



VoiceBox's Guide to Live Captioning Services



10 minute read time





Executive Summary

More than 18 million adults in the UK are deaf, have hearing loss, or have tinnitus.

At VoiceBox, we are on a mission to make content accessible to everyone, everywhere. As part of that journey, we want to help event organisers, audio visual engineers, producers and more take steps towards making their event more accessible.

While there has been progress over recent years, there is still work to be done!

Whether audience members are:

- Deaf
- Hard of hearing
- Non-native speakers
- Neurodivergent
- Dyslexic

They deserve to enjoy the full experience of an event.



One of the many ways events can be made more accessible is by using live captioning, as it enables a wide range of attendees to fully comprehend what is being said or sung if they can't hear it, or if they perhaps speak another language or prefer to read captions.

Nearly three in four people said that captions made a 'positive difference' to their experience (live events or to watch content), [data from Stagertext found](#).

Also, 35% said they made the experience "a lot better."

Some other data from [Caption Awareness Week](#) outlines:

- Almost half of the people surveyed say the number of captioned performances and events was not enough.
- More than three in four people (77%) say they are in favour of venues offering more captioned performances.
- A third of the public would be "more likely to increase their attendance at live shows" if more captioning were offered by live venues.



That being said, live captions aren't only for deaf people or those with hearing loss; other audience members may enjoy the captions if it's noisy or just for preference to follow the text on there.

But there is a bigger picture at play here beyond simply doing the right thing. Brands and events can benefit financially and reputationally here.

Firstly, your business opens itself up to [the Purple Pound](#) (spending power of disabled households in the UK). This amounts to £274 billion per year.

Investing in accessibility can provide a healthy return on investment (ROI), too. A [2022 report by Forrester](#) outlined that investing in accessibility and user improvements can return up to \$100 for every \$1 spent.

Secondly, creating a better experience will bring reputational benefits. This reputational benefit can travel via word of mouth. For example, the [2024 Euan's Guide Access Survey](#) found that 47% of respondents are "very likely" to share their experiences of disabled access with others.



On the flip side of this, the [Acquiai survey report in 2024](#) outlined that if a disabled person has accessibility issues while using digital platforms or services offered by brands, they will talk about it too.

Almost one-third (31%) would discuss the issue with family and friends.

One in five respondents (20%) would share their experience on social media, so brand reputation could be impacted by word of mouth and social platforms.

Word soon spreads in the disabled community about which events are accessible, and which ones aren't.

In this whitepaper, we will cover both types of live captions with the two different workflows.

We will also detail exactly how we provide live captions and the difference (or lack of) between in-person and remote projects.



Introduction

For the avoidance of doubt: live captions are a live text feed of what is being spoken or said at a live event, festival, or other occasion. They go on a screen or a projector to the side of or in front of the event. Live captions can also be displayed via a personal device and accessed via a QR code.

To be clear, live captions (also known as real-time subtitles, or live subtitling) are different from standard subtitles. They are live and produced in the moment, whereas subtitles are used for pre-recorded material and can be completed any time after the recording.

There are both human and AI workflows for live captions, which we will expand on further down.

Live captions are not scripted and they can't be edited. They appear between two and three seconds after the words are spoken, depending on how fast the speaker talks.

Therefore, the accuracy rates of human live captions (98%) are different to the accuracy rate of subtitles (99.99%).



Where are live captions used?

Live events:

- Conferences
- [Annual General Meetings](#)
- Live-streamed discussions/panels
- Music shows and festivals
- Comedy shows
- Theatre/performing arts
- Awards ceremonies
- Local council meetings

Broadcast and media:

- Live TV news
- Sports broadcasts
- Political debates and election coverage
- Live emergency announcements / breaking news

Education:

- Live-streamed sessions at university
- Live lectures



Section 1: Deciding what type of live captions you want

- Open live captions
- Closed live captions

And then deciding on what workflow you want to use:

- Human live captions
- AI live captions

Open vs closed live captions

Open live captions are a type of live captions that are burnt into a video. They are displayed at the bottom of the screen in the middle. The streamer (the person who pushes/sends the stream) can place the captions anywhere they like, but it's standard that they are at the bottom.



The streamer can decide on the font and colours, and then they push the feed to its destination with those caption settings.

You can choose as many lines of text as you want, but two or three is standard.

On the other hand, closed live captions are an extra layer of captions that are added to the video feed. They are an overlay on the video, and can be turned on and off.

For closed captions, we can only push two to three lines of captions max.

We decide (depending on whether the client chooses two or three lines), and viewers are stuck with the number of lines we push.

Flexibility in terms of position, font, text size etc., is subject to the destination video player (e.g. YouTube) and the settings of the platform.



What is best for you?

Whether a client chooses open or closed live captions is dependent on their end goal.

For example, when we deliver closed captions, clients think that some consumers will need captions and others won't. Therefore, they are giving audience members the option of having them on or off.

