I’d like to shift how we’re creating these digests. I want to go through each podcast in the database and specify the topic area that I think they belong to. Then I want to actually create different digests for each topic area, using the same daily pipeline python file, but creating a different output for each topic. Each topic area will have a different set of instructions for the headless claude code to create the topic-specific digest, so each topic group needs to be relatively static to make sure that it has a matching set of topic-specific instructions for creating the digest.

Each digest will have it’s own voice, it’s own music, it’s own structure.

Once each digest is written, the topic-specific transcripts that were used for it will be moved to ‘digested’ and that single digest will be converted to mp3 using the tts process we’ve established, then uploaded to the rss feed with a clear title – for example “Social commentary – New organizing ideas in 2025” or “AI News – Video generation tools” (those are just examples, not exact titles to be reused). Each day, the only digests that will be created are the digests that have a new episode from the 24 hours, so some days will have fewer digests and some days will have more (and potentially some days will have none).

On Friday’s, in addition to the daily digest for any news that has come out in the previous 24 hours, each topic area should also get a separate weekly digest that attempts to highlight major themes from that topic’s transcripts for the full week. This should not be a lengthy summary of everything discussed in every topic-specific transcript that week, but should identify key themes, topics, actions, thought leaders, product launches, or key news items from that week. The weekly digest should explicitly include key quotes from that topic’s transcripts from the week. Note that the weekly digest content will come entirely from the ‘digested’ transcripts and need to cross reference the topic assignments in the database to ensure that the weekly digest is topic-specific and not confusing the different topics for the week.

On Monday’s, when the script runs, it should include any episodes released since 6am the previous Friday. If ever the script runs and it has been more than 24 hours since the last time it ran, it should attempt to download all episodes since it last ran (up to 7 days prior to the day it is running, but no more than 7 days prior)

---

Get more aggressive about filtering out ads. These are ads:

You can see here that both c6c2e5f6 and 095fc94c were fully processed, but I don’t see transcripts for either of them in the transcripts folder and both still show in the audio cache (although the chunks were cleaned up):

🎯 Transcribing chunk 30/32: chunk\_030.wav

Python(97107) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 35s

🚀 Starting Parakeet MLX transcription...

Python(97108) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 30/32 - Progress: 43.0% (Elapsed: 15s, Est. remaining: 20s)

🔄 Chunk 30/32 - Progress: 85.9% (Elapsed: 30s, Est. remaining: 5s)

✅ Chunk 30 complete: 45.1s actual (RTF: 0.150x, chars: 4803)

🎯 Transcribing chunk 31/32: chunk\_031.wav

Python(97151) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 15s

🚀 Starting Parakeet MLX transcription...

Python(97153) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 31/32 - Progress: 99.0% (Elapsed: 15s, Est. remaining: 0s)

🔄 Chunk 31/32 - Progress: 99.0% (Elapsed: 30s, Est. remaining: 0s)

🔄 Chunk 31/32 - Progress: 99.0% (Elapsed: 45s, Est. remaining: 0s)

🔄 Chunk 31/32 - Progress: 99.0% (Elapsed: 60s, Est. remaining: 0s)

🔄 Chunk 31/32 - Progress: 99.0% (Elapsed: 76s, Est. remaining: 0s)

✅ Chunk 31 complete: 85.7s actual (RTF: 0.286x, chars: 5014)

🎯 Transcribing chunk 32/32: chunk\_032.wav

Python(97273) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 1.3min, estimated processing: 4s

🚀 Starting Parakeet MLX transcription...

Python(97275) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 32 complete: 15.1s actual (RTF: 0.196x, chars: 1425)

🔗 Step 4: Combining 32 transcription(s)...

🧹 Step 5: Cleaning up transcript...

🧹 Cleaning up transcript...

🧹 Cleaned transcript: removed 305 characters (potential commercials)

🗑️ Step 6: Cleaning up chunk files...

✅ Deleted chunk directory: audio\_cache/095fc94c\_chunks

✅ Transcription workflow complete!

• Total time: 27.0 minutes

• Transcript length: 158,228 characters

• Overall RTF: 0.173x

Analyzing audio for multi-speaker characteristics...

📻 Multi-speaker conversation characteristics detected

Error processing episode: name 'audio\_path' is not defined

Processing: How FASCISM Became Monetized

Downloading audio from https://anchor.fm/s/e8e55a68/podcast/play/107082614/https%3A%2F%2Fd3ctxlq1ktw2nl.cloudfront.net%2Fstaging%2F2025-7-19%2F405941925-44100-2-db9d35037d1f.mp3

Downloaded: audio\_cache/c6c2e5f6.mp3 (31259584 bytes)

🔄 Loading Parakeet MLX model...

✅ Parakeet MLX model loaded successfully

🎬 Starting robust transcription workflow for: c6c2e5f6.mp3

📊 Step 1: Analyzing audio file...

Python(97451) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📋 Analysis Results:

• Duration: 32.6 minutes

• Estimated processing time: 2.8 minutes

• Recommended chunks: 7

✂️ Step 2: File chunking...

Python(97452) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✂️ Splitting 32.6min audio into 5min chunks...

