LILT | CaptionHub

Starter Guide to Multilingual Video Subtitling

Introduction

In the modern digital age, video content has become vital to the global enterprise, playing an increasingly important role in connecting with customers, prospects, employees, and other stakeholders. Enterprise video content generation has experienced exponential growth over the last several years and shows no signs of slowing.

The market for Global Enterprise Video Platforms is anticipated to grow 1.7x between 2022 and 2030[1]. In fact, video is expected to account for 82% of all internet traffic by 2023[2], and digital advertising spend on video content is projected to double between 2019 and 2023[3].

1.7x

Expected growth for Global Enterprise Video Platforms, 2022 and 2030

As enterprises continue to invest in video content with increasingly limited budgets, content teams must maximize the ROI of the content they develop - and subtitling of video content provides a strong mechanism to do so. A recent study found that 80% of consumers are more likely to watch a video if it contains closed captioning[4]. In addition to improving content engagement, subtitling in multiple languages can increase the reach of the already-developed video content.

As video subtitling becomes a necessity for all content teams, this guide is designed to help these teams better understand the nuances and complexities of multilingual video subtitling and provide access to industry best practices for optimized workflows and ROI with AI.

1. Regional Research Reports, Enterprise Video Platforms Industry Market, 2022-2030; 2. Cisco, VNI Complete Forecast Highlights, 2023; 3 Statistia, . Digital video advertising spending in the U.S. 2019-2023, 2023; 4. “Break the Sound Barrier with Digital Video Captioning.” Verizon Media, 14 May 2019.

Three Key Trends Impacting the Rise of Captioning

Why are global content teams increasingly investing in multilingual subtitling?

1 Greater engagement in new geographic markets

By adding multilingual subtitles, companies can grow the overall impact of their video content by reaching new audiences in their language of choice. Customers prefer to purchase products when information is written in their native language, and this same principle applies to video. Whether localizing subtitles for product demo videos, ads for social media, or other key media content, localized subtitles amplify the reach of video content.

2 Emphasis on global employee learning

Video continues to be the preferred way to deliver highly engaging, dynamic training for global employees. Many companies provide training videos, webinars, and e-learning courses to enhance their employees' skills and knowledge. By incorporating subtitles and captions in all of the languages of their employee base into these resources, organizations can make the content more inclusive, increase comprehension, and ensure that all employees can fully engage with the material.

3 Renewed focus for accessibility

Another driving force behind the increased use of subtitles and captions is the global push for accessibility and engagement. Initiatives such as “Turn on the Subtitles” have gained momentum in recent years, advocating for the use of subtitles as a means to improve literacy rates among children. This movement has highlighted the benefits of including subtitles in educational content and shown their potential to enhance reading skills by associating spoken words with written text.

Getting Started: Subtitling and Localization

It can be overwhelming to know where to start when localizing multimedia. While there are multiple approaches available - including manual processing in a video editor - if you're facing a large volume of video, you'll want enterprise-grade features such as task assignment, best-in-class AI, and seamless integration with industry-leading online video platforms. In this guide, we'll focus on enterprise-grade solutions such as those offered by CaptionHub, LILT's integration partner.

The key steps for a standardized customer workflow are

1. Importing Video Files

3. Translation

2. Transcription

4. Formatting/Editing

5. Exporting/Rendering/Publishing

Let's review each of these items in a bit more detail.

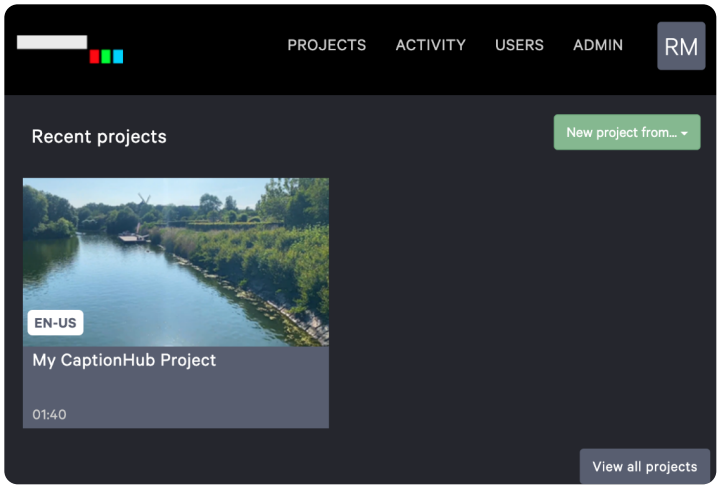
1 Importing Video Files

Start by importing the video file into the CaptionHub platform. This can be accomplished via several methods, including:

- **Manual Upload**: Uploading file by file from a computer
- **Link uploading**: Specifying a link
- **Programmatically**: Import a video via the API
- **Automated import**: For most enterprise-grade customers with large volumes of video to transcribe, direct import from an enterprise online video platform (OVP), will save vast amounts of time, and can be achieved through either:

Manually selecting the video for import	Creating an "Automation" process within CaptionHub to automatically import, transcribe, translate, and publish videos that match rules created by you

Once imported into CaptionHub, the video can be easily accessed in the dashboard. At this stage, it is a raw video file with no captions attached.



