

YOUR INSIDER BUSINESS RESOURCE FOR MORE THAN 30 YEARS

# scn

SYSTEMS CONTRACTOR NEWS

FUTURE

## AI FOR AV

### Industry Experts Debate Current Concerns and Future Applications

AVNETWORK.COM

JULY 2026



20



28



34

#### INDUSTRY INSIDERS

**20 Captioning on Campus**

Bill Bennett explains how AI is changing education landscapes.

**28 AI at the Edge**

Adam Teeple shows how AI helps situational awareness.

**34 Viewpoint**

James Knight encourages Pro AV to embrace AI certification.



Multiviewer



Video Wall



Switcher



Decoder

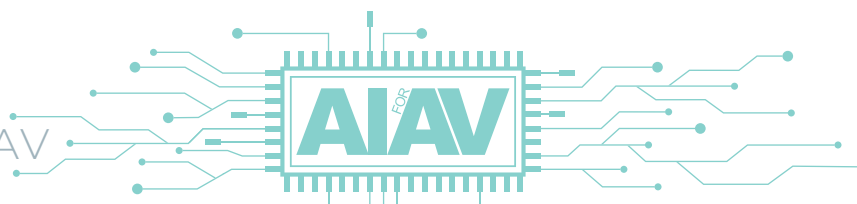


## Zio® W4008 | 1RU

The Ultimate All-in-One Video Processing Solution







# Accessibility on Campus

AI-based technologies are changing the captioning and translation landscape in education.

By Bill Bennett

Learning environments have endured a seismic shift over the past five years, as classroom and instructional models were transformed first through necessity and subsequently through new behaviors and preferences born through change. AV technology has played a substantial role along the way, from bringing physical and remote learners together through UC platforms to enhancing collaboration through streamlined connectivity and control technologies.

AI-based technologies have increasingly played a role in this changing classroom technology landscape, with beamforming microphones and camera tracking technology among the first to make a difference in our evolving education environments. Today, they have become powerful tools to help expand access to spoken word content across languages and models in real time with on-demand subsequent access just as easy (and expected) for the learners to access.

We've all become accustomed to live tools on our phones that can caption the words of the person in front of us, including real-time translations, and even re-speaking those words in a language we best understand. These capabilities (and much more) are possible today to help learning become more accessible, efficient, and comprehensible.

Besides all the "magic" AI can bring to maximize learning and understanding, there are also regulations

that have been increasing at federal, state, and local levels regarding accessibility. For example, "WCAG 2.1 Level AA" codifies the Department of Justice's final rule on ADA Title II with respect to online content for public entities, including K-12 districts and public universities. So, in many situations, captioning is required, not just helpful.

## Diversity Abounds

The community that makes up an education student body has evolved over the decades. Today, you'll find students of all ages and from diverse backgrounds as international as any United Nations event. With that comes all manner of hearing abilities and languages—but that doesn't mean any of those languages are what's being spoken in the room.

Add to that the art of multitasking by many of today's students, where they slice their time in the classroom (if they're in a room with others at all) across multiple activities and devices while lectures and classmate conversations unfold. Like me, some prefer to read transcripts rather than watch a video or listen to an audio recording; they may be a faster reader or their setting may just be more conducive to reading a transcript rather than experiencing a lecture in real time.

These are the dynamics that have propelled AI captioning or translation from advantageous to must-have. When evaluating these valuable solutions, buyers should consider aspects such as flexible pricing, ease of setup, accuracy, latency, allowance for

custom word libraries, word filtering and substitution, automated indication of speaker change, automated punctuation insertion, a documented API for automation, file-based access to transcripts and sidecar files (used for adding captions to videos), and a live data feed to get captions and translation out to other platforms and systems. That's quite a list, but they all matter when it comes to performance.

