

2025 Digital Justice Grant Webinar -- Data Ethics: Stewardship, Privacy, and Ownership

Welcome everyone. Good afternoon. My name is Keyanah Nurse. I'm the senior program officer of IDEA programs and the program lead for the ACLS Digital Justice grants program.

So welcome to the third webinar of our fall webinar series, which is called data ethics stewardship, privacy, and ownership. If you missed the first two webinars, one of which was a [general information session](#) about the program, and the second was a [conversation with current and former digital justice grantees](#) who have also served as reviewers in prior competitions. Those recordings are available now on the [ACLS digital justice supplementary materials web page](#). And I'm going to ask my colleague Katie Reis to pop that into the chat for folks if they're interested in consulting those recordings.

So, before I introduce our very esteemed interlocutors, I want to quickly say a few words about the intention behind today's session on data ethics. Now, some of you may already know that back in 2023, we hosted a session on data ethics featuring archivist Lae'l Hughes Watkins, who is at University of Maryland, College Park, and Professor Kim Gallen, who is at Brown. It was a really great conversation that outlines some fundamental frameworks around data ethics, as well as, the kinds of issues that prospective applicants might find helpful as they craft their application materials and talk about, you know, how data shows up and how questions of ethics show up in their applications. The recording to that conversation, as I said, is still available on the ACLS website, and I recommend folks check it out since so many of the points that we covered are still relevant today.

But a lot has changed since 2023. Particularly within this current landscape, I think we all must think more critically about what it means to, our what happens to our data when we use specific platforms and tools and the extent to which the ways proprietary tech may actually run counter to the stated values and ethos of a digital project oriented towards justice. And so, the goal of this session is really to build upon the foundation that that initial session on data ethics provided, but to also sort of provide a new set of considerations and critical questions that applicants should be mindful of addressing around data ethics within their application. So, I hope you find the following conversation generative and that it inspires thoughtful questions, which we will take in the second half of today's session during the larger Q&A.

So, the first interlocutor I'd like to introduce is Zakiya Collier. Zakiya Collier is a multi-disciplinary archivist, memory worker, and educator whose work and research explore archival practices that account for the material conditions of black life and the role of cooperative thought in the sustainability of cultural memory. Zakiya is currently a 2025

Create Change Artist-in-residence with The Laundromat Project, an adjunct professor at Queens College, which is a part of the City University of New York and New York University, and the program director for Archiving the Black Web.

We're also joined today by Yasmeen Shorish, who is a professor and director of data and scholarly communications at James Madison University Libraries. Her research focuses on epistemological justice and scholarly communications, data ethics and privacy, and the relationship between information access and power and privilege. So, I'm very excited to continue conversations that I've been having with these two folks that I'm very lucky to call colleagues and be in in community with. And so, as I get started, with the screen, I'm going to ask Zakiya and Yasmeen to also introduce themselves in their own words and just provide a minute or two of context of the work that you do. Maybe some share some things that you're currently working on or questions that you're that you're thinking about to get us started with the conversation.

All right. Hi everyone. I'm Zakiya Collier. Thank you, Keyanah for this opportunity to be here today, as well as for the beautiful introduction. I look forward to discussing all things data digital ethics with you all today. As Keyanah mentioned, I am the program director for [Archiving the Black Web](#), which is a newer program. We're in our first cohort of a web archiving school program, which teaches memory workers and community-based archives, as well as, folks from Black collecting institutions how to web archive, how to do so ethically and to use it as a tool in our toolbox to preserve Black history and culture. I'm also teaching two courses this year -- Intro to Archives and Information Activism. In both courses, really thinking about how to ground students in the communities that we are part of as we think about sort of the theoretical underpinnings of the archival and library field. but also really thinking about how ourselves, our communities, and loved ones are impacted by the decisions that we make and so how can we sort of use our skills as information professionals to make changes in our communities that we want to see. Similarly with my residency this year with The Laundromat Project, using art to think about ways that we can sort of have more autonomous archival practices within our communities and homes and do so collectively. So working on a lot of different things at the intersection of archives, digital archives, and data, as well as ethics.

Thank you. Yasmeen.