2 Transcription

Once the video is uploaded into the platform, the next step is to ensure that there are source-language captions. Available methods for this include:

1. **Pre-existing captions**: Upload a text log of what is said in the video (e.g., a plain text transcript)
2. **Timed text document**: Upload this document, which is essentially a plain text transcript with timing information (e.g. when the captions should appear in the video)
3. **Original language captions**: Use the caption editor and manually transcribe what is being said in the video or utilize one of CaptionHub's speech-to-text partners (see below for more details)

Speech-to-text

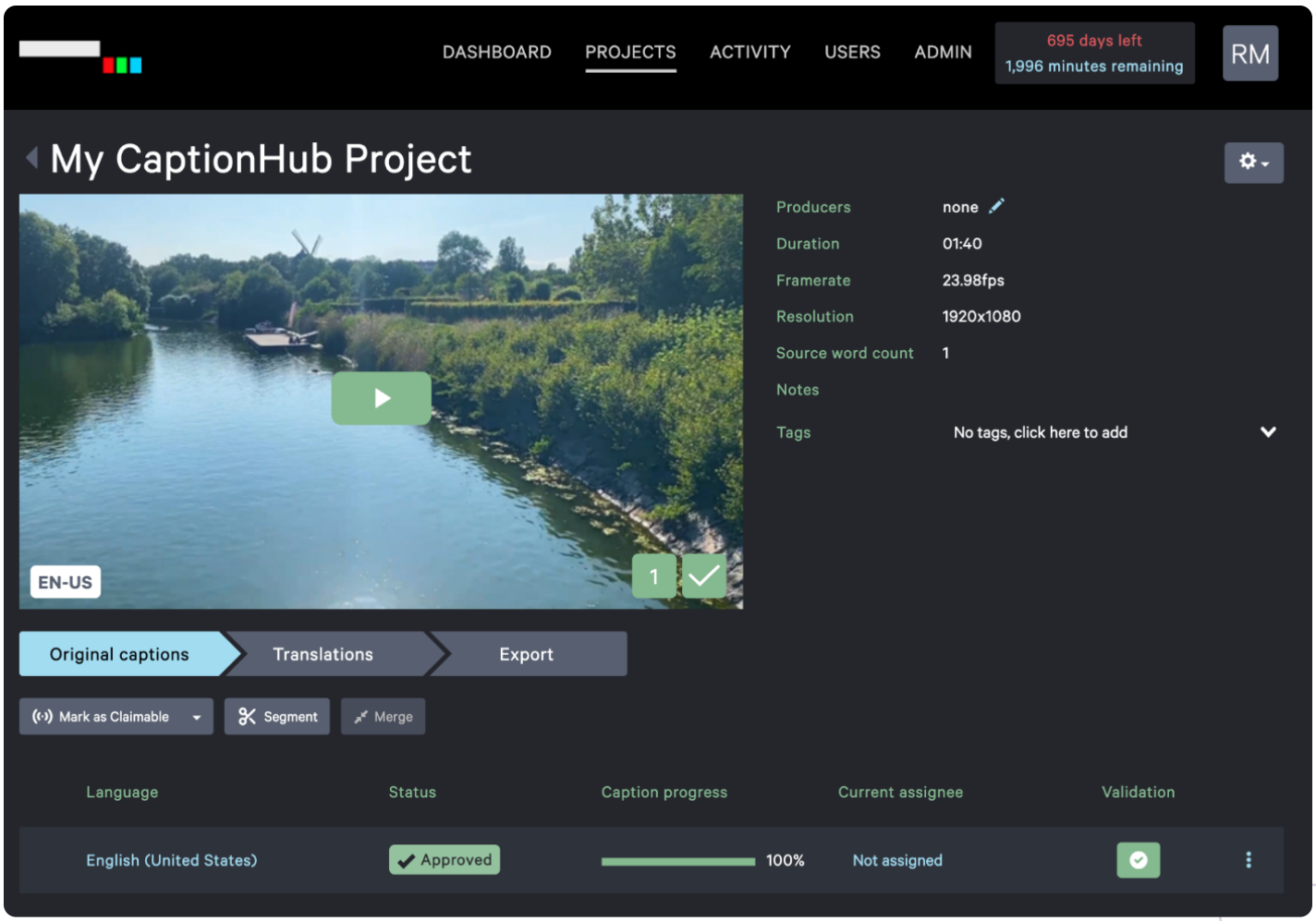
In speech-to-text, the video audio file is extracted and sent to one of CaptionHub's transcription providers for analysis and processing. It's possible to specify certain parameters, such as maximum lines per caption, minimum duration, maximum length, and more, all of which will be covered later in this guide. Once processed, the timed text captions are automatically populated within the CaptionHub platform, with captions timed to when they appear on screen using CaptionHub's proprietary technology.

