CHAPTER 21.

WE CAN SIMPLY RELY ON AI FOR ALT-TEXT ✨ *AI SHOULD SUPPORT BUT NOT LEAD ACCESSIBILITY MEASURES IN WRITING*

Brett Oppegaard
University of Hawai'i at Mānoa

A decade ago, maybe, people still could have built and maintained an accessible internet for audiences of all types of sensory abilities, including for those who are DeafBlind, blind, print dyslexic, or with low-vision (Ellcessor, 2014). But writing a simple and informative "alt-text"—let alone a more in-depth audio description—for each image published online back then proved to be way too high of a burden for the humans of the Internet Explorer age, before social media. In today's exponentially more massive media environment, generative artificial intelligence (GenAI)-powered machines realistically have become our only hope for alignment with our aspirational 50-year-old laws that literally require "comparable" media access for everyone, regardless of how well they can see.

Per those laws, such as Section 508 of the Rehabilitation Act, Americans collectively want all people to have access to the information encoded into our media, regardless of which sensory system gets used to process it. But in practice, and in contrast, we also communally allowed the unregulated emergence of the World Wide Web, and years later, took the almost identical let-it-grow stance on startups such as Facebook, YouTube, Twitter, etc., during their formative stages.

Too late now to put those information systems back into their boxes. By the time a person could have read these first few paragraphs aloud, other people around the world simultaneously will have uploaded about 500 hours of YouTube video, 2.4 million Snapchat "snaps," and 66,000 Instagram images (Domo, 2022), most with no accessible description for blind people. These content creators are adding to that backlog pile, minute by minute, and people who need or want non-visual means to access what is being produced are getting pushed farther to the edges of society every second. If only each content creator would also spend a few minutes writing a description of what they were sharing

DOI: https://doi.org/1010.37514/PER-B.2026.2777.2.21

online, the accessibility problem would be resolved without need for much further intervention. But they won't, and no one is going to make them.

Instagram, for example, as of this writing, does not *require* its content creators to provide an alt-text, let alone a more in-depth audio description of any of the tens of thousands of images being posted every minute. To reach the Instagram alt-text box, where such a description could be added, the content creator needs to manually select an "Accessibility" drop-down menu on the interface and then read the small text box's two lines of instructions, which say, "Alt text describes your photos for people with visual impairments. Alt text will be automatically created for your photos or you can choose to write your own" (Instagram, n.d.). With such sparse support, it's clear that Instagram and its parent company Meta, which also owns Facebook, is more interested in developing GenAI descriptions en masse to meet baseline legal standards than in training and nurturing a massive collective of human describers. The Perkins School for the Blind estimates only about 1 percent of all images online use alt-text for accessibility purposes in a useful way (Lewis, 2023). Therefore, I get it, humans had their chance. Now GenAI can take over, as its evangelists argue, and once and for all solve this societal issue.

Yet I tried this acclaimed built-in AI service on Instagram recently, and what alt-text did it generate for me when I posted a photo of a dog? … "Photo by (Account Name) on August 10, 2024." That was it. Where could I find this mostly unhelpful alt-text? In the ordinary web interface, it was invisible to me, not integrated, nor easily accessible. With a screen reader and by inspecting the page's code, though, I could locate those words, learning about the medium, i.e., a "photo" not a "reel," and when the photo was uploaded. But anyone hearing this alt-text surely must be wondering next: What did the photo show? That encoded information apparently is not the important part, just as an alt-text of "image_63478.jpg" on any website will clear the standards of automatic accessibility checkers. What about if we demand more and ask GenAI to write these descriptions for us in a way that actually provides information? What would it write? In other words, let us call in the machines to clean up our mess, and they can show us how to do it.

Our research team's initial tests with GenAI in the role of a writer, at the dawn of its development and mass public releases, in late 2022, though, created texts that sounded like nonsensical baby babble, generating descriptions of a portrait photo of United States Vice President Kamala Harris, for example, as "32-year-old woman is looking happiness" and "woman in blue blazer drinking from a glass," when Harris, who was not anywhere near 32 at the time, was wearing a N95 face mask, preventing her from showing much "happiness," and there was not a glass of liquid in the image, either (Bergin & Oppegaard, 2024, p. 10).