Yeah. Hi echoing those appreciations for being a part of this conversation here with all of you and really looking forward to what questions you might have in the second half. As was said, I'm the director of a department in the libraries at James Madison University, which is a public institution of Virginia. Dealing with data and scholarly communications. Some of the projects that I'm working on that I think will be really germane to this conversation; I had

been part of a Melon-funded project talking about collections as data. So we're looking at how we look at library and archive collections from a computational aspect, like what sort of considerations and projects can be done when you make those collections computationally available. So ethics featured prominently in that work. And in my institution we also have uh a Melon funded project that is with the [Furious Flower Poetry Center](#), which is the first and longest standing Black poetry center at a university in the United States. And we're in partnership with them in creating a living archive, and doing that very much from a community archives ethos and have been working with great people like Lae'l and [Shift Collective](#) and others to really ground that project and serve as a model, aspirationally, serve as a model for other institutions that wish to do the same kind of work. We've built on that work as a sub award with another project called AV Annotate which is coming out of Tanya Clement's work at UT Austin to see how one could add annotations to audiovisual objects. We're working with Furious Flower Poetry Center on that as well. I'm serving um as the PI for the sub award, and it has brought up some really interesting conversations that are directly relevant to today's topic and so I look forward to getting into that and speaking from that current experience with you all.

Wonderful. Yeah, as you were both giving your own introductions, I was sort of excited about the way some of the questions that we have created together kind of directly speak to the work that you're doing. And so one of the values that I've had particularly in this, you know, format of webinar where folks are in the field doing their work, but because they've also served as reviewers in prior competition, they're able to kind of provide some insight from multiple directions on the same topic. So how would you articulate this in an application? And also, what does this look like in practice in some of the projects that you're working on or consulting on? So, uh so thank you for those introductions.

The first question just to jump in related to some of the issues that you both sort of highlighted, I want to sort of introduce the idea of this politics of refusal that actually came up in the last webinar that we had with three digital justice grantees who have served as reviewers. And so in our last session, Jakeya Caruthers, who is a 2022 digital justice seed grantee, highlighted how a politics of refusal, so intentionally deciding not to collect data or to collect as little data as possible or as is necessary for their project, informed her how that politics of refusal informed her and her team's approach to data collection. And so not everything should be collected, not everything should be discoverable or accessible given the specific content at hand as a sort of quick way of understanding that politics. And so their project which at the time was called Decriminalized Survival: an Archive of Survival is a community engaged digital archive project that was designed to illuminate a kind of submerged legacy of feminist resistance to interpersonal but also carceral state violence. And so, I wanted to start here because it's a clear example of how, you know, a funded

digital justice project took a very clear ethical stance around carcerality and surveillance and how that informed the grant's approach to data collection. And so when we're thinking about this larger conversation of data ethics, I think it's important to start at the point of collection, right, of how we're going about even collecting our data. And so in your own work or even in projects that you've consulted on, can you talk about other examples of this intersection between, you know, data collection and issues of privacy, surveillance or preservation? How has that shown up for you?

Yeah, I can take a start going back even into work I was doing as a data librarian, right? Teaching data management. Ethics is always a component of that, and data privacy as well. And so with that framing you take the approach of sort of a risk assessment and you think, at least I encourage those to think about, you know what is the relative risk of holding data across whatever the characteristics might be of your of your project right and then approach it with the what is the least amount of data that I can collect. Because generally speaking, you know, the less that you that you hold, the risk will go down then to populations or individuals. I mean, this this actually holds true across any kind of entity that you're collecting data about. It doesn't even have to be people. Although that is certainly the risk that we're quite concerned about. So there's just a generalized best practice, if you use that terminology, to kind of turn away from the idea like if I gather all the data I can gather about something, then I'll be able to do the most robust research and ask all the questions I want to ask and have the most complete picture I can have of whatever the subject is you're looking at. That's not really true and a lot of times you end up with data that you don't really know what you're going to do with. So if you think about the the various attributes of the data that you're collecting, really ethically, you must hone in on what are the fewest attributes that I really need to fulfill this outcome that I'm trying to get to, right? Tie it tightly in that way. And then you are really putting yourself in a better position to mitigate harm and to reduce risk that can occur with data that that is collected without a good purpose, potentially with poor organization because you don't really know what you're going to end up doing with it. So that that's when you're collecting data, right? That's a very in-the-moment thing. If you're looking at existing digital objects and thinking about discoverability, right, and access, that's a little bit of a, you know, a different approach, I would say.