On-screen identification is a useful option for auto-populating a second track of captions with on-screen text - signposts, computer screens, text messages, etc. This second track of captions can also be sent over for translation once ready.

As with regular text captions, manual transcription can take place within the editor. This is useful if there are a few on-screen texts (e.g. one scene has a shop sign that requires translation).

Alternatively, and particularly if there is a large volume of OST, then automatic OST recognition may be preferable. Similar to speech-to-text, automatic OST recognition automatically scans the video content for text and inserts this into the second track of captions - ready for translation.

Once the original language captions are ready and timed to the video footage, we can progress to translation.



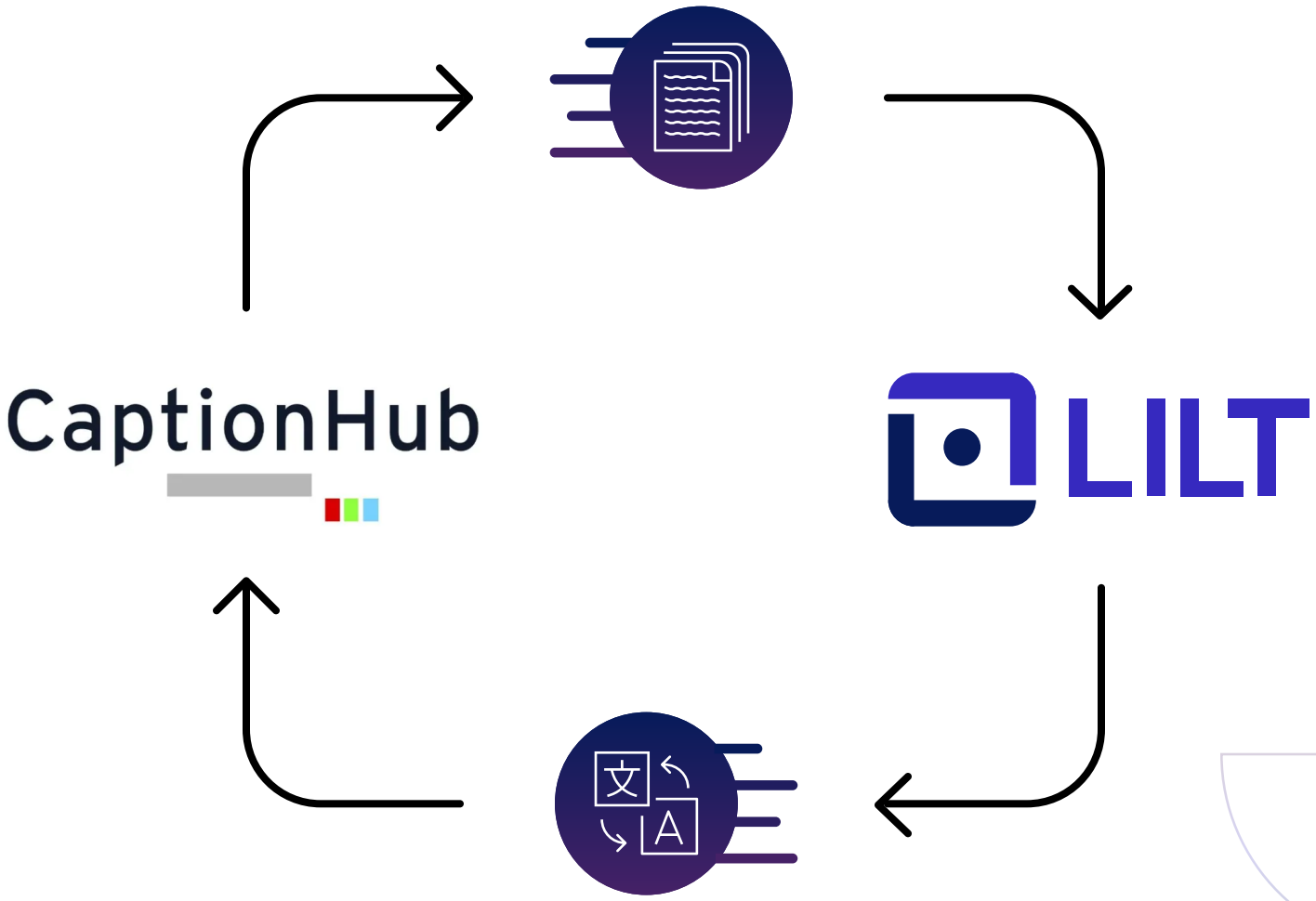
3 Translation

CaptionHub offers several methods for translating captioned source content. In some cases, pre-existing translations may already be available and can be simply uploaded into the video project. Alternatively, enterprises may have their own in-house linguists who will handle translation and can work directly in the CaptionHub interface.

However, the most common approach utilizes a CaptionHub translation solution partner, such as LILT, the leading AI solution for enterprise translation.

The automated LILT <>CaptionHub workflow offers a seamless and streamlined approach for this process. Simple steps activate this process:

1. Within the “users” section of the CaptionHub account, invite connectors@lilt.com as a Superuser
2. Go to the video project, add a translation, and assign connectors@lilt.com (with edit and view permissions enabled)
3. Once this is complete, CaptionHub will automatically send the original language captions directly to the LILT Platform
