3PLAYMEDIA

BEGINNER'S GUIDE TO **CAPTIONING**
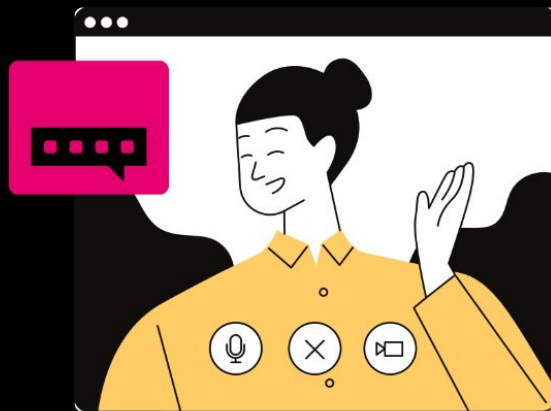
TABLE OF CONTENTS

WHAT

HOW

WHERE

WHY

CONCLUSION

WHAT

ARE CAPTIONS?

What are captions?

Defining captions

Captions are text that has been time-synchronized with the audio track and appear on-screen while a video is playing. Captions can be **open** or **closed**: closed captions aren't visible until activated by the viewer, while open captions are hard-coded into the video and visible to all viewers.

The French Chef with Julia Child was the first show to air with open captions in 1972. By the 1980s, captions were mandated by the Federal Communications Commission (FCC) for all broadcast television in the U.S. The purpose of captions is to display dialogue and describe relevant non-speech elements (like sound effects and speaker identifications) that are necessary to a viewer's understanding – especially if the viewer cannot hear the audio.

1972

Captions, subtitles, & transcripts

Though captions, subtitles, and transcripts may be similar, they actually refer to 3 distinct ideas:
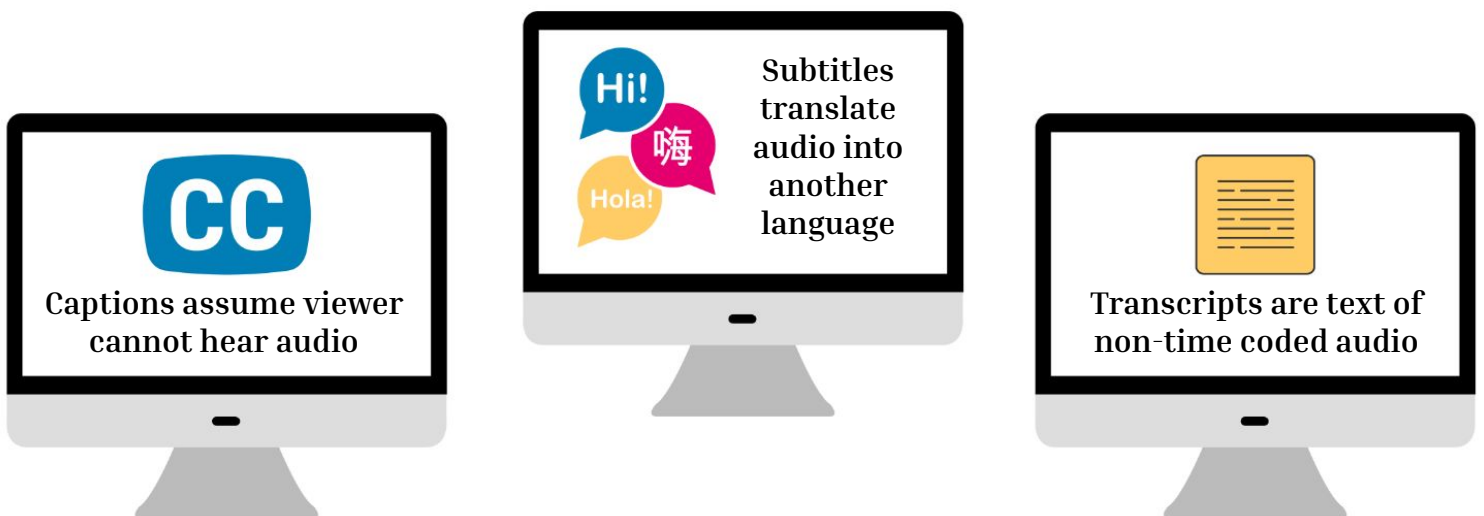
CAPTIONS

Assume the viewer cannot hear, and display time-synchronized words in the same language that is spoken in the video.