And I'll say a project I was aware of had wanted to improve archival description of populations of Black Americans from the early 20th century primarily for whom photographic records were poorly described, you know, as far as individuals go. And they wanted to use computer vision and do some aspect of facial recognition to improve the

models at that time that were performing you know very poorly on non-white individuals, to improve that kind of performance say. Me and my colleagues, you know, read through their proposal and just felt like it was a poorly articulated kind of argument when it when it came to recognizing what the inherent risks could be to populations today. To improve archival, you know, record description with softwares that were not locally controlled, right? And so I think this is a case where the good intentions of improving a record perhaps and even in you could argue creating visibility and representation and you know access to families and so on of these records. You really do kind of have a responsibility to at least think, well, what's the worst thing that could happen from this to two populations and really problem solve that all the way through so you're at least describing it and how you would mitigate that harm.

Yes, so I would completely definitely echo everything that Yasmeen just shared. I think just to build on this point around, you know not collecting that data that you may not need, also considering if you are the best steward for that data and the best long-term steward even if you are the collector, perhaps you are not also going to be the long-term steward for it. Maybe there are ways that the community members themselves could be equipped to be stewards or there is somewhere else or some technology that can be made available to them where they are the stewards. They are the ones that manage access to the data. So thinking about ways that the participants, in a data set, are more involved in controlling what happens to that data after the collecting.

But going back to thinking about if you are the best steward, it's important to think about context and you know, many of us are coming from academic contexts where there is an abundance of digital storage technologies available that are perhaps more available or less available in different under resourced communities. However, thinking about the context and what that data will be in relationship with other collections and other materials at that particular institution or institutional archive and thinking about ways to have that data set made available in other places or activated in other places. And considering, you know, whether or not to even do the collecting based on if you are the best steward for that data. And I think in this conversation around refusal, a couple of things kind of came to mind for me. One, thinking about [Documenting the Now](#) or Doc Now, which is a project that I've been a part of which is around ethical social media documentation or archiving. Doc Now existed as a, sort of, both a place that creates tools, as well as a community leader in sort of the ethical conversations around whether or not especially with data sets like social media where people are still actively a part of the conversations and you know if the data is not anonymized or protected in particular ways then there are ways that people can be targeted from the data that is collected or otherwise harmed. So thinking about, you know, the sort of minimum threshold for collecting, do you need to collect every tweet in that

hashtag or, you know, are there particular parts of the conversation where you can sort of add additional hashtags to narrow down, the data set or are there particular individuals that you're interested in rather than the entirety of a conversation, through the affordances that different social media platforms might provide. So do you need all the comments, a few comments? Do you need just the image? Do you need the person's profile? Or do you need 10,000 profiles? The answer is probably never 10,000 profiles. But you know, it's important to sort of consider all of those things in that conversation. I think also related to the sort of politics of refusal Dr. Tanya Sutherland's work also came to mind, and I'm thinking in particular her recent book [Resurrecting the Black Body](#), which is on race in the digital afterlife, and also the Feminist Manifesto which she mentions in that text as well which really goes through a list of refusals as it relates to data collecting and data management. Because I think it is important to name perhaps where it's important for something to be collected but perhaps it removes someone's autonomy to do so. Or perhaps you know the folks that are involved in that data set will not be able to speak to it or you know sort of have decision-making in what happens to it. That's not always possible in every data set and every situation but things to consider sort of broadly in that conversation.

Thank you both also for citing different kind of resources that folks can consult after this conversation if they're interested in sort of learning more. One of the things you know when I initially brought up this idea of how the politics of refusal came up in our last conversation to Yasmeen, you had you had framed it as sort of collecting as much information or as much data as necessary but not collecting everything. And I remember sitting with the feeling that oh my god that goes against my training like as a historian where you're constantly searching for as much as possible in the archive and sometimes sort of coming up against not having enough information or not having enough data. And as I was listening to, you know, both your comments and Zakiya's comments just now, it reminded me of how sort of when we're thinking about what justice looks like, at least in the context of this program and the sort of asked to be a little bit more critical around how you're engaging with the sort of topic of inquiry at hand at your project. I do think the projects that have been really successful, the ones that have been funded, kind of go beyond a kind of liberal notion of justice that just stops at inclusion and really interrogates what that means, right? And so I appreciate that the sort of, you know, the distance between a politics of inclusion and a politics of refusal is really that risk assessment, right, of like being included into what and for what purposes. And so it's that provides I think a really good segue into some of the comments that you just shared Zakiya about sort of thinking about the ways that certain pieces of data collection can impede on folk's autonomy right and so where does that practice of collection sort of sit especially as we're navigating different spaces of like

