Pickle Pod

An interactive podcast experience

Powered by

llElevenLabs

Problem

- Podcasts offer a great hands off experience. People use them when they go driving, go on a long hike or a walk.
- However, the experience is unnecessarily one directional. There is no interaction. Making learning new concepts more difficult when you simply just want to pause and hear answers to your questions by the Podcaster contextually.
- PicklePod allows you to communicate with the Podcaster real time:
    - Audience can ask any personalized questions to enhance your learning
    - Interactive learning results in better retention
    - Podcasters can gather valuable insights and feedbacks to their content
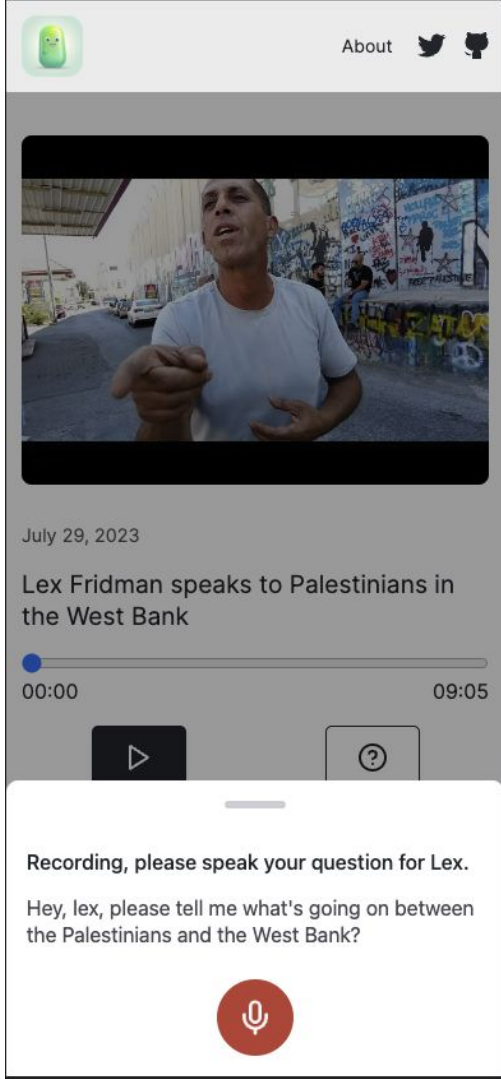
What we built

Demo

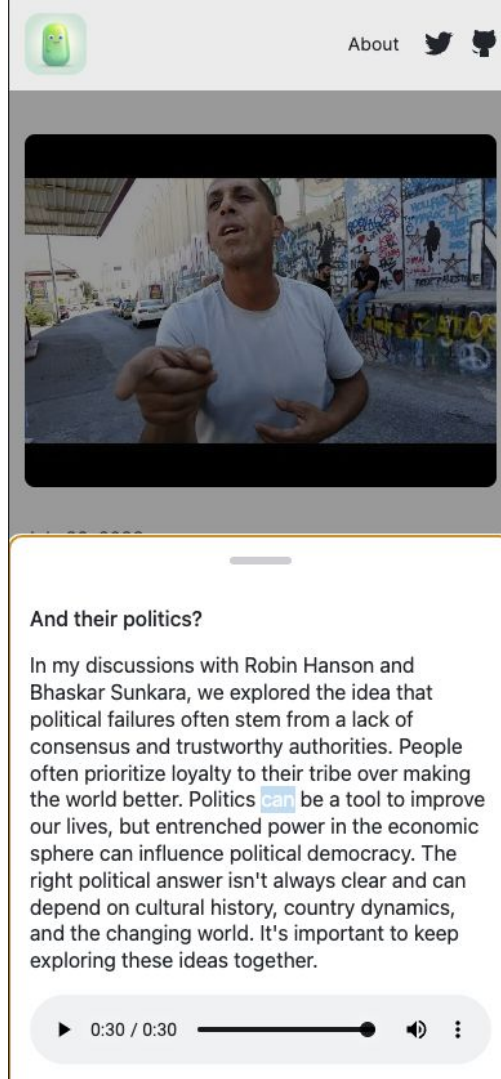July 29, 2023

Lex Fridman speaks to Palestinians in the West Bank

00:00   09:05

▷   ⑦

─

Recording, please speak your question for Lex.