SUBTITLES

Assume the viewer can't understand the language. and display a translation of the spoken words but exclude non-speech elements.

TRANSCRIPTS

Simply the text version of the audio in a video. Not time-synchronized, but often the first step in creating captions.

CC

Captions assume viewer cannot hear audio

Hi!
嗨
Hola!

Subtitles translate audio into another language

Transcripts are text of non-time coded audio

In some countries (like the UK) the word "subtitles" is used interchangeably to refer to captions & subtitles as defined above.

HOW

DO YOU CREATE CAPTIONS?

How do you create captions?

DIY captions with YouTube

One of the easiest ways to create captions on your own is through YouTube's caption editing interface. YouTube uses automatic speech recognition (ASR) to initially transcribe the audio of your video. Since ASR captions are computer-generated, they are generally only around 60-70% accurate. Nevertheless, they can save you a lot of time by generating a rough draft of the audio transcript.

Another (human!) layer of editing can usually fix any ASR mistakes, and you can edit YouTube's autocaptions in their Creator Studio. The studio gives you the ability to add or correct punctuation, correct misspelled words, adjust timings, and more in order to ensure your captions are as accurate a possible. When you're done, you can download the captions if you're planning to use the transcript elsewhere.

Consider captioning vendors

Creating your own captions is certainly a cost-saver, but it might make more sense to work with a professional captioning service – if your captioning needs include *many* videos or *lengthy* videos, consider hiring a captioning vendor!

A full-service captioning solution gives you peace of mind by taking care of all the logistics for you – including legal compliance, seamless automation, and accessibility from all angles. At 3Play Media, we offer strict turnaround SLAs according to your schedule & guarantee 99% average accuracy (no matter how quick you need a video captioned).

However, not all captioning solutions are created equal. When comparing vendors, learn more about their captioning processes so you can avoid crowdsourcing, untrained transcriptionists, and more (oh my!).

Quality standards & best practices

To create an equitable experience for viewers who cannot hear audio, captions and transcripts should be easy to read. The industry-wide standard for word-to-word spelling accuracy is at least 99%, and that includes proper punctuation, grammar, and accurate spelling. Captions should adhere to the following standards to comply with best practices:

NON-SPEECH AND SPEAKER IDs

In addition to accurate text transcription, captions should also include relevant non-speech sound effects and speaker identifications.

CAPTION APPEARANCE

Font should be non-serif. Only 1-3 lines of text should appear on-screen at one time, with a maximum of 32 characters per line. Captions should appear for a minimum of 1 second (but not linger too long after a speaker is finished), and should be located in the lower third of the screen. If captions are blocking any text or other information, they should move.

VERBATIM VS. CLEAN READ

In broadcast & entertainment, every scripted or intentional "um," pause, stutter, and stammer should be included for a verbatim transcription. In other content like lectures & live presentations, these sounds should be removed for a clean read transcript.

On accuracy rates

As we mentioned previously, the industry standard is that captions and transcripts must be at least 99% accurate. But how is accuracy measured when it comes to captioning? The chart below demonstrates the concept of caption accuracy more clearly:

Video Transcription Accuracy Rates

Word-to-word accuracy	1 of X words incorrect	8-word sentence accuracy	10-word sentence accuracy
50%	1 of 2	0%	0%
67%	1 of 3	4%	2%
75%	1 of 4	10%	6%
85%	1 of 7	27%	20%
90%	1 of 10	43%	35%
95%	1 of 20	66%	60%
98%	1 of 50	85%	82%
99%	1 of 100	92%	90%

FOR EXAMPLE:

With a word-to-word accuracy rate of 95%, you can expect 1 out of every 20 words to be inaccurate. This means every word in 1a sentence has a 1 in 20 chance of being wrong – a 10-word sentence with a 95% word-to-word accuracy rate only ends up being 60% accurate!