4. Translation occurs within LILT; once complete, translated captions are automatically sent back to the CaptionHub video project
5. Translations can now be viewed directly against source video, with in-situ amendments possible

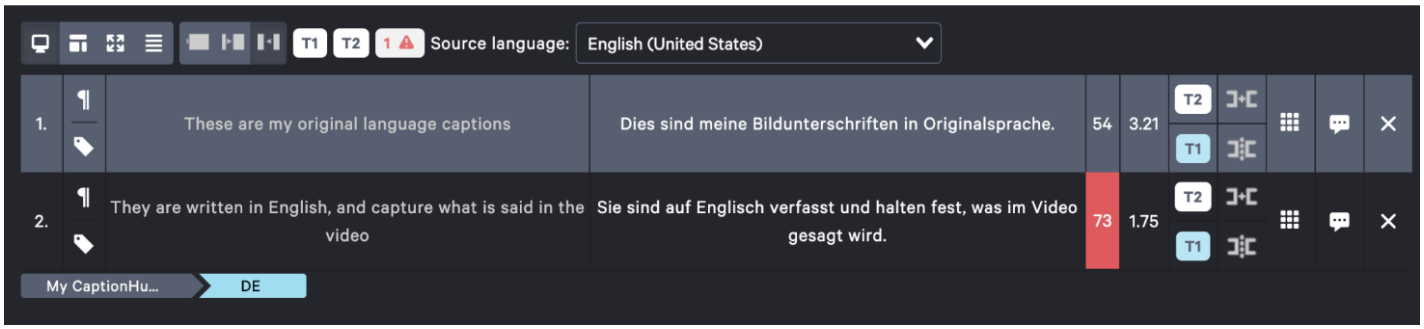


4 Formatting/Editing

Once the captions are translated in LILT with Contextual AI, they are automatically pushed back into the CaptionHub platform. At this point, the captions are ready to be reviewed and amended as preferred.

Within CaptionHub, each project can have specified caption limits, including:

- Minimum caption duration: how long captions must appear on screen
- Maximum caption duration: the length of time captions can remain on screen
- Maximum caption length: the number of characters captions have to remain within
- Maximum line count: how many lines can appear on screen at any given time
- Minimum frames between captions: any required gap between any two captions
- Maximum reading speed: a limit to the number of words per minute / characters per second that can appear on screen



In the example above, the original English captions are an appropriate caption length, but the German translation exceeds the character count limit and is flagged with a red color.

Some languages - particularly those that are faster/slower than the source language, or require more/fewer characters - may need to be adjusted in order to fit within the caption limits. Let's look at an example. Consider the following English caption:

Sorry, I need to leave; my train departs the station in 15 minutes and I don't wish to be late. Let's do this again soon!

Translating this sentence to German, in which words are up to 35% longer on average than in English:

Entschuldigung, ich muss gehen; Mein Zug fährt in 15 Minuten vom Bahnhof ab und ich möchte nicht zu spät kommen. Lasst uns das bald noch einmal machen. Auf Wiedersehen!