Hey, lex, please tell me what's going on between the Palestinians and the West Bank?
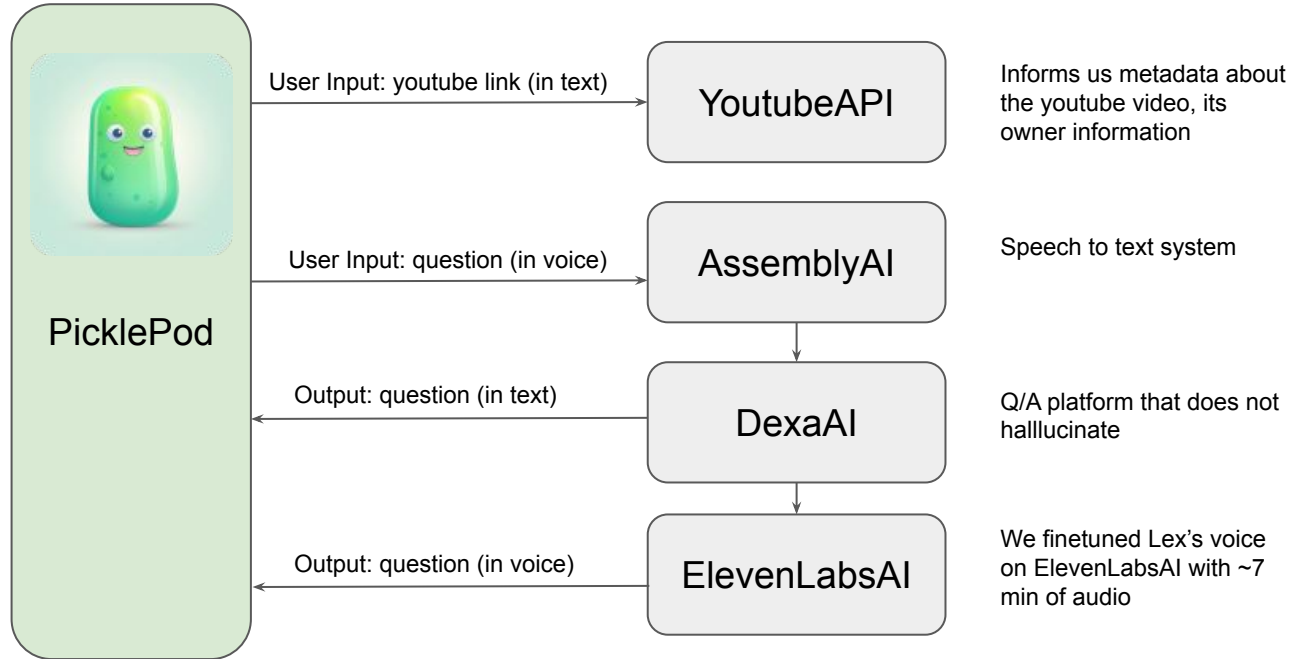
🎙

─

And their politics?

In my discussions with Robin Hanson and Bhaskar Sunkara, we explored the idea that political failures often stem from a lack of consensus and trustworthy authorities. People often prioritize loyalty to their tribe over making the world better. Politics can be a tool to improve our lives, but entrenched power in the economic sphere can influence political democracy. The right political answer isn't always clear and can depend on cultural history, country dynamics, and the changing world. It's important to keep exploring these ideas together.

▶ 0:30 / 0:30 ──────────●── 🔊 ⋮

Process

- Assembly AI is used for real-time speech-to-text, so users can ask the podcast host questions using their voice.
- Dexa AI is used to generate realistic answers to text queries in the style of Lex Fridman, with a custom private RAG API trained on the transcripts across all ~400 episodes of the Lex Fridman Podcast.
    - Dexa uses Pinecone hybrid search to retrieve the most relevant podcast chunks, applies cross-validation and re-ranking to ensure we only take into account transcript chunks that are relevant to the question, uses RAG + GPT-4 to generate an answer in the style of Lex Fridman, and then applies a post-processing step to minimize hallucination.
- ElevenLabs is used for text-to-speech to convert the resulting answer to audio using a custom voice trained on ~7 minutes of Lex Fridman's voice.

System

PicklePod

User Input: youtube link (in text) → YoutubeAPI — Informs us metadata about the youtube video, its owner information

User Input: question (in voice) → AssemblyAI — Speech to text system

Output: question (in text) ← DexaAI — Q/A platform that does not halllucinate

Output: question (in voice) ← ElevenLabsAI — We finetuned Lex's voice on ElevenLabsAI with ~7 min of audio

Future

- Enable for more audio sources rather than Youtube
- Allow for more creators voice (with their permission of course), we chose Lex because we are aware that Lex is generally okay with this.
- Handsoff interactions where you say, "hey hold up, lex", and then lex can pause in their speech to answer your question would feel much more magical.
- And others, please see github README.

Links

- Github: https://github.com/transitive-bullshit/pickle-pod
- Website: https://pickle-pod.vercel.app/
- Creators:
  - Travis Fischer: https://twitter.com/transitive_bs
  - Sasha Dog: https://twitter.com/hackgoofer