Our continuing follow-up tests have been showing improvements in GenAI's descriptive abilities, and the GenAI engines we have been testing seem to be gaining "comprehension" of the writing process, as a technical practice. For example, we use the same test image from a Parks Canada brochure and run it through GenAI brains over and over, with different prompts, and different GenAI models, to gauge bias and compare outcomes. This image shows a hiker in a park setting, and our proprietary programming and prompt engineering are able to generate an automated description of the image that started with these alt-text-like sentences: "A determined hiker walks through a rugged mountainous landscape with towering cliffs in the background. The sky is partly cloudy, casting shadows and light across the dramatic scene. Text in bold letters reads, "WE ARE EPIC ADVENTURES!"

As you can read, the writing style is pedestrian and repeats like that from description to description. It leaves out many details that the viewer of the image can easily ascertain. For example, from the description provided, can you generally learn what gender, race, or age the hiker is? What is that person's body shape, positioning, posture, attire, or expression? Is the person a teenage girl? Or a male senior citizen? How is the person dressed? How does the describer create a specific sense of place? How does the AI know the hiker is "determined" or the scene is "dramatic," without providing details to show, not tell? We are tracking improvements, and the example does show improvement from the Harris examples, but those are also like advancing from toddler talk to an early elementary school level of compositional sophistication. Beyond those present concerns, though, the questions have to be asked about how a GenAI will, or even can, make a call on describing a person's gender? Or race? Or age?

Such attributes usually are relatively simple for the sighted audience member to pick up visually, and then to describe with social sensitivity to others, but as of recent tests, the GenAI makers are mostly taking a purposeful pass on describing such fundamental pieces of information that generate a social identity. Because of the political peril of making public mistakes, and the embarrassing errors made already, GenAI not only is avoiding those areas of description now, despite the technical ability to address them, but I suspect GenAI will continue to avoid them for a long time.

But that does not mean people need to stay entirely on the sidelines and just watch all of this spotty development play out. GenAI for alt-text and for the more in-depth versions of it, called audio description—as of today (give it a try at no cost at our site: www.accesshound.com)—can offer significant support systems that still leave the writing part mostly to the humans. As others in this book have suggested, GenAI can help writers in many ways beyond the "writing" part itself, including Charles Bazerman's acknowledgement, for example, that GenAI

can "provide support at various moments within the process of idea exploration and text production" (2026).

In my studies of audio-description production practices, I concur and have found GenAI to be clearly capable of, for example, breaking compositional inertia in all levels of writers, such as by helping a person to get started on a complicated passage with a wild thrust into the vocabulary of the situation. Or just by holding a mirror up to a text and showing what it's all about when interpreted by a computer mind. GenAI can provide alternative ideas, too, even strange ones, helping a writer explore compositional options and pointing out as well what not to do. GenAI can interrogate a text, asking the writer tough questions, about everything from structure to sources to syntax. At least at this point, GenAI is not a great writer or even a good writer on its own. Yet GenAI has a lot of company in that regard. Try holding all of the humans around you to the same standards, in the same scenarios, and see the competition more clearly.

My tests demonstrate that GenAI has great potential not as a writer, per se, but to serve as an audio-description writing coach of sorts. That's something valuable in a writing-adjacent form, too. At this point, the machines are not ready for audio description straight up, but I also would say that audio description is not really ready for the machines, either. Audio description is a nascent academic field, with relatively few empirical studies to guide it, and very few people in the world who do it well, as enduring models of foundational standards. Most of what people know about this type of writing is drawn from everyday practitioner knowledge, passed from describer to describer (Koirala & Oppegaard, 2022), with ever-evolving complexities growing from the increasingly nuanced societal understandings of race, gender, age, etc. (Oppegaard & Miguel, 2024). In reflection, if the humans are still figuring out how to best describe visual media, the computers aren't going to be able do that sort of reckoning for us.