At a glance, it's clear that the same caption requires a higher number of characters (168 vs. 122) and words (29 vs. 25) in German than it does in English. Not only are the character and word counts higher, but the caption even stretches over 3 lines, which breaks the caption rules for this project. This can be enough to surpass the caption limits, particularly if the original English caption is already close.

Fortunately, CaptionHub has been designed with this consideration in mind. The source video is displayed alongside the captions, so that linguists are able to edit the captions in-situ. Any adjustments that need to be made can take the context of the video file into consideration, rather than amending the translations and hoping it's still relevant and matches the video footage.

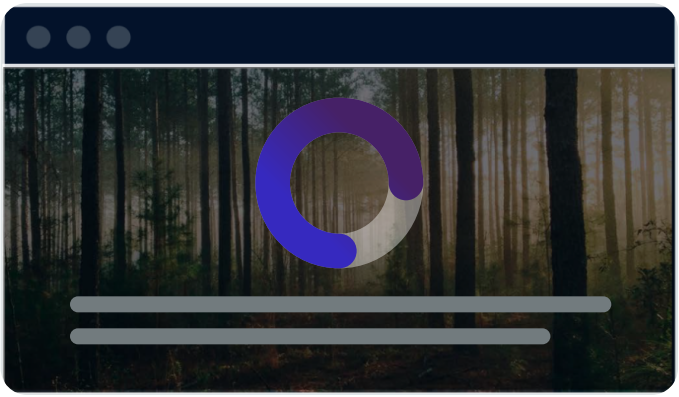
Once all captions have been checked, and all languages adhere to the caption limits of the project, it's time to publish.

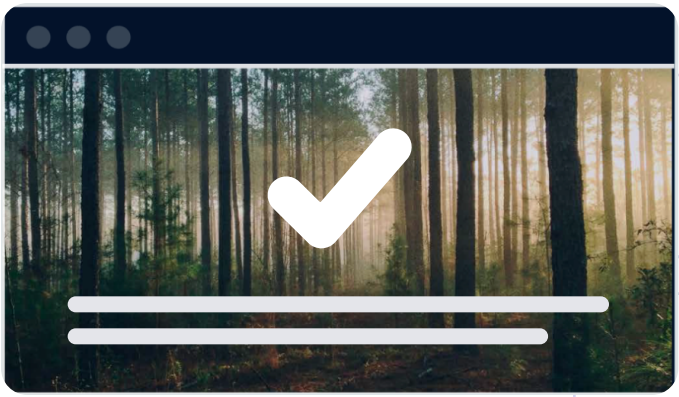
5 Exporting/Rendering/Publishing

Once translations adhere to your project specifications and have been approved, your videos with localized captions are ready to publish. As with importing, there are a few options for exporting finished captions, including:

1. **Downloading the subtitles**: A range of formats are available, including VTT, SRT, TXT, PDF, and more - useful if translations need to be imported into another software platform
2. **Rendering the video file**: Hardcoding the subtitles (also known as burning) into the video for exporting
3. **Creating a synthetic voice file**: Using speech-to-text integrations to generate an audio file of the captions- useful if there's a requirement to translate just the audio track
4. **Replacing the original audio track**: Utilizing the synthetic audio track in place of the original audio within the video file - to ensure always-on translated audio
5. **Publishing the subtitle track**: Exporting to an online video platform, ready for use wherever the video is streamed to - particularly useful for large volumes of video, or if you use the OVP for distribution

No matter which method, the localized project is now ready for distribution.





Spotlight: Localization with Human Verification

There are several strategies enterprises can employ to translate their subtitling content. Many dynamic teams choose to pursue an AI-powered solution to maximize efficiencies.

Today, many companies claim to offer AI technology, as “AI” cements its place as the latest buzzword across industries. In this environment, it can be difficult to cut through the noise and discern what a true AI translation solution is and offers. When deployed effectively, AI-powered translation result in less time spent on manual tasks, better customer experiences, and more consistent messaging across multiple channels for content teams.

