out that's an automatic $35 per head per meal four times a day thousands and thousands of people and then we made friends you know there there's a lot there's a pretty large filipino community in hawaii a lot of filipino soldiers from hawaii we made friends with the filipino workers who were there they would often go in like the back of the tents and set up their own like rice cookers and cook their own meals um which is where the real good food was uh but just started talking to them and getting to know them and ask like hey how much do you get paid and on average it was like oh i get paid like 500 bucks a month 500 bucks a month to go and do this work of of either you know cleaning out porta potties um picking up trash the dining facility
```

```
* 2:37:15 - O Block
* 2:41:11 - Crip Mac
* 2:44:23 - Aliens

## Introduction

Andrew Callaghan [(00:00:00)](https://youtube.com/watch?v=yEou104m_P0&t=0) There's two people in the back, two of her homegirls wearing sheisty masks. I'm like, "What are we doing? Where are we going?" She goes, "We're going to go film the riot. We're going to Lake Street." We drive down there, Kmart is burning, Target is burning, everything is on fire. She has the Sony a7, she gives me a microphone and she's like, "Go talk to that guy." That was the guy with a molotov cocktail in his hand who had just burned Kmart down. I go, "What should I ask him?" She goes, "What's on your mind?" I walk up to him and I'm like, "What's on your mind?"

Lex Fridman [(00:00:36)](https://youtube.com/watch?v=yEou104m_P0&t=36) The following is a conversation with Andrew Callaghan, host of Channel 5 on YouTube, where he does Gazelle style interviews with fascinating humans at the edges of society. The so-called vagrants, vagabonds, runaways, outlaws, from QAnon adherence to fish heads, O'Block residents, and much more. He created the documentary that I highly recommend called This Place Rules, on the undercurrents that led to the January 6th Capitol riots. This is the Lex Fridman podcast. To support it, please check out our sponsors in the description. Now, dear friends, here's Andrew Callaghan.

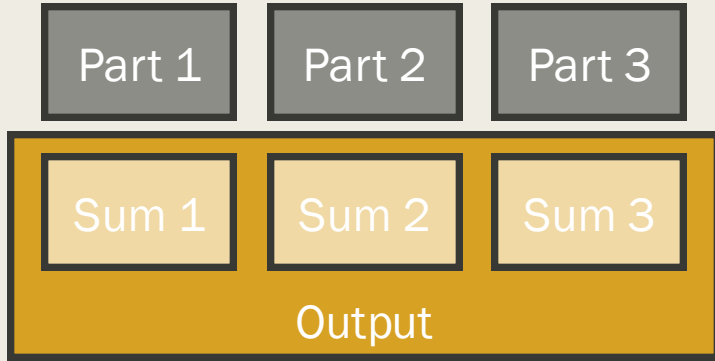
## Walmart

Andrew Callaghan [(00:01:18)](https://youtube.com/watch?v=yEou104m_P0&t=78) I tried to color match you though. Got the black and white going. I went to Walmart before this and got the Wrangler shirt with the Texas Longhorns Tee and everything.

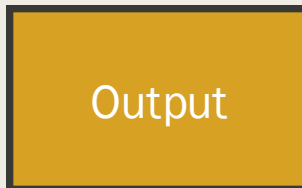
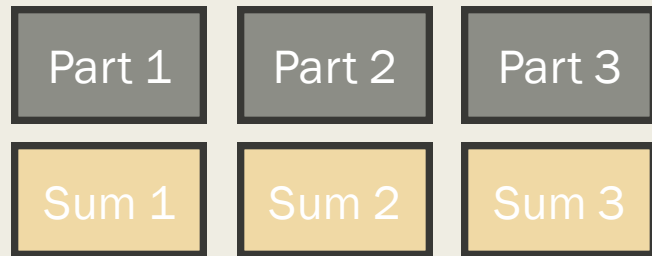
Lex Fridman [(00:01:25)](https://youtube.com/watch?v=yEou104m_P0&t=85) Is that where you shop, Walmart?
```

it's not as useful when the raw text already contains everything you need

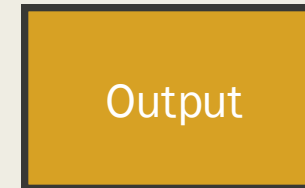
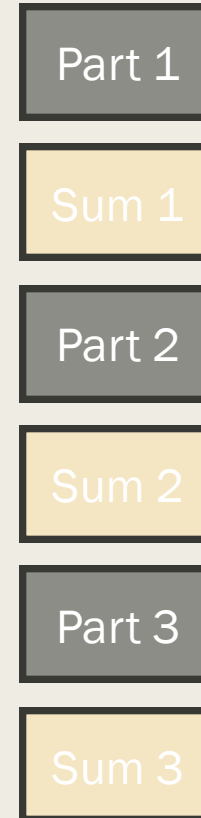
Chunk or MapReduce or Refine



Fewer API calls
Can be run in parallel



One extra API calls
Can be run in parallel



More API calls
Accumulated context
Must run in sequence

LLM or SLM



Quality



Llama 3.2 3B
Llama 3.1 8B
Openhermes 2.5 Mistral 7B



Claude
GPT 4o
GPT 4o mini

NotebookLM



Easy to use

Is it worth it?

A slightly deeper understanding of AI Apps, LLMs and SLMs

Better clarity on the word “scale” in an AI project

A few insights into what cost means in the era of AI

I’ve shortlisted some episodes I would like to listen to

I also found some unpopular opinions in some of the podcasts to broaden my perspectives

I have a website to read during my commute

I am doing a knowledge sharing to bounce ideas with great people :)