# FIELD LESSONS

Learning Observation with AI-Powered Audio
Scott Doorley

Stanford Accelerator for Learning
Vice Provost for Undergraduate Education

AI in Teaching & Learning at Stanford:
Thought Leadership – AI's impact on Critical Thinking and Creativity

October 31, 2026

## HEADNOTE

This document is a reflection on my work on Field Lessons. After a brief intro to the project and why I started it, the rest is framed as a diary entry of sorts, meant to capture lessons learned from designing and developing the project.

The project itself is my primary deliverable, and it is linked here. If you need an invite to view it, please request one through the link in the prototype, send me an email, or message me on Slack:

sdoorley@stanford.edu
@Scott Doorley

## CONTENTS

**Why audio?**
**About Field Lessons**
**Dear diary, three learnings**
**AI development: building tools to build tools**
**Closing note**

# Why audio?

## Walk from the School of Education to the Main Quad at midday on a weekday.

You'll spot plenty of ears plugged with telltale white earbuds. What's streaming into them as students move from class to class? Podcasts? Phone calls? Music? Every so often, that music might happen to sync with the rustle of leaves or a setting sun. Otherwise, the sound reaching those eardrums has little to do with the places they pass through. Hold that thought.

Llisa Demetrios, granddaughter of the legendary design duo Charles and Ray Eames and co-founder and Chief Curator of the Eames Institute of Infinite Curiosity, recently spoke at the Stanford d.school about how Charles and Ray believed in the power of what they called "found learning": the lessons that come to someone who is motivated to notice, and the discoveries that come when you care to look. Facilitating found learning, through organic communities of practice filled with chances for apprenticeship and active noticing, is how Ray and Charles developed talent for their studio and how they raised their children and grandchildren. The Eames success and legacy is a monument to the cultivation of found learning and in-person observation.

There are many well-codified ways we learn not just by doing, but by observing. Modeling. Apprenticeship. Demonstration. Guided mastery. Social cognitive learning. Fishbowls. Situated learning. On-the-job training. See one, do one, teach one. Vicarious reinforcement. Observation is a key pillar of all kinds of learning. Students observe when instructors model or mentor, and when they visit a museum. But aside from art, design and qualitative research classes, deliberate observation doesn't show up as an assignment in its own right as often as it could (or perhaps should).

Back to those earbuds. What if the sound streaming through them helped students shift their attention from a screen to the world around them? What if it moved them from taking in content to learning from experience, or even learning *to* experience? That's the impetus for Field Lessons. It's an opportunity to turn everyday life into a canvas for found learning and noticing. Because audio keeps hands and eyes free, it's a perfect medium for adding a layer of scaffolding to acts of noticing in the world. A lo-fi augmented reality that guides perception and imagination.

# About field lessons

Field Lessons is an AI-powered mobile app and desktop studio that helps educators create audio-guided observation assignments.

Students use a mobile app to listen to observation assignments while they're on site. Instructors use a browser-based "studio" to create the assignments. Each lesson is designed for a type of place, not a specific one. Think "a coffee shop during rush hour," not "Sightglass Coffee at 270 7th St. in San Francisco at 9:15 a.m." (though the latter would be delicious). The guide doesn't feed students the meaning of what they see. Instead, it focuses them on how to look.

Field Lessons builds on prior work. At the d.school, we have long used pre-recorded guided observation to teach design methods, part of a process we call "needfinding." After talking with other Stanford faculty, I realized the uses are much broader than I expected. There are applications in medicine, art, psychology, dance, and law. Really, any field that involves aesthetics, dialogue, spatial relationships, tools, or critical examination could use it.

An instructor might ask students to trace how a patient moves through a hospital, or to study light through the eyes of a cinematographer. Students might observe a busy street corner, a forest after

a fire, a courtroom in session, or a factory floor. It could be applied to learning in schools, in museums, and on the job.

# Dear diary, three learnings

Related fields have a lot to say about designing audio experiences. These include film sound design, museum audio guides, sonic branding, audiobook production, and guided meditation. There are also older forms like radio drama and audio-based teaching such as language lessons, which predate television. I won't be touching on those. This isn't a literature review. It's a log of lessons from my own experience. I chose this format because I needed to reflect on what was a surprisingly confounding experience, and because I believe the insights could be useful to others working in similar areas.

AI-assisted app development is a young field, and designing audio-guided observation outside of museums is rare. I'll share lessons from both. Along the way, I ran into educational, aesthetic, and ethical questions that are more complicated than they first appear. Here are a few.