The better AI captioning systems add a grammar filter to know the right spelling (for example, "they're," "their," or "there"), with latencies of three to four seconds for native captions. An additional second or two is acceptable for translations. The field of translation is evolving rapidly and getting better at managing word-order differences between languages.

Voice fingerprinting is also coming along fast. At first, in file-based material, a system could identify someone's voice in isolation so that the AI could ID the voice. It had time to process the file and provide their name along with their captioned words. Now, it's possible to do this in real time with live captions as people speak, thanks to high-speed GPU processing.

One of the easiest ways to make spoken word content easier for folks in the classroom to absorb and retain is with a caption video display. Tech by companies such as ENCO makes this simple, with a local NUC-type PC running totally on-prem, using no internet, taking a local in-room mic feed, and displaying text captions on a video monitor of any size in real time.

Further, those captions can be sent to mobile sites or



A caption video display is one of the easiest ways to make spoken word content easier for students in the classroom to absorb and retain.



apps or other devices and recorded as text transcripts. After all, it's just data. Then, those transcripts can be married with video or audio recordings of the sessions later, making them highly searchable for students who want to go back and review targeted material again. All this can happen while allowing the instructors to upload custom word libraries, so the AI systems have a leg up on knowing about complex or uniquely spelled words and technical jargon.

This same technology is available today to support students who may not speak the native tongue. In some cases, you might need an Internet connection for live translations, but more companies are providing entire on-prem live translation capabilities as well. Users can even display both the in-room language captions and various other live transitions on the same screens or on other screens or devices in the room. (And don't forget those multi-lingual transcripts.)

Keep in mind that many students also learn remotely. Today's technology can deliver these same live captions to where they learn, while still originating from the room where the little NUC computer resides. However, those captions and translations require relay through a public web server following standard datasec/cybersecurity policies. They can also be passed onto streaming encoders to feed streaming CDNs, a method traditionally closer to the proven CEA-608/708 closed captions that broadcasters use.

### The Third Rail

Captions and transcripts are very much the "third rail" of content. When coupled with audio and video (live or recorded), they can be the gateway to identifying essential nuggets of information within hours of long recordings. Searchable transcripts can quickly return results—and with the right content management system, users can instantly link to the point in a recording where something specific was mentioned.

While smaller institutions might need an in-room, off-network portable caption cart for occasional captioning and translation needs, what if there are dozens of rooms that require captions? Not a problem—remember, it's all just data.

A central cluster of campus servers can feed classrooms around a connected campus, employing useful signal routing capabilities such as audio over IP to get in-room audio to the central servers. The same infrastructure can bring the caption text data back to the room to show on a video monitor via a tiny Raspberry Pi or similar small computer, with its web browser looking back to the proper server.

You can also bank transcripts from all the sessions, allowing students to access them later (based on user authentication controls, since not everyone has access to every class). An overall "conductor" application can oversee when sessions are routed to a caption engine where the captions and transcripts can be accessed.

The ability to caption speakers isn't limited to the lecturer up front, of course, as ceiling-mount or other additional microphones can capture student questions

and comments that inevitably take place. Some caption systems can even label the speakers, for example, as "lecturer" and "student."

When learning is paramount, the best educational environments make available the course material and dialogues in as many ways as the learner can absorb. AI

tools are there to help make this highly affordable, flexible, and deployable across a variety of situations, buildings, technical environments, and other requirements. **SCN**

Bill Bennett is the product design and solutions manager for ENCO.



## SIMPLY SCALABLE

### ALLEN & HEATH AHM AUDIO MATRIX PROCESSORS

Available at Full Compass Systems, AHM is a family of powerful Audio Matrix Processors supported by a comprehensive ecosystem of scalable I/O, control and Dante solutions. Designed to excel in corporate, retail, hospitality and education environments, AHM matrix processors allow integrators to quickly create systems that build from a single boardroom to a complete campus.



PRO AUDIO STARTS WITH PRO SUPPORT  
800-356-5844 • fullcompass.com

