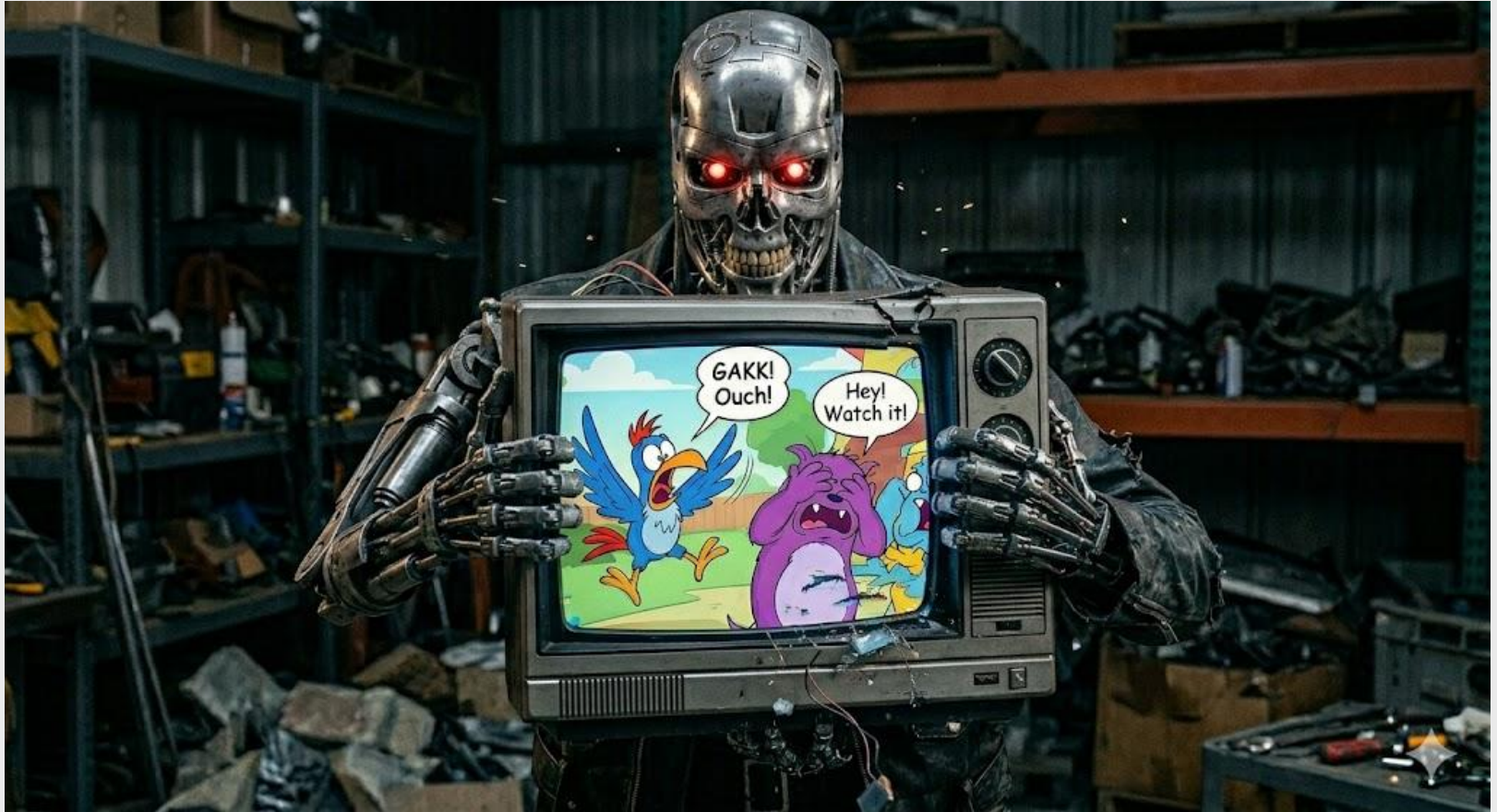


AI Video Summarization

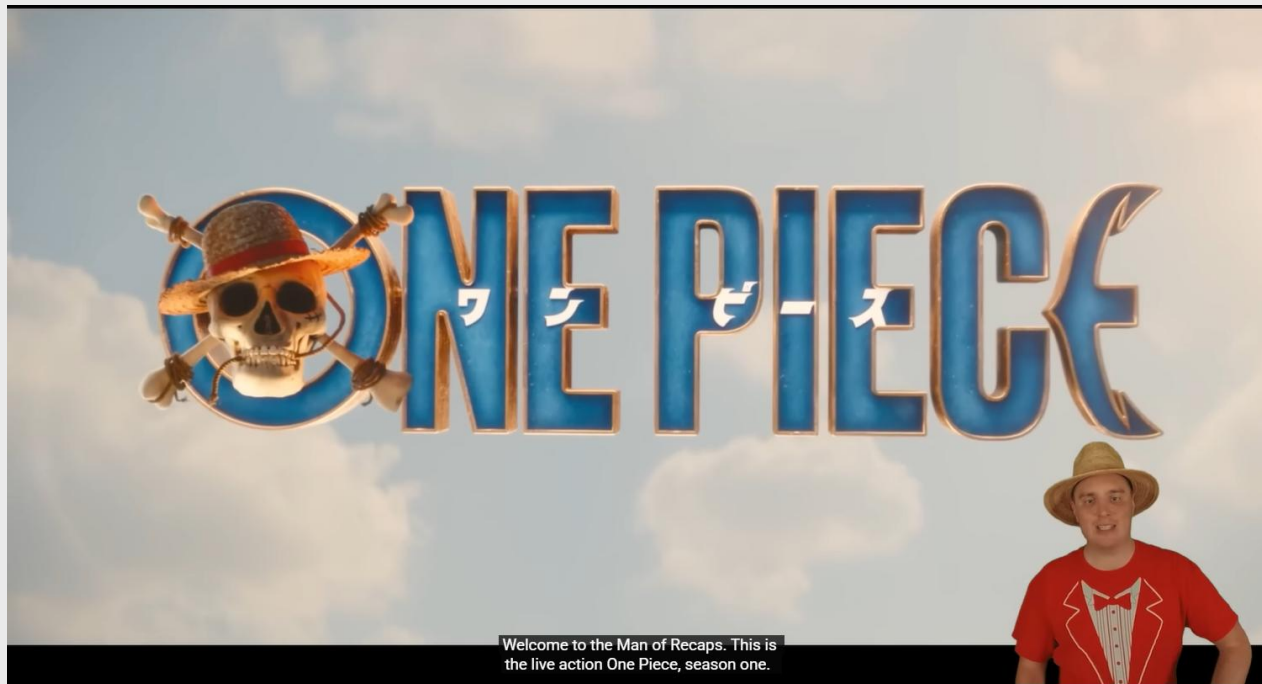


Why Summarize Video?

- **Time** – turn long-form content (films, lectures, sports) into short watchable summaries (movie trailers, high-light reels, etc)
- **Targeting** – highlight elements of a video that resonate with a particular audience
- **Discovery** – help viewers decide what to watch
- **Education** – lecture and documentary clips for study guides

Example: Man of Recaps

- Hosted recap video - cold open, episode or movie title, then voiceover over stitched clips from the source
- Similar channels on YouTube use this pattern for TV, anime, and film—search for recap or “story explained” formats.