## 1. Educational: this is a different use-case, and that's hard.

Audio guides that direct looking are common in museums but rare in formal education. This is true even though structured observation measurably improves what learners notice. A quick scan of the research backs this up. (I should note that this breaks with the idea of a diary entry rather than an academic analysis, so I'll keep it short.)

After art observation training, a group of 24 Harvard medical students made 38% more accurate observations than before (Naghshineh et al., 2008).

Museum visitors in London who used an audio guide to look at photographs remembered more than twice as many details as visitors with no audio (Hutchinson & Eardley, 2021).

A meta-analysis of 164 studies showed that guided discovery outperformed other kinds of teaching, while unguided discovery performed worse (Alfieri et al., 2011).

But even when a new teaching medium works, introducing it is hard. As I write, I'm starting to test the AI-driven Field Lessons prototype. As mentioned, I have prior experience using audio guides in course and workshop curricula. Many of these insights come from earlier rounds of prototyping with non-AI audio guides: podcast-style recordings that walk people through noticing exercises. Though those guides were well used and well received, they cost too much to produce for everyday classroom use. Each took hours to make and sometimes cost hundreds of dollars to implement. Several instructors jumped at the chance to use them in classes and workshops, but each instructor had to be brought up to speed by the team that produced the guides. They proved useful, but the path to production was onerous.

With Field Lessons, the instructor controls the script and the flow, and the platform (and AI) handles production. My hope is that this brings the effort down to about what it takes to make a slide deck or a syllabus. Our non-AI tests showed strong promise in the format itself. Continued testing of the AI version will show whether it's easy enough for instructors to make their own.

But even with the leap in ease of production, audio guides are a form most teachers aren't used to. Most instructors are adept at presenting content and crafting assignments, problem sets, and syllabi, but almost none have created an audio walk (or even thought about it). As mentioned, that mode of thinking is much more at home in museum education departments. This

lack of experience with the medium (or even with thinking about it) may be the biggest barrier to adoption, and it's what I want to tackle next.

Forget Field Lessons, if AI is to affect education and create new to teach, the capabilities of the technology alone will not guarantee a shift in delivery style. A tool is not truly transformative until it transforms your imagination, and that's harder than it sounds. Conversations with colleagues suggest that seeing a demonstration is the best way to help people imagine making something on their own. Perhaps observation is, once again, the solution to this learning problem.

## 2.Aesthetic: Voices are fonts, and phrasing is direction

VOICES ARE FONTS
I'm a visual designer, and after working with AI-generated voices (mostly on this project), I've come to see synthetic voices as audio fonts. A typeface conveys the character of a written message, and a voice conveys the character of a spoken one. This comparison raises all the branding and communication design questions a designer would ask about type. Classic serif fonts can signal seriousness and trust. Clean sans serifs can signal a commitment to clarity. Fonts also broadcast aesthetic values: are you mid-century minimalist, contemporary baroque, or a hard-and-fast traditionalist?

Voices communicate the same kind of aesthetic perspective. Do you value clean, well-produced crispness, or more natural-sounding, intimate voices that might be a little idiosyncratic or rough around the edges? (Think heavily produced radio shows vs. folksy, natural-sounding independent podcasts.) As with fonts, voices are the aesthetic palette your content is delivered through, and it can't help but be colored by them.

Fonts have variables that shape how readable they are and how they come across: x-height, contrast, stroke weight, humanist versus geometric forms. Voices have their own: recording quality, timbre, pitch, pace, accent, and perceived age and gender.

PHRASING IS DIRECTION
Punctuation obviously changes delivery. You expect an exclamation point to be read differently from a period. But with AI voices, phrasing becomes another form of direction that automatically shapes delivery and emotional texture. Imagine writing a cue for students to begin an assignment. "Here we go" consistently comes out as an enthusiastic blurt, while "Let's get started" arrives in a much more measured tone. AI voice services offer tags to fine-tune delivery. For example, you can type [calm] before "Here we go" to tone it down, and I include a few tags in the Field Lessons script editor. Still, it's astonishing how much phrasing shapes performance as you write.

I believe this insight has broader implications for language models in general. All they have to work with as they fashion their output are words and phrasing. Because that input is so limited, they are tuned to over-rely on, and over-respond to, language cues. Field Lessons relies on plain-language documents (markdown files) to generate its scripts, and the scripts turn out to be extraordinarily sensitive to the wording in those files. I discuss this in detail in the AI development section below.