📄 Creating chunk 1: 0m-5m

Python(97453) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 1 created: 4MB

📄 Creating chunk 2: 5m-10m

Python(97454) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 2 created: 4MB

📄 Creating chunk 3: 10m-15m

Python(97455) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 3 created: 4MB

📄 Creating chunk 4: 15m-20m

Python(97456) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 4 created: 4MB

📄 Creating chunk 5: 20m-25m

Python(97457) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 5 created: 4MB

📄 Creating chunk 6: 25m-30m

Python(97458) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 6 created: 4MB

📄 Creating chunk 7: 30m-35m

Python(97459) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

✅ Chunk 7 created: 2MB

✅ Created 7 chunks for processing

🎯 Step 3: Transcribing 7 chunk(s)...

🎯 Transcribing chunk 1/7: chunk\_001.wav

Python(97460) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 35s

🚀 Starting Parakeet MLX transcription...

Python(97461) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 1/7 - Progress: 43.0% (Elapsed: 15s, Est. remaining: 20s)

🔄 Chunk 1/7 - Progress: 86.0% (Elapsed: 30s, Est. remaining: 5s)

🔄 Chunk 1/7 - Progress: 99.0% (Elapsed: 45s, Est. remaining: 0s)

✅ Chunk 1 complete: 55.4s actual (RTF: 0.185x, chars: 5134)

🎯 Transcribing chunk 2/7: chunk\_002.wav

Python(97520) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 15s

🚀 Starting Parakeet MLX transcription...

Python(97522) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 2/7 - Progress: 99.0% (Elapsed: 15s, Est. remaining: 0s)

🔄 Chunk 2/7 - Progress: 99.0% (Elapsed: 30s, Est. remaining: 0s)

🔄 Chunk 2/7 - Progress: 99.0% (Elapsed: 45s, Est. remaining: 0s)

🔄 Chunk 2/7 - Progress: 99.0% (Elapsed: 60s, Est. remaining: 0s)

✅ Chunk 2 complete: 65.5s actual (RTF: 0.218x, chars: 5166)

🎯 Transcribing chunk 3/7: chunk\_003.wav

Python(97580) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 35s

🚀 Starting Parakeet MLX transcription...

Python(97581) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 3/7 - Progress: 43.0% (Elapsed: 15s, Est. remaining: 20s)

✅ Chunk 3 complete: 25.1s actual (RTF: 0.084x, chars: 5126)

🎯 Transcribing chunk 4/7: chunk\_004.wav

Python(97609) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 15s

🚀 Starting Parakeet MLX transcription...

Python(97610) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 4/7 - Progress: 99.0% (Elapsed: 15s, Est. remaining: 0s)

🔄 Chunk 4/7 - Progress: 99.0% (Elapsed: 30s, Est. remaining: 0s)

🔄 Chunk 4/7 - Progress: 99.0% (Elapsed: 45s, Est. remaining: 0s)

✅ Chunk 4 complete: 60.2s actual (RTF: 0.201x, chars: 5500)

🎯 Transcribing chunk 5/7: chunk\_005.wav

Python(97680) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 15s

🚀 Starting Parakeet MLX transcription...

Python(97681) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 5/7 - Progress: 99.0% (Elapsed: 15s, Est. remaining: 0s)

🔄 Chunk 5/7 - Progress: 99.0% (Elapsed: 30s, Est. remaining: 0s)

🔄 Chunk 5/7 - Progress: 99.0% (Elapsed: 45s, Est. remaining: 0s)

✅ Chunk 5 complete: 50.4s actual (RTF: 0.168x, chars: 5116)

🎯 Transcribing chunk 6/7: chunk\_006.wav

Python(97772) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 5.0min, estimated processing: 35s

🚀 Starting Parakeet MLX transcription...

Python(97775) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 6/7 - Progress: 43.0% (Elapsed: 15s, Est. remaining: 20s)

🔄 Chunk 6/7 - Progress: 86.0% (Elapsed: 30s, Est. remaining: 5s)

🔄 Chunk 6/7 - Progress: 99.0% (Elapsed: 45s, Est. remaining: 0s)

🔄 Chunk 6/7 - Progress: 99.0% (Elapsed: 60s, Est. remaining: 0s)

✅ Chunk 6 complete: 75.4s actual (RTF: 0.251x, chars: 5248)

🎯 Transcribing chunk 7/7: chunk\_007.wav

Python(97880) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

📊 Chunk duration: 2.6min, estimated processing: 8s

🚀 Starting Parakeet MLX transcription...

Python(97881) MallocStackLogging: can't turn off malloc stack logging because it was not enabled.

🔄 Chunk 7/7 - Progress: 99.0% (Elapsed: 15s, Est. remaining: 0s)

🔄 Chunk 7/7 - Progress: 99.0% (Elapsed: 30s, Est. remaining: 0s)

✅ Chunk 7 complete: 40.1s actual (RTF: 0.261x, chars: 2281)

🔗 Step 4: Combining 7 transcription(s)...

🧹 Step 5: Cleaning up transcript...

🧹 Cleaning up transcript...

🧹 Cleaned transcript: removed 6 characters (potential commercials)

🗑️ Step 6: Cleaning up chunk files...

✅ Deleted chunk directory: audio\_cache/c6c2e5f6\_chunks

✅ Transcription workflow complete!

• Total time: 6.4 minutes

• Transcript length: 33,577 characters

• Overall RTF: 0.196x

===