On speech recognition software

As we said, automatic speech recognition (ASR) software can help you transcribe spoken audio extremely quickly. ASR is commonly used in some live television captioning or to transcribe audio for some online video players, like YouTube.

Unfortunately, even with the latest technology, caption accuracy rates using ASR software rarely exceed 80-90%. ASR often makes errors that make sense acoustically, but linguistically – for example, mistaking a "four-story" building for a "forester" building. Similarly, since ASR cannot understand the context of spoken audio, it is more likely to mix up words like "can" and "can't," which could significantly impact a viewer's ability to understand.

Videos with multiple speakers, speakers who have accents, fast speakers, background noise, and other audio that is complicated or hard to understand will ultimately produce more captioning errors when relying solely on ASR.

? Most ASR programs also do not understand some proper nouns, names, and brands, which leads to some words going uncapitalized, while others are capitalized on accident. Speaker labels and punctuation – like commas and periods – are also commonly left out.

WHERE

DO YOU PUBLISH CAPTIONS?

Where you do publish captions?

Captions are supported on most devices that allow video publishing. Many common online video players and platforms support caption functionality, and some include advanced user controls that allow viewers to customize the on-screen appearance of captions.

With certain online video players (like YouTube and Vimeo), captions are published as a **side-car file**. A side-car file is a separate file containing captions and timing information that is uploaded along with the original video. When captions are uploaded like this, they play along with the video and can be turned on or off.

Sometimes, captions will be **encoded** in a video. In this case, captions are encoded into the original video and can be turned on or off. Common use cases for caption encoding include kiosk videos or other offline videos.

Open captions include no option to toggle captions, and cannot be turned off. Open captions are burned directly into the original video to create a single file.

Finally, when publishing video online, **caption integrations** between vendors and video platforms can automate and simplify the caption publishing process.

Common caption formats

The kind of caption file you need to publish depends on the video player you're working with. Below are examples of two of the most commonly used caption file formats, SRT and WebVTT.

SRT

```
1
00:00:00,000 --> 00:00:04,440
These are caption frames
from a sample file.

2
00:00:04,440 --> 00:00:06,570
They're timed well and
spelled wonderfully.

3
00:00:06,570 --> 00:00:08,540
Plus, it's rather short,
so examples are concise!
```

WebVTT

```
1
00:00:00.00 --> 00:00:04.440 align:middle line:90%
These are caption frames
from a sample file

2
00:00:04.440 --> 00:00:06.570 align:middle line:90%
They're timed well and
spelled wonderfully.

3
00:00:06.570 --> 00:00:08.540 align:middle line:90%
Plus, it's rather short,
so examples are concise!
```

WebVTT

A Web Video Text Track (WebVTT) file was originally based on SRT file, but allows additional functionality to include descriptions & metadata with each frame. This information is not visually displayed to the viewer.

SRT

The SubRip Subtitle (SRT) file format is very common for online video players & very easy to read. It's a basic text file that includes caption frame numbers, time codes, and lines of text.

WHY

SHOULD YOU CAPTION?

Why should you caption?

ACCESSIBILITY

Millions of Americans are either d/Deaf or hard of hearing, and captions make videos accessible to individuals who cannot hear the audio. Initially, this was the primary reason for the introduction of captions, but there are numerous other benefits in addition to accessibility.

VIEWER COMPREHENSION

Captions ensure ANY viewer can understand dialogue spoken in a video (not just those with hearing loss!), especially when thick accents, background noise, and unfamiliar words/phrases get involved. In A joint study with Oregon State University, we found that 80% of students say captions are helpful as a learning tool – regardless of hearing status. Captions can also aid English learning & comprehension for people who speak English as a second language.

