

Episode: AI Algorithms and Black Misrepresentation

Series: AI in the Classroom

Produced by: The Digital Futures Institute, Teachers College, Columbia University

Featuring: [Minne Atairu, Ed.D., Art & Art Education Department](#)

If you're a Black person, it's easy to decipher that the algorithm is basically misrepresenting what hair looks like. And for folks who are much more, describe themselves as AI engineers, they kind of understand what's going on on the backend and why it's choosing to represent Black figures in a certain way. So sometimes it's art and sometimes it's evidence of something that's not available in the data set.

[series graphic and brief futuristic music interlude]

So I've started working with generative AI in 2020. This was during the lockdown when I was thinking about my research and my work, but I didn't have materials to work with. And so AI at the time was the easiest tool at my disposal. It was available on my laptop and I could, you know, explore different things with it.

But then last year I was very interested when Midjourney was released to the public. I was very interested in how Midjourney would portray Black people. And so my first, one of my first prompts was: Generate the portraits of a Black woman wearing box braids. And so from there I started to observe how it was representing braids, whether it was mangled, whether it was plausibly like a braid that I would wear myself 'cause I wear braids. And over time I have continued to pursue that question.

I'm able to identify places where the algorithm is just failing to represent braids as it has become much more photorealistic. It also reveals what is much more problematic about it.

So there's a series that I call, Blonde Braid Studies. And so, scenario. So the idea behind the series is that we have twins. So we have twins, it's a portrait of two identical twins wearing blonde braids. But then what I have observed, especially with the algorithm, is that it's unable to generate one: identical twins. So it either generates: one twin is a darker skin tone, and then the second twin is a lighter skin tone. And so only in settings like that is it able to apply blonde braids on the figure.

I think there are many dimensions to my work, depending on who the observer or viewer is. If you're a Black Person, it's easy to decipher that the algorithm is basically misrepresenting what hair looks like. And for folks who are much more, describe themselves as AI engineers, they kind of understand what's going on on the backend and why it's choosing to represent Black figures in a certain way. So sometimes it's art, and sometimes it's evidence of something that's not available in the data set.

I have become much more aware of the tools. So I don't, when I'm working with a tool, I don't necessarily assume immediately that when I decide to, for example, write a text prompt about

blackness or Black figures or something that defines a Black person, that it'll generate that what, what I expect it to generate. And so I'm a bit apprehensive about working with the tools, but at the same time it's part of my practice and I think it's important to share it with the world.

There are researchers who are investigating how algorithms interact with questions about blackness, right? And it's not always necessarily visual in nature. There's research about how algorithms police Black people, predictive algorithms, um, leads to, you know, arrest and especially in communities of color. And so I, it intersects in many ways, but this is much more visual research, I would say, than it is anything else.

I'm also very interested in how art teachers might integrate generative AI in art lessons. So in a school setting, not in an informal space, what does it mean in terms of assessment? Organizing the class, the kinds of generative AI tools you want to use? Right. For example, I would say DALL·E 2 is much more safer for younger kids than Midjourney because with Midjourney you can easily jailbreak the algorithm, whereas DALL·E 2 has very tight guard rails. And so it's much more easier to work with that if you're working with younger kids.

So what we know, at least what the developers tell us today, is that some of these algorithms have guardrails that prevent you from generating, say, explicit content, right? But there are, there are ways in which you can jailbreak the algorithm and so bypass whatever guardrails they've put in place. But it's much more difficult to do that with DALL·E 2 because DALL·E 2 is just very tightly guard-railed.

So what I'm hoping would come out of the activity would be in terms of assessing your students, because that's important in classroom settings and in informal spaces of learning, you need to assess your students, you need to understand what you're learning about AI. That would be like the primary goal for me in terms of integration.

Yeah, creating an assessment and thinking of ways to also connect to already existing topics in the classroom. So instead of starting your lesson plan from scratch, how do you integrate it in already existing questions or themes you're already investigating in class? Say if you're investigating or thinking about cubism this week, perhaps there's a way you can interact with an algorithm to generate cubist related content.

I have received emails from art educators seeking, you know, guidance or just suggestions about how they might integrate it and other ways of using the tools. Like text-to-video, last week or two weeks ago, I had a workshop about text-to-video where I invited two artists to my class to just share their experience working with it, and also help guide students using the tool.

So in particular, when I, 'cause I teach in a school in Harlem, I'm thinking, so we've talked about box braids and braids. And so one theme that's very common in our classrooms is, like, representing humans. So if you had a classroom where you had Black kids and you had non-Black kids, and then your Black student is basically struggling to generate a portrait of someone who's wearing box braids, what does that mean in terms of assessment as a teacher?

Do you say your student doesn't quite understand how to write a prompt, or do you say that the student doesn't . . . I, I'm still thinking about these questions, right? It just means that your Black student will spend much more time generating a representation of someone that looks like them than your non-Black kids.