academic institution spaces, and then like working with community partners and digital justice projects often have to navigate what I think is that somewhat contentious space between those different sort of actors. And so, can you offer any insights or advice on how on like concretely how to do that? Like how to sort of design projects that are intentional about, you know, where ownership of project data, and outputs resides and why. and sort of you know listed some considerations in terms of whether or not you're the best steward for the data that you collect, but I'm wondering if you can share any additional thoughts on that, and Yasmeen mean as well.

Yeah. So yeah around this question of thinking about you know sort of transparency about your data outputs and making that widely available. It made me think of a collections as data project like Yasmeen mentioned that I was a part of that involved sort of turning an oral history project into linked open data, in order to sort of visualize and understand the relationships of jazz in central Brooklyn from the 1930s to the 1960s and sort of in conversation with some of the participants who were oral history participants who were still with us. Sort of one having to kind of debut this technology of like okay here is what is linked open data, this is how it works. This is how it exists in other places in our world. I'm based in in Brooklyn. So, here is New York City's open data portal where you can kind of go and see other like data sets that exist throughout the city that may or may not be relevant to you. You may or may not be included just based on how the data is collected and you know what involvement you may have. And sort of being in this like pedagogical mode initially of like I before I can ask your permission to do this, I need to make sure you feel informed about what I'm asking to do with your oral history interviews. Not transforming them in any way, in the traditional sense or changing your words, but extracting sort of the relationships from these oral history transcripts and kind of saying you know where someone says "we grew up together," I can at the bare minimum say these two people knew each other these two jazz musicians knew each other or these two jazz musicians were friends. Which you know in a larger scale we're able to build a visual model to say you know these artists knew each other. This person grew up with this person was influenced by this other person was educated by this person and so on and so forth. But beginning in a pedagogical place of making sure that you know any folks that may be participants understand what they are signing up to do. Then also sort of making sure they understand that the digital outputs are going to be in multiple places, in this case, this was a collaboration um between Weeksville Heritage Center and the Semantic Lab at Pratt Institute in their information department and wanting to make sure that they understood that this was going to be different places and on in Wiki data and you know publicly available but not saying anything outside of what was already in the oral history transcripts. I mean the sort of resulting outputs kind of live

digitally in those different spaces like on Weeksville's website, within the semantic lab, their servers, it's on wiki data. It's linked from the collections at data website and anywhere that they may have shared that information elsewhere. Sort of almost like in our main page for the project kind of creating this sort of link data or like hyperlinking of like these are all the places that you can find it. They are all connected to each other. They may present different parts of the data, but this is how they all communicate together. They're on GitHub. They're here. They're here. They're here. and making sure that there are multiple ways that people can access the data and use the data based on different skill sets and reasons for wanting to view the data. if it's just I know 'I'm included in it and I want to know what's going on' or even if I'm not planning to use it in any particular way or I want to check for, you know, I'm someone who's critical of this project and I want to analyze this data for myself. But making sure there's multiple access points and types of accessibility; considering the different types of potential audiences and users that may want to view those outputs.

Yeah, I've put a link into the [OSF site](#) where we've got like the project output for Weeksville in there under cohort one and others that I might reference in this session. I think in addition to those considerations, I also want to really underscore, you know, what Zakiya was talking about was essentially like the informed consent aspect of this. That it's important to write out, I think document, the expectations and responsibilities of the research team and anyone that's in that partnership or in that collaboration so that there's unambiguous shared understanding and to not I think it can be easy to forget what people who are outside of your immediate domain don't know about practices or institutional practices or you know research methodologies and so I think to like explicate that as much as possible is really critical to do this work, you know to do this work period, but particularly if you're applying a justice framework to it. I think that you should consider in your research plan, the time that it takes, if you are doing something in a co-creation model, right, with a community, what is that time that it takes for you to both understand one another and to co-educate one another on things? Of course, there's like obvious things like maybe terminologies, but again, like the methodologies, the practices, something that I think can happen for those of us that are, you know, doing a lot of our work in an institutional context, is a sort of inertial approach to a paternalistic over reliance on standard and methods. So having flexibility when you're working with communities in particular on a project together or you're working on a topic or a collection that that concerns a community. Have some flexibility in your approach. I'm not saying, you know, I come from a librarianship background. I have a science background. These are domains that have a lot of methods, a lot of standards. We have, you know, entire profession is