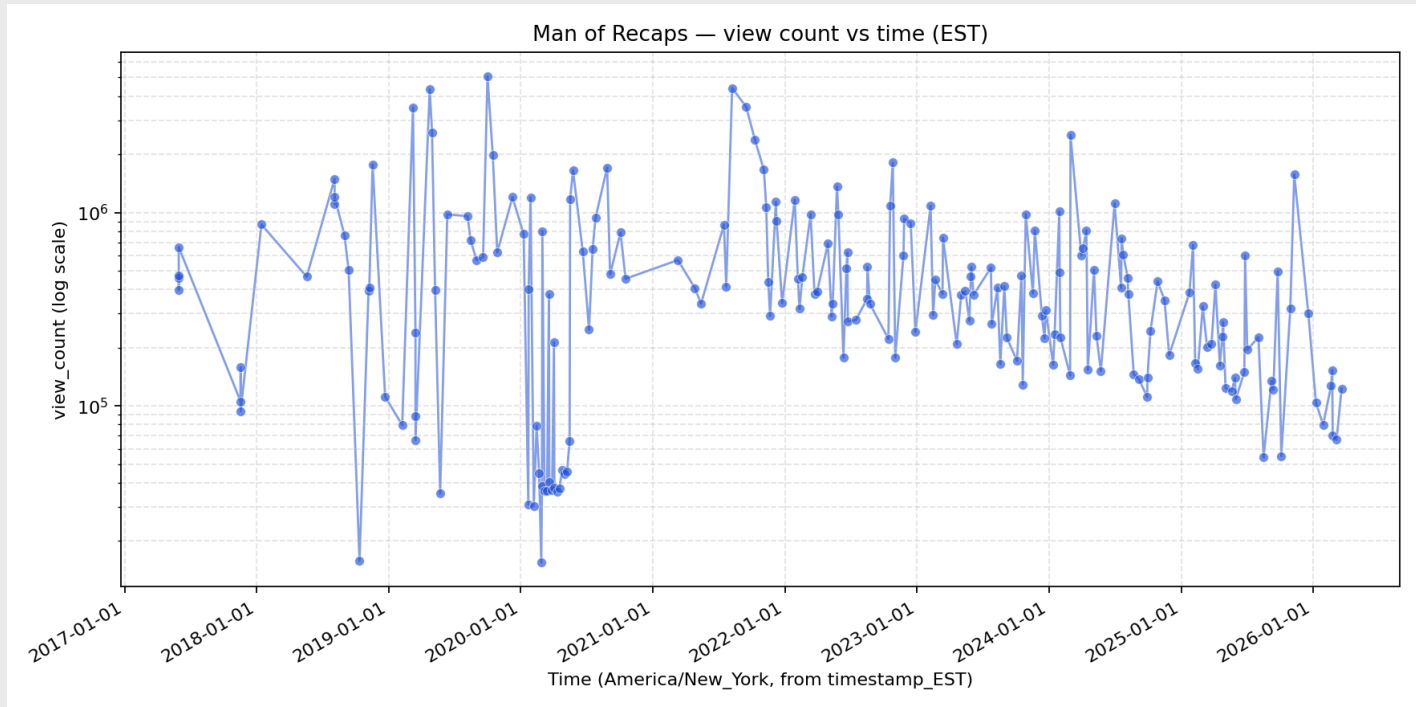


Summarization Channel

The image displays a grid of 12 video thumbnails, each representing a recap video for a different TV series or movie. Each thumbnail includes a title, a 'RECAP' label, a duration, and a view count with the time it was posted. The thumbnails are arranged in three rows and four columns.

Video Title	Duration	Views	Posted
Peaky Blinders RECAP: Full Series before The Immortal Man	18:45	123K	10 days ago
Outlander RECAP before the Final Season	29:28	66K	3 weeks ago
A Knight of the Seven Kingdoms RECAP: Season 1	8:00	70K	1 month ago
Paradise RECAP: Season 1	9:45	153K	1 month ago
The Night Agent RECAP: Season 2	6:53	127K	1 month ago
Bridgerton RECAP: Season 3	6:52	79K	2 months ago
The Pitt RECAP: Season 1	9:29	103K	2 months ago
Avatar RECAP: The Way of Water	5:44	300K	3 months ago
Stranger Things RECAP before the Final Season	18:13	1.5M	4 months ago
The Witcher RECAP before Season 4	11:22	318K	5 months ago
Fall Asleep to Harry Potter as a Bedtime Story	22:14	55K	5 months ago
Alice in Borderland RECAP before Season 3	13:41	494K	6 months ago