VIDEO SEARCH FUNCTIONALITY

Associating an interactive transcript with your video allows users the functionality to search a keyword within a transcript, and jump to the corresponding time in the video the word is spoken. An MIT survey found 97% of students found this video search feature enhanced their learning experience.

VIEWING FLEXIBILITY

In certain circumstances, users don't always want to play the audio of a video out loud on their devices – for example, at the library, on public transit, in the office, etc. Noise sensitive environments like this make captions necessary, especially without headphones. Facebook found that 80% of users react negatively when "feed-based mobile video ads play very loudly when people aren't expecting it," and captions can increase video ad view time by 12% (on average).

SEO: SEARCH ENGINE OPTIMIZATION

Since search engines can't watch your video, captions & transcripts provide metadata that Google can index, allowing it to read spoken audio content and giving the video an SEO advantage. Discovery Digital Networks found an overall increase of 7.32% in lifetime views after adding captions to YouTube videos, and This American Life found that 6.26% of all unique search traffic visitors landed on a page including transcripts. You can also get these SEO benefits by injecting the transcript into the HTML of a webpage, making your video more discoverable without publishing a transcript on the same page.

CONTENT LONGEVITY

The transcript of a video recording is transcribed can easily be used to create other media content. At the University of Wisconsin, 50% of students repurposed transcripts of in-class videos as study guides for their courses. On the flip side, professors can use transcripts to create additional course materials. Many businesses also repurpose transcripts of recorded events (like webinars) to create marketing content, including infographics, white papers, case studies, and more.

TRANSLATION

In addition to captions, video transcripts are the first step towards creating video subtitles, which make your video content accessible to global audiences. In order to do this, a transcript in the source language is required before it can be translated into the target language.

LEGAL COMPLIANCE

In many circumstances, captioning is required by disability and civil rights laws to ensure that video content is accessible to those who are d/Deaf or hard of hearing.

CAPTIONING LAWS & LAWSUITS

In order to provide the necessary accommodations to individuals with disabilities – as well as maintaining legal compliance and avoiding potential lawsuits – it's crucial to stay on top of the legal requirements surrounding web accessibility.

ANTI-DISCRIMINATION LAWS

United States anti-discrimination laws require closed captioning for video, and those laws include the **Americans with Disabilities Act** (ADA), **Sections 504 and 508** of the Rehabilitation Act, and the **21st Century Video Accessibility Act** (CVAA).
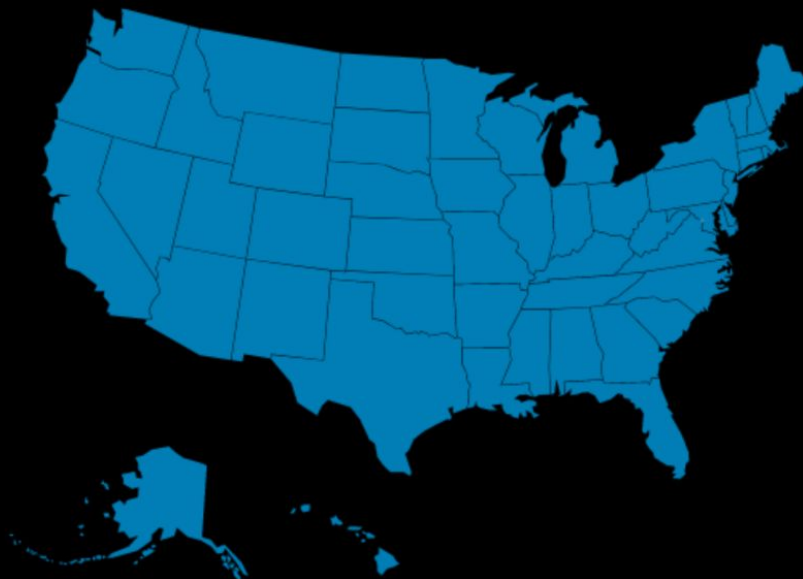
The Americans with Disabilities Act (ADA)

The ADA **guarantees equal opportunity for individuals with disabilities** in employment, services provided by state and local governments, places of public accommodation, commercial facilities, and transportation. Two sections of the act that most directly affect web accessibility are **Title II (Public Entities)** and **Title III (Places of Public Accommodation)**.