classification. I'm not I'm not saying that those are irrelevant and not important. But depending on the nature of the work that you're trying to do, you should allow for the possibility that you might need to augment them or be flexible with their application in order to center the human, right? Or to center the population. So I think that I have seen that can be an uncomfortable space. I think it's an uncomfortable space for a researcher who has a reasonable identity as an expert and has cultivated that expertise and applies that expertise to research domains, but I think that is also Keyanah you were saying you know like this is against everything I know as a historian. I think for many of us the ways that we were trained in in a domain are being really challenged by the current digital technology landscape and social technical situation of the United States and beyond honestly. So I think that we must recognize where that flexibility, that openness to okay, at least let me what would it mean to use a different way of description here? What would that really mean? what are the actual ramifications of deviating from this standard that my domain holds? You might find out if you if you think that all the way through that it's not that you know essential a change, right? it won't somehow make your project unusable or not have value to your peers and to the community. So I think those are those are things that take time. So that's my biggest thing is when you're when you're embarking on this work and you want to have that expansive view and that operate within the justice framework, you must allow for the time that it takes to interrogate the processes and the systems that we've relied on for years.

Yeah. I had two quick thoughts and then I'll, sorry to jump ahead of you, Zakiya, just to let people know that there is a recording of a session I hosted last year with Andre Brock where we talked about operationalizing digital justice projects and the biggest thing that came up was time, right, of just having to navigate different bureaucratic you know red tape and also just like the different sort of ways we move through time in academic institutions versus working with community partners and having to kind of reconcile the tension between those two things and because you know feasibility is a thing that reviewers look for in projects that is something to be really mindful of that things will just take longer than you perhaps anticipate. So, I wanted to flag that resource for folks if you want to watch that recording. Yeah.

Just something that came to mind as Yasmeeen was speaking and thinking about this sort of the ways that things will have to change at some point. There will be something that comes up; life, everything is unpredictable. I think one thing to also consider, especially as you as you're pursuing a digital justice project and naming justice as a value therefore and naming ethics as a value. I think it could be really useful to also continue down that train of thought of identifying project and individual project team members values and how those will be used to sort of make decisions when things do need to pivot. Rather than sort of having the

bureaucracy will do what it will, community does what it does. The administration does whatever it wants to do. So understanding your values and where you are willing to make changes, where you aren't, where something is actually altering the value system of your project as opposed to altering the methodology for getting there. What is bendable, what's not is important so that you don't have to try to do that work mid project and you have something to point back to as a way to make those decisions to decide on you know what types of tools you want to use and being able to say these are our values. This tool does not represent that. if we're going to go against that, we are making an active decision to do XYZ and documenting how and why that decision got made that was, you know, in opposition to the pre-identified values. I think in addition to that, I think what sparked that initially was, uh, Yasmeen, mentioning sort of this understanding of what each person is responsible for and what each person has access to and what is responsible for managing and like very clearly laying that out. I wanted to add thinking about the for how long, as well as, like contingency plans like jobs are not always permanent. Yeah, but just thinking about how long someone actually wants to be responsible for that thing. We can be committed to our projects and dedicated and love them so much and still say, "I do not want to be the steward of this data for the next 20 years of my life, so like let's figure out you know how this is a shared responsibility or you know what are the ways that it rotates the roles rotate or the responsibility rotates. who is like who's you know the direct contact for the particular storage solution or whose email address you know is attached to it," and making these as like very clear decisions that are documented and named.

Yeah, I really love that point and I was just thinking about how this is a session on data ethics and yet we're also kind of talking about these operational sort of portions of the projects because it directly kind of ties into that and so one of the you know big problems so big that there have been many commissions ACLS has had one itself on kind of thinking about the sustainability of digital projects right and so one of the things that this program particularly looks for projects that are thinking in creative but also project based ways about the issue of capacity building that's and how that's tied to sustainability. And so I very much encourage people to think about that as a kind of an ethical practice, right? Of thinking about how labor is distributed amongst your project team, thinking about success successors and succession planning, right? if the data that you're collecting you know is near and dear to you but also you know runs the risk of falling into the wrong hands for whatever reason thinking about what your project is going to look like in the future and who will be the stewards of it after you I think is a useful and important and ethical exercise.

