

Captions on the fly

By Yash, Senior UX mentor,
liquid galaxy



Make Auto generated captions native to HTML Video Element

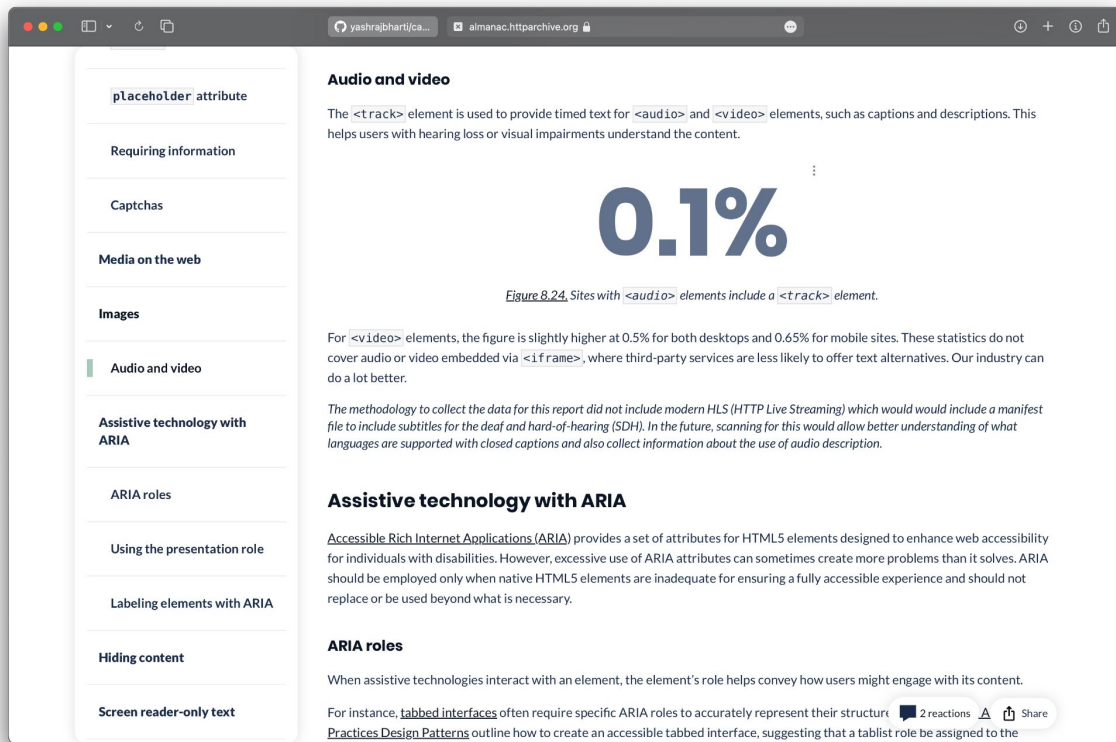
- Address the accessibility gap.
- Provide an implicit browser feature to generate captions.
- Allow user agents (UAs) to innovate and compete in improving the user experience.