> ## Places of public accommodation
>
> ...a private entity that owns, operates, leases, or leases to a place of public accommodation
>
> can include a wide range of entities – such as restaurants, hotels, pharmacies, & retail stores

– The ADA Questions & Answers

While the ADA doesn't specifically mandate or name video captioning as a legal requirement, the legislation's vague wording has been disputed in several lawsuits against organizations – including Netflix, Hulu, Amazon, and institutions like MIT, Harvard, and UC Berkeley – for failing to provide captions on videos. In these suits, the organizations & institutions were found to be in violation of Title III, which prohibits disability discrimination by "places of public accommodation."

The Rehabilitation Act: Sections 504 & 508

Section 504

This section of the Rehabilitation Act **declares and protects the civil rights of people with disabilities** by requiring that federal (and federally funded) organizations make accommodations for equal access. This means for users of a federal or federally funded website, product, service, etc. who are deaf or hard of hearing, video captioning must be provided.

Section 508

Signed into law in 1998, this section requires that federal agencies **make electronic communications and information technology equally accessible**. Therefore, all film, video, multimedia, and/or information technology produced or procured by federal agencies must include captions for audio. It is also worth noting that many state legislatures adopted 'little 508s' which directly require compliance with Section 508, so the same requirements can extend beyond federal agencies.

A refresh of Section 508 guidelines regarding information & communication technology (ICT) took effect in January 2018, and assigns specific WCAG 2.0 success criteria to reinforce & modernize captioning requirements.

What does the Section 508 refresh mean for captioning?

SECTION 508 CHANGES

- Categorization by functionality instead of product type
- Revisions to improve ICT usability, including interoperability with assistive technologies
- Clarification on types of ICT not covered
- Correlation of WCAG 2.0 Level A and AA standards to all Section 508 requirements