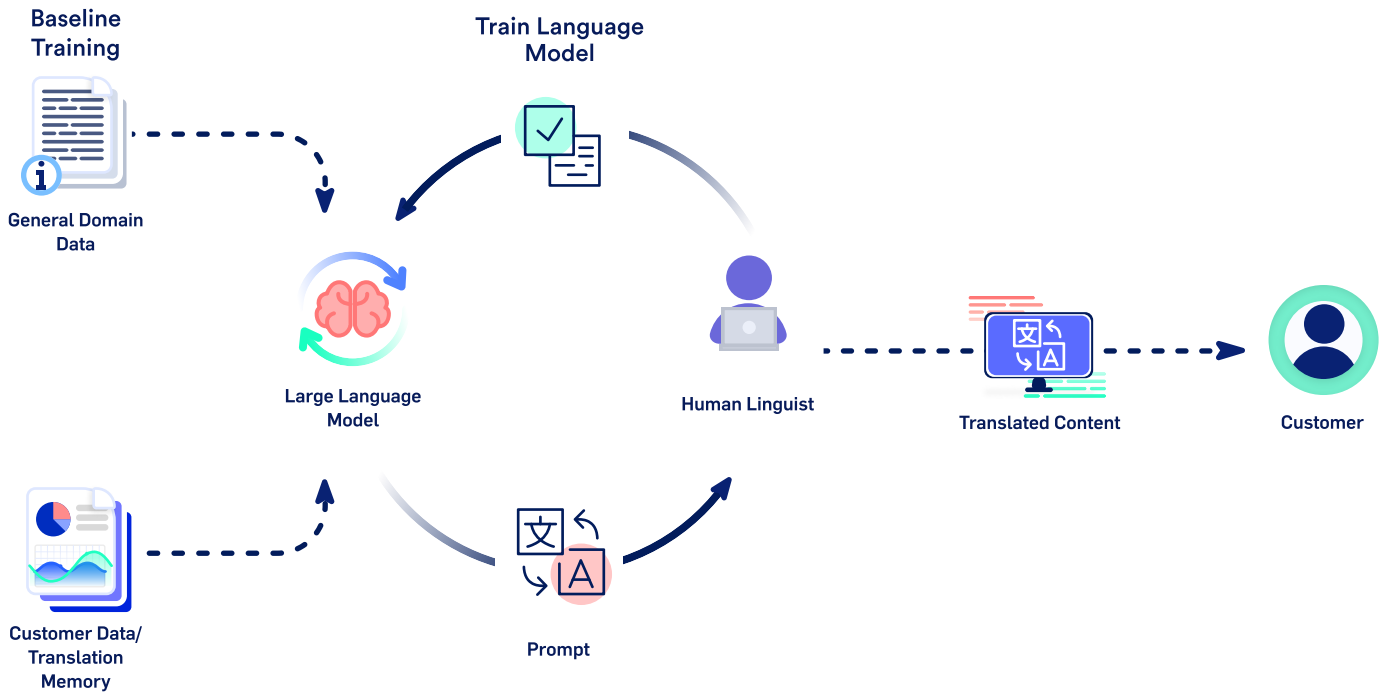
Whether working through a backlog of video content without multilingual subtitles or working to meet tight timelines for an upcoming marketing campaign, employing an AI translation strategy can ensure delivery at scale with high velocity. However, content creators need to be sure that they are not sacrificing their message and the impact of their content in pursuit of scale. To strike the right balance, employing an AI-powered strategy with human verification ensures quality-guaranteed translation delivered quickly and at scale.

Lilt Contextual AI Engine and In-Context Learning

The foundational technology behind LILT’s AI-powered translation solutions is the Contextual AI Engine, a set of domain-specific large language models instantly re-trained by linguist feedback. These large language models are purpose-built for translation and fine-tuned to ensure high accuracy and alignment with each customer’s brand voice.

LILT maintains separate contextualized models for each customer and content type (e.g., legal, product descriptions, creative marketing). Baseline models are trained on business-specific linguistic data provided by each customer (e.g., translation memories) and then these customer-specific models are then deployed to deliver suggested machine translation prompts to translators working in LILT Translate, the AI-enabled computer-aided translation (CAT) tool in the LILT Platform.

To deliver Verified Translation, skilled linguists interact with translation suggestions and deliver a quality-guaranteed translation to the customer via the LILT-CaptionHub Connector. This translator feedback also powers continued adaptation, as LILT's Contextual AI Engine learns in real time. The result is a continuous loop in which the human always has the final say on the translation, but receives increasingly better suggestions from the predictive machine translations.



