Subtitles, Captioning and Transcription Services:

Facts, Resources and Guidance for the HGSE Community

Multimedia, i.e., video and audio, delivers rich academic, research, and administrative content to our shared audiences. Captions and transcriptions further enrich these materials by providing textual overlays and supports that make multimedia resources more discoverable, searchable, and most importantly, more accessible, to the end user.

Historically, captions and transcriptions were largely human generated – a time consuming process that often-delayed delivery of multimedia content to the intended audience. In recent years, Automatic Speech Recognition (ASR) has matured dramatically, allowing machine-generated captioning and transcription to be created in minutes and seconds. ASR augmentation is natively supported in a variety of platforms and services in use across our community: most notably Panopto, Zoom, Microsoft Teams, and Adobe Premiere. These ASR services are generally very good, with accuracies ranging from 70% to 95%, depending largely on the format of the multimedia experience. There are occasions, however, when ASR falls well below these thresholds. Group discussions, contributors whose native language is not English, and inconsistent contributor volumes may have lower accuracies.

The need for these services is an evolving discussion. The University is committed to Digital Accessibility across all public-facing content as well as all digital content created after June 1, 2023.[i] While the commitment is broad, the need for accessibility compliance across every digital asset is far from absolute. In many cases, digital content does not need to be captioned, but doing so enriches the experience for learners. Additional guidance from Harvard's Digital Accessibility office is provided below to assist with understanding when captioning is or is not required.

Definitions

Subtitles. Although this term is used interchangeably with captions, subtitles are intended for viewers who do not understand the language in the video but are able to hear the audio.

Captions. Time-synchronized text that accompanies visual content designed for deaf and hard-of-hearing audiences. Captions are intended to communicate all audio information – which may include sound effects and other non-speech content.

Live Captioning. Real time captioning for streamed live events, i.e., Zoom Meetings and Webinars.

Audio Descriptions. Audio descriptions are used to describe the video for visually impaired or blind audience members.

Transcriptions. Intended strictly for audio, this is simply speech-to-text. Audio and video players will generally work to synchronize a transcription panel with the video content, should that be a preferred feature.

Translations. Conversion of spoken content in an audio file from one language to another.

Service Options

Harvard University and HGSE Information Technology offer a variety of solutions to meet the needs of our communities. Please review the descriptions of these services, their applicable use cases, and relevant constraints to better understand options available to you, your teaching teams, and staff.

In the following service descriptions, the term caption is used interchangeably with subtitles, despite the formal difference in their definition. Nearly all vendors will refer to captions and captioning in their product lines.

Automated Closed Captioning/Automated Subtitles. Machine-generated captions may be generated at no cost through several options. Real-time machine generated live captioning is natively supported in Zoom and Microsoft Teams. Pre-recorded assets can be uploaded to Panopto to generate captions. These products create “closed” captions, meaning that the captions/subtitles may be turned on/off by the audience. The accuracy of machine generated captions is dependent on the source content. Video with a single speaker on a dedicated microphone in a relatively low noise environment will generally produce highly accurate output, i.e., over 95%. Video with non-native speakers, overlapping conversation, and significant background noise creates challenges for automated captioning. Accuracy may drop significantly.

Self-Service. Captions can be manually added by anyone with editing privileges to the video. We recommend generating a first draft of the caption file using the resources above followed by a manual review by someone with experience/expertise in the subject area. In many cases, only minor edits are required. The additional benefit is the contextual understanding of the internal editor, i.e., someone who can add additional descriptive text, as needed.

Vendor-Outsourced Captioning. HGSE partners with two vendors for professional captioning services. These service providers typically work with our pre-recorded material and use ASR to generate a draft of the captioned output. The service provider will manually review the

captioning for accuracy and timing before the product is returned to us.[1] HGSE IT manages a central account with our vendors and requests for services are managed through a structured workflow. Please note that these services do not include descriptive text unless explicitly requested.

HGSE Captioning Guidance

HGSE defines our captioning standards as follows:

- **Tier 1**. Highest quality captioning. Highly accurate transcription, proper punctuation, speaker identification, and the identification of meaningful sounds other than speech. Includes audio descriptions, if required.
    - **Fulfilled through**: Self-service or Vendor
    - **Approximate Vendor Cost**: $11 to $16 per minute under existing Harvard contracts.
- **Tier 2**. Professional captioning. Highly accurate transcription, reviewed and corrected by people.
    - **Fulfilled through**: Self-service or Vendor
    - **Vendor Cost**: $1.60 per minute for standard service, i.e., 4 business day turnaround.
- **Tier 3**. Automatic Speech Recognition (ASR) Captioning. Transcription services provided through automated services. Review by people is recommended, but not required.
    - **Fulfilled through**: ASR and/or Self-service.

Multimedia assets created for students, instructors, or other learners with a documented accommodation request	Tier 1
Multimedia assets used in asynchronous content delivery for PPE courses	Tier 1 or 2
Multimedia assets used in academic (degree facing) courses repeatedly and over time, and/or for critical course content.	Tier 2
Multimedia assets included as part of asynchronous content delivery in academic (degree facing) coursework.	Tier 2
Multimedia assets created by HGSE to support co-curricular activities.	Tier 2
Multimedia content produced for public web sites and other public facing audiences.	Tier 2
Live streaming events.	Tier 2

[1] Enhanced services are available for difficult video and other needs.

Recordings resulting from automated lecture capture in HGSE classrooms and event spaces.	Tier 3
Recordings produced from Zoom webinars and meetings or Microsoft Team meetings.	Tier 3
Multimedia assets in academic courses used for optional content.	Tier 3

Service Level Agreements

Vendor SLA terms include a 4 business day turnaround period, with an accuracy of 99%. Expedited service is available, but please take note of the additional costs and constraints on duration of the asset being captioned.

Turnaround Service Levels	Max Duration per File	Surcharge
Expedited	2 hours	+ $0.50 per min
Rush	2 hours	+ $1.00 per min
Same Day	20 minutes	+ $2.00 per min
2 Hour	10 minutes	+ $5.00 per min

Funding

Professional captioning services are centrally managed and funded through HGSE IT. Audio Description surcharges attributable to a documented accommodation (ADA) requirement are funded separately.

Digital Accessibility Accommodation Requests

Individuals requiring accommodations to properly access and learn from multimedia assets may require enrichment to multimedia assets not covered by this guidance. Please consult directly with the HGSE Local Student Disability Coordinator to ensure your materials fully address these documented needs.

Captioning Request Workflow

All captioning requests must be made through our form-based workflow. The Multimedia Services team will manage the process of assigning approved captioning requests to the appropriate vendor. Completed captioning workflows will trigger an email to the requestor when they are completed.

https://harvard.az1.qualtrics.com/jfe/form/SV_03Bf7IjgP4iBbka

We recommend assets be uploaded to Panopto before a captioning request is made. Panopto provides the most direct method for making assets available to the vendor for captioning and has user friendly features for making edits of any final product.

---

[i] (Harvard University, 2023)