The Problem

Problem Statement

Only 0.5% of videos have captions

Source: Web Almanac by HTTP Archive

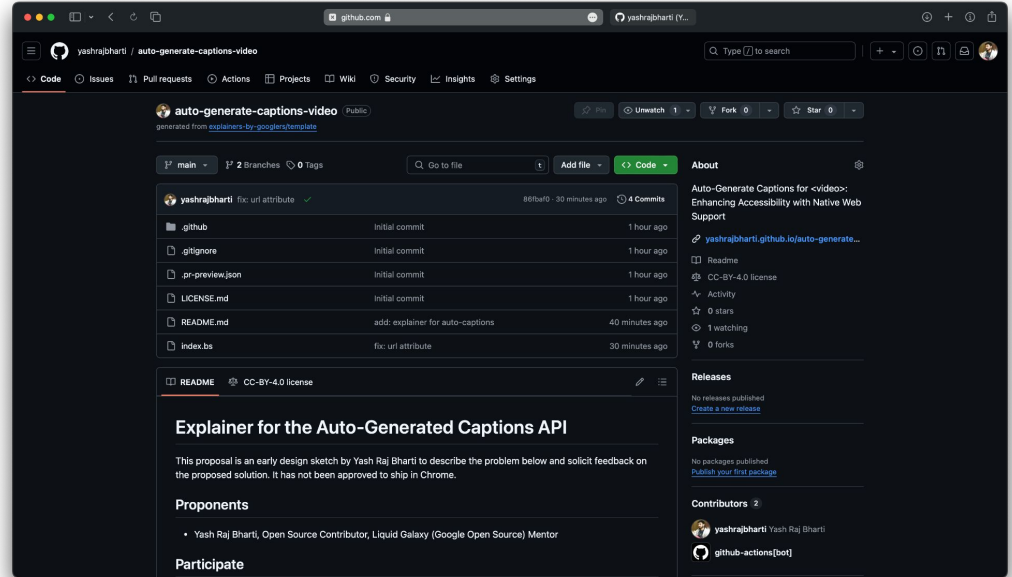


<https://almanac.httparchive.org/en/2024/accessibility#audio-and-video>

The Solution

The Explainer

- Details every important aspect, with lots of reviews
- It addresses the accessibility gap
- Overall feedback - positive



<https://github.com/yashrajbharti/auto-generate-captions-video>

From Chrome

- A new Chrome Gemini nano SST transcriber
- Existing Built-in APIs that can be used
 - Chrome translation API
 - Rewrite / summarizer API



The Feedback

Feedback

DEC 2, 2024




Thomas Steiner, PhD (He/Him) • 7:24 AM

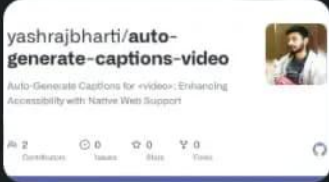
I can see how this could be useful. Questions to ask yourself: what if this leads to even less people manually creating subtitles? Does it make sense for every user who watches a video to run the same subtitle creation steps over and over again, or would it be better to just run this once on the server and then deliver the same subtitles to everyone? How would it deal with a translated language that has very long words? What if the generated subtitles don't fit the originally spoken parts? Just some random ideas to help you make this more solid.




Feedback


 **Adam Argyle** · 11:39 PM

i bet you could prototype this with a chrome extension that used the window.ai API to gen captions on the fly <https://github.com/yashrajbharti/auto-generate-captions-video>



GitHub - yashrajbharti/auto-generate-captions-video
github.com

 **Yash Raj Bharti** · 11:54 PM

Sure! 😊 Thanks for the feedback 

Feedback




I am OOO hanging with my family for the holiday

 1 

 **Dirk Ginader**  (He/Him) • 3:01 PM
I just stared your github repo

 **Yash Raj Bharti** • 3:02 PM
Have a happy holidays! 😊

 **Dirk Ginader**  (He/Him) • 3:02 PM
I truly love the idea of trying to fix this issue


Dirk Ginader  Active now   

yashrajbharti/auto-...
github.com


● lots of good thinking went in this already 👍

 **Dirk Ginader**  (He/Him) • 3:07 PM

● idea: to reduce unnecessary energy waste by producing the captions on demand over and over again could we propose a centralized storage?

 1  1 

(could be like a gist)

 **Yash Raj Bharti** • 3:14 PM
I find it nice, rather than running the caption generation over and over, to get it once through and deliver to everyone. I'll make a deeper research and add all perspectives on this, thanks for taking out the time. And yes, would love to see this getting done 😊

 **Dirk Ginader**  (He/Him) • 3:15 PM
love your energy on this! let's make it happen! 😊

Feedback



sideshowbarker  11:55 AM

December 16th, 2024 ▾

Hi Yash. What occurs to me about the problem case described in the explainer is that it seems like UAs don't need to have an explicit `track` element and new attribute in order to provide an opt-in feature for users to have auto-generated captions displayed.

What I mean is, it seems like if a `video` element has `track` element, then the UA can on its own provide a user option to users have generated captions displayed for the video. It's not clear to me what additional benefit there would be to users by having the document author specify that they want captions auto-generated.

I mean, in contrast, if the document author *does* have their own captions they want to provide to the user, then they can do that by including a `track` element. That seems like the entire (existing) meaning of the `track` element: It's a statement that "Captions exist for this video, and this element provides them."

So in that light, it seems like having a new case where a `track` element ends up stating "Captions *don't* exist for this video. I want to the UA to generate them." kind of runs counter to the entire purpose of the `track` element.