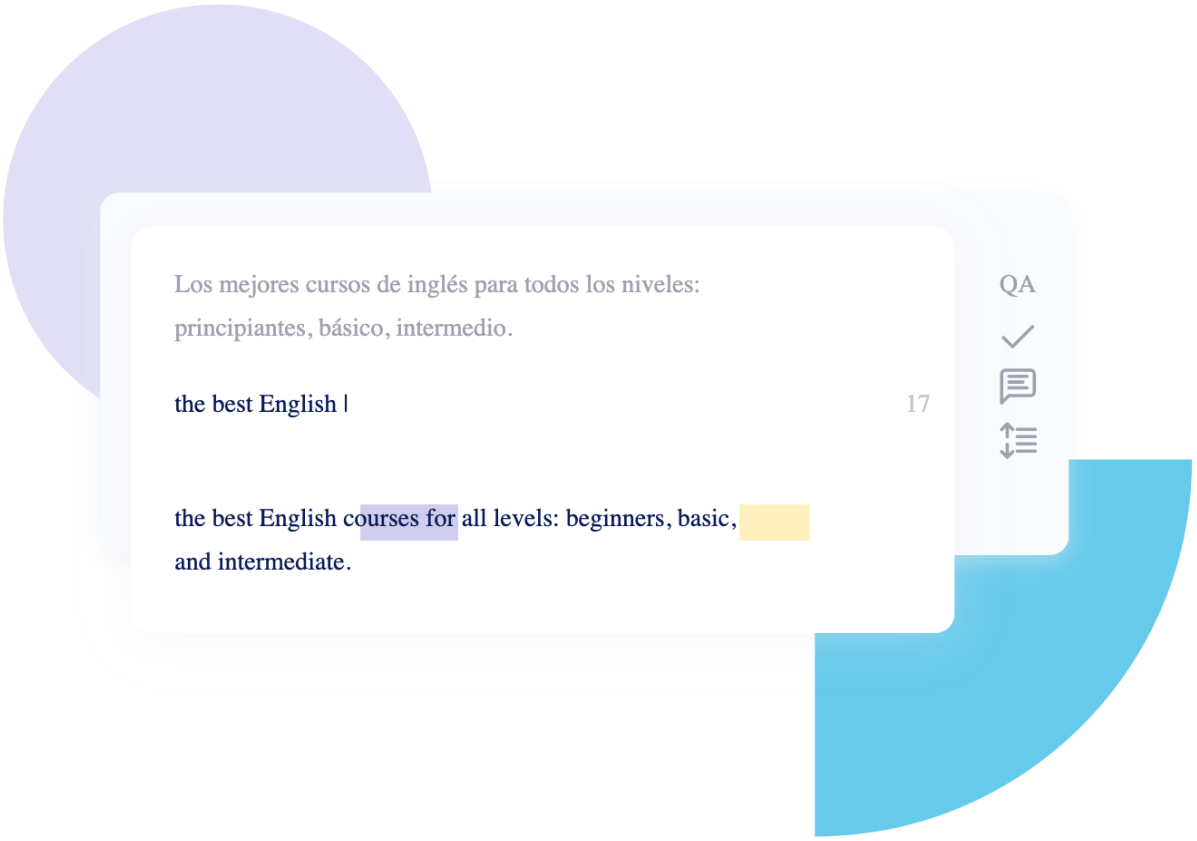
This is known as In-Context Learning (ICL), a state-of-the-art approach for achieving accuracy and efficiency in translation that has become the industry standard. With ICL, a machine translation model is trained on each new single example of the content type to be translated. This allows the model to be customized in real time to the specific content type without the need for a large dataset of training data and eliminates the need for manual retraining cycles.

What does this mean for enterprise content teams? Over time, the engine learns each organization's unique voice and frequently used terminology and phrases. Translation is delivered with brand specificity and higher accuracy. The continued real-time learning of the Contextual AI Engine enables translation at greater velocity, meaning that enterprise teams can translate rapidly to reach scale within budget. Replacing manual model retraining with ICL means that teams see immediate efficiency gains, rather than waiting for the next scheduled static model retraining.

Prediction

LILT's AI enables skilled linguists to work more efficiently for faster delivery due to the predictive nature of the technology. LILT's Contextual AI Engine analyzes existing translation data and makes predictive translation suggestions within LILT Translate. As translators make changes, the suggestions update in real time, dramatically increasing translator accuracy and efficiency. This allows translators to spend their time on new or particularly challenging translation work, while the system handles predictable and repetitive parts automatically.

As a result, linguists working in LILT Translate see substantial efficiency gains from one sentence to the next, as instant retraining powers improved suggestions across the rest of the document as well as all future projects for each specific organization.

