

A.I. and A.D. - OK or N/A?

Joel Snyder, PhD

Audio Description Associates, LLC

jsnyder@audiodescribe.com

Disclaimer. The presentation is not based on formal research but rather on the professional practice of AD by the presenter over four decades and an analysis of recent AI-based contributions to AD production.

Keywords: Audio Description, Artificial Intelligence.

1 Audio Description (AD)

Audio Description or AD is a literary art form. It's a type of poetry—a haiku. It provides a verbal version of the visual—the visual is made verbal, and aural (he points to his ear), and oral (he point to his mouth). Using words that are succinct, vivid, and imaginative, audio description conveys the visual image that is not fully accessible to a segment of the population—new estimates by the American Foundation for the Blind now put that number at over 31 million Americans alone who are blind or have difficulty seeing even with correction. And, of course, visual images are not often fully realized by the rest of us: sighted folks who see but who may not observe. Besides—it's great for anyone who's making a sandwich in the kitchen while the TV is on in the living room!

It's useful for anyone who wants to truly notice and appreciate a more full perspective on any visual event but it is especially helpful as an access tool for people who are blind or have low vision. Over the last four decades, I've written and voiced AD for thousands of arts events (performances—theater, dance, opera; media—feature films, nationally televised broadcasts, and streaming video; and myriad museum exhibitions) and I've also provided description at weddings, parades, rodeos, circuses, sports events, even funerals.

When I teach AD at sessions around the world, I focus a great deal on the crafting of the language used—but 99% of all AD is written to be heard. A voice talent trained in voicing techniques for AD is critical to the success of the effort to translate a visual image to the spoken word.

2 Artificial Intelligence and Audio Description

It seems as though we are on the cusp of AI dominance in all manner of endeavor. Speech synthesis is already employed by some companies who produce AD for broadcast television. I believe that the appropriate voicing of AD is done with nuance, attention to the images on screen, and an understanding of the phrasing used in the writing of the description. Indeed, in 2021, the American Council of the Blind passed a resolution noting its “full support for ... the use of human voices in the voicing of audio description for cinema and narrative video or streaming.” <https://www.acb.org/2021-resolutions#Resolution2122>

Can the spread of AI/speech synthesis be stopped or forestalled? Should it be? I decided to pose the question to a popular form of AI—Chat GPT. One might wonder how Chat GPT will evaluate its sister app’s usefulness.

I asked ChatGPT (Generative Pre-Trained Transformer): “Why is TTS inappropriate for audio description?” Its response reveals that AI apps are not necessarily supportive of each other! Read Chat GPT’s response:

“While TTS technology can be helpful for converting written text into spoken words, it is generally considered inappropriate for audio description due to a few key reasons:

‘TTS lacks nuance: TTS technology is not capable of providing the level of detail and nuance required for effective audio description. For example, TTS may not be able to convey the emotions or tone of a character’s voice, which is essential for understanding the story and characters.’

‘TTS may not accurately describe visual elements: Audio description involves describing the visual elements of a scene or event to provide context and understanding to the listener. TTS may not accurately describe these elements, which can lead to confusion or incorrect interpretation of the scene.’

‘TTS can be distracting: TTS technology can be distracting and disrupt the flow of the audio description. This can be especially true if the TTS voice is not natural-sounding or does not match the tone of the audio description.’

‘In summary, while TTS technology can be useful in many situations, it is generally not considered appropriate for audio description due to its lack of nuance, potential inaccuracies, and potential for distraction. Instead, trained human audio describers are generally used to provide high-quality audio description for media content.’”

3 Conclusion

AI apps are inevitable and will surely bring great advances to humanity. But—at least for the foreseeable future—an experienced and *human* voice talent is key to effective AD. And AD offers tremendous opportunities for voice artists—a way to do well by doing good.