However, if they want to enforce live captions for their audience, they will push for open captions as they can't be turned off and will be on the screen permanently.

“On paper, we could deliver both at the same time, but it would have to be via two different streams; one with open live captions and one with closed live captions. It's complex,” said Mohammed Khallof, Senior Project Manager at VoiceBox.

“Therefore, clients choose from one of the two options, and around 90%-95% of our live captioning projects are delivered as open live captions.”



“Closed captions are more complex because it depends on the destination, also known as the player (e.g. Vimeo or YouTube). These destinations have built-in closed captions (CC) buttons that we can deliver for.

“If the destination doesn’t have the closed captions capability and button, even if we deliver the closed captions, it will not show the captions.

“Whereas with open captions, it doesn’t matter if the destination supports captions or not because they are burnt in, and these projects are easier to deliver.

“But of course, it depends on the client's needs, and we work to that.

“Sometimes as well, we find that some audiences are not aware of the CC button and they may not click it and miss the captions, so open captions can ensure the audience sees the captions.”



Open vs Closed Live Captions

Service	What is it?	Position	Extras	Why?
Open live captions	A type of live captions that are burnt into a video	They can go anywhere the streamer likes, but standard is at the bottom	Font, colours, lines of text etc. can be chosen too	For when clients want to enforce live captions as they can't be turned off
Closed live captions	An extra layer of captions that are added to the video feed. They are an overlay on the video	Flexibility in terms of position and font etc. is subject to the destination and the settings of the platform	Two or three lines of text max	For when clients think some will use captions and some won't, so they give them the choice



Human vs AI live captions

Both human and AI live caption workflows can be delivered as either open or closed caption.

Human live captions are real-time captions completed by professional humans. They may be stenographers or respeakers.

We mainly work with trained stenographers, who will use a steno machine (a shorthand keyboard), as opposed to respeakers who repeat spoken words into a respeaking device.

Stenography is more accurate and faster than respeakers. And the use of respeakers has higher risks, such as if a microphone stops working.

AI live captions are completed by an AI engine using speech-to-text technology. Essentially, the AI detects the microphone and, as a result, 'hears' the event. The engine detects the speech and converts it to text, and so the captions are processed and delivered.

AI captions can be delivered via a URL, and the client can take the captions and embed them however they wish.



Which is better for the client?

As usual, this depends entirely on your preferences and needs.

AI live captions are cheaper and more accessible to smaller teams with smaller budgets.

However, as a result of the speed and being driven by AI, accuracy rates can vary between 70–85%.

This largely depends on the accent, audio quality and whether a glossary has been supplied.

We will also give honest advice about the accuracy of our captions and can offer a free consultation to explore captioning in more detail.

Whereas, with human live captions, the quality is better, and accuracy can be up to 98%.

Given the live nature of events, captions are never 100% perfect.



Which is better for the client?

As usual, this depends entirely on your preferences and needs.

AI live captions are cheaper and more accessible to smaller teams with smaller budgets.

However, as a result of the speed and being driven by AI, accuracy rates can vary between 70–85%. This largely depends on the accent, audio quality and whether a glossary has been supplied.

We will also give honest advice about the accuracy of our captions and can offer a free consultation to explore captioning in more detail.

Whereas, with human live captions, the quality is better, and accuracy can be up to 98%.

Given the live nature of events, captions are never 100% perfect.



Human vs AI Live Captions

Service	By who?	Accuracy	Use cases	Why?
Human live captions	Completed by a skilled stenographer	98%	Corporate meetings & conferences, public events & webinars, education & training, government & public sector, entertainment	For topics where accuracy and trust are essential. Also, to be professional
AI live captions	Completed by our AI engine	70% - 85%, depending on accent and audio quality	Conferences & large-scale events, education & training, community and public events	For smaller budgets or when you're hosting lots of events and need to keep costs down



The choice as to which is best suited to your business or event is based on a few factors.

Industry

If you are hosting a medical conference, for example, accuracy is crucial in the terminology so human live captions are best suited.

However, if it's a more laid-back AGM where exact accuracy isn't vital and the focus is more on providing a good general understanding of ongoings, then if budget doesn't allow for human, AI live captions are better suited here.

Here is a list of some industries where we believe AI or human live captions would be best:

- Medical: Human
- Legal: Human
- Formal AGM: human
- Relaxed AGM: AI
- Finance: Human
- Education and training: Either, but AI suits at scale
- Community and public events: AI



If you do choose either human or AI live captions, it would be useful to provide a glossary of key terms and spellings so that the quality is as accurate as possible, so that the stenographer or AI system will be able to understand the terms more easily.

Your values

If professionalism and trust are part of your core values, human live captions would be better suited to you.

“We had both tried and attended other online events that utilised live captioning that was done by automated systems, and it was often unreliable and didn’t make sense,” said Olivia James, former Community Cinema Coordinator, at Black Country Touring.