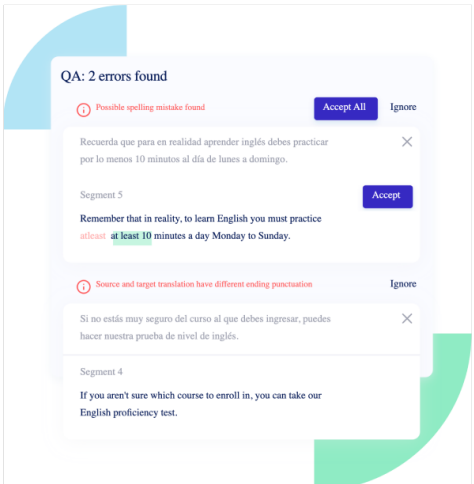


After the initial translation of subtitles is complete. LILT offers several review features in the workflow to ensure top quality and accuracy. One such feature is Auto Review.

Auto Review

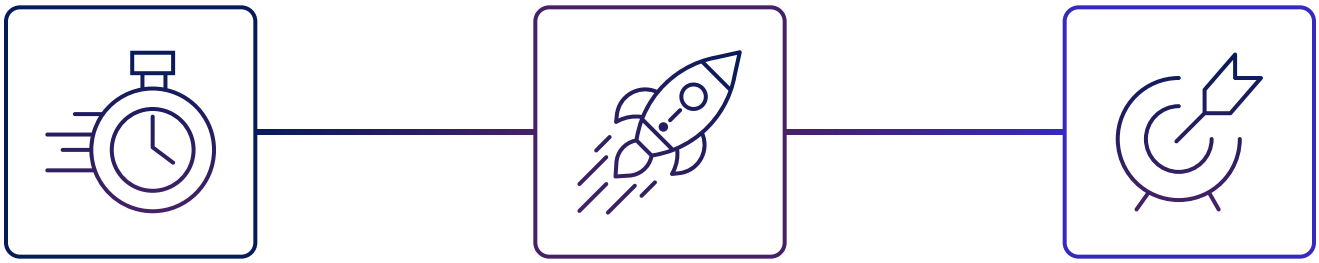
Auto Review is a tool that automates a series of standard QA checks to flag errors and warnings for translators. Auto Review catches any misspellings, leading and trailing whitespace, missing and inserted numbers, and doubled words, punctuation, or whitespace.

Ultimately, Auto Review is designed to make proofreading translated content more efficient and accurate for translators to ensure that the highest quality translation is delivered to content creators.



Optimizing Your Subtitling Workflow

When choosing the optimal subtitling workflow, there are several factors to consider. Even though there are many ways to assess and measure the efficiency of subtitling workflows, the three most impactful areas to consider are Speed, Scalability, and Accuracy.



Speed

Ability to localize media at higher speeds offers meaningful enterprise value. For certain types of content, such as videos for events, company announcements, or product launches, it is essential that teams can create multilingual video content within quick turnaround times. Even when there is not a tight timeline, having a high-velocity workflow enables teams to save time, freeing these skilled professionals to focus on tasks where they can provide the highest value.

With LILT and CaptionHub's AI-powered approach, manual and redundant steps are eliminated from team workflow through the streamlined Connector. LILT's Contextual AI Platform enables skilled linguists to translate at a higher velocity and thus teams can produce their multilingual video content substantially faster.

Scalability

A scalable workflow enables increases in volume without sacrificing quality or speed. In other words, what may work for one video, or isolated cases, may not work to localize 10, 50, or even 500 videos. A solution like the LILT-CaptionHub Connector allows for the standardization of workflows, leading to efficient scaling.

In comparison, a legacy workflow involves exporting and uploading original captions, translating, exporting translations, and re-importing into CaptionHub. While sufficient for a small number of videos, over time these process requirements quickly add up with larger volume, placing increased pressure on teams' time and budget.

As previously mentioned, the LILT-CaptionHub Connector streamlines team workflows. It enables translators to work at higher speeds and deliver at a more competitive price per word, enabling enterprise scale.

Accuracy

Accuracy is often the most important consideration for enterprise content teams. Depending on the use case, the accuracy requirements for translation will differ. For example, what may be acceptable for internal communications will often not suit large-scale product announcements or keynote presentations. LILT's Contextual AI Engine is able to deliver high accuracy, brand-specific translation at scale due to its self-learning nature. Optimizing subtitling workflows to focus on accuracy is key, as teams grow their multimedia localization processes and presence.

Conclusion

As businesses develop video content in increasing volumes, mastery of video subtitling becomes a foundational capability and requirement. There are important factors to consider to ensure that teams are able to meet the needs of their projects, timelines, and budgets.

Implementing a streamlined, efficient, and scalable workflow like the one provided by the LILT-CaptionHub Connector can differentiate video content from competitor offerings, creating a dynamic and highly engaging experience for global audiences.