Channel Views



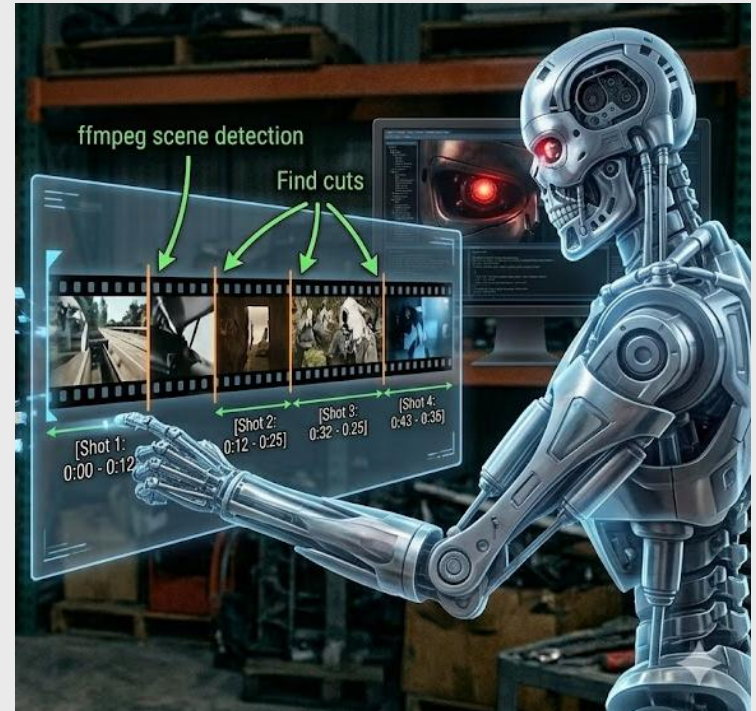
- **202 videos over 9 years**
- **Views per video:**
 - mean = 581K,
 - median = 378K,
 - max = 5M

Raw Inputs

- To summarize a video with AI we need two elements:
video file and **subtitle file**
- **Video file** – full episode or movie (e.g. mp4, mkv)
 - We will sample frames from this file to get a visual understanding of the shots
- **Subtitle file** – srt file with time-stamped dialogue
 - Can be extracted with ffmpeg library from video file sometimes
- Time-stamped dialogue + visual samples give enough signal for AI to describe each shot

Shot Detection

- Use ffmpeg scene detection (or similar) to find cuts: each interval is a shot with start/end times on the source timeline.
- Output: a manifest (e.g. shots.json) — timing only; hundreds of short shots for a feature film is normal.
- Parameters:
 - SCENE_THRESHOLD – lower bound on how different 2 frames need to be to declare a new shot
 - MIN_GAP_SECONDS – minimum duration of a shot



Example of shots.json

```
{
  "video_path":
  "C:\\Users\\tr2\\Dropbox\\ChatGPTs\\mgt575_apps\\2026Spring\\lecture_16_video_summarizat
  ion\\test\\video_raw\\Back.to.the.Future.Part.2.1989.1080p.Brrip.x264.Deceit.YIFY.mp4",
  "video_duration_seconds": 6472.4267,
  "scene_detection": {
    "scene_threshold": 0.35,
    "min_gap_seconds": 10,
    "num_detected_cuts": 157
  },
  "num_shots": 158,
  "shots": [
    {
      "index": 0,
      "start_seconds": 0.0,
      "end_seconds": 36.5417,
      "duration_seconds": 36.5417
    },
    {
      "index": 1,
      "start_seconds": 36.5417,
      "end_seconds": 47.7917,
      "duration_seconds": 11.25
    },
  ]
}
```