WCAG SUCCESS CRITERIA

The new standards modernizing the requirements for closed captioning of prerecorded video content are as follows:
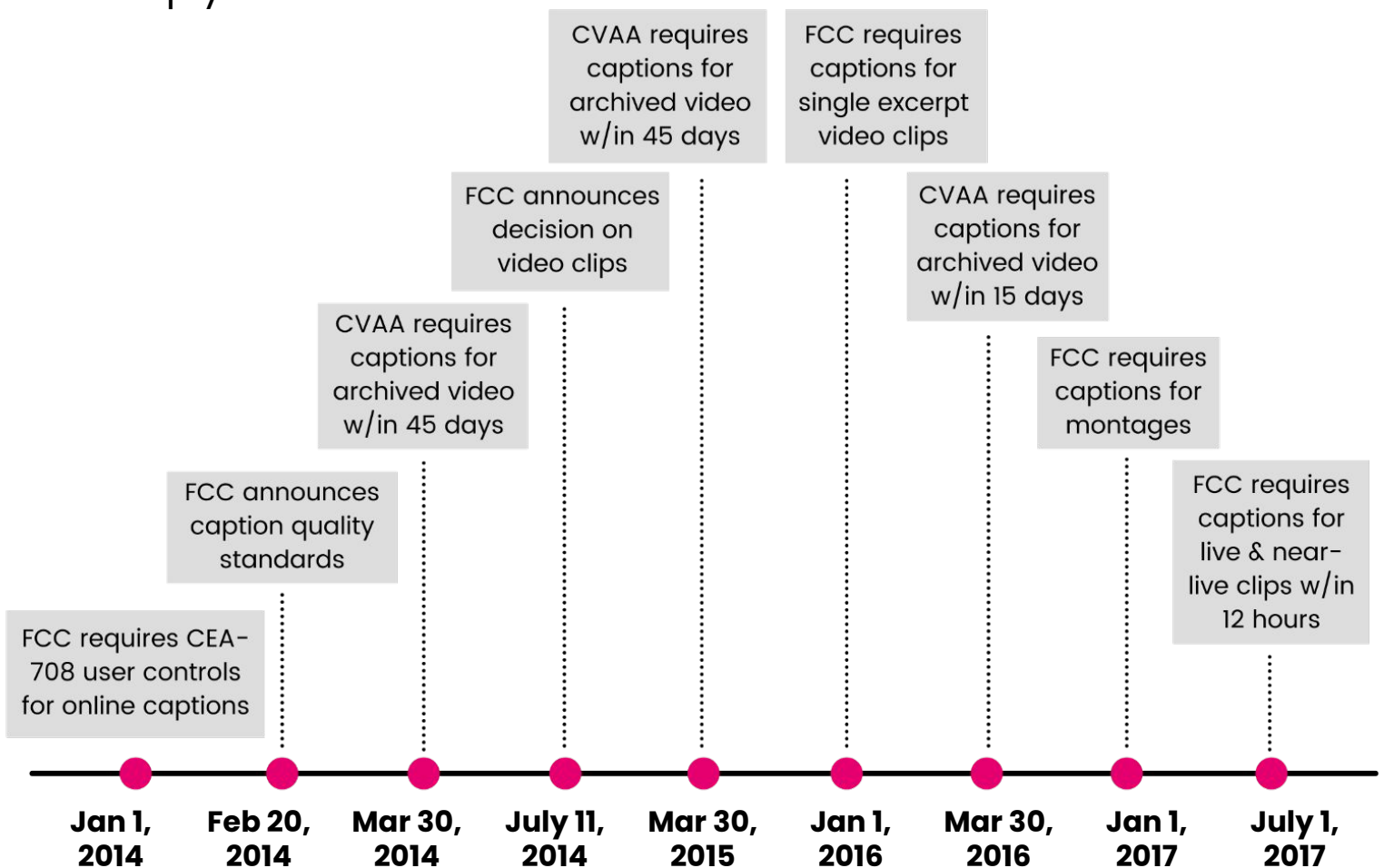
- Success criterion 1.22
- WCAG 2.0 Level A
- 508 Section 1194.22(b) and .24(c)
- Synchronized captions are provided for non-live, web-based video

WCAG 2.0 Guidelines

The Web Content Accessibility Guidelines (WCAG) 2.0 are the international standard for web accessibility & require captioning for time-based media (guideline 1.2). Under the Section 508 refresh, they also serve as the legal requirement that captions be provided for all prerecorded and live audio content in synchronized media. WCAG 2.0 compliance is the standard recommended by web accessibility experts, and is often required by state law.

21st Century Communications & Video Accessibility Act (CVAA)

Signed by Obama in 2010, the CVAA **outlines clear accessibility rules for internet video programming** and clips that previously aired on U.S. television. This act is enforced by the Federal Communication Commission (FCC) and has been implemented in phases over several years so video creators could adequately comply.

CVAA requires captions for archived video w/in 45 days

FCC requires captions for single excerpt video clips

FCC announces decision on video clips

CVAA requires captions for archived video w/in 15 days

CVAA requires captions for archived video w/in 45 days

FCC requires captions for montages

FCC announces caption quality standards

FCC requires captions for live & near-live clips w/in 12 hours

FCC requires CEA-708 user controls for online captions

| Jan 1, 2014 | Feb 20, 2014 | Mar 30, 2014 | July 11, 2014 | Mar 30, 2015 | Jan 1, 2016 | Mar 30, 2016 | Jan 1, 2017 | July 1, 2017 |

As of September 2013, all broadcast video that previously aired on U.S. television must include captions if then published on the internet. Stricter CVAA compliance deadlines have since been passed for video montages and live footage.

CAPTIONING LAWSUITS

National Association of the Deaf, et al v. Netflix

In 2012, the National Association of the Deaf (NAD) et al. argued that Netflix barred deaf users from enjoyment of their video streaming service by not including closed captions. Netflix unsuccessfully countered that ADA Title III did not apply to online businesses, however, the court ruled that Netflix stands as a "place of public accommodation." Therefore, ADA regulations do apply, and the case was settled after Netflix agreed to caption all videos by 2014.