That's why it seems better to me to just leave it up to the UA to decide what to do when there's no `track` element. It's a class of feature that UAs can actually compete on, to provide a better user-experience around. It doesn't seem to me like something we need to standardize any behavior for.



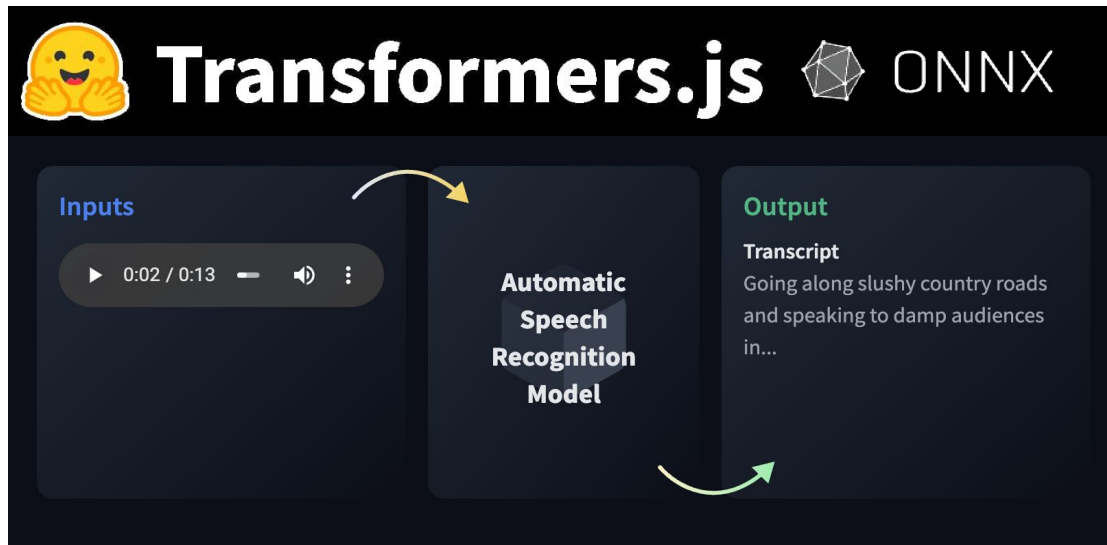
All that said, it's clear that your explainer addresses a real problem.



Live Demo

Demo Chrome Extension

- Works on the client side
- Uses Open neural network exchange format (onnx)
- Hugging face + Whisper to transcribe locally



<https://shorturl.at/HEJgH>

Usability Testing

Usability Test for Special needs

Speech-to-Text Transcriber: Append Mode

[Base Example](#) [See Example 1](#)



Clive Loseby

'Blind People Do Not Use The internet'
How can they? They cannot see the screen'

TEDx LukelyBrook

TEDx LukelyBrook
x = independently organized TED event

[MUSIC] >> Hi everyone, I'm Clive Loesby and I am passionate about changing the world. One website at a time, as Dale said, I was in

Usability Test (Hindi)

Feedback - positive

Pain points - amount of words
could sometimes be too long


Speech-to-Text Transcriber: Append Mode

[Base Example](#) [See Example 1](#)



यह वेबसाइट आपको सबसे अच्छा ऑनलाइन अनुभव देने के लिए कुकीज का उपयोग करती है। इस पर सहमति देने के लिए इस लिंक का अनुसरण करें। क्या सभी को पता है कि मैं किस बारे

Feedback (Sidenote)

 Pinned by sideshowbarker



Yash Raj Bharti 2:24 PM

I am currently raising the issue of Low Adoption of `<track>` Element, addressing that Only ***0.5% of web videos*** include them (Source: HTTPArchive [Web Almanac 2024 Report](#)) and would love to get any feedback to address the problem. The issue is raised here: <https://github.com/whatwg/html/issues/10866>



sideshowbarker  1:42 PM

Thanks for the update — that's all great news

Say Hi to Kenji for me



I know him well



Next Steps

What's next

- Building auto captions addresses a real problem
- An interesting challenge that can be solved by Chrome AI APIs
- I am passionate about it, and would keep contributing on it
- With a strong focus on good UI and leveraging client side AI, would love to help the Chrome team.