Building a Shot Catalogue

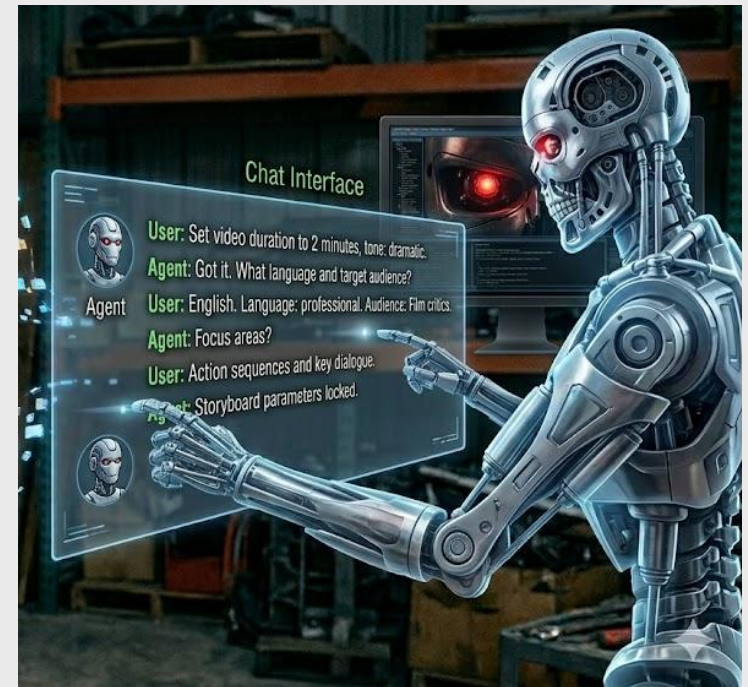
- For each shot: sample several **frames + subtitle lines + previous shot descriptions** → send to AI (e.g. Gemini).
- Output: `shots_description.json` — shot index, times, dialogue excerpt and AI generated description per shot.
- Approx. 15 minutes for 150 shots

Example of shots_description.json

```
{
  "index": 86,
  "start_seconds": 1986.8333,
  "end_seconds": 2008.375,
  "description": "Lorraine, Marty's future mother, prepares a pizza using a futuristic\n\"Black & Decker Hydrator\" appliance. She unwraps a tiny, foil-packaged pizza,\nplaces it on a tray into the machine, and it rapidly expands into a full-sized pizza.\n",
  "dialogue": "Hey, pizza. I'm hungry. All right. Just wait your turn. Grandma, when\nit's ready, could you just shove it in my mouth? Don't you be a smartass. Oh, great!\nThe Atrocity Channel. Hydrate level 4, please."
},
{
  "index": 87,
  "start_seconds": 2008.375,
  "end_seconds": 2018.8333,
  "description": "Lorraine serves the freshly hydrated pizza to her family, including\nolder Marty, Marty Jr., and Marlene, who are gathered around the dinner table.\nMeanwhile, the younger Jennifer (from 1985) secretly observes the family dinner from\na doorway.",
  "dialogue": "Mmm. Is it ready? Here you go. Oh, boy. Oh, boy. Mom, you sure can\nhydrate a pizza."
},
{
```

Storyboard Agent

- We use an agent with a chat interface to make our summary video storyboard
- Give agent the shot catalogue, a video style template (optional) and chat with agent to get video parameters (duration, tone, language, target audience, things to focus on)
- Output storyboard.json – ordered scenes with time stamps on source video and narration text for scenes



Example of storyboard.json

```
"scenes": [  
  {  
    "index": 0,  
    "start_time": 36.5417,  
    "stop_time": 47.7917,  
    "narration": "[smooth] Welcome to the Man of Recaps. [upbeat] Today we will be covering Back to the Future Part 2. [energetic] Get ready for a wild ride through time!",  
    "source_shot_indices": [  
      1  
    ]  
  },  
  {  
    "index": 1,  
    "start_time": 78.2083,  
    "stop_time": 86.0,  
    "narration": "[urgent] Doc Brown bursts in, telling Marty: \"You've got to come back with me!\"",  
    "source_shot_indices": [  
      4  
    ]  
  },  
]
```

Render Video

- **For each storyboard scene, extract the source video clip (ffmpeg)**
- **TTS (e.g. ElevenLabs) reads the narration**
 - emotion tags like [excited] steer delivery.
- **Mux audio with video**
 - adjust playback speed of video to match voiceover length
 - speed up or slow down video so no awkward silences
- **Concatenate scenes into final summary video**

Coding Session

- Make an agent to create AI video summaries
- Recaps, trailers, highlights