EEOC and NAD v. FedEx

In October 2014, the Equal Employment Opportunity Commission (EEOC) filed suit against FedEx Ground for ADA violations including the failure to provide closed captioned training videos, sign language interpreters, and other reasonable accommodations to their d/Deaf and hard of hearing employees nationwide. The NAD joined the suit in 2015, and a 2016 motion by FedEx to dismiss the case was denied. The case was settled in May 2020, with FedEx issuing a consent decree to pay over $3 million and implement programmatic relief moving forward.

What is a **consent decree**?

Consent decrees typically arise from legal settlements, and outline the steps the defendant organization agrees to take to remedy the situation. In turn, this gives other organizations a checklist with which to compare their own accessibility compliance.

What else can the **DOJ** do?

After receiving formal complaints about website accessibility issues, the Department of Justice (DOJ) or Department of Education's Office of Civil Rights (OCR) will investigate further and urge the organization to make the necessary changes. These are often referred to as **"dear colleague" letters**, and can be very helpful to organizations in a given industry as a detailed example of exactly what the DOJ and OCR consider satisfactory compliance criteria. Though these letters don't necessarily set a legal precedent, they do allow the industry to stay informed on the DOJ and OCR's major stances, as well as sharing the standards, requirements, and solutions they believe organizations should be following related to web accessibility.

After receiving a dear colleague letter, UC Berkeley removed their entire library of 20,000+ publicly accessible online course videos & audio files. This is the "all or nothing" approach – an unfortunate outcome of accessibility conflicts.

Conclusion

As technology continues to advance, video accessibility becomes increasingly important both in terms of legal requirements & the needs of our world. Captions have become more ubiquitous with the growth of online video, and legislation continues working to close the accessibility gap by requiring captioning in more places where video is found. As video continues to dominate the web and other public spaces, captioning will become less of an afterthought and more of a universally beneficial necessity in an accessible world.

3PLAYMEDIA ABOUT US

3Play Media provides cost-effective, premium quality captioning, transcription, and audio description to more than 2,000 customers across industries like higher education, enterprise, entertainment, media, and government. We offer a future-proofed solution for video accessibility that provides peace of mind every step of the way – including our flexible APIs, video player and platform automations, simple plugins, and an all-in-one, user-friendly platform.

3Play's Captioning Process

We used a patented captioning process which combines the best of both worlds: ASR and human editing. Our innovative technology allows us to offer competitive rates (including commitment-based discounts), and our multi-step quality assurance measures ensure we deliver premium quality captions, subtitles, and transcripts much more efficiently.
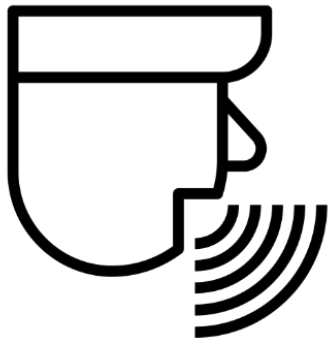
SPEECH RECOGNITION

HUMAN EDITING

HUMAN QUALITY REVIEW

Our average measured accuracy is 99.6%, and we guarantee the industry standard (99%) every time – even in instances of poor audio quality, multiple speakers, difficult content, or speakers with heavy accents. Our staff of 1,500+ transcriptionists gives us the flexibility to assign complex or technical content to transcriptionists with a discipline-specific expertise or a particular accent familiarity, so we can process a broad range of complex content with consistently high quality.

3PLAY MEDIA

3Play Media provides an innovative platform for closed captioning, transcription, and audio description services to make video accessibility easy. We're based in Boston, Massachusetts and have been operating since 2007.

Born in Boston

77 N Washington Street
Boston, MA 02114

Drop us a line!

Website: www.3playmedia.com

Email: info@3playmedia.com

Phone: 617-764-5189

Find us online:

Follow us **@3playmedia** on all platforms for more resources on web and video accessibility.