I want to switch gears slightly. This is somewhat related to what we were just talking about because I think at least in 2025 it's impossible to talk about data ethics or privacy without talking about the elephant in the room right, which is AI. Especially as more and more folks

are both interested I think in using you know large language models and agentic AI, but also as there are more and more financial incentives to do so as kind of not only tech companies, but also universities are kind of shifting resources in that direction. This you know I think begs the question of like what happens actually to data after it's collected and uploaded to these platforms, you know that the data is used to train these models but like what does that actually mean and then what? Then the question is then what?

So, I'm wondering if you can both share one to two just considerations that folks should keep in mind with respect to how they articulate uh their sensitivity or navigation of this reality in their in their application because of the kind of rapid pace that these things develop. And also, I think the kind of competition between these large competitions in terms of turning out the latest model. It's kind of hard to get a sense of concretely how some of these things work. But I think beginning the question around these are some things you should look out for, and this is how you could talk about some of those unknowns in your application um is a good I think a good place to uh to start in terms of this this topic right yeah

I will give my best attempt for today's answer about AI. So I mentioned this program the collections as data work that we did. We created a lot of documentation, but one of those documents were the Vancouver principles for collections of data, which were just a series of principles that were community informed to help stewards mainly of digital objects, digital collections, think through different things as they conduct their stewardship. And certainly, one of the things we put in there was that these collections would be used potentially for AI harvesting and AI training. And what are the implications of that? If you have it open to the web, what are you implicitly sort of allowing to be done with them? So you know this is an it's an interesting moment and shift in in my field at least which 15 years ago was all like very jazzed on open everything you know um lots of little quips about knowledge you know freely disseminated and available to everybody creates a more democratic globe, and I think there's you know a recognition in this moment that technologies exist now where you don't have to transform a digital object into another format to make it computationally available, right, it is available as it is natively digital for these language models or different types of training data to be used in ways that you cannot anticipate. And there are a lot of conversations occurring about, you know, is the right approach here mediated access, right? Do you create something that doesn't just allow immediate access, but you have to have some kind of authentication or some kind of request to view? Possibly, right? Do you put a statement? Do you put a right statement in there that forbids it? I mean, that really is the honor system in my opinion. Like not really

going to prevent harm in that case. But it is part of that conversation I think is what we've been talking about is that you, your collaborators, and your partners should all be in conversation around the values that Zakiya was talking about around what are unintended what could be what could be unintended uses of the outputs of this research project for example what are you know get someone on your team that's really good about thinking about the worst. That's what I say that's my superpower is I look at a thing I'm like okay, here's what the worst person in the world could do with this. Now how do we design against that? Right? It's just a form of design thinking that's like worst case scenario planning. You plan for the worst, and you design your system. So that it's very difficult to achieve that. And I think depending on where you sit in your project and your view of the technology, what could be done with AI could be a worst case scenario. So, this was a situation that actually was working with Furious Flower Poetry Center on their last conference. And we have a very standard sort of release because all of the conferences happen every 10 years and they're all they've been recorded. And so that's a lovely, you know, collection over the decades to see this this evolution and growth and these wonderful speakers. But then somebody was like, "But what if someone takes my image and does like a deep fake with it?" And you know, that's the level of concern that exists, right? Like is, every is this recording? You know, like anytime you've got something recorded now, every time you're out there on the internet in some way is there's this, you know, threat that people perceive of something can be done with this against my will. And so, you know, ultimately it's like, well, we cannot guarantee that such a thing would never happen, right? What we can do is say this is how this is going to be made available. And these are a lot of conversations that we've had because I'll be very honest with you as the director of the poetry center, Lauren Alleyne and I have had many conversations where the way we were talking about this project in August of 2024, and the way we were talking about this project in February of 2025 is significantly different. and the kinds of ways we are thinking about access and the kinds of ways we're thinking about the tension between historic visibility and what is American poetry, you know, and what is easily findable and and that kind of stuff. and I'm not being hyperbolic on this, like literal individual concerns for safety, health, employment, reputational risk because these are, these are poets. These are their words. They're intrinsically a part of who they are. you know, this this is a real generative and evolving conversation that we're having. So, I really have to say how important it is that you've already done good work to talk about your values, to build that relationship so that when the unexpected perhaps for something happens, you're able to have those conversations in a space that is productive, authentic, and humane. Hope that got to it. We went from AI to humane, but that it's all a piece of the arc, I think.