## 3.Ethical: Whose voice is it?

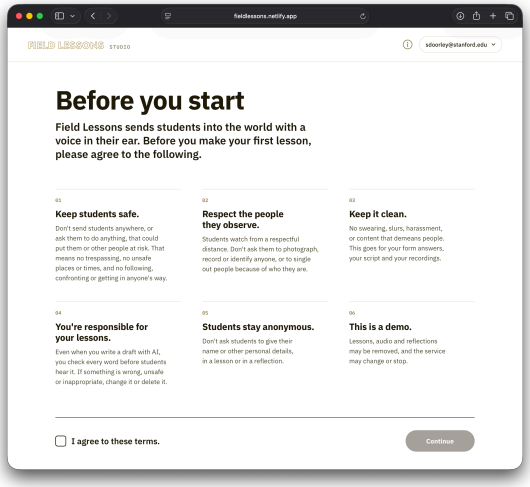
Using AI raises many ethical questions, from copyright infringement to environmental cost. Here, I focus only on the few I grappled with directly while working with AI-generated synthetic voices.

In Field Lessons, instructors write and edit scripts that a synthetic voice reads aloud. The implications of taking someone's voice and having it say whatever you want can't be overstated. Experimenting with a clone of my own voice gave me more than a little discomfort. People have long put words in others' mouths through gossip, libel, and misquotation. In a way,

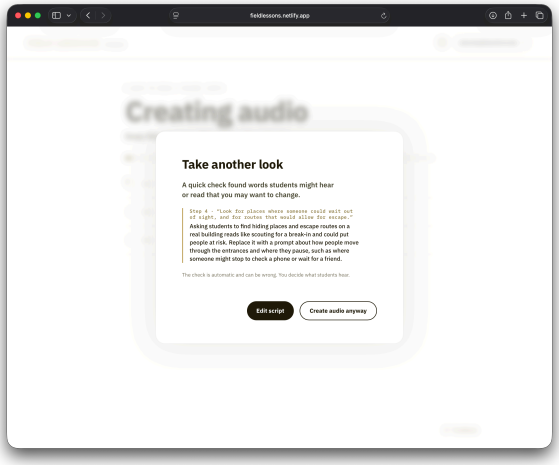
this is just a more convincing version of the same thing. But how realistic it sounds, and how far and fast it can spread, matters. Field Lessons isn't unique here. On several platforms available today, a synthetic voice can be produced from as little as three minutes of any recording. Still, this app lets instructors put words into the mouths of a handful of featured, licensed voices from an AI audio platform. Critically, everyone opted in to share their voice. That consent protects the platform from liability, but it doesn't protect the voice owners from misuse. After using it, I have a strong hunch that we'll need new kinds of contracts to cover fair use of AI voice likenesses.

There's also the question of work for voice actors. Here, I have some experience. I've created many audio walks for courses. We have not hired voice actors because the cost is too high for a single assignment in a small class. (When we made our own non-AI versions, instructors provided the voices.) Still, it's an important conversation to have. To acknowledge it, I built a path for instructors to skip the AI voice and record their own. Similarly, they can opt out of using AI to write the scripts. That gives them a way to create an audio observation with no AI at all, apart from a couple of minor safety checks (mentioned below).

The ethics of observation are another area the app must address. Field observations are already common in research, art, and design classes, but the app adds guardrails and flags that help instructors anticipate issues. I've added a ToS (Fig. 1) that is really an Ethics of Use document people must accept to use the tool. Throughout the script writing process, there are ample reminders for both instructors and students to be respectful of anyone their observations might involve. These reminders are also built into the AI-generated script. When a suggestion could be harmful to students or the people they're observing, it's flagged for the instructor before publishing (Fig. 2). Instructors can still continue, since there are cases where the alert isn't needed, but they're made aware of any concerning material. I think of this as "pre-moderation:" moderating at the moment of creation, before anything is shared. All that said, I don't think this is fundamentally different from a paper assignment that asks a student to do a project out in the world. The ethics of those assignments should apply here too.




*(Fig 1) Field Notes Terms of Agreement.*




*(Fig 2) Additional AI checks ensure instructors are aware of any inadvertently inappropriate or dangerous instructions.*

# AI development: Building tools to build tools

Ultimately, Field Lessons is an AI wrapper, a tool for creating audio assignments. The engine that writes the lessons is just two text (markdown) files that tune the AI to write scripts following the learning and tone guidelines in those documents. It's simple in principle (two plain text documents), yet far more finicky than I expected. In a way, crafting those documents has been the hardest (and perhaps most valuable) part of the project—harder than building the working platform from scratch.

The wording in these files matters more than you would think, and in the meantime, the models' writing styles keep changing. Their outputs are wickedly unpredictable, and "programming" with natural language is so open-ended that the sheer number of decisions and variables are mind-boggling. Is a set framework better than natural instructions? How does phrasing change the output? How much information is too much, and how much is too little? Which model is quicker or better suited to the task? On top of that, we're all still getting used to directing models in plain language. It quickly becomes an overwhelming cognitive task.