GenAI systems also cannot be a writer, in a traditional sense, because writing is more than just a mechanical content-creation process. Writing requires lived human experiences that create the unique paradigm that each person brings to ideas before expressing them. In the human sense, GenAI has no backstory, or life-altering epiphanies, or even personal opinions that uniquely flavor compositional choices of approach, and style, and words, which all swirl into a distinct writerly voice. GenAI has not been blind or had low-vision or even known anyone who has that lived experience. GenAI is a mirror of our age, our computational abilities, and our societal norms, reflecting the sensibilities of its creator's algorithm. It is a support system, like a calculator serves a mathematician.

For audio description to really work, though, it relies on the intimacy of the relationship of writer to listener. The social contract between them initiates a process in which a sighted person looks at an image and then—to the best

of that person's ability—describes it to a person who cannot see the image or at least cannot see it well. As a part of this contract, that describer inherently understands that the recipient can never directly check the description's qualities against the original image. The describer also knows that, for efficiency's sake, the recipient needs to unflinchingly trust the integrity of the effort, including believing in the describer's intent to be caring, comprehensive, and factually correct.

The latest accessibility fix-all technology, GenAI, certainly can create a convincing word-filled description, that sound like audio description, in a snap. But here is where we can learn from our mistakes with technologies in the past and insist that quality has to matter going forward, and it has to be required in the solutions. Someday soon, it's likely that one of the big GenAI providers will have "described" a huge collection of images, like at a Smithsonian museum, or something of that sort of stature; and then the dominoes will all start to fall in a race to describe the entire Internet. If GenAI is leading that automated movement, without any human quality-control checks, we will end up in a world covered with inadequate description masquerading as a problem solved.

While most people would get behind the idea that having a missing alt-text for a photo is something that should be addressed, for everyone's sake, we also can and should rally around the idea of human involvement in audio description. Once this theoretically nearing GenAI tipping point is passed, it's going to be harder to rally the troops around the idea that the effortlessly added audio description isn't good enough for the masses. GenAI can provide that impression of good enough; but people need to demand better from their information systems, whether personally affected or not, including accepting only descriptions that are complete, accurate, trustworthy, contextualized, and useful. In other words, people still have an important place in audio description. At the end of the day, people need to be the audio description writers not because computers can't write as many words, and faster, but because people need to be ultimately responsible for what's written.

REFERENCES

Bazerman, C. (2026). Writing means producing written words that look like they are from educated humans: Human writers can revise AI output as they think about purpose, audience, and context. In C. Basgier, A. Mills, M. Olejnik, M. Rodak, & S. Sharma (Eds.), *Bad ideas about AI and writing: Generative practices for teaching, learning, and communication*. The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PER-B.2026.2777.2.16

Bergin, D., & Oppegaard, B. (2024). Automating media accessibility: An approach for analyzing audio description across generative artificial intelligence algorithms.

Technical Communication Quarterly, 34(2), 169–184. https://doi.org/10.1080/10572252.2024.2372771

Domo. (2022). *Data never sleeps 10.0*. https://web-assets.domo.com/miyagi/images/product/product-feature-22-data-never-sleeps-10.png

Ellcessor, E. (2014). <ALT="Textbooks">: Web accessibility myths as negotiated industrial lore. *Critical Studies in Media Communication, 31*(5), 448–463. https://doi.org/10.1080/15295036.2014.919660

Koirala, S., & Oppegaard, B. (2022). The light bulb went on: A historiography-based approach to disentangling audio description's influential U.S. roots from its common practices. *Journal of Visual Impairment & Blindness, 116*(4), 1–12. https://doi.org/10.1177/0145482X221116903

Instagram. (n.d.). *Posts*. Instagram. Retrieved August 10, 2024, from https://www.instagram.com/instagram/

Lewis, V. (2023, July). *How to write alt text and image descriptions for the visually impaired.* Perkins School for the Blind. https://www.perkins.org/resource/how-write-alt-text-and-image-descriptions-visually-impaired/

Oppegaard, B., & Miguel, A. (2024). Audio description of gender: Self-description as an evocation of identity. *Perspectives, 32*(1), 43–58. https://doi.org/10.1080/0907676X.2022.2116990