Yeah, I share all of those sentiments. And I love being the worst case scenario person on any project. But we need those people who are going to yeah propose the worst case scenario but also do a deep dive into what we what may not be like visible at a first glance with these technologies that we interface with and use. And so I think I think Yasmeen I think you covered really thinking about what happens to the data that we may collect that is outside of our project scope that you know it just happens because it's available online it's open source.

I think I will speak to sort of the use with of AI within projects even though I have not myself had the pleasure or displeasure whichever it may be to um to use AI in a digital project. But I think just from generally understanding and having explored using new or different technologies and the example I'll use is Yasmeen mentioned Shift Collective earlier, which is a collective that I work with and, we are, we're at the end of a three-year research project on exploring decentralized storage technology. Through, uh for community- based archives and like what are the possibilities that can afford these organizations to be able to sort of create smaller networks, storage networks that are sort of outside of larger institutions that are independently managed and more autonomous and sovereign and reflect the values that these institutions hold. But sort of at the outset of this project, which is funded by the Filecoin Foundation for the Decentralized Web, really thinking about all of the connections that these particular technologies have in the world and some of the ways that they are impacting the world around us. And so where it may, you know, speed up a project by 3 months or it may offer a particular way to approach or analyze the data. When we maybe dig more deeply into some of the people that work at the organizations that create the tools that we may decide to use, we may realize that some of the decisions being made or some of the people that are making the tools hold oppositional values to what our projects purport to do, which purports to sort of have a justice oriented approach. And if your project for example is sort of looking at decriminalization, but a tool you're using feeds data to law enforcement, you're kind of not on the same page. Your technology doesn't align with your project values and goals. And so for us, in our project called Modeling Sustainable Futures, we had to do a year of research of really digging into the technologies themselves and how they operate, how where they get their money from, which are not fun questions to look into when you've already kind of like got your heart set on using a particular thing and then you find out all of the sort of ways that they are connected to things that contradict your values or contradict even your identity and personhood.

Yeah, having to make hard decisions in that way. But I think that is an important consideration, especially when working with marginalized or over extracted communities data, and not sort of proposing that this project will make new revelations or preserve their history in a particular way, while also putting them directly closer to something that they

are perhaps wanting to be away from. And if and if like decisions are made to continue with a particular technology where there are known sort of conflicts of values. I think needing to do the due diligence of communicating to the communities that this is a decision that's being made, and this is what is known about that. So that especially if if the data collecting involves sort of an agential process of someone, again back to informed consent making sure that people are able to consent to being a part of a data set that will then be used in a language model or a agential AI that will, yeah perhaps put them in harm's way. And so I think I'll end that statement.

And I think in addition to kind of thinking about worst case scenario as far as harm, potential harm or misuse. having a really good understanding especially in today's context of in our worst case scenarios, we should even think about things that perhaps are illegal because not everything that -- not us doing something illegal, but the ways that the data could be misused and the ways that people could be harmed are in ways that might not even exist in our current imagination. but it could be a reality and not in our imagination because we see it as like that's illegal. It would be illegal for someone to take the data and do that and it's like yes and people do illegal things all the time or people do unethical things all the time, and illegal and legal and ethical don't always mean the same thing at the same time. And so considering that, as well, when we're thinking about our worst case scenarios for either interacting with AI tools ourselves or the potential, for our data outputs to come in contact with AI tools post project.

Thank you for both sharing those considerations. You know, I definitely don't want you either of you to feel like you have to provide the answers for people because we are in the know, I think we're in a moment of all collectively trying to keep up with everything and find answers to things. And so this is also why I think with folks doing this sort of process of documentation is so critical to share out sort of like the challenges, the limitations, the kinds of conversations that you're having to sort of provide a way forward um, you know, for folks interested in in doing this work. So with that, I am going to open the floor to questions for the folks in the room, with the caveat of saying if you have any general questions about the digital justice grants program, I would ask that you email those to our inbox at digitaljustice@aclu.org. But any questions that you have related to the conversation that we've had this afternoon, related to the larger topic at hand of data ethics, you know we welcome those absolutely you can pop those into the chat or you can also raise your hand and sort of you know share your question. I'm actually going to stop the recording here.