“It was important for us, both out of respect for our audiences and also for our professionalism and dedication to inclusion, that the captions we provided were accurate and not distracting.”

If they’re not, and it’s more so speed and scale you’re looking for, AI live captions sound like they could be a better fit.



Section 2: How does VoiceBox deliver live captions?

This is the exact process that clients and potential clients may find useful to know about for human on-site live captions.

I spoke with our Senior Project Manager, Mohammed Khallof, to get all the correct information.

1. When the client reaches out, they need to give us all the information they can about their event.

This includes the date, the time, the number of sessions, and the run of show. This makes it possible for us to find live captioners available for all the sessions to make it all happen.

2. We assign the client a specific project manager.

At VoiceBox, we always allocate one project manager to take care of every event from start to finish.



That includes attending the rehearsals, being on a WhatsApp group to answer any questions, organising with the captioner and the client, receiving the preparation material from the client and sending it to the captioner — we take care of everything. The project manager will also be on the same live stream.

3. The project manager will organise a rehearsal.

This session or call will be with the client's tech team to agree on how they want to deliver the event, how they want to receive the captions, how we receive the audio feed and similar information.

During this rehearsal, we may be joined by an audiovisual technical team. Some clients do, some don't – it depends on how big the event is.

4. We need to receive the audiovisual feed of the client's stream.

The audiovisual team will take the audio and video from the mics and cameras to the mixer, and then they will send that feed to us using any platform they like.

We would suggest Teams, Google Meet or Zoom.



5. We join the event feed

Once we receive the feed, we can join the live stream with the captioner.

The captioner will listen and type in real-time, and the client will receive the captions on an HTML web page that they can embed on their screen or stream. This depends on the clients' requirements, which will be discussed in the rehearsal.

If, for some reason, that doesn't work, we can provide alternative solutions.

We use a speech-to-text tool for the captions that the client will receive. If they choose open live captions, this is fully customisable. The client can change:

- Font
- Colours
- Sizes
- And more!

The event unfolds with live captions in real-time with our full, dedicated support.



6. We support post-event

After the event, if the client wants the transcript, we can copy the script from our server, put it in a Word document and send it to the client.

It's important to note that these aren't transcripts with time codes. It's simply a copy of the live captions from real-time. We can also support with subtitling the live stream recording if the client wants.

For example, when [we helped the University of Liverpool](#) with its live conference for the National Postdoc Conference 2021, we supplied more than just a transcript.

As Dr Saneeya Qureshi, Head of Researcher Development and Culture at the University of Liverpool, said: “With VoiceBox, it wasn't just about providing a transcript. It was about supplying a platform where people who needed accessible features, such as those with dyslexia or English as a second language, could choose from different options.

“For example, slowing down caption speed or changing the size or colour of fonts.”

The event was praised as the gold standard of accessibility.



Section 3: In-person or remote? – what's the difference?

There isn't much difference between on-site and remote events for [live captioning projects](#).

For on-site events, note that the stenographer (selected by in-house vendor management) must attend the event in person, but the project manager is not needed on-site unless this is specifically requested and the client has a budget for it.

Ahead of on-site events, the stenographer must be sent the feed so they can listen to it from their location in the building.

On that note, you'll need to consider where you place the stenographer in the room. They can't be too close to the speakers, or the noise will get picked up by a microphone. Additionally, if they are too close to the audience, they will hear the clicking of the stenography machine.



The client will also need to ensure accommodation is available. It is up to them whether they book accommodation and supply food, or we pay for it and charge the client.

If remote live captioning is chosen, we just need the client to send us the feed so we can attend and see the stream remotely, along with the stenographer, so they can bring the captions to life or with our AI system.

During a remote event, our project manager is remote. Also, AI live captions are always remote.

When we completed a live captions project for [SAVE Britain's Heritage](#), Stephen Sheriff, Creative Producer, explained how they “were really pleased with how VoiceBox helped deliver remote live captions for our conference” and that “it was a great success.”

Ahead of the event, our team took time to explain how remote live captioning works, helping SAVE's team feel confident in the setup and delivery. Following the event, a full transcription was also provided, offering lasting value beyond the live sessions.



Conclusion

As we have outlined in this whitepaper, everybody wins with live captions. Here is a quick recap of exactly who benefits:

- Deaf and hard-of-hearing audience members who can better understand what is being said
- Non-native speakers of the language in which the event is spoken
- Neurodivergent audience members who may not be able to process visual and audio simultaneously
- Or someone who simply likes following text

And as we've mentioned, not only are you being more accessible, but there is also huge potential to appeal to the spending power of disabled households and win their trust and loyalty.

We hope you now feel better equipped to understand which kind of live captions you need, with which workflow, as well as how VoiceBox delivers live captions and more information about in-person or remote projects.



Request a free quote or call if you think we would be a good fit for your next live captioning project.

We can discuss live captioning and other ways of making your content more accessible, such as sign language interpreting, audio description, voice overs and subtitles.