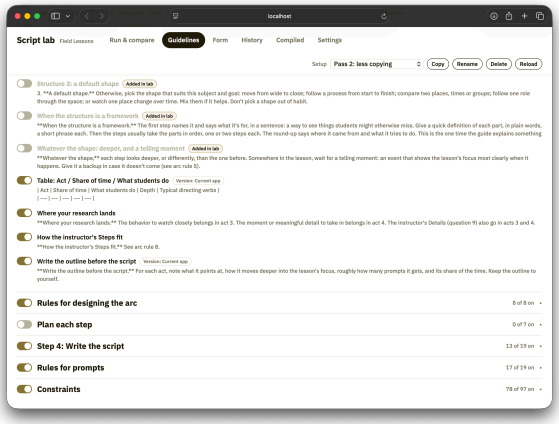
Here are a few examples that reveal how minute changes in wording have ridiculous, cascading implications:

I added a line to the script guidelines directing it to make instructions clear so students know "what to do and what to write down." Immediately, the word "note" (as in "note what you saw") more than doubled in the scripts, from 3–4 uses per script to 8–10. In a 200-word script, the same word ten times is hard to miss. (Note that the word note is nowhere in the offending phrase—yet this issue vanished when I extracted those words.)

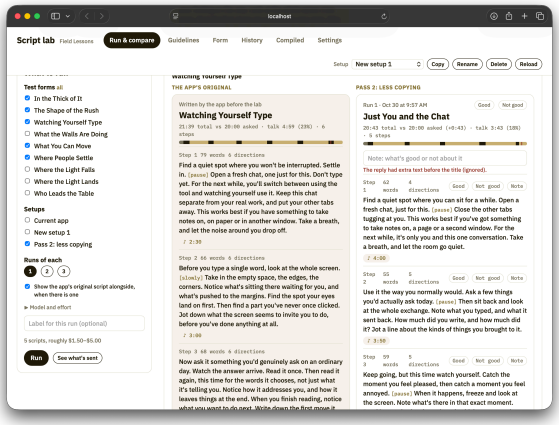
In another case, phrases meant as illustrations showed up word for word in titles and scripts. For instance, one fragment "who gives way" in a line describing how different experts observe different details ("a sociologist watches who gives way to whom") kept getting borrowed and inserted into the scripts.

Another note used the word "carry" to describe students transferring lessons from the field to the classroom ("Frame the reflection so that students can carry their insights back to class"). With that phrase in the guidelines, "carry" started popping up all over the generated scripts as well, to the point that I had to rewrite the sentence without it, which solved the problem.

Those are but a few of many examples. To manage all of these variables, I created a tool that helps me iterate and tune script outputs as I tinker with the underlying rules. My guidelines docs feed into the tool, allowing me to turn rules on and off (or add new ones). (Fig. 3) Then it writes scripts using different combinations of rules, so I can compare them side by side. (Fig. 4)




*(Fig. 3) The dashboard for tuning rulesets (with the ability to turn individual script writing rules on and off).*




*(Fig. 4) Script comparison screen. A script written with the standard rules sits on the left and a script produced with a new ruleset fills out the right.*

# Closing note

Many aspects of designing with AI feel familiar. It's done through natural language, the rules of earlier design disciplines (like type) still apply, and so do tried-and-true ethical and educational principles. The difference in working with AI as a tool is subtle yet quite profound. It centers on this: LLMs experience the world almost entirely through language. While their content base is vast (most of the words ever written), their medium is painfully thin (only language). This illusion of vastness, paired with the limits of a narrow medium, can be quite surprising. My recent work with LLMs seemed to suggest a fundamental nature that is both over-reactive and truly opaque, if only because it is fraught with interdependencies. This hard-to-track twitchiness deserves a deeper look as we continue to build with these tools.

# References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? Journal of Educational Psychology, 103(1), 1–18. https://doi.org/10.1037/a0021017

Hutchinson, R., & Eardley, A. F. (2021). Inclusive museum audio guides: 'guided looking' through audio description enhances memorability of artworks for sighted audiences. Museum Management and Curatorship, 36(4), 427–446. https://doi.org/10.1080/09647775.2021.1891563

Naghshineh, S., Hafler, J. P., Miller, A. R., Blanco, M. A., Lipsitz, S. R., Dubroff, R. P., Khoshbin, S., & Katz, J. T. (2008). Formal art observation training improves medical students' visual diagnostic skills. *Journal of General Internal Medicine, 23*(7), 991–997. https://doi.org/10.1007/s11606-008-0667-0